JISTA

Journal of Intelligent Systems: Theory and Applications

SEPTEMBER 2021

ISSN: 2651-3927

VOL 4 NO 2

ARTIFICIAL INTELLIGENT > MACHINE LEARNING > MULTI-AGENTS

https://dergipark.org.tr/en/pub/jista

Journal of Intelligent Systems: Theory and Applications
2021 Volume: 4 Issue: 2

Editorial Boards

Honorary Editors

Editor-In-Chief

Editors

Language Editor

Editorial Advisory Board

Editorial Assistants

Journal of Intelligent Systems: Theory and Applications
Volume: 4 Issue: 2

Contents

Research Articles

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) (2021) 65-75
DOI: 10.38016/jista.815823

A Turkish Question Answering System Based on Deep Learning Neural Networks

Cavide Balkı Gemirter [1*], Dionysis Goularas [2]

[1]Yeditepe University, Faculty of Engineering, Department of Computer Engineering, İstanbul/Turkey,

TEB ARF Teknoloji (Türk Ekonomi Bankası)

[2]Yeditepe University, Faculty of Engineering, Department of Computer Engineering, İstanbul/Turkey

cavidebalki.gemirter@std.yeditepe.edu.tr, goularas@cse.yeditepe.edu.tr

Abstract
In the domain of Natural Language Processing (NLP), despite the progress made for some common languages, difficulties persist for many others for the completion of particular NLP tasks. In this scope, the current study aims to explore these challenges by proposing a question answering (QA) system in the Turkish language. In particular, the system will generate the best answers in terms of content and length from questions that are based on a set of documents related to the banking sector. In order to achieve this goal, the system utilizes advanced artificial intelligence algorithms and large data sets. More specifically, BERT algorithm is used for the generation of the language model, followed by a fine-tuning procedure for performing a machine reading for question answering (MRQA) task. In this work, various experiments were conducted using original and translated data sets in an effort to solve the challenges that arise from morphologically complex languages as Turkish. Finally, the system achieved a performance that overall is applicable to a wider range than any other QA system in the Turkish language. The proposed methodology is not only proper to the Turkish language, but can also be adapted to any other language for performing various NLP tasks.

Keywords: machine reading comprehension, machine reading for question answering, deep learning, BERT.

Derin Öğrenme Sinir Ağlarına Dayalı Türkçe Soru Cevaplama Sistemi

Öz
Doğal Dil İşleme (NLP) alanında, yaygın diller için kaydedilen bazı ilerlemelere rağmen, diğer dillerde belli başlı NLP görevleri için zorluklar devam etmektedir. Bu kapsamda, mevcut çalışma Türkçe dilinde bir soru cevaplama (QA) sistemi önererek bu zorluklara çözüm araştırmayı amaçlamaktadır. Sistem, bankacılık sektöründen seçilen dokümanları kullanarak, sorulan sorulara içerik ve uzunluk açısından en iyi yanıtları üretecektir. Bu amaca ulaşmak için sistem, gelişmiş yapay zeka algoritmaları ve büyük veri kümeleri kullanır. Daha spesifik olarak, dil modelinin oluşturulması için BERT algoritması kullanılmış, ardından sistemin soru cevaplama (MRQA) becerisini arttırmak için bir iyileştirme (fine-tuning) uygulanmıştır. Bu çalışmada, Türkçe gibi morfolojik açıdan karmaşık dillerden kaynaklanan zorlukları çözmek için orijinal ve İngilizce'den çevrilmiş veri setleri kullanılarak çeşitli deneyler yapılmıştır. Son olarak, sistem, genel olarak Türkçe dilinde diğer tüm QA sistemlerinden genel olarak daha yeni bir yelpazede yüksek bir performans elde etmiştir. Önerilen metodoloji sadece Türk diline özgü olmayıp aynı zamanda çeşitli NLP görevlerini yerine getirmek için başka diğer dillerde de uyarlanabilir.

Anahtar Kelimeler: makine okuma anlama, soru cevaplama için makine okuma, derin öğrenme, BERT.

1. Introduction

In recent years, novel artificial intelligence algorithms proposed solutions in problems from various domains and outperformed previous methodologies and architectures. Even if some of the algorithmic ideas were not new, the dramatic increase of available data and process power, various parallelization techniques, cloud-based processing methodologies and the use of graphic processor units (GPU) allowed developing a series of different types of neural networks that demonstrated astonishing results. In the domain of Natural Language Processing (NLP), the use of deep

* Corresponding Author.
E-mail: cavidebalki.gemirter@std.yeditepe.edu.tr

Recieved : 25 October 2020
Revision : 03 March 2021
Accepted : 19 March 2021

neural networks outperformed almost all previous types of approaches. Their ability to create language models from large amounts of data is one of the main reasons for their success. Moreover, recent models managed to generate word representations that change according to their current context, thus giving a dynamic approach and increasing the system performance in various NLP tasks. These networks are configured and tested mainly for English and some other common languages because of the available data sets.

Despite some efforts to create multilingual models, there is still a lack of available data and little experience in the way that these algorithms should be trained and utilized for different languages. In particular, Turkish language has been proven to be challenging for Natural Language Processing because it is an agglutinative language with a derivational structure and morphologically rich. Consequently, the main motivation of this study is to explore the challenges of generating a Turkish language model and performing a particular NLP task, a question answering system (QA). Additionally, another motivation of this work is to explore the difficulties that could arise and propose adequate methods and guidelines for the generation of a language model and the completion of particular tasks in a language structurally different from English.

More specifically, the proposed system will be able to give the best and shorter answer to a question related to the banking sector. In order to achieve this goal, the system will be trained from a variety of data sets. In general, this task is called "Machine Reading for Question Answering" (MRQA) and it is essential for QA systems and search engines in general. To the best of our knowledge, in the Turkish language, a performant MRQA system doesn't exist, as most of the systems follow a semantic approach.

Today, in the domain of NLP the machine learning system that outperforms the state of the art is the Bidirectional Encoder Representatives from Transformers (BERT) (Devlin et al., 2018). Its success lies in the fact that in the generated language model words have a contextual, dynamic representation rather than a fixed one, resulting in a context-sensitive language model. Moreover, as its structure is agnostic and not configured for a particular task or domain, it can be fine-tuned with little effort and perform various NLP tasks. Although BERT proposed a multi-language model, its performance in the Turkish language is not satisfactory. In this context, the purpose of this study is first to create a model of Turkish language based on BERT and then use this model in order to generate a MRQA system oriented to the banking sector. Figure 1 presents a general overview of the study.

2. Related Work

In Turkish questioning (QA) systems, most of the research is focused on improving the skills of search engines by introducing two modules: one that ameliorates the structure of a user's query with specific preprocessing steps, and another that generates a selected list of the most appropriate search results (Amasyalı and Diri, 2005; Biricik et al., 2013; Çelebi et al., 2011; Er and Cicekli, 2013).

One of the tasks of the first module is to detect the question type with the help of a predefined table. Using specialized libraries for Turkish language processing like Zemberek (Akın and Akın, 2007) or Treebank (Eryiğit and Oflazer, 2006; Oflazer et al., 2003), the module analyzes the sentence morphologically and generates the stems of the words. The module can also create simplified variations of the query or eliminate prepositions, conjunctions, stop words, and replicates the query with synonyms of terms using the thesaurus.

In general, these studies utilize rule-based approaches. Their success is limited, and most of them are suitable for factoid questions only. To the best of our knowledge, there is still no approach that utilizes neural networks for a Turkish QA system. Today, NLP domain approaches that are based on neural networks outperform all rule-based systems. Neural networks solutions with pre-trained language representations were available with ELMo (Peters et al., 2018) and Generative Pre-trained Transformer (GPT) (Radford and Salimans, 2018). ELMo presented a bidirectional architecture but it was difficult to be adapted to different tasks. On the other hand, GPT required minimal architectural changes but it was unidirectional. In 2018, Bidirectional Encoder Representations from Transformers (BERT) was published by Google. BERT managed to have a bidirectional architecture by requiring minimal architectural changes for performing various NLP tasks. Within this way, BERT and its recent variations managed to achieve remarkable results.

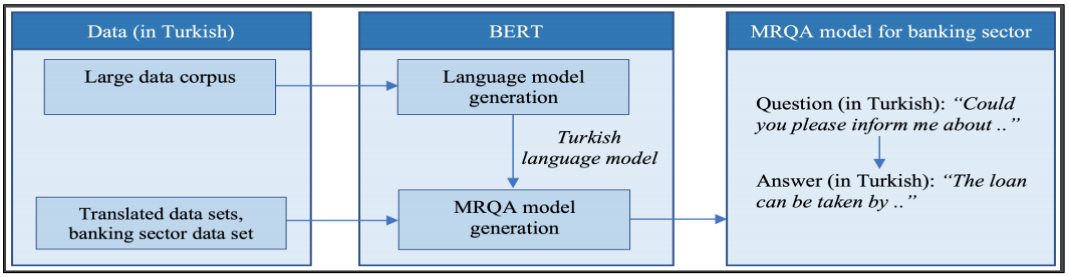


Figure 1. Diagrammatic representation of the study.

A recent competition for generalized MRQA tasks in English language (Fisch et al., 2019) included also solutions based on BERT architecture. In this contest, the presence of big multinational corporations demonstrated the increasing interest in the MRQA task worldwide. After its success in English, BERT has been implemented to other languages. Despite the fact that BERT proposes a multilingual model (mBERT), its performance is relatively low. For this reason, for a particular language a special effort has to be made for the generation of a new language model. In (Antoun et al., 2020) the authors generated a BERT model in the Arabic language named AraBERT and tested it in several NLP tasks, including QA. In tests conducted with the Arabic Reading Comprehension Dataset (ARCD), they utilized an English question answering datasets (SQuAD) translated to Arabic. When tested, AraBERT presented a similar performance to mBERT. There exists also a BERT model for the French language named CamemBERT (Martin et al., 2019), for the Korean language named KoBERT[1], and another for the Persian language named ParsBERT (Farahani et al., 2020) but these models do not report accuracy results on QA tasks. Finally, there is a model for the Chinese language trained by Google. This model was tested in the following Machine Reading Comprehension (MRC) datasets generated for Chinese: CMRC 2018, DRCD, CJRC (Cui et al., 2019).

Based on the above, it can be concluded that the existing QA systems in the Turkish language have limited success because of the approaches they utilize. Moreover, despite the progress made for systems in English language, there is still little progress for QA platforms in other languages. In the following section, the methodology for the creation of an MRQA system in the Turkish language is presented.

3. Methodology

3.1. Data sets

When a QA architecture incorporates neural networks, one of the challenges is to generate adequate data sets for the training procedure. In this work, special attention and effort were given in the choice and the generation of pertinent data sets. Two types were generated: one dedicated for a pre-training task for training the language model in Turkish and another for a fine-tuning task in order to perform a QA task for the banking domain. Table 1 presents the data sets used for training the language model. Here, the first data set is based on Wikipedia pages[2], the second on a news article collection in Turkish and the third on a corpus based on the specific domain of the final QA system prepared by the authors of this study. All sentences are in the Turkish language.

Table 2 presents the data sets utilized for the QA task or fine tuning the model. All documents were created based on the Stanford Question Answer Data Set (SQuAD) structure (Rajpurkar et al., 2016), published in 2016.

Table 1. Data sets for the fine-tuning task (QA system for the banking domain).

Name	Size	Content	Information
Wikipedia Corpus (Tr)	456.5 MB	4.5M sentences	Turkish Wikipedia dump 922335 pages (08/2019)
News Corpus (Tr)	2.5 GB	20M sentences	News articles collection in Turkish
Economy Corpus (Tr)	15.5 MB	270K sentences	Turkish economy blogs from Web

Table 2. Data sets for the fine-tuning task (QA system for the banking domain).

Name	Size	Content	Information
SQuAD (Tr)	24.42 MB	490 documents 20963 paragraphs 45872 questions 56117 answers	Q&A from paragraphs from Wikipedia articles. (Machine translation from English to Turkish)
NewsQA (Tr)	19.66 MB	8379 documents 8343 paragraphs 21270 questions 21270 answers	Q&A from articles from CNN news. (Machine translation from English to Turkish)
Banking Sector QA (Tr)	5 MB	679 documents 1637 paragraphs 17708 questions 17708 answers	Q&A from documents from the banking sector. (in Turkish)

[1] GitHub (2020). KoBERT GitHub Page [online]. Website https://github.com/SKTBrain/KoBERT [accessed 25 05 2020].

[2] Wikimedia (2020). Wikipedia Dump [online]. Website https://dumps.wikimedia.org/backup-index.html [accessed 25 05 2020].

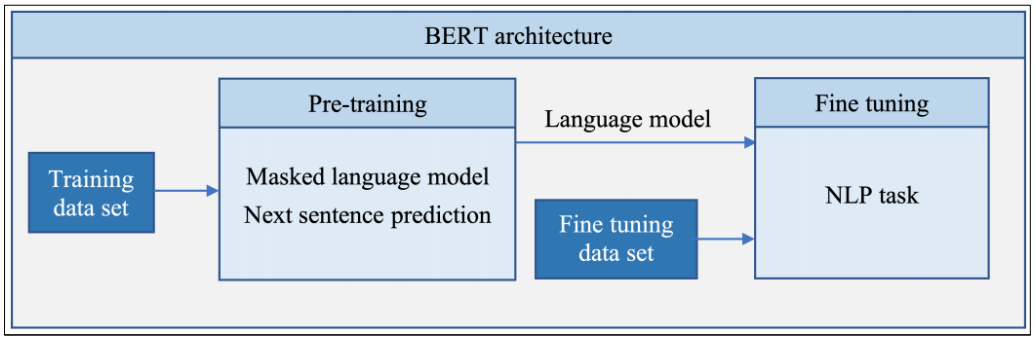


Figure 2. BERT architecture.

A SQuAD based data set includes a set of paragraphs accompanied with a set of questions and answers for each paragraph. The questions are related to the associated paragraph and the answers are generated from its text. The first data set is the original SQuAD data set translated automatically in Turkish. The second is the NewsQA data set (Trischler et al., 2016) translated also automatically in Turkish having also a similar structure with the SQuAD data set.

Finally, the third data set is created by a team working in a private Turkish bank and supervised by the authors of this study. This data set follows the SQuAD structure: a set of paragraphs with a set of related questions and answers for each paragraph.

3.2. BERT

In 2018 Google proposed the Bidirectional Encoder Representations from Transformers (BERT) neural network. First it generates a context-sensitive language model (pre-training task) and then it can perform a series of NLP tasks (fine tuning task). The language model is generated by applying two training procedures simultaneously. The first aims to predict a number of masked words from a sentence and the second aims to predict the following sentence. When the language model is generated, BERT can do a particular NLP task by using a supplementary data set. During the fine-tuning procedure, the weighs of the BERT network are slightly modified. Figure 2 presents the architecture of BERT. In this study, the BERT base model was utilized, having 110 million parameters, 12 transformer layers and 12 attention heads for each transformer layer (Vaswani et al., 2017).

3.2.1. Word Sense Disambiguation in BERT

Turkish is a morphologically rich language with a large number of suffixes and a variety of possible word positioning inside a sentence. In morphologically simpler languages such as English, POS tagging is a much more pertinent procedure. On the contrary, in agglutinating languages such as Turkish, the morphological disambiguation process is challenging. In this case, morphological disambiguation is crucial for finding the stems of the words. Otherwise, the neural network has difficulties in handling the suffixes and, as a result determining the connections between the words. BERT manages to overcome those problems by using:

- A subword--based embedding system.
- A Masked Language Model (MLM) and next sentence prediction training.
- Bidirectional transformers.

Although BERT cannot solve the morphological disambiguation problem of Turkish, the overall architecture solves the word sense disambiguation problem, as demonstrated in (Wiedemann et al., 2019) which is enough for performing a number of NLP tasks like text classification, machine translation, and question answering.

3.2.2. Subword-based embedding system

Word embedding methods became popular because they convert an input text into a numerical representation that can be used for mathematical operations in neural networks. Today, new generation word embedding methods capture the contextualized meaning even in cases with polysemy and the resulted word vector can vary according to the context. For example, the vector of the word ‘bank' is different when it is used in a sentence with a finance context and when it describes a seat in a park.

Until modern NLP solutions, morphological disambiguation was performed with rule-based solutions, such as Zemberek for Turkish. ‘Çekoslovakyalılaştıramadıklarımızdan mısınız?’ is the most unusual example in Turkish. Zemberek’s output is shown in Figure 3.

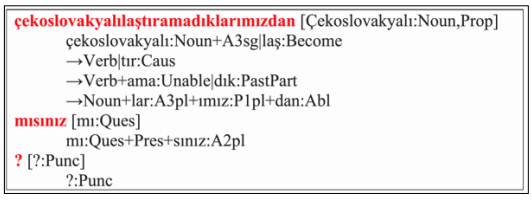


Figure 3. Morphological disambiguation result of 'Çekoslovakyalılaştıramadıklarımızdan mısınız?' using Zemberek.

WordPiece subword-based embedding (Wu et al., 2016) is a word segmentation algorithm that extracts subwords from a given data set. In the initial step, the WordPiece algorithm splits the corpus into characters, following by recursively combining them into subwords and calculating the loglikelihood of every candidate subword. After several passes over corpus, the algorithm generates a fixed-sized vocabulary by using the frequencies of combined subwords and picks the most frequent ones. The subword can be a word, a syllable, or a single character. Using the vocabulary file as a reference, most of the given text phrases can be tokenized, and the rest are marked as unknow tokens (UNK).

In summary, a desired subword vocabulary size is defined and after splitting words into characters, WordPiece generates the subwords progressively based on likelihood criteria, until a certain threshold is satisfied or the subword vocabulary size is reached. Although WordPiece is not the direct solution to the morphological disambiguation problem of the Turkish language, the result is satisfactory for the tokenization of the input sequences. The tokenization result of 'Çekoslovakyalılaştıramadıklarımızdan mısınız?' with WordPiece is presented in Figure 4. The symbols ## indicate that the sub word is a suffix. Consequently, by using these generated suffixes, the tokenizer can identify the suffixes in Turkish words and also cover many Out-Of-Vocabulary (OOV) words, thus giving an advantage over classical word embedding methods.

3.2.3. Masked Language Model (MLM) and Next Sentence Prediction training (NSP)

The training operation is based on two unsupervised tasks that are executed simultaneously: The Masked Language Model (MLM) and the next sentence prediction. The main logic behind the MLM is to try to learn the relationships of words in a language by randomly masking some tokens in the corpus and then try to predict the original ones with an attention bidirectional transformer network that handles the context both from left and right. After tokenizing the content with WordPiece tokenizer, the token embeddings are combined with positional embedding location for preventing long-distance mappings with unnecessary tokens in self-attentions. After Encoder & Decoder self-attention stacks, a Softmax classifier compares the predicted and original words and updates the weights of the network which builds the language model. In MLM, the system attempts to predict 15% of tokens in a sentence that are chosen randomly. During this procedure these tokens are replaced 80% by the token [MASK], 10% by a random word and 10% by the original word. In training, only 1.5% of all tokens are replaced with random words. Compared to traditional language models, as a result of masking process although the training time takes longer, the success of the results is satisfactory. Figure 5 shows examples of token replacements for the three cases.

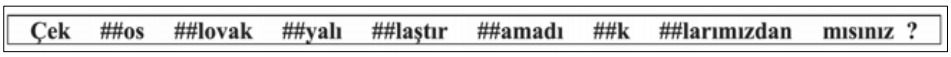


Figure 4. Tokenization result of 'Çekoslovakyalılaştıramadıklarımızdan mısınız?' using WordPiece.

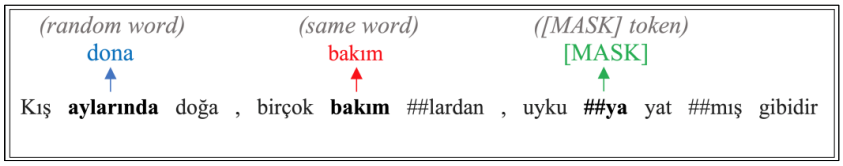


Figure 5. Token replacement examples: random word (blue), same word (red) and [MASK] token (green).

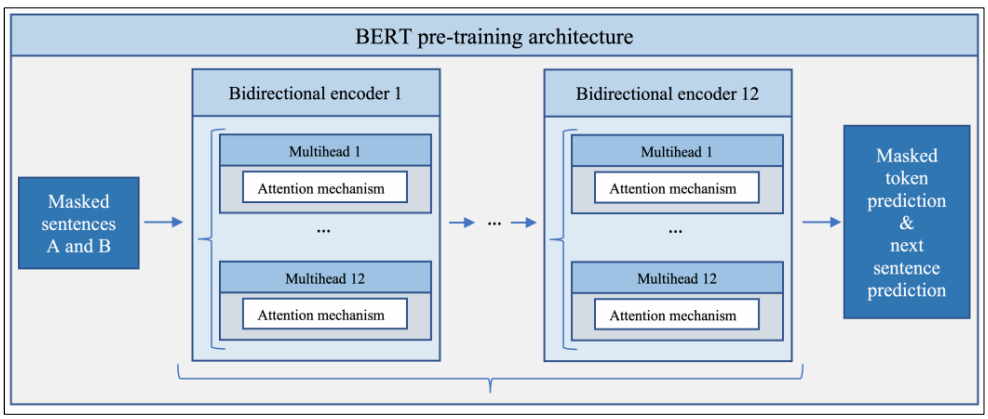


Figure 6. BERT pre-training architecture.

The percentage values are experimental. The reason of adding random words or keeping the original word is to reduce the consequences by the fact that the token [MASK] will not be present in the fine-tuning procedure. Next sentence prediction is a binarized task. In this task, the next sentence is replaced 50% by a random sentence. This procedure aims to predict the relation between sentences by Sigmoid classification and enlarge the contextual meaning that exists in isolated sentences.

3.2.4. Bidirectional transformers

Figure 6 presents the pre-training architecture of BERT. The input consists of unlabeled sentences pairs with masked tokens and the output layer predicts the masked tokens and the next sentence. BERT is using a stack of bidirectional transformer encoders and decoders (Figure 7). The transformer network has encoder and decoders stacks that contains self-attention mechanisms and Feed-Forward networks. This multi-layer architecture aims to generate word vectors that will be adaptable to the context of a sentence. In order to achieve this goal, the encoder architecture applies a word vector encoding that comprises three steps: the first is based on WordPiece encoding, the second on the position of a word in the sentence and the third on a mechanism of comparison of sentence words between them. This mechanism is described as *attention* because a word has an *attention* to the other words of its sentence.

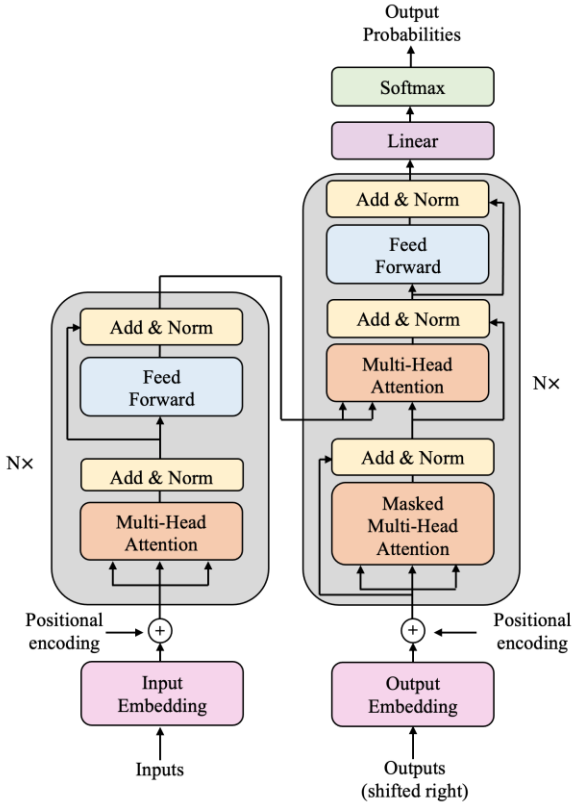


Figure 7. The transformer architecture (Vaswani et al., 2017).

Unlike a single context final state provided by a RNN model, the attention-based model has multiple hidden states, which considers all dependencies between every word in the input sequence. Although the connections are usually with previous or next words in the lower layers of the transformers, the semantic relations become more visible in the higher-level layers. In a traditional Encoder & Decoder architecture, if the input sequence is long, after a while. the model begins to forget some parts of the context Attention tries to solve this problem by focusing on finding the most critical input sections when summarizing the sequence. In an attention-based Encoder & Decoder architecture, there are weight matrices that keep the semantical relation densities of the words and highlight the significant parts of the sequence resulting in performance improvement of Decoder. In the transformer network, the output of every encoder is the input of the next encoder, like a chain. The output of the last encoder is the input of all the decoders. After converting the decoder's output to a logit vector at the top of the decoder stack, the Softmax layer calculates the probabilities, and the transformer picks the top-rated candidate. In summary, the *attention* is based on vector similarity measurements between vectors that are generated from the input vectors and a number of weight matrices.

The attention mechanism is applied many times with a multi-head approach and the resulted vectors are concatenated. The attention heads evaluate the same input of a given layer from different viewpoints. Every multi-head also includes a normalization process and an additional dense neural network. As the formation of a word vector takes into account the other words of the sentence, the system is able to generate context-sensitive word vectors.

3.2.5. Training procedure

The training procedures comprise the generation of a language model in the Turkish Language, a preprocessing of fine-tuning data sets and a fine-tuning procedure for training the QA system for the banking domain.

Pre-training for language model generation: The first step was to create a vocabulary of around 32.000 words. Among them around 30.000 was created from words existing in the data sets of Table 1 and another 2.000 words belonging to the finance sector were selected and added by the authors. WordPiece was utilized for token embedding. In BERT architecture additional embeddings are added based on the position of a word in a sentence and the sentence number. Two training procedures were done simultaneously: one for predicting the next sentence and another that is based on attempting to predict masked words in a sentence. During the training process, the three data sets of Table 1 were used together. Finally, a language model in the Turkish language is generated.

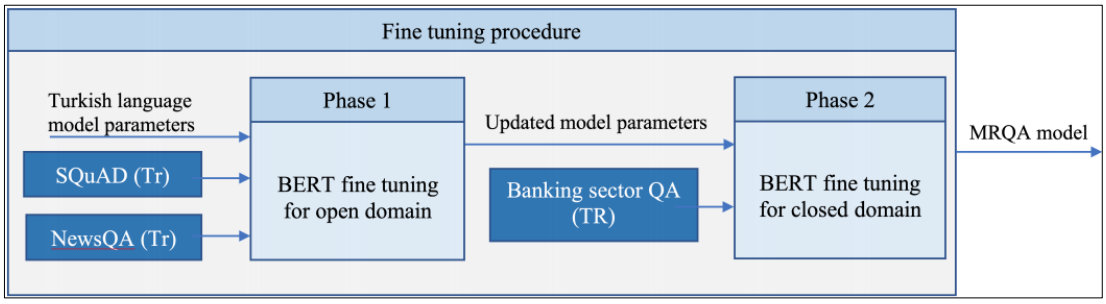


Figure 8. The two phases of the fine-tuning procedure.

Preprocessing of fine-tuning data sets: The automatic translation of data set documents in Table 2 resulted in some inconsistencies related to the SQuAD format. A pre-processing operation was applied aiming to fix problems related to an incorrect answer in a given paragraph or the absence of a start point in a sentence. The preprocessing required the development of some special procedures and in some cases a manual intervention.

Fine tuning for QA task: Fine tuning is applied in two phases. In the first phase, the fine tuning of the neural network which is already trained for a Turkish language model is done by using the SQuAD (Tr) and NewsQA (Tr) data sets. In this phase, the two data sets are combined together. The aim is to increase the QA skill of the system in general, or in other words, in an open domain. In the second phase, the model is trained with the updated parameters resulted from the previous fine tuning, by using the Banking Sector QA (Tr) data set. The second phase aims to increase the ability of the system to answer questions from a closed domain, the banking sector. Figure 8 shows an overview of the training phases together with the data sets they utilize.

3.3. System parameters

In the Pre-training task the following parameters are important to configure:

Maximum sequence length: Configures the maximum length of a sequence after the WordPiece tokenization. High values are necessary to learn positional embeddings in long sequences.

Maximum predictions per sequence and masked LM probability: when multiplied between them they define together the number of masked tokens in a sentence.

Do lower case and do whole word mask: Convert the sentences to lowercase and mask the tokens of a whole word instead of masking individual tokens, respectively. When these parameters are applied, the accuracy in general increases, especially for Asian languages. In the current study the vocabulary size is large enough to have most of the suffixed words in Turkish language. Consequently, the selection of an individual token masking or whole word masking doesn't alternate significantly the accuracy results.

Finally, for the fine-tuning task the important parameters to configure are the following ones:

Document stride parameter: This parameter allows to create training sentences examples that overlap between them. Within this way the next sentence training example will start in a given position in the previous sentence. This superposition is performed in a token level.

Maximum query length: The maximum number of tokens of a question. If a question is longer this number the rest will be truncated.

Maximum answer length parameter: The maximum length of an answer taken from the paragraph than belongs the related question. This parameter is character-based rather than token-based as it is the rest of the parameters.

3.4. Evaluation metrics

Similar to other machine reading comprehension and SQuAD studies, the exact match (EM) and F-Score will be used as the evaluation metrics. EM takes in account the predicted answer only if it is the same as the real answer and F-Score counts the predicted answers that they have an overlap with the real one.

4. Results

In this section, the results of the study are presented. In the beginning, the parameters that are achieving the best performance for answering questions for a specific domain, the banking sector are showed, followed by the errors types in errors and answers. Then, an evaluation of the study's model is performed as follows: first, the current model is compared with existing Turkish QA systems. Then its performance is evaluated in comparison with other BERT language models. Finally, the model is tested with other models in the Turkish language generated with BERT.

4.1. Neural network parameters evaluation

In order to better evaluate the accuracy of the system, different training parameters were tested. After experimenting with different values and considering the average and maximum length of paragraphs, questions, and answers in the data sets, it was observed that the following parameters are giving the best results for fine

tuning: maximum sequence length = 512, document stride = 256, maximum question length = 64, and maximum answer length = 64.

From the results it can be seen that a long sequence length (512) with a half overlapping (stride set to 256) has an important impact on the accuracy results.

Table 3 presents different combinations and accuracy results in terms of EM and F-Score.

4.2. Error types in questions & answers

In order to evaluate the success of the system in a real-world environment, the team that prepared the Banking Sector QA data set was asked to think in general rather than focusing on a particular question type or formulation. The examination of the results, revealed that the system responds correctly to the majority of the factoid questions, where the answer to the question is a single fact. Since there is a distinct difference between the exact match (EM) and F-Score metrics, a 3.2% of wrong answers which have zero EM (572 out of 17708 answers) was identified. Table 4 presents the type of errors and Table 5 gives examples for most of these errors. The errors were categorized based on the following reasons:

1) Multiple possible answers: Questions that have multiple possible answers cause 30% of the errors, as seen in the first row of the table, both the answer and the prediction are logically correct replies to the question.

2) Questions requiring interpretation: For some answers, it is necessary to interpret the entire paragraph. For the moment, deep neural networks don't have this capability.

3) Conditional answer: In some questions, the answer varies according to the circumstances, as seen in the third example, the answer is 'possible' only for the customers of a brand.

4) Questions requiring a list of elements: These questions require an answer, which is a set of elements, or a list.

5) Answers with syntax variations: Since Turkish is an agglutinating language, generating one unique correct answer in terms of syntax is a difficult task. Hence, sometimes a correct answer given by the system can be enlarged or reduced when compared to the predicted answer, thus leading to a zero EM.

6) Incorrect question: There are also incorrectly prepared questions and answers in the data set. Some of them have many spelling mistakes, some of them are logically incorrect.

7) Incorrect answers: 25% of the questions are not answered correctly, generally for the type of questions that the system has not during the training phase or for very long questions and answers which are truncated during training.

4.3. Comparison with other Turkish QA systems

In this study the model of the current study was compared with Turkish QA systems in open and closed domain. In open domain, the system should be able to answer generic questions and in a closed domain, the system answers questions from a specific domain.

Table 3. Different parameter combinations and results for Banking Sector QA (Tr) data set.

Maximum Sequence Length (token)	Document Stride (token)	Maximum Query Length (token)	Maximum Answer Length (character)	EM	F-Score
512	256	64	64	54,09	79,01
512	256	64	30	52,72	78,69
512	128	64	64	52,00	77,46
512	512	128	64	51,94	75,66
512	384	256	30	49,33	74,39
384	64	64	30	47,56	73,42
256	64	64	30	46,01	71,97
128	64	64	30	44,38	70,11

Table 4. Description of wrong answers with zero EM (3.2%).

Error ID	Description	Counts
1	Multiple possible answers	180
2	Questions requiring interpretation	85
3	Conditional answers	80
4	Questions requiring a list of elements	5
5	Answers with syntax variations	34
6	Incorrect question	49
7	Incorrect answers	139
	Total	**572**

Table 5. Examples for error types 1-5.

Error ID	Question	Real answer	Predicted answer
1	Kampanya için gerekli şartlar nelerdir?	Bankanın müşteri yada kredi kartına sahibi olmak	Kampanyadan yararlanmak isteyen müşteriler, üyeliklerini 31/12/2019 tarihine kadar aktifleştirmelidir.
2	Tarihi geçmiş belgeler için müşteri ne yapmalıdır?	Şubeleri ile görüşmeleri gerekmektedir.	güncellenmelidir
3	Parmak izi ile girişi tüm müşteriler kullanabilir mi?	sadece X marka telefonu olanlar	kullanabilir
4	İşlemler ne zaman fona dönüşür?	09:15, 11:15, 13:15, 15:15	günlük
5	Kampanya kuponları hangi sitelerde geçerlidir?	www.x.com	www.x.com'da

Table 6. Comparison of Turkish QA Systems. (1) Data sets translated to Turkish. (2) Mean Reciprocal Rank, considers the rank of the first correct answer in the list of possible answers. (3) Who, Where, When and What. (4) Author, capital, birth date, death date, language of country, birth place, death place. (*) Phase 1: Fine tuning the model using merged SQuAD (Tr) and NewsQA (Tr) data sets. (**) Phase 2: Fine tuning the model, which is already trained with phase 1, using Banking Sector QA (Tr) data set.

Study	Data sets	Domain	Metric	Results
BayBilmiş	TREC-9[1] and TREC-10[1]	Open	MRR[2]	0,313
Automatic QA for Turkish with Pattern Matching	Only Specific Questions[3]	Closed	Precision	0,79 (Average)
A Factoid QA System Using Answer Pattern Matching	Only Specific Factoid Questions[4]	Closed	MRR[2]	0,73
Current Study	SQuAD (Tr)[1] and NewsQA (Tr)[1*]	Open	EM	55,26
			F-Score	67,07
	Banking Sector QA (Tr)[**]	Closed	EM	54,09
			F-Score	79,01

Table 6 presents the results of this comparison in terms of Exact Match and F-Score. The performance of the system was measured in open and closed domain by taking in account its accuracy in the first phase of the fine-tuning procedure with the SQuAD (Tr) and NewsQA (Tr) data sets and its accuracy in the second phase where the Banking Sector QA (Tr) data set was additionally utilized for training.

The studies in Table 6 proposed solutions for certain types of questions, and the accomplished success rates have been achieved in particular data sets. Because of different evaluation metrics and test data sets, it is difficult to directly compare the results of the current study with the results of the previous QA systems in Turkish. But in general, based on the evaluation metrics (Mean reciprocal rank), and the type of data sets (question specific data sets) of these studies, it can be considered that the current study proposes a solution that is applicable to a wider range of QA's. The proposed method covers all types of questions, and there is no restriction in terms of question types or text for the evaluation data sets.

4.4. Comparison with other BERT language models

In order to evaluate the selected data sets and training procedure the current model was compared with models that used data sets in other languages. Table 7 presents the comparison with BERT models in the Arabic and Chinese Language in terms of EM and F-Score. Existing BERT models in the Persian, Korean and French language didn't present results for QA tasks.

Table 7. Comparison of BERT models with other languages. (1) Data sets translated to Turkish. (2) Multilanguage model published by Google. (3) Arabic Reading Comprehension Dataset, which was previously translated from SQuAD to Arabic. (4) Chinese specific model published by Google. (*) Phase 1: Fine tuning the model using merged SQuAD (Tr) and NewsQA (Tr) data sets.

Language	Model	Data sets	Domain	EM	F-Score
Arabic	mBERT	ARCD[3]	Open	34,2	61,3
	AraBERT	ARCD[3]		30,6	62,7
Chinese	BERT-Chinese[4]	CMRC		18,6	43,3
		DRCD		82,2	89,2
		CJRC		55,1	75,2
Turkish	Current Study	SQuAD (Tr)[1]		57,60	68,34
		NewsQA (Tr)[1]		48,01	59,86
		SQuAD (Tr)[1] and NewsQA (Tr)[1*]		55,26	67,07

Table 8. Comparison of Turkish base models. (1) Data sets translated to Turkish. (*) Phase 1: Fine tuning the model using merged SQuAD (Tr) and NewsQA (Tr) data sets. (**) Phase 2: Fine tuning the model, which is already trained with phase 1, using Banking Sector QA (Tr) data set.

Data sets	Model	Domain	EM	F-Score
SQuAD (Tr)[1] and NewsQA (Tr)[1*]	BERTurk	Open	57,43	69,36
	Current Study		55,26	67,07
	mBERT		54,52	65,74
Banking Sector QA (Tr)[**]	BERTurk	Closed	55,89	80,87
	Current Study		54,09	79,01
	mBERT		50,74	77,03

4.5. Comparison with other BERT Turkish models

The results show that the performance of this system is, most of the time, better than other existing models. The current Turkish language model was compared with two others, Google's multilingual version of BERT model, mBERT and another model in the Turkish language entitled BERTurk (Schweter, 2020). In Table 8, it can be observed that this model is better than mBERT and slightly inferior but still comparable with the BERTurk model (around 2% difference). The selection of larger and various data sets for the training procedure is one of the main reasons that generate those differences.

5. Discussion

In this work, the evaluation of a QA system was done in open and closed domains, based on a deep neural network that generates a model language with a context-sensitive vocabulary encoding. The results revealed the following findings:

In Turkish QA systems, deep neural network methodologies can cover a wider domain compared to other approaches. The results revealed that almost all other methods based on a semantic or rule-based approach are successful in specific types of QAs. This finding is expected and in general it is applicable to other fields apart from the NLP domain.

Language training data sets play a key role. The language model plays a significant role in the success of a QA system and the data sets used for generating have a major contribution. BERTurk performed slightly better because it used a wider training corpus. Special effort should be given by providing adequate data sets.

Translated data sets are adequate for fine tuning. The automatic translation of English data sets, even if they need a preprocessing step, gave high accuracy results during the fine tuning. According to authors' experience, this method can be applied to other languages and also other variations of deep neural networks based on bi-directional transformers.

The above finding let conclude that the proposed methodology can be successfully used in QA tasks for any language and any domain. Based on this experience, QA tasks can be carried out in an open or any closed domain, provided that existing data sets in English will be translated with a preprocessing procedure and an adequate data set will be generated for a particular domain. Generating a data set in SQuAD format for a closed domain even if it is time-consuming and requires human resources, it is still necessary for obtaining a closed domain QA system with high accuracy. Today, researchers are focused on increasing the accuracy of generic QA tasks in order to reduce the performance gap between open and closed domains. As a result, new variations of neural networks using bi-directional transformers like BERT are proposed, and new data sets are emerging for training purposes.

6. Conclusion

This study presented a QA system in the Turkish language for the banking domain. This approach required the use of various and large data sets. Even if BERT cannot solve the morphological disambiguation problem of Turkish, the overall architecture is sufficient for solving the word sense disambiguation problem. For QA tasks in open domain, a translation and preprocessing step of some data sets was necessary and for answering questions in the banking domain, the generation of a new data set was required. The experiments showed that the accuracy of the network can significantly vary according to the choice of the training parameters. To the best of our knowledge, this study is the first that proposes a framework in the Turkish language for a QA task in open and also in closed domain using deep neural networks. Additionally, the proposed methodology is applicable to any language and to any domain for performing QA tasks.

Acknowledgements

We would like to thank Prof. Dr. Kemal Oflazer for sharing the News Corpus (Tr), described in Section 3.1 and used in the pre-training task and the QA tasks for open and closed domain were generated using BERT neural network. We also thank TEB - Turk Ekonomi Bankasi and TEBArf for their support in generating the Banking QA data set and funding the TPU servers used in the study.

References

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint 2018, arXiv: 1810.04805.

Çelebi, E., Günel, B., Şen, B., 2011. Automatic question answering for Turkish with pattern parsing. INISTA 2011 - 2011 International Symposium on INnovations in Intelligent SysTems and Applications, pp. 389–393. https://doi.org/10.1109/INISTA.2011.5946098

Amasyalı, M.F., Diri, B., 2005. Bir soru cevaplama sistemi: Baybilmiş. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 1(1).

Er, N.P., Cicekli, I., 2013. A factoid question answering system using answer pattern matching. Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 854–858.

Biricik, G., Solmaz, S., Özdemir, E., Amasyalı, M.F., 2013. A Turkish Automatic Question Answering System with Question Multiplexing: Ben Bilirim. International Journal of Research in Information Technology (IJRIT) 1(6), 46–51.

Akın, A.A., Akın, M.D., 2007. Zemberek, an open source nlp framework for turkic languages. Structure, 10, 1–5.

Oflazer, K., Say, B., Hakkani-Tür, D.Z., Tür, G., 2003. Building a Turkish Treebank. Springer, pp. 261–277. https://doi.org/10.1007/978-94-010-0201-1_15

Eryiğit, G., Oflazer, K., 2006. Statistical dependency parsing for turkish, in: 11th Conference of the European Chapter of the Association for Computational Linguistics.

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, pp. 2227–2237. https://doi.org/10.18653/v1/n18-1202

Radford, A., Salimans, T., 2018. Improving Language Understanding by Generative Pre-Training. OpenAI, 1–12.

Fisch, A., Talmor, A., Jia, R., Seo, M., Choi, E., Chen, D., 2019. MRQA 2019 Shared Task: Evaluating Generalization in Reading Comprehension, pp. 1–13. https://doi.org/10.18653/v1/d19-5801

Antoun, W., Baly, F., Hajj, H., 2020. AraBERT: Transformer-based Model for Arabic Language Understanding. arXiv preprint 2020, arXiv:2003.00104.

Martin, L., Muller, B., Suárez, P.J.O., Dupont, Y., Romary, L., de la Clergerie, É.V., Seddah, D., Sagot, B., 2019. CamemBERT: a Tasty French Language Model. arXiv preprint 2019, arXiv:1911.03894.

Farahani, Mehrdad, Gharachorloo, M., Farahani, Marzieh, Manthouri, M., 2020. ParsBERT: Transformer-based Model for Persian Language Understanding. arXiv preprint 2020, arXiv:2005.12515.

Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z., Wang, S., Hu, G., 2019. Pre-Training with Whole Word Masking for Chinese BERT. arXiv preprint 2019, arXiv:1906.08101.

Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P., 2016. SQuad: 100,000+ questions for machine comprehension of text. EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings, pp. 2383–2392.

Trischler, A., Wang, T., Yuan, X., Harris, J., Sordoni, A., Bachman, P., Suleman, K., 2016. Newsqa: A machine comprehension dataset. arXiv preprint 2016, arXiv:1611.09830.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in Neural Information Processing Systems, 5999–6009.

Wiedemann, G., Remus, S., Chawla, A., Biemann, C., 2019. Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings. arXiv preprint 2019, arXiv:1909.10430.

Wu, Y., Schuster, M., Chen, Z., Le, Q. v., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint 2016, arXiv: 1609.08144.

Schweter, S., 2020. BERTurk - BERT models for Turkish. https://doi.org/10.5281/zenodo.3770924

Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 4(2) (2021) 76-85
DOI: 10.38016/jista.857881

Harris Şahinleri ve Balina Optimizasyon Algoritmalarının Kısıt İşleme Teknikleriyle Uygulaması: Karşılaştırmalı bir çalışma

Zeynep Garip[1]*, Murat Erhan Çimen[2], Ali Fuat Boz[2]

[1]Sakarya Uygulamalı Bilimler Üniversitesi, Bilgisayar Mühendisliği Bölümü, Sakarya, Türkiye

[2]Sakarya Uygulamalı Bilimler Üniversitesi, Elektrik ve Elektronik Mühendisliği Bölümü, Sakarya, Türkiye

zbatik@subu.edu.tr, muratcimen@subu.edu.tr, afboz@subu.edu.tr

Öz

Bu çalışmada kısıtlı optimizasyon problemlerin çözümünde, doğadan ilham alan meta-sezgisel algoritmaların etkisine odaklanılmıştır. Kısıtlı ihlal tekniklerinden olan ölüm ceza, statik ceza, dinamik ceza, bariyer fonksiyon ve Deb uygulanabilirlik kuralı Balina Optimizasyon (BOA) ve Harris Şahini Optimizasyon (HŞO) algoritmaları üzerinde test edilmiştir. Algoritmaların performansını test etmede kısıtsız ve kısıtlı benchmark fonksiyonları ve optimal güç akışı minimizasyon problemi kullanılmıştır. Ayrıca BOA ve HŞO algoritmaların optimal güç akışında gösterdikleri performanslarını karşılaştırmak amacıyla literatürde bulunan algoritmalarla kullanılmıştır. Sonuç olarak kısıt ihlal yöntemlerine entegre edilmiş algoritmaların kısıtlı optimizasyon problemlerin çözümünde etkili olduğu görülmüştür.

Anahtar kelimeler: Meta-sezgisel Algoritma; Kısıt İhlal Yöntemleri; Optimal Güç Akışı; Optimizasyon.

Application of Harris Hawks and Whale Optimization Algorithm with Constraint Handling Techniques: A comparative study

Abstract

This study focuses on the effect of meta-heuristic algorithms inspired by nature in solving constrained optimization problems. Death penalty, static penalty, dynamic penalty, barrier function and Deb feasibility rule, which are the constraint handling techniques, were tested on Whale Optimization (WOA) and Harris Hawk Optimization (HHO) algorithms. Unconstrained and constrained benchmark functions and optimal power flow minimization problem were used to test the performance of algorithms. Furthermore, in order to compare the performance of WOA and HHO algorithms in optimal power flow, it was used with algorithms found in the literature.. As a result, it has been observed that algorithms integrated into constraint handling methods are effective in solving constrained optimization problems.

Keywords: Meta-heuristic Algorithm; Constraint Handling Methods; Optimal Power Flow; Optimization.

1. Giriş (Introduction)

Kısıtlı optimizasyon problemlerini çözmede meta-sezgisel algoritmaların kısıtlı optimizasyona uyarlanması önemli bir görevdir. Kısıtlama yöntemleri ceza fonksiyonlarının varlığına bağlı olarak iki gruba ayrılır. İlki ceza temeli kısıtlı yöntemlerinde amaç fonksiyonuna uygulanabilir bölge dışında bulunan fonksiyona değeri ceza olarak eklenir. Bu yöntemlere ilave olarak Deb uygulanabilirlik yönteminin önerilmesi algoritmaların kısıtlı problemlerin çözümüne farklı bir bakış açısı kazandırmıştır. Literatürde ölüm cezası (Fan vd., 2019), statik ceza (Tsipianitis vd., 2020), dinamik ceza (Paszkowicz 2009), bariyer fonksiyonu (Matias vd., 2015), epsilon kısıtlama (Fang vd., 2020), stokastik sıralama (Bansal vd., 2009), Deb uygulanabilirlik kuralları (Babalik vd., 2018) gibi çeşitli kısıtlama teknikleri bulunmaktadır. Ayrıca kısıtlama teknikleri üzerinde yürütülen çalışmalarda sezgisel algoritmaların performanslarının arttırılmasına yönelik çalışmalar yapılmaya devam etmektedir.

Son zamanlarda popülasyon tabanlı meta-sezgisel algoritmaların optimizasyon performansları, basit ve etkili yapıları, adaptasyon ve uygulama kolaylığı nedeniyle

* Sorumlu yazar.
E-posta adresi: zbatik@subu.edu.tr

Alındı : 11 Ocak 2021
Revizyon : 15 Mart 2021
Kabul : 20 Mart 2021

araştırmacılar tarafından sıklıkla tercih edilmektedir (Babalik vd., 2018). Meta sezgisel algoritmalar keşif ve sömürü olmak üzere iki arama yöntemini içermektedir. Bu iki yöntemin dengelenmesi algoritmaların farklılaşmasını sağlar. Bununla birlikte arama sürecindeki rastgelelikten dolayı keşif ve sömürü arasında denge sağlanması kolay değildir. Ayrıca tüm sorunları çözecek evrensel bir algoritma yoktur. Başka bir deyişle bir algoritma bazı problemler için başarılı sonuçlar verirken diğerlerinde başarılı sonuçlar vermeyebilir.

Meta sezgisel algoritmalar, ayrık veya sürekli, tek veya çok amaçlı, kısıtlı veya kısıtsız optimizasyon problemlerinin çözümünde etkili bir şekilde kullanılmaktadır (Haklı 2019). Burada çözülmesi gereken önemli, sorun arama süresi boyunca en uygun değerin seçilmesidir (Chen vd., 2019). Kısıtlı optimizasyon problemlerine yapısal optimizasyon, mühendislik tasarımı, VLSI tasarımı, ekonomi gibi birçok alanda karşılaşılmaktadır (Akay vd., 2011). Son zamanlarda, doğadaki birey rekabetinden ve sürü işbirliğinden esinlenerek kısıtlı optimizasyon problemini çözmek için genetik algoritma (Jelovic vd., 2020), diferansiyel evrimi (Amaratunga vd., 2018), yapay arı koloni (Akay vd., 2011), parçacık sürüsü optimizasyonu (Kohler vd., 2019), çiçek tozlaşma algoritması (Fan vd., 2019), ağaç tohum algoritması (Babalik vd., 2019), sinüs kosinüs algoritması(Chen vd., 2020) ve ateş böceği algoritma (Tuba vd., 2014), Harris Şahini (Aljarah vd., 2019) ve Balina Optimizasyon algoritmaları (Lewis vd., 2016) gibi birçok meta sezgisel algoritma kullanılmıştır.

Literatürde kısıtlı optimizasyon problemlerini çözmek için çeşitli kısıtlama yöntemleri ve meta-sezgisel algoritmalar birlikte kullanılmaktadır. Babalık ve ark. tarafından kısıtlı optimizasyon tekniklerini çözmek amacıyla kısıtlama tekniklerinden biri olan Deb'in kuralları kullanılarak ağaç tohumu algoritması modifiye edilmiştir (Babalik vd., 2018). Karaboğa ve ark. kısıtlı problemleri çözmek için ABC algoritmasının seçim mekanizmasını Deb'in seçim mekanizmasıyla değiştirmiştir. Geliştirilen algoritmada kısıtlı optimizasyon problemlerin çözümünde uygunluk değerlerine göre uygulanabilir ve ihlallere göre uygun olmayan çözümlere olasılık değeri atanarak kısıtlamalar çözülmeye çalışılmıştır (Akay vd., 2011). Miranda-Varela ve ark. hedeflenen fonksiyonun değerine ve kısıt ihlallerinin toplamına yaklaşmak amacıyla kısıt yaklaşımları ve en yakın komşu regresyonu tabanlı diferansiyel algoritmaları arasındaki ilişki üzerinde durulmuştur (Miranda-Varela vd., 2018). Samanipour ve ark. dominat olmayan genetik algoritmayla kısıtlama tekniklerinin etkinliğini artırmak amacıyla onarım yaklaşımı önermişlerdir (Jelovic vd., 2020). Rodrigues ve ark. optimizasyon problemlerindeki kısıtları göz önünde bulundurarak, uyarlanabilir genişletilmiş dengeli bir sıralama yöntemi geliştirmişlerdir (Guimarães vd., 2018). Gandomi ve ark. tarafından önerilen yaklaşımda; kısıtlamaları ihlal etmeyecek şekilde uyarlanabilen ve değişkenin sınırlarını onararak uygulanabilir çözümlerin değerlendirilmesini sağlamışlardır (Deb vd., 2020).

Tablo 1. Literatür Karşılaştırması (Literature Comparison)

Yöntem	Kısıtlama Tekniği	Referans
BOA, HŞO	Ölüm cezası, Statik ceza, Dinamik ceza, Bariyer ceza ve Deb uygulanabilirlik kuralı	Bu çalışma
Ağaç Tohum Algoritması	Deb Kuralları	(Babalik vd., 2018)
Yapay Arı Koloni Algoritması	Deb seçim mekanizması	(Akay vd., 2011)
Diferansiyel Evrim Algoritması	Uygunluk kuralları, ε-kısıt methot, Stokastik sıralama, Çeşitlilik sağlama	(Miranda-Varela vd., 2018)
Genetik Algoritma	Uyarlanabilir eşik yaklaşımı, Deterministik Uygun Olmama Sıralaması	(Jelovic vd., 2020)
Genişletilmiş dengeli sıralama yöntemi	Deb Uygulanabilirlik kuralı, Stokastik Sıralama, Ceza Yöntemleri	(Guimarães vd., 2018)
İç nokta, Sıralı Karesel Programlama, Doğrudan Arama, Benzetim Tavlama, Parçacık Sürü Optimizasyonu, Genetik Algoritma, Diferansiyel Evrim Algoritması	Sınır Belirleme	(Deb vd., 2020)

Bu çalışmada ölüm cezası, statik ceza, dinamik ceza, bariyer ceza ve Deb uygulanabilirlik kısıtlama teknikleri ve kambur balinaların avcılık davranışından esinlenilen Balina Optimizasyon algoritması (BOA) ve Harris şahinlerinin yiyecek arama davranışından ilham alınan Harris Şahini Optimizasyon Algoritması (HŞO), kısıtlı optimizasyon problemlerini çözmek için uyarlanmıştır.

Bu çalışmanın başlıca katkıları aşağıdaki gibi özetlenebilir:

1. Ölüm cezası, statik ceza, dinamik ceza, bariyer ceza ve Deb uygulanabilirlik kısıtlama teknikleri incelenmiştir.
2. Kısıt işlemeye yöntemlerine HŞO ve BOA meta-sezgisel algoritmaları adapte edilmiştir.
3. Algoritmaların performansları kısıtsız ve kısıtlı fonksiyonlar ve optimal yük akışı mühendislik problemi üzerinde incelenmiştir.
4. BOA ve HŞO algoritmaların optimal güç akışında gösterdikleri performanslarını karşılaştırmak amacıyla literatürde bulunan HŞO (Islam vd., 2020), Gri Kurt Optimizasyon (GKO) (Islam vd., 2020), BOA (Islam vd., 2020), GA (Bouktir vd., 2008), Harris Şahini-

Diferansiyel Evrim (HŞODE) (Birogul, 2019) ve Modifiyeli diferansiyel evrim (Sayah vd., 2008) algoritmaları kullanılmıştır.

2. Optimizasyon (Optimization)

Optimizasyon problemlerinde belirli sınırlar arasındaki parametreler ile en uygun çözümü bulmaktır. Buna ek olarak optimizasyon işleminde tüm kısıtlamaları sağlayan birçok çözüm içerisinden en uygun çözüm ya da yaklaşık çözümün bulunması amaçlanmaktadır (Mert vd., 2021). Bu bölümde kısıtlı optimizasyon probleminin tanımı yapıldıktan sonra Harris şahin optimizasyon ve balina optimizasyon algoritmaları açıklanmıştır.

2.1. Kısıtlı problemlerin matematiksel modellenmesi (Mathematical modeling of constrained problems)

Optimize edilmek istenen amaç fonksiyonuna $f(X)$ eşitlik ve/veya eşitsizlik kısıdı (sırasıyla $h_i(X) = 0, g_i(X) \leq 0$) eklendiğinde problem kısıtlı optimizasyon problemi haline dönüşmektedir (Garcia vd., 2017). Eşitlik 1'de verilen çok değişkenli bir kısıtlı optimizasyon probleminde $f(X)$ fonksiyonun minimum olması istenmektedir. Bu problemin çözümünde N adet eşitlik $h(X)$ kısıtı ve M adet eşitsizlik $g(X)$kısıtlarında sağlaması gerekmektedir.

$$\begin{aligned} &minimum\ f(X) \quad X = [x_1, x_2, \dots, x_D] \in R^D \\ &h_i(X) = 0 \qquad (i = 1, 2, \dots, N) \\ &g_i(X) \leq 0 \qquad (j = 1, 2, \dots, M) \end{aligned} \tag{1}$$

Kısıtlı optimizasyon problemlerinde kısıt fonksiyonları amaç fonksiyonla aynı derece de etkilidir. Çünkü çözümün uygunluğu amaç fonksiyon kullanılarak hesaplanırken, çözümün uygulanabilirliği kısıtların ihlaline bağlıdır. Bu nedenle, bir çözümün uygulanabilirliği, uygunluk değerinden daha önemlidir (Babalik vd., 2018).

2.2. Balina Optimizasyon Algoritma (Whale Optimization Algorithm)

Balina optimizasyonu algoritması (BOA), optimizasyon problemlerinde kullanılmak amacıyla kambur balinaların avcılık davranışından ilham alınarak Marjalili ve Lewis tarafından önerilmiştir (Lewis vd., 2016). Yalnızca kambur balinalarda gözlenen yiyecek arama davranışı, kabarcık-ağ besleme yöntemidir. Balinalar, avlanma sırasında avı çevrelerken dairesel bir yol boyunca kabarcıklar oluşturur. Optimizasyonun gerçekleştirilmesi için, spiral kabarcık-ağ besleme davranışının işlem adımlarını gösteren sözde kodu Şekil 1'de verilmiştir.

Balina popülasyonunu oluştur ***Xi (i =1, 2, …, n)***
Her bir ajan uygunluk popülasyonu hesapla
X* =en iyi arama ajanı
while (t < maksimum iterasyon sayısı)
for her arama ajanı
a, A, C, l, ve p değerlerini güncelle
if (p < 0.5)
  if (|**A**| < 1)
Eş.(2) ile mevcut arama ajanın konumunu güncelle
  else if (|A| ≥1)
Rasgele arama ajanı seç (X_{rand})
Eş.(3) ile mevcut arama ajanın konumunu güncelle
  end if
else if (p≥0.5)
Eş.(8) ile mevcut arama ajanın konumunu güncelle
end if
end for
Herhangi bir arama ajanının arama alanının ötesine geçip geçmediğini kontrol edin ve değiştirin
Her bir arama ajanın uygunluğunu hesapla
Daha iyi bir çözüm varsa X* güncelle
t=t+1
end while
return X*

Şekil 1. BOA'nın sözde kodu (Pseudo Code of WOA)

2.1.1. Avı Kuşatma (Encircling prey)

Kambur balinalar avlanırken avının yerini bulabilir ve avı çevreleyebilir. BOA'da arama alanındaki en uygun tasarımın yeri önceden bilinmediğinden, mevcut en iyi aday çözümün hedef av olduğunu veya optimum duruma yakın olduğunu varsayar. En iyi arama ajanı tanımlandıktan sonra, diğer arama ajanları konumlarını en iyi arama ajanına doğru güncellemeye çalışacaktır. Kambur balinaların avını kuşatma davranışının matematiksel modeli Denklem 2 ve 3'de gösterilmiştir. Denklem 2-3'de bulunan $\overrightarrow{X}(t)$ ajanın konumunu, t iterasyon, $\overrightarrow{X^*}$ en iyi çözümü ifade ederken Denklem 4 ve 5'da $\vec{A}, \vec{C}$ yakınsama değerlerini temsil etmektedir. $\vec{r}$ [0,1] rastgele sayı ve $\vec{a}$ iterasyon boyunca 2'den sıfıra doğrusal olarak azalan vektörü göstermektedir.

$$\vec{D} = \left|\vec{C}\,\vec{X}^*(t) - \overrightarrow{X}(t)\right| \tag{2}$$

$$\vec{X}^*(t+1) = \left|\vec{X}^*(t) - \vec{A}.\vec{D}\right| \tag{3}$$

$$\vec{A} = 2\vec{a}\,.\vec{r} - \vec{a} \tag{4}$$

$$\vec{C} = 2.\vec{r} \tag{5}$$

2.1.2. Kabarcık-net saldırı yöntemi (Bubble-net attacking method)

Kambur balinaların kabarcık-net saldırma yönteminde avına doğru ilerken arama çevresini küçültme ve ava doğru spiral şeklinde yol alma bulunmaktadır. Balinaların Denklem 8'de bulunan $\vec{a}$ değerinin düşürülmesiyle arama çevrelerini küçülterek avı yakalama davranışlarını sergilerler. $\vec{A}$ değeri de $\vec{a}$ değerine bağlı olduğundan dolayı 2'den sıfıra doğrusal olarak azalır. Kambur balinaların

avını yakalarken oluşturdukları spiral şeklin matematiksel modeli Denklem 6 ve 7'de verilmiştir.

$$\overrightarrow{D'} = \left|\vec{X}^*(\mathrm{t}) - \overrightarrow{X(\mathrm{t})}\right| \quad (6)$$

$$\vec{X}^*(\mathrm{t}+1) = \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \vec{X}^*(\mathrm{t}) \quad (7)$$

Denklem 6 ve 7'de verilen D' balina ve en iyi av arasındaki mesafe, b logaritmik spiral sabiti, l ise [-1,1] arasında rastgele sayıdır. Kambur balinalar avına doğru hareket ederken yüzde 50 olasılıkla ya daralan hareket modelini ya da spiral hareket modelinden birini seçmektedir. Denklem 8'de bulunan p parametresi [0,1] aralığında rasgele sayıdır.

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(\mathrm{t}) - \vec{A}.\overrightarrow{D} & p < 0.5 \\ \overrightarrow{D'}.e^{bl}.cos(2\pi l) + \overrightarrow{X^*}(t) & p \geq 0.5 \end{cases} \quad (8)$$

2.2. Harris Şahini Optimizasyon Algoritması (Harris Hawks Optimization Algorithm)

Harris Şahini optimizasyon algoritması(HHO), optimizasyon problemlerinde kullanılmak amacıyla Harris şahinlerinin yiyecek arama davranışından ilham alınarak Ali Asghar Heidari ve arkadaşları tarafından önerilmiştir. Şahinler izleme, kuşatma ve saldırı özelliklerini kullanarak çeşitli aşamalarla işbirliğine dayalı yiyecek arama işlemini verimli bir şekilde gerçekleştirirler (He vd., 2020). Algoritmanın işlem adımlarını gösteren sözde kodu Şekil 2'de verilmiştir.

N boyutunda rasgele popülasyonu oluştur
while (Durdurma kriteri sağlanana kadar) **do**
Şahinlerin uygunluk değerlerini hesapla
Xtavsan konumu en iyi konum olarak ata
for (herbir şahin (X_i)) **do**
BAşlangıç enerjisi E_0 ve atlama kuvveti *J güncelle*
E_0=2rand()-1, J=2(1-rand())
Eş.(11) ile E'yi güncelle
if ($|E| \geq 1$) **then** ▷ Keşif aşaması
Eş.(9) ile konumu güncelle
if ($|E| < 1$) **then** ▷ Sömürü aşaması
if ($r \geq 0.5$ and $|E| \geq 0.5$) **then** ▷ Yumuşak kuşatma stratejisi- Eş.(12) ile konumu güncelle
else if ($r \geq 0.5$ and $|E| < 0.5$) **then** ▷ Sert kuşatma stratejisi- Eş.(13) ile konumu güncelle
else if ($r < 0.5$ and $|E| \geq 0.5$) **then** ▷ Hızlı saldırılarla yumuşak kuşatma stratejisi
Eş.(14) ile konumu güncelle
else if ($r < 0.5$ and $|E| < 0.5$) **then** ▷ Aşamalı hızlı saldırılarla sert kuşatma stratejisi- Eş.(16) ile konumu güncelle
Return *Xtavsan*

Şekil 2. HŞO algoritmasının sözde kodu(Pseudo Code of HHO)

2.2.1. Keşif aşaması (Exploration phase)

Harris şahinleri avını dev ağaçlarda veya telgraf direklerinde keskin bakışlarıyla arayarak keşif aşamasını gerçekleştirmektedir. HŞO algoritmasında, arama davranışı global keşif aşaması olarak kabul edilir. Global keşif stratejileri Denklem 9 ile matematiksel olarak ifade edilmektedir. Denklem 9'daki stratejilerden hangisinin seçileceğini olasılık değeri olan q değeri belirler.

$$x_i^{t+1} = \begin{cases} x_{rand}^t - r_1 * |x_{rand}^t - 2 * r_2 * x_i^t|, & q \geq 0.5 \\ \left(x_{tavşan} - x_{mean}^t\right) - r_3 * \left(lb + r_4 * (ub - lb)\right), & q < 0.5 \end{cases} \quad (9)$$

x_i^t değişkeni Harris Şahinin mevcut konumunu verirken x_i^{t+1} her bir iterasyondaki konum vektörüyken $x_{tavsan}(t)$ avın konum vektörüdür. r_1,r_2,r_3,r_4 ve q ise 0 ve 1 aralığındaki rassal sayılardır. ub ve lb, sırasıyla popülasyonun üst ve alt sınır değerleridir. x_{rand}^t mevcut popülasyondan rastgele seçilen bir şahini gösterirken, x_{mean}^t mevcut şahin popülasyonunun ortalama konum değerlerini vermektedir. Ortalama konum değeri t iterasyonda N şahin sayısı kullanılarak denklem 10 kullanılarak elde edilir (Aljarah vd., 2019).

$$x_{mean}^t = \frac{1}{n}\sum_{i=1}^{n} x_i^t \quad (10)$$

2.2.2. Keşiften sömürüye geçiş aşaması (Transition from exploration to exploitation)

HŞO algoritmasında global aramadan yerel sömürmeye geçiş kaçan avın enerjisine bağlı olarak oluşan enerji faktörü (E) ile kontrol edilmektedir. Enerji faktörü denklem 11 ile modellenmektedir. Bu modelde E kaçan avın enerjisini, T maksimum yineleme sayısını ve E_0 (-1, 1) avın enerjisinin başlangıç halidir.

$$E = 2E_0\left(1 - \frac{t}{T}\right) \quad (11)$$

Kaçan avın enerjisi $|E| \geq 1$ olduğunda, şahinler bir tavşanın yerini keşfetmek için farklı bölgelerde arama yaparlar böylece keşif aşaması gerçekleşir. $|E| < 1$ olduğu sömürü aşamasında ise çözümlerin komşuluğundan yararlanılmaktadır.

2.2.3. Sömürü Aşaması (Exploitation stage)

Harris şahinleri hedef avı bulduktan sonra avın etrafında bir çember oluşturur. Şahinler avın davranışa göre saldırı türünü belirlemektedir. Avın kaçma davranışlarına ve Harris'in şahinlerinin kovalamaca stratejilerine göre, saldırı aşamasını modellemek amacıyla dört olası strateji önerilmiştir. Stratejiler rastgele sayı(r) ve avın kaçan enerjisine(E) bağlıdır. r (0, 1) avın kuşatma halkasından kaçıp kaçamayacağına karar vermek için kullanılır.

Yumuşak kuşatma stratejisi $r \geq 0.5$ ve $|E| \geq 0.5$ olduğu durumdur. Bu durumda avın kaçma şansı yoktur, ancak kuşatma halkasından kaçacak kadar enerjisi olduğu için şahinler yumuşak kuşatma ile avlanırlar. Matematiksel modellemesi denklem 12'de verilmiştir. Δx^t

popülasyondaki uygun avdan mevcut arasındaki vektör mesafesidir. r_5 (0,1) 'de eşit olarak dağıtılmış rasgele bir sayı iken J ise avın kaçış sırasındaki atlama uzunluğunu gösterir.

$$x_i^{t+1} = \Delta x_i^t - E * \left| J * x_{tavşan} - x_i^t \right|$$
$$\Delta x_i^t = x_{tavşan} - x_i^t \quad (12)$$
$$J = 2 * (1 - r_5)$$

Sert kuşatma stratejisi r ≥ 0.5 and |E| < 0.5 olduğu durumdur. Avın kaçma şansı olmadığı için enerjisi de yetersizdir. Bu durumda şahinler sert kuşatma ile avlanırlar.

$$x_i^{t+1} = x_{tavşan} - E * |\Delta x_i^t| \quad (13)$$

Aşamalı hızlı saldırılarla yumuşak kuşatma stratejisinde r <0.5 ve | E | ≥ 0.5 olduğu durumdur. Bu durumda kaçmak için gerekli enerjiye sahip avın kuşatma halkasından kaçma şansı vardır. Bundan dolayı şahinler avı yakalamak amacıyla daha akıllı ve yumuşak bir kuşatma halkası oluşturacaktır. Bu strateji iki adımdan oluşmaktadır. Şahinler ava doğru konumu ilk adımla iyileşmezse ikinci adımla konumu güncellenmektedir. İlk adımda yumuşak kuşatma stratejisinde bulunan konum denklemi kullanılmaktadır. İkinci adım olan güncelleme modu Denklem 14 ile modellenmiştir. $s \in \mathbb{R}^{dim}$, 1xdim boyutunda rastgele bir vektördür. Levy fonksiyonu denklem 15 ile tanımlanmıştır. Burada u, v (0,1) arası rastgele sayı, β ise 1.5'tir.

$$z = \Delta x_i^t - E * \left| J * x_{tavşan} - x_i^t \right| + s * levy(dim) \quad (14)$$

$$levy(x) = 0.01x\frac{u-\sigma}{|\mu|^{\frac{1}{\beta}}}, \sigma = \left(\frac{\Gamma(1+\beta)xsin(\frac{\pi\beta}{2})}{\Gamma(1+\beta)x2^{\frac{\beta-1}{2}}}\right)^{\frac{1}{\beta}} \quad (15)$$

Aşamalı hızlı saldırılarla sert kuşatma stratejisinde |E| <0.5 ve r < 0.5 olduğu durumdur. Bu durumda kaçmak için gerekli enerjisi olmayan avın kuşatma halkasından kaçma şansı yoktur. Bundan dolayı şahinler ava saldırmadan önce sert bir kuşatma halkasıyla yakalayarak öldürmektedir. Matematiksel modellemesi 16 denklemi ile ifade edilmektedir.

$$x_i^{t+1} = \begin{cases} y, & if\ f(y) < f(x_i^t) \\ z, & if\ f(z) < f(x_i^t) \end{cases}$$

$$y = x_{tavşan} - E * \left| J * x_{tavşan} - x_{mean}^t \right|$$

$$z = y + s * levy(dim) \quad (16)$$

3. Kısıt İşleme Yöntemleri (Constraint Handling Methods)

Meta-sezgisel algoritmalar kısıtlı optimizasyon problemlerini çözmek için tasarlanmamıştır fakat bu sorunları çözmede çok etkilidirler. Kısıt işleme yöntemlerinin amaç fonksiyona eklenmesi arama uzayının uygun bölgelerinde arama yapmasını sağlar.

3.1. Ölüm Ceza Yöntemi (Death Penalty Methods)

Ölüm ceza yöntemi literatürdeki en kolay ve uygulanabilir kısıt işleme yöntemlerinden biridir. Bu yöntemde elde edilen olası çözümler uygun bölgede ise ceza değeri eklenmez. Aksi durumda yani olası çözümler içinde uygun bölgede bulunmazsa kısıtları ihlal eden çözümlere çok yüksek bir hata ataması yapılmaktadır. Yöntemin uygunluk fonksiyonu Denklem 17 ile modellenmektedir. Denklem 17'de verilen s değeri uygun bölge içerisindeki kısıt sayısını ifade etmektedir. K değeri ise çözümlerin uygun bölgede olması durumlarda eklenen yüksek bir değerdir.

$$\varphi(x) = \begin{cases} f(x) & kısıt\ ihlali\ yok \\ K - \sum_{i=1}^{s} \frac{K}{M} & kısıt\ ihlali\ var \end{cases} \quad (17)$$

Bu yöntemde kısıtlar ihlal edildiğinde, uygunluk fonksiyonunun değerine bakılmaksızın ihlaller aynı hata değeri ile değerlendirilir. Diğer bir deyişle, uygunluk fonksiyon değeri ve ihlallerin büyüklükleri önemli değildir.

3.2. Statik Ceza Yöntemi (Static Penalty Methods)

Statik ceza kısıt işleme yönteminde cezalandırma işleminde her bir ihlal değeri sabit bir sayı ile çarpılarak hesaplanmaktadır. Çözümlerin uygun bölmeye alınması için uygunluk fonksiyonu düzenlenerek Eşitlik 18'deki matematiksel model elde edilir. Bu eşitlikte λ ve μ sabit parametreler 1 ile ∞ arasında değeri değişmektedir.

$$\varphi(X) = f(X) + \lambda \sum_{i=1}^{N} |h_i(X)| + \mu \sum_{j=1}^{M} \max\{0, h_j(X)\} \quad (18)$$

3.3. Dinamik Ceza Yöntemi (Dynamic Penalty Methods)

Dinamik ceza yönteminde kısıtlama ihlal edildiğinde dinamik veya artan bir oranda ceza işlemi gerçekleşmektedir. Çözümlerin uygun bölmeye taşımak amacıyla ceza katsayısının güncellenmesi gerekmektedir. Ceza katsayısının güncellenmesi Eşitlik 19'da matematiksel modeldeki λ(t) iterasyon sayısına bağlı olarak dinamik olarak değişmektedir. λ(t) = $(\alpha t)^{\beta}$ matematiksel modelinde belirtildiği gibi $\lambda(t)$ değerinin dışında β değişkenin iterasyona bağlı olarak değişmesinin dinamik ceza yönteminin etkinliğini arttırmaktadır (Batık vd., 2019, Hatamlou vd., 2016).

$$\varphi(X) = f(X) + \lambda(t)\left(\sum_{i=1}^{N}(h_i(X))^2 + \sum_{j=1}^{M}\left(\max\{0, h_j(X)\}\right)^2\right) \quad (19)$$

Bu çalışmada kullanılan işlem parametreleri α = 0.5 ve β = 1.2 olarak alınmıştır. Fakat simülasyon çalışmalarında

amaç fonksiyonun çok yüksek değerlerde olduğu görüldüğünden dolayı hata fonksiyonun yetersiz olduğu belirlenmiştir. Bu sebepten dolayı $\lambda(t)$ değeri $\lambda(t) = 10^3(1 + (\alpha t)^{\beta})$ matematiksel modeli ile hesaplanmaktadır.

3.4. Bariyer Fonksiyon (Barrier Function)

Bariyer fonksiyon kısıt işleme yönteminde eşitlik kısıtları lagrange çarpanlarıyla sağlanmaktadır. Eşitsizlik kısıtları ise çözümler uygun olmayan bölgeye yaklaştıkça ceza fonksiyonu büyük veya sonsuz değerler almaktadır. Sonraki aşama işe elde edilen ceza fonksiyon değeri amaç fonksiyonun değerine eklenmektedir. Bariyer fonksiyonun matematiksel modellemesi Eşitlik 20 veya 21 ile sağlanmaktadır. Buna ek olarak bu fonksiyon değerleri hesaplanırken iterasyonlar sayısı artması durumunda $\mu(t) > 0$, $\mu \to 0$ ve $\mu(t) = 1/t$ veya $\mu(t) = 1/\sqrt{t}$ seçilmesi gerekmektedir.

$$\varphi(X) = f(X) + \mu(t)\sum_{j=1}^{M} -\log\left(g_j(X)\right) \tag{20}$$

$$\varphi(X) = f(X) + \mu(t)\sum_{j=1}^{M} \frac{1}{g_j(X)} \tag{21}$$

3.5. Deb'in Uygulanabilirlik Kuralı (Deb Feasibility Rule)

Deb'in uygulanabilirlik kuralında kısıtlı problemlerin çözümünde uygun veya uygun olmayan bölgeler için bir takım kurallar bulunmaktadır. Problemin çözümündeki her bir bireyin ihlal derecesi Eşitlik 22 kullanılarak hesaplanmaktadır. Hesaplanan değerlerin göre seçim yapılmaktadır. Debin önermiş olduğu kurallar göz önünde bulundurularak uygun olmayan bölgede bulunan tüm çözümler uygun bölgeye doğru hareket ettirilmekte veya yönlendirilmeye çalışılmaktadır.

$$\varphi(X) = \sum_{i=1}^{N}(h_i(X))^2 + \sum_{j=1}^{M}\left(\max\{0, h_j(X)\}\right)^2 \tag{22}$$

Debin önermiş olduğu bu uygulanabilirlik kuralları aşağıdaki gibidir.

1. İki çözümde uygun bölgede olduğu durumda uygunluk fonksiyon değeri küçük olan tercih edilir.
2. Çözümlerden biri uygun ve diğeri uygun olmayan bölgede olduğu durumda uygun bölgedeki çözüm tercih edilir.
3. Her iki çözümde uygun olmayan bölgede olduğu durumda kısıt ihlali az olan çözüm tercih edilir.

4. Deneysel çalışma ve tartışma
(Experimental study and discussion)

Bu çalışmada kısıt işlemeye yöntemlerine HŞO ve BOA meta-sezgisel algoritmaları adapte edilmiştir. HŞO ve BOA meta-sezgisel algoritmaların performansını değerlendirmek amacıyla 10 kısıtsız (Zhang vd., 2017) ve 5 kısıtlı (Wu vd., 2017) benchmark fonksiyonu ve mühendislik problemi olan optimal güç akışı (Amaratunga vd., 2018) kullanılmıştır. Bu problemler çok boyutlu, doğrusal, doğrusal olmayan ve ikinci dereceden çeşitli amaç fonksiyonlarını içermektedir. Buna ek olarak kısıtlı fonksiyonların sınırları da doğrusal veya doğrusal olmayan problemlerdir. HŞO ve BOA algoritmaların sürü boyutları 300 ve iterasyon sayıları maksimum 2500 olarak belirlenmiştir. Algoritmaların istatiksel sonuçlarını değerlendirmek amacıyla her bir algoritma ve kısıt işleme teknikleri 20 bağımsız çalışma gerçekleştirilmiştir.

4.1. Deney 1: Benchmark Problemleri (Test 1: Benchmark Problems)

Bu çalışmada ilk olarak sezgisel algoritmalar kısıtsız optimizasyon yöntemlerinde kullanılmış ardından kısıtlı problemlerde kullanılan kısıtlı işleme teknikleri ilave edilerek kısıtlı optimizasyon problemlerine uygulanmıştır. Kullanılan sezgisel HŞO ve BOA algoritmalarının performanslarının istatistiksel olarak karşılaştırılabilmek amacıyla kullanılan kısıtsız benchmark problemlerinin değişkenlerinin aralıkları ve değişkenlerinin boyut değerleri Tablo 2'de verilmiştir (Dzubera et.al. 1996).

BOA ve HŞO algoritmaların kısıtsız benchmark fonksiyonlarına uygulandığında elde edilen ortalama(ort) ve standart sapma(std) sonuçları Tablo 3'de verilmiştir. Tablo 3'de görüldüğü gibi HŞO algoritması BOA'ya göre on problem içerisinde (F_1, $F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9, F_{10}$) global minimuma daha yakındır. BOA sadece F_7 fonksiyonunda global optimuma yaklaşmıştır.

Tablo 2. Kısıtsız benchmark problemleri (Unconstrained benchmark problems)

Fonksiyon	Boyut	Aralık
$F_1(x) = -20exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) - exp\left(\frac{1}{n}\sum_{i=1}^{n} cos(2\pi x_i)\right) + 20 + e$	30	[-35,35]
$F_2(x) = \sum_{i=1}^{D} \lvert x_i sin(x_i) + 0.1x_i \rvert$	30	[-10,10]
$F_3(x) = \lvert x_1^2 + x_2^2 + x_1x_2 \rvert + \lvert sin(x_1) \rvert + \lvert cos(x_2) \rvert$	2	[-500,500]
$F_4(x) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2$	2	[-4.5,4.4]
$F_5(x) = 1 - cos\left(2\pi\sqrt{\sum_{i=1}^{D} \lvert x_i^2 \rvert}\right) + 0.1\sqrt{\sum_{i=1}^{D} x_i^2}$	30	[-100,100]
$F_6(x) = (x_1 + 1.7x_2)sin(x_1) - 1.5x_3 - 0.1x_4 cos(x_4 + x_5 - x_1) + 0.2x_5^2 - x_2 - 1$	5	[-100,100]
$F_7(x) = -e^{-0.5}\sum_{i=1}^{D} x_i^2$	30	[-1,1]
$F_8(x) = \left(10000\left\lvert\sum_{i=1}^{D} x_i\right\rvert\right)^{0.5}$	30	[-10,10]
$F_9(x) = \sum_{i=1}^{n} ix_i^4 + random[0,1]$	30	[-1,28, 1,28]
$F_{10}(x) = \sum_{i=1}^{D-1} \lvert 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \rvert$	30	[-30,30]

Tablo 3. Kısıtsız benchmark fonksiyonları için HŞO ve BOA algoritmalarından elde edilen istatiksel sonuçlar (Statistical results obtained from HŞO and BOA algorithms for unconstrained benchmark functions)

		F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
Optimal		***0***	***0***	***1***	***0***	***0***	***-529,871***	***0***	***0***	***0***	***0***
HŞO	Ort	**8,88E-16**	**1,04E-276**	**1**	**0**	**0**	**-528,351**	**3,06E-07**	**0**	**4,17E-06**	**8,13E-06**
	Std	**0,00E+00**	**0**	**0**	**0**	**0**	2,8238	**1,63E-22**	**0**	**4,18E-06**	**1,33E-05**
BOA	Ort	2,84E-15	0,2969	**1**	2,99E-18	0,1148	-529,83	**3,06E-07**	**0**	0,000137	23,7434
	Std	1,81E-15	1,3279	**0**	3,25E-18	0,0586	**0,02605**	**1,63E-22**	**0**	0,000168	0,2557

Kısıtlı optimizasyonda HŞO ve BOA algoritmalarının performanslarının karşılaştırmak amacıyla kullanılan eşitlik ve eşitsizlik kısıtları içeren benchmark problemlerinin değişkenlerinin aralıkları, değişkenlerinin boyutları ve fonksiyonun minimum değerleri Tablo 3'de verilmiştir. Ölüm ceza, statik ceza, dinamik ceza, bariyer ceza ve Deb kuralları kısıtlama tekniklerine uyarlanmış HŞO ve BOA algoritmalarından elde edilen ortalama, standart sapma deneysel sonuçları ve optimal değerleri Tablo 5'te verilmiştir. Tablo 5'e göre kısıt işleme tekniklerinden ölüm ceza yönteminde beş kısıtlı fonksiyonda da ($F_{11}, F_{12}, F_{13}, F_{14}, F_{15}$) HŞO global optimal sonuca yaklaşmıştır. Statik ceza yönteminF_{11}, F_{12} ve F_{13} kısıtlı fonksiyonlarda en iyi performansa HŞO algoritması sahipken F_{14} ve F_{15} fonksiyonlarında BOA en iyi sonucu vermiştir. Dinamik ceza yönteminde $F_{11}, F_{12}, F_{13}, F_{14}$ ve F_{15} kısıtlı fonksiyonlarda HŞO global optimal yaklaşmıştır. Dinamik ceza yönteminde $F_{11}, F_{12}, F_{13}, F_{14}$ kısıtlı fonksiyonların HŞO algoritması en iyi performans gösterirken F_{15} kısıtlı fonksiyonunda ise her iki algoritmada benzer sonuç vermiştir. Son olarak Deb uygulanabilirlik yönteminde ise sonuç dinamik ceza yöntemiyle aynıdır. Bu sonuçlara ek olarak kısıt işleme yöntemleri içerisinde algoritmaların ortalama ve standart sapma değerleri bakımından performansları karşılaştırıldığında HŞO algoritması daha iyi sonuç verdiği görülmektedir.

4.2. Deney 2: Optimal Güç Akış Problemi (Optimal Power Flow Problem)

Kısıt işleme tekniklerinin performanslarını ölçmek ve karşılaştırmak amacıyla bir çok araştırmacı tarafından bilinen ve optimizasyon literatüründe karşılaşılan yaygın bir problem olan Optimal Güç Akışı (OGA) ele alınmıştır. OGA problemindeki karmaşık güç akışı eşitlikleri, eşitsizlik kısıtlamaları, enerji sistemindeki fiziksel kısıtlamaları gibi kısıtların fazla olmasından

kaynaklanmaktadır. Bu karmaşıklık eşitlik veya eşitsizlik olabileceğinden bu değerler enerji üretim maliyetini etkilemektedir. OGA probleminde kısıtlar sağlanarak yakıt maliyetini minimum olacak şekilde minimizasyon gerçekleştirilecektir. Yakıt maliyetini minimum olarak hesaplanmasında ölüm ceza, statik ceza, dinamik ceza, bariyer ceza ve Deb uygulanabilirlik kısıtlama tekniklerine uyarlanmış HŞO ve BOA algoritmaları uyarlanmıştır.

Bu problemde literatürde çokça karşılaşılan IEEE30 baralı optimal güç akışı problemi ele alınmıştır. Bu problem amaç fonksiyonu, kısıtlar ve ağ yapısına bağlı olacak şekilde farklı şekilde modellenerek kullanılabilir. Bu çalışmada toplam yakıt maliyetini en aza indirecek şekilde Denklem 23'te görüldüğü şekilde matematiksel modeli çıkartılmıştır. Denklem 23'teki n sistemdeki toplam jeneratör sayısını, P_i i. bara da üretilen aktif güçleri, a_i, b_i ve c_i jeneratör yakıt maliyeti katsayılarını göstermektedir.

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{N} F_i(P_i) \\ &F_i(P_i) = a_i P_i^2 + b_i P_i + c_i \\ &P_i^{\min} \le P_i \le P_i^{\max} \end{aligned} \tag{23}$$

Üretilecek olan güç (P_i), tüketilecek güç (D) ve iletim kayıplarının (P_l) toplamı olmalıdır. Dolayısıyla iletim kayıplarının hesaplanması Denklem 24 ile hesaplanmaktadır.

$$\begin{aligned} &\sum_{i=1}^{N} P_i = D + P_l \\ &P_i = real\left(\sum_{j}^{N} V_i Y_{ij} \overline{V}_j\right) \quad i = 1,2,...,N \\ &Q_i = imag\left(\sum_{j}^{N} V_i Y_{ij} \overline{V}_j\right) \quad i = 1,2,...,N \end{aligned} \tag{24}$$

Tablo 4. Kısıtlı benchmark problemleri (Constrained bencmark problems)

Kısıtlı Fonksiyon Eşitlikleri	Boyut	F_{min}
$F_{11}(x) = 5\sum_{i=1}^{4} x_i - 5\sum_{i=1}^{4} x_i^2 - \sum_{i=5}^{13} x_i$ $g_1(x) = 2x_1 + 2x_2 + x_{10} + x_{11} - 10 \le 0 \quad g_2(x) = 2x_1 + 2x_3 + x_{10} + x_{11} - 10 \le 0$ $g_3(x) = 2x_2 + 2x_3 + x_{11} + x_{12} - 10 \le 0 \quad g_4(x) = -8x_1 + x_{10} \le 0$ $g_5(x) = -8x_2 + x_{11} \le 0 \quad g_6(x) = -8x_3 + x_{12} \le 0$ $g_7(x) = -2x_4 - x_5 + x_{10} \le 0 \quad g_8(x) = -2x_6 - x_7 + x_{11} \le 0$ $g_9(x) = -2x_8 - x_9 + x_{12} \le 0 \quad 0 \le x_i \le 1 \quad i = 1,2,\ldots,9$ $0 \le x_i \le 100 \quad i = 10,11,12 \quad 0 \le x_{13} \le 1$	13	-15
$F_{12}(x) = -(\sqrt{n})^n \prod_{i=1}^{n} x_i$ $h_1(x) = \sum_{i=1}^{n} x_i^2 - 1 = 0 \quad 0 \le x_i \le 1 \quad i = 1,2,\ldots,10$	10	-1.0005000
$F_{13}(x) = 5.3578547x_3^2 + 0.8356891x_1x_5 + 37.293239x_1$ $g_1(x) = 85.334407 + 0.0056858x_1x_5 + 0.0006262x_1x_4 - 0.0022053x_3x_5 - 92 \le 0$ $g_2(x) = -85.334407 - 0.0056858x_1x_5 - 0.0006262x_1x_4 + 0.0022053x_3x_5 \le 0$ $g_3(x) = 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + 0.0021813x_3^2 - 110 \le 0$ $g_4(x) = -80.51249 - 0.0071317x_2x_5 - 0.0029955x_1x_2 - 0.0021813x_3^2 + 90 \le 0$ $g_5(x) = 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3x_5 - 25 \le 0$ $g_6(x) = -9.300961 - 0.0047026x_3x_5 - 0.0012547x_1x_3 - 0.0019085x_3x_5 + 20 \le 0$ $78 \le x_1 \le 102$ $33 \le x_2 \le 45$ $27 \le x_i \le 45 \quad i = 2,3,5$	5	-3.06655x10^4
$F_{14}(x) = (x_1 - 10)^3 + (x_2 - 20)^3$ $g_1(x) = -(x_1 - 5)^2 + (x_2 - 5)^2 + 100 \le 0$ $g_2(x) = -(x_1 - 6)^2 + (x_2 - 5)^2 - 82.81 \le 0$ $13 \le x_1 \le 100$ $0 \le x_2 \le 100$	2	-6961.813875
$F_{15}(x) = (x_1)^2 + (x_2 - 1)^2$ Subject to $h_1(x) = x_2 - x_1^2 \quad -1 \le x_1, x_2 \le 1$	2	0.7499

Tablo 5 . Kısıtlı Benchmark Problemlerinin Optimizasyon Sonuçları (Optimization Results of Constrained Benchmark Problems)

Algoritma	Kısıt	Kriter	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}
			-15	***-1.0005***	***-3.06655x10^4***	***-6961.813875***	***0.7499***
BOA	Ölüm Ceza	Ort	-9,92257	-0,00329	-30375,4	-6963,21	0,749001
		Std	3,051379	0,014556	182,6252	1,407344	6,48E-07
HŞO		Ort	***-14,5482***	***-0,10744***	***-30580***	***-6964,14***	***0,749015***
		Std	0,878496	0,090828	144,3007	1,14E-06	4,39E-05
BOA	Statik	Ort	-9,24107	-1,86E-68	-30413,4	***-6963,47***	***7,49E-01***
		Std	2,054122	0,057546	168,4201	0,46082	7,01E-04
HŞO		Ort	***-14,7276***	***-0,01824***	***-30561,8***	-6961,81	1,00E+00
		Std	0,444701	4,55E-68	173,8887	4,13E-06	8,82E-17
BOA	Dinamik	Ort	-9,44737	-7,67E-69	-30429,3	-6961,24	9,73E-01
		Std	2,916546	1,71E-68	156,9226	0,364233	6,92E-02
HŞO		Ort	***-14,8243***	***-8,72E-03***	***-30640,3***	***-6961,81***	***9,11E-01***
		Std	0,218548	2,33E-02	53,86185	7,27E-07	9,99E-02
BOA	Bariyer	Ort	-7,43006	-0,00114	-30448,8	-6961,33	***1,00E+00***
		Std	4,990804	0,004272	150,9557	0,284288	0,00E+00
HŞO		Ort	***-14,8398***	***-0,00226***	***-30560,6***	***-6961,81***	***1,00E+00***
		Std	0,102318	0,010125	161,9542	2,03E-06	5,09E-17
BOA	Deb	Ort	-9,51114	-0,01077	-29725,3	-5607,19	***1,00E+00***
		Std	2,122866	0,026659	254,2305	2411,676	0
HŞO		Ort	***-11,6678***	***-0,01841***	***-30526,9***	***-6961,8***	***1,00E+00***
		Std	1,962415	0,05898	106,9022	0,042547	0

OGA yakıt maliyeti kısıt problemi, kısıt işleme tekniklerine uyarlanmış HŞO ve BOA algoritmaları ile çözümünden elde edilen minimum, ortalama, standart sapma deneysel sonuçları Tablo 6'da verilmiştir.

Tablo 6. OGA'da yakıt maliyet minimizasyonunun istatiksel sonuçlarının karşılaştırılması ($/sa) (Comparison of statistical results of fuel cost minimization in OPF)

Kısıt Yöntemi	HŞO			BOA		
	Min. *($/sa)*	Ort. *($/sa)*	Std	Min. *($/sa)*	Ort. *($/sa)*	Std
Ölüm Ceza	800.40	**805.0**	6.600	805.39	813.00	7.14
Statik Ceza	799.92	**804.0**	5.750	801.72	810.00	6.77
Dinamik Ceza	799.96	**802.0**	2.120	805.26	813.00	6.35
Bariyer Ceza	799.94	**802.0**	2.120	801.78	810.00	4.77
Deb Kuralı	800.37	**803.0**	1.908	801.82	805.38	2.24

Tablo 6'a göre OGA kısıt probleminde ölüm ceza, statik ceza, dinamik ceza, bariyer ceza ve Deb uygulanabilirlik kuralı kısıt işleme tekniklerinde HŞO global optimale yaklaşmıştır. Buna ek olarak kısıt işleme tekniklerinden HŞO algoritmasında dinamik ceza ve bariyer ceza, BOA algoritmasında ise Deb uygulanabilirlik kuralı daha etkili olduğu görülmüştür.

BOA ve HŞO algoritmaların optimal güç akışında gösterdikleri performanslarını karşılaştırmak amacıyla literatürde bulunan HŞO (Islam vd., 2020), Gri Kurt Optimizasyon (GKO) (Islam vd., 2020), BOA (Islam vd., 2020), GA (Bouktir vd., 2008), Harris Şahini-Diferansiyel Evrim (HŞODE) (Birogul, 2019) ve Modifiyeli diferansiyel evrim (Sayah vd., 2008) algoritmaları kullanılmıştır. Çalışmada kullanılan BOA ve HŞO algoritmaları literatürdeki algoritmalarla karşılaştırıldığında HŞO algoritması en iyi değer bakımından daha iyi sonuç vermiştir. Bunun nedeni bu algoritmaların rastlantısal olarak çalışması ve belirlenen iterasyon ve sürü sayıları farklı olması ve genellikle kısıt işleme tekniklerinin ve parametrelerinin değişebilmesinden kaynaklanmaktadır. Bu doğrultuda bağıl bir değerlendirme yapıldığında literatürde genellikle kısıtlı problemler ölüm, statik ve dinamik ceza fonksiyonları çözümlenmektedir.

Tablo 7. Optimal güç akışı için HŞO, GKO, BOA, GA, HŞODE ve MDA algoritmalarıyla elde edilen optimal çözümler (The minimum solutions obtained by the HŞO, GKO, BOA, GA, HŞODE, and MDA algorithms for optimal power flow)

Algoritma	**Minimum Değer ($/sa)**
HŞO (Islam vd., 2020)	801.8290
GKO (Islam vd., 2020)	801.8440
WOA (Islam vd., 2020)	801.8400
GA (Bouktir vd., 2008)	804.1000
HHODE (Birogul 2019)	800.9959
MDE (Sayah vd., 2008)	802.3700

Fakat Tablo 6'da görüleceği üzere bariyer ceza fonksiyonu ortalama olarak daha iyi sonuç vermiştir. Lakin bu çalışmada yapılan denemelerde sadece minimum değerler baz alınarak kıyas yapıldığında ise Tablo 6 ve Tablo 7 incelendiğinde Tablo 6'da bulunan HŞO ile kullanılan Statik Ceza fonksiyonu literatürdeki çalışmalardan daha iyi sonuç verdiği görülmektedir.

5. Sonuçlar (Results)

Bu çalışmada doğadan ilham alan meta-sezgisel optimizasyon algoritmaları için etkili bir kısıt işleme teknikleri sunmaktadır. Ölüm cezası, statik ceza, dinamik ceza, bariyer ceza ve Deb kuralları kısıtlama teknikleriyle BOA ve HŞO algoritmaları kısıtsız ve kısıtlı optimizasyon problemlerinin optimize edilmesinde kullanılmıştır ve performansları karşılaştırılmıştır. Ayrıca bu algoritmaların performansları literatürde var olan IEEE30 baralı optimal güç akışı minimizasyon problemine başarıyla uygulanmıştır. Deneysel sonuçlara göre kısıtlı ve kısıtsız benchmark problemlerinde ve optimal güç akışı mühendislik probleminde HŞO algoritmasının

performansı daha iyi olduğu istatistiksel olarak görülmüştür. Ayrıca kısıt işleme yöntemlerinin BOA ve HŞO algoritmaları üzerinde efektif olduğu görülmüştür.

Kaynaklar (References)

Akay B., Karaboga D., 2011. A modified Artificial Bee Colony (ABC) algorithm for constrained optimization problems. Applied Soft Computing, 11, 3021–3031.

Aljarah, I., Chen, H., Faris, H., Heidari, A. A., Mafarja, M., Mirjalili, S. 2019. Harris hawks optimization: Algorithm and applications. Future Generation Computer Systems, 97, 849–872.

Amaratunga, G.A.J., Biswas, P.P., Mallipeddi, R., Suganthan P.N., 2018.Optimal power flow solutions using differential evolution algorithm integrated with effective constraint handling techniques. Engineering Applications of Artificial Intelligence, 68, 81–100.

Babalik, A., Cinar, A.C., Kiran M.S., 2018, A modification of tree-seed algorithm using Deb's rules for constrained optimization. Applied Soft Computing, 63, 289–305.

Batık, Z. G., Boz, A. F., Çimen, M.E., Karayel D.2019. The Chaos-Based Whale Optimization Algorithms Global Optimization, Chaos Theory and Applications, 1, 51-63.

Birogul, S. 2019. Hybrid harris hawk optimization based on differential evolution (HHODE) algorithm for optimal power flow problem. IEEE Access, 7, 184468-184488.

Bouktir, T.,Slimani, L., Mahdad, B. 2008. Optimal power dispatch for large scale power system using stochastic search algorithms. International Journal of Power and Energy Systems, 28(2), 118.

Chen S., Gu Y., Jiang, S., Nouioua, M., Li Z., Zhang, S., 2019. FSB-EA: Fuzzy search bias guided constraint handling technique for evolutionary algorithm. Expert Systems with Applications, 119, 20–35.

Chen, H., Wang, M., Zha X., 2020. A multi-strategy enhanced sine cosine algorithm for global optimization and constrained practical engineering problems. Applied Mathematics and Computation, 369, 124872.

Deb, K., Gandomi, A.H., 2020 .Implicit constraints handling for efficient search of feasible solutions. Comput. Methods Appl. Mech. Engrg. 363, 112917.

Dzubera, J., Mathias, K., Rana S., D. Whitley, 1996. Evaluating Evolutionary Algorithms. Artif Intell, 85, 245-276.

Fan, Q.-W., He, X.-S., Karamanoglu M. and Yang, X.-S., 2019. Comparison of Constraint-Handling Techniques for Metaheuristic Optimization. ICCS 2019: Computational Science, 357-366.

Garcia R., Jacob, B. P., Lemonge, A., Lima, B., 2017. A rank-based constraint handling technique for engineering design optimization problems solved by genetic algorithms. Computers and Structures, 187, 77–87.

Guimarães, S., Lima, B., Rodrigues, M., 2018. E-BRM: A constraint handling technique to solve optimization problems with evolutionary algorithms. Applied Soft Computing, 72, 14–29.

Haklı H., 2019. A novel approach based on elephant herdıng optımızatıon for constraıned optımızatıon problems. Selcuk Univ. J. Eng. Sci. Tech., 7,405-419.

Hatamlou, A., Mirjalili, S., Mirjalili, S. M., 2016. Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Computing and Applications, 27, 495-513.

He W., Peng, X., Peng, X., Qu, C. 2020. Harris Hawks optimization with information Exchange, Applied Mathematical Modelling, 84, 52–75.

Islam, M.Z., Abdul Wahab, N.I., Veerasamy, V., Hizam, H., Mailah, N.F., Guerrero, J.M., Mohd Nasir M.N., 2020. A Harris Hawks Optimization Based Single-and Multi-Objective Optimal Power Flow Considering Environmental Emission. Sustainability, 12(13), 5248.

Jelovic, J., Samanipour, F., 2020. Adaptive repair method for constraint handling in multi-objective genetic algorithm based on relationship between constraints and variables. Applied Soft Computing Journal, 90, 106-143.

Kohler, M., Tanscheit, R., Vellasco, M.M.B.R., 2019. PSO+: A new particle swarm optimization algorithm for constrained problems. Applied Soft Computing, 85, 105865.

Lewis, A., Mirjalili, S., 2016. The whale optimization algorithm. Advances in Engineering Software, 95, 5167.

Mert, A., Tezel, B.T., 2021. A cooperative system for metaheuristic algorithms. Expert Systems with Applications, 165, 113976.

Miranda-Varela, M.,, Mezura-Montes, E. 2018. Constraint-handling techniques in surrogate-assisted evolutionary optimization. An empirical study. Applied Soft Computing Journal 73, 215–229.

Sayah, S., Zehar, K. 2008. Modified differential evolution algorithm for optimal power flow with non-smooth cost functions. Energy conversion and Management, 49(11),3036-3042.

Tuba, M. and Bacanin, N., 2014. Improved seeker optimization algorithm hybridized with firefly algorithm for constrained optimization problems. Neuro computing, 143, 197-207.

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) (2021) 86-93
DOI: 10.38016/jista.852099

Solution Representation in Proportionate Multiprocessor Open Shop

Zeynep Adak [1]

[1]Gebze Technical University, Department of Industrial Engineering, Kocaeli, Turkey

zeynepadak@gtu.edu.tr

Abstract
Proportionate multiprocessor open shop is considered in this study. It is a shop model where a set of jobs follow no predefined route to visit several stages with at least one having two or more parallel machines to carry out the same task. Proportionality means that processing times depend only on stages and are independent of jobs, hence is defined as stage-wise. The shop model has various application areas in industry but the literature on the field is still limited. In this study, a novel solution representation scheme is proposed for the proportionate multiprocessor open shop. The scheme is based on permutation of stages and encodes the cumulative number of job assignments to a stage. The proposed scheme is shown to generate higher quality random solutions compared to the common operation permutation representation for the shop model. The approach proposed in this study to design a solution representation for a scheduling problem is a new and favorable approach that takes into account the specific machine environment, job characteristics and objective function of the problem under consideration. This way of designing solution representation schemes would increase the solution quality or decrease the computational time required in solution algorithms for scheduling problems.

Keywords: Multiprocessor open shop, solution representation, implicit-stage representation, scheduling, proportionate

Orantılı Esnek Açık Atölye Tipinde Çözüm Gösterimi

Öz
Bu çalışmada orantılı esnek açık atölye tipi ele alınmıştır. Bu atölye modelinde yapılacak işlerin işlem istasyonlarını gezerken takip edecekleri bir rota bulunmaz ve bu istasyonların en az birinde aynı işlemi yapan iki veya daha fazla paralel makine bulunur. Orantılı ifadesi işlem sürelerinin istasyona bağlı olduğunu ve işten bağımsız olduğunu ifade eder. Böylece, orantılılık istasyon-bazlı olarak tanımlanmıştır. Bu atölye modelinin endüstride farklı uygulama alanları mevcuttur ancak bu alandaki literatür hala kısıtlıdır. Bu çalışmada, orantılı esnek açık atölye tipi için yeni bir çözüm gösterimi önerilmiştir. Önerilen gösterim istasyonların permutasyonuna dayanır ve bir istasyona yapılan kümülatif iş ataması sayısını şifreler. Önerilen bu yeni gösterim, bu atölye tipinde yaygın olarak kullanılan operasyon-permütasyonu gösteriminden daha iyi kalitede rastgele çözümler üretmiştir. Bu çalışmada çizelgeleme probleminde çözüm gösterimi tasarımında kullanılan yaklaşım yeni ve sonuçlar bakımından olumludur. Bu yaklaşım eldeki probleme özel makine ortamını, iş özelliklerini ve amaç fonksiyonunu dikkate alır. Çözüm gösterimi tasarlanmasında izlenen bu yol çizelgeleme problemlerinde kullanılan çözüm algoritmalarının daha yüksek kalitede çözüme ulaşmasını veya harcanan hesaplama zamanının kısaltılmasını sağlayacaktır.

Anahtar Kelimeler: Esnek açık atölye, çözüm gösterimi, istasyon gösterimi, çizelgeleme, orantılı

1. Introduction

Solution or schedule representation is a key element in dealing with scheduling of shop models. It is a way to represent a feasible schedule for the shop. Feasibility refers to the condition that a given schedule meets all model definitions in terms of machine environment and job characteristics and does not violate any constraints imposed in the model.

Scheduling problems are one large class of combinatorial optimization problems that are thoroughly studied in the literature. It is about creating a feasible schedule for a given shop model to minimize

* Corresponding Author.
E-mail: zeynepadak@gtu.edu.tr

Recieved : 01 January 2021
Revision : 23 March 2021
Accepted : 04 April 2021

a certain objective or multiple objectives. Since most of scheduling problems are NP-Hard in nature (Chen et al., 1998) they require unreasonable amount of computational time to find an optimal solution by enumerative algorithms. Instead, approximation algorithms and more recently search algorithms are used to find a good enough solution.

Search algorithms are named as metaheuristics and they mainly make an "educated" search in the large solution space of the problem. The search is educated since it uses problem knowledge and previous search experience. The search routine works around numerous solutions until a predefined termination criterion is met. Hence, solution (schedule) representation is an initial important step in those algorithms. Since the algorithm subroutines, neighborhood structures for instance, are defined over the solution representation, choosing a proper and an efficient one may even increase the solution quality of the algorithm and decrease the computational time required.

In this study, proportionate multiprocessor open shop (MPOS) with makespan minimization criterion is considered, and a novel efficient solution representation is proposed for the shop model. Inefficiencies due to the conventional representation in the literature are pointed out and the proposed representation is introduced in detail with its encoding and decoding procedures. Efficiency of the proposed representation is presented by making a basic random search for the solution of a set of benchmark instances. The results are compared with a random search by the conventional representation and it is shown that the solution quality reached with the proposed representation is remarkably higher.

The paper is organized as follows. The proportionate MPOS is defined in Section 2 and the literature is reviewed in terms of solution representations for the problem. Section 3 presents the conventional solution representation for the problem in the literature and states the inefficiencies it causes. The proposed solution representation is introduced in Section 4. Computational experiments and comparisons are presented in Section 5. Section 6 discusses the proposed representation and the results. Conclusive remarks are given in Section 7.

2. Proportionate Multiprocessor Open Shop

Multiprocessor open shop (MPOS) is a machine environment with a set of stages (machine centers), $S = \{1, 2, \dots, s\}$, where stage i has $m_i \geq 1$ machines in parallel and at least for one stage $m_i > 1$. Every stage carries out a different task and the machines of a stage are assumed to be identical in this study. There are a number of jobs, $J = \{1,2, \dots, n\}$, to be processed in the stages, but there is no route for the jobs to visit the stages and it is what makes the shop an *open* shop. A job is processed on a single machine and a machine processes a single job at a time. Operation O_{ji} is defined as the processing of job j on machine i. In this study, it is assumed that all jobs are processed in all stages and the objective function is to minimize the makespan, the time all operations of the shop are completed.

Proportionate MPOS is considered in this study. Proportionate property for MPOS is defined by Matta (2009) and it refers to processing times being based on stages and not both on stages and jobs. More precisely, a stage i processes any job in the same amount of p_i time, hence $p_{ji} = p_i \;\; j \in J, \;\; i \in S$, where p_{ji} is the processing time of job j in stage i. Thus, the scheduling problem considered in this study is represented in three-field notation as $O(P)|\, p_{ji} = p_i|C_{max}$.

The proportionate property definition for MPOS is in contrast with the widely accepted definition for proportionate flow, job, and open shops. Proportionality was introduced as job-wise for flow shops (Ow, 1985), defining $p_{ji} = p_j$. This job-wise proportionality definition has also been accepted in job shops and open shops (Pinedo, 2016). Proportionality was taken as machine-wise in rare studies in flow, job and open shops. However, it is always considered as stage-wise, $p_{ji} = p_i$, in MPOS literature due to the real applications of proportionate MPOS, as given next.

MPOS is seen in several industrial settings such as medical testing facilities in health care, auto repair and maintenance centers, electronics manufacturing, and inspection and quality control operations. The proportionate case, on the other hand, is more common in healthcare medical testing where it takes the same amount of time to carry out a test, independent of patients.

Gonzalez and Sahni (1976) showed that $O_m||C_{max}$ is NP-Complete for $m > 2$. The result applies for MPOS, $O(P)||C_{max}$, which is the generalized version of the classical open shop. Further, Mao (1995) showed that $O_2(P_k)||C_{max}$ is NP-Complete even when there are $k = 2$ machines at each stage. However, the complexity status of the proportionate case, $O(P)|\, p_{ji} = p_i|C_{max}$, has not been studied in the literature yet.

It should be noted that despite the various industrial applications of the shop model, the literature on MPOS is still very limited. Since the current study is about solution representation in proportionate MPOS, different approaches used in the literature to represent solutions in MPOS are reviewed here. A solution for the problem should supply the job routes to visit the stages as well as the machine sequences at every stage. The most common scheme used in representing schedules of MPOS is the operation permutation representation. It was first proposed by Liaw (2000) for the classical open shop and adapted to MPOS by Matta (2009). Later, it was also used by other researchers studying in MPOS (Naderi et al., 2011, Goldansaz et al., 2013, Azadeh et

al., 2014, Bai et al., 2016). Definition of the representation is given in the next section.

There were also a few other schemes used for solution representation in MPOS scheduling problem. Zhang et al. (2019) used a two-level hierarchical scheme where the first level supplied the information about job routes and the second level specified the machine to be used for the corresponding job and in the corresponding stage given by the first level. Abdelmaguid et al. (2014) used a disjunctive graph representation where every O_{ji} was represented by a separate node in the graph with a weight equal to p_{ji}. Stage visiting route for job j was represented by solid arcs between $O_{ji}, i \in S$, and processing sequences of machines at stage i were represented by dashed arcs between $O_{ji}, j \in J$. Abdelmaguid (2020) used two sets of vectors to represent a single solution of MPOS. One set of vectors had n vectors to define the stage visiting route for every job explicitly, and the other set of vectors had a total of $\sum_{i \in S} m_i$ vectors to define the processing sequence in every machine at every stage.

Interested reader is referred to Adak et al. (2020) for a detailed review of the literature on MPOS scheduling problem.

3. Operation Permutation and Its Inefficiencies in Proportionate MPOS

Operation permutation is a permutation of operations O_{ji}, as its name implies. A sample operation permutation is as the following for the sample 3-stage proportionate MPOS problem given in Table 1.

Table 1. Sample proportionate MPOS problem

$n = 4$

	Stage 1	Stage 2	Stage 3
p_i	5	2	2
m_i	2	1	1

$$O_{31}O_{21}O_{32}O_{12}O_{41}O_{23}O_{22}O_{42}O_{33}O_{13}O_{11}O_{43} \quad (1)$$

The permutation in (1) is decoded by reading it from left to right and assigning the respective operation to the first available machine at the respective stage. Accordingly, the schedule represented by (1) is given in Figure 1.

Several problems due to using operation permutation representation in proportionate MPOS are stated through the schedule encoded by the sample permutation. However, these problems do not apply only to the current sample permutation, but they rather represent the general nature of the schedules resulting from using the operation permutation scheme for the proportionate MPOS.

The schedule in Figure 1 has idle times in machines which causes a great increase in the makespan. Indeed, idle machine times are part of most optimum schedules. However, the idle times in the schedule in Figure 1 are undesirable (unnecessary) since there are jobs which can be started earlier and finished earlier in a machine without delaying start time of any job. This type of schedules is called inactive. However, an optimum schedule is known to be an active schedule.

The idle times are resulting from the job routes and machine sequences imposed by the permutation. The permutation encodes the following job routes and machine sequences:

Job routes to visit the stages: Job 1: [2-3-1]; Job 2: [1-3-2]; Job 3: [1-2-3]; Job 4: [1-2-3].

Job sequences in machines: Stage 1, machine 1: [3-4-1]; Stage 1, machine 2: [2]; Stage 2, machine 1: [3-1-2-4]; Stage 3, machine 1: [2-3-1-4]. Note that jobs can be interchanged in machines of stage 1, as long as the job order is preserved.

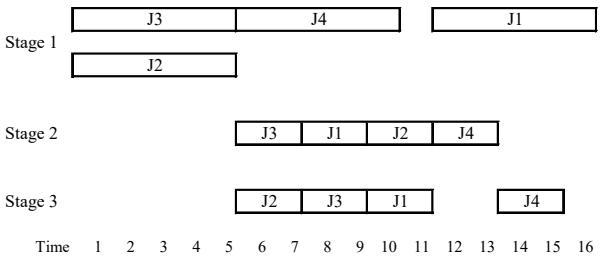


Figure 1. Schedule encoded by the sample operation permutation

The problem with the resulting schedule can be addressed by one of two approaches. First, it can be converted to an active schedule by moving the jobs to earlier times on machines, though not delaying any job's start time of processing on any machine. Applying this post-processing on the schedule, the enhanced schedule in Figure 2 is obtained. A 6 time-unit decrease in the makespan is achieved and an optimal schedule is received since stage 1 is working with full capacity until the completion time of the schedule. However, this post-processing causes inefficiencies in using the representation for the problem.

One inefficiency is due to the computational time required to convert an inactive schedule to an active one. Considering the thousands of schedules visited by search algorithms, performing such an additional time-consuming activity for every schedule generated is not favorable. Another inefficiency of a possible post-processing is that the ultimate schedule reached would be different from what the permutation encodes. This inconsistency between the encoded information and the decoded result poses a challenge in particularly memory-based algorithms. Those algorithms keep good solution characteristics -defined over solution

components represented by the permutation- in their memory and recall them later to build good quality solutions. Almost all metaheuristics use some kind of memory as part of their solution methodology.

Another approach to get rid of the unnecessary idle times on machines is to schedule a job as early as possible during decoding even it is in a later position in the permutation. This results in a non-delay schedule. However, the approach was shown (Naderi et al., 2011) not to increase the solution quality, even decrease it.

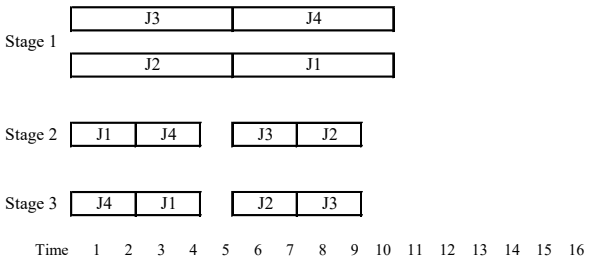


Figure 2. Enhanced schedule by post-processing

A second important problem with using operation permutation to represent proportionate MPOS schedules is the job restrictions present in the permutation. That is, the permutation forces a certain job to be scheduled for the respective stage. This reduces the flexibility present in the shop model and may lead to schedules with increased makespan values. To make the statement clearer, consider Figure 3 which contains 6 different schedules with the same makespan value of 10 for the sample problem. Even more schedules can be generated similarly. A stage processes any job in the same amount of time, and scheduling different jobs make no difference in the makespan as long as the schedule template is the same. This notion of templates for schedules are discussed further in the upcoming paragraphs.

Lastly, one implication drawn out by analyzing proportionate MPOS schedules is presented next, which constituted an important element in constructing the proposed representation.

Consider the dense schedule given in Figure 4 for a 2-stage 33-job problem. There are 23 machines at stage 1 and 10 machines at stage 2. Processing times are 10, and 4 time-units for stages 1 and 2, respectively. A dense schedule leaves no machine idle if there is any job waiting to be processed. Allocating 23 jobs to stage 1 at the beginning of the schedule makes machines of stage 2 wait idle until time 10 and to process 23 jobs after that time, which causes a makespan of 22. Instead, if the schedule in Figure 4 is enhanced as in Figure 5, stage 2 would be able to process another 10-job until time 10 and this will decrease the makespan by 2 time-units, leading to optimal makespan of 20.

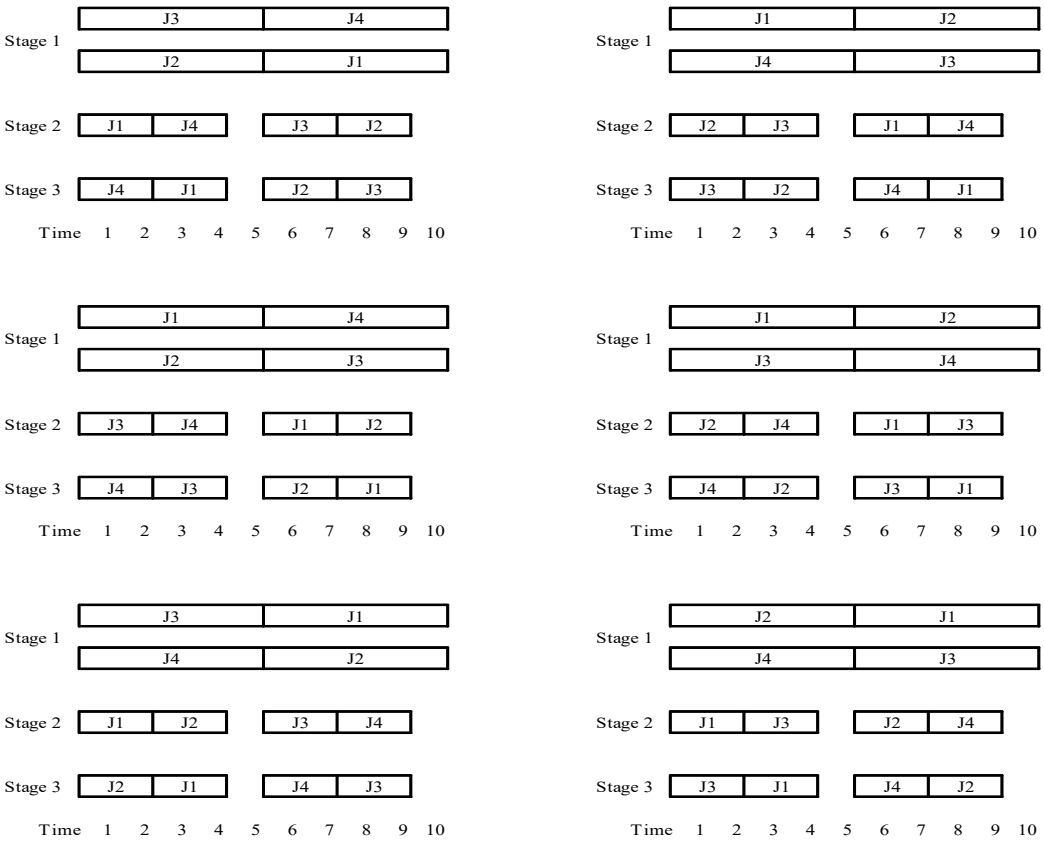


Figure 3. Different schedules with same makespan

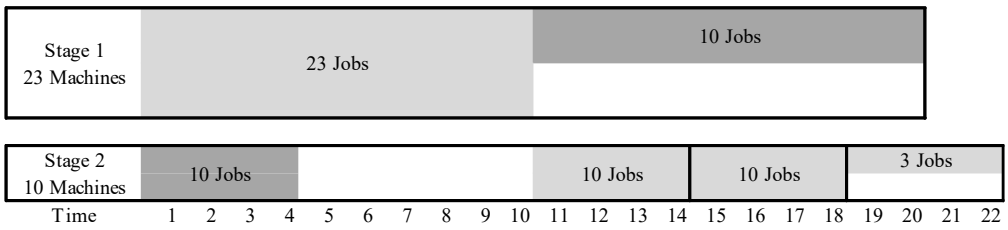


Figure 4. Dense schedule

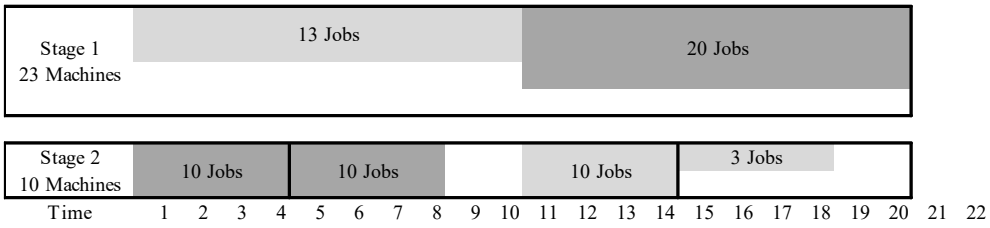


Figure 5. Enhanced schedule

An important implication to draw out from this analysis in schedules is the following. What makes a schedule better than another in terms of makespan objective is the number of jobs allocated to a stage in a block, where block is a p_i-time period at a stage. Thus, to differentiate between good and bad schedule characteristics one should hold this information in the solution representation. That would allow to correctly keep good solution characteristics in the memory of a search algorithm. Again, operation permutation fails to encode this information while representing a proportionate MPOS solution.

3.1. Schedule templates

The schematic schedule representations in Figures 4 and 5 are constructed as schedule templates which do not have explicit calls to job identities. Instead, there exists blocks of job bundles. Numerous different schedules can be generated using a schedule template by assigning different jobs to blocks. This template consideration for schedules is again a novel approach, and it constitutes an important base for the proposed solution representation. The schedule template approach is further discussed in the discussion section of the paper.

4. Novel solution representation: Implicit-stage permutation

To eliminate the inefficiencies caused by using operation permutation to represent a solution of a proportionate MPOS, a novel solution representation scheme is proposed here. This new representation is also able to encode good solution characteristics more effectively and allow for better memory models.

The proposed solution representation for the proportionate MPOS is based on a *stage permutation*. A stage permutation is introduced as a permutation of s stages where each stage repeats n times. Remind that it is assumed in this study that every job is processed at every stage. However, it is straightforward to modify the representation to allow for some jobs not to be processed at some stages. Simply, these stages would repeat less than n times in the permutation.

Following is a sample stage permutation for the sample problem in Table 1.

3 1 3 2 2 1 1 3 2 1 2 3 (2)

The stage permutation is decoded from left to right and *a job* is assigned to the respective stage. The job is selected from an eligible job set consisting of the jobs still waiting to be processed at that stage and are not currently processed at another stage. If there is more than one job in the eligible set, then the job with the lowest desirability across stages is selected. This way, it is aimed to allow for non-empty future eligible job sets as much as possible. If job desirability values are equal among the eligible set, then the jobs are assigned in numerical order. The selected job is assigned for the respective stage to earliest available machine.

The introduced stage permutation scheme for proportionate MPOS solves the flexibility issues mentioned before about the operation permutation as it includes no explicit job calls. This also enables active schedules, importance of which is also emphasized earlier in the paper. However, stage permutation does

not supply the information about the number of jobs to be allocated to a stage block, as it is shown to be a distinguishing characteristic of a schedule. Further, it includes repetition of elements, which makes it hard for memory models to extract patterns of good characteristics. Thus, the stage permutation scheme introduced here is proposed to be represented in a higher-level form where a number in permutation encodes the cumulative number of job assignments for the respective stage.

The stage permutation in (2) is converted to the representation in (3) using the encoding given in Table 2. This resulting form of solution representation is named as *implicit stage permutation*. It is implicit because a number in the permutation implicitly refers to two things: 1) the stage where a job is to be assigned, and 2) cumulative number of jobs assigned for that stage.

$$9 - 1 - 10 - 5 - 6 - 2 - 3 - 11 - 7 - 4 - 8 - 12 \quad (3)$$

The permutation is decoded as follows: 9: assign a *first* job to stage 3, 1: assign a *first* job to stage 1, 10: assign a *second* job to stage 3, 5: assign a *first* job to stage 2, and so on. The schedule given by the permutation is shown in Figure 6.

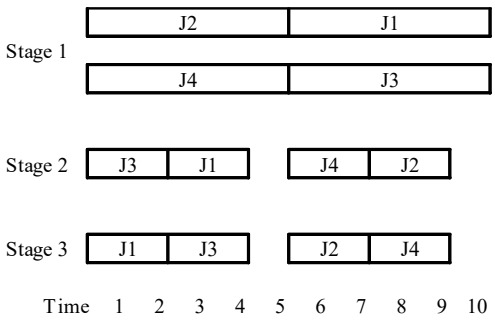


Figure 6. Schedule of the sample implicit stage permutation

Random solution generation

To have random implicit stage permutations, it would not be feasible to generate random numbers from 1 to $n \times s$. Because there is an ordering between cumulative number of assignments to a stage. That is, representation for a second assignment, for instance, should not precede the representation for the first assignment. Otherwise, the implicit stage permutation would not serve its purpose and become meaningless.

To generate random solutions, first a random stage permutation should be created and then it should be converted to an implicit stage permutation using Table 2. Creating a random stage permutation is straightforward as it is a permutation from 1 to s with every element repeating n times.

Table 2. Encoding to construct implicit stage permutation from a stage permutation

	1st assignment	2nd assignment	…	nth assignment
Stage 1	1	2	…	n
Stage 2	$n+1$	$n+2$	…	$2n$
Stage 3	$2n+1$	$2n+2$	…	$3n$
⋮	⋮	⋮	…	⋮
Stage i	$n(i-1)+1$	$n(i-1)+2$	…	$n \times i$
⋮	⋮	⋮	…	⋮
Stage s	$n(s-1)+1$	$n(s-1)+2$	…	$n \times s$

5. Computational Tests

The implicit stage permutation proposed in this study is aimed to be a solution representation for proportionate MPOS, particularly within a solution algorithm for the problem. For an effective ant colony optimization algorithm using the proposed solution representation see (Adak, 2020).

Developing a solution algorithm for the problem is out of the scope of this study. Rather, the favorable results due to the proposed solution representation are shown by random schedules. Random solutions are simple, quick but mostly inferior quality solutions for combinatorial optimization problems. It is highly unlikely to pick a good quality solution randomly from a very huge solution space. However, random solutions are used to initiate a search process in many algorithms. Starting with higher quality initial solutions may lead to increased performance or at least it decreases the computational time required to reach the best objective value.

In computational tests, quality of random solutions by the proposed solution representation are compared with the ones by the operation permutation representation. Table 3 reports the results of the computational tests. 12 test instances from a benchmark testbed (Matta, 2009) are used in the tests. 10 random solutions are generated for each of the instances, and minimum, average, and

maximum of the 10 solutions are given in the table. Optimum makespan values for the instances were reported in literature (Abdelmaguid, 2020) and are presented in the table. The table also gives the percent improvement in average makespan deviation from the optimum makespan, achieved by the proposed representation.

The results given in Table 3 suggests the ability of the proposed representation in generating higher quality solutions even in a random search. This is due to two features of the representation: 1) it does not enforce job identities in the permutation, thus does not restrict the flexibility already present in a proportionate shop model, and 2) it creates active schedules and avoids unnecessary idle times in the schedule.

Table 3. Comparison of quality of random solutions by operation permutation and by implicit stage permutation

					Operation-permutation (10 random permutations)			Implicit-stage permutation (10 random permutations)			
	Number of stages	Number of jobs	Max. number of machines in a stage	Optimum Makespan	Min. Makespan	Avr. Makespan	Max. Makespan	Min. Makespan	Avr. Makespan	Max. Makespan	% Improvement in deviation (Avr. makespan)
S4-P3	4	63	22	48	75	81	90	**51**	**54.3**	**57**	80.91
S4-P4	4	38	14	27	42	45.6	51	**36**	**36.3**	**39**	50.00
S4-P6	4	60	20	25	37	41.2	46	**26**	**27.8**	**28**	82.72
S4-P7	4	53	17	36	48	53.5	63	**36**	**36.5**	**38**	97.14
S8-P14	8	72	14	84	136	146	153	**86**	**90.2**	**96**	90.00
S8-P17	8	100	24	60	97	102.7	108	**60**	**62.1**	**64**	95.08
S8-P18	8	106	22	56	96	99.4	104	**58**	**61.2**	**66**	88.02
S8-P20	8	105	22	49	77	82	85	**51**	**52.8**	**58**	88.48
S16-P9	16	101	10	121	221	228.3	239	**126**	**133.9**	**143**	87.98
S16-P13	16	112	10	180	336	352.8	372	**195**	**207.6**	**213**	84.03
S16-P14	16	97	10	84	156	162.3	176	**90**	**94.7**	**101**	86.33
S16-P18	16	102	10	110	216	226.2	240	**122**	**125.6**	**130**	86.57

6. Discussion

The proposed representation for proportionate MPOS is an easy-to-use approach with straightforward encoding and decoding procedures. Complex representations of feasible solutions in attempting to solve a combinatorial problem makes it even harder and may render useless most of the time. Further, it may require additional computational time to decode a complicated representation and calculate the objective value, which is disadvantageous since search algorithms require to go over many solutions and assess the quality of every single solution. The simplicity of the proposed solution representation is very favorable in that aspect.

Instead of creating exact schedules where positions of jobs are explicitly stated, the proposed representation works through schedule templates. It states how many jobs would be allocated to distinct time slots, creates groups of jobs, and determines position of the groups at every stage in the final schedule. Different schedules having same template can be generated by using different rules of job selection from the eligible set during decoding the permutation. The template generating approach proposed in this study is particularly useful as it deals with solution families with same makespan value instead of single solutions. This decreases the problem size and facilitates searching the solution space.

It should be emphasized that the representation in this paper is proposed specifically for the proportionate MPOS with makespan criterion. The proposed representation encodes cumulative number of job assignments to a stage, since this information has a decisive role in makespan value.

However, a different objective function would require different considerations about the properties a solution representation should encode. Again, because of the proportionality of the stages the representation does not encode job identities. But if a non-proportionate shop is considered, then a solution representation should also encode job identities.

Since the proportionality definition is mostly jobwise in the literature for other shop models, the proposed

solution representation is not directly applicable in that problems. However, the proposed approach can be reconsidered for other proportionate models and can be adapted accordingly. The general idea in proposed solution representation is valuable in dealing with solutions of scheduling problems. It is based on constructing a representation that holds significant information about the machine environment, job characteristics and objective function. However, in scheduling literature, the common approach in selecting a scheme to represent solutions is based on generating complete, valid, and feasible solutions. The way of handling solutions as proposed in this study leads to a problem-specific approach for representation and results in more effective solution algorithms for scheduling problems.

7. Conclusion

A novel solution representation is proposed in this study for the proportionate multiprocessor open shop scheduling problem with makespan criterion. The representation, named implicit stage permutation, encodes the cumulative number of job assignments to a stage. It is a problem-specific scheme that focuses on the shop characteristics and the objective function. This feature of the proposed scheme makes it a powerful approach in representing solutions of the problem.

The performance of the proposed representation was compared with the conventional operation permutation representation for multiprocessor open shop scheduling problems by generating random solutions for benchmark instances from the literature. The proposed representation produced random solutions with a 11% deviation from the optimum in average while it was 77% in operation permutation, leading to an 84%-improvement in solutions. Its performance in random solutions is an indicator of its efficiency as a solution representation scheme for proportionate multiprocessor open shop problem.

An important conclusion to draw out from this study is that it suggests a new way of dealing with solutions in scheduling problems. More effective representation schemes can be constructed by considering the machine environment, job characteristics and the objective function of the scheduling problem at hand. Future studies should consider this approach in designing solution algorithms for various scheduling problems.

In future research, the proposed representation can be used to improve previous results on proportionate multiprocessor open shop problem with makespan criterion. Further, it can be adapted to other scheduling problems where machine-wise proportionality is considered.

References

Abdelmaguid, T. F. 2020. Scatter search with path relinking for multiprocessor open shop scheduling. *Computers & Industrial Engineering,* 141**,** 106292.

Abdelmaguid, T. F., Shalaby, M. A. & Awwad, M. A. 2014. A tabu search approach for proportionate multiprocessor open shop scheduling. *Computational Optimization and Applications,* 58**,** 187-203.

Adak, Z. 2020. *An ant colony optimization approach for the proportionate multiprocessor open shop.* PhD Doctoral thesis, Marmara University.

Adak, Z., Arıoğlu Akan, M. Ö. & Bulkan, S. 2020. Multiprocessor open shop problem: literature review and future directions. *Journal of Combinatorial Optimization,* 40**,** 547-569.

Azadeh, A., Hosseinabadi Farahani, M., Torabzadeh, S. & Baghersad, M. 2014. Scheduling prioritized patients in emergency department laboratories. *Computer Methods and Programs in Biomedicine,* 117**,** 61-70.

Bai, D., Zhang, Z.-H. & Zhang, Q. 2016. Flexible open shop scheduling problem to minimize makespan. *Computers & Operations Research,* 67**,** 207-215.

Chen, B., Potts, C. N. & Woeginger, G. J. 1998. A Review of Machine Scheduling: Complexity, Algorithms and Approximability. *In:* DU, D.-Z. & PARDALOS, P. M. (eds.) *Handbook of Combinatorial Optimization: Volume1–3.* Boston, MA: Springer US.

Goldansaz, S. M., Jolai, F. & Zahedi Anaraki, A. H. 2013. A hybrid imperialist competitive algorithm for minimizing makespan in a multi-processor open shop. *Applied Mathematical Modelling,* 37**,** 9603-9616.

Gonzalez, T. & Sahni, S. 1976. Open Shop Scheduling to Minimize Finish Time. *J. ACM,* 23**,** 665–679.

Liaw, C.-F. 2000. A hybrid genetic algorithm for the open shop scheduling problem. *European Journal of Operational Research,* 124**,** 28-42.

Mao, W. Multi-operation multi-machine scheduling. HPCN-Europe 1995, 1995 Berlin, Heidelberg. Springer, Berlin, Heidelberg, 33-38.

Matta, M. E. 2009. A genetic algorithm for the proportionate multiprocessor open shop. *Computers & Operations Research,* 36**,** 2601-2618.

Naderi, B., Fatemi Ghomi, S. M. T., Aminnayeri, M. & Zandieh, M. 2011. Scheduling open shops with parallel machines to minimize total completion time. *Journal of Computational and Applied Mathematics,* 235**,** 1275-1287.

Ow, P. S. 1985. Focused Scheduling in Proportionate Flowshops. *Management Science,* 31**,** 852-869.

Pinedo, M. L. 2016. *Scheduling: Theory, Algorithms, and Systems,* Cham, Springer International Publishing.

Zhang, J., Wang, L. & Xing, L. 2019. Large-scale medical examination scheduling technology based on intelligent optimization. *Journal of Combinatorial Optimization,* 37**,** 385-404.

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) (2021) 94-102
DOI: 10.38016/jista.772145

Automatic Ship Detection and Classification using Machine Learning from Remote Sensing Images on Apache Spark

Caner Özcan[1]*, Betül Dolapcı[2]

[1] Karabük University, Department Of Computer Engineering, Karabük, Turkey

[2]Kastamonu University, Department Of Computer Engineering, Kastamonu, Turkey

canerozcan@karabuk.edu.tr, bdolapci@kastamonu.edu.tr

Abstract
Ship detection and classification is very important for port and coastal security. Due to maritime safety and traffic control, high-resolution images of ships should be obtained. High resolution color remote sensing ship images taken from short distances provide advantages in ship detection applications. But the analysis of these high-dimensional images is complicated and requires long time. Dividing the image data into smaller blocks and representing them with a vector with distinctive and independent features facilitates the analysis process. For this reason, a block division method is applied first, dividing the image data into small pixel blocks. These obtained image blocks are also represented by the hybrid feature vectors. These feature vectors are created by adding the sub-features extracted from the color and texture properties of the images one after another. Using the obtained hybrid vectors, the images are classified using machine learning methods on Apache Spark. Classification studies were realized using Naive Bayes, Decision Trees and Random Forest methods in the MLlib. The analysis of the images was realized much faster with the clustering architecture created on Apache Spark platform. According to the obtained classification results, 99.62% classification success was achieved by using Random Forest method. In addition, an average of 3.4 times acceleration was achieved by running each method on 1 master + 4 workers clustering architecture on Spark.

Keywords: Apache Spark, Classification, Clustering, Machine Learning, Remote Sensing, Ship Detection

Apache Spark Makine Öğrenimi Kullanılarak Uzaktan Algılama Görüntülerinden Otomatik Gemi Tespiti ve Sınıflandırma

Öz
Gemi tespiti ve sınıflandırması, liman ve kıyı güvenliği açısından çok önemlidir. Deniz güvenliği ve trafik kontrolü nedeniyle, gemilerin yüksek çözünürlüklü görüntülerinin elde edilmesi gerekmektedir. Kısa mesafeden çekilmiş yüksek çözünürlüklü renkli uzaktan algılama gemi görüntüleri, gemi tespiti uygulamalarında avantaj sağlamaktadır. Fakat yüksek boyutlu bu görüntülerin analiz edilmesi süreci karmaşık ve uzun süreler gerektirmektedir. Görüntü verilerinin daha küçük parçalara bölünmesi ve bu parçalardan elde edilen ayırt edici ve bağımsız özelliklere sahip bir vektörle temsil edilmesi analiz işlemini kolaylaştırmaktadır. Bu nedenle, öncelikle görüntü verilerini küçük piksel bloklarına bölen bir blok bölümü yöntemi uygulanır. Elde edilen bu görüntü bloklarının da hibrit bir öznitelik vektörleri ile temsil edilmesi gerçekleştirilir. Bu öznitelik vektörleri, görüntülerin renk ve doku özelliklerinden çıkarılan alt özelliklerin birbiri ardına eklenmesi ile oluşturulur. Elde edilen hibrit vektörler Apache Spark'daki makine öğrenmesi yöntemleri ile kullanılarak görüntülerin sınıflandırılması sağlanmıştır. MLlib kütüphanesinde bulunan Naif Bayes, Karar Ağaçları ve Rastgele Orman yöntemleri kullanılarak sınıflandırma çalışmaları gerçekleştirilmiştir. Görüntülerin Apache Spark ortamında analiz edilmesi oluşturulan kümeleme mimarisi ile çok daha hızlı bir şekilde gerçekleştirilmiştir. Ayrıca her bir yöntemin Spark 1 master + 4 worker kümeleme mimarisi üzerinde çalıştırılması sonucu ortalama 3.4 kata yakın hızlanma sağlanmıştır.

Anahtar Kelimeler: Apache Spark, Sınıflandırma, Kümeleme, Makine Öğrenmesi, Uzaktan Algılama, Gemi Tespiti

* Corresponding Author.
E-mail: canerozcan@karabuk.edu.tr

Recieved : 21 July 2020
Revision : 04 Feb. 2021
Accepted : 23 Feb. 2021

1. Introduction

Today ship detection is very important in terms of maritime safety and maritime traffic. It is advantageous to work with ship images obtained by remote sensing due to coastal and port security (Yang et al., 2018). Also, ship detection and classification with remote sensing images is a crucial for military and civilian fields. In the literature, ship detection studies with remote sensing images are common in these fields (Liu et al., 2017). Another study in literature proposes a ship detection method from optical remote sensing images based on the network with visual attention (Bi et al., 2019). The type of image is crucial for correct feature extraction. Whether the image is a synthetic aperture radar (SAR) or optical remote sensing image requires different feature extraction techniques (Cavallaro et al., 2015). Less complex structure of color image data obtained from short distance provides an advantage in image detection and classification (Morillas et al., 2015). Local feature-based algorithms are used for object recognition in large-scale data obtained from satellite images (Ergul and Alatan, 2013).

A new detector called CenterNet++, working with SAR images, has been proposed (Guo et al., 2020). In this study, CenterNet++ method was developed to reduce complex background and increase detection capability. In another study developed using SAR images, ship detection was carried out according to the extraction of areas connected to water (Shi et al., 2019). Besides these, a new ship detection and classification method for complex sea surface is presented. (Wang et al., 2019).

A different approach presents a method for ship detection using satellite videos (Li et al., 2019). Another paper about vessel detection algorithms presents summarize of studies from optical spaceborne sensor images (Kanjir et al., 2018). Deep learning is used for autonomous ship detection in another paper. In this study, a novel hybrid deep learning method that combines a modified Generative Adversarial Network and a Convolutional Neural Network based detection approach is proposed for small ship detection (Chen et al., 2020). For some ship detection studies, images with land areas on the sea were used. The island filter is used for ship detection in the sea area with a land area in these studies (Wang et al., 2020).

Remote sensing is the technique of recording and examining the earth and ground resources without physical connection with them. In other words, remote sensing aiming to capture the earth images without any physical contact by means of aircraft and satellites and to obtain information through these images. The energy source used for remote sensing is either the sun or an artificial power source. Remote sensing technology has allowed the monitoring local and global environment for object detection (Yuan et al., 2020).

In this study, a hybrid feature vector has been developed for high performance classification and detection operations. The aim is to combine all the distinctive features of the image in a vector space and create a meaningful feature vector that will produce the correct result. Detection studies were carried out on remote sensing ship images in the marine environment. Before the feature vector was extracted on the images, the pre-processing was carried out. Noise removal and complex background cleaning were performed on the image. Pixel-based approaches make analysis difficult by considering unnecessary and non-distinctive variables (Morillas et al., 2015).

In this study, the block section is proposed to overcome the specified problem. Thanks to this block section, images are divided into small pixel blocks labeled as ship blocks or non-ship blocks. The classification of the blocks was carried out with the MLlib module of Apache Spark, which is used to classify large amounts of data. Naive Bayes, Decision Trees and Random Forest methods under this module have been applied.

In this approach, color and texture analyzes of the image are made and different features from both contents are combined in a hybrid vector. With the new hybrid vector formed, features are extracted from each block and then used for training and classification. In terms of efficiency of classification results, image blocks were analyzed in three different sizes and compared.

The rest of the article is organized as follows. Apache Spark Technology and architecture of cluster system are explained in Chapter 2. The detection process and the process of the formation of the feature vector, is detailed in Chapter 3. Machine learning classification algorithms of Spark used for fast data classification are explained in Chapter 4. The results of the analysis are presented in Chapter 5 with the tables and an evaluation is made by comparing the results of three different methods. Future work is presented in Chapter 6.

2. Apache Spark Clustering System

Apache Spark is an open-source library developed with Scala, which enables parallel processing on large data sets formed by high volume data. Spark has been developed as an alternative to the MapReduce method. Spark can be developed with Java, Scala, Python and R programming languages and supports SQL, data flow, machine learning and graphics processing.

The ability of the Hadoop Distributed File System (HDFS) and MapReduce method offered by the Hadoop environment to store data on multiple machines and to achieve parallel processing is faster and easier to achieve. It is due to the architecture that Apache Spark processes data faster and easier. Data is analyzed much faster using more than one machine. An abstraction method, defined as flexible distributed datasets (Resilient Distributed Datasets, RDD), is a collection of divided objects among a series of machines that allow lost data to be reproduced. With this method Spark

performs 10 times better than Hadoop in iterative machine learning operations and can analyze 39 GB of data interactively in under 2 seconds.

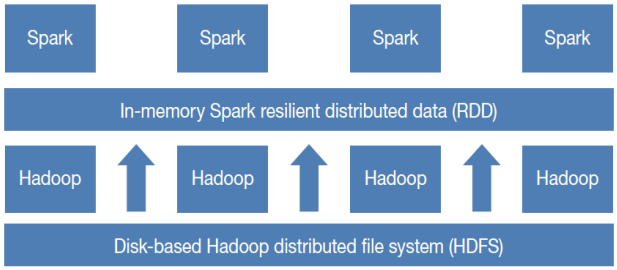


Figure 1. Spark and Hadoop in the cluster

Using Apache Spark clustering architecture, tasks can be distributed to computers in parallel. A Spark standalone cluster provides own web UI (User Interface) by monitoring cluster processes and running applications. It has a simple and efficient architecture. Standalone cluster consists of master and workers. Master is cluster manager that configures worker's processes and running applications. Workers start application's executors for tasks. Worker nodes communicate after completing their tasks in parallel and give the result of application to the master node.

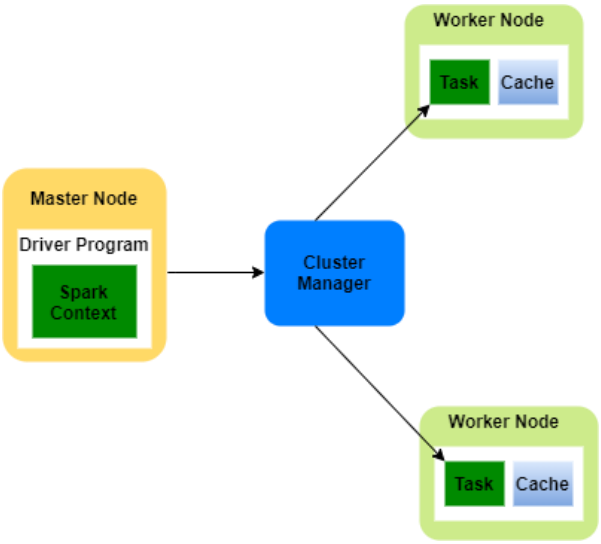


Figure 2. Architecture of Apache Spark cluster

In the literature, several methods have been developed with Spark to analyze various types of data. It was observed that text data was used for fast classification (Ogul et al., 2017) as well as image data for fast detection and classification (Ozcan et al., 2018). In the detection of objects in large-scale image data, Spark produces efficient results and can provide high performance in classification processes (Wang et al., 2020).

3. Feature Extraction Method

In this study firstly, a block section is applied to the images. After this step, features are extracted from the image blocks to be used as training data by extracting the color and texture features. These features are combined to create a hybrid feature vector. Then, the Naive Bayes, Decision Tree and Random Forest classifiers are trained based on the previously extracted feature vectors. As the last step after the classifiers have been trained, the classification between ship blocks and non-ship blocks has been carried out on the blocks of test images.

3.1. Block Division

Block-based approach provides more meaningful and holistic detection as opposed to pixel-based approach. Thanks to this method, which provides more homogeneous information depending on the color and texture content of the image, the rapid creation of the vector is also provided. In this study, features were obtained by applying the block section. The color images passed through the preprocessing stage are divided into 16x16, then 32x32 and then 64x64 pixel blocks and recorded in a folder. In Figure 3, some parts of an image divided into 64x64 blocks are shown.

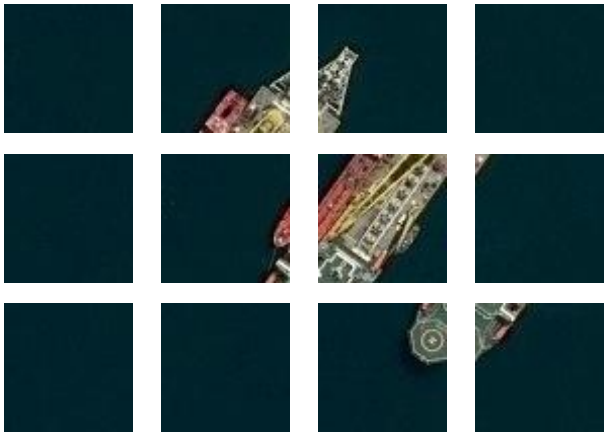


Figure 3. Example of an image divided into 64x64 pixel blocks

After the block division section, a binary mask application was applied for each block for labeling purposes. The purpose of this approach is to label images as ship or no ship in advance. The accuracy of the classification to be carried out in the next stages will be made by comparing with the labeled data. For this reason, it is very important to tag the data correctly. Ship blocks represent the pixel regions within the ship. Non-ship blocks consist of pixel areas outside the ship boundaries, such as water and sky areas. The reason for creating a binary mask is the labeling operations of the available image data. In image data, black regions are labeled 0 and white regions 1 (Figure 4.).

(a) (b)

Figure 4. (a) Original image. (b) Binary mask applied to the image

Compared to a pixel approach, this block division approach significantly reduces the complexity of the classification process because the number of elements to be classified is significantly reduced. Decreasing the number of elements is important in terms of dimension reduction of big data. In this study a detection was made on image data with image dimensions of 16x16 pixels, 32x32 pixels and 64x64 pixels, and comparisons were presented in tables in the Experimental Studies section.

3.2. Feature Extraction

In classification algorithms such as Naive Bayes, Decision Trees and Random Forests, it is an important decision to select the appropriate features to achieve the desired classification. This selection depends on the split blocks of the image and the type of images available (without pixel-based distortion). In case of ship detection in images obtained from short distance the marine environment provides useful visual features that can be used as a feature. This study proposes extracting color and texture features from each block of images. After analyzing whether the extracted features are meaningful in terms of data, they are added one after another and the feature vector is created. These features are then used during the training and classification stages.

3.2.1. Color Feature

Colors define the visual perception of pixels, tone distribution according to light and give information about their chromatic densities. In this approach, three different color spaces are evaluated: RGB (Red-Green-Blue), HSV (Hue-Saturation-Value) and L*a*b* defined by CIE (International Commission on Illumination). RGB is a color image area based on the color model commonly used in computer graphics because it works similar to the human visual system. In this model, primary colors are defined by red, green, and blue colors represented by the value of each of the RGB components (Morillas et al., 2015). HSV is a color space used in computer vision and image analysis with applications such as object recognition and image segmentation. One of the main advantages of HSV is the distinction between density and color information similar to that performed by the human brain. Hue (H) describes the shade of the color and its location in the color spectrum. Saturation (S) represents the purity of the tint according to a white reference. Value (V) is the measure of the brightness of the color, that is, the ratio of white in it (Morillas et al., 2015). The CIE LAB color space is based on the human perception of different wavelengths and can identify any color perceived by the average human observer. CIE LAB is a device-independent color compared to RGB and HSV which are device-dependent colors. At the CIE Lab, three parameters are represented by a sphere. The vertical axis L* represents lightness. The horizontal axis a* measures the difference between the red and green components, and the horizontal axis b* measures the difference between the blue and yellow components (Morillas et al., 2015).

Being able to use color spaces as features depends on the mean and standard deviation from each block for each color component. Mean:

$$\mu = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)}{M \times N} \quad (1)$$

Standard Deviation:

$$\sigma = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N} (I(x,y)-\mu)^2}{M \times N}} \quad (2)$$

where I(x, y) is the color component of the pixel in (x, y), M is the width of each block in pixels and N is the height of each block in pixels.

3.2.2. Texture Feature

Texture is a feature that represents the structure and spatial properties of pixels in a region. The texture can be characterized by the density properties of pixels and the spatial relationship between them on a gray level. Unlike color properties, texture properties describe region-based information instead of individual pixel. To extract texture features, images are first converted to grayscale, eliminating the hue and saturation information while preserving the brightness component. After this transformation, two types of texture features are extracted from each block: first-order-statistics (FS) and Gray Level Co-occurrence Matrices (GLCM).

In GLCM-based texture analysis, some statistical data provided by this algorithm are based on. Statistical data are explained in the table below. The FS-based statistical data is just like the other table (Gonzalez et al., 2003):

Table 1. Features of GLCM

Feature	Definition
Contrast	Density and gray level variations
Correlation	Gray level values linear dependence
Energy	Pixel homogeneity criterion
Homogeneity	Similarity criterion in different regions

Table 2. Features of FS

Feature	Definition
Mean	Pixel values average
Standard Deviation	Square root of variance information
Variance	Squared deviation from the mean
Distortion	Criterion of the asymmetry of its distribution
Entropy	Gray level spatial irregularity
Energy	Pixel homogeneity criterion

The reason why the four features above are preferred for feature extraction with GLCM algorithm is that these features combined with FS features best represent the gray level intensities. FS features are given in Table 2.

Table 3. Abbreviation for features used in feature extraction.

Feature	Definitions
RGB	Mean of RGB component
HSV	Mean of HSV component
LAB	Mean of LAB component
SD	Standard deviation of color components
FS	First Order Statistics
GLCM	Gray Level Co-occurrence Matrices

4. Classification using Spark MLlib

Machine learning algorithms under Spark used in image classification are computationally intensive. Spark contributes well to machine learning, as it supports fast in-memory computing and recursive querying of data. MLlib is Spark's scalable machine learning library (Lagerstrom et al., 2016). It consists of common learning algorithms and utilities such as classification, regression, clustering, collaborative filtering and dimensionality reduction. It also includes opportunities to model and train deep neural networks. Spark MLlib provides the use of an application programming interface in Java, Scala and Python, which facilitates integration with an existing Java application that uses OpenIMAJ for image extraction and classification (Han et al., 2006).

4.1. Naive Bayes

Naive Bayes (NB) algorithm is a controlled machine learning algorithm. It is a simple probability model for multiple classifications with the assumption of independence between features. NB assumes that each feature contributes independently to the possibilities assigned to a class. The NB classifier performs the analysis operations according to the formula below:

$$P(c|F) = (P(F|c)P(c))/(P(F)) \quad (3)$$

where P(c) and P(F) are the preliminary probabilities of events c and F, P(c|F) indicates the probability of event c occurring in the event of event F, P(F|c) indicates the probability of occurrence of event F when c event occurs. If P(F) probabilities are the same in all classes, it is aimed to maximize the dividend only. If P(c) probabilities are unknown, classes are assumed to be equal, and then we just maximize P(F|c). When many sets of data are given, computing P(F|c) will be computationally expensive. Reduction of the computational complexity in the evaluation of P(F|c)P(c) is only done with the naive assumption of class conditional independence using formula below:

$$P(c|F) \sim \prod_{k=1}^{n} P(c_k|F)) \quad (4)$$

It is partially more difficult to train the dataset with the NB algorithm, but it is a classification algorithm that works quite fast after training. It acts according to the condition of being the highest probability of a situation. The disadvantage is that the data is constantly changing. Because every new data will extend the training process (Kaya and Yıldız, 2014). Laplace smoothing was used in the Naive Bayes algorithm in this study and a parameter called lambda was used during the training as equaled to 1.

4.2. Decision Trees

The second method that provides the most effective results among machine learning algorithms is the Decision Tree (DT) algorithms. It can be used for classification and regression.

A decision tree; consists of knot, branch and leaf. The top part is the root, the path from the root to the other nodes is the branch and the last result through these branches is the leaf (Kavzoglu and Colkesen, 2010). With this algorithm, a series of questions are asked to the data to be trained, and the results are reached in line with the answers obtained. While forming a decision tree, it is calculated with the information gain and information gain rate approaches according to which criterion or attribute value of the branch in the tree (Ozcan et al., 2020). DT is a variant of a greedy algorithm that progresses in the form of dividing and conquering in a top-down repetition, applying a set of decision rules (Man et al., 2018). In this algorithm, a tree structure is created, and class tags are expressed in the leaves of the tree. The last tree predicts the same tag for all samples that reach the leaf node. Each section is determined by choosing the best separation from the set of possible divisions to maximize knowledge gain in a tree node. When the split selected in each tree node is applied to the T dataset of a split v, the arguments necessary to maximize knowledge gain are obtained by calculating IG (T, v). Here, two different measures (Gini impurity and entropy impurity) are proposed for classifying the dataset (Man et al., 2018). Gini impurity:

$$\sum_{a=1}^{C} f_a(1 - f_a) \quad (5)$$

is calculated as. Here, C is the number of unique tags, and f_a frekans is the frequency of tag a in a node. The impurity measure defined for entropy is as follows:

$$\sum_{a=1}^{C} -f_a \log(f_a) \quad (6)$$

Information gain is based on subtracting the main node impurity from the weighted sum of the two sub-node impurities. Information gain is defined as follows:

$$IG(T,v)=Gini(T)-\frac{N_{left}}{N}Gini(T_{left})-\frac{N_{right}}{N}Gini(T_{right}) \quad (7)$$

Here, the data set T with size N is obtained by dividing the sections and the terms T_{left} and T_{right} in sizes N_{left} and N_{right}. In this study, the maximum number of divisions for decision trees was chosen as 10. Maximum number of bins used for splitting features was chosen as 32.

4.3. Random Forest

The Random Forest (RF) method is one of the most successful machine learning models. RF is a community learning algorithm that comes together by decision trees to solve supervised learning tasks such as classification, has a good tolerance to noise and does not tend to over-sleep. Compared to the NB and DT approach, it provides much higher performance classification results. It combines multiple decision trees by producing stronger models to get a more accurate and stable estimate. The algorithm creates a model of multiple decision trees based on different data subsets using a random data sample during the training phase. This randomness constitutes an advantageous feature of the random forest model, which makes it more robust than a single decision tree and overcomes the problem of traditional data being overly compatible and similar (Man et al., 2018).

Overcompliance is defined as the model's over-learning and memorizing data while training on data. The RF approach generates and trains random subtrees from the dataset and feature vectors to overcome the problem of over-adaptation, a disadvantage of the DT approach. In this structure, each of which consists of different decision trees, the classification process is realized through the estimates with the highest votes.

The information gain (BK) and Gini index obtained using the attributes b to divide the sample set T are shown by the node division formula given below (Cortes and Vladimir, 1995):

$$BK(T,b)=Ent(T)-\sum_{n=1}^{V}-\frac{|T^{(n)}|}{|T|}Ent(T^{n}) \quad (8)$$

$$Gini(T,b)=\sum_{n=1}^{V}-\frac{|T^{(n)}|}{|T|}Gini(T^{n}) \quad (9)$$

Here (T^{n}) shows that in n branch node it contains all instances in T with the value of b^{n} in the b attribute. The number of trees in the random forest used for training was determined as 10. The classification number in the algorithm is determined as 6.

4.4. Training and Classification

In general, a supervised learning process consists of two stages: training and classification. The images to be classified are divided into two, as a training and test data at a predetermined rate. The training set consists of images used to train the machine learning classifier. In this approach, features are extracted from the blocks of these training images and combined in a hybrid feature vector. Before starting training, the created feature vector goes through normalization processes. The correct classification of each block during training is also provided through binary masks (Figure 4. (b)), in which the blocks are correctly labeled in advance.

After the image data is trained with machine learning algorithms, the classification processes are performed in the test set created by the images used for evaluation. The classifiers created during the training estimate the correct classification of the blocks in these test images and classify them as ship or non-ship blocks. Ship blocks and non-ship blocks are represented by white pixels and black pixels, respectively. The results of the classification analysis on the test images were examined by three different machine learning algorithms and evaluations were made. High performance in classifications made by these algorithms depends on the block size selected for analysis. The smallest blocks allow much more detail to be considered than the images. Compared to small block sizes, classifications using larger block sizes show less performance. Besides, factors such as brightness of the images, whether it is bright due to weather conditions, it is considered as a disadvantage in the classification stage that camouflage of the sea ships using shades similar to the colors of the sea to prevent them to be watched mostly by enemy forces. This disadvantage is solved by using texture features in addition to the features obtained from color.

5. Experimental Studies

In this study, a hybrid vector was obtained by extracting color and texture features from image contents. The length of the hybrid feature vector is 28x1. The results and classification success are evaluated through the different block sizes with the following tables and graphs. By creating the vector in different sizes, only the color spaces were first evaluated, then only the texture properties were evaluated, and classification procedures were performed. The classification results using different feature sizes were also evaluated.

When ship images with dimensions of 768x768 are divided into 16x16 block sizes, a total of 652032 block images are obtained for 283 images while 2304 block images are obtained from one image. When the same image data is divided into 32x32 block sizes, 163008 block images are obtained, and when they are divided into 64x64 block sizes, 40752 block images are obtained. In total, analysis operations were performed with 855792 block images obtained from 283 images.

The dataset consists of 283 images shared by the Airbus Company publicly. The dimension of the images was 768x768 pixels and stored as a jpg format. The training percentage of the data was determined as 70%, and the test percentage was 30%. The results of this study are presented in Table 4 and Table 5. The success of classification operations using the feature vector and label vector were calculated using the MLlib library of Apache Spark in the Eclipse Oxygen version environment. The results are taken from the GNU/Linux operating system distributions in Ubuntu 16.04 environment. Figure 5 shows example images from used in this study.

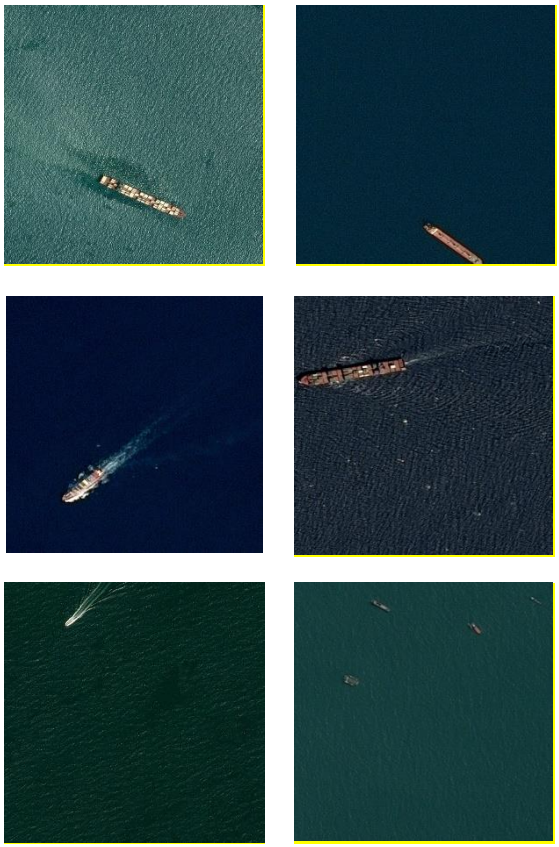


Figure 5. Example of ship images

As seen in the Figure 6, the best classification success has been achieved with RF algorithm. As the block sizes decrease, it gives much more successful classification criteria in DT and RF algorithms. When the graphic is analyzed, it is seen that the Random Forest approach gives the most successful accuracy. Again, according to the graph, Decision Trees are more successful than Naive Bayes approach.

In order to evaluate the classification performance criteria used in this study from a different perspective, the classification success was tested by dividing the hybrid vector created into several feature bases. For this, firstly, classification results were obtained with the vectors created in each color space. Then, SD and FS features were added to each color space vector and results were obtained. Lastly, GLCM features were added to measure which vector was obtained with a better classification result. The results obtained in these experiments are as follows:

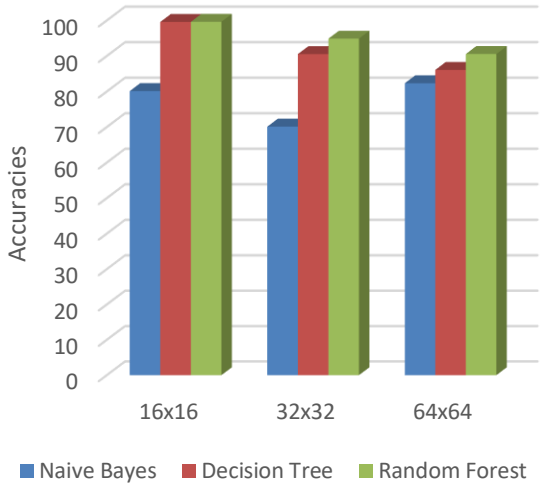


Figure 6. Accuracy results of ML algorithms in three different block sizes with the hybrid feature vector (%).

According to Table 4, the 28x1 dimension hybrid feature vector is divided into parts on ten different feature bases, which features contribute to the dominant degree of classification. According to the table, the vector with the highest performance in 3 different classification algorithms is our hybrid vector. In addition, although the classification performance decreased when trained with Naive Bayes algorithm, RGB+SD+FS, HSV+SD+FS, LAB+SD+FS vectors, it is seen that the classification success does not decrease below 94% when the RF algorithm is trained with all ten vectors.

Table 4. Accuracy results (%) of 3 different classification algorithms with piecewise feature vector.

Features	NB	DT	RF
RGB	95.03	96.60	97.59
HSV	98.72	98.75	98.84
LAB	93.36	94.03	94.38
RGB+SD+FS	95.62	98.64	98.72
HSV+SD+FS	93.72	98.63	98.70
LAB+SD+FS	94.39	98.52	98.81
RGB+SD+FS+GLCM	95.45	97.33	98.20
HSV+SD+FS+GLCM	97.12	97.83	98.93
LAB+SD+FS+GLCM	95.29	98.01	98.65
Our Hybrid Vector	80.12	99.58	99.62

After that, classification was made on image test data using Apache Spark clustering architecture. In this classification using the master worker architecture, the analysis of different image data sizes with different master-worker architecture was evaluated. The variation

of the Spark speed performance by data size and number of workers was observed. Time results developed using three different methods were evaluated. From left to right, as the number of workers decreased, the test processing time increased. This shows that parallel architecture is important in terms of time in big data processing.

In Table 5, Spark clustering architecture consist of master and workers has seen using three machine learning algorithms. It represents the vector number in million (M). The variation of the Spark speed performance by data size and number of workers was observed. The obtained results are presented in Table 5. Time results developed using three different methods were evaluated. From left to right, as the number of workers decreases it was observed that the test process time increased. With the NB algorithm, it is seen that there is 3.5 times increase in speed between 1 master+4 workers and 1 master+1 worker on 18 M data. With the DT algorithm, it is seen that there is a 3.3 times speed increase between 1 master+4 workers and 1 master+1 worker on 18 M data. Also, it is seen that the RF algorithm has a speed increase of 3.3 times between 1 master+4 workers and 1 master+1 worker on 18 M data. Although the time difference between the data size decreases and decreases proportionally with the data, it is seen that multiple workers defined processes are always very fast. As the data size passes to each lower row, the speed increase is observed when the classification times are reduced to a certain extent for evaluation. When it comes to big data, it is much more reasonable to prefer cluster architecture.

Table 5. Classification times (milliseconds) with NB-DT-RF methods using Master (Ms) - Worker (Wr) clustering architecture

Number of Vectors	Naive Bayes		Decision Tree		Random Forest	
	1 Ms+1 Wr	1 Ms+4 Wr	1 Ms+1 Wr	1 Ms+4 Wr	1 Ms+1 Wr	1 Ms+4 Wr
18 M	4939	1384	2549	763	25710	7680
15 M	3999	1179	2519	692	21075	7054
12 M	2972	1056	1925	617	17059	4400
9 M	2345	663	1566	383	12754	4410
6 M	1563	541	1134	315	8610	2454
3 M	882	271	596	199	4419	1398

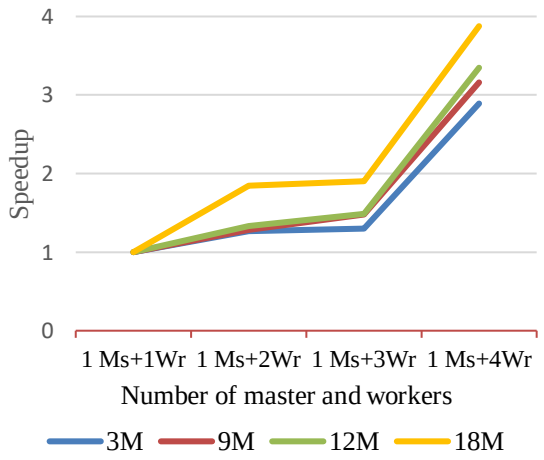


Figure 7. Classification testing phase speedups as a function of number of worker nodes

6. Conclusions

The aim of this study is to classify short distance images between ship and non-ship blocks using machine learning methods. The highest achievement was the Random Forest method with a rate of 99.62 %. In the comparative study between three color areas evaluated, the HSV+SD+FS+GLCM feature vector achieved the highest performance rate. It has been observed that when color and texture features are used together, higher success is achieved. Higher success can be achieved by adding the shape feature to the color and texture features. Thanks to the vector formed by adding the shape feature, ships can be classified according to their shapes and sizes. Evaluating more complex features using different machine learning methods and extracting features with deep neural networks can affect the performance of this study and provide much more efficient results.

Acknowledgment

This work was supported by the Scientific Research Projects Unit of Karabuk University under project number FYL-2019-2044.

References

Bi, F., Hou, J., Chen, L., Yang, Z., Wang, Y., 2019. "Ship Detection for Optical Remote Sensing Images Based on Visual Attention Enhanced Network". Sensors, 8, 4634-4646, 2015.

Cavallaro G, Riedel M, Richerzhagen M, Benediktsson JA, Plaza A. "On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector Machine Methods". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 4634-4646, 2015.

Chen, Z., Chen D., Zhang Y., Cheng X., Zhang M., 2020. "Deep learning for autonomous ship-oriented small ship detection". Safety Science, 130.

Cortes C, Vapnik Vladimir. "Support-Vector Networks", Machine Learning, 20, 273-297 (1995).

Ergul M, Alatan AA. “Geospatial Object Recognition From High Resolution Satellite Imagery”. 2013 21st Signal Processing and Communications Applications Conference (SIU), Haspolat, Turkey, 24-26 April 2013.

Gonzalez RC, Woods RE, Eddins SL. *Digital Image Processing using Matlab*. New Jersey, Prentice Hall, 2003.

Guo H, Yang X, Wang N, Gao X. “A CenterNet++ model for ship detection in SAR images”, Pattern Recognition, 112, 2020.

Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. Waltham, MA, USA: Elsevier, Second Edition, 2006.

Kanjir, U., Greidanus, H., Ostir, K., 2018. “Vessel detection and classification from spaceborne optical images: A literature survey”. Remote Sensing of Environment, 207 (1-26).

Kavzaoglu T, Colkesen İ. “Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği”, Electronic Journal of Map Technologies, 2, 2010.

Kaya C., Yıldız O. “Makine Öğrenmesi Teknikleriyle Saldırı Tespiti: Karşılaştırmalı Analiz”. Marmara Fen Bilimleri Dergisi

Li, H. Chen, L., Li, F., Huang M., 2019. “Ship detection and tracking method for satellite video based on multiscale saliency and surrounding contrast analysis”. Applied Remote Sensing 13 (2).

Li Y, Zhang H, Guo Q, Li X. “Machine Learning Methods for Ship Detection in Satellite Images”.

Liu, Y., Cuı, H.Y., Kuang, Z., Lı, G.Q., 2017. ITM Web of Conferences, 12.

Man W, Ji Y, Zhang Z. “Image Classification Based on Improved Random Forest Algorithm”, 2018 the 3rd IEEE International Conference on Cloud Computing and Big Data Analysis, 20-22 April 2018, Chengdu, China.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, L., et al. (2015). MLlib: Machine Learning in Apache Spark.

Morillas JRA, Garsia IC, Zolzer U. “Ship Detection Based on SVM Using Color and Texture Features”. 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 3-5 September 2015.

Nie T, He B, Bi G, Zhang Y, Wang W. “A Method of Ship Detection under Complex Background”. International Journal of Geo-Information, 2017.

Ogul IU, Ozcan C, Hakdaglı O. “Fast Text Classification with Naive Bayes Method on Apache Spark”. 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey.

Ogul IU, Ozcan C, Hakdaglı O. “Text Classification with Spark Support Vector Machine”. 1. Ulusal Bulut Bilişim ve Büyük Veri Sempozyumu, Antalya, 2017.

Ozcan C, Ersoy O, Ogul IU. “Classification of SAR Image Patches with Apache Spark Using GLCM Texture Features”. International Conference on Advanced Technologies, 3rd World Conference on Big Data, İzmir, 28 - 30 Nisan 2018.

Ozcan C, Ersoy O, Ogul IU. “Fast texture classification of denoised SAR image patches using GLCM on Spark”, Turkish Journal of Electrical Engineering & Computer Sciences,28, 2020.

Shi H, He G, Feng P, Wang J. “An On-Orbit Ship Detection And Classification Algorithm for SAR Satellite”, IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July-2 August. 2019.

Wang C, Pei J, Wang R, Huang Y, Yang J. “A new ship detection and classification method of spaceborne SAR images under complex scene”, 6th Asia-Pacific Conf. on Synthetic Aperture Radar (APSAR), Xiamen, China, 26-29 Nov. 2019.

Wang N, Chen F, Yu B, Qin Y. “Segmentation of large-scale remotely sensed images on a Spark platform: A strategy for handling massive image tiles with the MapReduce model”. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 137-147, 2020.

Wang. Z, Yang T., Zhang H. 2020. “Land contained sea area ship detection using spaceborne image”. Pattern Recognition Letters, 130 (125-131).

Yang, X., Sun, H., Fun, K., Yang, J., Sun, X., Yan, M., Guo, Z., 2018. “Automatic Ship Detection of Remote Sensing Images from Google Earth in Complex Scenes Based on Multi-Scale Rotation Dense Feature Pyramid Networks”. Remote Sens, 132, 10.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang Q., Wang, J., Gao, J., Zhang, L., 2020. “Deep learning in environmental remote sensing: Achievements and challenges”. Remote Sensing of Environment, 241.

Zaharia M, Chowdhury M, Franklin MJ, Shenker S, Medjahed I. “Spark: Cluster Computing with Working Sets”.

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) (2021) 103-112
DOI: 10.38016/jista.854250

Analysis of Turkish Sentiment Expressions About Touristic Sites Using Machine Learning

Muhammed Çağrı Aksu[1]*, Ersin Karaman[2]

[1]Artvin Çoruh University, Rectorate, Department of Informatics, Artvin, Turkey

[2] Ankara Hacı Bayram Veli University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, Ankara, Turkey

cagriaksu@artvin.edu.tr, ersin.karaman@hbv.edu.tr

Abstract
Analyzing data by inferring from unstructured data about customers is one of the main purposes of the tourism and many other industries as well. However, performing unstructured data analysis using traditional methods is quite inconvenient and costly. This can be overcome by using sentiment analysis, an area of application of text mining. Since there is no proven methodology for sentiment analysis, researchers often perform their studies by trial and error. Many studies on sentiment analysis have focused on comparing the preprocessing or the performance of various machine learning algorithms. Both for these reasons and since research on sentiment analysis with Turkish content is limited, this study aimed to determine the effects of labeling, stemming, and negation on the success of sentiment analysis using Turkish touristic site analysis. From the data set prepared for this study, 12 different variations were created according to labeling, number of classes, stemming, and negation. These data sets were classified using the algorithms Naive Bayes (NB), Multinominal Naive Bayes (MNB), k-Nearest Neighbor, and Support Vector Machines (SVM), often used in sentiment analyses, and the findings were compared.

Keywords: Text mining, sentiment analysis, machine learning, unstructured data analysis, classification, naive bayes, multinominal naive bayes, support vector machines

Turistik Mekanlar Hakkındaki Türkçe Duygu İfadelerinin Makine Öğrenmesi Yöntemleri ile İncelenmesi

Öz
Müşteriler ile ilgili yapılandırılmamış verilerden çıkarımlar yaparak bu verileri analiz etmek birçok sektör için olduğu gibi turizm sektörü için de temel amaçlardandır. Yapılandırılmamış veri analizinin geleneksel yöntemlerle gerçekleştirilmesi oldukça zahmetli ve maliyetli olmaktadır. Metin analizi uygulama alanlarından biri olan duygu analizi kullanılarak bu sorunun üstesinden gelinebilmektedir. Duygu analizi çalışmalarında henüz kanıtlanmış bir metodoloji bulunmadığı için araştırmacılar genellikle deneme yanılma yoluyla çalışmalarını yürütmektedirler. Duygu analizi alanında yapılan birçok çalışma duygu analizi ön işlemlerinin ya da farklı makine öğrenimi algoritmalarının performanslarının karşılaştırılması üzerinedir. Hem bu nedenlerden dolayı hem de Türkçe içeriklerle gerçekleştirilmiş duygu analizi çalışmalarının kısıtlı olmasından dolayısıyla bu çalışmada Türkçe turistik mekân incelemeleri kullanılarak duygu analizi ön işlemlerinden etiketleme, köklerine ayırma ve olumsuzlaştırma işlemlerinin duygu analizinin başarısına olan etkileri tespit edilmeye çalışılmıştır. Bu nedenle bu çalışma için hazırlanan veri setinden etiketlenme şekline, sınıf sayısına, köklerine ayırma ve olumsuzlaştırma durumlarına göre 12 farklı varyasyon oluşturulmuştur. Oluşturulan bu veri setleri duygu analizi çalışmalarında sıklıkla kullanılan Naive Bayes (NB), Multinominal Naive Bayes (MNB), k-Nearest Neighbor ve Support Vector Machines (SVM) algoritmalarıyla sınıflandırılarak elde edilen sonuçlar karşılaştırılmıştır.

Anahtar Kelimeler: Metin madenciliği, duygu analizi, makine öğrenmesi, yapılandırılmamış veri analizi, sınıflandırma, naive bayes, multinominal naive bayes, destek vektör makineleri

* Corresponding Author.
E-mail: cagriaksu@artvin.edu.tr
** This study is derived from the first author's doctoral thesis.

Recieved : 05 Jan 2021
Revision : 09 Mar 2021
Accepted : 16 Apr 2021

1. Introduction

Structured data obtained by traditional methods are insufficient for dynamically analyzing changing customer trends (Esen & Türkay, 2017). Therefore, new mechanisms have emerged that allow the dynamic analysis of changing customer trends and support the decision-making of managers in organizations. These mechanisms enable data-driven decision-making using unstructured data sources such as social media posts, blogs, and web server logs (Provost & Fawcett, 2013). Most of the data used in organizations consist of unstructured data, with estimations of around 80% (Beal; Blumberg & Atre, 2003; Lohr, 2012). Therefore, it is of great importance to analyze unstructured data and to extract meaningful information that will be useful for the organization. Through sentiment analysis, which is one of the sub-research areas of text mining, people's opinions, evaluations, attitudes, and sentiments about products, services, and activities can be analyzed (B. Liu, 2012). In other words, unstructured data can be analyzed using sentiment analysis.

Sentiment analysis can be performed in two ways, namely, dictionary-based method and machine learning method (Can & Alataş, 2017). In the dictionary-based method, there is a dictionary called sentiment dictionary that contains a large number of words, and each word has a sentimental polarity, that is, the word's positivity, negativity, or neutrality degree. In this method, the sentimental polarity of the text is calculated by searching each word, the sentimental state of which is to be determined, in the sentiment dictionary (Baccianella et al., 2010; Esuli & Sebastiani, 2006; Kuvd., 2006). Then, data of unknown classes are classified using this model (Pangvd., 2002). In some sentiment analysis studies conducted with machine learning, the classes of the data are determined manually by the researchers (Kulcu & Dogdu, 2016; K.-L. Liu, Li, & Guo, 2012). Some studies perform a labeling process using metadata. These metadata are often scores ranging from one to five or the emojis in the text to be analyzed (Chang, Ku, & Chen, 2019; Gezici & Yanıkoğlu, 2018; Taecharungroj & Mathayomchan, 2019; Türkmenoglu & Tantug, 2014).

Aydoğan and Akcayol (2016), Özyurt and Akçayol (2018), Can and Alataş (2017) reviewed recent studies on sentiment analysis based on their methods, areas of application, and data sets. Considering these reviews and other research in the literature (Bilgin & Şentürk, 2017; Çoban et al., 2015; Kaynar et al., 2016; Kızılkaya, 2018; Meral & Diri, 2014; Salur et al., 2019; Toçoğlu, 2018; Türkmenoğlu, 2015), it is observed that the studies on sentiment analysis are often focused on the classification of examinations, tweets, or comments in a certain area using various machine learning techniques and the comparison of these classification results. It is also found that sentiment analysis studies in the context of tourism are quite limited and there are suggestions to conduct further analysis including Turkish content. Thus, here, it was aimed to create a data set that allows the sentiment analysis of texts about Turkish touristic sites, to determine the effect of labeling on the success of the classification, and to examine the effect of tokenization, stemming, and negation on the success of classification.

To achieve the aims of the study, it is first necessary to create a data set. To create a data set suitable for the subject of the current sentiment analysis, reviews on Tripadvisor, Google Maps, and Foursquare, websites that are frequently mentioned in research on tourism and that allow tourists to make comments or reviews about the places they visit, were used. From these sites, reviews by tourists for 203 touristic sites in the cities of Trabzon, Artvin, and Rize in the Eastern Black Sea region of Turkey were obtained. A total of 49031 reviews were obtained on 10.03.2020. This data consisted of reviews by tourists, the date of their reviews, and scores ranging from one to five given by the tourists. Using these data, the data sets described in the third section were created. Using each of these data sets, machine learning models were created using the classic machine learning algorithms NB, MNB, k-NN, and SVM. The machine learning models were validated by 10-repetitive cross-validation and the findings were obtained by the f-scores of the models.

The second section describes sentiment analysis, the method of the study. The third section explains how the experiments were performed. The fourth section demonstrates the findings. Finally, the last section discusses these findings.

2. Materials and Methods

This section briefly explains how the data are obtained, the aforementioned machine learning algorithms, and the method, sentiment analysis.

2.1. Data Collection

As stated, the data was obtained from the websites of Tripadvisor, Google Maps, and Foursquare to create the data sets. First, a list of touristic sites in the cities of Trabzon, Artvin, and Rize in the Eastern Black Sea region of Turkey was obtained from karadeniz.gov.tr. Then, these touristic sites were searched on the websites and the pages containing comments or reviews on them were obtained. The links to these pages were sent to the

Datashake[1] data scraping[2] service with a php script. The reviews were then recorded in the database. Finally, the data in the database were applied the processes specified in the experiment section and the dataset(s) was created.

2.2. Sentiment Analysis

The process of automatically discovering some previously unknown information from different written sources is called text mining (Hearst, 2003). Text mining is divided into seven areas of application as text clustering, text classification, web mining, information extraction, natural language processing, concept extraction, and information retrieval and each area of application has its specific features (Miner et al., 2012). Sentiment analysis is located at the intersection of opinion extraction and document classification (Miner et al., 2012). Sentiment analysis, also known as opinion mining, is a field of study that analyzes people's views, evaluations, attitudes, and feelings about assets such as products, services, organizations, individuals, activities, topics, and their characteristics (B. Liu, 2012). Sentiment analysis is often done by classifying the sentiment in a text in a binary (positive-negative) or ternary (positive-negative-neutral) form (Şeker, 2016). Sentiment analysis can be performed in two ways: dictionary-based method and machine learning method (Can & Alataş, 2017). In the dictionary-based method, each word in the text is searched in dictionaries with predetermined polarities, the opinion score of the text is calculated, and classification is performed (Baccianella, Esuli, & Sebastiani, 2010; Esuli & Sebastiani, 2006; Ku, Liang, & Chen, 2006). In the machine learning method, a machine learning model with labeled data is created and data with unknown classes are tried to be classified using the created model (Pang, Lee, & Vaithyanathan, 2002). In this study, sentiment analysis was performed using the machine learning method, as it has been stated to be the superior method in the literature (Özyurt & Akçayol, 2018).

Since many classification algorithms used in sentiment analysis by machine learning cannot work with categorical data, they must be converted into numerical data. In sentiment analysis, the conversion of categorical data is often carried out by the bag-of-word (BOW) method ("bag-of-word model," 2007; Harris, 1954). In this method, every single term in the text (a word, a sentence, or a certain number of characters) is considered an attribute and the frequency of each term in the text is assigned as the value of the attribute. Thus, categorical text data is converted into digital form.

In many sentiment analysis studies, the data is preprocessed, consisting of steps such as tokenization, normalization, stemming, stop words removal, and term weighting to increase classification success and to reduce the attribute size (Aydoğan&Akcayol, 2016; Çoban, 2016; Çoban et al., 2015; Meral & Diri, 2014; Saad, 2010; Türkmenoğlu, 2015). These steps are briefly explained below.

Tokenization: In this step, the text to be classified is divided into terms by various methods. A term can contain one word or multiple words. Using a method called n-grams, the text can be divided into word-based or character-based terms. In word-based n-gram, the number of n words is treated as a term and in character-based n-gram, the number of n characters is treated as a term.

Normalization: There may be typos in texts on social media. Normalization is a natural language processing procedure that corrects spelling errors. In Turkish text classification studies, normalization is often performed using the Zemberek-NLP natural language processing tool.

Stemming: In the classification of a given text, each word in the text is taken as an attribute; therefore, it is aimed to reduce the number of attributes by stemming. Stemming is a natural language processing operation that is often performed using the Zemberek-NLP natural language processing tool in Turkish (Akın & Akın, 2007).

Stop words removal: While classifying a text, removing stop words is performed to reduce the number of attributes in many studies (Çoban, 2016; Kaynar et al., 2016; Meral & Diri, 2014). Stop words are often those that do not affect the sentiment of the sentence, such as prepositions and conjunctions (Sevindi, 2013).

Term weighting: In the bag-of-word approach, words with high frequency become dominant and cannot provide much information for the model (Waykole & Thakare, 2018). In other words, terms that appear very often in the text may not have any distinctive significance. However, they can have a high weight value. To prevent this, the frequency of the terms is rescaled by the TF-IDF method (TermFrequency – InverseDocumentFrequency) considering how often the terms occur in all texts (Spärck Jones, 2004). The TF value is calculated by the formula in Eq. 1, the IDF value by the formula in Eq. 2, and the TF-IDF by the formula in Eq. 3.

$$tf_{ij} = \frac{F_{ij}}{\sum_i F_j} \tag{1}$$

$$IDF_j = \log(\frac{D}{df_j}) \tag{2}$$

[1] Datashake is a web service that provides reviews and comments from over 85 websites using data scraping. Using the data scraping APIs offered by Datashake, users can easily access reviews and comments on websites like Tripadvisor, Foursquare, and Google Maps (Datashake, 2021).

[2] Data scraping is the process of obtaining desired data from unstructured website content by software using data sets that are suitable for automatic processing (Data Scraping, 2021).

$$w_{i,d} = tf_{ij} * IDF_j \quad (3)$$

where i = text index, j = term index, F = frequency, df_j= number of texts containing j, and D = the number of texts.

Text classification studies in the literature have used different weighting techniques beside TF-IDF such as A-TF, B-TF, LA-TF, L-TF, Knowledge gain, and Chi-square (Sevindi, 2013; Yıldız, 2016).

2.3. Naïve Bayes Classifier

Naïve Bayes is a classification technique based on the probability theory of Bayes (1763). It is based on the assumption that each attribute to be used in classification is independent of each other. It has been used in text classification studies since the early 1960s (Maron, 1961). The Naïve Bayes classifier briefly estimates the class with the highest probability by calculating the probabilities of all cases for each class. The classifier works as follows.

1. Let *X* be a vector that is tried to be predicted and consists of n attributes.

$$X = (x_1, x_2 \dots x_n) \quad (4)$$

2. Let there be m classes in the data set represented by $C_1, C_2 \dots C_m$.
3. The classifier calculates the value with the highest successive probability $P(X|C_i)$ among all classes, as in Eq. 5, to find out which class the vector *X* belongs to.

$$P(C_i|X) = \frac{P(X|C_i)\,P(C_i)}{P(X)} \quad (5)$$

   a- $P(C_i)$, is calculated as in Eq. 6 by dividing the number of elements in the C_i class by the number of all elements.

$$P(C_i) = \frac{C_i}{|C|} \quad (6)$$

   b- $P(X|C_i)$, is calculated as in Eq. 7 since *X* is an n-element property vector. Since the x_i values are considered independent of each other, there is no need to calculate the *P(X)* value.

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \quad (7)$$

4. As a result, the C_i class, which has the largest $P(X|C_i)\,P(C_i)$ value, is determined as the class of *X*.

2.4. Multinominal Naïve Bayes

Multinominal Naïve Bayes is widely used because it is a fast, easy-to-apply, and effective method (Rennie, Shih, Teevan, & Karger, 2003). It is calculated with the formula in Eq. 5 like MNB and NB. In the MNB classifier, different from the NB, the $P(x_k|C_i)$ value is calculated as shown in Eq. 8.

$$P(x_k|C_i) = \frac{N_{ki}}{N_i} \quad (7)$$

where N_{ki} indicates the total frequency of x_k in samples with class C_i containing the x_k attribute. N_k indicates the total frequency of the features included in the samples in the same class (Çoban, 2016; Rennie et al., 2003).

2.5. k-Nearest Neighbor

In this classification method, the k samples nearest to the sample to be classified are calculated by some distance measurement methods and the plural class of the calculated samples is assigned as the class. The k value is a parameter entered by the expert making the classification. Entering a too large k value may cause dissimilar records to be collected together and entering a too small value may cause some records to be assigned to different classes (Khan, Ding, & Perrizo, 2002).

2.6. Support Vector Machines

This is a machine learning algorithm developed by Cortes and Vapnik (1995) for two-group classification problems. Although it was developed for two-group classification problems, it can also be used in multi-group classification problems with planar separation mechanisms in three-dimensional space and hyperplanar separation in multi-dimensional space (Güran, Uysal, & Doğrusöz, 2014).

2.7. Evaluation of Classification Results

In classification processes performed with machine learning, metrics such as accuracy, recall, precision, and f-score are often used to evaluate classification success (Altunkaynak, 2017; Köse, 2018; Parlar & Özel, 2016). These metrics are briefly explained below.

Accuracy: This is a metric that shows what percentage of classified records has been correctly classified. It is obtained by dividing the correct number of classified records by the total number of records. Accuracy is calculated as shown in Eq. 9 in a two-class (positive-negative) classification process (Köse, 2018).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where TP (True Positive) = number of records correctly classified as positive, TN (True Negative) = number of records correctly classified as negative, FP (False Positive) = number of records falsely classified as positive, and FN (False Negative) = number of records falsely classified as negative.

Recall: Also called the true positive rate (TPR), recall is a metric that shows how many of the true positives were correctly classified. Recall is calculated as shown in Eq. 10 (Köse, 2018).

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

Precision: Precision is calculated as shown in Eq. 11 by dividing the number of correctly classified records by the total estimated number of positives. Precision is a metric that specifies how many of the positives predicted by the classification model were true positives (Shung, 2020).

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

F-Score: This is an evaluation indicator that takes into account both the recall and the precision metrics. It is based on the efficiency criteria of Rijsbergen (1979). It is calculated as shown in Eq. 12 by taking the harmonic means of precision and recall ("F1 score," 2006; Miner et al., 2012).

$$F - Score = \frac{2 * Precision * recall}{Precision + recall} \quad (12)$$

3. Experimental Setup

An experiment was carried out in this section to achieve the aims stated in the introduction. The steps for this experiment are given below.

3.1. Data Preparation

Since the study is a sentiment analysis study, the data should be prepared for analysis. The following operations were carried out for this preparation.

Identifying Turkish Reviews: Since only reviews written in Turkish were to be used in the study, it was necessary to determine the language in which the reviews were written. Many software or services do this. The language values of the data were determined using the language detection method in the Google Translation API[3] and only reviews detected to be in Turkish were used Google, 2019).

Cleaning and Lower Casing: All other characters in the text such as punctuation marks, numbers, and special characters are cleared. All characters are converted to lowercase letters.

Term Normalization: The data is normalized as explained in Section 2.2 using Zemberek-NLP to eliminate spelling errors.

Stemming and Negation: All words are separated using Zemberek-NLP to reduce the number of attributes.

Since Turkish is an agglutinative language, negativity in words is made by suffixes. A negative word can become positive as a result of stemming. Therefore, negative words should be found and appropriate negativity suffixes such as "sız/siz," "me/ma," and "lı/li" should be added to their roots. This process is called negation (Gezici & Yanıkoğlu, 2018). This process was applied to all words that needed to be negated in the sentiment analysis.

Stop Words Removal: Finally, all stop words defined by Sevindi (2013) were deleted. The data were prepared for sentiment analysis. After removing non-Turkish reviews and all stop words, the number of reviews decreased from 49,031 to 37,912.

3.2. Preparation of Data Sets

A number of data sets were created to achieve the aims mentioned in the introduction. To create these data sets, the 37,464 reviews prepared for sentiment analysis in section 3.1 were read one by one by the researcher and the positive, negative, and neutral comments were determined. In other words, manual labeling was performed. The data were also labeled according to the scores. Reviews with scores of four and five were labeled as positive, those with three as neutral, and those with one and two as negative. As a result of the automatic labeling process according to scores, 3458 records were labeled as negative, 5556 as neutral, and 28898 as positive. As a result of manual labeling, 5614 records were labeled as negative, 5658 as neutral, and 26640 as positive.

After the labeling process, data sets were created by taking an equal number of samples from each class. The data sets were created according to the number of classes, the status of labeling, the stems of the words, and negation. The data sets are presented in Table 1. A systematic code was given to each data set for easy use.

As seen in Table 1, data sets are first separated according to their labeling status to determine how the type of labeling affected classification success. Since both two-class and three-class sentiment analysis will be performed with the data sets, they are also divided according to the number of classes. To determine the effect of stemming and negation on classification

[3] Google Translation API is a translation and language detection service that supports over 100 languages (Google, 2019).

Table 1. Data sets used in the study

Labeling Status	Number of Classes	Stemming and Negative Condition	Data Set Code
Labeled Manually	Two (Positive, Negative)	Not Stemmed	M2-DS1
		Stemmed and Not Negated	M2-DS2
		Stemmed and Negative	M2-DS3
	Three (Positive, Negative, Neutral)	Not Stemmed	M3-DS1
		Stemmed and Not Negated	M3-DS2
		Stemmed and Negated	M3-DS3
Auto Labeled by Score	Two (Positive, Negative)	Not Stemmed	S2-DS1
		Stemmed and Not Negated	S2-DS2
		Stemmed and Negated	S2-DS3
	Three (Positive, Negative, Neutral)	Not Stemmed	S3-DS1
		Stemmed and Not Negated	S3-DS2
		Stemmed and Negated	S3-DS3

success, the data sets are divided into three according to the stems of the words and negation.

3.3. Experiment

The data sets were classified using the classic machine learning algorithms NM, MNB, k-NN, and SVM over the WEKA[4] software. The adjustments made in the WEKA software for the classification are listed below.

- The String to Word Vector filter was used to convert the texts to numeric data.
- TF-IDF was used for term weighting.
- The texts in the data sets are divided into word-based n-grams. Tests were conducted with values of 1-gram, 2-gram, and 3-gram.
- In the k-Nearest Neighbor algorithm, the value of k was considered as three, considering the uses in the literature (Silahtaroğlu, 2013). The default parameters of the WEKA software are used in other classification algorithms.
- All data sets were validated by 10-fold cross-validation and classification results were obtained.

4. Results

With the data sets mentioned in the previous section, 144 tests were carried out. With the tests performed, the

Table 2. Classification results of the two-class sentiment analysis process

Labeling Type	Data set	n-gram	NB	MNB	SVM	K-NN
Manually Labeled	M2-DS1	1-gram	0,88	0,93	**0,94**	0,79
		2-gram	0,87	**0,94**	**0,94**	0,63
		3-gram	0,87	**0,94**	0,93	0,62
	M2-DS2	1-gram	0,86	0,92	**0,94**	0,81
		2-gram	0,87	**0,94**	**0,94**	0,71
		3-gram	0,87	**0,94**	**0,94**	0,70
	M2-DS3	1-gram	0,86	0,93	**0,95**	0,81
		2-gram	0,87	0,94	**0,95**	0,70
		3-gram	0,87	**0,94**	**0,94**	0,69
Average F-score			0,87	**0,94**	**0,94**	0,72
Automatically Labeled by Score	S2-DS1	1-gram	0,84	0,87	**0,91**	0,36
		2-gram	0,85	0,87	**0,92**	0,35
		3-gram	0,85	0,87	**0,92**	0,35
	S2-DS2	1-gram	0,86	0,86	**0,91**	0,36
		2-gram	0,86	0,87	**0,93**	0,35
		3-gram	0,86	0,87	**0,93**	0,35
	S2-DS3	1-gram	0,86	0,86	**0,91**	0,36
		2-gram	0,86	0,87	**0,93**	0,35
		3-gram	0,86	0,87	**0,93**	0,35
Average F-score			0,86	0,87	**0,92**	0,35

[4] Weka (Waikato Environment for Knowledge Analysis) is an open-source software used in data mining and machine learning (Witten, Frank, Hall, & Pal, 2016).

Table 3. Contribution of stemming and negation to classification success in two-class sentiment analysis

Data Set	n-gram	NB	MNB	SVM	K-NN
M2-DS2 – M2-DS1 (Stemming)	1-gram	-0,02	-0,01	0	0,03
	2-gram	-0,01	-0,01	0	0,08
	3-gram	-0,01	-0,01	0	0,08
M2-DS3 – M2-DS2 (Negation)	1-gram	0,01	0	0	0
	2-gram	0,01	0	0,01	-0,01
	3-gram	0,01	0	0	-0,01
S2-DS2 – S2-DS1 (Stemming)	1-gram	0,02	-0,01	0	0,01
	2-gram	0,01	-0,01	0	0
	3-gram	0,01	-0,01	0,01	0
S2-DS3 – S2-DS2 (Negation)	1-gram	0	0	0	0
	2-gram	0	0	0	0
	3-gram	0	0	0	0

contribution of labeling, stemming, and negation to classification success were measured, and then the classification success of different classifiers was compared. As a result of these tests, 144 machine learning models were created. The f-score values showing the classification success of machine learning models are given in the tables.

When Table 2 is examined, it is seen that the most successful two-class sentiment analysis classification is performed with the SVM classifier. It is seen that the SVM and MNB algorithms achieved very close classification results with manually labeled data sets. In manually labeled two-class data sets (M2-DS1, M2-DS2, M2-DS3), average f-scores were 0.94 for SVM and MNB, 0.87 for NB, and 0.72 for k-NN. In automatically labeled two-class data sets (S2-DS1, S2-DS2, S2-DS3), average f-scores were 0.92 for SVM, 0.87 for MNB, 0.86 for NB, and 0.35 for k-NN. These results show that classification with manually labeled two-class data sets was more successful than classification with automatically labeled two-class data sets according to scores.

No preprocessing was applied to the data in the data sets ending with DS1. The words in the data sets ending with DS2 were applied stemming. The words in the data sets ending with DS3 were applied both stemming and negation. Besides, each data set was classified into tokens with parameters of one, two, and three grams and the classification process was carried out. The difference between the classification results of data sets ending with DS2 and DS1 shows the contribution of stemming to classification success. Similarly, the difference between the classification results of data sets ending with DS3 and DS2 shows the contribution of negation to classification process. In the light of this information, when Table 3 is examined, it is seen that stemming manually labeled two-class data sets with the k-NN classifier had a significant contribution to classification success. On the other hand, in the classification of automatically labeled two-class data sets with the NB, SVM, and k-NN classifiers, stemming had little contribution. In the classification of manually labeled two-class data sets, negation had a general contribution to classification success. However, the same does not apply to automatically labeled two-class data sets, where negation was found to have no contribution.

In Table 4, the classification results obtained for three-class sentiment analysis are given. When the table

Table 4. Three-class sentiment analysis classification results

Labeling Type	Data set	n-gram	NB	MNB	SVM	K-NN
Manually Labeled	M3-DS1	1-gram	0,70	0,72	**0,79**	0,61
		2-gram	0,70	0,74	**0,80**	0,49
		3-gram	0,70	0,74	**0,79**	0,49
	M3-DS2	1-gram	0,70	0,72	**0,79**	0,64
		2-gram	0,70	0,73	**0,79**	0,55
		3-gram	0,70	0,73	**0,79**	0,54
	M3-DS3	1-gram	0,70	0,72	**0,80**	0,64
		2-gram	0,70	0,74	**0,80**	0,54
		3-gram	0,71	0,73	**0,79**	0,54
Average F-score			0,70	0,73	**0,79**	0,54
Automatically Labeled by Scores	S3-DS1	1-gram	0,59	0,63	**0,70**	0,20
		2-gram	0,59	0,64	**0,70**	0,19
		3-gram	0,59	0,63	**0,69**	0,18
	S3-DS2	1-gram	0,62	0,62	**0,71**	0,22
		2-gram	0,62	0,63	**0,71**	0,19
		3-gram	0,62	0,62	**0,70**	0,19
	S3-DS3	1-gram	0,63	0,63	**0,71**	0,23
		2-gram	0,62	0,63	**0,71**	0,19
		3-gram	0,62	0,63	**0,70**	0,19
Average F-score			0,61	0,63	**0,70**	0,20

Table 2. Contribution of stemming and negation processes to classification success in the three-class sentiment analysis process

Data set	n-gram	NB	MNB	SVM	K-NN
M3-DS2 – M3-DS1 (Stemming)	1-gram	0	-0,01	0	0,03
	2-gram	0	0	0	0,06
	3-gram	0	-0,01	0	0,05
M3-DS3 – M3-DS2 (Negation)	1-gram	0	0	0,01	0
	2-gram	0	0	0,01	-0,01
	3-gram	0	0	0	0
S3-DS2 – S3-DS1 (Stemming)	1-gram	0,03	-0,01	0,01	0,03
	2-gram	0,03	-0,01	0,01	0
	3-gram	0,02	-0,01	0,01	0
S3-DS3 – S3-DS2 (Negation)	1-gram	0,01	0,01	0	0
	2-gram	0,01	0	0	0
	3-gram	0,01	0	0	0

is examined, it is seen that the most successful classifier in the three-class sentiment analysis was the SVM. In manually labeled three-class datasets (M3-DS1, M3-DS2, M3-DS3), average f-scores were 0.79 for SVM, 0.73 for MNB, 0.70 for NB classifier, and 0.54 for k-NN. In automatically labeled three-class data sets (S3-DS1, S3-DS2, S3-DS3), average f-scores were 0.70 for SVM, 0.63 for MNB, 0.61 for NB, and 0.20 for k-NN. These results show that classification with manually labeled three-class data sets was more successful than classification with automatically labeled three-class data sets.

Table 5 shows the contributions of stemming and negation in three-class sentiment analysis to classification success. When Table 5 is examined, it is seen that stemming had a significant contribution to classification success in the classification of manually labeled three-class data sets using the k-NN classifier. However, stemming had little contribution in the classification of automatically labeled three-class data sets using the NB, SVM, and k-NN classifiers. It is also seen that negation contributed to classification success in the processes with the NB, MNB, and SVM classifiers.

5. Conclusion

In this study, a sentiment analysis was carried out with the machine learning method using Turkish reviews for touristic sites. The classification success of the NB, MNB, SVM, and k-NN classifiers was compared in the sentiment analysis, the effects of stemming and negation on classification success were investigated, and the effect of the type of labeling on classification success was measured. Both two-class (positive-negative) and three-class (positive-negative-neutral) sentiment analyses were performed. As a result, the most successful classification result in two-class sentiment analysis was reached using the SVM classifier with an f-score of 0.95. The most successful classification result in three-class sentiment analysis was again reached using the SVM classifier with an f-score of 0.80. It was concluded that the classification results were more successful in the sentiment analysis of manually labeled data sets compared to data sets automatically labeled according to scores. It was determined that stemming significantly contributed to classification success, especially in the k-NN classifier. It was observed that stemming had little contribution to classification success in the NB and SVM classifiers, and a negative effect on classification success in the MNB classifier. Negation resulted in a general increase in classification success in the NB, MNB, and SVM classifiers.

Considering that similar studies in the literature reported f-score values of 0.78 – 0.92 for two-class sentiment analyses and 0.59 – 0.78 for three-class sentiment analyses, it can be said that the results are quite successful (Çoban et al., 2015; Kaya et al., 2012; Kaynar et al., 2016; Velioğlu et al., 2018; Yıldırım et al., 2015). While the findings suggest that the sentiment analysis model created here is feasible for Turkish touristic site reviews, using the data sets in further research and comparing the findings will result in better interpretations.

In the sentiment analysis performed here, the bag-of-word method was used for word representation and classic machine learning algorithms were used as classifiers. Future studies should aim to measure the classification performance of the data sets created here using different word representation methods such as fastText, word2Vec, or glove, along with different machine learning techniques such as artificial neural networks.

References

Akın, A.A., and Akın, M.D., 2018. Zemberek-NLP. Available at: https://github.com/ahmetaa/ zemberek-nlp

Altunkaynak, B., 2017. Veri Madenciliği Yöntemleri ve R Uygulamaları, 2., Seçkin Yayıncılık, Ankara.

Aydoğan, E., and Akcayol, M. A. 2016. A comprehensive survey for sentiment analysis tasks using machine learning techniques. International Symposium on Innovations in Intelligent Systems and Applications, 2-5 August 2016, Sinaia, Romania, pp: 1-7.

Baccianella, S., Esuli, A., and Sebastiani, F. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Seventh

International Conference on Language Resources and Evaluation, 17-23 May, Valletta, Malta, pp: 2200-2204.

Bag-of-Words model 2007. Available at: https://en.wikipedia.org/wiki/ Bag-of-words_model

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Royal Society, 53, 370-418.

Beal, V., n.d. Unstructured data. Available at: https://www.webopedia.com/TERM/U/unstructured_data.html

Bilgin, M., and Şentürk, İ. F. 2017. Sentiment analysis on Twitter data with semi-supervised Doc2Vec. International Conference on Computer Science and Engineering, 5-8 October 2017, Antalya, Turkey, pp: 661-666.

Blumberg, R., and Atre, S., 2003. The problem with unstructured data. Available at: http://soquelgroup.com/wp-content/uploads/2010/01/dmreview_0203_problem.pdf

Can, Ü., and Alataş, B., 2017. Duygu analizi ve fikir madenciliği algoritmalarının incelenmesi. International Journal of Pure and Applied Sciences, 3(1), 75-111.

Chang, Y.-C., Ku, C.-H., and Chen, C.-H., 2019. Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. International Journal of Information Management, 48, 263-279.

Çoban, Ö., 2016. Metin sınıflandırma teknikleri ile türkçe twitter duygu analizi (Master's Thesis), Atatürk University.

Çoban, Ö., Özyer, B., and Özyer, G. T. 2015. Sentiment analysis for Turkish Twitter feeds. 23nd Signal Processing and Communications Applications Conference, 16-19 May 2015, Malatya Turkey, pp: 2388-2391.

Cortes, C., and Vapnik, V., 1995. Support-vector networks. Machine Learning, 20(3): 273-297.

Data Scraping, 2021. Available at: https://en.wikipedia.org/wiki/Data_scraping

Datashake, 2021. Available at: https://www.datashake.com/

Esen, M. F., and Türkay, B., 2017. Turizm endüstrilerinde büyük veri kullanımı. Journal of Tourism and Gastronomy Studies, 4 (4), 92-115.

Esuli, A., and Sebastiani, F., 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. Fifth International Conference on Language Resources and Evaluation, 22-28 May 2006, European Language Resources Association, Geona, Italy, pp: 417-422.

F1 score, 2006. Available at: https://en.wikipedia.org/wiki/F1_score

Gezici, G., and Yanıkoğlu, B., 2018. Sentiment analysis in Turkish. In: Turkish Natural Language Processing, K. Oflazer & M. Saraçlar (Eds.), Springer International Publishing, Cham.

Google, 2019. Google Translation API. Available at: https://cloud.google.com/translate/

Güran, A., Uysal, M., and Doğrusöz, Ö., 2014. Destek vektör makineleri parametre optimizasyonunun duygu analizi üzerindeki etkisi. Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering, 16 (48), 86-93.

Harris, Z. S., 1954. Distributional structure. Word, 10 (2-3): 146-162.

Hearst, M., 2003. What is text mining. Available at: https://people.ischool.berkeley.edu/~hearst/text-mining.html

Kaya, M., Fidan, G., and Toroslu, I. H., 2012. Sentiment analysis of Turkish political news. International Joint Conferences on Web Intelligence and Intelligent Agent Technology, 4-7 December 2012, Macau, China, pp: 174-180.

Kaynar, O., Görmez, Y., Yıldız, M., and Albayrak, A., 2016. Makine Öğrenmesi Yöntemleri ile Duygu Analizi. International Artificial Intelligence and Data Processing Symposium, 17-18 September 2016, Malatya, Turkey, pp: 234-241.

Khan, M., Ding, Q., and Perrizo, W., 2002. K-nearest neighbor classification on spatial data streams using p-trees. 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 6-8 May 2002, Berlin, Germany, pp: 517-518.

Kızılkaya, Y. M., 2018. Duygu analizi ve sosyal medya alanında uygulama (Doctoral Disertation), Bursa Uludağ University.

Köse, İ., 2018. Veri madenciliği teori uygulama ve felsefesi, Papatya Yayıncılık Eğitim, İstanbul.

Ku, L.-W., Liang, Y.-T., and Chen, H.-H., 2006. Opinion extraction, summarization and tracking in news and blog corpora. AAAI Spring Symposium, 27-29 March 2006, CA, USA, pp: 100-107.

Kulcu, S., and Dogdu, E., 2016. A scalable approach for sentiment analysis of Turkish tweets and linking tweets to news. International Conference on Semantic Computing, 4-6 February 2016, CA, USA, pp: 471-476.

Liu, B., 2012. Sentiment analysis and opinion mining. Morgan & Claypool Publishers, Williston.

Liu, K.-L., Li, W.-J., and Guo, M., 2012. Emoticon smoothed language models for twitter sentiment analysis. Twenty-Sixth AAAI Conference on Artificial Intelligence, 22-26 July 2012, Ontario, Canada, pp: 1678-1684.

Lohr, S., 2012. The age of big data. Available at: https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Maron, M. E., 1961. Automatic Indexing: An Experimental Inquiry. J. ACM, 8(3), 404-417.

Meral, M., & Diri, B., 2014. Sentiment analysis on Twitter. 22nd Signal Processing and Communications Applications Conference, 23-25 April 2014, Trabzon, Turkey, pp: 690-693.

Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., and Delen, D., 2012. Practical text mining and statistical analysis for non-structured text data applications. Academic Press, Waltham, MA.

Özyurt, B., and Akçayol, M. A., 2018. Fikir madenciliği ve duygu analizi, yaklaşımlar, yöntemler üzerine bir araştırma. Selcuk University Journal of Engineering, Science and Technology, 6(4), 668-693.

Pang, B., Lee, L., and Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. Conference on Empirical Methods in Natural Language Processing, 6-7 July 2002, Philadelphia, USA, pp: 79-86.

Parlar, T., and Özel, S. A., 2016. A new feature selection method for sentiment analysis of Turkish reviews. International Symposium on Innovations in Intelligent Systems and Applications, 2-5 August 2016, Sinaia, Romania, pp: 1-6.

Provost, F., and Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. Big data, 1(1), 51-59.

Rennie, J. D. M., Shih, L., Teevan, J., and Karger, D. R., 2003. Tackling the poor assumptions of naive bayes text classifiers. International Conference on Machine Learning, 21-24 Agust 2003, Washington, DC, USA, pp: 616-623.

Rijsbergen, C. J. V., 1979. Information Retrieval. Butterworth-Heinemann.

Saad, M. K., 2010. The impact of text preprocessing and term weighting on arabic text classification (Master's Thesis), The Islamic University.

Salur, M. U., Aydın, İ., and Alghrsi, S. A., 2019. SmartSenti: A twitter-based sentiment analysis system for the smart tourism in Turkey. International Artificial Intelligence and Data Processing Symposium, 21-22 September 2019, Malatya, Turkey, pp: 1-5.

Şeker, S. E., 2016. Duygu analizi. Management Information Systems Encyclopedia, 3(3), 21-36.

Sevindi, B. İ., 2013. Comparison of supervised and dictionary based sentiment analysis approaches on Turkish text (Master's Thesis) Gazi University.

Shung, K. P., 2020. Accuracy, Precision, Recall or F1? Available at: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

Silahtaroğlu, G., 2013. Veri madenciliği: Kavram ve algoritmaları. Papatya Yayıncılık Eğitim, İstanbul.

Spärck Jones, K., 2004. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 60 (5), 493-502.

Taecharungroj, V., and Mathayomchan, B., 2019. Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75, 550-568.

Toçoğlu, M. A., 2018. Lexicon-based emotion analysis in Turkish (Doctoral Dissertation), Dokuz Eylül University.

Türkmenoğlu, C., 2015. Türkçe metinlerde duygu analizi (Master's Thesis), Istanbul Technical University.

Türkmenoglu, C., and Tantug, A. C., 2014. Sentiment analysis in Turkish media. International Conference on International Conference on Machine Learning, 21-26 June 2014, Beijing, China.

Velioğlu, R., Yıldız, T., and Yıldırım, S., 2018. Sentiment analysis using learning approaches over emojis for Turkish tweets. 3rd International Conference on Computer Science and Application Engineering, 20-23 September, Sanya, China, pp: 303-307.

Waykole, R. N., and Thakare, A., 2018. A Review of feature extraction methods for text classification. IJAERD, 5 (04), 351-254.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J., 2016. The weka workbench. Online appendix for "data mining: practical machine learning tools and techniques". Morgan Kaufmann, Burlington, MA.

Yıldırım, E., Çetin, F. S., Eryiğit, G., and Temel, T., 2015. The impact of NLP on Turkish sentiment analysis. TBV Journal of Computer Science and Engineering, 7(1), 41-51.

Yıldız, O., 2016. Metin madenciliğinde anahtar kelime seçimi bir üniversite örneği, Journal of Management Information Systems, 2(1), 29-50.

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 113-126
DOI: 10.38016/jista.933005

Smartness and Strategic Priority Assessment in Transition to Mobility 4.0 for Smart Cities

Hakan İnaç [1*], Ercan Öztemel [2], Mehmet Emin Aydın [3]

[1]Marmara University, Institute of Pure and Applied Sciences, Department of Engineering Management, İstanbul/Turkey

[2]Marmara University, Faculty of Engineering, Department of Industrial Engineering, İstanbul/Turkey

[3]University of the West of England, Department of Computer Science and Creative Technologies, Bristol/UK

hkninac@gmail.com, eoztemel@marmara.edu.tr, mehmet.aydin@uwe.ac.uk

Abstract
Digital transformation of urban transportation services attracts significant attention within the scope of smart city studies. It is known as a strategic action with many investment opportunities for the future cities, therefore it attracts the attention of enterprises and institutions. A sound and sustainable digital transformation requires substantial decision support in order to customize the needs and priorities of cities, especially in developing countries, where a lack of knowledge and skills is the case. Mobility 4.0 is recognized as the most up-to-date state of art technology and vision for urban transportation in the age of smart cities. Mobility as a Service (MaAS) is one of the most prominent Mobility 4.0 components that needs to be developed with an integrated management approach, which is available for research and development investments that have defined conceptually yet. In this study, the smartness level of city transport services is identified with a four-level transition approach. The transition from Mobility 3.0 to Mobility 4.0 for cities has been carried out with conceptual integrated management design of MaAS and demonstrated with a case study. The study also adapted a series of new strategically targeted transformation priorities for cities with reference to research and development strategies.

Keywords: Digital Transformation, MaAS, Mobility 4.0, Transportation Management, Nominal Group Technique, Smart City.

Akıllı Şehirler İçin Mobilite 4.0 Geçişinde Akıllılık ve Stratejik Öncelik Değerlendirmesi

Öz
Kentsel ulaşım hizmetlerinin dijital dönüşümü, akıllı şehir çalışmalarında öne çıkan ve büyük ilgi gören bir konudur. Bu dönüşüm geleceğin şehirleri için birçok yatırım fırsatı içeren stratejik bir eylem olarak bilinir, bu nedenle şirketlerin ve kurumların ilgisini çekmektedir. Güçlü ve sürdürülebilir bir dijital dönüşüm, özellikle bilgi ve beceri eksikliğinin söz konusu olduğu gelişmekte olan ülkelerde, şehirlerin ihtiyaçlarını ve önceliklerini özelleştirmek için etkin bir karar desteği gerektirir. Mobilite 4.0, akıllı şehirler çağında kentsel ulaşım için en güncel teknoloji ve vizyon olarak kabul edilmektedir. Bir hizmet olarak hareketlilik (MaAS), kavramsal olarak henüz tanımlanmış araştırma ve geliştirme yatırımları için mevcut olan entegre bir yönetim yaklaşımıyla geliştirilmesi gereken en önemli Mobilite 4.0 bileşenlerinden biridir. Bu çalışmada, şehir içi ulaşım hizmetlerinin akıllılık seviyesi dört aşamalı bir geçiş yaklaşımı ile belirlenmiştir. Şehirler için Mobilite 3.0'dan Mobilite 4.0'a geçiş, MaAS'ın kavramsal entegre yönetim tasarımı ile örneklenmiş ve bir vaka çalışması ile gösterilmiştir. Çalışma da ayrıca, araştırma ve geliştirme stratejilerine referansla şehirler için stratejik olarak hedeflenmiş bir dizi yeni dönüşüm önceliği ile ilgili önerilerde bulunulmuştur.

Anahtar Kelimeler: Dijital Dönüşüm, MaAS, Mobilite 4.0, Ulaşım Yönetimi, Nominal Grup Tekniği, Akıllı Şehir.

* Corresponding Author.
E-mail: hkninac@gmail.com

Recieved : 5 May 2021
Revision : 16 May 2021
Accepted : 15 June 2021

1. Introduction

Digital transformation is a crucial instrument of smart cities, which implies digitizing business processes using information and communication technologies (ICT) infrastructure, while a digital system supported by artificial intelligence technologies is known as an intelligent system. Emerging technologies help facilitate widespread use of smart services in urban life and city management, noting that use of smart systems is a transformation process that requires investment and takes time (Anttiroiko et al., 2014). In this regard, policy makers need to plan strategically focused transformation. It is known that credible assessment models are required to identify smartness level of digitalisation processes and to evaluate the efficiency, efficacy and sustainability of Intelligent Transport Systems (ITS).

Digital transformation is an ongoing process, which introduces new services while updating many existing services. Particularly, the new services are initiated in the conceptual level, either through definitions originate from use cases or from prototypes, which remains uncertain and risky in comparison to the mature and well-defined services. There are cities that decide to take initiative for investing in infrastructures required for various services as part of digital transformation, while there are many others which may not afford to exercise new developments due to imposed uncertainties and risks of digital services owing to various city-specific circumstances, such as budgets, knowhow etc. Some cities realize the digital transformation process with sustainable city-specific strategies by experimenting with prototypes of digital services with pilot applications. Others may choose a risk-free digital transformation strategy with certain services which already have defined infrastructure service requirements and related process. In this respect, the determination of digitalisation targets for transportation services and the choice of a digital transformation strategy compatible with cities are investigated within the scope of this study.

The digitalization of urban transport systems is addressed in the literature with the concepts of smart city and Mobility 4.0. There are a number of city indicators for smart city assessment. Using various combinations of these, it is possible to comment on the smart transformation of cities. The most important causes of emissions in the cities emerge from the transportation services (Colvile et al., 2001). Traffic congestion, lack of public transport capacity, and lack of infrastructure are the most frequently complained problems (Redman et al., 2013). According to studies conducted in European Union countries, fossil fuel consumption, which corresponds to 24% of total greenhouse gas emissions and 1/3 of total energy consumption, originates from the transportation sector (Foltýnová et al., 2018).

Smart cities have been evaluated with a general maturity approach that implies situational analysis for current position and evaluations for the implementation of new technologies in the literature. The studies exploit weighted average method using indicators determined as smart city evaluation model (Tay et al., 2018). Being a smart city is not an objective, but a means to achieve a goal (Akpınar, 2019). The approaches presented with maturity models make an assessment of the current situation. However, these studies do not provide decision support as they do not set digital transformation targets for cities according to their potential.

On the other hand, the concept of Mobility 4.0 can be viewed as an implementation of Industry 4.0 to perform digital transformation of transportation services. The historical progress of Mobility 4.0 demonstrates how the stages of mobility develops alongside the emergence of smart city concepts, trends and smartness level. The development steps of the Mobility 4.0 are specified in terms of transportation services and utilized as a key performance indicator suggesting the level of smartness. We propose measurable objectives in determining the level of smartness specific to cities, rather than comparing the current level of digitalisation of cities with various criteria. This paper elaborates a digital transformation framework that takes into account the characteristics, capacities and specific strategies of cities based on expert opinion for decision makers responsible for the digital transformation of cities.

The proposed framework evaluates the digital transformation on component basis. The smartness level of each component is calculated by the model according to expert opinions utilising smart city indexes. Investments by local governments in research and development (R&D) are needed for the digital transformation of conceptual services. Mobility as a service (MaAS) can be considered in this context. Additionally, in this study, the transition process of MaAS, which is conceptually defined only in literature, is analysed for different categorized cities with the transformation of digitization trends and Mobility 4.0 components. This approach assists to ascertain which city needs a priority of smartness transformation in terms of Mobility 4.0 components of MaAS first. A strategic evaluation is made according to the R&D investment potentials for the components under consideration and it is recommended to rank transformation priorities.

Expert views suggest that a healthier digital transformation process requires starting form strategic level, where the common strategies of cities with respect to the digital transformation of transportation services are developed with leading, sustainable and cost-efficient integral systems. For a sustainable ecosystem, it is necessary to reduce the use of private transportation offering comfortable and highly capacitated public transportation and increase mobility, while, for a habitable environment, cities should prioritise

accessible, comfortable, transparent and economical transportation services that prevent noise and noise pollution.

The framework introduced in this work evaluates the qualities and potentials of cities imposes working principles to be considered as a road map. The following contributions are introduced by this article:

(1) The concept of Mobility 4.0 is summarized with its components, explained with its historical development, user trends and the level of smartness in digitization. This is useful to grasp the essences of the digital transformation of cities in transportation services.

(2) An integrated management approach associated with Mobility as a service (MaAS) in this respect is considered based on R&D investment decisions, which is required for smart transportation platforms and structural transformation.

(3) The concept of the level of smartness in transportation systems is presented. It proposes a four-stage transformation model for digital transformation defining a 4-tier operation and management process to evaluate integrated components.

(4) The general flowchart/road map of the proposed approach is included.

(5) The strategically focused prioritization approach is explained and exemplified.

(6) The implementation is comparatively evaluated.

2. Literature Review and Background

Modern transportation faces prominent challenges such as congestion, pollution, greenhouse gas emissions, the scarcity of oil, growing last mile inefficiency in increasing a city's density, growing transport demand, which appear to impose causal effects upon increasing inefficiency in the transportation systems. Therefore, it enforces the policymakers and industrial leads to address these issues for sustainability concerns (Müller and Liedtke). The use of advanced technology and systematic management approach will increase efficiency and reduce pressure in the use of limited resources. The main strategy to be implemented in the solution of transportation-related city problems is to design people-oriented, sustainable and liveable/habitable systems (Angelidou et al., 2018). Smart cities provide opportunities to use integrated service management and common infrastructure (Razaghi and Finger, 2018).

Mobility 4.0 concept is defined as "Integrated Urban Mobility", which aims to provide effective transport services by integrating all kind of emerging technologies in transportation (Sochor et al., 2015, Smith et al., 2018). Private and public sector cooperation (İnaç and Dönmez, 2018), mobile data collection (Şentürk and Coulibally, 2020), use of cloud technology (Cottrill, 2020) and artificial intelligence supported service systems are prioritised in the digital transformation of transportation systems components. There is no "one size fits all" Mobility 4.0 and it is critical to consider local specifications for successful implementation (Butler et al., 2020).

It is not trivial to identify the smartness level of smart cities since there are many elements and dimensions that characterize a smart city. It is equally difficult to define and propose a universal standard approach to reach an assessment level with the characteristics of cities around the world. That suggests that smart city assessments should take into account that cities have different visions and priorities to achieve their goals (Albino et al., 2015).

The following studies relevant to transportation components have mainly been carried out on transformation, which can be considered related to intelligent transportation systems. The studies and the components help determine the needs for a general framework of digital transformation.

- ✓ Traffic jam (Mouchili et al., 2018)
- ✓ Lack of parking lots (Ma and Xue, 2020)
- ✓ Accessibility (Mouchili et al., 2018)
- ✓ Traffic Safety (Mouchili et al., 2018)
- ✓ Comfort problems (Lytras and Visvizi, 2018)
- ✓ Noise pollution (Bello et al., 2019)
- ✓ Visual pollution (Bello et al., 2019)
- ✓ Air pollution (Kumar et al., 2015)
- ✓ Limited public transportation capacity (Lytras and Visvizi, 2018)
- ✓ Logistics distribution problems (Crainic et al., 2009, Yaman and Bayğın, 2020).

Many problems are linked by cause and effect relationship. Therefore, it is necessary to develop integrated approaches. A model containing expert opinion that evaluates the digital transformation needs of transportation components together has been developed in this study.

Essentially, there is no definitive approach to the measurement of smart city outcomes, although approaches should be using current indexes and their values for build a framework. The indices help simplifying the work of the engineers of the national and local authorities without requiring general knowledge of the subject and marketing the results to the public opinion with possible policy implementations.

The suggested indexes in the evaluation of the digitalization of transportation are as follows: Traffic Efficiency, Mobility, Traffic Safety, Pollution and Public Transportation Capacity (Kaparias et al., 2011).

The capabilities of cities, the structure of the city, its geographical conditions and the socio-economic structure of the region will naturally affect the digitalization goals. Indices suitable for these qualifications should be added to the evaluation criteria.

Four different development methods were used in the selection of smart city evaluation indicators. First, developers often choose peer-reviewed literature to choose smart city sizes and their indicators are weighting of composite indexes (Pringle, 2011). The second is rational expert opinions, using a combination of interrogative surveys, consultation workshops, the Delphi method and the Analytical Hierarchy Process

(AHP) (Giffinger et al., 2007). Thirdly, it relies on the eco-design approach to involve various stakeholder groups (Manville et al., 2014). Fourth and finally, the best practices of smart city assessment are evaluated. Figure 1 shows which methodology is often used in such studies (Sharifi, 2020).

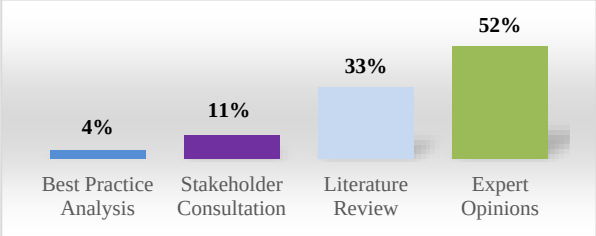


Figure 1. Major methods for development of assessment digital transformation (Sharifi, 2020).

Cities are classified into specific clusters for comparison with respect to various criteria such as economic development level, population, geographical location, ICT infrastructure. The most common classification criterion used is the economic development level of the countries. In this classification, cities are classified as developed country cities, developing countries cities and cities in underdeveloped countries (Kemeny and Storper, 2015). Another classification is done with population. For example, cities with a population of over 10 million are called megacities, while others are ranked as super city, middle city and small city (Ngo et al., 2019). Geographical location suggests that cities are classified as European cities, Far East cities, South American cities, African cities (Nagy, 2016). Meanwhile, in order to identify innovation and creativity level of countries, a classification has been made for countries according to the number of patent applications (Khayyat and Lee, 2015). We adopted a hybrid approach, in this study, to classify the cities using existing criteria in which the economic power, infrastructure opportunities, innovation provision and the impact of the population are taken into account.

The hybrid approach proposed in this study is inspired of research and development (R&D) activities to adapt of new technologies. The idea is to consider smart city digital transformations process the same as the new product R&D strategies that are widely accepted in production and service systems. Companies in existing new product strategies; Classified by "First-to-market", "Follow-the-leader" and "Me-too" strategies. (Morse et al., 2014). Likewise, cities can be categorised into such three classes with respect to their economic strength, innovation potential, and human resources. While cities are setting their strategic objectives are recommended to follow similar strategies;

- ✓ **"First-to-market"** class can be reserved for cities with high innovation potential in developed countries with a population of over 10 million, such as London, Seoul, and New York. In this study, this group of cities are referred to as **"Leading City"** that have R&D investment capabilities and are able to invest in digitalisation-wise future-oriented services defined as concepts. Cities in this group have R&D investment potential.
- ✓ **"Follow-the-leader"** class of cities have lower population and high potential such as Amsterdam and Barcelona. In this study, they are referred to as **"Sustainable City"** that are classified as the cities capable of investing in prototype development for products and are following sustainable strategies for digital transformation with exercising through pilot applications.
- ✓ Cities, such as Istanbul and Moscow, with high potential but limited economic power are defined to be in **"Me-too"** group. In this study, this group of cities are referred to as **"Cost-efficient City"**. They transform developed and accepted digital services in cities.

This paper introduces the use a 4-tier model developed in the assessment of the smartness level based on expert opinion. It has been suggested to choose indexes used in smart city evaluations as indicators. The process flow diagram of the proposed model is presented in Figure 5.

Digitalisation offers substantial benefits to effectively manage infrastructure resources for the solution of transportation problems in the city life but takes a long time. Expert judgments and an implementation model are needed to decide which transport component should primarily be digitalised. The economic power and ICT capacities of the cities also need to be evaluated, as transformation decisions require new investment.

3. Proposed Approach for Digital Transformation

Similar to the digital transformation of transportation in many ways, it is associated with a widely used industrial concepts such as Industry 4.0.

Industry 4.0 is the most up-to-date version of industrial periods, which requires to understand its difference from previous industrial periods. The main technology of Industry 4.0 is the Cyber-Physical System (CPS), which is defined as the combination of physical and cybernetic systems (Lee et al., 2015, Klingenberg and Antunes, 2017). It is due to some important technical advances in Cyber-Physical Systems, the Internet, embedded systems, computer science and artificial intelligence (Hellinger and Seeger, 2011).

3.1. Transition of Mobility 4.0

The transformation of transportation services has been conducted with emerging technologies and

innovations in-line with the development of the industry. The new version of mobility, also known as the new generation, is called "Mobility 4.0". At the time of traditional transport, the trends consisted of road transport by private vehicles and public transport travel by intercity railways. In recent times, multimodal transport services have developed with public-private cooperation. Nowadays, Mobility 4.0 allows the use of advanced technologies for efficient management of transport resources. With the concept of Mobility 4.0, the new trend of transportation management is to increase the sharing economy and mobility. The development of mobility and the concept of Mobility "4.0" is presented in Figure 2.

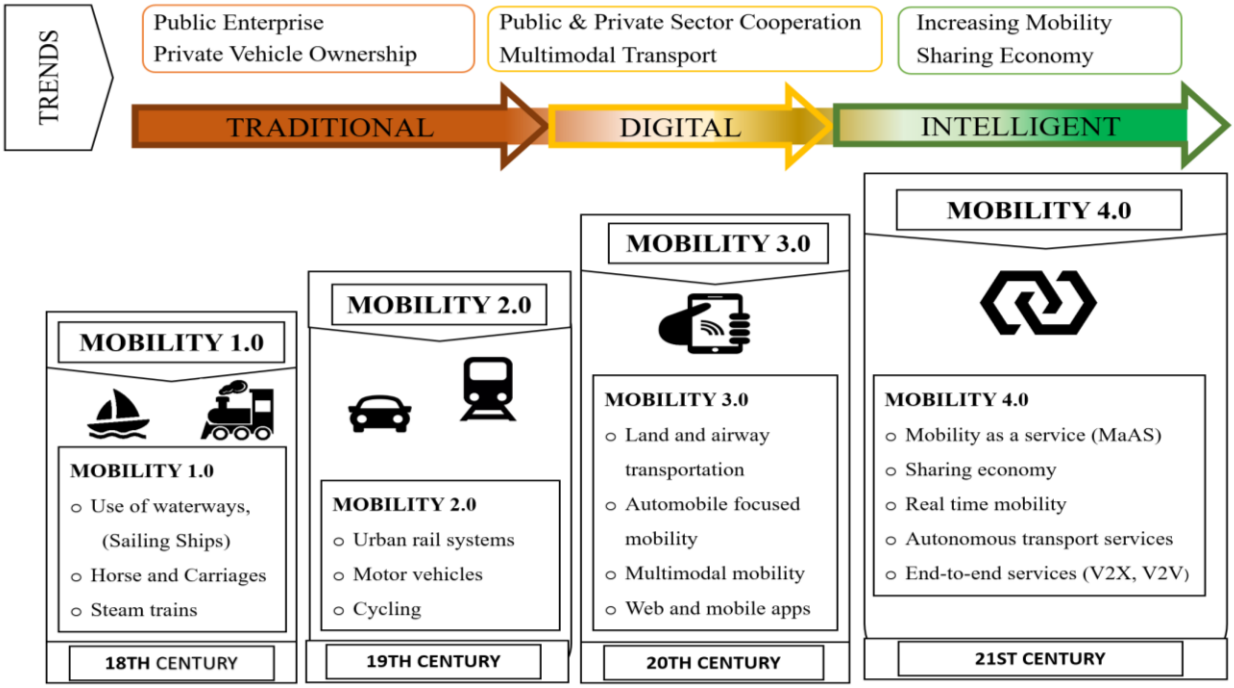


Figure 2. Development of transportation and the concept of Mobility 4.0

Until the end of the 18th century, in the first period of the industry, waterways and sailing ships were used in water transport. In this period, rail transport with steam-powered trains had begun to develop. On the land, horse and horse drawn carriages were used for transportation. Transportation services also changed with the invention of fossil fuel engines in the 19th century. In 1863 the word's first underground railway was opened in London (Halliday, 2013). Bicycles, automobiles, rail systems and maritime transport began to be widely used. These developments and the emerging transportation services until the 20th century are named as the traditional period in this article. Of course, these services constitute the basis of transportation components today. Some transportation components, which are considered to have no need for digital transformation, are expected to serve with traditional management.

In the 20th century, with the 3rd Industrial Revolution, electro mechanics and computer technologies appeared. In this period, land, water, rail and air transportation services developed. In the last years of this period, web-based systems started to be offered in transportation services. The use of electronic ticket systems and web services transportation services has also become widespread. Huge data has begun to be collected for mobile network usage and mobile applications and mobility. Digital services started to be used widely.

In the 21st century, stakeholders' expectations from transportation services have changed with the concepts of Smart city and Industry 4.0. An integrated management approach has been adopted in order to meet the conflicting expectations of the citizens, private sector and state administrations. This period, which is adopted as the Mobility 4.0 phase, covering today and the near future, is called the period of intelligent transportation systems in the study. In the period of Mobility 4.0; cyber physical system components such as artificial intelligence, big data and internet of things are expected to be widely used in transportation services. This period adopts the use of integrated transportation systems in the proposed model approach.

3.2. Components of Mobility 4.0

It is expected that many existing services will change in terms of standards and qualification with the digital

transformation of transportation systems. In this context, digital transformation should propose a re-organization that includes new components in service management. It is recommended to make a transformation assessment with these new components created by digital integrated management.

Advances in ICT enable the integrated management of many transport components. Today, transportation service components are managed in an integrated manner in digitalized cities with developed economies. Integrated transportation management and prominent transportation components are shown in Figure 3.

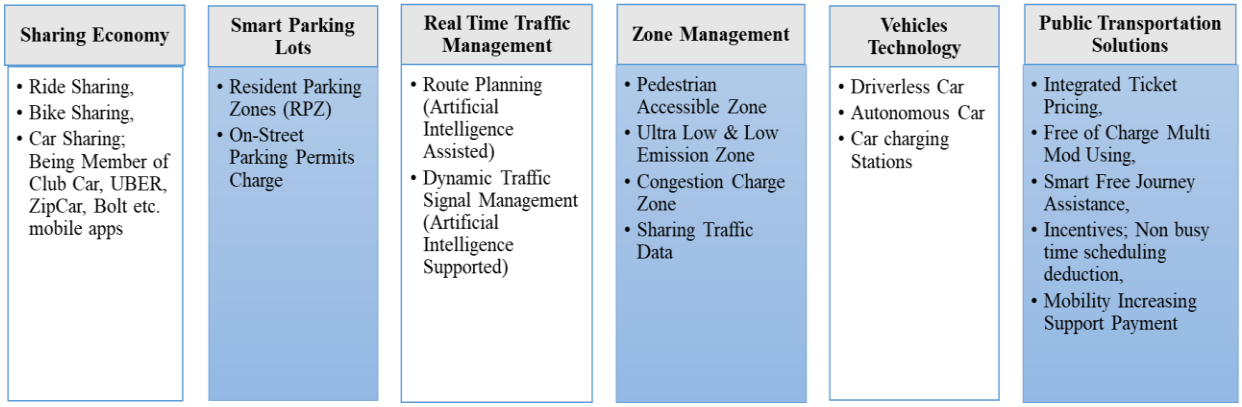


Figure 3. Components of Mobility 4.0 & integrated transportation management

Expectations from transportation services; is changing with the developing technology and the digital world. In the near future, individual car ownership will be replaced by car-sharing economy systems. The diffusion of shared transportation can significantly change the vehicle ownership rate: each car sharing vehicle is estimated to remove 9 to 13 vehicles from the roads (Greenblatt and Shaheen, 2015), as most private cars are used less than 10% of the time (Fagnant and Kockelman, 2014). It is an accepted practice in developed countries that cities are restricted as low emission zones and only vehicles with reduced emission will be allowed access (Wolff and Perry, 2010). Due to the restricted parking lot and traffic congestion, demand suppression policies are implemented with the congestion charging for private vehicles (Metz, 2018).

Given the demand for daily and hourly limited city parking, lots must be suppressed. Cities like London are a thriving commercial and residential area – consequently parking controls are necessary to maintain traffic flow and protect public spaces. Using car parks is often preferable to parking on the street as many roads are reserved for resident permit holders only. Car parks may also allow parking for longer periods (Clayton et al., 2014).

In the coming years, we will see vehicle-to-everything vehicles (V2X) interacting with its environment using relevant technologies. In 2040, it is estimated that 40% of the journeys will be made by autonomous vehicles (Bagloee et al., 2016). In order to encourage public transportation and increase mobility, fixed payment options should be applied in various free travel and unlimited use of public transportation for certain periods.

All these practices should be implemented to ensure effective management of limited resources and to develop sustainable transportation policies by increasing mobility.

3.3. Mobility as a Service (MaAS) and Related Integrated Management

Mobility as a Service (MaAS) is the most significant concept considered as a prominent transportation service component for leading cities to invest in R&D in digital transformation. MaAS has only been conceptually identified in the literature, and is suggested to be developed as final products or services. Mobility as a Service is a service concept to reach the intelligent transportation system (Yang and Lee, 2019). Its purpose is to connect and bring together service modes on a single platform to improve the travel experience.

Passengers using MaAS only pay for the entire trip at once, and the maximum benefit is achieved for the optimization of the service. The resources of the service operators are integrated into the entire transportation system.

In order for cities to overcome the transport challenges they face, they need to increase mobility. As the mobility increases, the efficiency of transportation services increases. Advanced modern technology tools (artificial intelligence, internet of things, emotional intelligence technology, big data and analytics, etc.) are widely used in business models developed to increase mobility.

Sharing economy, multimodal and intermodal transportation segment are prominent issues to increase mobility. In the transition from Mobility 3.0 to Mobility 4.0, cities need to plan resources for this infrastructure return, on the other hand, it is possible to obtain the maximum economic benefit from business model opportunities.

It is recommended that cities transform their transport management infrastructure according to the MaAS system to offer an economic advantage. This transformation is enabled by integrated service management. Figure 4 provides a management transformation proposal that will ensure common planning and interactive management.

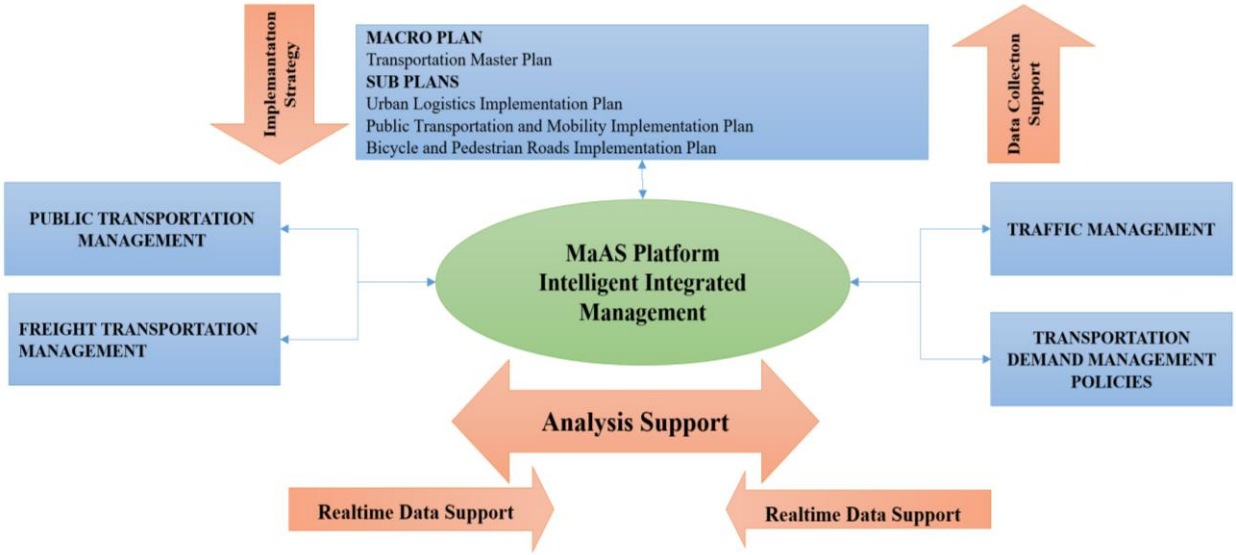


Figure 4. Proposed Frame for MaAS Platform and Integrated Management

A common MaAS platform to be established according to the proposed framework must be systems that take implementation strategies from the transportation master plan and sub-plans and provide real-time data support to these plans. It should ensure that the major components of transport management are integrated together to share data and use common infrastructure. The level of smartness required by each transport component should be determined.

The common MaAS Platform is possible with end-to-end encrypted data sharing and data security using block chain technology. It is essential to ensure the security of commercial and personal data and service integration. Intelligence level should be determined for each transportation component. Investment decisions for the management of services and the priority order of these investments should be determined with a strategically focused approach specific to cities.

3.4. Level of Smartness in Transportation Systems

In order to express the transition from "Traditional" to "Intelligent" for urban mobility, a 4-tier operational management level is defined as part of this study. Figure 5 shows a comprehensive approach, a scale, to evaluate operational management of transport transition process in which the working level of operational management is determined. This scale is developed within the scope of the study. For instance, the transition from "Digital" to "Intelligent" has been defined as a concept working in line with the transition from Mobility 3.0 to Mobility 4.0.

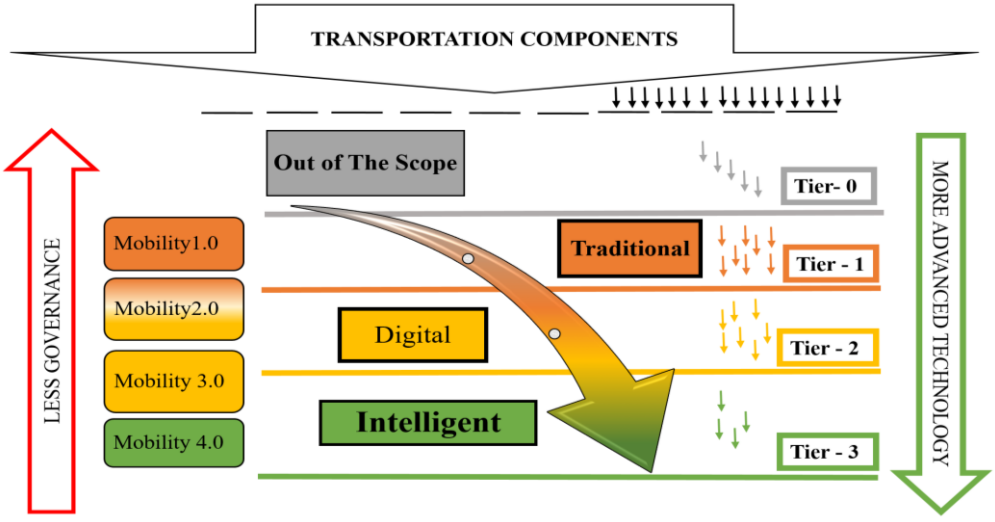


Figure 5. Services operated and managed levels of tiers

The 4-tier model developed to identify the smartness level of a transportation system is used in the assessment of the digital transformation of a city. The assessment should be made for each component of the city's integrated transport system. The smartness of a component is defined in 4 levels as seen in Figure 5 and is determined following the expert evaluation.

"Tier-0" is the lowest level and defined as "Out of the Scope", which means that there is no defined smart service in the management of the transportation component. This transportation service may not be available in the city before. For example, for a city without a rail system, rail system line optimization is not required.

The second level, "Tier-1", is defined as "Traditional" referring that there is no need for public investment in the management of the service. The current level of smartness is considered to be sufficient for the management of the service. The services identified at this smartness level usually use digital services provided by the private sector mobile services.

"Tier-2" level is defined as "Digital" indicating that there is a need to develop a web- or mobile-based digital service to execute the service. The smartness in the digital transportation system is expected to include the sample capabilities listed below.

- ✓ It is expected to have a real-time and shared traffic information system (with mobile and web support).
- ✓ Public transport should provide route inquiry and travel planning services.
- ✓ Electronic ticket application is expected to be included, where subscription initiation and contactless payments should be available.
- ✓ Services such as remote payment, debt inquiry and registration capabilities for all transportation components should be available.
- ✓ An integrated payment system in multimodal public transport should be provided.

The final level, "Tier-3", is defined and named as "Intelligent". This means that a cyber-physical system should be in place for the execution of the service. Systems empowered with artificial intelligence (AI), using big data and internet of things (IoT) technology infrastructure may be required. The integrated intelligent transportation system is expected to avail the following capabilities as listed below.

- ✓ Transport and traffic systems is monitored in real time and enhanced with artificial intelligence.
- ✓ A real-time information system to collect, process and analyse the current/ongoing transport and traffic with big data and data analytics capabilities.
- ✓ Coordination of transport and traffic-related incidents, information sharing with stakeholders and incident management services.
- ✓ Providing real-time travel advice to public transport users.
- ✓ Analyse and schedule the impact of planned construction and maintenance activities on transport and traffic.

3.5. Process Flow Chart of the Proposed Approach

The model offers components to be used as decision-support systems for decision makers and also makes suggestions for determining the level of smartness based on the decision maker's preferences and decisions. A decision maker is allowed to choose the indexes as the set of preferences among the predetermined indexes, then determines the working values. At this stage, the model is structured in accordance with the choices of the decision makers. The level of smartness of each identified transportation component is identified based on expert views and index values. Once the smartness level is identified, the targeted digitalization rates of the city under consideration are determined. Finally, the model makes a priority ranking with strategy-focused assessment. The decision maker can decide on the transformation of the transportation management components with the strategy opted among the predetermined ones. The model can be rerun to generate different configurations to diversify offered solutions. The process flow chart of the proposed approach is shown in Figure 6.

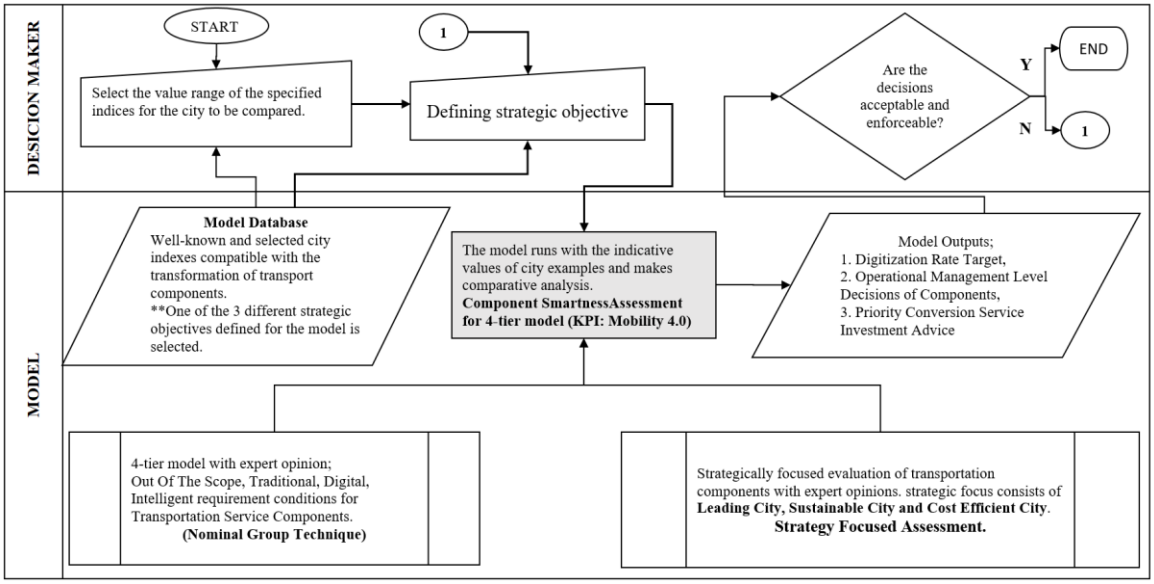


Figure 6. The process flow chart of the proposed approach

The model outputs 3 results. The first result is the proportional target suggested for the operational management smartness levels of transport services (out of the scope, traditional, digital and intelligent) for the city under-consideration. The second result is the digitalisation suggestion for the operational service level of each transport service component. The last one, it is the prioritization of transport components to be transformed by comparative strategic focus assessment. If the model outputs are suitable for the decision maker, the results are used as a digital transformation goal. Otherwise, the decision maker can re-run the system with different configuration options for a new assessment with variations and indexes.

Strategic priorities suitable for the conditions of the cities should be determined together with the model setup. Strategic focused assessment is used for prioritizing the components whose smartness level is determined. As a result of the model, investment priority, among the components evaluated at the same level of smartness, may change depending on the strategy. The decision maker decides on the digitization investment priority ranking by choosing the appropriate strategy.

The scope of this paper is emphasised on identifying the smartness level of transportation systems of smart cities using expert views. The experts, whose views were consulted in this study, have been selected among those who work in the public and private sectors and have expertise in transport management in various capacities including engineers, specialists and junior/ senior managers, who are actively on duty either in academia or in industry. It is important to note that the expert-views have been elicited from twenty-one experts whilst the development of conceptual and reference models.

In the expert opinion method, the nominal group technique was used. As a first step, the factors connected with the digitisation of transport services have been identified, indicators of smart cities in the literature that could express these factors have been selected, and then the experts were asked to suggest sufficient conditions for the smartness levels proposed within the model. After identifying component smartness level, a comparative assessment as to dedicate strategic investment priority was performed using R&D product development approach. A comparative assessment of the components for which the level of smartness was identified as a strategic priority with Mobility 4.0 components was conducted as a sample.

4. A Case Study

The proposed four-level model approach is based on the concept of Mobility 4.0 to determine the level of smartness. The level of smartness, which is considered as traditional, means that the services in the Mobility 2.0 stages are widely used for the city. The meaning of smartness level evaluated "digital" refers to Mobility 3.0 and Mobility 4.0 transition period. It should be understood that cities that determine urban transportation services at the level of smartness "intelligence" are managed with the concept of Mobility 4.0. Figure 7 illustrates how the model will measure the level of smartness of cities with different capacities.

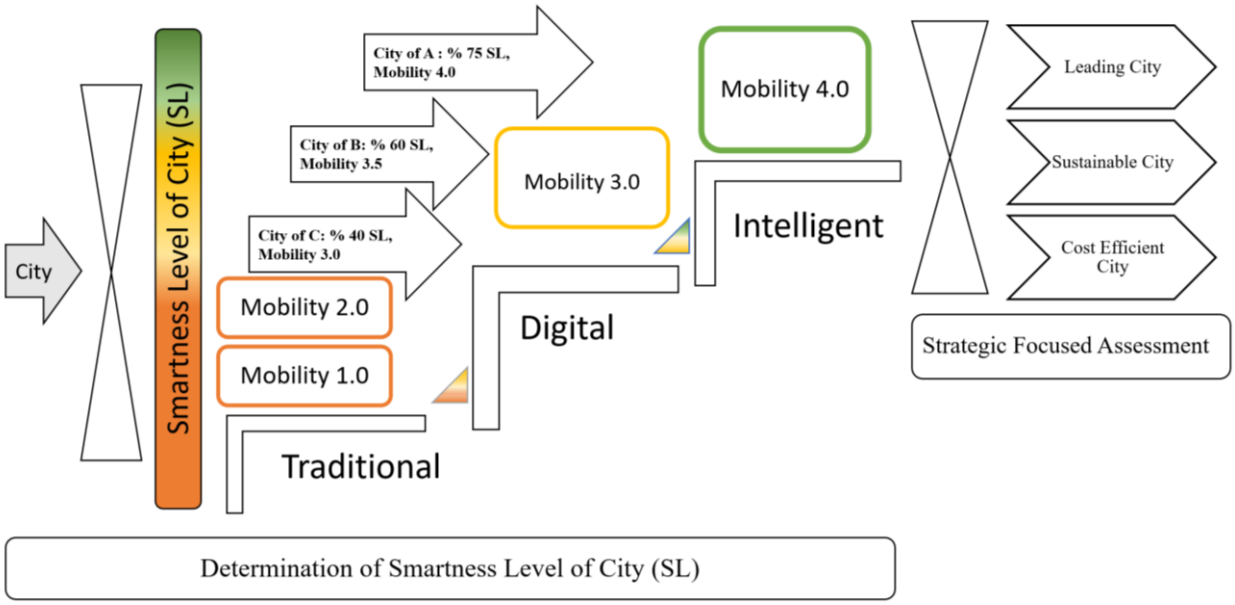


Figure 7. The level of smartness of cities with different capacities

It is assumed that the selected sample cities "A", "B" and "C" will set targets at various levels of smartness in the model assessment. These results can be achieved by assessing the smartness levels of each of the transport components of these cities. With the parameters of the index to be chosen and the opinions of experts, the level of intelligence of each transport component is determined by the proposed framework model. When some examples of transport components are assessed using the model, the level of smartness indicated in Figure 8 is suggested for selected cities.

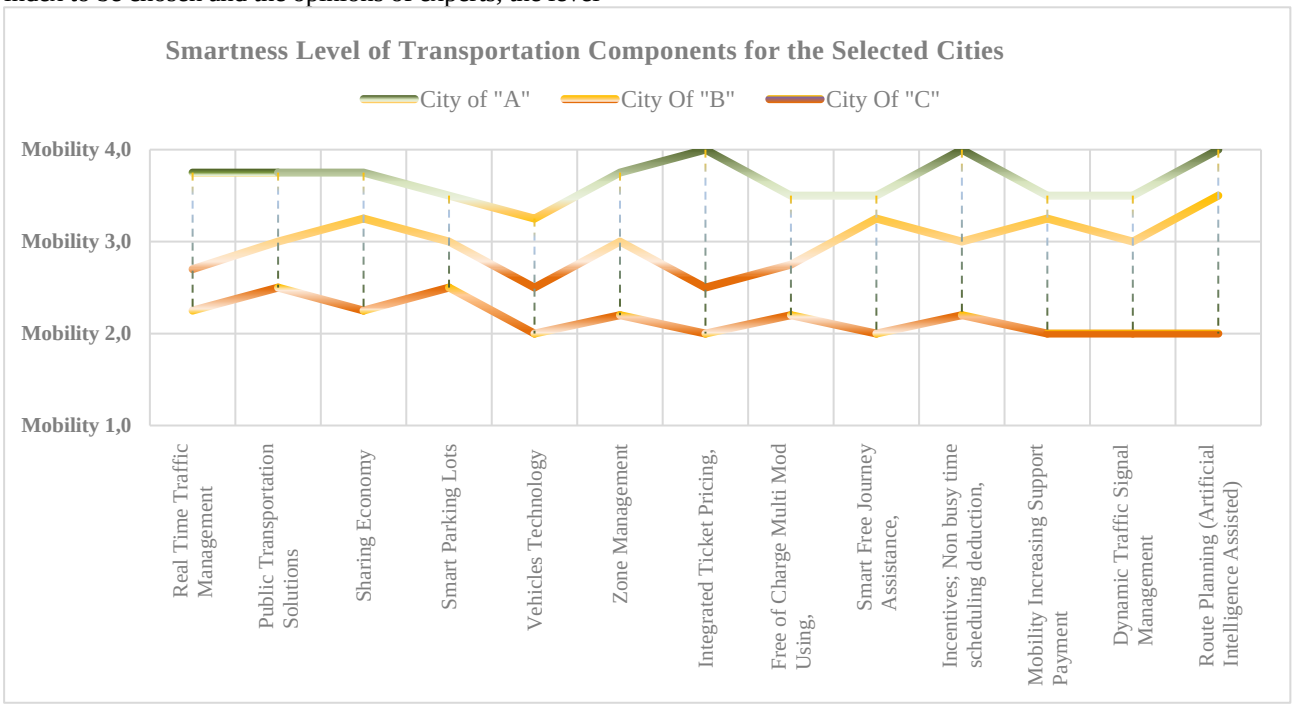


Figure 8. The level of smartness of cities with different capacities

With the conditions indicated, it is recommended by the model that City A serves at 75% smartness (SL) and Mobility 4.0 level. This means that investing in the Cyber Physical System (CPS) has to be prioritised for these transport components. However, there seems to be a need for Mobility 4.0 services in the management of certain transportation components for City B. A detailed review of these services is recommended.

Once the smartness rate and the digitalization proposals for the transportation components of cities under consideration, the transportation components to invest in are required a strategically focused assessment. A strategically focused prioritization study is carried out among the components by using multiple decision-making techniques. In this study, 3 different strategic focuses are determined for decision makers to choose one among; "Leading City", “Sustainable City” and "Cost Efficient City ". The preferred strategic focus is used to set investment priorities for digitalization.

"Leading City" imposes a strategic priority that prioritizes citizens' expectations and is likely to have a disruptive impact on existing transport operators. It is a strategy that only economically powerful cities should choose. Extra investment is required to change or remove existing licenses and authorizations. By creating research and development (R&D) funds, opportunities are provided for new innovative projects. It is recommended to select cities that lead digitalization with leading city characteristics. Examples are more common in developed economies. Cities such as London, Paris, New York, Tokyo, and Seoul are examples of this category. The characteristic feature of this strategy is the R&D investment towards digital transformation.

“Sustainable City” is a strategic approach that aims to implement digitalization with the effective management of existing resources. It focuses on changing citizen habits in solving problems in the city. Pedestrian-oriented solutions such as parking lot management, cycling, and extending pedestrian areas are prioritized. In general, cities that choose this strategy, with strong infrastructure and human resources, follow leading cities. Cities such as Amsterdam, Barcelona, Copenhagen, Milan, Rome and Glasgow can be cited as examples.

"Cost Efficient City" offers suitable strategies for cities that expect income from digital transformation. Cities that choose this strategy need to invest in income-generating and relatively low infrastructure. Long-term financing of digital transformation is recommended with authorization and licensing revenues after the first stage of transformation. However, this strategic approach is not suitable for leading cities. Digital transformation applications should be implemented as the adaptation of successful practices that have been tried and accepted in other cities. Cities such as Istanbul, Moscow and Beijing can be given as examples

In Figure 9, the assessment of R&D investments required for the development of MaAS by type of strategy is presented in terms of main transport components.

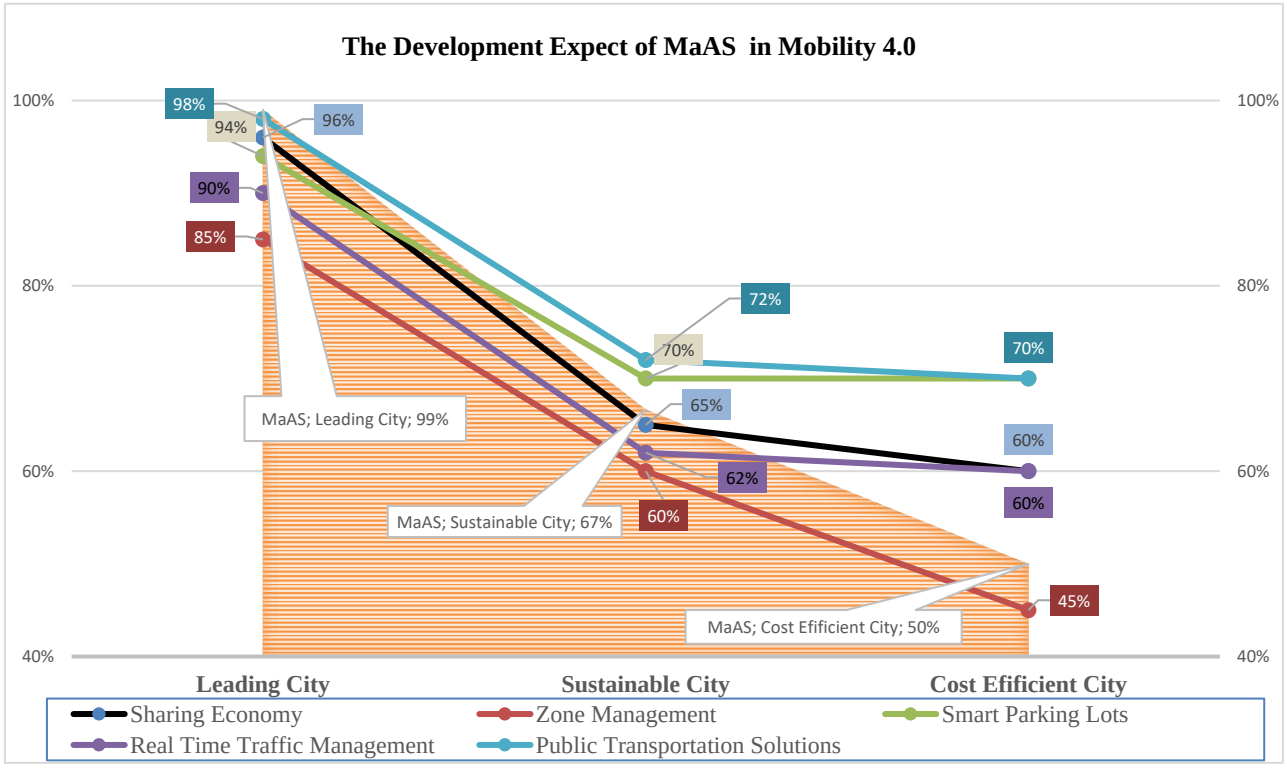


Figure 9. Strategic assessment for MaAS development with main component

It is expected that MaAS would evolve as part of the Mobility 4.0 components. However, according to expert assessments, cities should target different levels of development according to their strategic priorities for the digitization goal, each of which requires ICT and R&D investment. As seen from Figure 9, the simultaneous execution of the development of all

components can be determined as a target for a city which is a leading city.

Of course, it should be known that cities make decisions together with all stakeholders. According to experts, the reason for delaying the full commissioning of a technologically advanced service such as "zone management" is commercial and social concerns. Figure 9 shows us that investment priorities and digitization rates have changed strategically effectively when some of the Mobility 4.0 services, which have advanced technological infrastructure, are considered together with MaAS, which is still at the conceptual stage

“Intelligent" cities, such as City "A", with a goal of digitalization at the Mobility 4.0 level and targeting over 75% smartness, need to invest in digitization and R&D. These cities are considered to be leading cities. All transport components are assumed to be holistically simultaneous conversion goals. The "Leading City" strategic focus would be a suitable choice for these cities. They are expected to make the R&D investments needed for MaAS. These cities should have smartness transportation management targets in excess of 75 %.

It is recommended that "Digital" cities such as City "B" with a digitalization target above Mobility 3.5 and aiming for over 60% smartness choose the "Sustainable City" strategy. In these cities, the primary transport component transformation are Public transport management and real-time traffic management components. Once prototype applications are developed for MaAS, these cities are expected to develop various innovative services with pilot cases.

It is suitable for cities that transition from traditional systems to digital systems such as city "C", which has a digitalization target above Mobility 3.0 and aims for over 40% smartness, to carry out digitalization with the "Cost Efficient City" strategy. The transportation components planned for primary transformation in these cities are low-cost and high income-generating features. Parking areas management, real-time traffic management is the primary conversion component. MaAS is a high-potential conceptual approach to complement its development as a service inspired of cloud solutions. In the next few years, it would evolve with the R&D investments of leading cities and the pilot cases that sustainable cities will bring in diversity. Mature MaAS services will be used by cities in this category without investing in R&D.

The decision maker makes the prioritization assessment of the transportation components by choosing the appropriate one from the strategy-based evaluation made by the experts for the province it evaluates.

4. Conclusions

The assessment of digitisation is a present and new field of study in the literature. There are some studies that compare the digitization rates of cities with maturity models in smart city concepts. Existing research does not include comprehensive models to define a vision of digitization for cities. This study was developed with the idea that the city is unique and needs its own digitization objectives.

In this article, the concepts of determining city-specific digitalization targets covers Mobility 4.0 and suggest smartness levels, according to the its components and determining investment priorities with a strategic assessment for MaAS Concept Development are presented.

Literature review confirms how difficult is to define the transformation situation of smart cities with a universal system. Priorities, targets and visions identified for cities play very important role in this process. Studies conducted within the scope of digital transformation of transportation services are compatible with "Mobility 4.0" concepts in literature research. Providing services with integrated and interactive management in this concept constitutes the future perspective of transportation services. The maturity levels of services transformed by digitalization vary. A strategic approach is necessary to assess mature services, prototype services and services that are still in conceptual stage. MaAS is a major powerful conceptual component of digitalization relevant to Mobility 4.0. It is recommended that cities assess MaAS transformation in terms of transport components with a well-known strategy to seize opportunities. This study has been conducted by evaluating that known index values can be used in the digital transformation of transportation services. The potential development of the MaAS concept in the digitalization process is evaluated using the developed 4-tier smartness level approach. It also provides proposals for strategic ranking of priority investments with an R&D approach. The approach can be used as a roadmap for cities that want to transform Mobility 4.0 and want to define a MaAS strategy.

Acknowledgment

The Scientific and Technological Research Council of Turkey (TÜBİTAK) Science Fellowship and Grant Programmes Department supported this study, which produce from PhD Study of Hakan INAC (Application Number: 1059B141900373).

References

Akpınar, M., 2019. Smart City Applications in Digital Age: State-Of-Art Review and Critique. Journal of Information Systems and Management Research, 1(1), 37(42).

Albino, V., Berardi, U., Dangelico, R. M. 2015. Smart cities: Definitions, dimensions, performance, and initiatives. Journal of urban technology, 22**,** 3-21.

Angelidou, M., Psaltoglou, A., Komninos, N., Kakderi, C., Tsarchopoulos, P., Panori, A. 2018. Enhancing sustainable urban development through smart city applications. Journal of Science and Technology Policy Management.

Anttiroiko, A.V., Valkama, P., Bailey, S. J. 2014. Smart cities in the new service economy: building platforms for smart services. AI & society, 29**,** 323-334.

Bagloee, S. A., Tavana, M., Asadi, M., Oliver, T. 2016. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. Journal of modern transportation, 24**,** 284-303.

Bello, J. P., Silva, C., Nov, O., Dubois, R. L., Arora, A., Salamon, J., Mydlarz, C., Doraiswamy, H. 2019. Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution. Communications of the ACM, 62**,** 68-77.

Butler, L., Yigitcanlar, T., Paz, A. 2020. Barriers and risks of Mobility-as-a-Service (MaaS) adoption in cities: A systematic review of the literature. Cities**,** 103036.

Clayton, W., Ben-Elia, E., Parkhurst, G., Ricci, M. 2014. Where to park? A behavioural comparison of bus Park and Ride and city centre car park usage in Bath, UK. Journal of Transport Geography, 36**,** 124-133.

Colvile, R., Hutchinson, E. J., Mindell, J. & Warren, R. 2001. The transport sector as a source of air pollution. Atmospheric environment, 35**,** 1537-1565.

Cottrill, C. D. 2020. MaaS surveillance: Privacy considerations in mobility as a service. Transportation Research Part A: Policy and Practice, 131**,** 50-57.

Crainic, T. G., Ricciardi, N., Storchi, G. 2009. Models for evaluating and planning city logistics systems. Transportation science, 43**,** 432-454.

Fagnant, D. J., Kockelman, K. M. 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transportation Research Part C: Emerging Technologies, 40**,** 1-13.

Foltýnová, H. B., Attard, M., Melo, S. 2018. Topical collection on the role of planning towards sustainable urban mobility. Springer.

Giffinger, R., Fertner, C., Kramar, H., Meijers, E. 2007. City-ranking of European medium-sized cities. Cent. Reg. Sci. Vienna UT**,** 1-12.

Greenblatt, J. B., Shaheen, S. 2015. Automated Vehicles, On-Demand Mobility, and Environmental Impacts. Current Sustainable/Renewable Energy Reports, 2**,** 74-81.

Halliday, S. 2013. Underground to everywhere: London's underground railway in the life of the capital, The History Press, ISBN 978 0 7524 9551 4.

Hellinger, A., Seeger, H. 2011. Cyber-Physical Systems. Driving force for innovation in mobility, health, energy and production. Acatech Position Paper, National Academy of Science and Engineering, 1.

İnaç, H., Dönmez, C. 2018. Evaluation of Value Capture Strategies for Public Private Partnership Projects in Railway System Projects to be Built in Istanbul City, International Journal of Advances in Engineering and Pure Sciences, 30(3), 286-298. DOI: 10.7240/marufbd.396649.

Kaparias, I., Bell, M., Tomassini, M. 2011. Key performance indicators for traffic management and intelligent transport systems. ISIS, 14.

Kemeny, T., Storper, M. 2015. Is specialization good for regional economic development? Regional Studies, 49**,** 1003-1018.

Khayyat, N. T., Lee, J.D. 2015. A measure of technological capabilities for developing countries. Technological Forecasting and Social Change, 92**,** 210-223.

Klingenberg, C., Antunes, J. 2017. Industry 4.0: what makes it a revolution. EurOMA 2017**,** 1-11.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., Britter, R. 2015. The rise of low-cost sensing for managing air pollution in cities. Environment international, 75**,** 199-205.

Lee, J., Bagheri, B., Kao, H.-A. 2015. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing letters, 3**,** 18-23.

Lytras, M. D., Visvizi, A. 2018. Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research. Sustainability, 10**,** 1998.

Ma, X., Xue, H. 2020. Intelligent smart city parking facility layout optimization based on intelligent IoT analysis. Computer Communications, 153**,** 145-151.

Manville, C., Cochrane, G., Cave, J., Millard, J., Pederson, J. K., Thaarup, R. K., Liebe, A., Wissner, M., Massink, R., Kotterink, B. 2014. Mapping smart cities in the EU.

Metz, D. 2018. Tackling urban traffic congestion: The experience of London, Stockholm and Singapore. Case Studies on Transport Policy, 6**,** 494-498.

Morse, L.C., Babcock, D.L., Murthy, M. 2014. Managing engineering and technology, Pearson.

Mouchili, M. N., Aljawarneh, S., Tchouati, W. Smart city data analysis. Proceedings of the First International Conference on Data Science, E-learning and Information Systems, 2018. 1-6.

Müller, I. S., Liedtke, G. Understanding structural barriers of innovations towards sustainable transportation.

Nagy, D.K. 2016. City location and economic development. Princeton University, mimeograph.

Ngo, H. Q.T., Nguyen, T.P., Nguyen, H., 2019. H. Hardware Design for Intelligent IoT Approach to Optimize Parking Slots. 2019 International Conference on Advanced Computing and Applications (ACOMP), IEEE, 171-175.

Pringle, P., 2011. AdaptME toolkit: Adaptation monitoring and evaluation. Oxford, UK: UKCIP. accessed 29, 03-21.

Razaghi, M., Finger, M. 2018. Smart governance for smart cities. Proceedings of the IEEE, 106**,** 680-689.

Redman, L., Friman, M., Gärling, T., Hartig, T. 2013. Quality attributes of public transport that attract car users: A research review. Transport policy, 25**,** 119-127.

Sharifi, A. 2020. A typology of smart city assessment tools and indicator sets. Sustainable cities and society, 53**,** 101936.

Smith, G., Sochor, J., Karlsson, I. M. 2018. Mobility as a Service: Development scenarios and implications for public transport. Research in Transportation Economics, 69**,** 592-599.

Sochor, J., Strömberg, H., Karlsson, I. M. 2015. Implementing mobility as a service: challenges in integrating user, commercial, and societal perspectives. Transportation research record, 2536**,** 1-9.

Şentürk, İ., Coulıbaly, S., 2020. Mobile Data Collection in Smart City Applications: The Impact of Precedence-based Route Planning on Data Latency. Journal of Innovative Science and Engineering, 4 (1), 22-34. DOI: 10.38088/jise.713809.

Tay, K.C., Supangkat, S. H., CORNELIUS, G. & ARMAN, A. A. The SMART initiative and the Garuda smart city framework for the development of smart cities.

2018 International Conference on ICT for Smart Society (ICISS), 2018. IEEE, 1-10.
Wolff, H., Perry, L. 2010. Policy monitor: trends in clean air legislation in Europe: particulate matter and low emission zones. Review of Environmental Economics and Policy, **4,** 293-308.
Yaman, O., Bayğın, M., 2020. UHF-RFID, Smart Cargo Management, Real Time Tracking. Journal of Intelligent Systems: Theory and Applications, 3 (2), 38-45. DOI: 10.38016/jista.762685
Yang, W., Lee, S.P. 2019. Service Model Research of Bicycle-sharing based on Mobility-as-a-Service (MaaS). Journal of Service Research and Studies, 9**,** 19-40.

Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 4(2) (2021) 127-135
DOI: 10.38016/jista.932762

Belediye Hizmetlerin Değerlendirilmesinde Duygu Analizi Yaklaşımı: Sakarya İli Örneği

Feyza Nur Uyaroğlu Akdeniz[1]*, Halil İbrahim Cebeci[2]

[1]Sakarya Üniversitesi, Bilgisayar Mühendisliği Bölümü, Sakarya, Türkiye

[2]Sakarya Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, Sakarya, Türkiye

feyzanuruyaroglu@gmail.com, hcebeci@sakarya.edu.tr

Öz

İnternetin hızla yayılımı, sosyal medyaların da ortaya çıkışını hızlandırmıştır. Dünyanın farklı bölgelerinde bulunan insanlar, sosyal medyalar aracılığıyla saniyeler içinde etkileşim kurabilmekte ve düşüncelerini hızla yayabilmektedir. Sosyal medyalar bireysel ve örgütsel olarak etkili bir iletişim ağı sağlamaktadır. Kullanıcılar, bu platformlarda tüketici veya vatandaş olarak almış oldukları hizmetleri olumlu veya olumsuz olarak değerlendirebilmektedirler. Bunun sonucunda vatandaşlarla hizmet aldıkları kuruluşlar arasında bir etkileşim ağı oluşmuştur. Bu etkileşim ağında vatandaşların sıklıkla belediye hizmetlerini de yorumladığı gözlemlenmiştir. Halka en yakın yönetim birimi olan belediyelerin hizmetleri vatandaşlar tarafından sosyal medyalar aracılığıyla değerlendirmelere tabi tutulmaktadır. Halkla ilişkiler boyutuyla ele alındığında özellikle Twitter'da bireyler bizzat belediyeler ve belediye başkanlarıyla iletişim kurabilmektedir. Bu durum vatandaş-belediye arasında farklı bir iletişim zemini sağlamakta ve çözüm odaklı bir etkileşim ağı kazandırmaktadır. Bu çalışma, Sakarya özelini kapsayarak; Twitter kullanıcılarının Sakarya Büyükşehir, Adapazarı, Serdivan, Erenler ve Hendek belediyeleri hakkında paylaşımları incelenerek, vatandaşın belediyelere karşı tutumları analiz edilmek için hazırlanmıştır. Belediye hesaplarına atılmış tweet metinleri toplanmıştır**.** Elde edilen bu veriler öncelikle Makine Öğrenmesi yöntemleriyle önceden belirlenmiş gruplara atanmış, daha sonra Makine Öğrenmesi ve Derin Öğrenmesi yaklaşımlarından biri olan Turkish Bert yöntemi ile Duygu Analizi gerçekleştirilmiştir. Vatandaşların belediyelere karşı olan tutumu ortaya çıkmış, olumlu ve olumsuz yanıyla değerlendirmeye tabi tutulmuştur. Yapılan bu çalışmayla belediyelerin vatandaşlardan gelen olumlu veya olumsuz geri bildirimlerle, vatandaş memnuniyetinin sağlanması ve gelen taleplere daha hızlı cevap verilebilmesi sağlanmıştır. Belediyelerin vatandaşın olumsuz tutumlarını, yapacakları çözüm odaklı çalışmalarla olumlu yönde değiştirmesi bu çalışmanın ana amacını oluşturmaktadır. Twitter aracılığıyla belediyelerin vatandaşlarla ikili diyaloglarını geliştirmesi, istek, talep ve şikâyetlere çözüm üretmesi, katılımcı demokrasiyi güçlendirmesi gibi başlıklarla topluma katkı sağlanacağı düşünülmektedir.

Anahtar kelimeler: Duygu Analizi, Makine Öğrenmesi, Turkish Bert, Twitter, Yerel Yönetimler.

Sentiment Analysis Approach in the Evaluation of Municipal Services: The Case of Sakarya Province

The fast span of the internet has accelerated the emergence of social media people in all over the world to affect each other in seconds through social media and spread their thoughts quickly. Social media provide an effective communication network individually and organizationally. Users can evaluate the services they have received as consumers or citizens on these platforms as positive or negative. As a result, an interaction network has been formed between the citizens and the organizations they receive services from. Reconsidering this interaction network, citizens frequently interpret municipal services. The services of the municipal corporations, the managing agency, are evaluated by citizens through social media. Considering the public relations aspect, individuals can personally communicate with municipalities and mayors, especially on Twitter. This situation provides a different communication ground between the citizen and the municipality and provides a solution-oriented interaction network. This study covers the Sakarya special; The shares of Twitter users about the Metropolitan Municipality, Adapazarı, Serdivan, Erenler and Hendek municipalities were examined, and the attitudes of citizens towards municipalities were prepared to be analyzed. The texts of tweets sent to municipal accounts were collected. These obtained data were first assigned to predetermined groups with Machine Learning methods, and then Emotion Analysis was performed with the Turkish Bert method, one of the Machine Learning and Deep Learning approaches. Citizens' attitude towards municipalities has emerged and has been evaluated with its positive and negative aspects.With this study, municipalities were provided with positive or negative feedback from citizens, ensuring citizen satisfaction and responding to requests faster. The principal aim to

* Sorumlu yazar.
E-posta adresi: feyzanuruyaroglu@gmail.com

Alındı : 4 Mayıs 2021
Revizyon : 8 Haziran 2021
Kabul : 21 Haziran 2021

this paper municipalities change the negative attitudes of the citizens positively with the solution-oriented studies they will do. It is thought that through Twitter, municipalities will contribute to the society with topics such as improving their bilateral dialogue with citizens, finding solutions to requests, demands and complaints, and strengthening participatory democracy.

Keywords: Sentiment Analysis, Machine Learning, Turkish Bert, Twitter, Local Government.

1. Giriş (Introduction)

Her geçen gün sosyal medyaya yönelik artan ilgi, bireysel ve örgütsel davranışların etki alanı genişletmektedir. Bireysel, ticari, siyasi vb. birçok amaç için kullanılan sosyal medya platformları bireyler ve kurumlar arası etkin bir etkileşim zemini oluşturmaktadır. Mahalle muhtarından, devlet başkanlarına, ufak bir bakkaldan, dünya devi e-ticaret firmalarına varana dek her kesim bu platformları etkileşim açısından bir kazanım olarak görmektedir. Sosyal medyaya ilgi artışının sebebi, alışılmış medyadan farklı olarak kullanıcı ile karşılıklı etkileşim özelliğine sahip olmasıdır. Etkileşim özelliğinin çok katmanlı bir iletişim imkânı sağlaması, kullanıcıların iletişimlerini her iki yönde de aktif bir şekilde kullanmasına katkı sunmaktadır.

Twitter, 2006'da Amerika'da ortaya çıkmış ve kullanılmaya başlanmıştır (Işık, 2018). "Tweet" adı verilen, kullanıcıların duygularını, mesajlarını paylaşmaya yarayan sosyal medya araçlarından biridir. Son zamanlarda oldukça popülerdir ve güncel konuların takibinin en kolay olduğu sosyal kullanım aracıdır. Belediyelerin Twitter hesapları; kararların, yapılacak uygulamaların, aylık etkinliklerin ve yapılan hizmetlerin duyurularının paylaşıldığı ortam olmanın ötesinde; yapılacak yeni hizmetlerde ve alınacak kararlarda vatandaşın bireysel olarak dâhil edilmesini ve anlık haber paylaşımını sağlayan bir imkân da sunmaktadır.

Vatandaşların günlük yaşamında veya diğer özel işlerinde yerel yönetimler içerisinde en fazla etkileşime geçtiği birimler bağlı bulundukları belediyelerdir. Belediyeler vatandaşla hızlı, doğrudan ve etkin bir iletişime geçemedikleri için eleştiri altında kalabilmektedir. Fakat sosyal medyanın hayatımıza girmesiyle ve hızlıca kullanıcı sayısı artmasıyla belediyeler de bu artışa kayıtsız kalmayıp aktif bir şekilde sosyal medya kullanımına başlamışıyla birlikte, vatandaş-belediye ilişkilerinde olumlu sonuçların alındığı belirgin bir dönüşüm yaşanmaya başlanılmıştır.

Teknoloji geliştikçe artan sosyal medya kullanımı ile yazılı metinler, ses kayıtları, video ve fotoğraflar şeklinde büyük miktarda veri üretilmektedir. Bu büyük miktardaki verilerden saklı kalan yararlı bilgilerin çıkarılmasını sağlayan işlem Veri Madenciliğidir. Veri madenciliği, büyük ve karmaşık verilerin nizamını, verinin analizi ve yazılım yöntemleri kullanılması ile ilgilidir. Metin madenciliği, dijital ortamlarda işlenmemiş halde bulunan doğal dil metinlerini temel veri olarak işleyen veri madenciliği yöntemidir. Yapılandırılmış veri tabanları yerine, doğal dil metinlerinden yararlı bilgiyi meydana çıkarması sebebiyle veri madenciliğinden farklılık gösterir (Hearst, 1999). Doğal dil işleme metotlarından biri olarak isimlendirilen metin madenciliği bilgisayarların insan dilini yorumlamasına yardımcı olmaktadır (Çınar, 2019). Metin veri setinin analiz edilip değerlendirilme aşamasında, sözcük frekans dağılımı, duygu analizi (sentiment analysis), sözcük bulutu (wordcloud) gibi farklı metin madenciliği yöntemleri uygulanır (Çınar, 2019).

Bu çalışmayı ortaya çıkaran ana etken, duygu analizi ve sosyal medya analizi yöntemlerinin, çok geniş kapsamlı ve saklı kalmış yorumları, yarı otomatik değerlendirme kabiliyetleri ele alındığında doğrudan bireylerin Twitter'da Belediye Hizmetleri hakkındaki paylaşımlarını ortaya çıkarma ve değerlendirme imkânı sunmasıdır.

2. Duygu Analizi (Sentiment Analysis)

Duygu analizi terimi ilk defa 2003'de Tetsuya ve Jeonghee tarafından bahsedilmiştir. Çalışmada pozitif veya negatif olarak ikiye ayırıp gruplandırmak yerine, belirli konular için pozitif veya negatif ile ilişkili duyguları belirten bir duygu analizi çalışması yapılmıştır (Nasukawa ve Yi, 2003).

Duygu analizi, bir varlığa karşı metinlerde ifade edilen fikirlerin, duyguların ve tutumun hesaplamalı bir çalışmasıdır (Ravi ve Ravi, 2015). Diğer bir tanım ile Duygu analizi; çeşitli konularla ilgili yapılan yorumları fikirleri, duyguları ve davranışları saptamak, ortaya çıkarmak ve sınıflandırmaktan ibarettir (Korkusuz, 2019). Doğal dil işlemenin bir alt dalı olan duygu analizi, 21 yıldır araştırmacılar için etkin bir çalışma sahası haline gelmiştir (Tuzcu, 2020).

Duygu analizinin ana amacı yazılı metinlerde yer alan işlenmemiş bilgilere erişip bu bilgileri işleyerek metinlerde geçen duyguları meydana çıkarmaktır. Duygu analizi yapılırken farklı sınıflandırma yöntemleri kullanılabilmektedir. Duygu analizi sınıflandırma yöntemleri makine öğrenimi ve sözcük tabanlı yöntem olmak üzere ikiye ayrılmıştır (Onan, 2017).

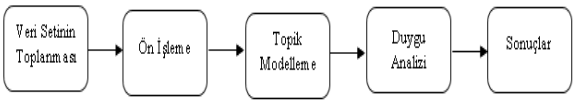
Sözcük tabanlı duygu analizinde temel mantık, duyguyu anlatan anlamsal bir sözlük veri tabanı kullanılarak duygu analizi yapılmasıdır (Kızılkaya, 2018).

Geleneksel makine öğrenimi algoritmalarını yürüten, makine öğrenimi yaklaşımı dil ile ilgili konuları kullanır. Bu yaklaşımda istatiksel ve matematiksel metotlar edinilerek, var olan eğitim verilerinden anlamlı veri çıkarımı yapılır daha sonra ise yeni oluşturulan verilerin çıktıları üzerine öngörüler yapabilmek için çıkarılan bu anlamlı veriler kullanılır. K En Yakın Komşu Algoritması, Destek Vektör Makineleri, Karar Ağaçları, Naive Bayes sınıflandırıcı, Lojistik Regresyon ve Yapay Sinir Ağları gibi birçok makine öğrenmesi yöntemleri geliştirilmiştir (Albayrak,

2018). Bu algoritmaların bazıları tahmin, kümeleme ve sınıflandırma yapabilme yeteneğine sahiptir (Atalay ve Çelik, 2017). Makine öğrenmesinin en başlıca amacı doğru tahminler yapabilmektir (Çağlayan, 2018).

3. Uygulama (Application)

Yapılan duygu analizi modeli; veri setinin toplanması, ön işleme, topik modelleme ve derin öğrenmeli duygu analizi uygulama adımlarından oluşmaktadır.



Şekil 1. Uygulama yöntemi akış şeması (Application method flow chart)

3.1. Veri Setinin Toplanması (Collection of Data Set)

Yapılan çalışmada; Sakarya'da en çok takipçisi olan 5 belediye; Sakarya Büyükşehir, Adapazarı, Serdivan, Erenler ve Hendek Belediyesi için Twitter platformundan Türkçe olarak belirli bir sorgu kelimesini içeren tweetler toplanmıştır. Bu işlemler için Python programlama dili snscrape kütüphanesi kullanılmıştır. Twitter'dan her belediye için atılmış yaklaşık 10.000 Türkçe tweet üzerinde işlem yapılmıştır.

Python, 1991'de Guido Van Rossum tarafından bulundu. Python, mevcut bir programlama dilidir ve çözüm oluşturmak için kolay bir yol sağlayan bilgisayar programıdır (Sarlan, Nadam ve Basri, 2014). Python programlama dili, veri bilimi ve makine öğrenimi gibi uygulamalar için kullanılabilir. Python çok ünlü açık kaynak kodlu, ücretsiz, yüksek seviyeli ve dinamik bir programlama dilidir (Küçük ve Arıcı, 2018; Bhavsar ve Manglani, 2019).

3.2. Ön İşleme (Pre-Processing)

Toplanılan tweet metninden yer alan bazı karakterler analiz aşamasında sonuca bir anlam katmadığı için tweet metnine ön işlem yapılması gereklidir. Tweet metninde bazı karakter veya kelimeler gereksiz kalmaktadır (Akgül, Ertono ve Diri, 2016). Tweetler '#' ile başlayan hashtaglerden, '@' ile başlayan kullanıcı isimlerinden, rt'lerden, web adreslerinden (http), sayısal ifadelerden, noktalama işaretlerinden ve en az sayıda geçen kelimelerden arındırılıp, temizlenmiştir. Metindeki büyük harfler küçük harflere dönüştürülmüştür. Belirlenen stopwords, yani etkisiz kelimeler, tweet metninden silinmiştir. Örneğin: Acaba, ama, belki, çok, daha, eğer, gibi, için, nasıl, ne vb. Çalışmada bu adımlar için NLTK (Natural Language Toolkit) kütüphanesi kullanılmıştır. NLTK, Python'da insan dili verisiyle programlar düzenlemek için geliştirilmiştir (Tuzcu, 2020).

Her kelimeyi köklerine ayırmayı sağlayan Turkish Stemmer modülü kullanılarak metinlerdeki kelimeler köklerine ayrılmıştır. Aynı kelime kökleri birbirinden değişik ekler aldığında farklı kelimeler gibi algılanmaması için, örneğin; paylasimlar-paylasim, sakaryanın–sakarya, faturaların-fatura gibi sadece kök kelimeler analize dâhil edilmiştir (Tunçeli, 2019).

Belirtilen ön işleme adımları ile lüzumsuz görülen kelime ve karakterler tweet metninden çıkarılmış ve analizlerin daha kolay gerçekleştirilmesi sağlanmıştır.

Textblob Kütüphanesi kullanılarak tweet metinleri tek tek kelimelere ayrılmıştır. Textblob, metin analizleri yapmak için kullanılan Python kitaplığıdır. Konuşma parçası etiketleme, isim tümcesi çıkarma, duygu analizi, çeviri ve sınıflandırma yapma gibi doğal dil işleme (NLP) görevlerini gerçekleştirmek için basit bir API sağlar (Nausheen ve Begum, 2018).

Ön işleme yapılmış ve tek tek ayrılan her kelime için bir frekans oluşturulmuştur. En sık kullanılan kelimelerin frekansı en fazladır. Örneğin Sakarya Büyükşehir Belediyesi için frekansı en fazla olan kelimeler "teşekkür 354, sakarya 332, ilgili 330'dur. Frekansı en fazla olan kelimeler belirlenip sık geçen kelimeler olarak listelenmiştir. Sık geçen kelimeler bir araya getirilerek kelime bulutu oluşturulup görselleştirilmiştir. Kelime Bulutu incelenerek, olumlu veya olumsuz duyguları ifade eden sıfatlar, isimler, fiiller belirlenip, fikir sahibi olunulabilir (Büyükeke, Sökmen ve Gencer, 2020).

Şekil 2'de Sakarya Büyükşehir Belediyesi hakkında atılan tweetler için kelime bulutu oluşturulmuştur. Kelimeler arasında vatandaşların, belediyeye olan talepleri "istiyoruz" gibi temel ifadelerle taleplerde bulunduğu gözlemlenmiştir.

Şekil 2. Sakarya Büyükşehir Belediyesi tweetlerine ait kelime bulutu (Word cloud of Sakarya Metropolitan Municipality tweets)

Şekil 3'te Adapazarı Belediyesi hakkında atılan tweetler için kelime bulutu oluşturulmuştur. Atılan tweetlerde, vatandaşların kurumsal hesaba yönelik **"**başkanım, lütfen" gibi ifadelerle talep ve istekte, "teşekkürler" ifadesiyle ise olumlu geri bildirimler yönelttiği gözlemlenmektedir.

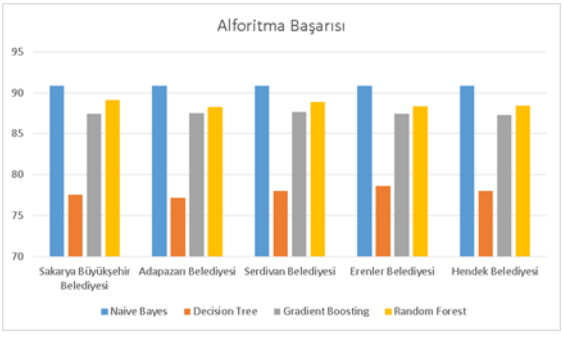
Şekil 3. Adapazarı Belediyesi tweetlerine ait kelime bulutu (Word cloud of Adapazarı Municipality tweets)

3.3. Topik Modelleme (Topical Modeling)

Tweet metinlerini sınıflandırmak ve makine öğrenmesi modellerin kullanılması için örnek veri seti oluşturulmuştur. Fen İşleri, Kültür İşleri, Sosyal Hizmetler, Spor Hizmetleri ve Çevre Hizmetleri alanında 5 farklı kategori için elle veri etiketlenmesi yapılmıştır. Her kategori için 700 tweet metni etiketlenmiştir (Sarıman ve Mutaf, 2020). Toplanan örnek veri setindeki her bir kelime sayısallaştırılarak aktarılmıştır.

Model öncesi, K katlamalı Çapraz Doğrulama yöntemi kullanılarak öğrenme süreci oluşturulmuştur. Veri seti, k=10 değeri alarak 10 eşit bölüme ayrılmıştır. Seçilen her bölüm hem eğitim hem de test için kullanılmıştır. Böylelikle öğrenmedeki hatalar en aza indirilmeye çalışılmıştır.

Örnek veri setinin %80'i modeli eğitmek için, %20'si ise modeli test etmek için kullanılmıştır. Kurulan model makine öğrenmesi yöntemlerinden faydalanarak eğitilmiştir. Uygulamada 4 farklı makine öğrenmesi algoritma kullanılmıştır. Bunlar; Naive Bayes, Decision Tree, Gradient Boosting ve Random Forest'tır. Şekil 4'te her algoritma için başarı ölçütleri verilmiştir. En başarılı algoritma Naive Bayes olduğu için makine öğrenmesi yöntemi olarak Naive Bayes seçilmiştir. Twitter'dan toplanan örnek veri seti, hazırlanan bu 5 kategoriden en uygun kategoriye Naive Bayes kullanılarak atanmıştır.



Şekil 4. Makine öğrenmesi algoritmaların başarısı (The success of machine learning algorithms)

Bayes teoreminden esinlenerek oluşturulan Naive Bayes yöntemi, makine öğrenme metotları içinde kolaylıkla gerçekleştirilen, metin sınıflandırmada kullanılan ve en başlıca kabul gören öğrenme algoritmalarından biridir (Tuzcu, 2020). Bayes yöntemi olasılık kurallarına göre nitelendirilmiş birçok ölçümle, öğretilmiş veri seti kullanılarak yeni oluşturulan veri setinin hangi sınıfta olduğunu saptamak için çalışmaktadır. Naive Bayes yöntemini kullanılan diğer metotlardan farklı kılan en mühim özelliği, olasılık hesaplamalarıyla durumu değerlendirmesidir. Bu yöntemi diğer metotlardan ayıran en önemli başarısı ise öğretilmiş verilerin çok olmasından kaynaklanmaktadır (Karamanlı, 2019; Özkul, 2019).

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} \qquad (1)$$

Bu formülüzasyonda (Karamanlı, 2019); P(X)= X durumunun gerçekleşme olasılığı, P(Y)= Y durumunun gerçekleşme olasılığı olarak tanımlanır. P(X/Y)= Y öznitelik vektörünün oluşması durumunda X vektörünün belirlenen bir sınıfa ait olma olasılığı olarak tanımlanır. Yürütülen bu formülde payda da bulunan P(X)*P(Y|X) teriminin sonucunun sıfır olması durumu gerçekleşebilir. Bu durumda, sıfır yerine eşik değer denilen bir değer düzenlenir ve işlem uygulanır (Özkul, 2019).

Naive Bayes: İstatistiksel bir sınıflandırma algoritmasıdır, Bayes teoreminin bağımsızlık önermesiyle basitleştirilmiş halidir (Meral ve Diri, 2014). Naive Bayes, X nesnesi verilen her bir Y sınıfının son olasılığını P(Y|X) tahmin etmek için örnek verilerdeki bilgileri uygulamak için bir modelleme sağlar. Tahminlere sahip olunduğunda, bunları uygun sınıflandırmalar yapabilmek için kullanılabilir (Webb, 2016).

3.4. Derin Öğrenmeli Duygu Analizi (Deep Learning Emotion Analysis)

Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Turkish Bert her bir kelimeyi ayrı ayrı değerlendirmek yerine, diğer kelimelerle olan ilişkileriyle birlikte değerlendirir. Turkish Bert bir transfer Derin Öğrenme Algoritma modelidir. Transfer öğrenme daha önceden öğretilmiş verinin aynı sorunları çözmek için başka bir modele aktarılarak kullanılan Yapay Sinir Ağları algoritmasıdır (Sevli ve Kemaloğlu, 2021). Türkçe için doğal dil işleme tekniği olan Turkish Bert modeli ile metni oluşturan kelimeleri tek tek değerlendirmek yerine, kelimelerin önünde ve arkasındaki kelimelerle veya aynı anlamlara gelen kelimelerle incelenmiştir (Urhan, 2020). Kelimeler pozitif veya negatif duygularına göre skor değeri almıştır. Her bir tweet, hesaplanan duygu skorlarına göre olumlu veya olumsuz olarak değerlendirilmiştir. Bu değerler sonucuna göre tweetler, olumlu veya olumsuz olarak gruplandırılmıştır.

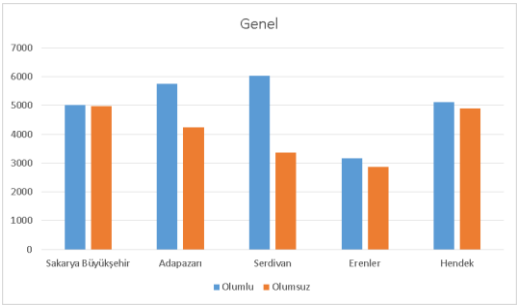
Seaborn kütüphanesi, yoğun olarak istatistiksel modellerin görselleştirilmesinde kullanılmaktadır. Verileri özetler, genel dağılımları gösterir ve görselleştirmeler sunmaktadır (Uslu ve Akyol, 2021)

Tweet metinleri; Naive Bayes sınıflandırma algoritması ile Seaborn kütüphanesi kullanılarak veri görselleştirmesiyle daha anlaşılır hale getirilmiştir (Can ve Alatas, 2017; Ayan, Kuyumcu ve Ciylan, 2019; Üçükkartal, 2020).

4. Deneysel Sonuçlar (Experimental Results)

Twitter platformundan Sakarya Büyükşehir, Adapazarı, Serdivan, Erenler ve Hendek Belediyeleri hakkında atılan tweet metinleri toplanıp, veri seti oluşturulmuştur. Veri setine ön işleme aşamaları uygulanmıştır. Ön işlemden geçmiş veriler, makine öğrenmesi ve derin öğrenmesi yaklaşımlarından biri olan Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Tweetler, olumlu ve olumsuz durumlarına göre duygu skorları hesaplanarak değerlendirilip, görsellenmiştir.

Şekil 5'te her belediye için Twitter'dan toplanmış tweetlerin olumlu ve olumsuz değerleri görülmektedir. Sakarya Büyükşehir Belediyesi için olumlu ve olumsuz tweet sayılarının birbirine yakın değerler olduğu; Adapazarı, Serdivan, Erenler ve Hendek Belediyeleri için olumlu tweet sayısı olumsuz tweet sayısından daha fazla olduğu görülmektedir.



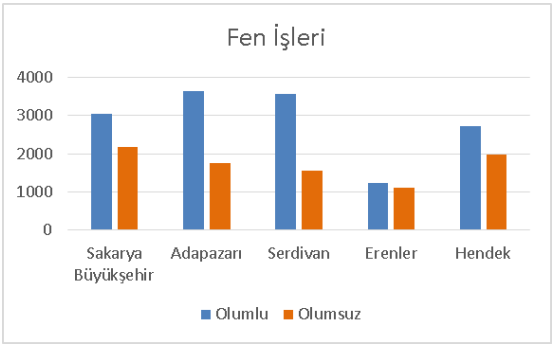
Şekil 5. Belediyeler hakkında atılan toplam tweetlerin sınıflandırılması (Classification of total tweets posted about municipalities)

Belediyeler hakkında atılan tweetler makine öğrenmesi yöntemlerinden Naive Bayes sınıflandırıcısı kullanılarak; Fen İşleri, Kültür İşleri, Sosyal Hizmetler, Spor Hizmetleri ve Çevre Hizmetleri kategorilerine ayrılmıştır.

4.1. Fen İşleri (Technical Works)

Belediyelerin Fen İşleri Birimleri; yol bakım ve onarım işlemleri başta olmak üzere; altyapı ve çeşitli üstyapı hizmetlerinin yürütülmesini sağlar. Fen İşlerinin iş bölümlerini oluşturan, "altyapı, çukur, yol, asfalt, kaldırım ve kanalizasyon" terimlerini içeren tweet metinleri etiketlenerek örnek veri seti hazırlanmıştır. Belediyeler hakkında toplanan tweetler, makine öğrenmesi yöntemlerinden Naive Bayes sınıflandırma algoritması kullanılarak her belediye için ayrı olarak Fen İşleri kategorisine atanmıştır. Fen İşleri kategorisine atanan tweetler Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Şekil 6'da her belediyenin Fen İşleri alanında atılan tweetlerin olumlu ve olumsuz değerleri görülmektedir.

Her belediye için olumlu tweet sayısının daha fazla olduğu görülmektedir. Kullanıcılar tarafından atılan; olumlu olarak sınıflandırılan tweetler belediyelerin hizmetleri için teşekkür mahiyeti taşıdığı, olumsuz olarak sınıflandırılan tweetlerin ise şikâyet ve istek bildirimi olduğu görülmüştür.

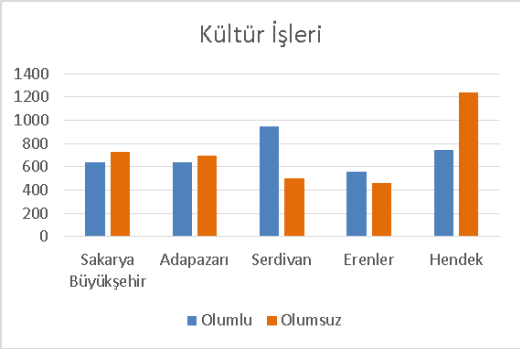


Şekil 6. Fen İşleri hakkında atılan tweetlerin sınıflandırılması (Classification of tweets posted about Technical Works)

4.2. Kültür İşleri (Cultural Services)

Belediyelerin Kültür İşleri Birimleri; kültür hizmetlerinin yaygınlaştırılması, geliştirilmesi ve il içinde yapılacak her türlü kutlamalarla ilgili çalışmaları yürütülmesini sağlamaktadır. Kültür İşlerinin iş bölümlerini oluşturan "tiyatro, bilet, seminer, sergi, konferans ve konser" terimlerini içeren tweetler etiketlenerek örnek veri seti hazırlanmıştır. Belediyeler hakkında toplanan tweetler, makine öğrenmesi yöntemlerinden Naive Bayes sınıflandırma algoritması kullanılarak her belediye için ayrı olarak Kültür İşleri kategorisine atanmıştır. Kültür İşleri kategorisine atanan tweetler Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Şekil 7'de her belediyenin Kültür İşleri alanında atılan tweetlerin sonuçları olumlu ve olumsuz değerleri görülmektedir.

Serdivan Belediyesi için olumlu tweet sayısı olumsuz tweet sayısından daha fazla olduğu; Sakarya Büyükşehir, Adapazarı ve Erenler Belediyesi için olumlu ve olumsuz tweet sayılarının birbirine yakın değerler olduğu; Hendek Belediyesi için ise olumsuz tweet sayısı daha fazla olduğu görülmektedir. Kullanıcılar tarafından atılan; olumlu olarak sınıflandırılan tweetler belediyelerin etkinlikleri için teşekkür mahiyeti taşıdığı, olumsuz olarak sınıflandırılan tweetlerin ise şikâyet ve istek bildirimi olduğu görülmüştür.

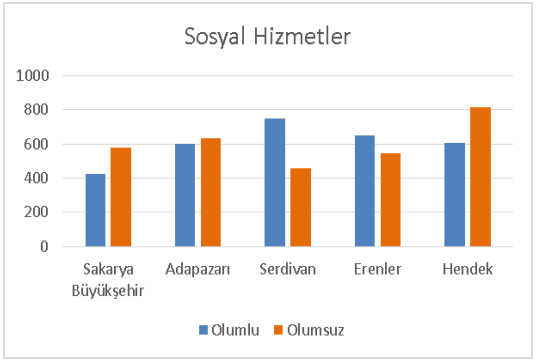


Şekil 7. Kültür İşleri hakkında atılan tweetlerin sınıflandırılması (Classification of tweets posted about Cultural services)

4.3. Sosyal Hizmetler (Social Services)

Belediyelerin Sosyal Hizmetler Birimleri; ailelere, kadınlara, yaşlılara ve çocuklara yönelik eğitici kurs ve danışmanlık hizmetleri verilmesi, engelli vatandaşlar için kolaylaştırıcı çalışmaların yürütülmesini sağlamaktadır. Sosyal Hizmetler iş bölümlerini oluşturan "kadın, çocuk, aile, yaşlı, engelli, eğitim kursları ve yardım hizmetlerini" terimlerini içeren tweetler toplanarak örnek veri seti hazırlanmıştır. Belediyeler hakkında toplanan tweetler, makine öğrenmesi yöntemlerinden Naive Bayes sınıflandırma algoritması kullanılarak her belediye için ayrı olarak Sosyal Hizmetler kategorisine atanmıştır. Sosyal Hizmetler kategorisine atanan tweetler Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Şekil 8'de her belediyenin Sosyal Hizmetler alanında atılan tweetlerin sonuçları olumlu ve olumsuz değerleri görülmektedir.

Serdivan ve Erenler Belediyesi için olumlu tweet sayısı olumsuz tweet sayısından daha fazla olduğu; Adapazarı Belediyesi için olumlu ve olumsuz tweet sayılarının birbirine yakın değerler olduğu; Sakarya Büyükşehir ve Hendek Belediyesi için ise olumsuz tweet sayısı daha fazla olduğu görülmektedir. Kullanıcılar tarafından atılan; olumlu olarak sınıflandırılan tweetler belediyelerin yardımları için teşekkür mahiyeti taşıdığı, olumsuz olarak sınıflandırılan tweetlerin ise şikâyet, istek ve eksik görülen konuların bildirimi olduğu görülmüştür.

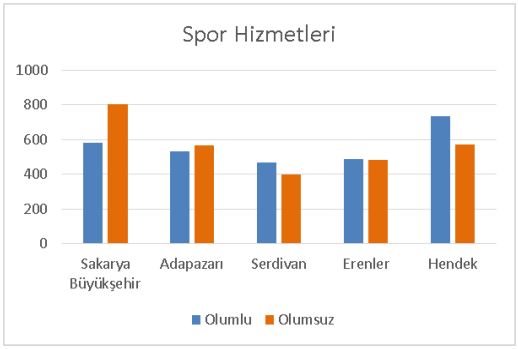


Şekil 8. Sosyal Hizmetler hakkında atılan tweetlerin sınıflandırılması (Classification of tweets posted about Social Services)

4.4. Spor Hizmetleri (Sport Services)

Belediyelerin Spor Hizmetler Birimleri; Spor faaliyetlerinin gelişmesini, yaygınlaşmasını ve katılımı sağlayacak çalışmalar yürütülmesini sağlamaktadır. Spor Hizmetler iş bölümlerini oluşturan "spor, futbol, boks, tenis, karate, olimpiyat, tesis, turnuva ve saha" terimlerini içeren tweetler etiketlenip örnek veri seti hazırlanmıştır. Belediyeler hakkında toplanan tweetler, makine öğrenmesi yöntemlerinden Naive Bayes sınıflandırma algoritması kullanılarak her belediye için ayrı olarak Spor Hizmetleri kategorisine atanmıştır. Spor Hizmetleri kategorisine atanan tweetler Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Şekil 9'da her belediyenin Spor Hizmetleri alanında atılan tweetlerin sonuçları olumlu ve olumsuz değerleri görülmektedir.

Sakarya Büyükşehir Belediyesi için olumsuz tweet sayısı olumlu tweet sayısından daha fazla olduğu; Adapazarı ve Erenler Belediyesi için olumlu ve olumsuz tweet sayılarının birbirine yakın değerler olduğu; Serdivan ve Hendek Belediyesi için ise olumlu tweet sayısının daha fazla olduğu görülmektedir. Kullanıcılar tarafından atılan; olumlu olarak sınıflandırılan tweetler belediyelerin hizmetleri için teşekkür mahiyeti taşıdığı, olumsuz olarak sınıflandırılan tweetlerin ise istek ve eksik görülen konuların bildirimi olduğu görülmüştür.

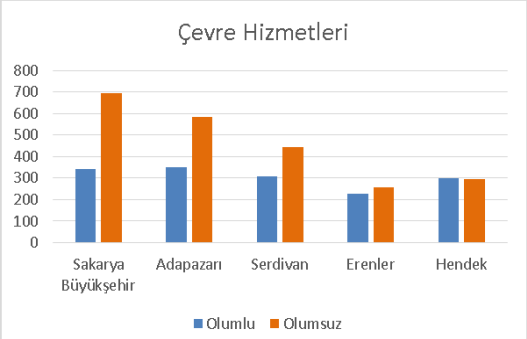


Şekil 9. Spor Hizmetleri hakkında atılan tweetlerin sınıflandırılması (Classification of tweets posted about Sports Services)

4.5. Çevre Hizmetleri (Environmental Services)

Belediyelerin Çevre Hizmetleri Birimleri; çevrenin korunmasına, temizlenmesine, geliştirilmesine dair projeler tasarlayıp, çalışmalar yürütülmesini sağlamaktadır. Çevre Hizmetleri iş bölümlerini oluşturan "çöp, temizlik, kirli, çevre, park ve bahçe" terimlerini içeren tweetler etiketlenerek örnek veri seti hazırlanmıştır. Belediyeler hakkında toplanan tweetler, makine öğrenmesi yöntemlerinden Naive Bayes sınıflandırma algoritması kullanılarak her belediye için ayrı olarak Çevre Hizmetleri kategorisine atanmıştır. Çevre Hizmetleri kategorisine atanan tweetler Turkish Bert yöntemi ile duygu analizi gerçekleştirilmiştir. Şekil 10'da her belediyenin Çevre Hizmetler alanında atılan tweetlerin sonuçları olumlu ve olumsuz değerleri görülmektedir.

Sakarya Büyükşehir, Serdivan ve Adapazarı Belediyesi için olumsuz tweet sayısı olumlu tweet sayısından daha fazla olduğu; Erenler ve Hendek Belediyesi için olumlu ve olumsuz tweet sayılarının birbirine yakın değerler olduğu görülmektedir. Kullanıcılar tarafından atılan; olumlu olarak sınıflandırılan tweetler belediyelerin hizmetleri için teşekkür mahiyeti taşıdığı, olumsuz olarak sınıflandırılan tweetlerin ise şikâyet ve eksik görülen konuların bildirimi olduğu görülmüştür.



Şekil 10. Çevre Hizmetleri hakkında atılan tweetlerin sınıflandırılması (Classification of tweets posted about Environmental Services)

5. Tartışma (Discussion)

Sakarya ilinde bulunan 5 belediye için; Fen İşleri alanında; olumlu bir algı olduğu görülmektedir. Kültür İşleri alanında; Serdivan Belediyesi için olumlu; Sakarya Büyükşehir, Adapazarı ve Erenler Belediyesi için olumlu ve olumsuz algının birbirine yakın olduğu, Hendek Belediyesi için ise olumsuz bir algı olduğu görülmektedir. Sosyal Hizmetler alanında; Serdivan ve Erenler Belediyesi için olumlu bir algı olduğu; Adapazarı Belediyesi için olumlu ve olumsuz algının birbirine yakın olduğu, Sakarya Büyükşehir ve Hendek Belediyesi için ise olumsuz bir algı olduğu görülmektedir. Spor Hizmetleri alanında atılan tweetler de; Sakarya Büyükşehir Belediyesi için olumsuz bir algı; Adapazarı ve Erenler Belediyesi için olumlu ve olumsuz algının birbirine yakın olduğu; Serdivan ve Hendek Belediyesi için ise olumlu bir algı olduğu görülmektedir. Çevre Hizmetleri alanında atılan tweetlerde; Sakarya Büyükşehir, Adapazarı ve Serdivan Belediyesi için olumsuz bir algı olduğu; Erenler ve Hendek Belediyesi için olumlu ve olumsuz algının birbirine yakın olduğu görülmektedir. Çıkan sonuçlar göz önünde bulundurulduğunda, belediyelerin kimi hizmetlerinin Twitter'daki algısı olumsuz seyrederken, kimi hizmetlerinin de olumlu yönde seyrettiği ortaya çıkmıştır.

Twitter üzerindeki duygu analizi bazlı yapılan çalışmalardan biri olan "Visualization of Local Municipal Satisfaction by Twitter Data Analysis" adlı makalede duygu analizi yöntemi ile Kanada Montréal'deki ilçelere ait tweetler toplanarak vatandaş memnuniyetini belirlemek için duygu analizi gerçekleştirilmiştir. Çalışma sonucunda Verdun Beldesi sakinlerinin belediye hizmetlerinden en yüksek düzeyde memnuniyet duydukları ortaya çıkmış; Anjou sakinlerinin ise bu hizmetlerden daha az memnuniyet duyduğu görülmüştür (Zarei, Nik-Bakht ve Hammad, 2019).

Twitter üzerindeki diğer bir çalışma ise "New York City Street Cleanliness: Applying Text Mining Techniques to Social Media Information" adlı makalede duygu analizi yöntemi ile New York şehrinde atılan tweetler toplanarak vatandaş memnuniyetini belirlemek için duygu analizi gerçekleştirilmiştir. Çalışma sonucunda çoğunlukla NYC sokaklarının temizliği hakkındaki görüşünün olumsuz olduğunu görülmüştür (Duan, Codesso, Alzamil, ve Vasarhelyi, 2020). Metin madenciliğindeki gerçekleşecek yeni gelişmeler sayesinde, Türkçe metinler için duygu analizi yaklaşımı çok daha verimli sonuçlar verecektir.

6. Sonuçlar (Conclusions)

Bu çalışmada Twitter üzerinden ilgili belediye hizmetlerine dair atılan tweetlere duygu analizi çalışması yapılmıştır. Belediyelerin hizmet alanlarındaki çalışmalarının Twitter'daki algısı gözlemlenmiştir.

Sakarya'da en çok takipçisi olan Sakarya Büyükşehir, Adapazarı, Serdivan, Erenler ve Hendek Belediyeleri için atılan tweetler örneklem olarak alınmıştır. Her Belediye için 31 Mart 2019 tarihinden sonra paylaşılmış 10.000 Türkçe tweet veri seti snscrape kütüphanesi kullanılarak oluşturulmuştur. Python programının Türkçe diline tam uyum sağlayamaması, verilerin görselleştirilmesinde ve duygu analizinde bazı sorunlarla karşılaşılmasına sebep olmuştur. Kelimelerde bulunan yapım eklerinin çalışmadan çıkarılması, sadece kelime köklerinin çalışmaya alınması, Türkçe dilinin sondan eklemeli bir dil olmasından dolayı kelime kökünden başka anlama gelecek yeni türetilen kelimelerin çalışmada yer almaması analizi etkilemiştir.

Vatandaşların Twitter'dan istek, şikâyet ve taleplerini hızlı ve şeffaf bir etkileşim süreciyle belediyelerle paylaşması, bu karşılıklı etkileşim sonucunda daha hızlı ve nitelikli bir çözüm ağı meydana getirecektir. Bu etkileşim sonucunda belediyeler de sorunların zamanında çözülmesini, kaynakları daha iyi dağıtılmasını, operasyonların daha etkin bir şekilde yönetilmesini ve kamu politikasını iyileştirilmesini sağlayacaktır. Çalışma boyunca ortaya çıkan veriler ve beraberinde gerçekleştirilen analizler göstermiştir ki Twitter, katılımcı demokrasinin güçlenmesine katkı sağlayacaktır.

Gelecek çalışmalarda veri setinin genişletilmesi veya önemli tarih aralıklarında atılan tweetlerin analizi gibi daha kapsamlı veri kümeleri ile farklı analiz sonuçları elde edilebilecektir.

Kaynaklar (References)

Akgül, E.S., Ertono, C. ve Diri, B., 2016. Twitter Verileri ile Duygu Analizi. Pamukkale Universitesi Muhendislik Bilim Dergisi, 22(2), 106-110.

Albayrak, A., 2018. 'Duygu Analizinde Farklı Vektör Temsil Yöntemleri ve Sınıflayıcıların Karşılaştırılması', Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi, Sivas,Türkiye.

Atalay, M. ve Çelik, E., 2017. Büyük Veri Analizinde Yapay Zeka ve Makine Öğrenmesi Uygulamaları. Sosyal Bilimler Enstitüsü Dergisi, 9(22), 155-172.

Ayan, B., Kuyumcu, B., Ciylan, B., 2019. Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti. Gazi Üniversitesi Fen Bilimleri Dergisi GU J Sci, Part C,. 7(2), 495-502.

Bhavsar, H., ve Manglani, R., 2019. Sentiment Analysis of Twitter Data using Python. International Research Journal of Engineering and Technology (IRJET), 6 (3), 510-511.

Büyükeke, A., Sökmen, A.ve Gencer, C., 2020. Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma. Journal of Tourism and Gastronomy Studies, 8(1), 322-335.

Can, U., Alatas, B., 2017. Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi. Int. J. Pure Appl. Sci. 3(1), 75-111.

Çağlayan Akay, E., 2018. Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. Social Sciences Research Journal, 7(2), 41-53.

Çınar, A., 2019. Sınıflandırma Algoritmaları ile Bir Metin Madenciliği Uygulaması, Pdf, Marmara Üniversitesi, Türkiye.

Duan, H.K., Codesso, M.M., Alzamil, Z. ve Vasarhelyi, M. A., 2020. New York City Street Cleanliness: Applying Text Mining Techniques to Social Media Information. 1 September, Available at SSRN: https://ssrn.com/abstract=3706506

Işık, O., 2018. İST 491 İleri İstatistik Projeleri, Pdf, Haccetepe Üniversitesi, Türkiye.

Karamanlı, Ersin 2019. 'Makine Öğrenmesi Algoritmaları Kullanarak Metin Madenciliği ve Duygu Analizi ile Müşteri Deneyiminin Geliştirilmesi', Yüksek Lisans Tezi, İstanbul Üniversitesi, İstanbul,Türkiye.

Kızılkaya, Y.M., 2018. 'Duygu Analizi ve Sosyal Medya Alanında Uygulama', Doktora Tezi, Uludağ Üniversitesi, Bursa, Türkiye.

Korkusuz, Rıza 2019. 'Futbola İlişkin Twitter Paylaşımlarının Duygu Analizi', Yüksek Lisans Tezi, Trakya Üniversitesi, Edirne,Türkiye.

Küçük, D. ve Arıcı, N., 2018. Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine Bir Literatür Çalışması. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 2(2), 76-86.

Meral, M. ve Diri, B., 2014. 'Twitter Üzerinde Duygu Analizi', 2014 IEEE 22nd Signal Processing and Communications Applications Conference (SIU 2014), Nisan 23 – 25, Trabzon, Türkiye.

Nasukawa, T.ve Yi, J., 2003. 'Sentiment Analysis: Capturing Favorability Using Natural Language Processing', Proceedings of the 2nd international conference on Knowledge capture, Ekim, pp.70-77.

Nausheen, F. ve Begum, H.S., 2018. 'Sentiment Analysis to Predict Election Results Using Python', Proceedings of the Second International Conference on Inventive Systems and Control (ICISC), Coimbatore, India.

Onan, A., 2017. Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi. Yönetim Bilişim Sistemleri Dergisi, 3(2), 1-14.

Özkul, Mübin 2019. 'Metin Madenciliği Yöntemi ile Akıllı Ulaşım Sistemleri Algısının Duygu Analizi ile Ölçülmesi', Yüksek Lisans Tezi, Bandırma Onyedi Eylül Üniversitesi, Bandırma,Türkiye.

Ravi, K.ve Ravi, V., 2015. 'A survey on opinion mining and sentiment analysis: tasks, approaches and applications', Article in Knowledge-Based Systems, pp.14-46.

Sarıman, G., Mutaf, E., 2020. COVID-19 Sürecinde Twitter Mesajlarının Duygu Analizi Int. J. Pure Appl. Sci. 7(10).

Sarlan, A., Nadam, C. ve Basri, S., 2014. 'Twitter Sentiment Analysis'. International Conference on Information Technology and Multimedia (ICIMU), Perak, Malaysia.

Sevli, O. ve Kemaloğlu, N., 2021. Olağandışı Olaylar Hakkındaki Tweet'lerin Gerçek ve Gerçek Dışı Olarak Google BERT Modeli ile Sınıflandırılması. Veri Bilim Dergisi, 4(1), 31-37.

Tunçeli, O., 2019. Turkish Stemmer for Python, 18 Haziran https://github.com/otuncelli/turkish-stemmer-python.

Tuzcu, S., 2020. Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. ESTUDAM Bilişim Dergisi, 1(2), 1-5.

Urhan, M 2020. BERT Algoritması Nedir? Örneklerle BERT Algoritması, media release, 13 Aralık 2020, https://zeo.org/tr/blog/bert-algoritmasi-nedir.

Uslu, O., Akyol, S., 2021. Türkçe Haber Metinlerinin Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması. ESTUDAM Bilişim Dergisi, 2(1), 15-20.

Üçükkartal, H.K., 2020. Twitter'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması. ESTUDAM Bilişim Dergisi, 1(2), 10-13.

Webb, I.G., 2016. Naïve Bayes, Encyclopedia of Machine Learning and Data Mining, pp.1-2.

Zarei, F., Nik-Bakht, M., Hammad, A., 2019. Visualization of Local Municipal Satisfaction by Twitter Data Analysis. CSCE Annual Conference, June 12 – 15, Laval, Kanada.

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 136-141
DOI: 10.38016/jista.946954

Convolutional Neural Network Approach to Predict Tumor Samples Using Gene Expression Data

Büşra Nur Darendeli [1], Alper Yılmaz [2*]

[1,2] Yıldız Technical University, Faculty of Chemical and Metallurgical Engineering, Department of Bioengineering, İstanbul/Turkey

bndarendeli@gmail.com, alyilmaz@yildiz.edu.tr

Abstract

Cancer is threatening millions of people each year and its early diagnosis is still a challenging task. Early diagnosis is one of the major ways to tackle the disease and lower the mortality rate. Advancements in deep learning approaches and the availability of biological data offer applications that can facilitate the diagnosis and characterization of cancer. Here, we aimed to provide a new perspective of cancer diagnosis using a deep learning approach on gene expression data.
In this study, RNA-Seq data of approximately 30 different types of cancer patients the Cancer Genome Atlas (TCGA) study, and normal tissue RNA-Seq data from GTEx were used. The input data for the training was transformed to RGB format and the training was carried out with a Convolutional Neural Network (CNN). The trained algorithm is able to predict cancer with 97% accuracy, using gene expression data. In conclusion, our study shows that the deep learning approach and biological data have a huge potential in the diagnosis and identification of tumor samples.

Keywords: cancer, CNN, gene expression, RNA-Seq, TCGA

Gen İfade Verilerinde Konvolusyonel Sinir Ağı Kullanılarak Tümör Örneklerinin Tahmini

Öz

Kanser her yıl milyonlarca insanı tehdit eden, erken teşhisi hala mümkün olmayan yaygın bir hastalıktır. Erken teşhis, kanserle baş etmenin ve ölüm oranını düşürmenin en önemli yollarından biridir. Derin öğrenme yaklaşımlarındaki gelişmeler ve biyolojik verilerdeki artış, kanserin teşhisini ve karakterizasyonunu kolaylaştırabilecek uygulamalar sunmaktadır. Bu çalışmada, gen ifade verilerini kullanarak derin öğrenme yaklaşımı ile kanser teşhisine yeni bir bakış açısı sağlamayı amaçladık.
30 farklı kanser çeşidine ait RNA-Seq verisi Kanser Genom Atlası (TCGA) adlı kaynaktan normal dokuların RNA-Seq verileri GTEx adlı kaynaktan temin edilip model eğitiminde kullanılmıştır. Gen ifade verileri RGB formatına dönüştürülüp Konvolusyonel Sinir Ağı (CNN) eğitimi için kullanıldı. Eğitilen model, gen ifade verilerine dayanarak kanseri %97 doğrulukla tahmin edebilmektedir. Sonuç olarak çalışmamız, derin öğrenme yaklaşımının ve biyolojik verilerin tümör örneklerinin tanısında büyük bir potansiyele sahip olduğunu göstermektedir.

Anahtar Kelimeler: CNN, Gen İfadesi, Kanser, RNA-Seq, TCGA.

1. Introduction

The deep learning approach has emerged by designing computer models that can perform the learning process as a result of interconnected layers based on the human brain, such as neurons. As a result of the development of data science and especially the rapid increase in biological data in the last decade, neural networks have begun to play important roles in the interpretation of biological data for the diagnosis and treatment of diseases (Esteva *et al.* 2019). Cancer, one

* Corresponding Author.
E-mail: alyilmaz@yildiz.edu.tr

Recieved : 02 June 2021
Revision : 04 July 2021
Accepted : 04 July 2021

of the biggest health problems in the world, is one of the diseases in which deep learning approaches were widely applied.

Since cancer is a disease with high genomic heterogeneity and phenotypic plasticity, its diagnosis and treatment are quite challenging (Persi *et al.*2020). Thanks to the developments in medical technologies, various forms of medical data are abundant for cancer patients and these data are suitable for deep learning-based approaches for diagnosis or treatment of cancer.

Image-based methods, such as radiology and histopathology, are commonly used for cancer diagnosis thus image-based deep learning approaches have thrived in supervised learning applications of cancer prediction. CT, MRI, histopathology data have been used in deep learning algorithms for the diagnosis of many types of cancers, including breast cancer (Zuluaga-Gomez *et al.* 2020; Gour *et al.* 2020; Zhang *et al.*; Hu *et al.* 2020; Bejnordi *et al.* 2018; Couture, *et al.* 2018), prostate cancer (Swiderska-Chadaj *et al.* 2020; Hartenstein *et al.* 2020; Duran-Lopez *et al.* 2020; Yoo *et al.* 2019; Nagpal *et al.* 2019; Arvaniti *et al.* 2018), lung cancer (Kanavati *et al.* 2020; Lai *et al.* 2020; Parnian *et al.* 2020), colon cancer (Jiang *et al.* 2020), head and neck cancer (Fontaine *et al.* 2020), and skin cancer (Tschandl *et al.* 2020, Esteva *et al.* 2017). These image-based studies facilitated clinical decision making especially in the detection of tumors in the early stages.

In addition to the image-based approaches, biological data such as gene expression (Dolezal *et al.* 2020) and gene mutations (Jiao *et al.* 2020) have also been used for training deep learning models to diagnose cancer. Gene expression data and deep learning approaches are integrated to tackle various challenges such as estimation of survival times of individuals with cancer (Ramirez *et al.* 2021), determination of biomarker genes (Xie *et al.* 2021), assurance of effective therapeutics for cancer treatment (Zeng *et al.* 2021), classification of cancer subtypes (Binder *et al.* 2021, Galili *et al.* 2021, Ahn *et al.* 2018). Ahn *et al.* developed a deep learning algorithm using publicly available gene expression databases to classify the samples as normal or tumor and high predictive scores were obtained. All of these studies show that by using gene expression data and deep learning approaches together, critical information will be revealed about the mechanism of cancer.

In our study, The Cancer Genome Atlas (TCGA) dataset with RNA-Seq data of approximately 30 different types of cancer patients and a dataset obtained by curation of GTEx data including RNA-Seq analysis of normal tissues was used. The input data for the training was converted to RGB format and the training was carried out with the CNN algorithm. The trained algorithm can predict cancer and normal patients with 97% accuracy, based on gene expression data. Our results suggest that gene expression data have the potential to make inferences on cancer by mapping gene expression content to RGB space.

2. Methods

2.1 Dataset Preparation

Data was downloaded from the UCSC Xena platform (UCSC Xena), which includes RNA-Seq data from various resources including, TCGA and GTEx. Label distribution of selected datasets is shown in Table 1.

Table 1. Distribution of training dataset labels.

Datasets	***Normal***	***Tumor***
TCGA	727	9750
GTEx	7429	0

Data labels (Normal, Tumor) have been extracted from phenotype information of selected samples. Gene IDs were converted from Entrez ID to ENSEMBL IDs using the BioMart online tool (BioMart).

The differentially expressed gene list (LINCS Harmonizome)(Rouillard et al. 2016) was used to select 1024 genes that show the highest up-regulation or down-regulation count throughout the whole dataset. Expression data for selected genes have been using as input for training.

2.2 Conversion of Inputs to Images

Gene expression values have converted into (R, G, B) format before the training step. RGB values are obtained by converting gene expression value into 24-bit long binary and then using the first 8 bits for R (red), second 8 bits for G (green), and third 8 bits for B (blue) (Figure 1). For each sample, a 32x32x3 3D Numpy array was prepared.

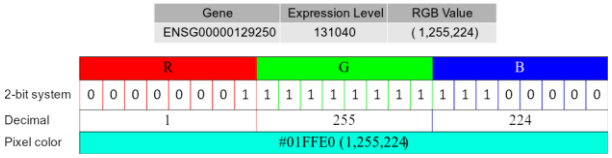


Figure 1. Conversion of gene expression value to RGB format.

2.3 CNN Architecture

The CNN architecture shown in Table 2 has been using for training. The architecture includes eight convolution layers, four dropout layers, one global average pooling layer. Each convolution layer consists of 3x3 kernels.

ReLU has been using as an activation function and, to overcome overfitting, dropout rates of 0.2 or 0.5 used. The final layer has a Sigmoid as an activation function.

Table 2. CNN Architecture

```
Model: "TCGA CNN"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 32, 32, 96)        2,688
dropout (Dropout)            (None, 32, 32, 96)        0
conv2d_1 (Conv2D)            (None, 32, 32, 96)        83,040
conv2d_2 (Conv2D)            (None, 16, 16, 96)        83,040
dropout_1 (Dropout)          (None, 16, 16, 96)        0
conv2d_3 (Conv2D)            (None, 16, 16, 192)       166,080
conv2d_4 (Conv2D)            (None, 16, 16, 192)       331,968
conv2d_5 (Conv2D)            (None, 8, 8, 192)         331,968
dropout_2 (Dropout)          (None, 8, 8, 192)         0
conv2d_6 (Conv2D)            (None, 8, 8, 192)         331,968
activation (Activation)      (None, 8, 8, 192)         0
dropout_3 (Dropout)          (None, 8, 8, 192)         0
conv2d_7 (Conv2D)            (None, 8, 8, 192)         37,056
activation_1 (Activation)    (None, 8, 8, 192)         0
dropout_4 (Dropout)          (None, 8, 8, 192)         0
conv2d_8 (Conv2D)            (None, 8, 8, 2)           386
global_average_pooling2d     (None, 2)                 0
activation_2 (Activation)    (None, 2)                 0
=================================================================
Total params: 1,368,194
Trainable params: 1,368,194
Non-trainable params: 0
_________________________________________________________________
```

3. Results

3.1. Retrieved Input Images

Since gene expression data have been converted into RGB format, visualizing the expression layout for any sample as possible. In Figure 2, sample images for Normal and Tumor samples are presented. The images do not reveal any apparent pattern for the naked eye. However, convolutional layers are able to pick regions or patterns formed by neighboring pixels so gene expression data was passed through convolution layers. Please note that gene expression data was converted into RGB format but they are not saved as images before training. The training was performed on a 32x32x3 3D multidimensional array for each sample.

3.2. CNN Training

The deep learning architecture shown in Table 2 has been using for the training of 17,906 samples having evenly distributed normal and tumor labels. Samples were split into Train: Test with 80:20 ratio. After 40 epochs the accuracy has reached 97.7%. The accuracy and loss plots of the test and training samples are shown in Figure 3.

3.3 Performance Measurement

Figure 4 shows the ROC curve of the model. The AUC value of our model was found to be 0.97. Additional performance measures were calculated from the confusion matrix generated by test sample predictions. Our model had 98% precision and 98% recall for tumor prediction (Table 3).

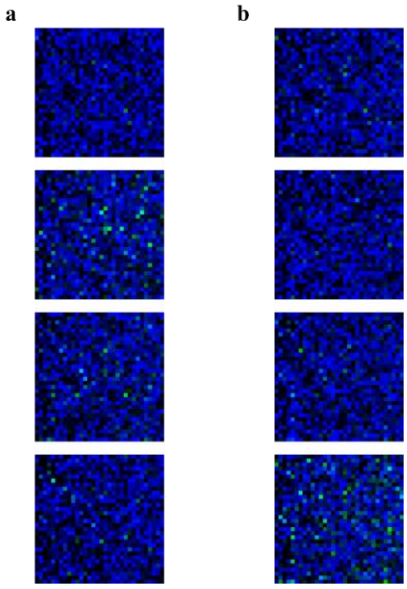


Figure 2. Visualization of gene expression data as image. 4 sample images from (a) Normal tissue data and (b) Tumor tissue data generated by converting gene expression levels of 1024 selected genes using RGB mapping.

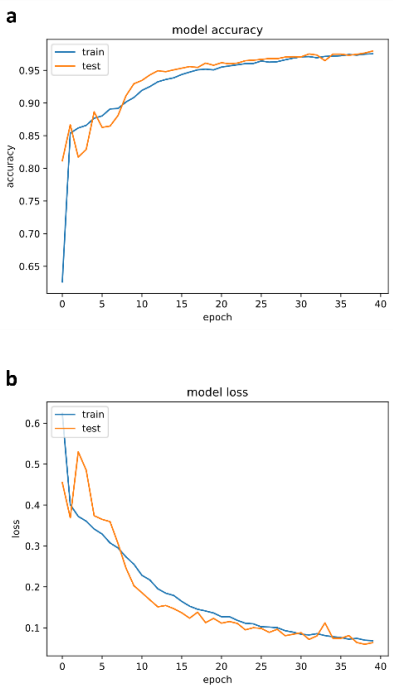


Figure 3. Model accuracy (a) and loss (b) plots.

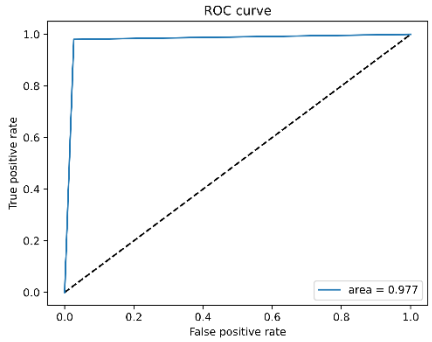


Figure 4. The ROC curve of CNN model test predictions for tumor and normal classification.

Table 3. Performance measurements based on confusion matrix

	Tumor Prediction
Accuracy	0.98
Precision	0.97
Recall	0.98
F1-Score	0.98

In literature, several different approaches use gene expression data to classify tumor and normal samples ranging from simpler machine learning approaches to complex deep learning networks. These approaches usually start with pre-processing the gene expression data with an irreversible manipulation (normalization) and even mapping data points to a different domain (PCA, t-SNE, etc.). Our method involves a minimal and reversible change to gene expression data. The RGB mapping is reversible and does not require normalization or any dimensional reduction techniques. Table 4 compares our approach with several different approaches both in pre-processing and classification steps. Although Elbashir et al. study (Normalization + CNN) has the highest accuracy, the sample used in their study is problematic so our approach has better results overall.

Please note that Elbashir et al uses a smaller and unbalanced TCGA dataset (only Breast Cancer dataset, containing 113 Normal, 1095 Tumor samples). Their accuracy starts from 91% and reaches 98.7% and, due to dominating the number of tumor samples, their model has a tendency to pick "tumor" as a label irrespective of the sample being predicted, explaining their very low sensitivity and full precision scores. In our case, our dataset is larger and balanced (8156 Normal vs. 9750 Tumor) and, our accuracy starts from 54% and then reaches 97.7%.

4. Conclusions

Due to its complex biological microenvironment, cancer has many difficulties in diagnosis and treatment. The difficulties caused by this complexity can be overcome with ever-increasing RNA-Seq data. The vast number of expression data sets combined with deep learning models have the potential to help diagnose cancer cases.

In this study, we proposed an approach to process gene expression in a reversible manner that does not require normalization. After RGB mapping of expression data, the processed data can be treated as image data and be subject to convolutional neural network learning.

Since our approach retains each pixel as an individual gene, segmentation analysis which reveals important pixels has the potential to reveal important genes for cancer development. Moreover, the strength of RGB mapping should be tested for not only tumor prediction but also tumor stage prediction.

Table 4. Comparison of our model with other studies. SVM; support vector machine, t-SNE; t-distrubuted stochastic neighbor embedding.

Expression Preprocessing	**Classification**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**References**
RGB mapping	CNN	97,73%	97,66%	97,80%	98,00%	0,975	Our method
Normalization	CNN	98,76%	91,43%	100,00%	100,00%	0,955	Elbashir et al.
Normalization	Stacked Denoising	94,78%	94,04%	97,50%	97,20%		Danaee et al.
Normalization	AlexNet	96,69%	96,89%	94,12%	99,54%	0,955	Elbashir et al
t-SNE	SVM	100,00%	100,00%	51,00%	95,96%	0,97	Elbashir et al

References

Ahmed, O., & Brifcani, A. (2019, April). Gene Expression Classification Based on Deep Learning. In 2019 4th Scientific International Conference Najaf (SICN) (pp. 145-149). IEEE.

Ahn, T., Goo, T., Lee, C. H., Kim, S., Han, K., Park, S., & Park, T., 2018. Deep learning-based identification of cancer or normal tissue using gene expression data.

In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1748-1752). IEEE.

Arvaniti, E.,Fricker, K.S.,Moret, M.,Rupp, N.,Hermanns, T.,Fankhauser, C.,Wey, N., Wild, P.J.,Rueschoff, J.H. and Claassen, M., 2018. Automated Gleason grading of prostate cancer tissue microarras via deep learning. Scientific reports,8(1), pp.1-11.

Bejnordi, B.E., Mullooly, M., Pfeiffer, R.M., Fan, S.,Vacek, P.M., Weaver, D.L., Herschorn, S., Brinton, L.A., van Ginneken, B., Karssemeijer, N. and Beck, A.H., 2018. Using deep convolutional neural networks to identify and classify tumor associated stroma in diagnostic breast biopsies. Modern Pathology, 31(10), pp.1502-1512.

Binder, A., Bockmayr, M., Hägele, M., Wienert, S., Heim, D., Hellweg, K., ... & Klauschen, F. (2021). Morphological and molecular breast cancer profiling through explainable machine learning. Nature Machine Intelligence, 1-12.

Couture, H.D., Williams, L.A., Geradts, J., Nyante, S.J., Butler, E.N., Marron, J.S., Perou, C.M., Troester, M.A. and Niethammer, M., 2018. Image analysis with deep learning to predict breast cancer grade, Erstatus, histologic subtype, and intrinsic subtype. NPJ breast cancer, 4(1), pp.1-8.

Danaee, P., Ghaeini, R., & Hendrix, D. A. (2017). A deep learning approach for cancer detection and relevant gene identification. In Pacific symposium on biocomputing 2017 (pp. 219-229).

Dolezal, J.M., Trzcinska, A., Liao, C.Y., Kochanny, S., Blair, E., Agrawal, N., Keutgen, X.M., Angelos, P., Cipriani, N.A. and Pearson, A.T., 2020. Deep learning prediction of BRAF- RAS gene expression signature identifies noninvasive follicular thyroid neoplasms with papillary-like nuclear features. Modern Pathology, pp.1-13.

Duran-Lopez, L., Dominguez-Morales, J.P., Conde-Martin, A.F., Vicente-Diaz, S. and Linares- Barranco, A., 2020. PROMETEO: A CNN-Based Computer-Aided Diagnosis System for WSI Prostate Cancer Detection. IEEE Access, 8, pp.128613-128628.

Elbashir, M. K., Ezz, M., Mohammed, M., & Saloum, S. S. (2019). Lightweight convolutional neural network for breast cancer classification using RNA-seq gene expression data. IEEE Access, 7, 185338-185348

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), pp.115-118.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. and Dean, J., 2019. Aguide to deep learning in healthcare. Nature medicine, 25(1), pp.24-29.

Fontaine, P., Acosta, O., Castelli, J., De Crevoisier, R., Müller, H. and Depeursinge, A., 2020. The importance of feature aggregation in radiomics: a head and neck cancer study. Scientific Reports, 10(1), pp.1-11.

Galili, B., Tekpli, X., Kristensen, V. N., & Yakhini, Z., 2021. Efficient gene expression signature for a breast cancer immuno-subtype. Plos one, 16(1), e0245215.

Gour, M., Jain, S. and SunilKumar, T., 2020. Residual learning based CNN for breast cancer histopathological image classification. International Journal of Imaging Systems and Technology.

Hartenstein, A., Lübbe, F., Baur, A.D., Rudolph, M.M., Furth, C., Brenner, W., Amthauer, H., Hamm, B., Makowski, M. and Penzkofer, T., 2020. Prostate Cancer Nodal Staging: Using Deep Learning to Predict 68 Ga-PSMA-Positivity from CT Imaging Alone. Scientific Reports, 10(1), pp.1-11.

Hu, Q., Whitney, H.M. and Giger, M.L., 2020. Adeep learning methodology for improved breast cancer diagnosis using multiparametric MRI. Scientific Reports, 10(1), pp.1-11.

Jiang, D., Liao, J., Duan, H., Wu, Q., Owen, G. Shu, C., Chen, L., He, Y., Wu, Z., He, D. and Zhang, W., 2020. A machine learning-based prognostic predictor for stage III colon cancer. Scientific reports, 10(1), pp.1-9.

Jiao, W., Atwal, G., Polak, P., Karlic, R., Cuppen, E., Danyi, A., De Ridder, J., van Herpen, C., Lolkema, M.P., Steeghs, N. and Getz, G., 2020. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. Nature communications, 11(1), pp.1-12.

Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F., Yamazaki, K., Takeo, S., Iizuka, O. and Tsuneki, M., 2020. Weakly-supervised learning for lung carcinoma classification using deep learning. Scientific Reports, 10(1), pp.1-11.

Li, Z., Zou, D., Tang, J., Zhang, Z., Sun, M., & Jin, H., 2019. A comparative study of deep learning-based vulnerability detection system. IEEE Access, 7, 103184-103197.

Lai, Y.H., Chen, W.N., Hsu, T.C., Lin, C., Tsao, Y. and Wu, S., 2020. overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. Scientific reports, 10(1), pp.1-11.

Marra, F., Gragnaniello, D., & Verdoliva, L., 2018. On the vulnerability of deep learning to adversarial attacks for camera model identification. Signal Processing: Image Communication, 65, 240-248.

Mencattini, A., Di Giuseppe, D., Comes, M.C., Casti, P., Corsi, F., Bertani, F.R., Ghibelli, L., Businaro, L., Di Natale, C., Parrini, M.C. and Martinelli, E., 2020. Discovering the hidden messages within cell trajectories using a deep learning approach for in vitro evaluation of cancer drug treatments. Scientific reports, 10(1), pp.1-11.

Nagpal, K., Foote, D., Liu, Y., Chen, P.H.C., Wulczyn, E., Tan, F., Olson, N., Smith, J.L. Mohtashamian, A., Wren, J.H. and Corrado, G.S., 2019. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. NPJ digital medicine, 2(1), pp.1-10.

Parnian, A., Arash, M., Tyrrell, P.N., Cheung, P., Ahmed, S., Plataniotis, K.N., Nguyen, E.T. and Anastasia, O., 2020. DRTOP: deep learning-based radiomics for the time-to-event outcome prediction in lung cancer. Scientific Reports (Nature Publisher Group), 10(1).

Persi, E., Wolf, Y.I., Horn, D., Ruppin, E., Demichelis, F., Gatenby, R.A., Gillies, R.J. and Koonin, E.V., 2020. Mutation–selection balance and compensatory mechanisms in tumour evolution. Nature Reviews Genetics, pp.1-12.

Ramirez, R., Chiu, Y. C., Zhang, S., Ramirez, J., Chen, Y., Huang, Y., & Jin, Y. F., 2021. Prediction and interpretation of cancer survival using graph convolution neural networks. Methods.

Rouillard, A. D., Gundersen, G. W., Fernandez, N. F., Wang, Z., Monteiro, C. D., McDermott, M.G., & Ma'ayan, A., 2016. The harmonizome: a collection of processed datasets

gathered to serve and mine knowledge about genes and proteins. Database, 2016.

Shon, H. S., Yi, Y., Kim, K. O., Cha, E. J., & Kim, K. A. (2019). Classification of stomach cancer gene expression data using CNN algorithm of deep learning. Journal of Biomedical and Translational Research (JBTR), 20(1), 15-20.

Sinha, S., & Saranya, S. S., 2021. One Pixel Attack for Fooling Neural Networks. Annals of the Romanian Society for Cell Biology, 8405-8412.

Su, J., Vargas, D. V., & Sakurai, K., 2019. Attacking convolutional neural network using differential evolution. IPSJ Transactions on Computer Vision and Applications, 11(1), 1-16.

Swiderska-Chadaj, Z., de Bel, T., Blanchet, L., Baidoshvili, A., Vossen, D., van der Laak, J. and Litjens, G., 2020. Impact of rescanning and normalization on convolutional neural network performance in multi-center, whole-slide classification of prostate cancer. Scientific Reports, 10(1), pp.1-14.

Tran, N.H., Qiao, R., Xin, L., Chen, X., Shan, B. and Li, M., 2020. Personalized deep learning of individual immunopeptidomes to identify neoantigens for cancer vaccines. Nature Machine Intelligence, pp.1-8.

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J. and Paoli, J., 2020. Human–computer collaboration for skin cancer recognition. Nature Medicine, 26(8), pp.1229-1234.

Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., ... & Leung, E. L. H., 2021. Early lung cancer diagnostic biomarker discovery by machine learning methods. Translational oncology, 14(1), 100907.

Yoo, S., Gujrathi, I., Haider, M.A. an Khalvati, F., 2019. prostate cancer Detection using Deep convolutional neural networks. Scientific Reports, 9.

Zeng, B., Glicksberg, B. S., Newbury, P., Chekalin, E., Xing, J., Liu, K., ... & Chen, B., 2021. OCTAD: an open workspace for virtually screening therapeutics targeting precise cancer patient groups using gene expression features. Nature Protocols, 16(2), 728-753.

Zhang, Y., Chan, S., Park, V.Y., Chang, K.T., Mehta, S., Kim, M.J., Combs, F.J., Chang, P., Chow, D., Parajuli, R. and Mehta, R.S., 2020. Automatic Detection and Segmentation of Breast Cancer on MRI Using Mask R-CNN Trained on Non–Fat-Sat Images and Tested on Fat-Sat Images. Academic Radiology.

Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S. and Zerhouni, N., 2020. A CNN-based methodology for breast cancer diagnosis using thermal images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp.1-15.

Research Article.
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 142-154
DOI: 10.38016/jista.889991

Short-Term Wind Power Prediction Approach Based On Bayesian Optimization and Ensemble Learning

Kübra Yazıcı [1*], Semra Boran [2]

[1] Turkish-Deutsche University, Faculty of Engineering, Department of Industrial Engineering, İstanbul/Turkey

[2] Sakarya University, Faculty of Engineering, Department of Industrial Engineering, Sakarya/Turkey

kubra.yazici@tau.edu.tr, boran@sakarya.edu.tr

Abstract

In wind energy studies, predicting the short-term energy generation amount for wind power plants and determining the production offer to be placed on the market play an important role. In this study an hourly short-term wind power estimation of a wind turbine located in Turkey with an installed power of 3600 kW has been made. Estimation results were evaluated on a seasonal and annual basis. New hybrid models have been developed for short-term wind power prediction, consisting of Bayesian Optimization (BO), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Decision Tree (DT), stacking, and bagging algorithms. In the proposed prediction approach, it is aimed to reduce prediction errors by combining different regression algorithms with the BO method and ensemble algorithms. Unlike other wind prediction studies, BO was used for the first time in the hyperparameter selection of the regression algorithms selected as the basic learner in the study. Bayesian optimized decision tree (BO-DT) with the lowest error values among the base learners, and Bayesian optimized gaussian process regression (BO-GPR) combined with bagging and stacking. The efficiency of ensemble learning algorithms was measured by the statistical measurement methods Normalized Absolute Mean Error (NMAE), Normalized Root of Mean Squares Error (NRMSE), and determination coefficient (R^2). According to the results, the bagging method created with the BO-DT took the annual average NRMSE, NMAE, R^2 criteria of 11.045%, 4.880%, 0.899, respectively, and the model with the best performance was selected in terms of both annual and seasonal results.

Keywords: Renewable energy, Wind power prediction, Bayesian Optimization, Ensemble Learning

Bayes Optimizasyonu Ve Topluluk Öğrenmesine Dayalı Kısa Dönem Rüzgar Gücü Tahmin Yaklaşımı

Öz

Rüzgar enerjisi çalışmalarında, rüzgâr santralleri için kısa dönem enerji üretim miktarının tahmini ve piyasaya verilecek üretim teklifinin belirlenmesi önemli bir rol oynamaktadır. Çalışmada Türkiye'de bulunan ve kurulu gücü 3600 kW olan rüzgar türbinin saatlik kısa dönem rüzgar enerjisi tahmini yapılmıştır. Tahmin sonuçları mevsimsel ve yıllık olarak değerlendirilmiştir. Kısa dönem rüzgar gücü tahmini için bayes optimizasyonu, destek vektör regresyonu, gauss süreç regresyonu, karar ağacı, stacking ve bagging algoritmalarının birleşiminden oluşan yeni hibrit modeller geliştirilmiştir. Önerilen tahmin yaklaşımında farklı regresyon algoritmaları ile bayes optimizasyon yöntemi ve topluluk algoritmaları birleştirilerek tahmin hatalarının azaltılması amaçlanmıştır. Çalışmada temel öğrenen olarak seçilen regresyon algoritmalarının hiper parametre seçiminde diğer rüzgar tahmin çalışmalarından farklı olarak ilk defa bayes optimizasyonu kullanılmıştır. Temel öğreniciler içerisinde en düşük hata değerlerine sahip bayes algoritması ile optimize edilmiş karar ağacı ve gauss süreç regresyonu, torbalama ve istifleme ile birleştirilmiştir. Topluluk öğrenmesi algoritmalarının etkinliği istatistiksel ölçüm yöntemleri olan Normalize Mutlak Ortalama Hata (NMAE), Normalize Ortalama Hata Kareleri Kökü (NRMSE) ve determinasyon katsayısı ($\mathbf{R^2}$) ile ölçülmüştür. Sonuçlara göre bayes algoritması ile optimize edilmiş karar ağacı ile oluşturulan torbalama yöntemi yıllık ortalama NRMSE, NMAE, $\mathbf{R^2}$ kriterleri sırasıyla 11.045 %, 4.880 %, 0.899 değerlerini almış ve hem yıllık hem de mevsimlik sonuçlar açısından en iyi performansa sahip model seçilmiştir.

Anahtar Kelimeler: Yenilenebilir enerji, Rüzgar gücü tahmini, Bayes Optimizasyonu, Topluluk öğrenmesi

* Corresponding Author.
E-mail: kubra.yazici@tau.edu.tr

Recieved : 02 March 2021
Revision : 22 May 2021
Accepted : 12 July 2021

1. Introduction

The need for energy is increasing day by day due to the increase in population, as well as the developments in the technology and automation sector developing in recent years. However, providing energy from fossil fuels both harms human health and causes irreparable environmental problems. Increases in population, urbanization and energy demand require sustainable energy management, technological infrastructure and innovative applications. With the developments in the last 20 years, it is estimated that the 2% growth in energy need will increase by 1.3% in the next 20 years, which requires incentives to continuously support the sector with innovative systems, products and investments (Eroğlu, 2019). In the next 10 years, especially alternative energy use will be solar, wind, etc. It will depend on the resources, bringing along important developments in the energy production industry (Öz & Alyürük, 2020). Wind energy is the fastest-growing energy type among renewable energy sources globally and the most invested energy type in the last 6 years. In 2019 approximately 15% of the electricity demand in Europe, in Turkey about 7% is obtained from wind power plants (YEKDEM, 2020). Wind power plant installed power in 2023 in Turkey total power installed power (100,000 MW) 20%. It is aimed to reach a value of 20,000 MW, corresponding to the amount of energy (Şenol & Koç, 2015). Considering the wind potential and the 2023 targets, although the current installed capacity of wind power plant is low, significant developments have been achieved in the last decade. The electricity consumption in Turkey, which reached 290.4 kilowatt-hours in 2019, is expected to reach 375.8 TWh with an annual average increase of 4.8% (ETKB, 2019). This increase in energy consumption increases foreign dependency and causes a current account deficit for countries that cannot produce the energy they consume and meet this need with imports. When examining Turkey's current account deficit, this deficit constitutes a significant portion of energy imports(Bağcı,2019).

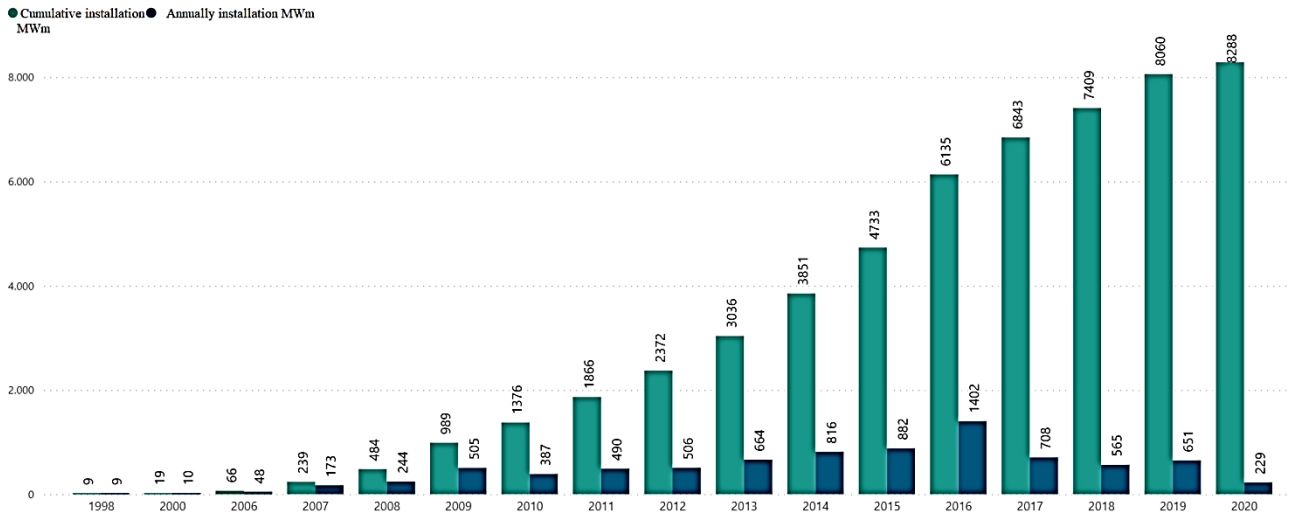


Figure 1. Installed wind power plants by years

In Figure 1, the last twelve years for Wind Power Plants in Turkey (1998-2020) is given to the development of total installed capacity (TÜREB, 2021). As shown in the graph, installed wind power has increased 10 times in the last 10 years and exceeded 8 GW. The ratio of electricity generated from wind energy to total electricity generation in Turkey in 2020 was 8.4%. A ratio of the 8.4% of electricity consumed in Turkey is produced from wind energy. It is aimed to increase this value to 20% in 2023.

Electricity generation with wind power has a highly variable profile. For this reason, energy prediction models based on wind prediction have an important place in the reliable, economical and quality operation of wind energy resources. Wind energy prediction models are used in power systems planning, reserve planning, maintenance and repair planning, and bidding in the electricity market. Thanks to prediction models, power plants can increase their revenues by reducing prediction errors in the day-ahead market, thus reducing energy imbalances in the electricity market and consequently, costs (Karık et al., 2017). Especially, short-term wind prediction play an important role in day-ahead electricity trade, planning the day-ahead electricity system, determining the required reserve amount, and making unit commitment decisions (Kerem, 2018). Thanks to short-term wind prediction, problems such as excessive production planning and allocation of excess reserves can be avoided, reducing operating costs and integrating more wind energy into the system.

Better predictive models for wind power significantly reduce the need for conventional power

plants to control energy. At this point, among prediction models, especially machine learning methods play a major role in the successful integration of wind power generation into electrical networks.

In this study, short-term wind power prediction four models based on the Bayesian optimization algorithm (BOA) and ensemble learning that can assist system operators and producers in more accurate production planning and accurate price proposal in electricity markets are proposed. Ensemble learning algorithms have been preferred in this study because it provides the opportunity to obtain higher performance models by combining more than one regression algorithm. In the parameter setting of regression algorithms, unlike other methods, the BOA, which uses the information obtained from previous experiments and works quickly, is used. Decision tree, Gaussian process regression and Support vector regression, which are frequently used in short-term wind power prediction in the literature, are used as regression algorithms. Among the optimized regression algorithms, Bayesian optimized Gaussian process regression (BO-GPR) and Bayesian optimized decision tree (BO-DT) have been chosen as the base learner for ensemble learning. In bagging and stacking methods, which are ensemble learning methods, different combinations of BO-GPR and BO-DT were tried and the results were evaluated seasonally and annually in terms of NRMSE, NMAE and R^2 performance criteria. The developed models were used for short term wind energy estimation of a wind turbine operating in Yalova. The model consisting of the combination of BO-DT and bagging algorithm, which has the best performance criteria for four seasons, has shown that it is the most effective model in short term wind power with annual average NRMSE 11.045%, NMAE 4.880% and R^2 0.899.

The remainder of this work is organized as follows. Section 2 summarizes the literature on short-term wind power prediction studies. Section 3 and Section 4 introduces the methods used in the study and proposed approach, respectively. Section 5 shows the implementation stages of the proposed models in the study and the analysis results on the determined wind turbine. And finally, Section 6 summarizes the results of the proposed short-term wind prediction models and their contribution to the literature.

2. Literature Survey

There are many studies of machine learning and ensemble learning applications in short-term wind power prediction in the literature. Lee and Baldick proposed 52 artificial neural networks (ANN) models and 5 Gaussian process regressions (GPR) models based on the GPR model and ANN for prediction at 48 hours time horizon. Artificial neural network (ANN) sub-models predicted future wind power over a 48-hour period based on past and predicted wind power data. Parallel to the ANN, GPR predict only from historical wind power data. The most appropriate prediction from more than one prediction value formed for the same hour is determined by the decision process (Lee & Baldick, 2014).

Chen et al. (2014), combined NWP (Numerical weather prediction) model and GPR models to predict short term wind power. Wind power was forecasted based on the relationship between the corrected data after the wind speed data received in the proposed model were corrected with Gaussian process. The data set on which the model has applied consists of wind speed, wind direction, temperature, pressure, humidity and wind power and is in a time period of 10 minutes. The proposed model was compared with classical wind prediction methods such as Multilayer perceptron (MLP) and ARIMA, and an improvement between 9% and 14% was achieved in error values compared to the artificial neural network.

Li et al. (2018), proposed a hybrid model based on Support vector machine (SVM) using Wavelet transform (WT) and Cuckoo search (CS) methods.

In the proposed model, Fourier, Gaussian and polynomial fitting was used to deal with the missing and erroneous data, and then the original signal was eliminated with WT. With the CS optimization, core function of the SVM and the penalty factor are optimized, and the prediction accuracy is increased. The model created has been applied to the data set consisting of hourly wind speed, wind direction, and wind power belonging to China's wind power plant. The past 60 days (1440 hours) data was used for the training, and the next 3 days (72 hours) data was used for the test. According to the results, the proposed model has fewer error values than ARIMA, Support Vector Regression (SVR) and Back Propagation Neural Network (BPNN).

Fu et al. (2019), put forward a SVM model optimized with the improved CS method on the data set consisting of hourly wind speed, wind direction, temperature and wind power. The classic CS method has been optimized to prevent the local optimum's easy capture in large data sets. The proposed model was applied in two data sets, 500 and the other 400 training data, and tested with 40 data. According to the results, the proposed model gave better results than the SVM optimized with classical CS, and it was shown that the number of training data was effective on the prediction.

Ma & Zhai (2019), established a 2-stage hybrid model consisting of WT, Feed-forward artificial neural network (FFANN) and Ant colony optimization (ACO) methods. In the first stage of the model, meteorological parameters such as wind direction, temperature, pressure and wind speed were predicted. In the second stage, wind speed predictions from the first stage were predicted with an ANN optimized with the ant colony and wind power sub-series. Wind power prediction results are obtained by applying the predicted sub-series WT. The proposed model has been more successful than ANN optimized with genetic algorithms, ANN

optimized with Particle swarm optimization in the next 24-hour wind power prediction.

Li et al. (2020), proposed a hybrid model consisting of a SVM and improved Dragonfly algorithm (IDA) has been proposed for short term wind power forecasting. The IDA is used to select the optimum parameters of the SVM. The proposed model was tested on a data set consisting of hourly wind speed, wind direction and wind power in 2017 of a wind farm located in France. In the hybrid model proposed for short-term wind forecasting, the past 6 days (144 hours) data are determined as training data, and the next 2 days (48 hours) data are determined as test data. The proposed model gave better short-term wind power forecasting results than ANN and GPR methods.

Recently, ensemble learning applications have increased in wind power forecasting by combining multiple forecasting methods, offering more generalized performance and reliability.

Heinermann & Kramer (2016), applied the k nearest neighbor (k-NN), Decision tree (DT) and SVM methods to short-term wind power. DT and SVM methods that predict with less error are combined with heterogeneous ensemble learning. With the created heterogeneous ensemble learning models, both an increase in prediction accuracy and a decrease in runtime were achieved.

Ahmad et al. (2018), applied Random Forest (RF) and Extra Trees (ET) method to investigate the effect of ensemble learning methods on improving calculation costs and prediction accuracy in hourly wind power prediction. The results were compared with the SVM. According to the prediction performance results of the methods, RF and ET, which are tree-based ensemble learning methods, gave better results than the SVM. In addition to the prediction accuracy, the ET method has less training and testing time than the other two methods.

Banik et al. (2020), used Boosting, Gradient Boosting and Extreme gradient boosting (XGBoost) methods as predictors, outliers were removed in the data set consisting of hourly data from 2014, and the relationship between wind power and meteorological parameters was determined with the Pearson correlation heat map. According to the results, a strong correlation was found between wind power and wind speed, wind direction, temperature and humidity, and other parameters were not included in the model. Then, 5-fold validation was applied to the models and predictions were made in two different time horizons, 1 hour and 1 week. According to R^2 and RMSE criteria, XGBoost ensemble gave the best results compared to other methods.

Tahir et al (2018), proposed a two-layer stacked ensemble-based model consisting of Random Forest, Support vector machine and Radial Basis Function Neural Network methods for very short-term wind power forecasting. The output values obtained with the support vector machine, random forest and radial basis function neural network models in the first layer of the model are combined with the support vector machine in the second layer and wind power are forecasted. According to the results, the proposed stacked ensemble-based model has been more successful than the classical support vector machine, linear regression, regression tree, random forest in very short-term wind power forecasting.

In this study, SVR, GPR and DT were applied in order to predict short-term wind power on a data set consisting of hourly wind speed, wind direction, temperature, pressure, relative humidity and wind power. In order to optimize the regression algorithms, unlike the optimization algorithms such as CS, ACO, PS, which are used in wind prediction studies in the literature, the BO was used. In order to reduce the prediction error of optimized regression algorithms and provide an improved prediction performance, these algorithms are combined with the ensemble learning methods, bagging and stacking. In the literature, DT and GPR algorithms, which are frequently used as single models in wind power prediction, are combined with the stacking algorithm in the study and different models are proposed.

3. Materials and Methods

3.1. Gaussian Process Regression

Gaussian process regression (GPR) is a kernel-based nonparametric probabilistic model. The Gaussian process estimates the final probability distribution based on a previous probability distribution and updates the previous probability distribution based on training data. A Gaussian process model predicts response variables with the new input vector and training data by defining the relationship between input variables and target variables (Rasmussen & Williams, 2006). Regression model function:

$$y= f(x) + \varepsilon \qquad \varepsilon \approx N(0, \sigma_n^2 \ ) \tag{1}$$

where y is the target variable, and x is the input variable. The average of ε, which is additive noise, is 0 and it is assumed to be normally distributed (Wan & Sapsis, 2017). A Gaussian process mean function is expressed by $m(x)$ and covariance function $k(x,x')$:

$$m(x) = E[f(x)] \tag{2}$$

$$k(x,x') = E[(f(x) - m(x))\ (f(x') - m(x'))] \tag{3}$$

The model is defined as follows:

$$f(x) \approx GP(m(x), k(x,x')) \tag{4}$$

Assuming that the training dataset of the Gaussian model is represented by $D = \{(x_i, y_i) = |i =$

$1,2,\ldots\ldots N|\}\, x \in R^d$, $y \epsilon R$ where x is the input y is output, R^d, d-dimensional R is the one-dimensional vector space. The Gaussian distribution with D, a linear combination of Gaussian parameters, is given in Equation 5.

$$y \approx GP(m(x), k(x, x') + \sigma_n^2 I_n) \quad (5)$$

Where I denote the identity matrix. The covariance function is the central component in the GPR model, so function selection is critical. It is the most commonly used quadratic exponential in the literature from different functions such as linear, exponential, matern, rational quadratic, squared exponential. This can be explained by the fact that the function is infinitely differentiable and therefore, uniformly distributed. Because of these properties, the squared exponential kernel function is frequently used in studies on energy estimation (Heo & Zavala, 2012). In this study, the squared exponential function used as the covariance function is shown in Equation 6.

$$k(x\, i, x\, j \mid \theta) = \sigma_f^2 exp\left[-\frac{1}{2}\ \frac{(x\, i - x\, j)^t\ (x\, i - x\, j)}{\sigma_l^2}\right] \quad (6)$$

In the equation, σ_f refers to the signal standard deviation, and σ_l refers to the feature data's length.

3.2. Support Vector Regression

The support vector machine (SVM) is one of the most frequently used methods for predicting renewable energy generation (Zendehboudi et al., 2018). The method provides flexible control over the model complexity. Thus it gives successful results in solving nonlinear problems even in a small training set (Deng et al., 2018). The SVM aims at minimizing the structure risk in contrast to conventional experimental risk reduction. It tries to minimize the upper limit of education error and generalization error consisting of the sum of a confidence interval (Dong et al., 2005).

x_i represents the vector of input parameters, and y_i represents the output value (*i* represents data point *ith* in the data set). In this case, the sample set can be defined as $\{(x_i, y_i)\}_{i=1}^n$ where N represents the total number of samples.

$$y = f(x) = W.\emptyset(x) + b \quad (7)$$

In Equation 7, *W* denotes the feature vector, *b* is the intersection vector, and ∅ (x) is the high dimensional space. In Equation 8, the regulated risk function is given to estimate the *W* and *b* coefficients (Li et al., 2009).

$$Minimise = \frac{1}{2}|W|^2 + C\frac{1}{N}\sum_{i=1}^N L_\varepsilon(y_i, f(x_i)) \quad (8)$$

$$L_\varepsilon(Y_i, f(x_i)) = \begin{cases} 0, & |y_i - f(x_i) \le \varepsilon| \\ |y_i - f(x_i) \le \varepsilon| & others \end{cases} \quad (9)$$

$|W|^2$ is the regulated term, while C is the penalty factor determining the balance between pattern smoothness and training. The second term of Equation 8 is empirical error measured by the density loss function ε given in Equation 9. To estimate *W* and *b*, the equation given above is converted into the objective function in Equation 10.

$$Minimise\ \zeta_1\zeta_1^*\ W\ b: \frac{1}{2}|W|^2 + C\ \frac{1}{N}\sum_{i=1}^N(\zeta_1 + \zeta_1^*) \quad (10)$$

$$Subject\ to: \begin{cases} y_i - W.\emptyset x_i - b \le \varepsilon + \zeta_1 \\ W.\emptyset x_i + b \le \varepsilon + \zeta_1^*,\ i = 1,2,\ldots.N \\ \zeta_1 \ge 0\ \ \zeta_1^* \ge 0 \end{cases}$$

Here ζ_1^* refer to artificial variables. Equation 10 is written as when the kernel function is $K(x_i, x_j)$:

$$Minimise\ \{a_i\}\{a_i^*\} = \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N(a_i - a_i^*).(a_j - a_j^*).K(x_i, x_j) - \varepsilon\sum_{i=1}^N(a_i - a_i^*) + \sum_{i=1}^N y_i\,(a_i - a_i^*) \quad (11)$$

$$Subject\ to: \begin{cases} \sum_{i=1}^N (a_i - a_i^*) = 0 \\ (a_i - a_i^*) \in [0, C] \end{cases}$$

In Equation 11, a_i and a_i^* represent Lagrange factors, and *i* and *j* represent different examples. In this case, Equation 7 is expressed as follows (Li et al., 2009):

$$y = f(x) = \sum_{i=1}^N (a_i - a_i^*)\, K(x_i, x_j) + b \quad (12)$$

3.3. Decision Tree

Decision tree (DT) is frequently used in classification and regression problems because of their low computational costs and easy interpretation. Since short-term wind power prediction is a regression problem, the study focuses on the principle of regression trees. The DT is usually a binary tree in each node where a decision criterion is defined, taking into account a particular feature of the test model. A tag is assigned to each leaf node in the tree, so the practitioner can easily understand the tree's decisions (Hastie et al., 2009; Bishop, 2006).

Building regression trees is the process of creating an iterative tree by selecting the most appropriate features and split points based on the minimum square error (Breiman, 1984). Thanks to the tree's ability to be adjusted according to the properties of the data set, there is no need to pre-set the function structure, and it can be worked with both discrete and continuous variables. A simple decision tree is a binary tree consisting of two leaf nodes, a root node, and a branch. The decision tree

is denoted by $h(x; a_m)$, where a_m is the characteristic variable and split point in the m-th iteration. The examples given are $R = \{x_i , y_i\}_i^N$ and x_j are continuous variables. In the *R* set, x_j takes n different values. The values are written as $\{x_j^1, x_j^2, \ldots . x_j^n\}$ on an increasing level and the set *R* is divided into R^+ and R^- according to the division point s. If the value of x_{ij} is less than the s value, it is included in the R^- set, if it is large, it is included in the R^+ set (Equation 13).

$$R^+(j,s) = \{x_{ij} | x_{ij} \geq s\}$$
$$R^-(j,s) = \{x_{ij} | x_{ij} < s\} \quad (13)$$

The predicted value in each set must be equal to the output value or all samples' mean. The predicted value for the set R_m with the number of data N_m is calculated c_m with $Equation$ 14.

$$c_m = \frac{1}{N_m} \sum_{x_j \in R_m} y_i \quad (14)$$

The set consisting of all possible values of the division points s for each value of the continuous variable x_{ij} is shown by Equation 15.

$$S_{x_j} = \left\{ \frac{x_j^i + x_j^{i+1}}{2} \; | 1 \leq i \leq N-1 \right\} \quad (15)$$

To find the appropriate feature x_{ij} and the split point s, all the split points for all properties should be examined and selected as the final split point with minimum loss. Here the loss can be calculated by Equation 16.

$$\Psi(j+s) = \sum_{x_j \in R^+} (y_i - c_s^+)^2 + \sum_{x_j \in R^-} (y_i - c_s^-)^2 \quad (16)$$

Finally, the optimal feature variable and the split point are written as in Equation 17.

$$(j^*, s^*) = argmin_{j,s} \sum_{x_j \in R^+} (y_i - c_s^+)^2 + \sum_{x_j \in R^-} (y_i - c_s^-)^2 \quad (17)$$

3.4. Bayesian Optimization Algorithm

Hyperparameter selection plays an important role in the success of machine learning algorithms. Grid search, which is frequently used in hyperparameter optimization, becomes complicated when the parameter space size is high while performing a comprehensive search in simple models (Cornejo-Bueno et al., 2018; Alade et al.,2019). Since the random search algorithm works by randomly sampling the search field, it does not use the information obtained from previous experiments. This situation creates a problem, especially in hyperparameter selection problems where the function is unknown, and the cost of running an experiment is high. BOA comes into play at this point (Wang et al, 2012). BOA estimates the posterior distribution of the objective function using the Bayes theorem and determines the hyperparameter combination of the next example according to this distribution. Unlike the random search, it uses all the information obtained from the previous experiment and tries to find the parameter to bring the result to the global maximum. To avoid native optima, exploration and exploitation need to be changed. The gain function is defined to encrypt this exchange. The gain function returns the utility estimates of the candidate points for the next step of *f(x)* and selects $x_{(t+1)}$, which produces the maximum utility.

In this study, the expected improvement (EI) is used as the gain function. The value x looking for the global minimum of the given function f (x) is obtained as in Equation 18:

$$x = argmax\, E(max\{0, f_{t+1}(x) - f(x^+)\} | D_t \quad (18)$$

Where *D* represents the number of x components. When a Gaussian process is used, it is expressed for EI as:

$$EI(x) \begin{cases} (\mu(x) - f(x^+))\varphi(z) + \sigma(x)\,\varphi(z), if\ \sigma(x) > 0 \\ 0, \qquad if \sigma(x) = 0 \end{cases} \quad (19)$$

$$z = \frac{\mu(x) - f(x)^+}{\sigma(x)} \quad (20)$$

3.5. Bagging

In the bagging approach is aimed to create independent predictors by using the samples of the training set and the average or voting of the output of the prediction methods. When more than one similar dataset is created by re-sampling with preloading, the variance of the output error decreases and overfitting is solved (Breiman, 1996). Breiman (1996) showed also that bagging method gives better results than the single tree in both classification and regression problems.

In bagging, in the first step, N new training data of n dimensions are created, in which n of the n samples are selected homogeneously by changing from the original data set. Then, each tree in the community is individually trained with relevant new training data. For example, in this study, 50 trees were used in bagging tree models. In the last step, the average of all predictions is calculated to make a final forecast. In the bagging trees model, the prediction is defined as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \quad (21)$$

Here f_i shows the generating trees, *i* shows the boot data and each tree model created is trained on the boot data.

3.6. Stacking

The stacking method, known as a batch generalization, is a method of different levels that combines different prediction models in a single model. The approach aims to minimize the errors by reducing the bias of the generalizers by introducing an optimal learning system with the concept of meta-learning (Wolpert, 1992; Van der Laan et al., 2007).

The stacking approach consists of two levels, level-0 and level-1. In Level-0, different prediction models are trained, and the output variable is predicted. The predictions obtained at Level-0 are used as input for Level-1. The model in Level-1 is called a meta-model and learns with the previous level models that give the best prediction of each model of the previous level (Shamaei & Kaedi, 2016; Serbes et al., 2015; Petropoulos et al., 2017). The number of levels is not limited to 2 in the stacking method. According to the method's working principle, the n-level model uses the prediction of the n-1 level models. In the stacking method, the change in the prediction results is due to the diversity of models at different levels. This is because models with different generalization rules tend to produce different results (Mendes-Moreira et al., 2012).

3.7. Model performance evaluation

Different performance criteria such as Normalized Absolute Mean Error (NMAE), Normalized Root of Mean Squares Error (NRMSE) and determination coefficient (R^2) have been used to test the success of models used in short-term wind prediction. The performance criteria used are given in Equation 22-24.

$$NRMSE\ (\%) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2}}{C}\ 100 \tag{22}$$

$$NMAE\ (\%) = \frac{\frac{1}{N}\sum_{i=1}^{N}|f_i - y_i|}{C}\quad 100 \tag{23}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(f_i - y_i)^2}{\sum_{i=1}^{N}(f_i - \bar{f}_i)^2} \tag{24}$$

Performance criteria are proportioned to the amount of installed power to measure the accuracy of wind energy prediction models in percentage terms and not be affected by the characteristics of the wind turbine (Chen et al., 2013). In the equations, *N* is the number of data, f_i is the actual value, y_i is the predicted value, and $\bar{f}_i$ is the average actual value. In Equation 22 and Equation 23, *C* represents the installed power. The turbine capacity used in the study is 3600 kW.

4. Proposed Approach for Short Term Wind Power Prediction

In the study, a hybrid approach consisting of BO, GPR, DT, SVR, bagging and stacking methods is proposed for short term wind prediction. The phases of the proposed approach are as follows:

Step 1. Data Collection and Data Preprocessing

- The data set consisting of SCADA data and meteorological parameters is rearranged so that the time step is 1 hour.
- The data for 15 days randomly selected from January, April, July and October are defined as test data and the remaining 6981 data as training data.
- Missing data is completed using the k-NN algorithm.
- The data is brought to the 0-1 range by applying max-min normalization to the data set. The formula used for normalization is given in Equation 25:

$$x_{nor,i} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{25}$$

Step 2. Bayesian Optimization of base learners

- BO is applied to find optimum hyperparameters of SVR, GPR and DT algorithms.
- Prediction results of GPR, DT and SVR algorithms are obtained with optimized hyperparameter values and denormalization is applied.
- The algorithm or algorithms with the least error values are selected as base learner for ensemble learning according to NRMSE, NMAE and R^2 performance criteria.

Step 3. Ensemble Models

- Ensemble learning models (bagging and stacking) are created with different combinations of algorithms selected from among the optimized algorithms. The estimation results of hourly test data determined by the created ensemble learning models are obtained.
- Denormalization is applied to the obtained prediction results.

Step 4. Comparison of developed models

- Prediction results of Bayesian optimized GPR, SVR, DT, bagging and stacking models created by working with SVR, GPR, DT algorithms trained with default parameters are obtained.
- The denormalized prediction results are compared seasonally and annually in terms of NRMSE, NMAE and R^2 criteria.

5. Application

5.1. Data Collection and Data Preprocessing

In the study, a wind turbine located in Yalova was chosen for the application. The turbine characteristics are as follows: capacity 3600 kW, rotor diameter 116.8 meters, cut-in wind speed 3 m / s, cut-out wind speed 25 m / s. The data are wind speed, wind direction, temperature, air pressure, relative humidity, and wind power for 2018. The parameters used in the studies in the literature were taken into consideration while creating the data set in this study (Chen et al., 2014; Banik et al., 2020, Ahmad et al., 2018). The wind turbine scada dataset on the Kaggle website was used in the article for short term wind power prediction (Erisen, 2019). Meteorological data were obtained from on the National Weather Service Center Environmental Forecast Climate (NOAA) website (NOAA, 2018). The data obtained have been rearranged so that the time interval is 1 hour and consists of 8399 samples. The 228 missing data points in the data set were completed with the k-NN algorithm (Gao et al., 2020).

The fact that the parameters used have different value ranges affects the prediction accuracy and causes uncertainty. In order to prevent this situation, the normalization process was applied to the data set by using Equation 25. While normalizing the wind direction data, it was first converted to radians, then sine and cosine values were calculated (Li et al., 2020).

In wind power prediction studies, it is seen that the data have a periodic pattern between seasons. For this reason, in many studies in the literature, predictions were made by creating training and test data containing data for each season, considering seasonal characteristics (Esfetang & Kazemzadeh, 2018; Zheng et al., 2017; Acikgoz et al., 2020). In this study, randomly selected 15-day data for January, April, July and October were used for the test data, and the remaining data were accepted as training data.

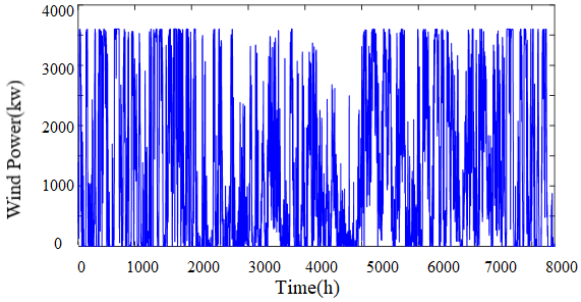


Figure 2. Wind power time series in 1-h time period in 2018

5.2. Bayesian Optimization of base learners

MATLAB 2020a program was used to create short-term wind power prediction models. GPR, DT and SVR, which are the selected prediction algorithms, were first applied with default parameters. The hyperparameter values that minimize the loss of 5- fold cross-validation with the BOA investigated. In the study, expected improvement is used as the acquisition function, which expresses how the parameter space should be investigated during BO.

In the first step of BO, initial hyperparameter values are assigned to the model. The model learns 4/5 of the data and tests it over 1/5. In each iteration, a new hyperparameter vector is created and the loss function value (log(1+loss)) is calculated. After the iterations are completed, the best hyperparameter vector that makes the loss function minimum is selected. With the selected hyperparameter vector, the final model is trained on the whole data set and the prediction results are obtained. Optimum values of sigma parameter in GPR, kernel function, box constraint, epsilon parameters in SVR and min leaf size and max num split parameters in DT were investigated.

Table 1 expresses the DT models in the BO iterations. Fig. 3 shows the minimum objective plot versus the number of function evaluations for the different decision tree models.

Table 1. The Bayesian optimization iterations for the DT model

Iter	Eval Result	Objective:log (1+loss)	Objective Runtime	BestSoFar (observed)	BestSoFar (Estim.)	MinLeafSize	MaxNumSplits
1	Best	0.011502	0.051844	0.011502	0.011502	467	15
2	Accept	0.023687	0.036668	0.011502	0.012333	1748	922
3	Best	0.0092397	0.063725	0.0092397	0.0092528	20	152
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
36	**Best**	**0.0088459**	**0.070322**	**0.0088459**	**0.0089**	**3**	**220**
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
48	Accept	0.0095316	0.056205	0.0088459	0.0088483	79	37
49	Accept	0.0093518	0.063937	0.0088459	0.0088493	28	504
50	Accept	0.010085	0.054857	0.0088459	0.0088499	2	17

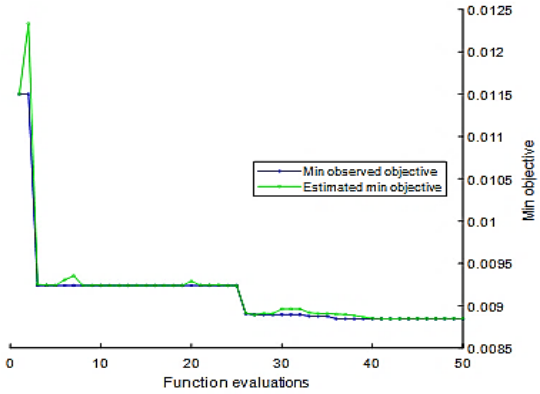


Figure 3. The minimum objective versus the number of function evaluations for the DT

According to the optimization results, the model reached the minimum objective function value by taking Min leaf size 3, and Max num split 220 in the 36-th iteration.

The Table 2, expresses the GPR models in the Bayesian optimization iterations. The Fig. 4 shows the plot of the minimum objective versus the number of function evaluations for the different GPR models.

Table 2. The Bayesian optimization iterations for the GPR Model

Iter	Eval Result	Objective: Log (1+loss)	Objective Runtime	BestSoFar (observed)	BestSoFar (Estim.)	Sigma
1	Best	0.0089561	60.936	0.0089561	0.0089561	0.030826
2	Accept	0.084427	73.087	0.0089561	0.013786	0.00016228
3	Accept	0.022513	66.347	0.0089561	0.015651	0.0017231
.	.	.	.	.	.	.
.	.	.	.	.	.	.
11	**Best**	**0.0068647**	**67.936**	**0.0068647**	**0.0069051**	**0.11505**
.	.	.	.	.	.	.
.	.	.	.	.	.	.
28	Accept	0.060073	65.317	0.0068647	0.0069193	0.011038
29	Accept	0.0069398	30.578	0.0068647	0.0069199	0.052232
30	Accept	0.022756	57.746	0.0068647	0.0069209	0.0029126

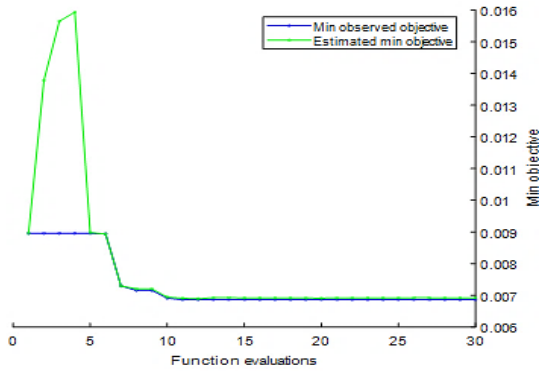


Figure 4. The minimum objective versus the number of function evaluations for the GPR

According to the results, the model reached the optimum result and minimum cross-validation loss in the 11-th iteration. Sigma took the value of 0.11505 in the optimum model.

The Table 3, represents the SVR models in the Bayesian optimization iterations. The Fig. 5 shows the plot of the minimum objective versus the number of function evaluations for the SVR model.

Table 3. The Bayesian optimization iterations for the SVR model

Iter	Eval result	Objective:log(1+loss)	Objective runtime	BestSoFar observed	BestSoFar estim.	BoxConstraint	Epsilon	KernelFunction
1	Best	0.12055	0.98747	0.12055	0.12055	0.082336	17.448	Linear
2	Best	0.018887	39.963	0.018887	0.02644	0.0039777	0.0013106	Polynomial
3	Accept	0.12055	0.24973	0.018887	0.024742	61.741	0.94565	Polynomial
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
25	**Best**	**0.0096151**	**1036.7**	**0.0096151**	**0.0096458**	**399.1**	**0.01481**	**gaussian**
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
28	Accept	0.0096409	826.89	0.0096151	0.009604	841.24	0.064184	gaussian
29	Accept	0.12055	0.67599	0.0096151	0.0096046	0.0010709	47.876	polynomial
30	Accept	0.12055	0.67059	0.0096151	0.0096049	0.0010016	45.804	gaussian

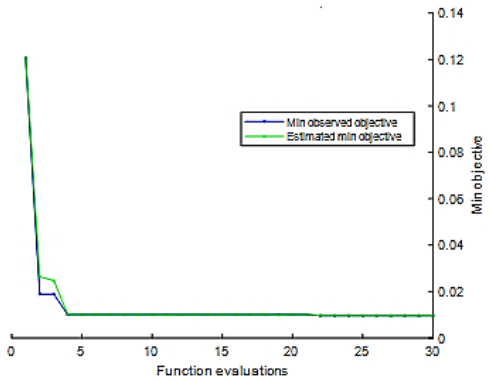


Figure 5. The minimum objective versus the number of function evaluations for the SVR

The SVR model has reached the optimum hyperparameter values in the 25-th iteration. The Kernel function of the optimum model is gaussian and Box constraint, Epsilon hyperparameters values are 399.1, 0.01481, respectively.

5.3.Selection and training of base learners

Table 4. Comparison of different regression algorithms used as base learner

Base Learner	Performance criteria		
	NRMSE (%)	NMAE (%)	R^2
BO-GPR	12.863	5.862	**0.873**
BO-SVR	13.821	5.700	0.845
BO-DT	**12.512**	**5.694**	0.866

BO has been used to improve base learner prediction performance. Table 4 shows a comparison of the BO-DT, BO-SVR and BO-GPR algorithms with BO as the base learner. According to NRMSE, NMAE, R^2 performance criteria, BO-GPR and BO-DT models have higher prediction performance than BO-SVR. Since no choice could be made between BO-GPR and BO-DT in terms of performance criteria, it was decided to use both as base learner in the study.

5.4. Ensemble Models

Table 5 lists the ensemble learning algorithms and parameters of the algorithms examined in this article. In the bagging algorithm, the BO- DT that generates fewer error predict compared to BO-GPR and BO-SVR was chosen as the basic learner. The bagging model prediction result was obtained by averaging 50 BO- DT. Bayesian optimized decision trees were trained with examples that are homogeneously selected from the original data set. In the stacking models, 3 different models were created by considering 2 levels. The first level BO-DT, BO-GPR and bagged tree model results are combined with the second level linear regression (LR) and BO-DT. It is aimed to create a model with better performance than both methods by combining BO-GRP and BO-DT models with meta learner in stacking models. For this purpose, the LR algorithm is also used as a meta learner, as it is a fast and less complex method, which increases the stacking performance, except for the BO-DT model.

Table 5. Parameters of all prediction models

Model	Parameters
BAG	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), Number of tree:50, Bagging size percent: 50%
STACK1	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), BO-GPR (Sigma: 0.11505) Meta learner: BO-DT
STACK2	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), BO-GPR (Sigma: 0.11505) Meta learner: LR (Lamda: 1.8584e-05, Learner: leastsquares)
STACK3	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), BO-GPR, Bagged tree Meta learner: LR (Lambda: 1.8584e-05, Learner: leastsquares)

5.5. Experimantel Results

Table 6 contains seasonal performance criteria and the average values of all models used in the study.

When Table 6 is examined, it is seen that BO reduces the prediction error of the basic learners for each season. In addition, it has been demonstrated that the optimized GPR, SVR, DT of bagging and stacking algorithms improve the predictive performance of bagging and stacking models. The corresponding performance improvement was measured by the R^2, NRMSE and NMAE performance criteria. The ensemble learning algorithms created in the study were more successful in short-term wind power prediction than single algorithms for all four seasons. Considering the seasonal performance, it is seen that prediction errors are higher in winter and spring seasons. The reason for this is the fluctuations and sudden changes in the parameters during these seasons. Since this difference is less in summer and autumn seasons, the performance of the models is also better in this direction. When the annual and seasonal prediction results are examined, the average R^2, NRMSE, NMAE values of the bagging model with the best performance among the suggested methods were measured as 0.899, 11.045% and 4.880%, respectively.

Table 6. Seasonal and average results of models for 1-h prediction

		Methods									
Seasons	Performance criteria	GPR	DT	SVR	BO-GPR	BO-DT	BO-SVR	BAG	STACK1	STACK2	STACK3
Winter	R^2	0.868	0.707	0.502	0.869	0.809	0.723	**0.870**	0.801	0.782	0.823
	NRMSE(%)	13.563	21.902	27.518	13.460	15.940	21.459	**13.356**	17.666	17.039	15.875
	NMAE(%)	8.499	9.773	18.913	8.299	9.065	10.820	**8.259**	9.041	9.359	8.889
Spring	R^2	0.730	0.776	0.770	0.803	0.824	0.842	**0.888**	0.882	0.874	0.887
	NRMSE(%)	19.872	17.832	17.745	19.174	15.592	13.856	**13.830**	13.953	13.954	13.894
	NMAE(%)	8.709	7.916	10.354	8.512	6.320	5.407	**5.388**	5.636	5.408	5.980
Summer	R^2	0.881	0.919	0.801	0.915	0.925	0.910	**0.926**	0.910	0.922	0.918
	NRMSE(%)	10.203	8.124	14.931	9.694	8.051	10.578	**8.001**	9.067	8.089	8.231
	NMAE(%)	3.876	3.992	8.950	3.864	3.481	3.513	**3.320**	3.802	3.321	3.689
Fall	R^2	0.907	0.876	0.861	0.908	0.908	0.905	**0.912**	0.900	0.909	0.906
	NRMSE(%)	9.202	11.890	12.373	9.124	10.464	9.390	**8.992**	9.342	9.034	10.034
	NMAE(%)	2.790	4.963	6.861	2.771	3.912	3.059	**2.650**	3.564	4.139	3.366
Average	R^2	0.846	0.819	0.733	0.873	0.866	0.845	**0.899**	0.873	0.872	0.883
	NRMSE(%)	13.210	14.937	18.141	12.863	12.512	13.821	**11.045**	12.507	12.029	12.009
	NMAE(%)	5.968	6.661	11.270	5.862	5.694	5.700	**4.880**	5.511	5.531	5.482

6. Conclusions

In the study, prediction models based on BO, machine learning and ensemble learning algorithms were created for short term wind power prediction. In the models, analyzes were carried out by considering both statistical (SCADA records) and physical (meteorological parameters) data. The missing data in the data set created in the first step of the study were completed with the K Nearest Neighbor algorithm, and test and training data sets containing data for each season were created. For this purpose, the months of January, April, July and October were chosen to represent the winter, spring, summer and autumn seasons. In the second step, hyperparameters of the DT, SVR and GPR algorithms selected as base learner for wind power prediction were optimized with the Bayesian algorithm. The Bayesian optimized decision tree (BO-DT) that reaches the least prediction error was chosen in the bagging model. Stacking models include the Bayesian optimized GPR (BO-GPR) in addition to the BO-DT. When the results of the analysis are analyzed seasonally and annually, it is seen that the ensemble learning algorithms are more successful in wind power prediction than single methods. The annual average R^2, NMAE and NRMSE values are 0.899, 4.880%, 11.045%, respectively, and the average calculation time of the bagging model performs better than other wind energy prediction models, is less than 10 seconds.

The conclusions of the study can be summarized as follows:

- BOA was used for the first time in short-term wind energy prediction and has increased the prediction performance of regression algorithms.
- In the study, GPR and DT algorithms, which gave successful results in wind power prediction studies, were combined with the stacking algorithm for the first time and a model with higher performance than both methods was provided.
- Successful results were obtained by adding BO to the bagged tree model, which is also included in wind power prediction studies in the literature.
- It has been shown that the prediction errors of the models change depending on the seasons. All models created have fewer prediction errors in the summer and autumn seasons compared to the winter and spring seasons.

In future studies, the proposed model will be developed as follows:

- Optimum hyper parameter values of bagging and stacking algorithms, which are meta-learners, will also be investigated with BOA.
- In order to improve the data quality, which directly affects the prediction performance of the model, outliers in the data set will be detected and cleaned.

References

Acikgoz, H., Yildiz, C., Sekkeli, M., (2020). An extreme learning machine based very short-term wind power forecasting method for complex terrain. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 42(22), 2715-2730, DOI:10.1080/15567036.2020.1755390.

Ahmad, M.W., Mourshed, M., Rezgui, Y., (2018). Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy, 164, 465-474.

Alade, I.O., Abd Rahman, M.A., Saleh, T.A., (2019). Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. Solar Energy, 183, 74–82.

Bağcı, E., (2019). Türkiye'de Yenilenebilir Enerji Potansiyeli, Üretimi, Tüketimi ve Cari İşlemler Dengesi İlişkisi. R&S-Research Studies Anatolia Journal, 2 (4), 101-117. DOI: 10.33723/rs.501940.

Banik, R., Das, P., Ray, S., Biswas, A., (2020). Wind power generation probabilistic modeling using ensemble learning techniques. Materials Today: Proceedings, 26, 2157–2162.

Bishop, C., (2006). Pattern Recognition and Machine Learning. Information Science and Statistics. Springer.

Breiman, L., (1996). Bagging predictors. Machine learning, 24, 123–140.

Breiman. L., Friedman, J.H., Olshen, R.A., (1984). CART: Classification and Regression Trees. Biometrics, 40, 358–380.

Chen, N., Qian, Z., Nabney, I., Meng, X., (2014). Wind power forecasts using Gaussian processes and numerical weather prediction. IEEE Trans. Power Syst., 29(2), 656-665.

Chen, N., Qian, Z., Meng, X., (2013). Short-term wind power forecasting using Gaussian processes. Proceedings of 23rd international joint conference on artificial intelligence, 3-9 August 2013. pp. 2790-2796.

Cornejo-Bueno, L., Garrido-Merchán, E.C., Hernández-Lobato, D., Salcedo-Sanz, S., (2018). Bayesian optimization of a hybrid system for robust ocean wave features prediction. Neurocomputing, 275, 818–828.

Deng, H., Fannon, D., Eckelman, M.J., (2018). Predictive modeling for US commercial building energy use: a comparison of existing statistical and machine learning algorithms using CBECS microdata. Energy Build, 163, 34-43.

Don, B., Cao, C., Lee, S.E., (2005). Applying support vector machines to predictbuilding energy consumption in tropical region. Energy Build, 37 (5). 545-553.

Erisen, B., (2019). Wind turbine scada dataset, Version 3. Available Online: https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset (accessed on 10 August 2020).

Eroğlu, M., (2019). Enerji Hukuku. İstanbul: BAU.

Esfetang, N. N. & Kazemzadeh, R., (2018). A novel hybrid technique for prediction of electric power generation in wind farms based on WIPSO neural network and wavelet transform. Energy, 149, 662-674.

ETKB. (2019). Enerji ve Tabii Kaynaklar Bakanlığı, Available Online: https://enerji.gov.tr/bilgi-merkezi-enerji-elektrik (accessed on 4 January 2021).

Fu. C., Li. G.-Q., Lin. K.-P., Zhang. H.-J., (2019). Short-term wind power prediction based on improved chicken algorithm optimization support vector machine. Sustainability, 11, 512.

Gao, Z., Shi, J., Li, H., Chen, C., Tan, J., Liu, L., (2020). Substation Load Characteristics and Forecasting Model for Large-scale Distributed Generation Integration. IOP Conf. Series: Materials Science and Engineering, 782, 032044 doi:10.1088/1757-899X/782/3/032044

Hastie, T., Tibshirani, R., Friedman, J., (2009). The Elements of Statistical Learning: Data Mining. Inference. and Prediction. Second Edition. Springer Series in Statistics. Springer.

Heinermann, J. & Kramer, O., (2016). Machine learning ensembles for wind power prediction. Renewable Energy.89, 671-679.

Heo, Y. & Zavala, V.M., (2012). Gaussian process modeling for measurement and verification of building energy savings. Energy Build, 53, 7-18.

Karık, F., Sözen, A. & İzgeç, M.M., (2017). Rüzgâr gücü tahminlerinin önemi: Türkiye elektrik piyasasında bir uygulama. Politeknik Dergisi. 20(4), 851-861.

Kerem, A., (2018). Rüzgar Parametrelerinin Değişiminin İzlenmesi ve Yapay Zeka Algoritmaları Kullanarak Tahmini. Doktora Tezi. Gazi Üniversitesi

Lee, D. & Baldick, R., (2014). Short-term wind power ensemble prediction based on Gaussian processes and neural networks. IEEE Trans Smart Grid, 5(1), 501-510.

Li, C., Lin, S., Xu. F., Liu, D., Liu, J., (2018). Short-term wind power prediction based on data mining technology and improved support vector machine method: a case study in Northwest China. J. Cleaner Prod., 205(4), 909-922, 10.1016/j.jclepro.2018.09.143

Li, L.-L., Zhao, X., Tseng, M.-L., Tan, R.R., (2020). Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. J Clean Prod. 242, 118447. 10.1016/j.jclepro.2019.118447

Li, Q., Meng, Q., Cai, J., Yoshino, H., Mochida, A., (2009). Applying support vector machine to predict hourly cooling load in the building. Appl Energy, 86(10), 2249- 2256.

Ma, Y.-J. & Zhai, M.-Y., (2019). A dual-step integrated machine learning model for 24h-ahead wind energy generation prediction based on actual measurement data and environmental factors. Appl Sci., 9, 21-25. 10.3390/app9102125

Mendes-Moreira, J.A., Soares, C., Jorge, A.M., Sousa, J.F.D., (2012). Ensemble approaches for regression: A survey. ACM Comput. Surv., 45(1), 1-10. http://dx.doi.org/10.1145/2379776.2379786

NOAA, 2018. National Weather Service Center Environmental Forecast Climate. Avaliable Online: https://www.ncdc.noaa.gov/cdo-web/ (accessed on 15 October 2020).

Öz, S. & Alyürük, M., (2020). Energy Sector Overview and Future Prediction for Turkey. Journal of Industrial Policy and Technology Management. 3(1). 2020. 59-69

Petropoulos, A., Chatzis, S.P., Siakoulis, V., Vlachogiannakis, N., (2017). A stacked generalization system for automated forex portfolio trading. Expert Syst. Appl., 90, 290–302. http://dx.doi.org/10.1016/j.eswa.2017.08.011

Rasmussen. C. E. &Williams, C. K. I., (2006). Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts.

Serbes, G., Sakar, B.E, Gulcur, H.O., Aydin, N., (2015). An emboli detection system based on dual tree complex wavelet transform and ensemble learning. Appl. Soft Comput., 37 (2015) 87–94. http://dx.doi.org/10.1016/j.asoc.2015. 08.015

Shamaei, E. & Kaedi, M., (2016). Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. Appl. Soft Comput., 45, 187–196. http://dx.doi.org/10.1016/j.asoc.2016.03.009

Şenel, M.C. & Koç, E., (2015). Dünyada ve Türkiye'de Rüzgâr Enerjisi Durumu-Genel Değerlendirme. Mühendis ve Makina. 56(663), 46-56.

Tahir, M., El-Shatshat, R., Salama, M.M.A., (2018). Improved stacked ensemble based model for very short-term wind power forecasting. In: Proceedings – 53rd International universities power engineering conference (UPEC). Glasgow. 4–7 September.

TÜREB, (2021). Türkiye Rüzgar Enerjisi Birliği RES Veritabanı, Available Online: https://www.tureb.com.tr/ (accessed on 31 December 2020).

Van der Laan, M.J., Polley, E.C., Hubbard, A.E.,

(2007). Super learner. Statistical Applications in Genetics and Molecular Biology, 6(1) http://dx.doi.org/10.2202/1544-6115.1309

Wan, Z.Y. & Sapsis, T.P., (2017). Reduced-space Gaussian Process Regression for data-driven probabilistic forecast of chaotic dynamical systems. PhysicaD, 345, 40–55.

Wang, C., Li, S., Zhu, M., (2012). Bayesian network learning algorithm based on unconstrained optimization and ant colony optimization. Syst Eng Electron, 5, 784–90.

Wolpert, D.H., (1992). Stacked generalization. Neural Netw. 5 (2), 241–259. http://dx.doi.org/10.1016/S0893-6080(05)80023-1.

YEKDEM, (2020). Yenilenebilir Enerji Kaynakları Destekleme Mekanizması (YEKDEM), Available Online:https://www.enerjiportali.com/wpcontent/uploads/2020/09/KSD_YEKDEM_2.09.2020-1-1.pdf (accessed on 8 January 2020).

Zendehboudi, A., Baseer, M.A., Saidur, R., (2018). Application of support vector machine models for forecasting solar and wind energy resources: a review. J. Clean. Prod., 199, 272-285.

Zheng, D., Shi, M., Wang, Y., Eseye, A. T., Zhang, J. (2017). Day-Ahead Wind Power Forecasting Using a Two-Stage Hybrid Modeling Approach Based on SCADA and Meteorological Information, and Evaluating the Impact of Input-Data Dependency on Forecasting Accuracy. Energies, 10(12), [1988]. https://doi.org/10.3390/en10121988

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 155-167
DOI: 10.38016/jista.918795

Analysis of the Parameters that Affect the Measurements of Reflection Coefficients and Evaluation of the Effects of Parameters for K Nearest Neighbors-Based Liquid Classification

Ebru Efeoglu [1], Gurkan Tuna[2*]

[1] İstanbul Gedik University, Faculty of Economics, Administrative and Social Sciences, Department of Management Information Systems, İstanbul/Turkey

[2] Trakya University, Edirne Vocational College of Technical Sciences, Edirne/Turkey

ebru.efeoglu@gedik.edu.tr, gurkantuna@trakya.edu.tr

Abstract
In this study, microwave spectroscopy method has been used in liquid measurements and K nearest neighbors algorithm has been used for classifying liquids. For this aim, firstly an experimental setup consisting of a vector network analyzer, a patch antenna and a bottle have been built to measure the reflection parameter of each liquid used in classification experiments. The aim of this study is to examine both the parameters that may affect the measurements taken with the proposed system and the algorithm parameters that may affect the performance in the classification of liquids and the effects of these parameters. Measurements have been taken by leaving different distances between the antenna and the liquid in order to examine whether the distance of the liquids to the antenna affects the measurement result, and if so, what effect. For examining the parameters of K nearest neighbors algorithm that may affect the classification, the scattering parameters of different liquids measured using the patch antenna have been used as microwave dataset. In addition, the effect of container type has been analyzed. Performance tests have been conducted by weighting and without weighting the algorithm, by measuring the accuracy rate when different numbers of nearest neighbors and different distance metrics have been used. The results reveal that the classification made by applying weighting is more successful than the classification made without weighting regardless of the number of nearest neighbors and used distance metrics.

Keywords: Microwave measurement, circular patch antenna, K nearest neighbors algorithm, liquid classification, distance metric, weighting, number of nearest neighbors.

K En Yakın Komşular Tabanlı Sıvı Sınıflandırması İçin Yansıma Katsayıları ve Parametrelerin Etkilerinin Değerlendirilmesi

Öz
Bu çalışmada sıvı ölçümlerinde mikrodalga spektroskopi yöntemi kullanılmış ve sıvıların sınıflandırılmasında K en yakın komşular algoritması kullanılmıştır. Bu amaçla, öncelikle sınıflandırma deneylerinde kullanılan her bir sıvının yansıma parametresini ölçmek için bir vektör ağ analizörü, bir yama anteni ve bir şişeden oluşan deney düzeneği oluşturulmuştur. Bu çalışmanın amacı, hem önerilen sistemle alınan ölçümleri etkileyebilecek parametreleri hem de sıvıların sınıflandırılmasında performansı etkileyebilecek algoritma parametrelerini ve bu parametrelerin etkilerini incelemektir. Sıvıların antene olan mesafesinin ölçüm sonucunu etkileyip etkilemediğini, etkiliyorsa etkisini incelemek için anten ile sıvı arasında farklı mesafeler bırakılarak ölçümler yapılmıştır. Sınıflandırmayı etkileyebilecek en yakın komşu algoritmasının parametrelerini incelemek için, yama anten kullanılarak ölçülen farklı sıvıların saçılma parametreleri mikrodalga veri seti olarak kullanılmıştır. Ayrıca kap tipinin etkisi analiz edilmiştir. Farklı sayıda en yakın komşu ve farklı mesafe ölçütleri kullanıldığında doğruluk oranı ölçülerek ağırlıklandırılarak ve algoritma ağırlıklandırılmadan performans testleri yapılmıştır. Sonuçlar, ağırlıklandırma uygulanarak yapılan sınıflandırmanın, en yakın komşu sayısına ve kullanılan uzaklık ölçütlerine bakılmaksızın ağırlıklandırma yapılmadan yapılan sınıflandırmaya göre daha başarılı olduğunu ortaya koymaktadır.

Anahtar Kelimeler: Mikrodalga ölçümü, dairesel yama anten, K en yakın komşular algoritması, sıvı sınıflandırması, mesafe ölçüsü, ağırlıklandırma, en yakın komşu sayısı.

* Corresponding Author.
E-mail: gurkantuna@trakya.edu.tr

Recieved : 17 April 2021
Revision : 07 June 2021
Accepted : 12 July 2021

1. Introduction

The measure of a liquid being a flammable liquid is the flash point. Flash point is the lowest temperature at which a liquid will emit sufficient vapor to form an air-flammable mixture. Alcohols cannot be distinguished visually because they are 100% liquid and colorless liquids, and they are in a small group of chemicals that can spontaneously ignite (Q. Chen, Kang, Zhou, & Wang, 2017). The flash point of alcohol-water solutions diluted with water will increase and as the flash point value of flammable liquids increases, the risk of fire hazards decreases (Cheremisinoff, 1999), However, these liquids, besides their flammable properties, also contain toxic substances and endanger human health. For instance, drinking methanol accidentally causes serious health problems (Slaughter, Mason, Beasley, Vale, & Schep, 2014). From this perspective, the classification of liquids is important to manage the hazards of chemicals and take necessary measures.

Classification of flammable and explosive liquids using THz time history spectroscopy in classification of liquids (Tan et al., 2017), characterization of aqueous alcohol solutions in bottles and determination of the alcohol content of an aqueous solution were performed (Jepsen, Jensen, & Møller, 2008). An electronic nose using machine learning was proposed to detect mixtures of water, methanol and ethanol (Hayasaka et al., 2020). Flammable liquids were detected using the X-ray spectroscopy method (Orachorn, Chankow, & Srisatit, 2019), (H. Chen, Hu, Wang, Xu, & Hou, 2020). Raman spectroscopy method was used for screening and determination of methanol content in ethanol-based products (Wirasuta et al., 2019).

Liquids differ in complex permeability and reflection and transmission coefficients. Microwave frequency bands can be used to determine complex permeability, reflection and transmission coefficients of liquids and to characterize liquids. Liquid characterization is also important for food safety and quality. For fruit quality control, non-destructive control experiments with microwave method were performed (Jawad et al., 2017). It was also used to calculate the permeability, reflection coefficient, S_{11}, and transmission coefficient, S_{21}, (Li, Haigh, Soutis, Gibson, & Sloan, 2018), (Jiang, Ju, & Yang, 2016) of the liquids. Microwave measurement method is fast, non-hazardous and not affected by environmental conditions (Li, Haigh, Soutis, Gibson, & Sloan, 2017b). It was used to measure the permeability of thin layer materials (Borisov & Karpenko, 2001) and to measure the parameters of silicon (Yurchenko, Novikov, & Kitaeva, 2012).

There are many microwave measurement methods used in fluid measurements in the literature, such as open-ended coaxial probe techniques (Li et al., 2017b) and Free space method (Jose, Varadan, & Varadan, 2001). When the coaxial probe method is used, the probability of inaccuracy in solid material measurements is high. The cost of the measurement method to be used is also important. For example, Time-Domain Reflectometers (TDRs) are expensive (Venkatesh & Raghavan, 2005). For Free space technique, measurements vary according to the choice of the horn antenna, the design of the specimen holder and the geometry and location of the specimen. An improperly determined measurement location and an unsuitable sample geometry increase the likelihood of erroneous measurements (Li, Haigh, Soutis, Gibson, & Sloan, 2017a).

K Nearest Neighbors (KNN) algorithm does not require a training step and is resistant to noisy training data (Bhatia, 2010). Therefore, it is commonly used as a basic classifier in many field problems (Jain, Duin, & Mao, 2000). KNN is known as instance-based learning. In it, training samples are stored exactly and the classification of an unknown, i.e., a new test sample, takes into account the similarity between the samples in the training set. The similarity is, for example, its proximity to the data in the training set. The distance metric is used to decide which member of the training set is the nearest. Once the nearest training sample is found, the class is estimated for the test sample (Chakrabarti et al., 2008). For this, the test sample is compared with the records that are most similar in the current training set at hand (Larose & Larose, 2014). In the literature, there are several distance metrics. However, the most commonly used distance function for KNN is the Euclidean distance metric. The microwave measurement methods and KNN algorithm were used together for different purposes including classification of kidney stones (Saçlı et al., 2019), detection of deep tissue injuries (Moghadas & Mushahwar, 2018) and detection of breast cancer (Aydın & Kaya Keleş, 2017).

Although the performance of KNN has been heavily studied, it has not been evaluated for classifying alcoholic liquids with different distance metrics, the different number of nearest neighbors, and the weighting process. Different from the literature, in this study it is evaluated whether weighting application using different distance metrics and changing the number of nearest neighbors can affect the performance of KNN algorithm when it is used for alcoholic liquid classification made with microwave datasets. Another issue examined in the study is the parameters that may affect the measurements. Determining these parameters and paying attention to them while making measurements ensure more reliable measurements. The remainder of this paper is as follows. First, the factors affecting microwave measurements are examined and the effects of different parameters on the measurement results are presented. Parameters used for KNN algorithm and their implementation are described in Section 2. Experimental setup of this study which was used to collect and use microwave measurement data is explained in Section 3. Finally, Section 4 concludes this paper.

2. Parameters of K Nearest Neighbors Algorithm for Liquid Classification and Their Implementation

In KNN, the similarities of the data to be classified with the data in the training set are computed. As a result of the computation, the data to be classified is assigned to the nearest classes in the training set. The nearest number of neighbors and similarity function criteria affect the performance of KNN (Kresse & Danko, 2012). In KNN, training samples are defined with n-dimensional numerical properties. Each sample shows a point in n-dimensional space. Thus, all training samples are stored in an n-dimensional sample space. The objective is to find the nearest *k* training samples to the unknown sample. The distance between two points, such as X= (x_1, x_2,…., X_n) and Y= (y_1, y_2, ……., Y_n) is expressed by different distance metrics (Chakrabarti et al., 2008).

Basic parameters of KNN algorithm are distance metric, number of nearest neighbors, *k*, and weighting application. *k* expresses the number of neighbors, and classification is made based on this value. For instance, if *k* value is set to 1, the nearest 1 neighbor is taken into consideration and the tested sample is assigned to the class where this neighbor is located. In this study, four different distance metrics will be used.

As given in (1) Minkowski distance metric is calculated by summing the absolute difference between the two points by taking the *p* prime in the distance criterion. Then 1/*p* of this sum is taken. This equation gives Euclid distance if *p* value is set to 2, Manhattan distance if $p \to \infty$, and Chebyshev distance if *p* value is set to 1 (Kresse & Danko, 2012).

$$\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \quad (1)$$

Euclidean distance metric, defined as a straight line distance between two points in any number of dimension spaces, is calculated by taking the square root of the sum of the squares of the differences between the respective coordinates of each point, as given in (2) (Kresse & Danko, 2012).

$$\left(\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\right) \quad (2)$$

As given in (3) Manhattan distance metric calculates the linear distance between actual vectors using the sum of absolute differences (Kresse & Danko, 2012).

$$\left(\sum_{i=1}^{n}|x_i - y_i|\right) \quad (3)$$

Finally, Chebyschev distance metric, also known as the maximum value distance, is calculated using (4) (Rey, Kordon, & Wells, 2012).

$$lim_{p\to\infty}\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} = max_{i=1}^{n}|x_i - y_i| \quad (4)$$

In KNN, weight values are assigned to all neighbors. The weight values of neighboring samples that are closer to the sample to be classified in the weighting application are higher than the other neighbors. Generally, the most preferred method of assigning weight is the method in which the weight of each neighbor is taken in 1/*d*. Here *d* represents the distance between neighbors (Doad & Bartere, 2013).

2.1. Performance of the Classifier

There are several performance metrics used to evaluate how well a classifier is performing at the end of the classification process (Chakrabarti et al., 2008). The metrics used in this study were confusion matrix, accuracy, precision, recall, Kappa, Area Under the Receiver Operating Characteristic (ROC) Curve (AUC), Matthews Correlation Coefficient (MCC) and Root Mean Square (RMS). Confusion matrix is often used to determine the performance of the classification model with a series of test data with actual values known (Chakrabarti et al., 2008), (Larose & Larose, 2014). True Positive (TP) values, i.e. actual alcoholic liquids, are positive values that have been predicted correctly. True Negative (TN) values, i.e. actual non-alcoholic liquids, are negative values that have been predicted correctly. These values indicate that, for the selected sample, the actual class is the same as the predicted class. They are the diagonal elements of the matrix and are shown in green in Figure 1. False Positive (FP) values, i.e. non-alcoholic liquids misclassified as alcoholic liquids, and False Negative (FN) values, i.e. alcoholic liquids misclassified as non-alcoholic liquids, occur when the actual class is different from the predicted class. That is, they indicate the number of incorrectly classified samples. They are shown in red in Figure 1. The increase in TP and TN values and the decrease in FN and FP values indicate that the classification performance is good.

Using the confusion matrix, the accuracy value can be calculated as in (5). Accuracy is the ratio of accurately estimated samples to the total number of samples. High accuracy rate is an indicator of high classification performance.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (5)$$

		Predicted	
		TP (Alcoholic Liquids)	TN (Non-Alcoholic Liquids)
Actual	TP (Alcoholic Liquids)	TP	FN
	TN (Non-Alcoholic Liquids)	FP	TN

Figure 1. Confusion matrix

Other performance metrics are also calculated using a confusion matrix. For example, precision is calculated using the left side of the matrix (Equation (6)). It is a measure of the precision of the classification algorithm. Recall, which is a measure of the integrity of the classification algorithm, is calculated using (7).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN + TP}} \quad (7)$$

The harmonic mean of precision and recall values gives F-measure value (8). It is difficult to compare the two models with low recall and high precision and vice versa. In this case, the value of F-measure is checked.

$$\text{F\_Measure} = \frac{2*\text{Recall}*\text{ Precision}}{\text{Recall+ Precision}} \quad (8)$$

Kappa value is used to measure how much agreement there is between the classification made as a result of the classification and the actual classifications in a dataset.

$$\text{Kappa} = \frac{\text{P(x)}-\text{P(y)}}{1-\text{P(y)}} \quad (9)$$

where P(x) is a value that shows probabilistic accuracy of the classification algorithm and P(y) is the weighted average of the probability of classifications made in the same dataset.

In a ROC curve, the horizontal axis shows the false positive rate (FPR), the vertical axis the correct positive rate (TPR). The area under this curve (AUC) is used as the classification metric. FPR and TPR values are calculated using (10) and (11), respectively.

$$\text{FPR} = \frac{FP}{FP+TN} \quad (10)$$

$$\text{TPR} = \frac{TP}{TP+FN} \quad (11)$$

MCC is used as a measure of the quality of binary classifications in machine learning and is calculated using (12).

$$MCC = \frac{TPxTN-FPxFN}{\sqrt{(TP+FP)x(TP+FN)x(TN+FP)x(TN+FN)}} \quad (12)$$

RMS is used to scale the differences between the actual values and the values predicted by the model. It is determined by taking the square root of the mean square error and calculated using (13).

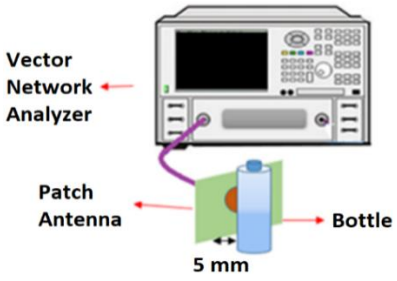
$$\text{Rms} = \sqrt{\frac{1}{\text{n}}\sum_{\text{k=1}}^{\text{n}}(\text{T}_{\text{ik}} - \text{A}_{\text{k}})^2} \quad (13)$$

where T_{ik} is the predicted value and A_k is the objective value. If the error value approaches zero, it means that the correct prediction of the classification algorithm increases.

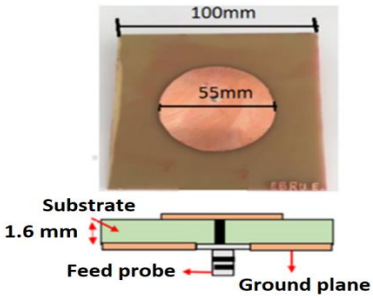
When the values of accuracy, precision, recall, F-measure, AUC, MCC and Kappa are 1, it indicates perfect classification. Therefore, these values are desired to be as close to 1 as possible.

3. Experimental Setup for Collecting and Using Microwave Measurement Data

Scattering parameters (S parameters) describe the electrical behavior of linear electrical networks when they are exposed to various steady-state stimuli by electrical signals. The measurement system shown in Figure 2 consists of a vector network analyzer (VNA) and a circular patch antenna that can send signals between a specific frequency band and record the reflection coefficient of the reflected signals. The reflection coefficient (S_{11}) of the reflected signals from the source is expressed as the ratio of the amplitude of the reflected signal to the amplitude of the transmitted signal. The resonance frequency of the antenna fed with the 50 Ohm SMA (SubMiniature version A) feed probe is 1.5 GHz. The measurement setup used in this study is given in Figure 2. The reflection parameter (S_{11}) of each liquid was measured so that the distance between the antenna and the bottle remains 5 mm without touching the patch antenna to the liquids in 0.5 liter pet bottles with the microwave measurement device. Measurements were made between 1.42-1.54 Ghz.



(a)



(b)

Figure 2. a) Experimental setup, b) Schematic view of the patch antenna

There were 56 S_{11} values for each liquid. Then, these values were divided into two classes as alcoholic and non-alcoholic liquids by using KNN algorithm. The liquids used in the study were non-alcoholic liquids such as cola, soap, shampoo, water, milk, bath cream, shower gel, ice-tea (peach), cherry juice, ayran and alcoholic liquids such as cologne, whiskey, white wine, raki.

Apart from these liquids, Ethanol, Methanol, 1-Propanol, Isopropanol and their aqueous solutions with different volume concentrations were used. The total number of liquids used was 54, including 44 alcoholic and 10 non-alcoholic. In analyzing the effect of the type of container in which the liquid is on the measurement result in liquid measurements, measurements were taken by placing the liquids in glass and plastic bottles and the measurement results were compared. The aim of the experiments was to determine the factors affecting S parameter measurements and to examine the effects of KNN parameters on classification performance in order to characterize the liquid with high accuracy in measurements. In determining the factors affecting the measurements in the experiments, the effect of the distance of the bottle to the antenna and the effect of the container that the liquid is in were analyzed. The parameters used to examine the effects of KNN parameters on classification performance are *k* value, distance metric and weighting application. In order to examine the classification parameters, a separate classification was made for each parameter in the classification of alcoholic and non-alcoholic liquids using microwave data. The best values were tried to be determined by comparing the classification results.

3.1. Results and Discussion

In order to find out whether the distance between the antenna and the liquid affects the microwave measurement data, measurements were taken by leaving different distances between the antenna and the liquid. The measurement of liquids in pet bottles was made by leaving 5 mm, 1 mm and 25 mm between antenna and liquid. The results of alcohol measurements taken for three different distances are given in Figure 3. This step was repeated for non-alcoholic liquids to examine the effect of measurement distance, and the results are given in Figure 4. As it can be seen in the figures, it is seen that the resonance peak increases with the increase of the distance between the antenna and the liquid in the measurements of all liquids. Since the resonance peak gives its highest value in the air environment, an increase in the value of the peak as it moves away from the antenna in liquid measurements indicates that the sensitivity of the antenna decreases. In other words, as the liquid moves away from the antenna, the sensitivity of the antenna to detect the liquid decreases. In order to measure the liquid accurately, the distance between the liquid and the antenna should be as small as possible. Another parameter whose effect on measurement data is examined is the container effect. For this, measurements were taken using different containers and the results are given in Figure 5 and Figure 6. Glass bottles and plastic bottles were used as container types in the measurements.

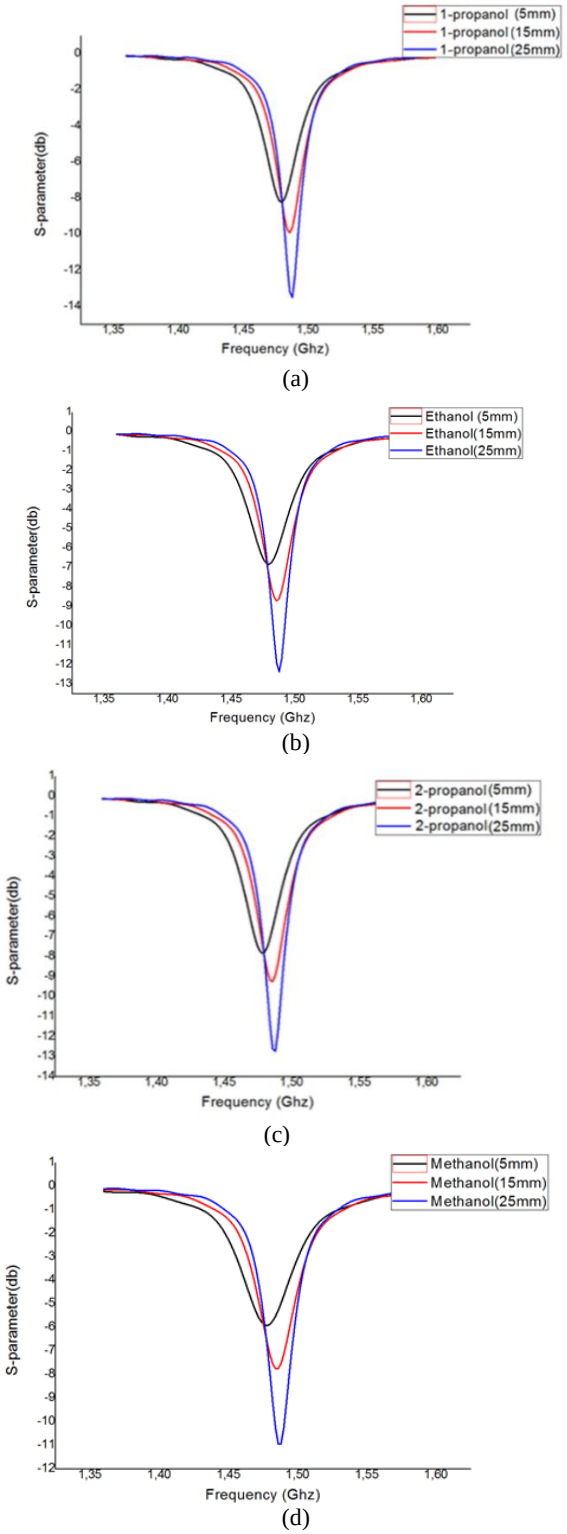


Figure 3. S parameter measurements of alcohols at different antenna-bottle distances a) Isopropyl (1-propanol) b)Ethanol c)Isopropanol (2-propanol) d)Methanol

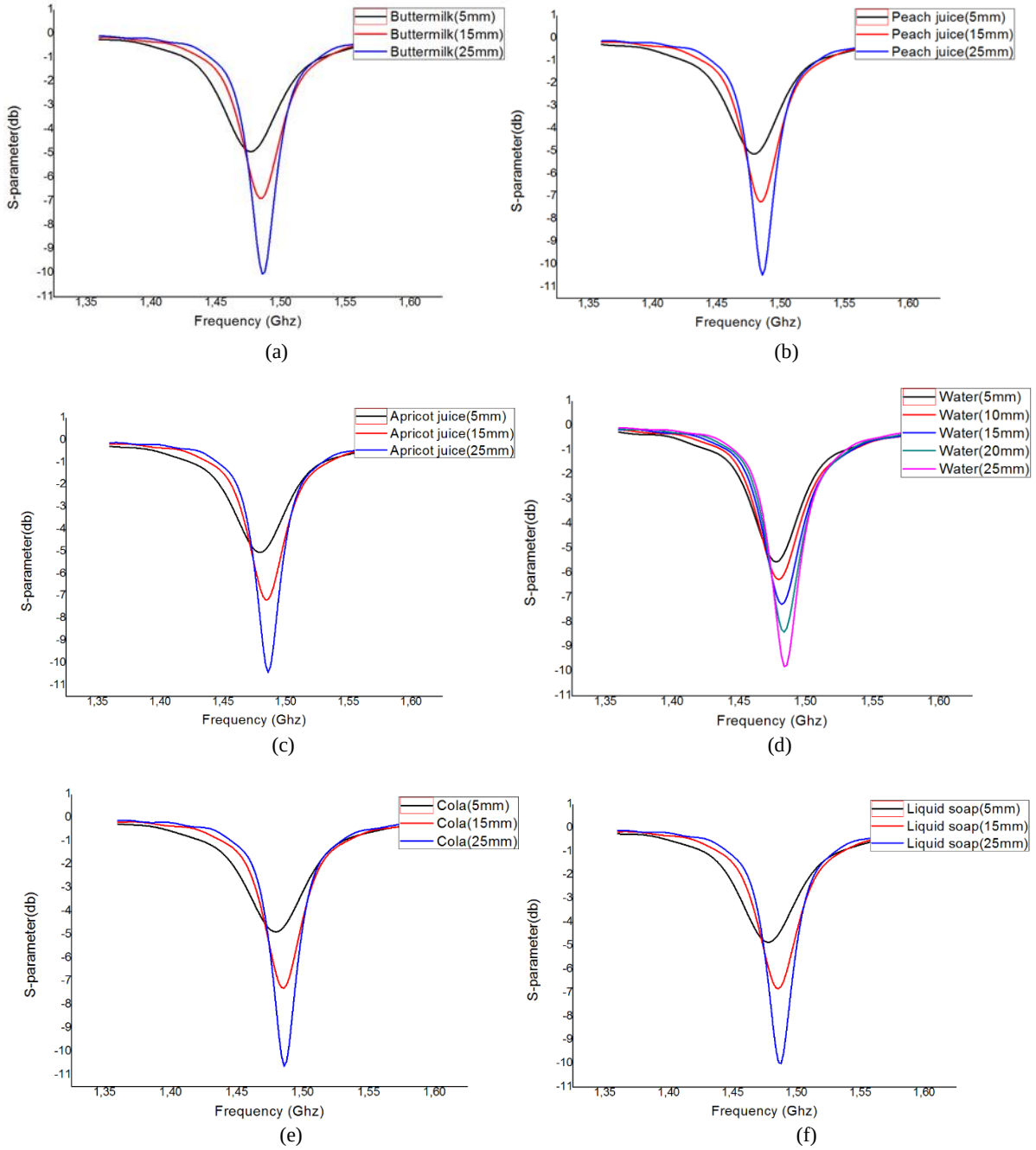
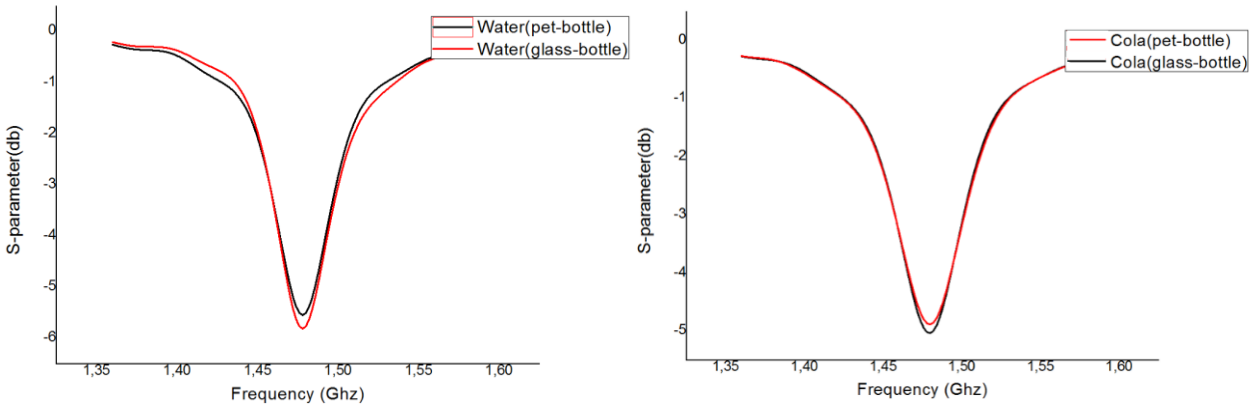


Figure 4. S parameter measurements of everyday liquids at different antenna-bottle distances a) Buttermilk b) Peach juice c) Apricot juice d) Water e) Cola f) Liquid soap



(a) (b)

Figure 5. S parameter measurements of everyday liquids in different bottles a) Water b) Cola

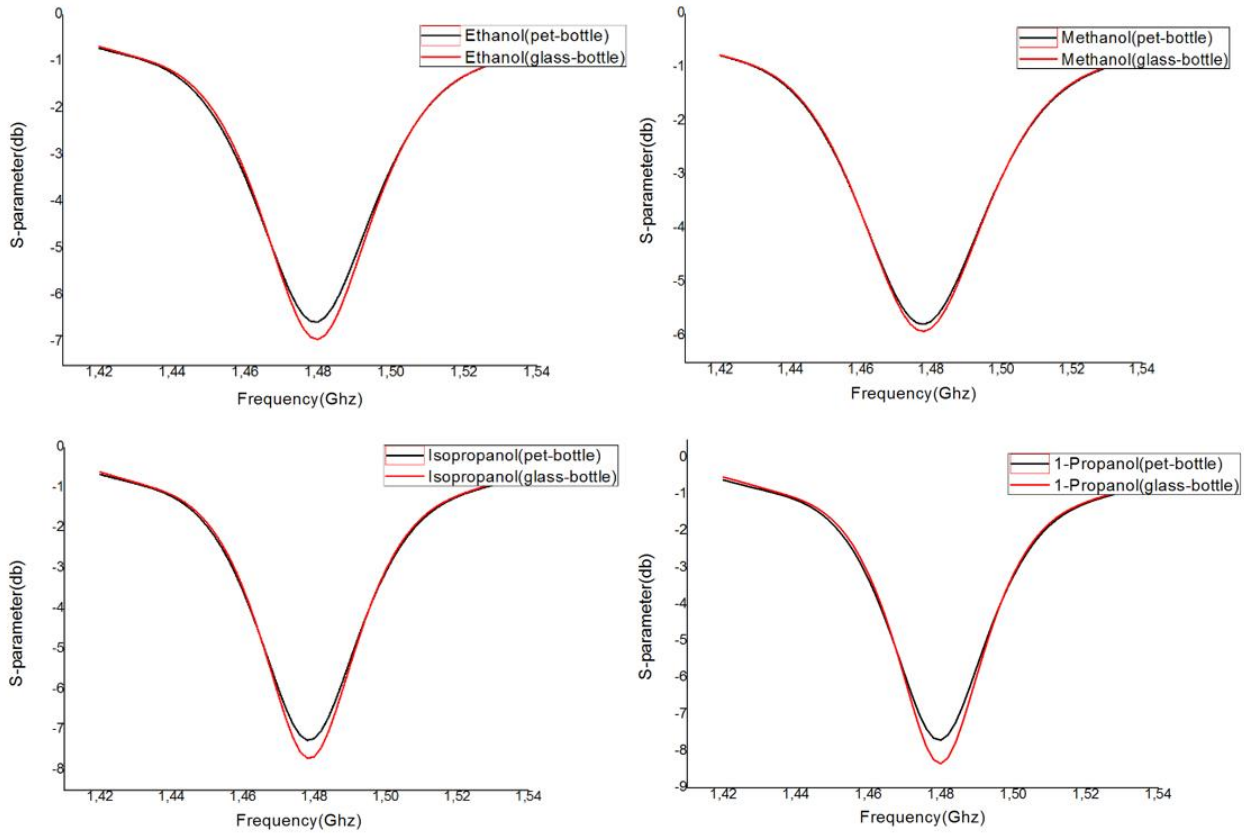


Figure 6. S parameter measurements of alcohols in different bottles

The measurement results taken using different containers show us that the resonance peak gives a higher value in the measurements taken using the glass container (bottle). S_{11} parameter measurements of the liquids were used in the study and S_{11} parameter is the reflection coefficient. Signals reflected from the object to be measured are detected by the antenna. Therefore, taking measurements in (using) a highly reflective container like glass will suppress the reflected signals from the liquid. Therefore, since the signals reflected from the liquid are important for us, it is more appropriate to use pet bottles with low reflectivity in liquid measurements. It was understood that the measurement results, the distance and the type of container used in microwave measurements affect the results. Classifications were made using different parameters to investigate the KNN parameters affecting the classification. The dataset used in the classification studies is given in Figure 7.

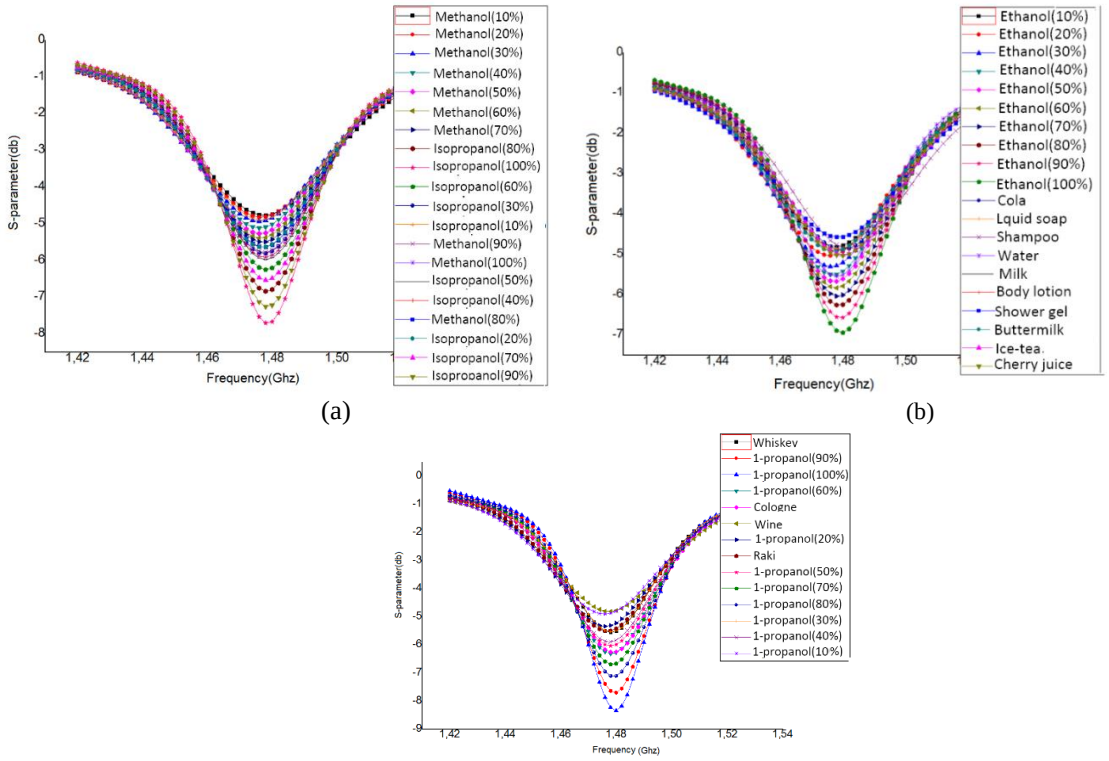


Figure 7. S parameter measurement results used in classification

For each liquid, the measurements were repeated twice and a dataset was created. The dataset was created from 108 measurement data belonging to 54 different liquids. Then, using WEKA, 10-fold cross-validation was applied. As shown in Figure 8, the dataset was divided into 10 parts, 9 parts were used for the training and 1 part was used as the test data.

Cross Validation	1st group	2nd group	3rd group	4th group	5th group	6th group	7th group	8th group	9th group	10th group
1st	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
2nd	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
3rd	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
4th	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
5th	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
6th	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
7th	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
8th	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
9tn	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
10th	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

Figure 8. 10-fold cross-validation

After the cross-validation, tradeoff between *k* value and the accuracy rate of the classifier for different distance metrics is shown in Figure 9. As can be seen, the accuracy rate of the classifier mostly decreased as the number of neighbors increased for all the distance metrics. The highest accuracy rate was 100% and the lowest accuracy rate was 90%. As can be seen in Figure 10 and Figure 11, the number of correctly classified liquids was 108 when *k* was set to 1 for different distance metrics, while the number of correctly classified liquids decreased to 97 when Chebyshev distance metric was used and *k* was set to 7. Likewise, although there was no misclassified liquid when *k* was set to 1, it was seen that a total of 11 liquids were misclassified when *k* was set to 7. After the classification experiments with these neighbor numbers and distance metrics, weighting was applied. As a result of this, for all the k values and distance metrics 100% accuracy was obtained and all the liquids were classified correctly. The confusion matrices obtained for different *k* values and the distance metrics are given in Figure 12. When the confusion matrix is analyzed, it can be seen that all the liquids were correctly classified when *k* was set to 1 and Euclidean distance metric was preferred. However, when *k* was set to 2 and Euclidean distance metric was preferred, 2 alcoholic liquids were incorrectly classified as non-alcoholic liquids and 6 non-alcoholic liquids were incorrectly classified as alcoholic liquids.

When the confusion matrix is analyzed, it can be seen that the average number of alcoholic liquids that was predicted incorrectly for different *k* values was 6.8 when Euclidean distance metric was preferred, was 6.8 when Chebyshev distance metric was preferred, was 7.2 when Manhattan distance metric was preferred, and finally it was 6.8 when Minkowski distance metric was preferred. This reveals that the distance measure which is the least affected by the change of *k* value in the detection of alcoholic liquid is Manhattan distance.

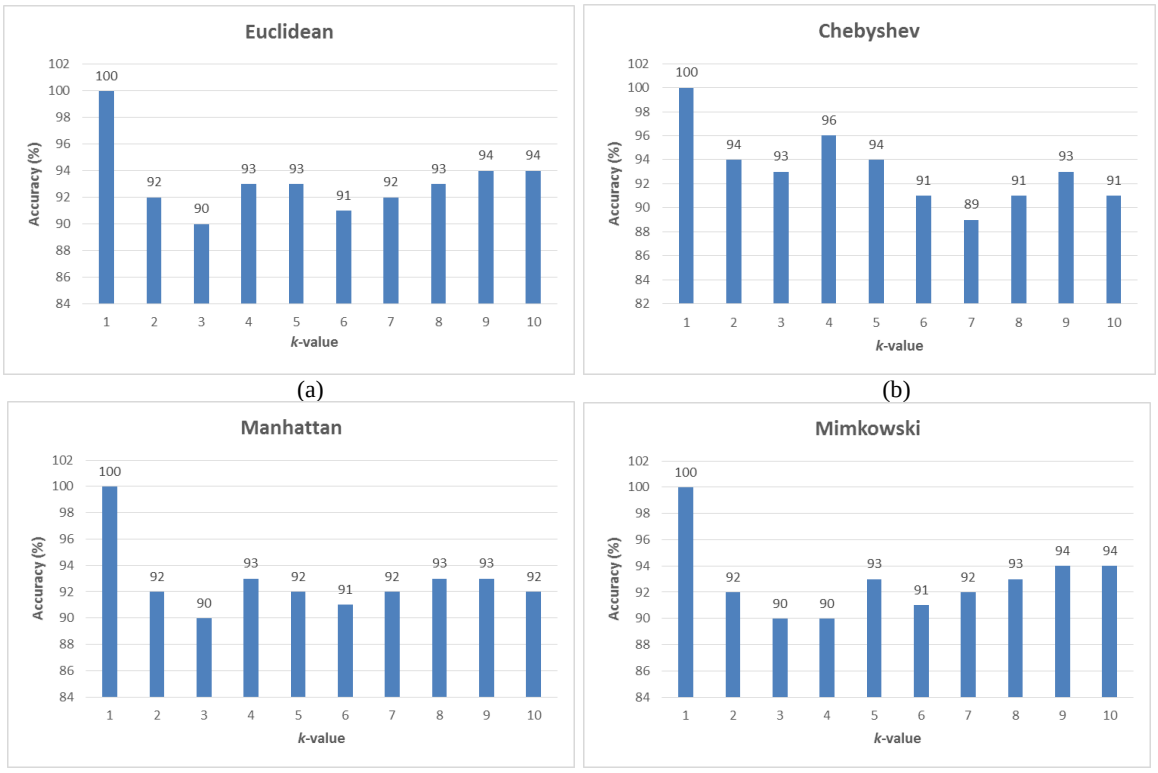


Figure 9. Tradeoff between *k* value and accuracy rate when the distance metric was a) Euclidean, b) Chebyshev, c) Manhattan, d) Minkowski

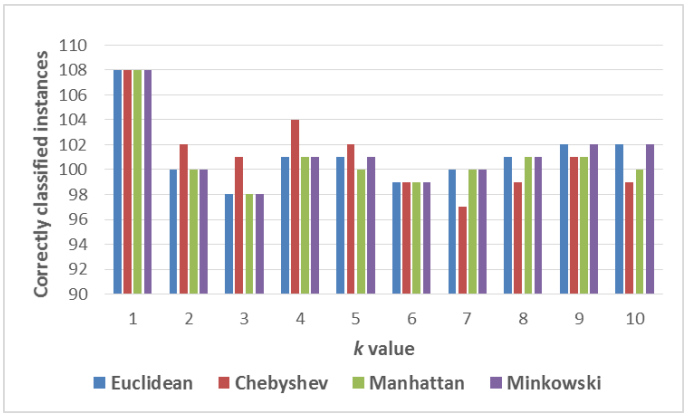


Figure 10. Tradeoff between *k* value and the number of correctly classified instances

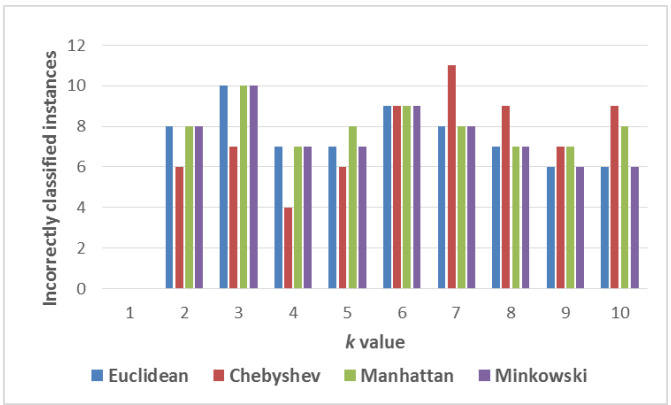


Figure 11. Tradeoff between *k* value and the number of incorrectly classified instances

	k:1			*k*:2			*k*:3			*k*:4			*k*:5		
Euclidean		**Predicted**														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	88	0	TP	86	2	TP	86	2	TP	86	2	TP	84	4
		TN	0	20	TN	6	14	TN	8	12	TN	5	15	TN	3	17
	k:6				*k*:7			*k*:8			*k*:9			*k*:10		
		Predicted														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	83	5	TP	85	3	TP	86	2	TP	86	2	TP	86	2
		TN	4	16	TN	5	15	TN	5	15	TN	4	16	TN	4	16
	k:1				*k*:2			*k*:3			*k*:4			*k*:5		
Chebyshev		**Predicted**														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	88	0	TP	86	2	TP	86	2	TP	86	2	TP	84	4
		TN	0	20	TN	4	16	TN	5	15	TN	2	18	TN	2	18
	k:6				*k*:7			*k*:8			*k*:9			*k*:10		
		Predicted														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	83	5	TP	83	5	TP	84	4	TP	84	4	TP	83	5
		TN	4	16	TN	6	14	TN	5	15	TN	3	17	TN	4	16
	k:1				*k*:2			*k*:3			*k*:4			*k*:5		
Manhattan		**Predicted**														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	88	0	TP	86	2	TP	86	2	TP	86	2	TP	84	4
		TN	0	20	TN	6	14	TN	8	12	TN	5	15	TN	4	16
	k:6				*k*:7			*k*:8			*k*:9			*k*:10		
		Predicted														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	83	5	TP	85	3	TP	86	2	TP	86	2	TP	86	2
		TN	4	16	TN	5	15	TN	5	15	TN	5	15	TN	6	14
	k:1				*k*:2			*k*:3			*k*:4			*k*:5		
Minkowski		**Predicted**														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	88	0	TP	86	2	TP	86	2	TP	86	2	TP	84	4
		TN	0	20	TN	6	14	TN	8	12	TN	5	15	TN	3	17
	k:6				*k*:7			*k*:8			*k*:9			*k*:10		
		Predicted														
	Actual		TP	TN		TP	TN		TP	TN		TP	TN		TP	TN
		TP	83	5	TP	85	3	TP	86	2	TP	86	2	TP	86	2
		TN	4	16	TN	5	15	TN	5	15	TN	4	16	TN	4	16

Figure 12. Confusion matrix for different *k* values and different distance metrics

When the confusion matrix is considered to examine the average number of non-alcoholic liquids predicted correctly, it can be seen that the results were different. The average number of non-alcoholic liquids that was predicted correctly for different *k* values was 15.6 when Euclidean distance metric was preferred, was 16.5 when Chebyshev distance metric was preferred, was 15.2 when Manhattan distance metric was preferred, and finally it was 15.6 when Minkowski distance metric was preferred.

Performance metrics obtained from classifications using different *k* values and different distance metrics are given in Table 1. When *k* was set to 1, all the performance metrics reached the maximum value and RMS was at the minimum value. With the increase in *k* value, there was a decrease in the performance metrics. When *k* was set to 7, the highest decrease was seen in Precision and Recall values when using Chebyshev distance metric. In the classifications made, an increase in RMS was observed with the increase of the *k* value.

For Chebyshev distance metric, RMS became 0.09 when *k* was set to 1 and it became 0.22 when *k* was set to 10. The lowest Kappa value was 0.65 and it was obtained when *k* was set to 3 for Euclidean, Manthattan and Minkowski distance metrics and when *k* was set to 7 for Chebyshev distance metric.

Since the number of alcoholic and non-alcoholic data used in this classification study was not equal, the dataset was an unbalanced dataset. For unbalanced datasets, the use of MCC values is recommended (Chicco & Jurman, 2020). When MCC values obtained for different *k* values were examined, the average MCC values were computed as 0.78 (for Euclidean), 0.78 (for Chebyshev), 0.76 (for Manhattan) and 0.78 (for Minkowski). This result shows that the distance metric most sensitive to the value of *k* was Manhattan.

Table 1. Performance metrics for different distance metrics

Distance metric	Metric	*k*=1	*k*=2	*k*=3	*k*=4	*k*=5	*k*=6	*k*=7	*k*=8	*k*=9	*k*=10
Euclidean	Precision	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	Recall	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	F-measure	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	MCC	1	0.74	0.66	0.77	0.79	0.72	0.74	0.77	0.81	0.81
	AUC	1	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
	KAPPA	1	0.73	0.65	0.77	0.78	0.72	0.74	0.77	0.80	0.80
	RMS	0.09	0.17	0.20	0.21	0.22	0.22	0.22	0.21	0.18	0.18
Chebyshev	Precision	1	0.94	0.93	0.96	0.94	0.91	0.89	0.91	0.93	0.91
	Recall	1	0.94	0.93	0.96	0.94	0.91	0.89	0.91	0.93	0.91
	F-measure	1	0.94	0.93	0.96	0.94	0.91	0.89	0.91	0.93	0.91
	MCC	1	0.81	0.77	0.87	0.82	0.72	0.65	0.71	0.79	0.72
	AUC	1	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.98	0.97
	KAPPA	1	0.80	0.77	0.87	0.82	0.72	0.65	0.71	0.78	0.72
	RMS	0.09	0.15	0.17	0.17	0.19	0.20	0.22	0.21	0.19	0.21
Manhattan	Precision	1	0.92	0.90	0.93	0.92	0.91	0.92	0.93	0.93	0.92
	Recall	1	0.92	0.90	0.93	0.92	0.91	0.92	0.93	0.93	0.92
	F-measure	1	0.92	0.90	0.93	0.92	0.91	0.92	0.93	0.93	0.92
	MCC	1	0.74	0.66	0.77	0.75	0.72	0.74	0.77	0.77	0.74
	AUC	1	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
	KAPPA	1	0.73	0.65	0.77	0.75	0.72	0.74	0.77	0.77	0.73
	RMS	0.09	0.17	0.20	0.20	0.22	0.22	0.22	0.21	0.19	0.19
Minkowski	Precision	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	Recall	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	F-measure	1	0.92	0.90	0.93	0.93	0.91	0.92	0.93	0.94	0.94
	MCC	1	0.74	0.66	0.77	0.79	0.72	0.74	0.77	0.81	0.81
	AUC	1	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98
	KAPPA	1	0.73	0.65	0.77	0.78	0.72	0.74	0.77	0.80	0.80
	RMS	0.09	0.17	0.20	0.21	0.22	0.22	0.22	0.21	0.18	0.18

4. Conclusion

The classification of liquids, which is important to manage the hazards of chemicals, is an interesting research topic in recent years. Therefore, various methods were proposed for liquid classification. As it is easier to implement than other algorithms, KNN algorithm was often preferred for classification problems. However, the use of KNN algorithm for classifying liquids that contain alcohol is still limited. In this study, the parameters affecting S_{11} parameter measurements were analyzed in order to characterize the liquid in the most appropriate way. In addition, the application of weighting in the performance of KNN algorithm for classification, the use of different number of nearest neighbors and different distance metrics was examined.

It was observed that the increase in the number of the nearest neighbors reduced the classification performance. Although generally reducing the value of the nearest neighbor number increases the algorithm's sensitivity to noisy data, the low number of nearest neighbors led to the better results due to the very low noise in the S_{11} parameters. Considering the effect of distance metrics, when the nearest neighbor number value was 1, all the distance metrics led to the same result.

After weighting was applied, all the liquids were classified correctly with 100% accuracy. By applying weighting, the performance of the classification can be made independent of distance metrics and *k* values. Therefore, it is recommended to apply weighting to KNN algorithm in the classification made with S_{11} data. Moreover, the results obtained in this study made it clear that in order to increase the sensitivity of the measurements, it is recommended to make liquid measurements in pet bottles with low reflective

properties and to take measurements by keeping the liquid as close to the antenna as possible.

References

Aydın, E. A., & Kaya Keleş, M. (2017). Breast cancer detection using K-nearest neighbors data mining method obtained from the bow-tie antenna dataset. *International Journal of RF and Microwave Computer-Aided Engineering, 27*(6), e21098.

Bhatia, N. (2010). Survey of nearest neighbor techniques. *arXiv preprint arXiv:1007.0085*.

Borisov, V., & Karpenko, A. (2001). Using of the Michelson microwave interferometer for the measurement of permittivity of thin-layer materials. *Russian journal of nondestructive testing, 37*(8), 597-599.

Chakrabarti, S., Cox, E., Frank, E., Güting, R. H., Han, J., Jiang, X., . . . Neapolitan, R. E. (2008). *Data mining: know it all*: Morgan Kaufmann.

Chen, H., Hu, Z., Wang, P., Xu, W., & Hou, Y. (2020). *Application of spectral droplet analysis method in flammable liquids identification.* Paper presented at the 2019 International Conference on Optical Instruments and Technology: Optical Sensors and Applications.

Chen, Q., Kang, G., Zhou, T., & Wang, J. (2017). *Fire hazard analysis of alcohol aqueous solution and Chinese liquor based on flash point.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Cheremisinoff, N. P. (1999). *Handbook of industrial toxicology and hazardous materials*: CRC Press.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*, 6.

Doad, P., & Bartere, M. (2013). A Review: Study of Various Clustering Techniques. *International Journal of Engineering Research & Technology, 2*(11), 3141-3145.

Hayasaka, T., Lin, A., Copa, V. C., Lopez, L. P., Loberternos, R. A., Ballesteros, L. I. M., . . . Lin, L. (2020). An electronic nose using a single graphene FET and machine learning for water, methanol, and ethanol. *Microsystems & Nanoengineering, 6*(1), 1-13.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence, 22*(1), 4-37.

Jawad, H., Lanteri, J., Migliaccio, C., Pichot, C., Platt, I. G., Tan, A. E.-C., . . . Woodhead, I. M. (2017). *Microwave modeling and experiments for non destructive control improved quality of fruit.* Paper presented at the 2017 IEEE Conference on Antenna Measurements & Applications (CAMA).

Jepsen, P. U., Jensen, J. K., & Møller, U. (2008). Characterization of aqueous alcohol solutions in bottles with THz reflection spectroscopy. *Optics express, 16*(13), 9318-9331.

Jiang, Y., Ju, Y., & Yang, L. (2016). Nondestructive in-situ permittivity measurement of liquid within a bottle using an open-ended microwave waveguide. *Journal of Nondestructive Evaluation, 35*(1), 7.

Jose, K., Varadan, V., & Varadan, V. (2001). Wideband and noncontact characterization of the complex permittivity of liquids. *Microwave and Optical Technology Letters, 30*(2), 75-79.

Kresse, W., & Danko, D. M. (2012). *Springer handbook of geographic information*: Springer Science & Business Media.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4): John Wiley & Sons.

Li, Z., Haigh, A., Soutis, C., Gibson, A., & Sloan, R. (2017a). Evaluation of water content in honey using microwave transmission line technique. *Journal of Food Engineering, 215*, 113-125.

Li, Z., Haigh, A., Soutis, C., Gibson, A., & Sloan, R. (2017b). Microwaves sensor for wind turbine blade inspection. *Applied Composite Materials, 24*(2), 495-512.

Li, Z., Haigh, A., Soutis, C., Gibson, A., & Sloan, R. (2018). A simulation-assisted non-destructive approach for permittivity measurement using an open-ended microwave Waveguide. *Journal of Nondestructive Evaluation, 37*(3), 1-10.

Moghadas, H., & Mushahwar, V. K. (2018). Passive microwave resonant sensor for detection of deep tissue injuries. *Sensors and Actuators B: Chemical, 277*, 69-77.

Orachorn, P., Chankow, N., & Srisatit, S. (2019). An Alternative Method for Screening Liquid in Bottles at Airports Using Low Energy X-ray Transmission Technique. *Radiation environment and medicine: covering a broad scope of topics relevant to environmental and medical radiation research, 8*(2), 77-84.

Rey, T., Kordon, A., & Wells, C. (2012). *Applied data mining for forecasting using SAS*: SAS Institute.

Saçlı, B., Aydınalp, C., Cansız, G., Joof, S., Yilmaz, T., Çayören, M., . . . Akduman, I. (2019). Microwave dielectric property based classification of renal calculi: Application of a kNN algorithm. *Computers in biology and medicine, 112*, 103366.

Slaughter, R., Mason, R., Beasley, D., Vale, J., & Schep, L. (2014). Isopropanol poisoning. *Clinical toxicology, 52*(5), 470-478.

Tan, X., Huang, S., Zhong, Y., Yuan, H., Zhou, Y., Xiao, Q., . . . Qi, C. (2017). *Detection and identification of flammable and explosive liquids using THz time-domain spectroscopy with principal component analysis algorithm.* Paper presented at the 2017 10th UK-Europe-China Workshop on Millimetre Waves and Terahertz Technologies (UCMMT).

Venkatesh, M., & Raghavan, G. (2005). An overview of dielectric properties measuring techniques. *Canadian biosystems engineering, 47*(7), 15-30.

Wirasuta, I. M. A. G., Dewi, N. K. S. M., Purwaningsih, N. K. P. A., Heltyani, W. E., Aryani, N. L. P. I., Sari, N. M. K., . . . Ramona, Y. (2019). A rapid method for screening and determination test of methanol content in ethanol-based products using portable

Raman spectroscopy. *Forensic Chemistry*, *16*, 100190.

Yurchenko, A. V., Novikov, A., & Kitaeva, M. V. (2012). A resonator microwave sensor for measuring the parameters of Solar-quality silicon. *Russian journal of nondestructive testing*, *48*(2), 109-114.

Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 4(2) (2021) 168-176
DOI: 10.38016/jista.909178

Elektrokardiyografi Sinyali ile Bazal Metabolizma Hızının Cinsiyet Bazlı Yapay Zekâ Tabanlı Tespiti

Ahmet Enes Yıldırım[1,a], Ömer Faruk Kadıoğlu[1,b], Hamza Kavak[1,c], Kadir Salman[1,d], Muhammed Kürşat Uçar[1,e*]

Zeliha Uçar[2], Mehmet Recep Bozkurt[1,f]

[1]Sakarya Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Sakarya, Türkiye

[a]yildirim_1132@hotmail.com, [b]omerfaruk_kadioglu@hotmail.com, [c]hamza.kavak@ogr.sakarya.edu.tr, [d]kadir.salman@ogr.sakarya.edu.tr, [e]mucar@sakarya.edu.tr, [f]mbozkurt@sakarya.edu.tr

[2]İstanbul Okan Üniversitesi, Sağlık Bilimleri Enstitüsü, Beslenme ve Diyetetik Bölümü, İstanbul, Türkiye

zelihaguvenc@hotmail.com

Öz

Bazal Metabolizma Hızı (BMH) günlük harcanan ve alınması gereken enerji hakkında bilinmesi gereken en önemli unsurlardan biridir. Literatürde genellikle kalorimetreler ve birtakım denklemler tarafından tespit edilmektedir. Bu çalışmada BMH tahmini için elektrokardiyografi (ECG) sinyalleri ile yapay zekâ tabanlı bir model oluşturulmuştur. Öncelikle bireylerden toplanan ECG sinyalleri gürültülerden temizlenip filtrelenmiştir. Daha sonra özellik çıkartılıp özellik seçme algoritmaları yardımıyla azaltılmıştır. Elde kalan özelliklerle yapay zekâ algoritmaları sayesinde BMH tahmininde bulunulmuştur. Erkekler için R = 0.91, kadınlar için R = 0.99 değerlerine sahip modeller oluşturulmuştur. Performans değerlendirme kriterleri de göz önüne alınarak en iyi model kadınlar için de erkekler için de Linear Regression modeli seçilmiştir. Tüm bu sonuçlara bakıldığında günlük hayatta BMH tahmini için önerilen modelin kullanılabileceği belirlenmiştir.

Anahtar kelimeler: Yapay Zekâ, Bazal Metabolizma Hızı, Elektrokardiyografi Sinyali

Gender-Based Artificial Intelligence Based Detection of Basal Metabolic Rate by Electrocardiography Signal

Abstract

Basal Metabolic Rate (BMR) is one of the most important factors that should be known about the energy consumed and taken daily. It is usually determined by calorimeters and some equations in the literature. In this study, an artificial intelligence-based model was created with electrocardiography (ECG) signals for BMR prediction. First of all, ECG signals collected from individuals are cleaned from noise and filtered. Later, the feature was removed and reduced with the help of feature selection algorithms. BMR predictions have been made with the remaining features thanks to artificial intelligence algorithms. Models with R = 0.91 for men and R = 0.99 for women were created. Considering the performance evaluation criteria, the Linear Regression model was chosen as the best model for both women and men. Considering all these results, it was determined that the proposed model could be used for BMR estimation in daily life.

Keywords: Artificial Intelligence, Basal Metabolic Rate, Electrocardiography Signal

* Sorumlu yazar.
E-posta adresi: mucar@sakarya.edu.tr

Alındı : 03 Nisan 2021
Kabul : 21 Ağustos 2021

1. Giriş (Introduction)

Bazal Metabolizma Hızı (BMH) genellikle yaşamı sürdürmek için gerekli temel süreçlerle uyumlu minimum metabolizma hızı olarak kabul edilir (Speakman, Król, and Johnson 2004). Günlük harcanan ve alınması gereken enerji hakkında bilinmesi gereken en önemli parametrelerden biridir. Ancak günlük hayatta ölçümünü yapmak oldukça zahmetli ve maliyetli bir iştir. Literatürde kullanılan en yaygın BMH ölçüm yöntemi dolaylı kalorimetredir. Dolaylı kalorimetre, oksijen tüketimi ve karbondioksit üretimi ölçümlerinden bazal metabolizma hızını tespit eder (Ferrannini 1988).

Bir diğer BMH tespiti için kullanılan yöntem de BMH denklemleridir (Müller et al. 2001). Bu denklemlerin BMH tespitinde kullanılması oldukça pratiktir. Ancak yapılan araştırmalar gösteriyor ki denklemlerin sonucunda elde edilen değerler, gerçek BMH değerlerinden oldukça farklıdır (Flancbaum et al. 1999). Tüm bu durumlar göz önüne alındığında BMH tespiti için yeni yöntemlere ihtiyaç duyulduğu gözlemlenmiştir.

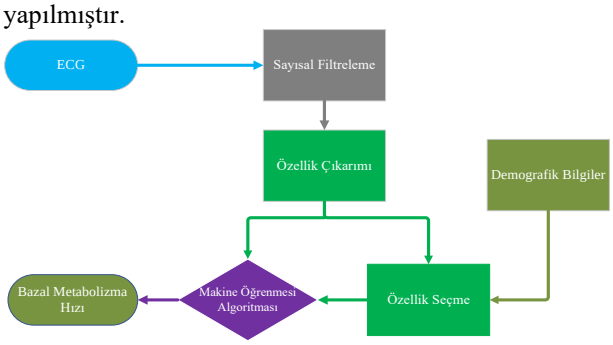
Bu çalışmanın amacı maliyetli ve ölçümü zor olan yöntemler yerine yapay zekâ tabanlı güvenilir yeni bir yöntem geliştirmektir. Bu çalışmada, literatürdeki yöntemlerin aksine kalorimetreler yerine elektrokardiyografi (ECG) sinyalleri kullanarak BMH tahmini yapmaktır. ECG sinyali kalbin elektriksel faaliyeti sonucu oluşan ve deriden elektrotla ölçülebilen biyopotansiyel sinyallerdir (Nur Göz et al. 2017). Bu çalışmada 324 farklı kişiden ECG sinyalleri alınarak 3 farklı filtreleme sonucunda gürültüleri en aza indirilmiş bir sinyal elde edilmiştir. Bu sinyalden çıkarılan 23 özelliğe demografik bilgiler de eklenerek toplamda 27 özellik elde edilmiştir. Spearman özellik seçme algoritması yardımıyla makine öğrenmesinde kullanılacak özellikler belirlenmiştir. Makine öğrenmesi algoritmaları kullanılarak seçilen özellikler yardımıyla BMH tahmini yapılmıştır ve performans değerlendirme kriterleri göz önüne alınarak en iyi sonuç veren model belirlenmiştir. Bütün süreç kadın ve erkek için ayrı ayrı tekrar edilmiştir ve cinsiyet bazlı modeller geliştirilmiştir.

Bu çalışma literatürle kıyaslandığında birçok yenilik içermektedir. (1) Literatürde BMH hesaplamaları için denklemler kullanılırken, bu çalışmada yapay zekâ tabanlı algoritmalar kullanılmıştır. (2) Ölçümlerde zahmetli ve maliyetli kalorimetreler kullanılırken bu çalışmada ECG sinyalleri kullanılmıştır. (3) Özellik seçme algoritmaları kullanılarak modelin doğruluk oranları arttırılmıştır ve güvenilirliği yüksek bir tahmin modeli oluşturulmuştur. Tüm bu çalışmalar göz önüne alındığında literatüre yenilik katmaktadır.

Çalışmada kadın/erkek/kadın-erkek olmak üzere 3 farklı model geliştirilmiştir. Erkekler için geliştirilen model R = 0.91, kadınlar için R = 0.99, kadın-erkek için R = 0.87 değerlerine sahiptir. Sonuçlar göz önüne alındığında cinsiyet bazlı oluşturulan modellerin doğruluk oranlarının daha yüksek olduğu gözlemlenmiştir.

2. Materyal ve Yöntem (Material and Method)

Araştırmada izlenen yol şu şekilde açıklanabilir (Şekil 1). İlk olarak bireylerden alınan ECG sinyalleri filtrelenmiştir ve bu filtrelenmiş sinyallerden özellik çıkarımı yapılmıştır. Daha sonra bu özelliklere demografik bilgiler eklenmiştir. Bir sonraki adımda özellik seçme algoritması kullanılarak özellik sayısı azaltılmıştır. Son olarak bu özelliklerden makine öğrenmesi algoritmaları sayesinde BMH tahmini yapılmıştır.



Şekil 1. Çalışma akış diyagramı (work flow diagram)

2.1. Verilerin toplanması (Data collection)

Bu çalışmada kullanılan veriler Sakarya Üniversitesi Sağlıklı Beslenme ve Danışmanlık Birimi'ne beslenme ve danışmanlık hizmeti için Eylül 2019 – Şubat 2020 tarihleri arasında başvuru yapan gönüllülerden alınmıştır. Çalışmanın yapılabilmesi için gerekli etik kurul onay ve veri kullanım izni alınmıştır.

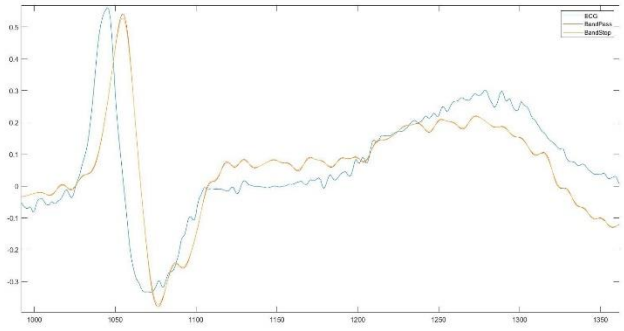
Çalışmada 324 bireyden alınan ECG sinyali ve BMH verileri toplanmıştır. Bu veriler ve sinyaller kullanılarak BMH'nın yüksek doğruluk oranında tahmin edilmesi amaçlanmaktadır.

Cinsiyete göre demografik bilgiler ve BMH dağılımları farklılık gösterdiği için BMH'nın cinsiyet bazlı hesaplanması gerektiği anlaşılmıştır. Bu sebeple, BMH'nın hesaplanması hem kadın ve erkek için ayrı ayrı, hem de tüm veriler kullanılarak yapılmıştır.

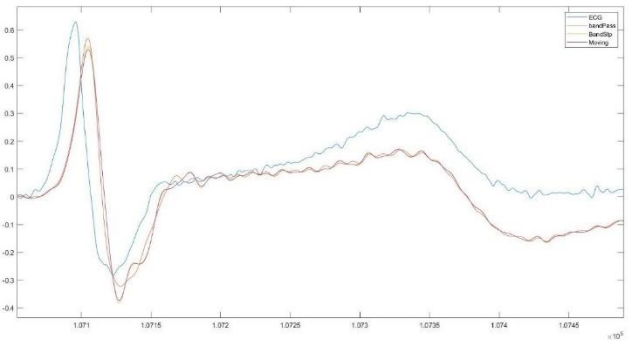
2.2. Sinyal önişleme (Signal preprocessing)

ECG sinyali üzerinde oluşan gürültülerin temizlenmesi amacıyla sayısal filtre tasarlanmış ve uygulanmıştır. Gürültü temizleme işlemi için 3 farklı filtre çeşidi kullanılmıştır. İlk önce IIR Chebyshev Type 2 Band Geçiren Filtre (0.25-100 Hz) kullanılmıştır (Şekil 2). Ardından IIR Chebyshev Type 2 Band Durduran Filtre (49-51 Hz) kullanılmıştır (Şekil 2). En son olarak ise gürültülerin en aza indirilmesi için Moving Filtresi uygulanarak gürültüler en aza indirilmiştir (Şekil 3).

Toplam 324 adet ECG verisini filtreleyerek bir sonraki adımda kullanmak için hazırlanmıştır.



Şekil 2. Moving filtresi henüz uygulanmamış sinyal (Moving filter not yet applied signal)



Şekil 3. Moving filtresi uygulanmış sinyal (Moving filter applied signal)

2.3. Özellik çıkarımı (Feature extraction)

Filtreleme işleminden sonra en iyi filtrelenmiş sinyaller dosyasını seçilmiştir. Seçilen dosyada her bir sinyal bir kişiye ait olan 324 sinyalin özelliklerini formüllerine göre hesaplayarak sinyallerin özellikleri çıkarılmıştır. Bu özelliklere ek olarak yapay zekanın bazal metabolizma hızını daha sağlıklı hesaplayabilmesi için bu özelliklere ek olarak kişilerin yaşı, boyu, kilosu gibi özellikleri de ekleyerek toplamda 23'ü formül hesaplamasıyla bulunan (Çizelge 1) sinyal özellikleri 4 tanesi kişisel özellik olmak üzere toplamda 27 özellik çıkarılarak özellik çıkarma işlemi sonlandırılmıştır.

2.4. Spearman özellik seçme algoritması (Spearman feature selection algorithm)

Orijinal veri setini temsil edebilecek en iyi altkümenin seçimi olarak tanımlanmaktadır. Özellik seçimi (diğer adıyla nitelik seçimi veya değişken seçimi), kullanılan algoritmaya göre özellikleri değerlendirerek veri setindeki n adet özellik arasından en iyi k adet özelliği seçme işlemidir (Forman 2003). Özellik seçimi, ilgilenilen problem için en faydalı ve en önemli özellikleri seçerek veri kümesindeki özellik sayısının azaltılmasını amaçlamaktadır. Bu çalışmada Spearman korelasyon katsayısı tabanlı özellik seçme algoritması kullanılmıştır.

Tablo 1. Sinyallerden çıkarılmış özellikler ve formülleri (Properties and formulas extracted from signals)

No	Özellik	Formül
1	Basıklık	$\frac{\sum_{i=1}^{n}(x(i)-\bar{x})^4}{(n-1)S^4}$
2	Çarpıklık	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{(n-1)S^3}$
3	IQR	$IQR = iqr(x)$
4	DK	$DK = (S/\bar{x})100$
5	Geometrik Ortalama	$G = \sqrt[n]{x1 + \cdots + xn}$
6	Harmonik Ortalama	$H = n/(\frac{1}{x_1} + \cdots + \frac{1}{x_n})$
7	Activity- Hjort Parameters	$A = S^2$
8	Maksimum	$x_{max} = max(x_i)$
10	Ortalama Mutlak Sapma	$MAD = mad(x)$
11	Minimum	$x_{min} = min(x_i)$
12	Merkez Anlar	$CM = moment(x, 10)$
13	Ortalama	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} = \frac{1}{n}(x_1 + \cdots \cdot + x_n)$
14	Ortalama Eğri Uzunluğu	$\mathrm{CL} = \frac{1}{n}\sum_{i=2}^{n} \lvert x_i - x_{i-1} \rvert$
15	Ortalama Enerji	$E = \frac{1}{n}\sum_{i=1}^{n} x_i^2$
16	RMS-(Karekök Ortalama)	$X_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert^2}$
17	Standart Hata	$S\bar{x} = S/\sqrt{n}$
18	Standart Sapma	$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})}$
19	Şekil Faktörü	$SF = X_{rms}/\ (\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i \rvert})$
20	Tekil Değer Ayrışımı	$SVD = svd(x)$
21	%25 Kesilmiş Ortalama	$T25 = trimmean(x, 25)$
22	%50 Kesilmiş Ortalama	$T50 = trimmean(x, 50)$
23	Ortalama Teager Enerjisi	$TE = \frac{1}{n}\sum_{i=3}^{n}(x_{i-1}^2 - x_i\, x_{i-2})$

İki sayısal ölçüm arasında doğrusal bir ilişki olup olmadığını, varsa bu ilişkinin yönünü ve şiddetinin ne olduğunu belirlemek için kullanılan bir istatistiksel yöntemdir. Bu sebeple çalışmamızda veri dağılımları normal dağılımdan uzak olduğundan Spearman

Korelasyonu kullanılmaktadır. Yönteme ait denklemler şu şekilde özetlenebilir:
Çift Yönlü Test
$H0$: ρs=0
$H1$: $\rho s \neq 0$

Test İstatistiği
$r_s = 1 - \frac{6\Sigma\, di^2}{n(n^2-1)}$ (1)

di=ui−vi
ui: 1. Örneklemdeki i. ölçümün sıra numarası
vi: 2. Örneklemdeki i. ölçümün sıra numarası
Ret Bölgesi
rs>rs, a/2 veya rs<−rs, a/2 ise $H0$ reddedilir.

Denklemde elde edilen r_s ilişki düzeni gösterir ve özellik seçiminde temel parametre olarak kullanılır. Her özelik için r_s değeri hesaplandıktan sonra en iyi %20 özellik seçilerek sistem tekrar modellenmiştir.

2.5. Makine öğrenmesi (Machine learning)

Makine öğrenmesi algoritmaları, kullandığı tekniklere göre denetimli ve denetimsiz öğrenme olmak üzere iki gruba ayrılırlar. Denetimsiz öğrenme, sadece giriş verilerinin olduğu bir öğrenme yöntemidir. Genellikle etiketsiz veriler üzerinde çalışılırken tercih edilen bir uygulamadır. Denetimli öğrenme ise giriş ve çıkış verilerinin ikisini de barındırmaktadır ve etiketlenmiş veriler kullanılarak öğrenme sağlanmaktadır. Giriş ve çıkış verileri bilindiği için öğrenme süreci denetlemeye tabi tutulur. Bu yüzden denetimli öğrenme olarak adlandırılmaktadır. Denetimli öğrenme, sınıflandırma ve regresyon olarak gruplandırılabilir. Regresyon, bir bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi belirlemeye çalışan, nicel değişkenleri tahmin etmek için kullanılan bir denetimli öğrenmedir.Sınıflandırma ise tahmin edilen veriyi kategorilere atama ile ilgili bir denetimli öğrenmedir (Akay 2018).

Bu çalışmada kullanılan regresyon modelleri Lineer Regresyon, Gauss Süreç Regresyon, Destek Vektör Makineleri, Karar Ağaçları, Ensemble'dir. Bu sonuçlar çeşitli performans değerlendirme kriterleri göz ölüne alınarak karşılaştırılmıştır. Karşılaştırma sonucunda en iyi sonucu veren model belirlenmiştir.

Verilerin %80'i eğitim, %20'si test sürecinde kullanılmıştır.

2.6. Performans değerlendirme kriterleri (Performance evaluation criteria)

Çalışmada kullanılan performans değerlendirme kriteri olarak R^2, MSE, RMSE ve MAE kullanılmıştır. Bu performans kriterlerinden R^2, modelin açıklayıcılık katsayısıdır. Bu katsayı, modelin tahmin ilişkisiyle doğru orantılıdır.

MSE, RMSE ve MAE birer hata ölçüleri olmaları sebebiyle modelin performansıyla ters orantılıdır. Yani düşük değerler, yüksek performansı göstermektedir (Wang and Xu 2004).

2.6.1. Ortalama karesel hata (Mean squared error – MSE)

MSE, modelin hatalarının karelerinin ortalamasını ifade eder. Tahmin edilen değerler gerçek değerlere ne kadar yaklaşırsa MSE değeri de o kadar küçülür.

e_i, gerçek değer (y_i) ile tahmin edilen değer ($\hat{y}_i$) arasındaki farktır.

$$e_i = y_i - \hat{y}_i \quad (2)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} {e_i}^2 \quad (3)$$

2.6.2. Ortalama karesel hata karekökü (Root Mean Squared Error – RMSE)

RMSE, tahmin hatalarının ne kadar yayıldığının ölçüsüdür. MSE değerinin kareköküdür.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {e_i}^2} \quad (4)$$

2.6.3. Ortalama mutlak hata (Mean absolute error – MAE)

MAE, veri kümesi içindeki her bir örnek için gerçek değer ile tahmin edilen değer arasındaki farkın mutlak değerinin ortalamasıdır.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |e_i| \quad (5)$$

2.6.4. Açıklayıcılık katsayısı (Coefficient of determination)

R^2, bağımlı değişkendeki değişimin ne kadarının bağımsız değişken tarafından açıklanabildiğini gösteren bir katsayıdır. 1'e yakın olması beklenir.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

3. Sonuçlar (Conclusions)

Bu çalışmada amaç ECG sinyali ile yapay zekayı kullanarak bireylere ait BMH değerini yüksek doğruluk oranında tahmin etmektir. Çalışmada izlenen yol şu şekildedir: İlk olarak bireylerden toplanan ECG sinyalleri filtrelenmiştir. Bu filtrelenmiş sinyallerden 23 adet özellik çıkarılmıştır ve demografik bilgiler eklenmiştir. Daha sonra kullanılacak özellikler özellik seçme algoritmaları yardımıyla belirlenmiştir.

Belirlenen özellikler kullanılarak BMH tahmini için makine öğrenmesi algoritmaları kullanılmıştır. Son olarak modellerin karşılaştırılması için performans değerlendirme kriterleri kullanılmıştır.

Çalışmada, filtrelenmiş ECG sinyalinden özellikler çıkarılarak BMH tahmin edilmiştir. Bu tahminler ayrıca cinsiyet bazlı da yapılmıştır. Tahmin sonuçlarına bakıldığında, cinsiyet bazlı yapılmış tahminlerin daha iyi sonuç verdiği belirlenmiştir. Bütün veriler kullanıldığında (kadın-erkek) R = 0.86 bulunmuştur (Çizelge 4). Cinsiyet bazlı yapılan tahmin sonucunda erkek için R = 0.85 (Çizelge 2), kadınlarda ise R = 0.98 bulunmuştur (Çizelge 3). Buradaki sonuçlar göz önüne alındığında BMH tahminin cinsiyet bazlı tahminlerde daha iyi sonuçlar verdiği belirlenmiştir.

Sinyalden elde edilen bütün özellikler Spearman Özellik Seçme Algoritması kullanılarak azaltılmış ve BMH tahmin modeli hem cinsiyet bazlı hem de bütün veriler kullanılarak yeniden oluşturulmuştur. Cinsiyet gözetmeden yapılan tahmin modelinde R = 0.87 başarı oranı elde edilirken (Çizelge 7), cinsiyet bazlı tahmin modelinde erkekler için R = 0.91 (Çizelge 5), kadınlar için R = 0.99 bulunmuştur (Çizelge 6). Özellik seçimi öncesi bulunan R değerleri ile özellik seçimi sonrası elde edilen R değerleri karşılaştırıldığında, özellik seçme algoritmalarının elde edilen başarı oranını arttırdığı belirlenmiştir.

Tablo 2. Erkekler için tüm özellikler kullanılarak yapılmış tahmin sonuçları (Estimation results using all features for men)

Tüm Özellikler (Erkek)	RMSE	R^2	MSE	MAE	R
Linear Regression (Interactions Linear)	-	-	-	-	-
Linear Regression (Linear)	82,88	0,73	6869	27,844	0,854
Linear Regression (Robust Linear)	86,163	0,71	7424	21,133	0,842
SVM (Linear SVM)	86,099	0,71	7413,1	29,29	0,842
Gaussian Process Regression (Squared Exponential GPR)	87,775	0,69	7704,4	32,039	0,83
Gaussian Process Regression (Matern 5/2 GPR)	88,196	0,69	7778,5	32,41	0,83
Gaussian Process Regression (Rational Quadratic GPR)	87,775	0,69	7704,4	32,039	0,83
Stepwise Linear Regression	90,349	0,68	8162,9	36,431	0,824
Gaussian Process Regression (Exponential GPR)	89,297	0,68	7973,9	37,292	0,824
SVM (Quadratic SVM)	93,936	0,65	8824	40,095	0,806
SVM (Medium Gaussian SVM)	96,34	0,63	9281,4	52,088	0,793
SVM (Cubic SVM)	100,06	0,6	10011	43,383	0,774
SVM (Coarse Gaussian SVM)	109,41	0,52	11971	64,112	0,721
Ensemble (Boosted Trees)	111,81	0,5	12502	79,813	0,707
Tree (Medium Tree)	113,18	0,49	12810	48,804	0,7
Tree (Coarse Tree)	118,43	0,44	14025	59,449	0,663
Tree (Fine Tree)	120,51	0,42	14523	53,442	0,648
Ensemble (Bagged Trees)	139,85	0,22	19558	113,52	0,469
SVM (Fine Gaussian SVM)	151,17	0,09	22851	115,89	0,3

Tablo 3. Kadınlar için tüm özellikler kullanılarak yapılmış tahmin sonuçları (Estimation results using all features for women)

Tüm Özellikler (Kadın)	RMSE	R^2	MSE	MAE	R
Linear Regression (Interactions Linear)	-	-	-	-	-
Gaussian Process Regression (Squared Exponential GPR)	25,977	0,98	674,8	21,297	0,989
Gaussian Process Regression (Rational Quadratic GPR)	25,977	0,98	674,8	21,297	0,989
SVM (Linear SVM)	29,69	0,97	881,52	21,266	0,984
Gaussian Process Regression (Matern 5/2 GPR)	26,949	0,97	726,27	22,216	0,984
Linear Regression (Robust Linear)	34,975	0,96	1223,3	23,162	0,979
Linear Regression (Linear)	39,74	0,95	1579,3	31,505	0,974
Stepwise Linear Regression	37,43	0,95	1401	29,285	0,974
SVM (Quadratic SVM)	45,931	0,93	2109,6	35,588	0,964
Gaussian Process Regression (Exponential GPR)	53,074	0,9	2816,8	38,396	0,948
SVM (Medium Gaussian SVM)	83,709	0,76	7007,2	55,0731	0,871
Ensemble (Boosted Trees)	87,375	0,74	7634,5	69,402	0,86
Tree (Fine Tree)	96,378	0,68	9288,8	61,679	0,824
SVM (Coarse Gaussian SVM)	97,32	0,67	9471,2	75,636	0,818
SVM (Cubic SVM)	107,66	0,6	11590	67,981	0,774
Ensemble (Bagged Trees)	119,73	0,5	14336,7	82,219	0,707
Tree (Medium Tree)	133,39	0,38	17794	86,503	0,616
Tree (Coarse Tree)	133,18	0,38	17736	94,504	0,616
SVM (Fine Gaussian SVM)	155,64	0,16	24224	113,71	0,4

Tablo 4. Kadın – erkek için tüm özellikler kullanılarak yapılmış tahmin sonuçları (Estimation results using all features for men and women)

Tüm Özellikler (Kadın-Erkek)	RMSE	R^2	MSE	MAE	R
Linear Regression (Interactions Linear)	-	-	-	-	-
Linear Regression (Linear)	84,934	0,74	7213,8	63,327	0,86
Gaussian Process Regression (Squared Exponential GPR)	84,774	0,74	7186,6	61,157	0,86
Gaussian Process Regression (Matern 5/2 GPR)	84,585	0,74	7154,7	61,359	0,86
Gaussian Process Regression (Exponential GPR)	85,275	0,74	7271,9	62,132	0,86
Gaussian Process Regression (Rational Quadratic GPR)	84,774	0,74	7186,6	61,157	0,86
Linear Regression (Robust Linear)	87,334	0,72	7627,2	65,576	0,848
Stepwise Linear Regression	87,312	0,72	7623,4	67,402	0,848
SVM (Linear SVM)	89,967	0,71	8094	64,769	0,842
SVM (Quadratic SVM)	89,779	0,71	8060,3	67,073	0,842
Tree (Fine Tree)	94,131	0,68	8860,7	68,348	0,824
SVM (Medium Gaussian SVM)	100,17	0,64	10034	76,594	0,8
Tree (Medium Tree)	101,17	0,63	10234	74,288	0,793
SVM (Coarse Gaussian SVM)	106,13	0,59	11264	75,166	0,768
Ensemble (Bagged Trees)	108,23	0,58	11713	79,183	0,761
Tree (Coarse Tree)	115,19	0,52	13269	82,01	0,721
SVM (Cubic SVM)	118,92	0,49	14141	89,597	0,7
Ensemble (Boosted Trees)	120,28	0,48	14468	92,828	0,692
SVM (Fine Gaussian SVM)	158,77	0,09	25207	119,8	0,3

Tablo 5. Erkekler için seçilmiş özellikler kullanılarak yapılmış tahmin sonuçları (Estimation results using selected features for men)

Özellik Seçimi Sonrası (Erkek)	RMSE	R^2	MSE	MAE	R
SVM (Cubic SVM)	-	-	-	-	-
Linear Regression (Robust Linear)	41,755	0,84	1743,5	19,106	0,916
SVM (Linear SVM)	41,426	0,84	1716,1	18,553	0,916
SVM (Quadratic SVM)	42,868	0,83	1837,7	21,616	0,911
SVM (Coarse Gaussian SVM)	47,944	0,79	2298,7	26,141	0,888
Gaussian Process Regression (Exponential GPR)	47,718	0,79	2277	27,251	0,888
SVM (Medium Gaussian SVM)	49,063	0,78	2407,1	28,219	0,883
Gaussian Process Regression (Rational Quadratic GPR)	49,964	0,77	2496,4	28,405	0,877
Tree (Medium Tree)	51,241	0,76	2625,7	33,965	0,871
Gaussian Process Regression (Matern 5/2 GPR)	51,174	0,76	2618,8	30,306	0,871
Linear Regression (Linear)	52,191	0,75	2723,9	33,058	0,866
Linear Regression (Interactions Linear)	52,516	0,74	2758	32,105	0,86
Stepwise Linear Regression	52,517	0,74	2758	32,969	0,86
Gaussian Process Regression (Squared Exponential GPR)	53,927	0,73	2908,1	32,678	0,854
Tree (Fine Tree)	69,971	0,55	4854	44,723	0,741
Ensemble (Bagged Trees)	70,754	0,53	5006,1	47,88	0,728
Tree (Coarse Tree)	86,786	0,3	7531,7	69,151	0,547
SVM (Fine Gaussian SVM)	86,482	0,3	7479,2	59,048	0,547
Ensemble (Boosted Trees)	103,22	0,01	10655	86,678	0,1

Tablo 6. Kadınlar için seçilmiş özellikler kullanılarak yapılmış tahmin sonuçları (Estimation results using selected features for women)

Özellik Seçimi Sonrası (Kadın)	RMSE	R^2	MSE	MAE	R
Linear Regression (Interactions Linear)	-	-	-	-	-
Linear Regression (Robust Linear)	19,025	0,99	361,97	12,796	0,994
SVM (Linear SVM)	17,309	0,99	299,6	11,796	0,994
Linear Regression (Linear)	26,185	0,98	685,66	22,097	0,989
Gaussian Process Regression (Squared Exponential GPR)	33,432	0,96	1117,7	26,242	0,979
Gaussian Process Regression (Matern 5/2 GPR)	33,448	0,96	1118,7	26,077	0,979
Gaussian Process Regression (Rational Quadratic GPR)	33,432	0,96	1117,7	26,242	0,979
SVM (Quadratic SVM)	38,096	0,95	1451,3	28,163	0,974
Stepwise Linear Regression	42,587	0,94	1813,6	29,613	0,969
Gaussian Process Regression (Exponential GPR)	47,351	0,93	2242,1	33,013	0,964
SVM (Medium Gaussian SVM)	71,894	0,83	5168,7	53,372	0,911
SVM (Cubic SVM)	94,385	0,71	8908,6	58,26	0,842
Tree (Coarse Tree)	97,264	0,7	9460,3	72,294	0,836
Ensemble (Boosted Trees)	99,623	0,68	9924,8	79,491	0,824
Tree (Medium Tree)	112,68	0,59	12697	81,429	0,768
SVM (Coarse Gaussian SVM)	112,52	0,59	12662	89,811	0,768
Tree (Fine Tree)	113,55	0,58	12893	76,206	0,761
Ensemble (Bagged Trees)	118,06	0,55	13939	85,497	0,741
SVM (Fine Gaussian SVM)	161,8	0,16	26180	124,83	0,4

Tablo 7. Kadın- erkek için seçilmiş özellikler kullanılarak yapılmış tahmin sonuçları (Estimation results using selected features for men and women)

Özellik Seçimi Sonrası (Kadın-Erkek)	RMSE	R^2	MSE	MAE	R
Linear Regression (Interactions Linear)	-	-	-	-	-
Linear Regression (Linear)	86,275	0,77	7443,4	63,878	0,877
Linear Regression (Robust Linear)	85,866	0,77	7373	63,638	0,877
SVM (Linear SVM)	85,128	0,77	7246,7	63,016	0,877
Gaussian Process Regression (Squared Exponential GPR)	87,952	0,76	7735,6	65,767	0,871
Gaussian Process Regression (Matern 5/2 GPR)	87,647	0,76	7682,1	65,916	0,871
Gaussian Process Regression (Rational Quadratic GPR)	87,953	0,76	7735,6	65,767	0,871
Stepwise Linear Regression	88,807	0,75	7886,7	67,225	0,866
Gaussian Process Regression (Exponential GPR)	91,971	0,73	8458,7	65,723	0,854
SVM (Coarse Gaussian SVM)	94,226	0,72	8878,6	68,319	0,848
SVM (Quadratic SVM)	99,105	0,69	9821,8	71,61	0,83
Ensemble (Bagged Trees)	107,56	0,64	11570	79,021	0,8
Tree (Fine Tree)	108,3	0,63	11729	80,284	0,793
Tree (Medium Tree)	107,74	0,63	11607	76,41	0,793
SVM (Medium Gaussian SVM)	112,71	0,6	12704	74,559	0,774
Ensemble (Boosted Trees)	115,53	0,58	13347	87,643	0,761
Tree (Coarse Tree)	120,82	0,54	14598	84,6	0,734
SVM (Fine Gaussian SVM)	157,97	0,21	24955	105,4	0,458
SVM (Cubic SVM)	169,95	0,09	28882	94,595	0,3

Elde edilen sonuçlar göz önüne alındığında en iyi modeli belirlemek için performans değerlendirme kriterlerinden RMSE, MSE ve MAE değerlerinin 0'a yakın olması istenirken R ve R2 değerlerinin 1'e yakın olması istenmektedir. Özellik seçimi sonrası bulunan değerlere bakıldığında en iyi modelin Linear Regression modeli olduğu anlaşılmaktadır (Çizelge 5-6-7). ECG sinyali kullanılarak bir çalışma yapıldığında Linear Regression modeli cinsiyet bazlı uygulanırsa performans olumlu bir şekilde etkilenecektir.

4. Tartışma ve Sonuç (Discussion and Conclusion)

Araştırmada bulunan BMH değerleri ECG sinyalinin işlenmesiyle elde edilmiştir. Bireylerden toplanan ECG verileri işlenerek makine öğrenmesi algoritmaları yardımıyla BMH değerlerine ulaşılmıştır.

BMH günlük alınması gereken kalori miktarı hakkında en önemli yol göstericilerden biridir. Günümüzde BMH tespiti için kullanılan yöntemler oldukça zahmetli ve maliyetli olduğu için alternatif yöntemler aranmaktadır.

Literatürde BMH tespiti için genellikle dolaylı kalorimetre kullanılmaktadır (Arslan 1984). Dolaylı kalorimetre, kişinin oksijen alıp vermesini inceleyerek BMH ölçümü yapmaktadır (LIU, LU, and CHEN 1995). Oluşturulan yeni tahmin modeli, literatürdeki yöntemle kıyaslandığında daha pratik ve maliyetsiz bir yöntem olduğu belirlenmiştir.

Dolaylı kalorimetre ile bulunan BMH değerlerinin doğru olabilmesi için deneklerin uyması gereken bazı koşullar vardır (12 saat aç kalınması, deney sırasında hareket edilmemesi, uyku düzeni vb.) (Tverskaya et al. 1998). Ancak oluşturulan yeni yöntemde herhangi bir koşul istenmemektedir. Bu da kişiden kaynaklı hataların minimuma indirilmesini sağlamaktadır. Bununla birlikte de yöntemin doğruluk oranı artmaktadır.

Sonuç olarak ECG sinyalleri ile BMH tahmini yapan bir sistem geliştirilmiş olup gündelik hayatta kullanılabileceği düşünülmektedir. Bu sistemin avantajları zahmetsiz ve maliyetsiz olması, kişi kaynaklı hataları ortadan kaldırması ve ispatlanmış doğruluk oranı sayesinde güvenilir sonuçlar elde eden bir sistem olmasıdır. Tüm bu değerlendirmeler dikkate alındığında literatüre katkı sağlayacağı düşünülmektedir.

Çıkar Çatışması

Yazar, bu makalenin yayınlanmasıyla ilgili herhangi bir çıkar çatışması olmadığını beyan eder.

Bilgilendirme

Finansal Destek

Bu çalışma Sakarya Üniversitesi Bilimsel Araştırma Projeleri Komisyonu Tarafından Desteklenmiştir. Proje Numarası: 2019-5-19-244.

Etik Kurul Onayı

Bu çalışmanın gerçekleştirilebilmesi için Sakarya Üniversitesi Tıp Fakültesi'nden eti kurulu onayı alınmıştır. Onay tarih ve numarası 03/05/2019 - 71522473/050.01.04/112.

Veri Kullanım İzni

Bu çalışmanın gerçekleştirilebilmesi için Sakarya Üniversitesi Rektörlüğü'nden veri toplama ve kullanım izni alınmıştır. İzin tarih ve numarası 27/08/2019 - 35955870/604.02.

Kaynaklar (References)

Akay, Ebru Çağlayan. 2018. "Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi." *Social Sciences Research Journal* 7(2): 41–53.

Arslan, Perihan. 1984. "BAZAL METABOLİZMA HIZININ SAPTANMASINDA KULLANILAN YÖNTEMLERİN KIYASLANMASI VE BAZAL METABOLİZMA ENERJİSİNİN VÜCUDUN YAĞSIZ DOKU KÜTLESİ İLE İLİŞKİSİ ÜZERİNDE BİR ARAŞTIRMA." *Beslenme ve Diyet Dergisi*.

Ferrannini, Eleuterio. 1988. "The Theoretical Bases of Indirect Calorimetry: A Review." *Metabolism* 37(3): 287–301.

Flancbaum, Louis et al. 1999. "Comparison of Indirect Calorimetry, the Fick Method, and Prediction Equations in Estimating the Energy Requirements of Critically Ill Patients." *The American Journal of Clinical Nutrition* 69(3): 461–66. https://academic.oup.com/ajcn/article/69/3/461/4694179 (January 30, 2021).

Forman, George. 2003. 3 Journal of Machine Learning Research *An Extensive Empirical Study of Feature Selection Metrics for Text Classification*.

LIU, HSIU YING, YI FA LU, and WEI JAO CHEN. 1995. "Predictive Equations for Basal Metabolic Rate in Chinese Adults. A Cross-Validation Study." *Journal of the American Dietetic Association* 95(12): 1403–8.

Müller, B., S. Merk, U. Bürgi, and P. Diem. 2001. "Berechnung Des Grundumsatzes Bei Schwerer Und Morbider Adipositas." *Praxis* 90(45): 1955–63. https://europepmc.org/article/med/11817239 (January 30, 2021).

Nur Göz, Ayşe et al. 2017. *EKG İŞARETLERİ YARDIMIYLA BİR OTOMATİK ARİTMİ DEDEKTÖRÜ*. Fatih Sultan Mehmet Vakıf Üniversitesi Biyomedikal Elektronik Tasarım, Uygulama ve Araştırma Merkezi. http://acikerisim.fsm.edu.tr:8080/xmlui/handle/11352/2491 (January 30, 2021).

Speakman, John R., Elzbieta Król, and Maria S. Johnson. 2004. "The Functional Significance of Individual Variation in Basal Metabolic Rate." In *Physiological and Biochemical Zoology*, The University of Chicago Press , 900–915. https://www.journals.uchicago.edu/doi/abs/10.1086/427059 (January 30, 2021).

Tverskaya, Rozalia, Russell Rising, Debra Brown, and Fima Lifshitz. 1998. "Comparison of Several Equations and Derivation of a New Equation for Calculating Basal Metabolic Rate in Obese Children." *Journal of the American College of Nutrition* 17(4): 333–36. https://pubmed.ncbi.nlm.nih.gov/9710841/ (January 30, 2021).

Wang, Wenjian, and Zongben Xu. 2004. "A Heuristic Training for Support Vector Regression." *Neurocomputing* 61(1–4): 259–75.

Research Article
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 177-194
DOI: 10.38016/jista.976919

An Inductive-ROC Based Model for Determining Purchasing System Requirements

Tijen Över Özçelik [1*], Orhan Torkul[2]

[1,2] Sakarya University, Faculty of Engineering, Department of Industrial Engineering, Sakarya/Turkey

tover@sakarya.edu.tr, torkul@sakarya.edu.tr

Abstract
The success of system development studies depends on the completeness and accuracy of the requirements collected from system user and stakeholders. Mistakes and deficiencies in the requirements determination process lie in the failure of many system development studies. The dynamism and complexity of today's systems make the requirements determination process even more complicated, and the requirement determination phase, which is the primary stage of system development studies, gains excellent importance. Although there are many techniques, strategies, and models in the literature to improve the requirements Determination process, there are very few studies on their multiple-use, classification of needs, and their experimental realization in a real system. In this study, a system development study was carried out using the System Requirements Model (SRM), which has a flexible structure that allows the system needs to be determined with the use of multiple techniques, at the same time classifies the identified requirements, can decide on the determination levels, and allows the work to be stopped if a sufficient level of needs is determined The developed model consists of three stages: requirement identification, rulemaking/technical/needs matrix creation and classification of needs. In the needs determination process, collection, presentation, verification, and sort of needs. In the second stage of the model, an analogy was made, and the needs and techniques were grouped with the ROC algorithm, and the rule-based RULES-3 algorithm was used to universalize the model. In the third stage of the model, needs were determined based on technical/technical combinations.

Anahtar Kelimeler: Requirements Determination, Requirements Determination Techniques, Stoping Rules, Requirements Engineering, Rank Order Cluster, Inductive Learning, RULES-3.

Satın Alma İhtiyaçlarını Belirlemede Endüktif-ROC Temelli Bir Model

Öz
Sistem geliştirme çalışmalarının başarısı sistem kullanıcıları ve paylaşımcılarından toplanan ihtiyaçların tamlığı ve doğruluğuna bağlıdır. Birçok sistem geliştirme çalışmasının başarısızlığında da ihtiyaç belirleme sürecinde ki hatalar ve eksiklikler yatmaktadır. Günümüz sistemlerinin dinamikliği ve karmaşıklığı ihtiyaç belirleme sürecini daha da güçleştirmekte ve sistem geliştirme çalışmalarının öncelikli aşaması olan ihtiyaç belirleme aşaması büyük önem kazanmaktadır. İhtiyaç belirleme sürecinin iyileştirilmesi amacı ile literatürde birçok teknik, strateji ve model bulunmasına karşın bunların çoklu olarak kullanımı, ihtiyaçların sınıflandırılması ve gerçek bir sistemde deneysel olarak gerçekleştirilmesine yönelik çok az sayıda çalışma bulunmaktadır. Bu çalışmada çoklu teknik kullanımı ile sistem ihtiyaçlarının belirlenmesine olanak sağlayan, aynı zamanda belirlenen ihtiyaçları sınıflayan, belirlenme seviyelerine karar verebilen ve yeterli seviyede ihtiyaç belirlenmesi durumunda çalışmanın durdurulmasına izin veren esnek bir yapıya sahip Sistem İhtiyaçları Modeli (SİM) kullanılarak bir sistem geliştirme çalışması gerçekleştirilmiştir. Geliştirilen model ihtiyaç belirleme, kural çıkarma/teknik/ihtiyaç matrisi oluşturma ve ihtiyaçların sınıflanması şeklinde üç aşamadan oluşmaktadır. İhtiyaç belirleme sürecinde; ihtiyaçların toplanması, sunulması, doğrulanması ve sınıflanması gerçekleştirilmiştir. Modelin ikinci aşamasında bir analoji gerçekleştirilerek ROC algoritması ile ihtiyaç ve teknikler gruplanmış ve modelin evrenselleşmesi için kural tabanlı RULES-3 algoritması kullanılmıştır. Modelin üçüncü aşamasında ise teknik/teknik kombinasyonları bazında ihtiyaçlar belirlenmiştir.

Anahtar Kelimeler: İhtiyaç Belirleme, İhtiyaç Belirleme Teknikleri, Durdurma Kuralları, İhtiyaç Mühendisliği, Derece Sıralama ve Kümeleme, Endüktif Öğrenme, RULES-3.

* Corresponding Author.
E-mail: tover@sakarya.edu.tr

Recieved : 01 Aug 2021
Revision : 25 Aug 2021
Accepted : 06 Sep 2021

1. Introduction

With the developments in technology, information systems have changed the processes of enterprises and have become the most important component of enterprises. In a world where data increases rapidly and information becomes more valuable than anything else, it is only possible for businesses to gain competitive advantage and maintain their existence if they have an effective information system (Laudon, 2014). Thanks to the reports produced by the information systems, the decision-makers (managers) can make decisions at both strategic, tactical, and operational levels, while the objectivity and accuracy of the decisions can increase. At the same time, thanks to these systems, productivity can be increased, and costs can be reduced, and superiority can be achieved against competitors thanks to new products, services, and processes. In this context, digitalization and information systems play a significant role for businesses (Cascio and Montealegre, 2016). The more necessary and important it is for businesses to be digitized and knowledge-based, the more complex and difficult it is to develop such systems. Because of rapidly changing, highly dynamic business environments, a large number of stakeholders and needs make it difficult to develop these systems. In the literature, it is stated that the stage with the highest effort and cost in system development projects is analysis, and it is argued that the problems and failures are caused by the need identification, which is the primary stage of the analysis process (Watson and Frolick, 1993; Davis, 1982; Byrd et al.,, 1992; Vessey and 1993; Wetherbe, 1991; Whitten and Bentley, 2007; Vessey and Conger, 1994). Needs-based problems point to the inadequacy of the developed system, thus rendering the effort and cost-ineffective (Alvarez, 2002; Sommerville at al., 1998; Guinan at al., 1998).

Requirements determination in the literature; is defined as collecting and modeling information for the functions required for the desired system by system developers/analysts (Sommerville at al., 1998). According to another definition, it is the form of understanding and solving the user requirements and problems for the desired system, collecting information for this purpose by system analysts, meeting the system and user requirements with a user-centered approach (Lazar at al., 2000). The concept of requirements engineering is also used in some studies to determine needs and focuses on the methods used in system development studies to extract and validate the set of needs (Hanish at al., 2001). According to the literature, there is a great deal of interest in the needs identification process and the problems experienced in this process.

In this study, the requirements determination process needs inference techniques and stopping rules based on the collected needs assessment are examined in detail. In order to improve the needs determination process, a flexible model was presented and applied to the determination of purchasing process needs. The model is a flexible model that can use the techniques defined in the literature together, thus determining the needs ultimately, classifying the identified needs and deciding to what extent they have been determined, and allowing the determination of needs to be stopped when necessary. The validity of the model was also tested with statistical methods.

2. Literature Review

Identifying requirements in information system development studies is an activity that requires a great deal of effort and is equally challenging. Identifying/identifying, and verifying needs are seen as the two most critical tasks in system development studies (Hanish at al., 2001). It is emphasized that the identification of the requirement is the phase of the system development process that should be best defined (Lazar at al., 2000; Davidson, 2002; Darke and Shanks, 1997; Browne and Pitts, 2004; Freeman, 2004) and it is one of the most challenging information system development activities (Watson and Frolick, 1993; Davis, 1982; Browne and Rogich, 2001; Janz, at al.). Studies on needs determination are in the literature; Identification of information needs, extraction of needs, needs Determination techniques, or needs engineering.

In the system development process, requirements extraction, analysis, validity, and management are intertwined. Extracting needs is also about discovering the needs of the stakeholders. It is basically based on collecting information about the existing system and the desired system. In this activity, it means performing the processes related to the needs repeatedly. The discovery of knowledge and requirements associated with this repetitive and other activity is also called inference. Each stage in this process requires preparation, practice, and analysis. In this sense, system developers and analysts should work on which inference technique should be selected in which process. For this purpose, it is necessary to understand the existing system, organizational structure and application area, and constraints, determine the needs and the characteristics of the current needs, analyze the Users and stakeholders, choose the techniques, tools, and approaches to be used in the needs process. Finally, the requirements should be determined to include all resources (Carrizo at al., 2014). For this reason, it is vital to determine the actual participants of the system and their needs. In addition, mutual interviews with the stakeholders, the experience of the analysts, and the methods and techniques they prefer are decisive in determining the highquality requirements related to the field and the system in the inference phase (Shams at al., 2019).

There is a direct relationship between the success of the developed systems and the complete and sufficient system requirements. In addition, the quality of the system requirements determination process is very much related to the communication and interaction

among the users of the system, analysts, and stakeholders. Incompatibility and communication problems between these groups fail in system development studies (Freeman, 2004). Interviews with the users of the system and senior managers are generally very effective in determining the features of the desired system/reducing the needs. However, the Joint Application Development (JAD) technique is recommended when there are differences in knowledge and perspectives between the users of the system and the analysts (Duggan, 2003). Although it is emphasized in the literature that accurate and complete information needs play a vital role in the planning and implementation of business information systems, it is stated that it is often difficult to collect information needs wholly and accurately (Davis, 1982). This difficulty is due to three reasons;

- ✓ Constraints on people such as problem solvers and information processors,
- ✓ Diversity and complexity of IT needs,
- ✓ It can be shown as the complexity of the interaction between the user and the analysts in defining the needs (Davis, 1982).

The best solution to overcome such problems is advocated for using several different methods and strategies (Hickey and Davis, 2004). Strategies for determining information needs; questioning, deriving from an existing information system, synthesizing the system's characteristics, and discovering from the experience of a developing system. While one of these strategies is determined as the primary strategy, the others should be preferred as supporting strategies. There are different methods and techniques for each strategy. For example, to query; questionnaires, for derivation from an existing information system; existing software, for synthesizing from the characteristics of the system used; process analysis is for discovering from the experience of a developing system; prototyping is recommended (Davis, 1982). Requirements inference techniques; It is highly related to the characteristics of people, processes, and projects. Each technique has different characteristics related to these application areas. For this purpose, an approach that allows the selection of requirements acquiring techniques to achieve the best result in the needs inference process is presented, and in the first stage of the approach, three dimensions as project, people, and process are defined. In the second stage; Three P matrices (3PM) were created, and the relationships between inference techniques and three dimensions were demonstrated. In the third and final stage, the generation of a criteria map for the selection of the inference technique is explained (Tiwari and Rathore, 2017).

System/software development studies are processes based on the solution of user needs and the satisfaction of sharing purposes. Some requirements may be intense, complex, and based on multidisciplinary needs. At the same time, the success of traditional acquiring methods in meeting the needs depends on the users. Because these methods mainly adopt a technical approach rather than a socio-technical point of view and focus on system constraints. Success in information system development depends on identifying the social, organizational, and technical characteristics of the system. At the same time, success in system development is a result based on the acceptability of the system by the users of the system. For this purpose, a socio-technical requirement acquiring process that enables the systematic determination of user needs has been proposed and demonstrated by an experimental study. In the study, the quality of the questions, the readability of the interviews, and the effect of the model on the success of the analyst in terms of socio-technical understanding of the field were evaluated (Wahbeh at al., 2019).

Requirements engineering and system development studies are about discovering the needs of users and stakeholders of the system. Although there are many techniques used by analysts for this purpose, one of the most frequently used is interviews (Bano at al., 2018). Interviews are particularly effective in extracting non-implicit information. At the same time, the analysts' prior knowledge of the field affects the process positively and contributes to understanding the requirements and increasing communication. However, excessive content knowledge can sometimes lead to prejudices (Hadar at al., 2012). Another issue that has been studied related to the determination of needs is the stopping rules. Stopping rules are about when to stop the need inference and information Gathering process. For this purpose, a needs classification that can be used in information system development studies was carried out (Byrd et al.,, 1992), and this classification was later developed (Browne and Rogich, 2001). In another study, the need classes were organized at four levels; goal, process, task, and information-level needs. These requirements can be used in many other problem areas and in any system development work (Browne and Rogich, 2001; Browne and Pitts, 2004). Thanks to these requirements groups, the collected needs can be effectively measured in terms of quality and quantity (Browne and Pitts, 2004). The aim here is to classify the requirements using the identified needs and the analyst's experience. Requirement classes are used as stopping rules to terminate the Process (Pitts and Browne, 2004). The study, starting from the idea that the purpose of the analyst in a design process is to collect enough information, and this information should be measured; carried out a classification of needs. This classification template includes problem areas of assets believed to be critical to the successful design of information systems. This template is also used as stopping rules, thus preventing the extraction of missing or excess information (Browne and Pitts, 2004). A systematic mapping study carried out indicates that there are many studies with requirements acquiring techniques, but there are not many studies based on evaluating the performance of these techniques (Pitts and Browne, 2004).

Improving the process of determining the requirements, in other words, it is said that not determining the needs more than or completely will lead to the development of more efficient and effective information systems in terms of gain, cost, and customer satisfaction, but it is not possible to eliminate the problems of determining the information needs suddenly and completely. This is a difficult process. Therefore, in order to improve the needs assessment process, it should be done step by step within the scope of a model by providing a better understanding of the stages of information collection, presentation, and validation. According to the needs, determination research results; The ability to overcome all problems, and the successful conclusion of system development studies depend on determining the set of needs in the most appropriate and complete way. For this reason, many scientists and researchers have carried out research on this subject and developed many techniques and tools for determining requirements. However, although there are many studies on requirements determination, requirements identification/development techniques, strategies, and requirement classification, application studies on comparison of requirements identification techniques and modeling the requirements identification process are very limited. Research on requirements assessment shows that there is a requirement for models for the comparison of needs assessment techniques and the integrated use of these techniques and that there is a gap in this area. Scientists generally argue that the use of multiple techniques is appropriate for their needs, but they do not provide a model for how to use multiple techniques. As a result, there is a need for a model and applications that can overcome the problems that may arise from incomplete needs collection in the needs determination process, use more than one technique together, and decide on the adequacy by measuring the collected needs. For this purpose, a flexible model was presented in the study and applied in determining the purchasing process needs. The model is a flexible model that can use technical combinations together, thus improving the needs determination process, classifying the determined needs with the stopping rules, and allowing the need determination to be stopped when necessary.

2.1. System Requirements Model (SRM)

In this section, a flexible model developed considering the problems that may be encountered in the needs identification process is presented. This model has three stages, and these stages are; needs determination process, rulemaking/technical/needs matrix creation, and the last step is the classification of requirements. The model is shown schematically in Figure 1. The requirements identification process, which is the first stage of the model, consists of four sub-stages. These; gathering the requirements, presenting the gathered requirements, verifying the requirements presented by the experts in the field, and classifying the requirements. At the stage of meeting the requirements, questionnaires, observation, interview, software review, literature review, and document analysis methods were used.

In order to present the needs, flowchart, associative information map, decision map, scenario, and similarity diagram techniques were used. After the needs presented with the techniques defined in the model are verified by the field experts, through a template created from the literature; It is classified into four levels as goal, process, task, and informatics. Thanks to this classification, a template was tested for the adequacy of the identified requirements and contributed to the verification of the identified requirements. In the second stage of the model; By making an analogy, the ROC algorithm, which is widely used in the grouping of parts and machines in manufacturing, was used to group the requirements and techniques. In order to universalize the model and increase its usability, an artificial intelligence-based RULES-3 algorithm was used.

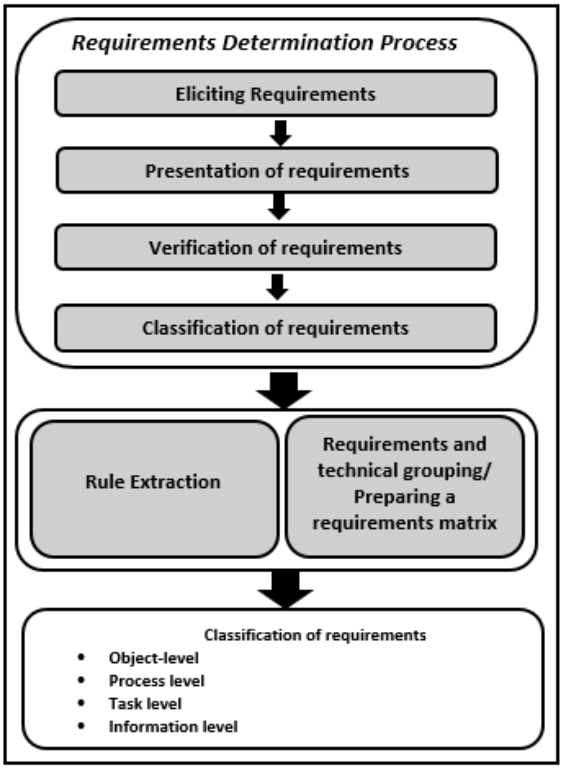


Figure 1. Modeling system requirements [SRM] (Över, 2006; Över Özçelik and Torkul, 2019)

In the third and final stage of the model; For system development studies, a generalized model based on four different need levels and allowing to see a need on the basis of technique/technique combinations have been developed (Över, 2006; Över Özçelik and Torkul, 2019).

3. Using SRM in Determining Purchasing System Requirements

In this section, the sub-processes of SIM; The stages of determining requirements, making rules, creating a technical/requirements matrix, and separating the requirements according to their levels, are explained. The application study was carried out in the Purchasing department of a company that produces wagons. The system requirements model (SRM), whose validity was tested with the pilot application, was applied to the real business environment (Över, 2006; Över Özçelik and Torkul, 2019).

In order to determine the requirements of the new system to be established, a survey was conducted with the employees of the existing system and interviews with senior employees. The handbooks of the current system and some purchasing system software in the market were examined. Matches between problems, products, and processes are provided in order to meet the goals of the desired system and the needs of the users and to produce solutions to existing problems.

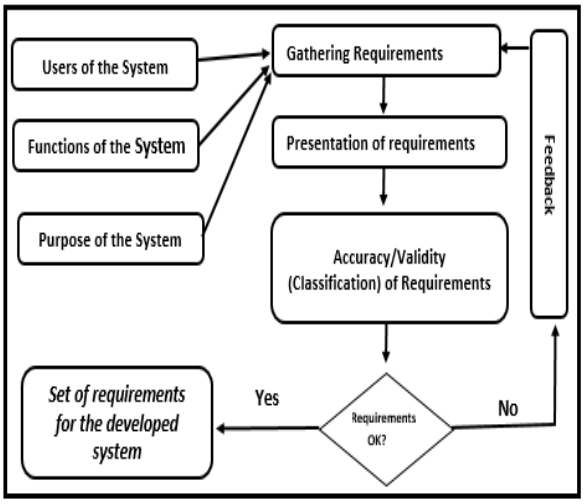


Figure 2. The development process of the new system

This perspective and operation are modeled in Figure 2. In the study, firstly, the existing purchasing system of the enterprise was examined, the problems and malfunctions in the existing structure were determined, it was concluded that a new system should be designed in order to overcome these problems, and for this purpose, the set of requirements required for the new system was determined. It is aimed to design an integrated structure by examining all information flows of the purchasing department with other basic functions within the enterprise. Figure 3 shows the information flows of the purchasing department with other departments. In order to identify and eliminate the problems experienced in the current system, an ideal Purchasing System development study for the enterprise was carried out step by step using SRM.

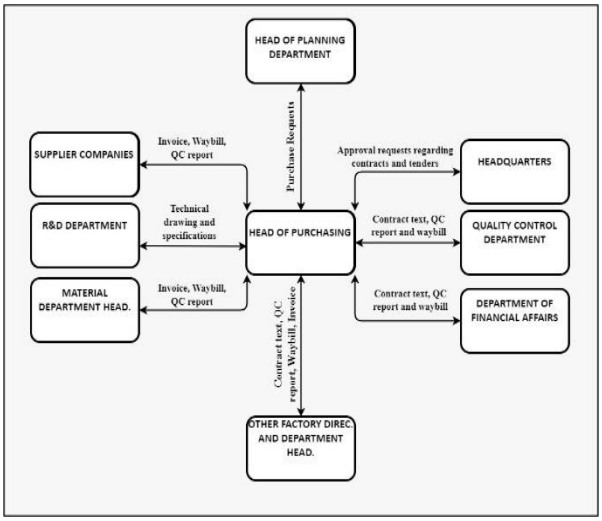


Figure 3. Information flows of the current Procurement System with other basic Functions

3.1. Eliciting system requirements

Within the model; Interviews with field experts/users of the system, survey, examination of system procedures/handbooks/documents, needs collection/information gathering process in the light of the software used in the market and literature information was carried out. The reason for using these methods in an integrated manner is to avoid a lack of information in the data collection process. In the information extraction process, 20 system users were used. These are the users of the system in effect at the enterprise.

3.2. Presentation of System Requirements

Gathered requirements; Flow chart, Evocative knowledge map, Decision map, Scenario, and Similarity diagram technique were presented, and their validity was confirmed by system users. The arguments put forward during the sessions were grouped and evaluated as needs. The set of confirmed requirements was accepted as the requirement set of the purchasing process, and 40 requirements were determined.

- Presenting the requirements of the purchasing system with the flow diagram technique

These techniques, which were selected based on the literature (Browne and Ramesh, 2002) to overcome the conceptual and behavioral problems seen in all three stages of the requirements assessment process and to improve the needs assessment process, were applied in the following order. In Figure 4, the stages and needs of a purchase request are shown with the flow chart technique. In this process, a total of 14 needs were identified.

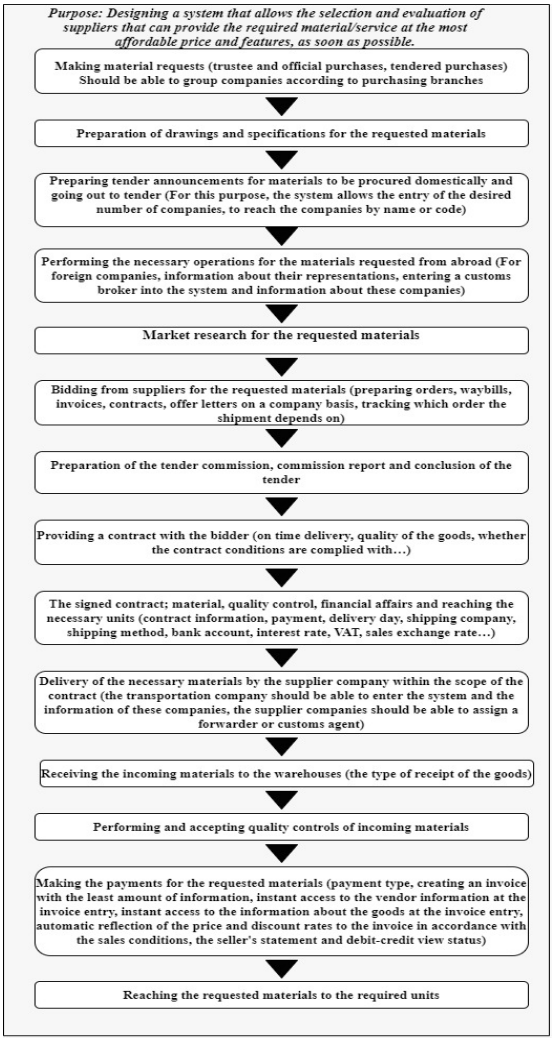


Figure 4. Presentation of requirements with flow chart technique (Över, 2006)

- *Presenting the requirements of the purchasing system with the Evocative knowledge map technique*

Another technique used in the presentation of the gathered requirements is the Evocative knowledge map technique. Figure 5 shows the evocative information map prepared for the purchase request. With the help of this technique, 9 needs were expressed at the first level and the needs at the second level were ignored based on the expert opinion.

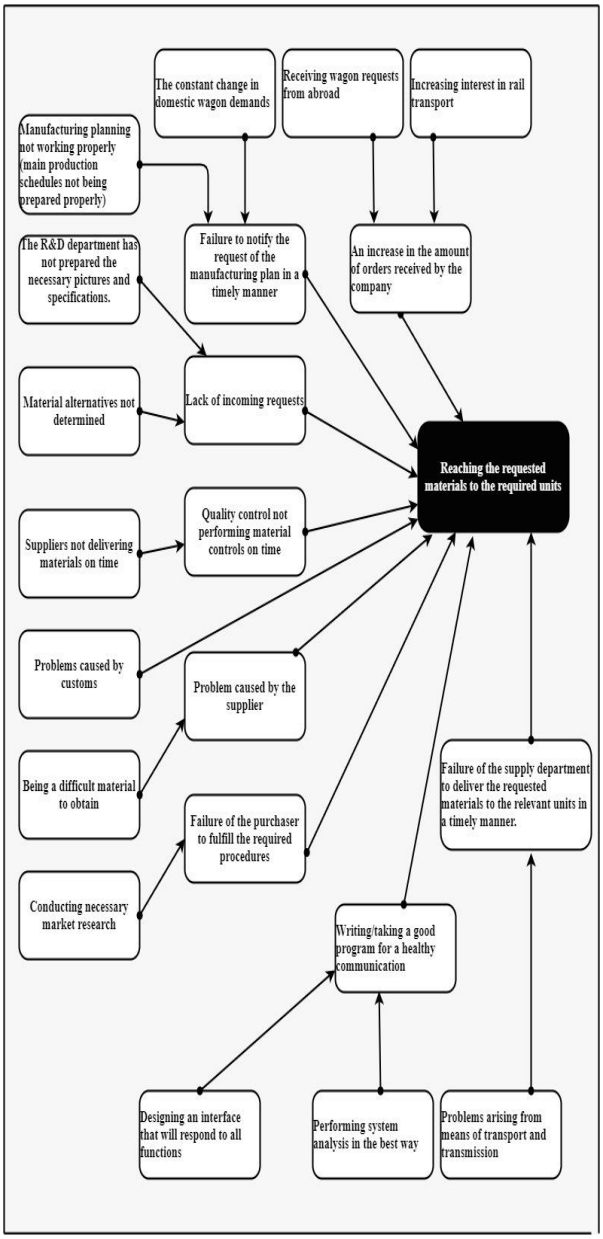


Figure 5. Presentation of requirements with an evocative knowledge map (Över, 2006)

- *Presenting the requirements of the purchasing system with the decision mapping technique*

The decision mapping technique is one of the preferred techniques to overcome the remembering and communication problems in the information gathering, presentation, and verification stages. A total of 19 requirements were identified with the help of this technique. The decision map prepared in Figure 6 is modeled and shown.

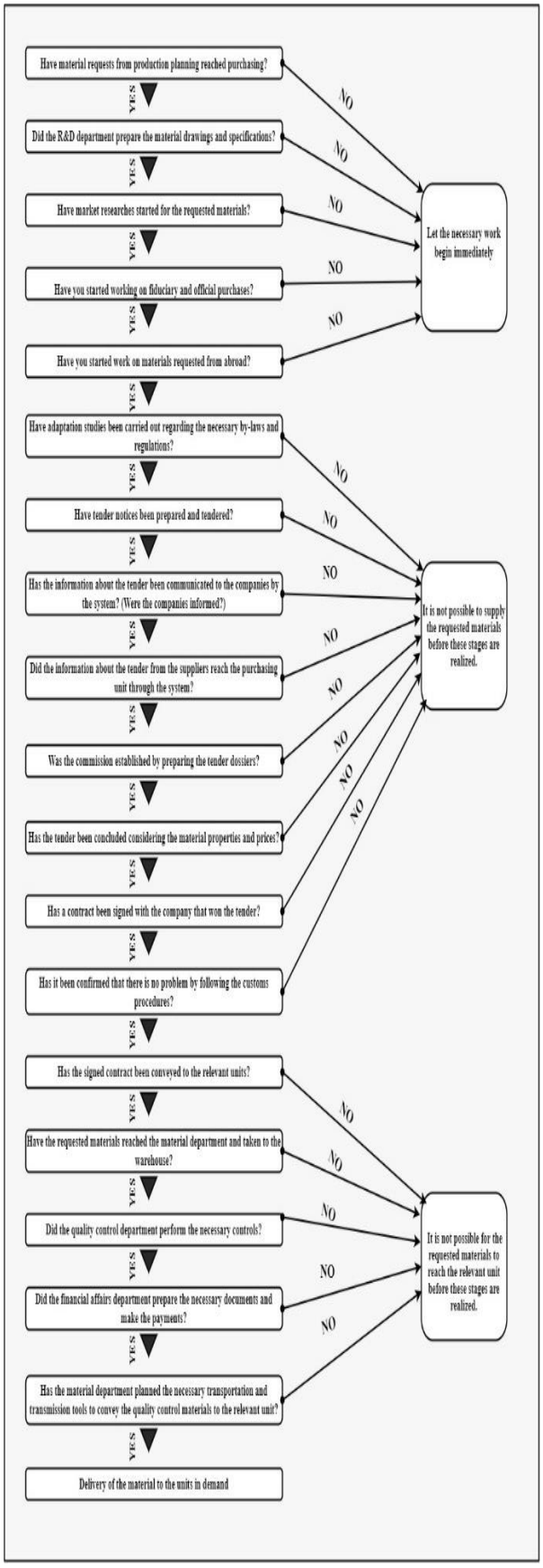


Figure 6. Presentation of requirements with the decision map technique (Över, 2006)

- *Presenting the purchasing process requirements using the scenario technique*

The scenario technique was preferred and applied based on the literature (Sutcliffe, 2003) to overcome the problems arising from behaviors caused by automated actions. Scenarios consist of events and actors. The scenario written for the purchasing system in the enterprise where the application is made is given below. The requirements were expressed by developing a scenario covering the purchasing process. It was accepted that every activity was carried out as it should be without any problems in this scenario. Within the framework of the planned scenario, new orders to the business were conveyed to Manufacturing Planning by Sales/Marketing, and Manufacturing Planning requested materials from the material department to realize the new orders, but materials that were not available were in question.

For this reason, material requests were sent to purchasing. Let these materials be x, y, and z. Let x be a material to be purchased by the trustee. Material y should be a material that requires official purchase, and material z should be a material that needs to be procured from abroad. Manufacturing Planning should convey the specifications and pictures of these materials to the purchase while requesting these materials' supply because these are important in the appropriate procurement of materials. The system to be designed should include all these situations.

- *Presenting the purchasing process requirements using the similarity diagram technique*

The similarity diagram was preferred in the information gathering and validation stages to overcome the diversity/complexity of needs and communication problems. To determine the needs by utilizing the similarity diagram technique, first of all, a group of users of the current purchasing system and experienced people in system development was formed. Discussions were organized, and brainstorming was carried out through group sessions. During these sessions, the discussions and speeches regarding the system were systematically noted, and the arguments put forward in the group were listed as shown in Table 2. These arguments, which are given in the form of a table, are arranged in accordance with the logic of the use of the similarity diagram and the sub-systems that should be covered by the system to be realized based on the arguments that have been determined.

Table 1. Scenario prepared for purchase request (Över, 2006)

Event No	Actors	Actions
1	**Production planning office**	Requesting materials
2	**R&D department**	Preparation of drawings and specifications of materials requested by manufacturing planning
3	**Contract execution and follow-up officer foreign trade officer**	Incoming demand for material x and y
4	**Research tender preparation branch supplier**	Incoming request for material Z (To take necessary actions for the supply of requested materials)
5	**Research tender preparation branch**	Conducting market research for the requested materials, preparing tender announcements, and publishing [Price lists, maturity group, discount rate, and sales condition information about the products sold by each company, minimum and maximum order quantities to be received from companies with an order].
6	**Supplier**	Bidding of the firms participating in the tender for the requested goods
7	**Rapporteur**	Preparing the tender commission and commission report and finalizing the tender
8	**Contract execution and follow-up officer supplier**	Preparation of the contract with the companies that won the tender [The system should give warnings about the companies in line with the information and impressions when necessary]. [The ability to search by code or name of the goods for which a purchase request is created, to directly access the detail information of the goods for which a purchase request is created, to access the previous purchase request and delivery date of the goods for which a purchase request is created, to approve the purchase request, to monitor which order the shipment is connected to, to monitor the stock status in all warehouses at the order entry].
9	**Supplier**	Delivering the requested materials to the company according to the prepared contract
10	**Contract execution and follow-up officer**	The signed contract; material, quality control, financial affairs and relevant units [Introductory information of the company (name, title, address, telephone, fax, capital, product quantity and types, number of workers...), information about company managers or officials, Containing information about the group or holding company to which the firm is affiliated].
11	**Material department**	Receiving incoming materials to warehouses
12	**Quality control**	Performing quality control of incoming materials
13	**Material department**	Delivering the requested materials to the required units
14	**Financial affairs**	Carrying out the ordered and incoming materials [Automatic calculation of all discount and VAT rates, Automatically bringing prices based on companies, goods, or unique formulas, Calculating the cost in the desired foreign currency].

Table 2. Arguments developed within the group for the realization of the purchase request (Över, 2006)

Person	Arguments
Person A)	We need a new system to carry out purchasing transactions and to make healthier decisions regarding this issue.
Person B)	But we don't want a system that will make decisions for us, do we? In other words, we are talking about a system that will help us decide. In short, we are the ones who will make the final decision.
Person A)	You are right, but to make easier and healthier decisions and perform more accessible purchasing functions, we must organize our choices and define some criteria to facilitate this decision-making task.
Person C)	In addition, the criteria we use must be consistent with each other. I think that the criteria we will determine will be of great importance in evaluating the performance of suppliers.
Person A)	An interface should be designed so that it allows us to perform all our functions in the best possible way. E.g., Ability to perform company transactions in the desired foreign currency (calculation ability), group companies according to certain features and sectors, prepare orders, delivery notes, invoices, contracts, letters of offer based on companies, define the account code of the company, define the desired number of addresses for shipment or document sending to companies should include fields such as
Person B)	I think this interface is more the work of the program, so system analysis should be done very well so that the requirements of the whole system can be determined and the system will respond to all our requests.
Person E)	I believe it will be very beneficial for us in keeping our data in electronic form. In this way, we will avoid keeping repetitive data through a shared database.
Person A)	To realize all these, work should be started to convert all our data into the same format.
Person E)	Then we can start by determining the fields in the system interface, the criteria, and the reports we intend to receive from the system, for example; An interface should be designed that allows us to see the previous orders and delivery dates of the ordered goods, the final delivery and delivery date of the order, the authorized person who placed the order, the opportunity to distribute the orders to different companies, the opportunity to see important messages about the seller at the order entry and to determine the order priority, and to evaluate the companies based on these. .
Person D)	If we enter the goods/services to be supplied to this system, is it possible for the system to give us the criteria for this good/service?
Person A)	The criteria, which are essential to us and that we have determined before, must be entered into the system. Therefore, when we enter the goods/services to be procured into the system, it is possible to access its criteria and other information.
Person B)	It's impossible to do this without an expert system.
Person A)	But we do not need an expert system to realize what we are talking about.
Person E)	I read somewhere that some simple expert systems can make more effective decisions than managers in making decisions about the goods/services to be procured.
Person F)	An expert system or decision support system may not be suitable for us. Because; we are an official institution, and we can't depend only on such systems while making some decisions.
Person C)	I do not know that, and I do not trust computers that can make such decisions. In addition, some situations require us to make decisions independently of the system due to urgent conditions such as public procurement laws and regulations.
Person A)	We need a system that puts the decisions in order but leaves the final choice up to us; you know, this is a decision support system. Because, thanks to the reports we get from here, we can make healthier decisions. In other words, we will tell the system which criteria we should use, and it will give us the information that will enable us to decide which goods/services we can obtain at what time and at what price, and from where?
Person E)	I know that; Expert systems have advantages in terms of consistent use of criteria; we should use them consistently, make fair decisions, and get the most appropriate goods/services.
Person F)	We need a system that puts the decisions in order but leaves the final choice up to us; you know, this is a decision support system. Because, thanks to the reports we get from here, we can make healthier decisions.
Person A)	Yes, I agree; I think that a computer cannot fully evaluate some conditions and features for suppliers. It's easy; then, we'll put these exceptions and features into a scoring system.
Person C)	In other words, we will tell the system which criteria we should use, and it will give us the information that will enable us to decide which goods/services we can obtain, for what time and at what price, and from where?

During the sessions, the arguments were grouped according to the system they covered or what they expressed, and the statements in the groups were evaluated as requirements. The structure and similarity diagram technique of all these systems/subsystems and the needs of a purchasing process are given in Figure 7.

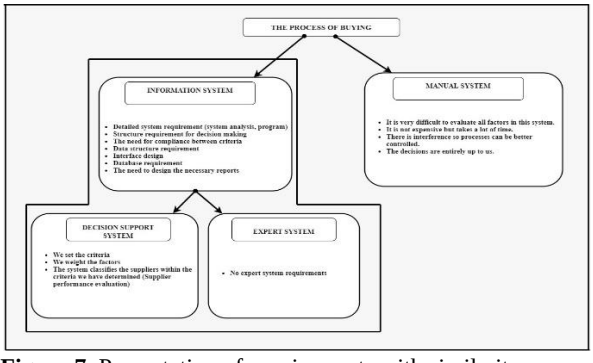


Figure 7. Presentation of requirements with similarity diagram technique (Över, 2006)

Table 3. Identified requirements and techniques used for purchase request (Över, 2006), [Requirements 1/16]

Requ. No	Requirements	Techniques				
		Flow diagram	Evocative knowledge map	Decision map	Scenario technique	Similarity diagram
1	Requesting materials	*		*	*	
2	Preparation of drawings and specifications of materials requested by manufacturing planning	*		*	*	
3	Incoming demand for material x and y (Preparing tender notices for materials to be procured domestically and bidding)	*		*	*	
4	Incoming request for material Z (To take necessary actions for the supply of materials requested from abroad)	*		*	*	
5	Conducting market research for the requested materials	*		*	*	
6	Bidding of the firms participating in the tender for the requested goods	*		*	*	
7	Preparing the tender commission and commission report and finalizing the tender	*		*	*	
8	Preparing the contract with the companies where the tender remained	*		*	*	
9	The signed contract; To ensure that it reaches the material, quality control, financial affairs, and related units	*		*	*	
10	Delivering the requested materials to the company according to the prepared contract	*		*	*	
11	Receiving incoming materials to warehouses	*		*	*	
12	Performing quality control of incoming materials	*		*	*	
13	Carrying out the payment transactions of the requested and incoming materials	*		*	*	
14	Delivering the requested materials to the required units	*		*	*	
15	An increase in the number of orders received by the company		*			
16	Manufacturing scheduling not reporting the request on time		*			

Continuation of Table 3. The requirements and techniques used for the purchase request **[Requirements No. 17/40]**

Requ. No	Requirements	Techniques				
		Flow diagram	Evocative knowledge map	Decision map	Scenario technique	Similarity diagram
17	Lack of incoming requests		*			
18	Quality control not performing material controls on time		*			
19	The problem caused by customs		*			
20	The problem caused by the supplier		*			
21	Failure of the purchaser to fulfill the required procedures		*			
22	Failure of the supply department to deliver the requested materials to the relevant units in a timely manner.		*			
23	Writing/taking a good program for a healthy communication		*			
24	Have you started to work on fiduciary and official purchases?			*		
25	Have adaptation studies been carried out regarding the necessary by-laws and regulations?			*		
26	Has the information about the tender been communicated to the companies by the system?			*		
27	Did the information about the tender from the suppliers reach the purchasing unit through the system?			*		
28	Has the material department planned the necessary transportation and transmission tools to convey the quality control materials to the relevant unit?			*		
29	Detailed system requirement					*
30	Structure requirement for decision making					*
31	The requirement for agreement between criteria					*
32	Data structure requirement					*
33	We set the criteria					*
34	We weigh the factors					*
35	No expert system requirements					*
36	The system classifies the suppliers within the criteria we have determined (Supplier performance)					*
37	Fatabase requirement					*
38	Interface design requirements					*
39	Identifying information formats					*
40	The need to design the necessary reports					*
	Identified requirement in total	*14/40*	*9/40*	*19/40*	*14/40*	*12/40*

Table 3; includes requirement numbers, needs identified for the purchasing process, and identifying techniques. For example, need No. 1; presented by the flowchart, decision map, and scenario technique.

Asterisks (*) corresponding to the needs under the Techniques indicate by which methods the requirements are presented. Any requirement that was presented by several techniques at the stage of creating the needs set was accepted as if it was presented only once. When evaluated in this way, the purchasing process needs to set 40 elements.

3.3. Verification of system requirements

At this stage, the requirements expressed with integrated techniques were examined by the employees of the purchasing department or analysts with knowledge of the purchasing system, and their accuracy and validity were decided. For this purpose, first of all, an informative study was carried out on the need for presentation techniques to the system users, namely field experts. Later, the requirements that were transferred to the presentation techniques were examined by 12 purchasing system users (field experts) and 12 experienced analysts in the sessions held. These sessions were held in ten sessions, twice for each presentation technique. Sessions for each presentation technique lasted approximately 45 minutes. It has been decided that the system is sufficient to meet all the requirements of the system by the users and expert analysts.

Their field knowledge for system users and their experience for analysts have an important place in this verification. The output of this stage is the set of requirements of the system required to be installed to realize the purchase request. The elements of this cluster consist of 40 needs identified with the help of presentation techniques. The most critical issue to be considered in system development studies is determining the system needs most wholly and accurately. Because the functions of the system and the realization of the users' wishes or reaching the goal are based on a correctly and completely determined set of needs. Therefore, these needs need to be measured with a method. For this purpose, a template for need classes was used in the study.

3.4. Classification of the requirements identified for the system

Identifying requirements is seen as the most important step in the successful conclusion of information system development studies. For this reason, the most crucial issue to be considered in system development studies is to determine the system needs most accurately and completely. Realizing system functions and user requests or reaching the goal is only possible with a wholly and correctly determined set of needs. For this purpose, it is suggested that these verified needs should be measured by a method (Browne and Pitts, 2004). In the study, a template for the need classes suggested in the literature was used, so it was possible to decide on the completeness of the need set.

The classification includes four different levels. Needs are classified according to these four different levels and their sub-levels, confirming their completeness and completeness. Table 4 below shows the distribution of the requirements determined for the purchasing system at four primary levels. These levels are; purpose, process, task, and informatics. The needs expressed through presentation techniques are given in need classes based on their sequence numbers.

Table 4. Classification of the requirements determined for the purchasing system according to their levels

Requirement Levels	Requirement Numbers
Goal Level Requirements	1, 5, 6, 8, 12, 15, 16, 17, 18, 20, 21, 22, 23, 25
Process Level Requirements	3, 4, 9, 10, 14
Task-level requirements	2, 7, 11, 13, 19, 24, 26, 27, 28
Information Level Requirements	29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40

When the 40 requirements that make up the purchasing system needs are classified as general need levels, purpose, process, task, and information level, 14 of them were objective, 5 were processed, 9 were task, and 12 were informatics. The requirements set of the purchase request were classified using the general need level template and then the lower levels of this general level template. It was time to implement the second stage of the model. At this stage, the ROC algorithm was used in grouping the purchasing system requirements.

3.5. Using the ROC Algorithm to Determine System Requirements

In the second stage, the need/presentation techniques were grouped, and the families of requirements were formed, and the determination of which techniques were more effective in determining the number of needs, and the rates of determination of the requirements determined based on the technique/technical combination were carried out. For this purpose, the ROC algorithm was used (Över, 2006; Över Özçelik and Torkul, 2019; King, 1980].

- Creating a requirement/presentation techniques matrix for the system

According to the ROC algorithm application logic; After running the algorithm by arranging rows and columns, with the help of the requirements/presentation technique matrix whose elements are 0-1, to gather similar needs together and thus to group presentation techniques to form clusters on the diagonal axis of the matrix. In this way, just as machine groups and part

families are formed by bringing together the same operations by the logic of group technology, need families, and technical groups are formed in this study due to the same logic. Due to the group technology application logic, there is a certain number of parts and a certain number of machines. A certain number of needs are identified in this study, and a certain number of techniques are used to determine them. Just as a certain logic is used during a system development work, even if it serves many different purposes, the system is developed, and its users have particular needs. In this study, to determine these general needs, it has been tried to select the technique/technical combinations to express them. The ROC algorithm also helps in this regard.

While applying the ROC algorithm in the study, matrices were used, need determination techniques were placed in the matrix rows, and determining requirements were placed in the columns. Figure 8a and Figure 8b below, T1 and T2 show the techniques used to determine needs, and P1, P2, P3, P4, and P5 indicate the needs. techniques; T1, T2, T3,….Tn, the requirements are; It is expressed as İ1, İ2, İ3,……In. For example, if the T1 technique determines the need for I1, "1" is written; if not, "0" is written. In this way, the requirements determined by all techniques are entered in order. By running the ROC algorithm, similar needs come together due to the working logic of the algorithm and requirement families. Technical/technical combinations are created that follow each other diagonally.

	i1	i2	i3	i4	i5
T1	1	1	1		
T2			1	1	1

Figure 8a. ROC algorithm was applied to the combination of T1-T2 techniques.

	i3	i2	i1	i5	i4
T1	1	1	1		
T2	1			1	1

Figure 8b. ROC algorithm applied to the combination of T1-T2 techniques.

Figure 8b was obtained by applying the ROC algorithm to the matrix formed by the use of T1 (flow chart) and T2 (connotation information map) techniques in Figure 8a. As shown in Figure 8b, the requirements I1, I2, I3 are expressed by the T1 technique from the requirement family/group 1, while the needs I4, I5 are defined by the T2 technique from the need family/group 2.

- Evaluation of requirements identification success criteria for the system

The pilot application uses the ROC algorithm [35, 36]; It has been observed which technique/technical combinations determine which level requirements and technique/technical combinations determine all needs. The success criteria developed within the model's scope were used to evaluate the created need groups and technical groups belonging to the purchasing system needs (Över, 2006; Över Özçelik and Torkul, 2019).

These success criteria are;

- Need Determination Efficiency Measure of Techniques (t_e)
- Need Verification Efficiency Measure of Techniques (t_v) and
- Average Technical Efficiency Measure (t_{ov}).

In Table 5, Table 6, Table 7, and Table 8 below, the Requirements Determination Efficiency Measure of Techniques (t_v), Needs Verification Efficiency Measure of Techniques (t_{ov}), and Average Technical Efficiency Measure (t_{ov}) values for the purchasing system are given. Symbols T1 represents the flowchart, Associative information map T2, Decision map T3, Scenario technique T4 and Affinity diagram technique T5. As mentioned above, the values calculated through Success Criteria for Determining System Requirements are arranged and tabulated, techniques/technique combinations are given in the first column, values in the second column, values in the third column, and values in the fourth column. Table 5 below shows the success criteria values obtained by using only the techniques for the purchase request. Here, the values of t_e, t_v and t_{ov} for five techniques that are valid within the scope of the model are given.

Table 5. Success criteria for purchase requisition calculated using only techniques

Tech./Tech. Combinations	t_e	t_v	t_{ov}
T1	0.35	0	0.35
T2	0.22	0	0.22
T3	0.48	0	0.48
T4	0.35	0	0.35
T5	0.30	0	0.30

In Table 8, Flowchart/Associative information map/Decision map/Similarity diagram, Associative information map/Decision map/Scenario/Similarity diagram techniques, which have the highest value among the need identification efficiency measure t_e values of the techniques (T1-T2- T3-T5), (T2-T3-T4-T5) and (T1-T2-T3-T4-T5) combinations formed by the use of Flowchart, Associative information map, Decision map, Scenario and Similarity diagram

techniques 1.00 to 40 determined all your requirements. The combination of T1-T2-T3-T4, which is formed by using the techniques as a quadruple, determined 28 needs with its lowest value.

Requirements verification efficiency measure of techniques (t_V) Flowchart / Associative information map / Decision map diagram / Scenario, Flowchart / Decision map diagram / Scenario / Similarity diagram (T1-T2-T3-T4), (T1-T3-T4-T5)) techniques in four and using a combination of Flowchart, Associative information map, Decision map, Scenario and Similarity diagram techniques (T1-T2-T3-T4-T5) techniques in a quintet, between 0.70 and 14 repetitive needs were determined. T1-T2-T3-T5 and T2-T3-T4-T5 technical combinations take the highest value of 0.25 for the average efficiency measure value t_{OV} of the techniques and 25% of the needs per technique T1-T2-T3-T4-T5 technical combination. It was seen that it determined 20% of the needs per technique with 0.20.

3.6. Inductive Learning Approach and Use of RULES-3 Algorithm

The second stage of the SRM involves creating a need/needs determination techniques matrix and parallel with the rulemaking. With the application of the ROC algorithm, the response was obtained to determine which technique/technique combination is effective in determining the needs, the number of repetitive requirements accepted as need verification, and the amount of need determined by each technique on average. But the aim of the study lies in finding out which techniques determine the need at which level and to what extent. However, it is not possible to see this clearly by applying the ROC algorithm. In addition, it is aimed to use need levels as stopping rules in the study. For this purpose, the RULES-3 algorithm was also used in parallel with ROC in the second stage of SRM.

Table 9. The training set obtained for the purchasing system

Techniques/ Combinations	**Characteristic Goal level (W_1: 0.36)**	**Characteristic Process level (W_2: 0.14)**	**Characteristic Task level (W_3: 0.18)**	**Characteristic Information level (W_4: 0.32)**	**Class Range**
	Linguistic Expression	Linguistic Expression	Linguistic Expression	Linguistic Expression	Linguistic Expression
T1	**L**	**VH**	**N**	**VL**	**NM**
T2	**N**	**VL**	**VL**	**VL**	**NM**
T3	**N**	**VH**	**VH**	**VL**	**LM**
T4	**L**	**VH**	**N**	**VL**	**NM**
T5	**VL**	**VL**	**VL**	**VH**	**NM**
T1-T2	**VH**	**VH**	**N**	**VL**	**PM**
T1-T3	**N**	**VH**	**VH**	**VL**	**LM**
T1-T4	**L**	**VH**	**N**	**VL**	**NM**
T1-T5	**N**	**VH**	**N**	**VH**	**PM**
T2-T3	**VH**	**VH**	**VH**	**VL**	**PM**
T2-T4	**VH**	**VH**	**N**	**VL**	**PM**
T2-T5	**N**	**VL**	**VL**	**VH**	**PM**
T3-T4	**N**	**VH**	**VH**	**VL**	**LM**
T3-T5	**N**	**VH**	**VH**	**VH**	**MM**
T4-T5	**L**	**VH**	**N**	**VH**	**PM**
T1-T2- T3	**VH**	**VH**	**VH**	**VL**	**PM**
T1- T2- T4	**VH**	**VH**	**N**	**VL**	**PM**
T1- T2- T5	**VH**	**VH**	**N**	**VH**	**MM**
T1- T3- T4	**N**	**VH**	**VH**	**VL**	**LM**
T1- T3- T5	**N**	**VH**	**VH**	**VH**	**MM**
T1- T4- T5	**L**	**VH**	**N**	**VH**	**PM**
T2- T3- T4	**VH**	**VH**	**VH**	**VL**	**PM**
T2- T3- T5	**VH**	**VH**	**VH**	**VH**	**FM**
T2- T4- T5	**VH**	**VH**	**N**	**VH**	**MM**
T3- T4- T5	**N**	**VH**	**VH**	**VH**	**MM**
T1-T2- T3-T4	**VH**	**VH**	**VH**	**VL**	**PM**
T1-T2- T3-T5	**VH**	**VH**	**VH**	**VH**	**FM**
T1-T2- T4-T5	**VH**	**VH**	**N**	**VH**	**MM**
T1-T3-T4-T5	**N**	**VH**	**VH**	**VH**	**MM**
T2- T3-T4-T5	**VH**	**VH**	**VH**	**VH**	**FM**
T1-T2- T3-T4-T5	**VH**	**VH**	**VH**	**VH**	**NM**

Inductive learning aims at a specific-to-general approach. In this study, starting from a specific example, a model that can be used in whole general system development studies is proposed, and an inductive

learning approach is preferred (Aksoy, 2005/113-120). In using the inductive learning approach, the output is either a decision tree or a set of rules. What is expected from an inductive learning algorithm is to obtain the most general rules possible. Because; These decision trees or rules will guide the system development studies in all different areas that will be carried out later. One of the algorithms used for this purpose is the RULES-3 algorithm (Aksoy, 2005/113-120; Aksoy, 2005/121-132). The preference of this algorithm is that it can classify the most samples by checking the obtained rules one by one and select the most general rules.

The steps to be realized during the implementation of the inductive learning approach and the RULES-3 algorithm were applied step by step as given below.

- Determination of samples,
- Calculation of the representative values of the characteristics of the samples,
- Assigning these representation values to defined class ranges,
- Weighting of the characteristics,
- Determining the class values of the samples,
- Assignment of linguistic variables corresponding to numeric values,
- Creation of the training set after all the above steps and
- It is the extraction of rules from the training set with the help of the RULES-3 algorithm.

As a result of these steps, a training set with 31 examples was prepared, consisting of 5 techniques and combinations of these techniques with 2, 3, 4, and 5. While creating the training set, four different characteristics were used. These express need levels and are named as purpose, process, task, and informatics. These characteristics must have certain weights or degrees of importance due to the RULES-3 algorithm usage logic. These degrees of importance were determined based on the literature.

In the literature, the requirements are divided into four general levels: goal, task, process, and informatics. Also, goal-level needs are divided into ten sub-levels, process-level needs 4, task-level needs 5, and information-level needs nine sub-levels. These sub-levels were collected and accepted as 28 basic needs. Based on this acceptance, the weight falling on each need class has been determined. In this study, it has been accepted that all of these requirement definitions represent a need, they constitute a set of needs, and all of them are equal to 1. Based on this acceptance and evaluation, the weight of each requirement is 0.036. This ratio is multiplied by the number of needs per general class and for the purpose; 0.36 for the process; 0.14 for the task; 0.18 and informatics; It was found to be 0.32.

Characteristics/requirements are expressed numerically first, and then these numerical values are transformed into linguistic expressions by giving specific definition ranges (Över, 2006; Över Özçelik and Torkul, 2019). As a result of all these stages, a training set was created for the purchasing system shown in Table 9. (linguistic expressions; Low/L, Normal/N, Very low/VL, Very high/VH, Never met/NM, Partially Met/PM, Fully Met/FM, Most Met/MM, Little Met/LM). The RULES-3 algorithm was used to transform the information in training set into regular rules. The rules obtained here can be used in the classification of new samples in future studies without the need for a training set. The rules produced by the RULES-3 algorithm are given in Table 10.

Table 10. Rules generated by the RULES-3 algorithm (Över, 2006)

Rule 1	if; goal N and process VL and task VL and information VL in case Class NM
Rule 2	if; goal VL and process VL and task VL and information VH in case Class NM
Rule 3	if; goal L and process VH and task N and information VL in case Class NM
Rule 4	if; goal N and process VH and task N and information VH in case Class PM
Rule 5	if; goal N and process VL and task VL and information VH in case Class PM
Rule 6	if; goal VH and process VH and task N and information VL in case Class PM
Rule 7	if; goal N and process VH and task VH and information VL in case Class LM
Rule 8	if; goal L and process VH and task N and information VH in case Class PM
Rule 9	if; goal VH and process and task VH and information VL in case Class PM
Rule 10	if; goal VH and process VH and task N and information VH in case Class MM
Rule 11	if; goal N and process VH and task VH and information VH in case Class MM
Rule 12	if; goal VH and process VH and task VH and information VH in case Class FM

In the proposed model, by using the techniques together, the requirement determination rates of the techniques were measured, the analogy was realized with the help of an algorithm that has not been used in this field, the success criteria for determining the needs were developed, and a general decision-making structure was developed about which technique/technique could meet the needs at which level and to what extent. At the same time, it is ensured that the need classes are used as stopping rules in determining the needs. When the techniques are examined; It can be said that the flow chart technique is very effective in determining process-level needs and determines all the needs at this level, it is low in terms of purpose-level needs, normal in terms of task-level needs, and very low in terms of information-level needs. Associative knowledge map technique; While goal-level requirements are normal, they have very low identifying efficiency in terms of process, task, and information-level requirements. As a result, it is a system that determines how the proposed model in the studies of determining the system needs can meet the needs at the defined level (purpose, process, task, and informatics) of the technical/technical combinations described in the model. The model is straightforward to use and flexible to be applied to any environment.

At the same time, obtained with the model, the numerical value and linguistic variables of the need determination rates in case the need classes are used as stopping rules, the relationship between the techniques' needs determination efficiency measures, and the techniques were used as stopping rules evaluated. In addition, the "need identification effectiveness measure of the techniques" (t_e) values and the numerical value/linguistic variables of the technique/technical combinations in training set in terms of determining the needs were evaluated by means of the t-test, and it was examined whether there was a significant difference between them. It is possible to say that there is no statistically significant difference at the 5% significance level (95% confidence interval) between the class range and values. As a result, ROC and RULES-3 algorithms give similar results based on the need-determination efficiency success criterion of the techniques. In addition, the correlation analysis shows that the combination, which is determined as the most effective and determines the needs in the entire need set (all 40 requirements), has the most efficient value in terms of requirements determination efficiency.

If the Associative information map, Decision map, and Similarity diagram (T2-T3-T5) techniques are used in the system requirements determination studies, all the needs are met. This combination obeys rule 12 produced by RULES-3. This combination also determines all the requirements as defined in the ROC algorithm. In conclusion; Associative information map (T2) to identify goal-level needs, Flowchart, Decision map and Scenario (T1, T3, and T4) to identify process-level needs, Decision map (T3) to identify task-level needs, and Similarity diagram (T5) to identify information-level needs technique can be said to be the most effective technique.

4. Results

When the applied system requirements model (SRM) is evaluated, all the needs are met if the Associative information map, Decision map, and Similarity diagram (T2-T3-T5) techniques are used. This combination obeys rule 12 produced by RULES-3. This combination also determines all the requirements as defined in the ROC algorithm. Also; Associative information map (T2) to identify goal-level needs, Flowchart, Decision map and Scenario (T1, T3, and T4) to identify process-level needs, Decision map (T3) to identify task-level needs, and Similarity diagram (T5) to identify information-level needs technique can be said to be the most effective technique.

Table 10 below shows the 12 rules produced by RULES-3 for the purchase request. These 12 rules cover all 31 examples given in the training set above. With the help of this created training set and the rules generated from it, inference-based modeling/inductive learning, a general approach was adopted starting from a specific example, and a model was developed that could give beneficial results in other systems. As a result, the model was generalized.

Information; This study was derived from a Ph. D. thesis titled An Inductive-ROC Based Model in Requirement Determination.

References

Aksoy, M. S., (2005). Dynamic System Modelling Using Rules3 Induction Algorithm. *Mathematical and Computational Applications*, Vol. 10, No. 1, pp. 121-132.

Aksoy, M. S., (2005). Pruning Decision Trees Using Rules3 Inductive Learning Algorithm. *Mathematical and Computational Applications*. Vol. 10, No. 1, pp. 113-120

Alvarez, R., (2002). Discourse Analysis of Requirements and Knowledge Elicitation Interviews. *Proceedings of the 35th Hawaii International Conference on System Sciences*, IEEE Computer Society.

Bano, M., Zowghi, D., Ferrari, A., Spoletini & P., Donati, B. (2018). Learning from Mistakes: An Empirical Study of Elicitation Interviews performed by Novices. *International Requirements Engineering Conference* – Banff, Alberta Canada.

Browne, G.J., & Pitts, M.G. (2004). Stopping Rule Use During Information Search in Design Problems. *Organizational Behavior and Human Decision Processes*. 95, 208–224.

Browne, G.J., & Rogich, M.B. (2001). An empirical investigation of user requirements elicitation: comparing the effectiveness of prompting techniques. *Journal of Management Information Systems*. Vol. 17 No 4, pp. 223-249.

Browne, G.J., Ramesh, V., "Improving information requirements determination: a cognitive perspective", Information & Management, 39 (2002), pp. 625-645

Byrd, T. A., Cossick, K. L., & Zmud, R. W. (1992). A Synthesis of Research on Requirements Analysis and Knowledge Acquisition Techniques. *MIS Quarterly*. 16 (1), 117-138.

Carrizo, D., Dieste, O., & Juristo, N. Systematizing requirements elicitation technique selection. (Access date; 30.06.2021), http://oa.upm.es/37447/1/INVE_MEM_2014_195480-3.pdf

Carrizo, D., Ortiz, C., & Aguirre, L. (2016) What do researchers mean by "the right requirements elicitation techniques". *Revista chilena de ingeniería*. vol. 24 Nº 2, pp. 263-273.

Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*. (3), 349-375.

Darke, P., & Shanks, G. (1997). User viewpoint modelling: understanding and representing user viewpoints during requirements definition. *Information Systems Journal*. Vol. 7, pp. 213-239.

Davidson, E.J. (2002). Technology frames and framing: a socio-cognitive investigation of requirements Determination. *MIS Q*. 26 (4) 329–358.

Davis, G. B. (1982). Strategies for Information Requirements Determination. *IBM Systems Journal.* 21(1), 4-30.

Duggan, E.W. (2003). Generating systems requirements with facilitated group techniques. *Hum. Comput. Interact*. 18 (4), 373–395.

Freeman, L., A. (2004). The Effects Of Concept Maps on Requirements Elicitation And System Models During Information Systems Development. *Concept Maps: Theory, Methodology, Technology Proc. of the First Int. Conference on Concept Mapping*. Pamplona, Spain.

Guinan, P. J, Cooprider, J. G., & Faraj, S. (1998). Enabling Software Development Team Performance During Requirements Definition: A Behavioral Versus Technical Approach. *Information Systems Research*. Vol. 9, No. 2.

Hadar, I., Soffer, P., & Kenzi, K. (2012). The role of domain knowledge in requirements elicitation via interviews: an exploratory study. *Springer-Verlag London Limited*.

Hanish, J., Thanasankit, T., & Corbitt, B. (2001). Exploring the cultural and social impacts on the requirements engineering processes – highlighting some problems challenging virtual team relationships with clients. *Journal of Systems and Information Technology*. Vol. 5, No.2, pp. 1-19.

Hickey, A. M., & Davis, A., M. (2004). A Unified Model of Requirements Elicitation. *Journal of Management Information Systems*. Vol. 20, No. 4, pp. 65–84.

Janz, B.D., Frolick, M.N., & Wetherbe, J.C. Human Perception: A Challenge to Organizational Process Optimization. (Access date; 30.06.2021) https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.563.6642&rep=rep1&type=pdf

King, J.R., (1980). Machine Component Grouping Formation in Group Technology. *International Journal of Management Science*. Vol. 8(2), pp. 193-199.

Laudon K. C., & Laudon J. P. (2014). Management Information Systems Managing the Digital Firm, Thirteenth Edition, Pearson Education Limited, https://repository.dinus.ac.id/docs/ajar/Kenneth_C.Laudon,Jane_P_.Laudon_-_Management_Information_Sysrem_13th_Edition_.pdf (Access date;; 6.06.2021)

Lazar, J., Hanst, E., Buschwalter, J., % Preece, J. (2000). Collecting User Requirements in a Virtual Population, WebNet Journal: Internet Technologies. *Applications and Issues*. Vol. 2 No. 4, pp. 20-27.

Över Özçelik, T., & Torkul, O. (2019). Bilişim Sistemi Geliştirmede İhtiyaç Belirleme. *Iksad Publications* – Ankara / Turkey, ISBN: 978-625-7954-15-0.

Över, T., (2006). İhtiyaç Belirlemede Endüktif-Roc Temelli Bir Model. Doktora Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

Pitts, M. G. & Browne, G. J. (2004). Stopping Behavior of Systems Analysts During Information Requirements Elicitation. *Journal of Management Information Systems*. 21:1, 203-226

Shams, F., Ammar, H., & Bhutta, M. (2019). Selection of Software Requirements Elicitation Techniques- A Systematic Review. *Journal of Multidisciplinary Approaches in Science*. 1, Issue 1, 8-17.

Sommerville, I., Sawyer, P., & Viller, S. (1998). Viewpoints for Requirements Elicitation: a practical approach. *Third International Conference on Requirements Engineering*. IEEE Computer Society Press, pp. 74-81.

Sutcliffe, A. (2003). Scenario-based requirements engineering. Paper presented at: Proceedings *11th IEEE international requirements engineering conference*. IEEE, pp. 320-329.

Tiwari, S. & Rathore, S. S. (2017). A Methodology for the Selection of Requirement Elicitation Techniques. arXivPrepr. arXiv1709.08481.

Vessey, I. & Conger, S. (1993). Learning to Specify Information Requirements: The Relationship

Between Application and Methodology". *Journal of Management Information Systems*, 10 (2), 177-201.

Vessey, I. & Conger, S. (1994). Requirements specification: learning object, process, and data methodologies. *Communications of the ACM*. 37 (5), 102-113.

Wahbeh, A., Sarnikar, S., & El-Gayar, O. (2019). A socio-technical-based process for questionnaire development in requirements elicitation via interviews. *Requirements Engineering*. 1-21, (Access date; 30.06.2021), https://scholar.dsu.edu/bispapers/136/

Watson, H. J. & Frolick, M. N. (1993). "Determining Information Requirements for an EIS". *MIS Quarterly*. 17(3), 255-269.

Wetherbe, J.C. (1991). Executive Information Requirements: Getting It Right. *MIS Quarterly*. Vol. 15, pp. 51-65.

Whitten, J.L. & Bentley, L.D.. (2007). Systems Analysis and Design Methods. McGraw-Hill/Irwin, 7th Edition.

Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 4(2) (2021) 195-202
DOI: 10.38016/jista.901670

Eritematöz Skuamöz Hastalıkların Teşhisinde Makine Öğrenme Algoritmaları Performanslarının Değerlendirilmesi

Gürkan Bilgin[1], Ahmet Çifçi[2*]

[1,2] Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Mimarlık Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Burdur, Türkiye

gbilgin@mehmetakif.edu.tr, acifci@mehmetakif.edu.tr

Öz

Eritematöz skuamöz hastalıkların ayırıcı tanısı dermatolojide önemli problemlerden biridir. Hepsi birbirleri arasında, çok az farkla eritem ve ölçeklenmenin klinik özelliklerini paylaşmaktadırlar. Bu gruba dâhil olan hastalıklar; sedef hastalığı, seboreik dermatit, liken planus, gül hastalığı (pityriasis rosea), kronik dermatit ve pityriasis rubra pilaris olarak sınıflandırılabilmektedir. Tanı için genellikle biyopsi gereklidir ancak ne yazık ki bu hastalıklar pek çok histopatolojik özelliği de paylaşmaktadır. Diğer taraftan, son yıllarda özellikle bilgisayar teknolojisindeki gelişmeler ve yapay zekâ teknolojileri, biyomedikal alanda kendine geniş bir uygulama alanı bulmuştur. Tıbbi cihazlarda bilgisayar teknolojilerinin kullanılmasıyla daha hassas, daha hızlı, insandan kaynaklanan hataları minimize eden cihazlar geliştirilmektedir. Dolayısıyla, bu çalışmada, makine öğrenme algoritmaları deri hastalıklarının sınıflandırılması ve tahmininde ne kadar etkili olmaktadır onun araştırılması yapılmıştır. Bu çalışmada, 366 hastaya ait 33 nitelikten oluşan deri doku örnekleri, Destek Vektör Makineleri (Support Vector Machines - SVM), Topluluk Öğrenme Algoritmaları (Ensemble Learning Algorithms - ELA), Karar Ağaçları (Decision Trees - DT) ve k-En Yakın Komşuluk (k-Nearest Neighborhood - k-NN) ile sınıflandırılmış ve en yüksek doğruluk değerleri kaydedilmiştir. Buna göre deri hastalıklarının ayrıştırılması ve sınıflandırılması ile ilgili etkiler araştırılmıştır. SVM ile bu veri setinde, önceki tüm çalışmalardan daha yüksek olan %99.73'lük bir doğruluk elde edilmiştir.

Anahtar kelimeler: Deri Hastalıkları, Destek Vektör Makinesi, Topluluk Öğrenme Algoritmaları, Karar Ağaçları, k-En Yakın Komşuluk.

Evaluation of Machine Learning Algorithms Performance in Diagnosis of Erythematous Squamous Diseases

Abstract

Differential diagnosis of erythematous squamous diseases is one of the important problems in dermatology. They all share the clinical picture of erythema and scaling among each other, with little difference. The diseases included in this group can be classified as psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis and pityriasis rubra pilaris. Biopsy for diagnosis, but unfortunately these diseases also share many histopathological features. Technologies related to some other technologies have found wide application in the biomedical field. With the use of computer technologies in medical devices, more sensitive, faster, faster, cutting devices are developed. Therefore, it has been investigated how effective machine learning algorithms are in classifying and predicting skin diseases. In this study, skin tissue samples consisting of 33 attributes belonging to 366 patients, Support Vector Machines (SVM), Ensemble Learning Algorithms (ELA), Decision Trees (DT), k-Nearest Neighborhood (k-NN) were classified with algorithms and the highest knowledge information was recorded. Accordingly, the effects related to the separation and classification of skin diseases have been investigated. SVM has achieved an accuracy of 99.73% which is higher than all the previous studies on this dataset.

Keywords: Skin Diseases, Support Vector Machine, Ensemble Learning Algorithms, Decision Trees, k-Nearest Neighborhood.

* Sorumlu yazar.
E-posta adresi: acifci@mehmetakif.edu.tr

Alındı : 23 Mart 2021
Revizyon : 22 Haziran 2021
Kabul : 16 Eylül 2021

1. Giriş (Introduction)

Deri, en büyük immünolojik organlardan biridir ve hem dış hem de iç faktörlerin yanı sıra doğuştan gelen ve uyarlanabilir immün yanıtlardan etkilenir (Fonacier vd., 2010). Ayrıca deri, canlıyı ultraviyole ışınlarından, mikrobiyal, kimyasal ve fiziksel saldırılardan koruyan bir bariyerdir (Lee vd., 2006). Anatomileri ve işlevleri açısından önemli ölçüde farklılık gösteren epidermis, dermis ve hipodermis adı verilen üç katmandan oluşur. Epidermis, gözle görülebilen ve çoğunlukla keratinositlerden oluşan dış tabakadır. Dermis, derinin epidermis ile deri altı doku arasında bulunan kıl kökleri, yağ bezleri ve ter bezleri ihtiva eden katmanıdır. Yapısal bileşenleri kollajen, elastin ve hyaluronik asittir (Tabassum ve Hamdani, 2014; Dumas ve Ntambi, 2018; Yousef vd., 2021).

Deri hastalıkları, yenidoğanlardan yaşlılara kadar her yaşı etkileyen ve sıkça ortaya çıkan bir sağlık sorunu olup, çeşitli şekillerde zararlara neden olur. Deri hastalıklarının çoğu da benzer bazı semptomlar göstermektedir. Bu yüzden aralarındaki farkları bilmek önemlidir. Eritematöz skuamöz hastalıkların ayırıcı tanısı dermatolojide önemli problemlerden biridir. Hepsi birbirleri arasında, çok az farkla eritem ve ölçeklenmenin klinik özelliklerini paylaşmaktadırlar. Bu gruptaki hastalıklar; sedef hastalığı, seboreik dermatit, liken planus, pityriasis rosea, kronik dermatit ve pityriasis rubra pilaris'dir (Guvenir vd., 1998).

Sedef hastalığı kalıcı, bulaşıcı olmayan ve iltihaplı bir hastalıktır. Sedef hastalığı olan hastalar, hastalığa genetik bir yatkınlığa sahiptir. Lezyonlar tipik olarak kafa derisi, gövde, dirsekler ve dizlerde görülür (Lyons ve Ousley, 2014). Sebore olarak da adlandırılan seboreik dermatit uzun süren bir deri hastalığıdır. Kırmızı, pürüzlü, yağlı, tahriş olmuş deri semptomları arasındadır (Borda ve Wikramanayake, 2015). Liken planus deri, mukoza zarı ve tırnakları etkileyen idiyopatik papüloskuamöz bir deri hastalığıdır. Kesin insidansı ve yaygınlığı bilinmemektedir. Genellikle orta yaşlı bireyleri etkilemektedir (Gurusamy ve Selvaraj, 2016). Pityriasis rosea veya gül hastalığı sağlıklı çocuklarda ve yetişkinlerde sıklıkla görülen, genellikle 1-2 ay süren iyi huylu akut papüler döküntüdür (Lyons ve Ousley, 2014). Kronik dermatit deride tahriş ve kızarıklık ile kendini gösteren, derinin kötüleşmesine neden olan bir deri iltihabıdır (Nedorost, 2012). Pityriasis rubra pilaris idiyopatik, papüloskuamöz inflamatuar bir deri hastalığıdır. Turuncu-kırmızı pullu plaklar içine birleşmiş hiperkeratotik foliküler papüller, koruyucu adalar ve palmoplantar keratoderma ile karakterizedir (Wang vd., 2018).

Eritematöz skuamöz hastalıkları doğru teşhis etmek için çeşitli teknikler kullanılarak birçok çalışma yapılmıştır. Guvenir ve ark. (1998), eritematöz skuamöz hastalıklarının ayırıcı tanısı ile ilgili ilk çalışmayı yapmışlardır. Oylama özelliği aralıkları (Voting Feature Intervals - VFI5) adlı yeni bir sınıflandırma algoritması geliştirmişler ve eritematöz skuamöz hastalıkları veri seti için doğruluk oranını %99.2 olarak bulmuşlardır. Fidelis ve ark. (2000), genetik algoritma kullanarak veri seti için doğruluk oranını %95 olarak elde etmişlerdir. Karabatak ve İnce (2009), eritematöz skuamöz hastalıklarının teşhisi için birliktelik kuralları ve sinir ağları tabanlı yeni bir özellik seçme yöntemi sunmuşlardır. Önerdikleri yöntemin doğru sınıflandırma oranı %98.61'dir. Xie ve Wang (2011), SVM tabanlı ardışık ileri yönde seçim algoritması kullanmışlar ve 21 özellikten oluşan azaltılmış veri seti için %98.61 sınıflandırma doğruluğu elde etmişlerdir. Çifci ve ark. (2014), sedef ve liken planus deri hastalıkları olmak üzeri iki grup üzerinde çok katmanlı yapay sinir ağı oluşturmuşlar ve ağın korelasyon katsayısını 0.98 olarak bulmuşlardır. Parikh ve Shah (2017), sınıflandırıcı olarak SVM ile sarmal yöntemini kullanarak, 34 özellikten oluşan orijinal veri setinin alt kümesinin 20 özelliği için %97.27 doğruluk elde etmişlerdir. Elsayad ve ark. (2018), hem klinik hem de histopatolojik özellikleri kullanarak eritematöz skuamöz hastalıklarının teşhisi için Ki-Kare Otomatik Etkileşim Algılama (Chi-Squared Automatic Interaction Detection - CHAID) karar ağaçlarının ve bunların torbalanmış (Bagging) ve artırılmış (Boosting) topluluklarının uygulamasının deneysel sonuçlarını sunmuşlardır. Modelin doğruluk oranını %93.81 olarak elde etmişlerdir. Idoko ve ark. (2018), eritematöz skuamöz hastalıklarının ayırıcı tanısı için bulanık sinir ağı (Fuzzy Neural Network - FNN) tanımlama sistemi tasarlamış ve performansını sınıflandırma ve regresyon ağacı (Classification and Regression Tree - CART), birliktelik kuralları ve bulanık c-ortalamalar kümeleme (Association Rules and Fuzzy c-Means Clustering - ARFCMC), uyarlamalı sinirsel bulanık çıkarım sistemi (Adaptive Neuro-Fuzzy Inference System - ANFIS), ve topluluk sınıflandırıcılar (Ensemble of Classifiers - AEC) ile karşılaştırmışlardır. Chaurasia ve Pal (2020), lojistik regresyon (logistic regression - LR), lineer diskriminant analizi (linear discriminant analysis - LDA), k-NN, SVM, CART ve Gaussian Naïve Bayes (NB) gibi doğrusal ve doğrusal olmayan algoritmaları dermatoloji veri setine uygulamışlardır. LR ile %97.94 doğruluk oranına ulaşmışlardır. Verma ve ark. (2020), pasif agresif sınıflandırıcı (Passive Aggressive Classifier - PAC), LDA, radius komşular sınıflandırıcı (Radius Neighbors Classifier - RNC), Bernoulli Naïve Bayes (Bernoulli Naïve Bayesian - BNB), NB ve ekstra ağaç sınıflandırıcı (Extra Tree Classifier - ETC) gibi altı makine öğrenimi sınıflandırma tekniğini deri hastalıklarının tahminini sınıflandırmak için kullanmışlar ve doğruluğu artırmak için torbalama, AdaBoost ve gradyan artırma sınıflandırıcıları olmak üzere üç topluluk tekniği uygulamışlardır. RNC'de uygulanan gradyan artırma topluluğu yöntemi ile %99.68'lik doğruluk elde etmişlerdir. Rashid ve ark. (2020), öznitelik seçimi için işbirlikçi birlikte evrim algoritmasının etkisini, yaygın olarak kullanılan altı farklı makine öğrenimi sınıflandırma algoritması, yani

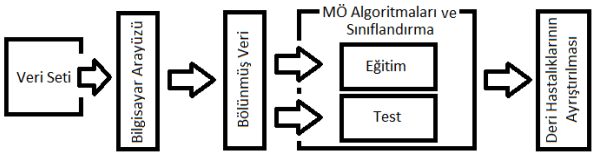
NB, SVM, k-NN, J48, rastgele orman (Random Forest - RF) ve LR üzerinde analiz etmişlerdir. NB ve RF algoritmalarında %97.54 başarı oranına ulaşmışlardır. Putatunda (2020), dermatoloji veri setine hem yeni bir hibrit derin öğrenme yaklaşımı olan Derm2Vec'i hem de derin sinir ağlarını diğer geleneksel makine öğrenme (MÖ) yöntemleriyle birlikte uygulamıştır. Derm2Vec, derin sinir ağları ve ekstrem gradyan arttırma algoritmasının ortalama doğruluk skorunu sırasıyla %96.92, %96.65 ve %95.80 olarak bulmuştur. Shastri ve ark. (2021), eritematöz skuamöz hastalıklarının analizi ve tahmini için gradyan topluluk öğrenme yaklaşımı olan GBoost'u kullanmışlardır. Sonuç olarak %99.45 gibi yüksek bir tahmin başarısına ulaşmışlardır.

Bu çalışmanın ana odak noktası sedef hastalığı, seboreik dermatit, liken planus, pityriasis rosea, kronik dermatit ve pityriasis rubra pilaris gibi farklı eritematöz skuamöz hastalık sınıflarını içeren dermatoloji veri setidir. Hekimler tarafından tanısı konulmaya çalışılan bu hastalıkların tespit edilmesi ve birbirleri arasında ayrıştırılmasının yapılabilmesi için, dermatoloji veri setine dört farklı MÖ yöntemi uygulanarak en yüksek oranda doğru sınıflandırmanın elde edilmesi hedeflenmiştir. Ayrıca yöntemlerin performans karşılaştırmaları yapılmıştır. Çalışmada, MÖ metotları olarak SVM, ELA, DT ve k-NN metotları kullanılmıştır.

Çalışmanın bu bölümünde konu hakkında daha önce gerçekleştirilmiş benzer çalışmalar özet olarak verilmiş, ayrıca çalışmanın hedefleri ve motivasyonundan bahsedilmiştir. İkinci bölümde ise veri setinin içeriği hakkında bilgiler sunulmuş ve önerilen sınıflandırma yöntemleri açıklanmıştır. Üçüncü bölümde analizlerde kullanılan yöntemlerin performansları tartışılmış ve önceki çalışmalar ile kıyasları yapılmıştır. Dördüncü bölümde ise bu çalışmanın sonuçları verilmiştir.

2. Materyal ve Yöntem (Material and Method)

Çalışmada bahsi geçen sınıflandırma sisteminin genel blok şeması Şekil 1'de gösterilmiştir. Bloklara bakacak olursak eritematöz skuamöz hastalıklara sahip hastalardan oluşturulan veri seti, tasarlanan bir bilgisayar arayüzüne hekim tarafından girilecektir. Sistem arayüze sunulan veri setini sayısal değerlere dönüştürüp MÖ metodu girişlerine uygun hale getirecektir. Sınıflandırma için veri seti sırasıyla SVM, ELA, DT ve k-NN metotları ile eğitilecek ve test verileri kullanılarak her bir algoritmanın performansı incelenecektir. Hastalık tahmin performanslarının değerlendirmesi için çapraz doğrulama, duyarlılık ve özgüllük değerleri kullanılacaktır.



Şekil 1. Çalışmanın genel blok şeması (General block diagram of the study)

2.1. Veri Tabanı (Database)

Çalışmada kullanılan veri seti UCI (University of California, Irvine) makine öğrenme veri tabanından alınmıştır (Dua ve Graff, 2019). Gerçek veri tabanı, 33'ü doğrusal değerli ve biri nominal olan 34 öznitelik içermektedir. Çalışmada sadece 33 özellik dikkate alınmıştır. Hastanın yaşı olan son özellik çalışmaya dâhil edilmemiştir. Bu 34 niteliğin, 12 tanesi klinik nitelikleri, geri kalan 22 nitelik ise histopatolojik nitelikleri oluşturmaktadır. Bu niteliklerin açıklamaları Tablo 1'de verilmiştir.

Tablo 1. Özniteliklerin açıklaması (Description of attributes)

Nitelik no	Nitelik açıklaması	Değerler
Klinik nitelikler		
1	Cilt kızarıklığı	0-3
2	Ciltte pullanma	0-3
3	Kesin sınırlar	0-3
4	Kaşıntı	0-3
5	Koebner fenomeni	0-3
6	Çokgen papüller	0-3
7	Foliküler papüller	0-3
8	Oral mukozal tutulum	0-3
9	Diz ve dirsek tutulumu	0-3
10	Kafatası derisi tutulumu	0-3
11	Aile öyküsü	0-1
34	Yaş	Doğrusal
Histopatolojik nitelikler		
12	Melanin tutamama	0-3
13	Sızıntıdaki eozinofiller	0-3
14	PNL sızması	0-3
15	Kabarcıklı derinin fibröz dejenerasyonu	0-3
16	Eksositoz	0-3
17	Akantoz	0-3
18	Hiperkeratoz	0-3
19	Parakeratoz	0-3
20	Rete çıkıntılarının Çomaklaşması	0-3
21	Rete çıkıntılarının uzaması	0-3
22	Yüksek kabarcıklı üstderinin incelmesi	0-3
23	Sponjiyöz çıbanı	0-3
24	Munro mikroapsesi	0-3
25	Fokal hipergranüloz	0-3
26	Pürüzlü katmanın yok olması	0-3
27	Temel katmanın kofullaşması ve zedelenmesi	0-3
28	Sponjiyoz	0-3
29	Retenin testere dişi görünümünde olması	0-3
30	Foliküler boynuz tıkacı	0-3
31	Perifoliküler parakeratoz	0-3
32	İltihaplı mononükleer sızması	0-3
33	Bant şeklinde sızma	0-3

Tablo 1'de gösterilen aile öyküsü niteliği bu hastalıklardan herhangi biri ailede gözlenmişse 1, aksi halde 0 değerine sahiptir. Diğer klinik ve histopatolojik niteliklere 0 ile 3 aralığında bir değer verilir. Burada 0

özelliğin mevcut olmadığını, 3 mümkün olan en büyük değeri ve 1 ile 2 ise göreceli ara değerleri gösterir.

Dermatoloji veri seti altı türden 366 örnek içeren eritematöz skuamöz hastalıklarla ilgilidir: Sedef hastalığı (112 hasta), seboreik dermatit (61 hasta), liken planus (72 hasta), pityriasis rosea (49 hasta), kronik dermatit (52 hasta) ve pityriasis rubra pilaris (20 hasta). Bu hastalıklara ilişkin görseller Şekil 2'de verilmiştir.

Şekil 2. Eritematöz skuamöz hastalıklar (Lyons ve Ousley, 2014) (Erythematous squamous diseases) (a) Sedef hastalığı, (b) Liken planus, (c) Kronik dermatit, (d) Seboreik dermatit, (e) Pityriasis rosea, (f) Pityriasis rubra pilaris

2.2. Önerilen Yöntemler (Proposed Methods)

Çalışmada, hastalardan elde edilen klinik ve histopatolojik niteliklerden oluşturulmuş veri seti SVM, ELA, DT ve k-NN metotları gibi farklı MÖ algoritmalarına sırasıyla uygulanmıştır. MÖ sistemine verilerin uyarlanmasını temsil eden şematik çizim Şekil 3'te gösterilmiştir.

Şekil 3. MÖ metotlarının sisteme uygulanmasının şematik gösterimi (Schematic representation of the application of ML methods to the system)

2.2.1. Destek Vektör Makinaları (Support Vector Machines)

SVM, çok fazla verilerin bulunduğu sınıflandırma problemlerinde fazla kullanılan bir ayrım metodudur. Bu ayrım işlemini bir kümenin tüm veri noktalarını başka bir kümenin tüm noktalarından ayırt eden en uygun hiper düzlemi belirleyerek gerçekleştirmektedir. En uygun hiper düzlemin seçilmesi için marjın kavramının en büyük olması gerekmektedir. Marjın bölgesi, içerisinde herhangi bir veri noktası bulundurmayan ve hiper düzleme paralel olan maksimum genişlik anlamına gelmektedir. Hiper düzleme en yakın verilerden geçen noktalar destek vektörleri olarak isimlendirilir. SVM'de, ikili sınıflandırma fonksiyonlarının karar kuralı denklem 1'deki şekilde ifade edilebilmektedir (Abakar ve Yu, 2014; Subasi ve Gursoy, 2010; Mohammed vd., 2016; Huang vd., 2018).

$$f(z) = sign\left[\sum_{j=1}^{n} a_j y_j x_j z + b\right] \qquad (1)$$

Denklem (1) 'de, xj destek vektörlerini, aj optimum Lagrange çarpanlarını, yj ikili sınıf etiketlerini ifade eder. Parametre b ise fonksiyonun bias parametresi olarak bilinir. Bu çalışmada ise SVM ikili sınıflandırıcılar ile oluşturulan çok sınıflı hata düzeltme çıktı kodları (HDÇK) modeli kullanılmıştır. Bu modelde SVM ikili sınıflandırıcıları birebir kodlama tasarımı ile gerçekleştirilmiştir. HDÇK modeli üç veya daha fazla sınıf ile sınıflandırma sorununu bir dizi ikili sınıflandırma problemine indirger. HDÇK modeli diğer çoklu sınıflandırma problemleri ile karşılaştırıldığında sınıflandırma doğruluğunu iyileştirebilmektedir (Fürnkranz, 2002).

2.2.2. Topluluk Öğrenme Algoritmaları (Ensemble Learning Algorithms)

ELA, eğitim verileri ile birden fazla temel modelin tahmin sonuçlarını birleştirme yoluyla karar vermeyi amaçlar (Zhou, 2012). Buradaki amaç tek bir modele göre daha güçlü ve genellenebilir sonuçlar elde etmektir. Tahmin sonuçlarının bir araya getirilerek sonuç üretilmesi için çeşitli yöntemler mevcuttur. Çalışmada ELA'nın farklı metodları test edilmiş ve en yüksek doğruluğu AdaBoostM2 metodu ile sağlanmıştır. AdaBoost algoritması Boosting algoritmasının daha genel bir versiyonudur (Freund ve Schapire, 1996).

2.2.3. Karar Ağaçları (Decision Trees)

DT makine öğrenme metodları içerisinde insan düşünme mantığına en yakın yapılardan biridir. DT mantık olarak bulunulan durumun bilgi kazancını maksimum yapabilmek için rastgelelik değerini minimize edecek seçimler yapmaya çalışır. Bunu sağlamak için her düğümde hata fonksiyonun tekrardan hesaplar ve minimum hataya sahip durumu seçer. Herhangi bir ağaca benzer DT yapısı temel olarak üç parçadan oluşur. Bunlar yapraklar, dallar ve düğümlerdir. Bu yapıda, kullanılan özellikler düğümler ile temsil edilmektedir. Yapraklar ile dallar ise, ağaç

yapısının kalan diğer bileşenlerini oluşturmaktadır. DT çalışma prensibi, eldeki veriyi küçük parçalara ayırarak en kısa sürede sonucu elde etmek üzerinedir. DT'de, her bir düğüm bir sınıfı belirtmekte veya düğümdeki test verilerini oluşturan olası çıktılara göre numune alanını ayırarak bir test bölümü oluşturmaktadır. Ayrılan alt kümelerin her biri, yeni bir alt ağaç ile çözülerek yeni bir alt sınıflandırma problemi ortaya çıkaracaktır. Yapraklar olarak adlandırılan düğümler, sonuç düğümünün sınıfını içerir. Yaprak düğümü olmayan noktalar ise karar düğümleri olarak tanımlanır.

2.2.4. k-En Yakın Komşuluk Metodu (k- Nearest Neighborhood Method)

k-NN algoritması, kullanım kolaylığına sahip ve kolay anlaşılır temel makine öğrenimi algoritmaları arasında popülerdir. Metodun isminde de var olan "k" değeri, sınıflandırma yapabilmek için veri kümesindeki sonuçlardan kendisine en yakın seçilecek komşu sayısı olarak belirlenir. En yakın özellikleri belirleyebilmek için veri noktaları arasındaki uzaklığın belirlenmesi gereklidir. Bunun için farklı yöntemler vardır. Mesafe Chebyshev, Euclid, Hamming, Manhattan veya Minkowski yöntemleriyle ölçülebilmektedir.

3. Deneysel Sonuçlar (Experimental Results)

Yapılan çalışma süresince, hastalardan elde edilen veriler SVM, ELA, DT ve k-NN algoritma girişlerine uygulanmış ve doğruluk başarıları rapor edilmiştir. Doğrulukların birbirleri ile karşılaştırılması çapraz doğruluk yöntemleri ile test edilmiştir. Buna göre, doğrulama işlemi 3, 5 ve 10 kat çapraz doğrulama (KÇD) işlemleri gerçekleştirilmiştir. Her bir doğrulama işleminde doğruluk değerleri hem tüm veri tabanı ve hem de test verileri için hesaplanması yapılmıştır. Çalışmada hesaplanan doğruluk değeri aşağıda verilen denklem 2 ile hesaplanmaktadır:

$$Doğruluk(\%) = \frac{DT}{DT+YT} \times 100 \qquad (2)$$

Bu formülde, DT parametresi, doğru tahmin edilen değer sayısını gösterirken, YT parametresi ise yanlış tahmin edilen değer sayısını göstermektedir. Algoritmalar arasında en yüksek doğruluk değerini sağlayan makine öğrenme algoritması en uygun sınıflandırma metodu olarak belirlenmiştir. Daha sonra ise en yüksek doğruluk değerini sağlayan algoritma çıkışında elde edilen sınıflar karmaşıklık matrisi ile gösterilmiştir. Çalışmada, çoklu sınıflandırma işlemi olduğu için hassasiyet ve duyarlılık analizleri yerine karmaşıklık matrisi gösterimi yer almaktadır.

Çalışmada Bayesian optimizasyon algoritması kullanılmış ve deneyler sonucunda maksimum karar bölme sayısı 20, minimum yaprak düğüm gözlem sayısı ise 1 seçilmiştir. Yine Bayesian optimizasyon algoritması kullanarak en başarılı yöntem Minkowski olarak belirlenmiştir. Močkus (Močkus, 1974) tarafından geliştirilmiş Bayesian optimizasyonu, makine öğrenmesi alanında etkili bir tekniktir ve değerlendirilmesi pahalı bir amaç fonksiyonu için en iyi seçimdir. Bayesian optimizasyon algoritması, *x* için bir skaler amaç fonksiyonu olan *f(x)*'i minimize etmeyi amaçlayan bir kara kutu tekniğidir. Fonksiyonun deterministik veya stokastik olmasına bağlı olarak, çıkış aynı *x* girişi için farklı olacaktır. Bayesian optimizasyon algoritmasında genellikle Gauss süreci (Gauss Process), olasılık modeli olarak seçilir (Rasmussen, 2003). Minimizasyon süreci üç ana bileşenden oluşur: *f(x)* amaç fonksiyonu için bir Gauss süreç modeli, amaç fonksiyonunun her yeni değerlendirmesinden sonra Gauss modelini değiştiren bir Bayesian güncelleme süreci ve bir *a(x)* kazanım fonksiyonu. Bu kazanım fonksiyonu, bir sonraki değerlendirme noktasını belirlemek için maksimize edilir. Bu fonksiyonun rolü, amaç fonksiyonunda beklenen gelişmeyi ölçerken, onu artıracak değerleri atmaktır (Brochu, 2010).

Analizlerde kullanılan makine öğrenme algoritmalarının performanslarının karşılaştırıldığı doğruluk değerleri Tablo 2'de gösterilmektedir.

Tablo 2. Metotların ortalama k-kat çapraz doğrulama karşılaştırmaları (Mean k-fold cross validation comparisons of methods)

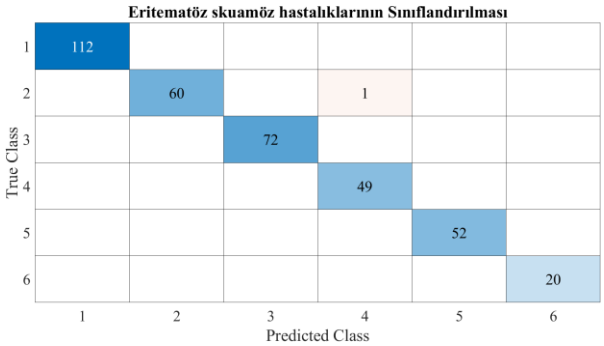
KÇD	Metot	Doğruluk (%)	
		Test Verileri (%)	Tüm Veri (%)
3-KÇD	SVM	99,73	99,73
	ELA	96,72	98,91
	DT	95,08	97,36
	k-NN	96,99	98,99
5-KÇD	SVM	99,73	99,73
	ELA	95,91	99,18
	DT	94,27	97,38
	k-NN	96,17	99,24
10-KÇD	SVM	99,73	99,73
	ELA	96,73	99,67
	DT	94,81	97,62
	k-NN	96,70	99,67

Tablo 2'de gösterilen sonuçlara göre, SVM algoritmasının en etkili metot olduğu açıkça görülmektedir. Tabloda, çapraz doğrulama işlemleri 3, 5 ve 10-KÇD uygulanarak elde edilen sonuçlar gösterilmiştir. Diğer taraftan doğruluk değerleri karşılaştırıldığında gerek test verileri yönünden ve gerekse tüm veriler yönünden, en yüksek tahmin doğruluk değerinin her KÇD bölümünde SVM metodunun üstün ve doğruluk yüzdesinin %99.73 şeklinde olduğu açıktır. Böylelikle, çalışmada gerçekleştirilen optimizasyon teknikleri ve deneysel

çalışmalar, eritematöz skuamöz hastalıkların ayırıcı tanısında SVM yönteminin çok etkili olduğu ve klinik uygulamalarda tanı ve teşhise yardımcı bir yöntem olarak kullanılabileceğini ortaya sermektedir. Bununla beraber, DT metodunun ise bu tür sınıflandırma için diğer metotlardan çok daha yetersiz kaldığı ortaya çıkmaktadır.

Bu çalışmada elde edilen sonuçlar literatürde gerçekleştirilmiş olan diğer çalışmalar ile karşılaştırıldığında, SVM yönteminin hayli başarılı bir yöntem olduğu burada da ortaya çıkmaktadır. SVM sonucunda elde edilen karmaşıklık matrisi Tablo 3'te gösterilmektedir. Bu tabloya göre Sedef hastalığı sınıf 1'i, seboreik dermatit sınıf 2'yi, liken planus sınıf 3'ü, pityriasis rosea sınıf 4'ü, kronik dermatit sınıf 5'i ve pityriasis rubra pilaris sınıf 6'yı göstermektedir. Zaten doğruluk çok yüksek olmakla birlikte algoritma sadece seboreik dermatit hastalığından bir tanesini pityriasis rosea olarak değerlendirmiştir ve bu da doğruluğun çok yüksek olduğunu göstermektedir.

Tablo 3. SVM ile elde edilmiş karmaşıklık matrisi (Confusion matrix obtained by SVM)



Dermatoloji veri seti için sonuçların diğer yayınlanmış yöntemlerle karşılaştırılması Tablo 4'te verilmiştir.

Tablo 4. Dermatoloji veri seti için sınıflandırma yöntemlerinin karşılaştırılması (Comparison of classification methods on the dermatology dataset)

Çalışma	Yöntem	Doğruluk (%)
Guvenir ve ark. (1998)	VFI5	99.2
Fidelis ve ark. (2000)	Genetik Algoritma	95
Karabatak ve İnce (2009)	Birliktelik Kuralları + Sinir Ağları	98.61
Xie ve Wang (2011)	SVM	98.61
Çifci ve ark. (2014)	Çok Katmanlı Yapay Sinir Ağı	98
Parikh ve Shah (2017)	SVM	97.27
Elsayad ve ark. (2018)	CHAID + Bagging ve Boosting	93.81
Idoko ve ark. (2018)	CART	94.84
	ARFCMC	75.96
	ANFIS	95.50
Chaurasia ve Pal (2020)	AEC	97.32
	FNN	98.37
	LR	97.94
	LDA	96.22
	k-NN	85.57
	SVM	92.10
	CART	93.50
	NB	89.02
	PAC	97.38
Verma ve ark. (2020)	LDA	98.98
	RNC	99.68
	BNB	95.36
	NB	98.66
	ETC	97.65
Rashid ve ark. (2020)	NB	97.54
	SVM	96.99
	k-NN	95.62
	J48	95.90
	RF	97.54
	LR	96.99
Putatunda (2020)	Derm2Vec	96.92
	Derin Sinir Ağı	96.65
	XGBoost	95.80
	DT	93.10
	Yapay Sinir Ağı	74.29
	SVM	82.13
	RF	51.13
	NB	92.68
	k-NN	79.30
Shastri ve ark. (2021)	GBoost	99.45
Bu çalışma	**SVM**	**99.73**

Dermatoloji veri seti ile ilk çalışmayı yapan Guvenir ve ark. (1998), VFI5 adlı yeni bir sınıflandırma algoritması ile doğruluk oranını %99.2 olarak bulmuşlardır. Genetik algoritma kullanan Fidelis ve ark. (2000), doğruluk oranını %95 olarak elde etmişlerdir. Birliktelik kuralları ve sinir ağları tabanlı yeni bir özellik seçme yöntemi sunan Karabatak ve İnce (2009), önerdikleri yöntemin doğru sınıflandırma oranını %98.61 olarak bulmuşlardır. Xie ve Wang (2011), SVM tabanlı ardışık ileri yönde seçim algoritması ile %98.61 sınıflandırma doğruluğu elde etmişlerdir. Çifci ve ark. (2014), çok katmanlı yapay sinir ağı oluşturarak ağın korelasyon katsayısını 0.98 olarak bulmuşlardır. Parikh ve Shah (2017), sınıflandırıcı olarak SVM ile sarmal yöntemini kullanarak %97.27 doğruluk elde etmişlerdir. Torbalama ve artırma uygulanarak oluşturulan CHAID karar ağacı algoritmasını kullanan Elsayad ve ark. (2018), %93.81 doğruluk oranına ulaşmıştır. Idoko ve ark. (2018), tasarladıkları FNN tanımlama sistemi ile %98.37 doğruluk oranı elde etmişlerdir. Dermatoloji veri setine birçok yöntem uygulayan Chaurasia ve Pal (2020), LR ile %97.94'lük doğruluk oranına ulaşmışlardır. Verma ve ark. (2020) ise en yüksek doğruluk oranına %99.68 ile RNC yöntemiyle ulaşmışlardır. NB ve RF yöntemleriyle en yüksek doğruluk oranını elde eden Rashid ve ark. (2020), %97.54 olarak doğruluk oranını bulmuşlardır. Putatunda (2020), dermatoloji veri setine tablodan da görüleceği

üzere birçok yöntem uygulamıştır. Derm2Vec (%96.92), derin sinir ağları (%96.65) ve XGBoost Sınıflandırıcı (%95.80) yöntemlerinde en yüksek skorlara ulaşmıştır. Aynı veri seti üzerinde çok yeni bir çalışmada, Shastri ve ark. (2021), GBoost kullanarak %99.45 gibi yüksek bir tahmin başarısına ulaşmışlardır. Tablodan da açıkça görülebileceği üzere dermatoloji veri seti için bu çalışmada elde edilen sonuçlar literatürde gerçekleştirilmiş olan diğer çalışmalar ile karşılaştırıldığında, SVM yönteminin hayli başarılı bir yöntem olduğu ortaya çıkmaktadır.

4. Sonuçlar (Conclusions)

Bilindiği üzere, dermatologlar deri hastalıklarının teşhisini ilgili alanın görsel incelemesine ve ilgili alandan alınan biyopsi ile histolojik örneklerin değerlendirilmesine dayandırırlar. Burada teşhisin doğruluğu hekimin deneyimine büyük ölçüde bağlıdır. Dolayısıyla insan faktörlü yanlış teşhislerin önlenmesi için bilgisayar destekli teşhis araçlarına ilgi gün geçtikçe artmaktadır. Makine öğrenmesi yöntemleri ve araçları geliştikçe, erken teşhisle birlikte maliyetler düşecek ve dermatoloji alanındaki rolü vazgeçilmez hale gelecektir. Bu çalışmada, eritematöz skuamöz hastalıkların teşhisi ve ayrıca daha iyi sınıflandırma doğruluğu için çeşitli makine öğrenmesi metotları kullanılmıştır. Bu çalışmanın literatüre olan temel katkısı, deri hastalıklarının erken teşhisi için yapılan en iyi algoritmayı bulmaktır. Çalışmada, SVM algoritması %99,73 değeri ile en iyi doğruluk sonucunu vermiştir. İkinci en iyi doğruluk değeri ise k-NN metodu ile çapraz doğrulama işlemi 10-KÇD seçilerek %99,67 elde edilmiştir.

Kaynaklar (References)

Abakar, K.A., Yu, C., 2014. Performance of SVM based on PUK kernel in comparison to SVM based on RBF kernel in prediction of yarn tenacity. Indian Journal of Fibre & Textile Research, 39, 55-59.

Borda, L.J., Wikramanayake, T.C., 2015. Seborrheic dermatitis and dandruff: a comprehensive review. Journal of Clinical and Investigative Dermatology, 3(2), 1-22.

Brochu, E., Cora, V. M., De Freitas, N., 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599.

Chaurasia, V., Pal, S., 2020. Machine learning algorithms using binary classification and multi model ensemble techniques for skin diseases prediction. International Journal of Biomedical Engineering and Technology, 34(1), 57-74.

Çifci, A., İlkuçar, M., Bozkurt, M.R., Uyaroğlu, Y., 2014. Liken planus ve sedef deri hastalıklarının çok katmanlı yapay sinir ağı kullanılarak sınıflandırılması. Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 5 (1), 30-36.

Dua, D., Graff, C., 2019. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Dumas, N.S, Ntambi, M.J., 2018. A discussion on the relationship between skin lipid metabolism and whole-body glucose and lipid metabolism: systematic review. Journal of Cell Signaling, 3(3), 191.

Elsayad, A.M., Al-Dhaifallah, M., Nassef, A.M., 2018. Analysis and diagnosis of erythemato-squamous diseases using CHAID decision trees. IEEE 15th International Multi-Conference on Systems, Signals & Devices (SSD). 19-22 March 2018, Yasmine Hammamet, Tunisia, pp. 252-262.

Fidelis, M.V., Lopes, S.H., Freitas, A.A., 2000. Discovering comprehensible classification rules with a genetic algorithm. Proceedings of the 2000 Congress on Evolutionary Computation, 16-19 July 2000, La Jolla, CA, USA, pp. 805-810.

Fonacier, L.S., Dreskin, S.C., Leung D.Y.M., 2010. Allergic skin diseases. Journal of Allergy and Clinical Immunology, 125(2), 138-149.

Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference, July 1996, pp. 148-156.

Fürnkranz, J., 2002. Round robin classification. The Journal of Machine Learning Research, 2, 721-747.

Gurusamy, L., Selvaraj, U., 2016. Clinicopathological study of lichen planus in a tertiary care center. International Journal of Scientific Study, 4(1), 244-247.

Guvenir H.A., Demiroz G., Ilter N., 1998. Learning differential diagnosis of eryhemato-squamous diseases using voting feature intervals. Artificial Intelligence in Medicine, 13, 147-165.

Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., Xu, W., 2018. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics-Proteomics, 15(1), 41-51.

Idoko, J.B., Arslan, M., Abiyev, R. 2018. Fuzzy neural system application to differential diagnosis of erythemato-squamous diseases. Cyprus J Med Sci, 3(2), 90-97.

Karabatak M., Ince M.C., 2009. A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases. Expert Systems with Applications, 36(10), 12500-12505.

Lee, S.H., Jeong, S.K., Ahn, S.K., 2006. An update of the defensive barrier function of skin. Yonsei Medical Journal, 47(3), 293-306.

Lu, Y., Tian, Q., 2009. Discriminant subspace analysis: An adaptive approach for image classification. IEEE Transactions on Multimedia, 11(7), 1289-1300.

Lyons, F., Ousley, L., 2014. Dermatology for the advanced practice nurse. Springer Publishing Company, New York.

Močkus, J., 1974. On Bayesian methods for seeking the extremum. Optimization Techniques IFIP Technical Conference. 1-7 July 1974, Springer, Berlin, Heidelberg, pp. 400-404.

Mohammed, M., Khan, M.B., Bashier, E.B.M., 2016. Machine learning: algorithms and applications. CRC Press.

Nedorost, S.T., 2012. Generalized dermatitis in clinical practice. Springer Science & Business Media, Berlin.

Parikh, K.S., Shah, T.P., 2017. Feature selection paradigm using weighted probabilistic approach. International Journal of Advanced Science and Technology, 100, 1-14.

Putatunda, S., 2020. A hybrid deep learning approach for diagnosis of the erythemato-squamous disease. IEEE International Conference on Electronics, Computing and Communication Technologies. 2-4 July 2020, Bangalore, India, pp. 1-6.

Rashid, A.N.M.B., Ahmed, M., Sikos, L.F., Haskell-Dowland, P., 2020. A novel penalty-based wrapper objective function for feature selection in Big Data using cooperative co-evolution. IEEE Access, 8, 150113-150129.

Rasmussen, C.E., 2003. Gaussian processes in machine learning. Summer School on Machine Learning. Springer, Berlin, Heidelberg, pp. 63-71.

Shastri, S., Kour, P., Kumar, S., Singh, K., Mansotra, V. 2021. GBoost: A novel Grading-AdaBoost ensemble approach for automatic identification of erythemato-squamous disease. International Journal of Information Technology, 13(3), 959-971.

Subasi, A., Gursoy, M.I., 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert systems with applications, 37(12), 8659-8666.

Tabassum, N., Hamdani, M., 2014. Plants used to treat skin diseases. Pharmacognosy Review, 8(15), 52-60.

Verma, A.K., Pal, S., Kumar, S., 2020. Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study. Applied Biochemistry and Biotechnology, 190(2), 341-359.

Wang, D., Chong, V.C-L., Chong, W-S., Oon, H.H., 2018. A review on pityriasis rubra pilaris. American Journal of Clinical Dermatology, 19(3), 377-390.

Xie, J., Wang, C., 2011. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Expert Systems with Applications, 38(5), 5809-5815.

Yousef, H., Alhajj, M., Sharma, S., Anatomy, 2021. Skin (Integument), Epidermis. Treasure Island, FL: StatPearls Publishing.

Zhou, Z.H., 2012. Ensemble Methods: Foundations and Algorithms. New York, USA, Chapman and Hall/CRC.

Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 4(2) (2021) 203-211
DOI: 10.38016/jista.920659

Nano Sistemlerde Hücreler Arası Haberleşmenin Analiz Edilmesi

İbrahim Işık[1*], M. Emin Tağluk[2], Esme Işık[3]

[1,2] İnönü Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Malatya, Türkiye

[3] Turgut Özal Üniversitesi, Tıbbi Hizmetler ve Teknikler Bölümü, Malatya, Türkiye

ibrahim.isik@inonu.edu.tr, mehmet.tagluk@inonu.edu.tr, esme.isik@ozal.edu.tr

Öz

Farklı canlı türleri farklı zeka seviyelerine sahip olduğundan, akıl ve zekanın beyinle bir bağlantısının olabileceği düşünülmektedir. Akıl, bilinç ve zaka insanlık tarihinden beri bilim insanlarını etkileyen en etkileyici kavramlar arasında olmuştur. Ancak bilinci, yalnızca sinir sisteminin veya beynin bir ürünü olarak görmenin de uygun olmayacağı düşünülmektedir. Beynin, yaratıcılık, düşünce ve duygu vs. gibi organizasyonlardan sorumlu olduğu bilinmektedir. Bilincin varlığı için ise sinir sistemi veya beynin gerekli olmadığı birçok çalışma tarafından daha önce ispatlanmıştır. Bunun en basit örneği bitkilerin herhangi bir sinir sistemi veya beyni olmamasına rağmen güneşe yönelebilme bilinçlerinin olmasıdır. Bilinçli bir nano haberleşme modelin tasarlanabilmesi için öncelikle biyolojik canlıların ürettiği sinyallerin tasarlanan bu model ile elde edilmesi gerektiği düşünülmektedir. Çünkü canlıların bütün hücrelerinin bu sinyaller (aksiyon potansiyel) aracılığı ile birbiri arasında iletişim kurduklarına inanılmaktadır. Bu sebeple bu çalışma kapsamında, biyolojik hücrelerin ürettiği aksiyon potansiyel sinyaline neredeyse birebir benzer sinyal üreten elektronik bir devre tasarlanmıştır. Üretilen aksiyon potansiyelin gerçek bir nöron hücresinin ürettiği aksiyon potansiyele yakın olabilmesi için literatürdeki elektronik modellerde kullanılan elemanlar incelendikten sonra böyle bir sistem geliştirilmiş ve bu sistemde kullanılan parametrelerin değeri de yine üretilen aksiyon potansiyelin benzerliğini arttıracak şekilde uzun süren denemeler sonunda hassas bir şekilde tespit edilmiştir. Daha sonra ise tasarlanan model ile iki hücrenin birbiri ile haberleşmesinden elde edilen veriler incelenmiştir.

Anahtar kelimeler: Yapay zeka, Yapay bilinç, Nano sistemler, Unipolar junction transistör

Analyzing of Intercell Communication in Nano Systems

Abstract

Since different living species have different intelligence levels, it is thought that intelligence and mind may have a connection with the brain. Mind, consciousness and intelligence have been among the most influential concepts that have influenced scientists since human history. However, it is thought that it would not be appropriate to view consciousness only as a product of the nervous system or the brain. Your brain, creativity, thoughts and emotions etc. It is known to be responsible for such organizations. It is previously defined in several studies that the brain or nervous system is not necessary for the presence of consciousness. The simplest example of this is that plants do not have any nervous system or brain, but have the consciousness to turn towards the sun. In order to design a conscious nano communication model, it is thought that the signals produced by biological cells must be obtained with this designed model. Because it is believed that all the cells of living things communicate with each other through these signals (action potentials). For this reason, within the scope of this study, an electronic circuit that produces an almost same signal to the action potential signal produced by biological cells has been designed. In order for the produced action potential to be close to the action potential produced by a real neuron cell, such a system was developed after examining the components used in the electronic models in the literature, and the value of the parameters used in this system was determined precisely after long trials, again increasing the similarity of the produced action potential. Then, the data are obtained from the communication of two cells with each other with the designed model were analyzed.

Keywords: Artificial intelligence, Artificial consciousness, Nano system, Unipolar junction transistor

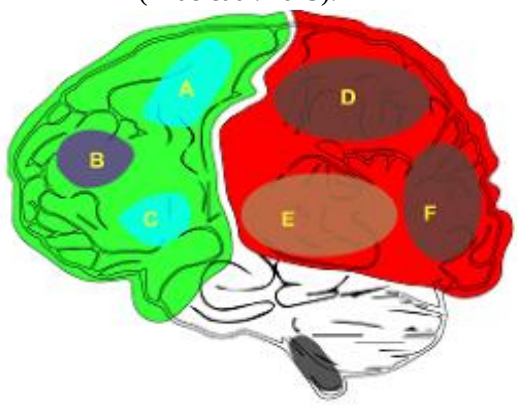
* Sorumlu yazar.
E-posta adresi: ibrahim.isik@inonu.edu.tr

Alındı : 19 Nisan 2021
Revizyon : 20 Temmuz 2021
Kabul : 9 Ağustos 2021

1. Giriş (Introduction)

1.1. Biyolojik bilinç (Biological consciousness)

Bilinç kavramı, bilimsel ve klinik öneme sahip nörobilim ve nörobiyolojik bilimin bir konusudur. Bilinç ağı ile ilgili olarak bilinç araştırmalarındaki en son ilerleme (Zhao et al. 2019) tarafından nöro cerrahi perspektifinden incelenmiştir. Bu çalışmaya göre bilincin iki temel özelliği vardır. Bunlar, uyanıklık ve farkındalık. Tıp literatüründe bilinç, Şekil 1'de gösterildiği gibi A, B, C, D, E ve F korteksi gibi bölümlere ayrılabilir. Burada A, birincil motor korteksi, B, dikkat veya çalışma belleği, C, sözlü rapor, D, diğer bilinç içeriği, E İşitsel bilinci ve F görsel bilinci göstermektedir (Zhao et al. 2019).



Şekil 1. Bilincin beyindeki nörobiyolojik dağılımı (Zhao et al. 2019) (Neurobiological distribution of consciousness in the brain)

Beynin tüm bu korteks bölümleri, özellikle prefrontal ve posterior oksipital korteksler ve klostrum olmak üzere, bilinç oluşumunda önemli bir rol oynar. Beyindeki korteksi Şekil 1'de gösterildiği gibi varsayarsak, bilincin tek bir beyin bölümünden (korteksinden) kaynaklanmadığını bunun yerine yerel olarak oluştuğunu kabul etmemiz gerekir. Bilincin nörobiyolojik mekanizmalarını anlamak için, bilinç oluşumunun tüm kritik çekirdek işlevlerini ve temel beyin parçalarının serebral korteksinin birbirine bağlaması gerekir (Zhao et al. 2019). Bilincin, DNA'sında kodlanmış içgüdüsel talimatlara göre çalışan birbirine bağlı sinir hücresi ağları olduğuna inanılmaktadır. Ayrıca her nöronun, tüm beyinde neler olup bittiğinin farkında olmadan sadece bazı fizik yasalarını takip ederek görevini yerine getirdiği bilinmektedir (Marchetti 2018).

1.2. Yapay zeka (Artificial intelligence)

Son yıllarda, articial intelligence-yapay zeka (AI-YZ) ve makine öğrenimi alanlarında dikkate değer bir ilerleme kaydedilmiştir. Önümüzdeki yıllarda, akıllı otonom, tıp, genetik, ilaç tasarımı ve robotik gibi çeşitli alanlardaki YZ uygulamalarının daha da artacağı öngörülmektedir. Yapay zekanın 1950'lerde ortaya çıkışından bu yana insanlar, kendileri gibi öğrenen ve düşünen makineler yapmak istemişlerdir (Lake et al. 2017). Literatürde tıp, mühendislik ve hatta din alanlarında farklı bilinç tanımına sahip birçok çalışma bulunmaktadır (Bilal Er and Aydilek 2019; Kinouchi, MacKin, and Hartono 2018; Xu et al. 2019). Günümüz teknolojisi ile bilincin tam olarak mekanizmasını veya çalışma prensibini tanımlamanın imkansız olduğu düşünülmektedir. Bilim insanları uzun bir süredir, yapay sinir ağı (YSA) gibi matematiksel, mantıksal veya benzetim çalışmaları ile bilinci tanımlamak için birçok çalışma yapmışlardır. Yapay zeka olarak da adlandırılan YSA ile beyin bilincinden ilham alınarak bilinçli modellerin geliştirilmeye çalışıldığı sistemler son zamanlarda literatürde görülmektedir(Lake et al. 2017; Zhao et al. 2019). Stephen Hawking, 2014 yılında BBC'de, yapay zeka teknolojileri hakkında yaptığı bir konuşmada "Yapay zekanın gelişmesinin insan ırkının sonunu getirebileceğini düşünüyorum" demiştir. Tesla'nın CEO'su Elon Musk ise, 2017'de "Yapay zeka teknolojisi insan uygarlığının varlığı için temel bir risktir" demiştir (Kak n.d.). YZ'dan önce yapay sinir ağlarının (YSA) tanımlanması gerekmektedir çünkü YZ sistemleri YSA kullanılarak geliştirilmektedir.

Yapay sinir ağlarını modellemek için denetimli veya denetimsiz öğrenme algoritmaları kullanılmaktadır. Mevcut çeşitli sinir ağı mimarileri ve öğrenme algoritmaları arasında, Kohonen'in self organizing map (SOM) (Banfield and Raftery 1992; Barreto et al. 2002; Eter and As 2002), en önemli sinir ağı modellerinden biridir. Bu model, denetimsiz öğrenme algoritmasına sahip, retina - korteks haritalaması ile motive edilen ilişkisel bellek modeli için geliştirilmiştir. SOM, girişten çıkışa haritalama sağlayabilir. SOM, sensörlerden beyin korteksine kadar topografik haritalamanın matematiksel bir modelidir. (Yin 2008), SOM'u biyolojik olarak inceleyerek ve Hebbian öğrenmeye dayalı retina - korteks haritalamasının basitleştirilmiş ve teorik bir matematiksel modeli olduğu gösterilmiştir.

Başka bir YSA modeli, "System applying High Order Computational Intelligence" (SHOCID) projesi ile (Neukart et al. 2012) tarafından önerilmiştir. Kortikal YSA'lar, girdi verilerinin yalnızca bir veya daha fazla yatay gizli katmanda (veya kortekste) değil, birkaç paralel gizli katmanda işlenmesiyle karakterize edilir. Giriş, YSA'nın her bir korteksinde ileriye doğru beslenir (Neukart et al. 2012). Yapay zeka konusu yaklaşık 1960'dan beri çalışılan ve insan zekasının yapay olarak bilgisayarlarda yeniden üretilmesine amaçlayan bir konudur. Zeka, sadece bir insan karakteri değildir veya insan beyniyle sınırlı değildir. Zeka, doğal ya da fiziksel bir olgunun bir biçimi olarak düşünülmelidir. Dinamik denge fikrinin otonom uyarlamalı sistemlerde bilgi işlemeye uygulanmasıyla "zekayı ne yapar" sorusunun çözülebileceği düşünülmektedir. Biyolojik bilinç, YSA ve YZ açıkladıktan sonra yapay bilinç tanımının yapılması daha doğru olacaktır.

Biyolojik bilinçten ilham alınarak geliştirilen yapay bilinç (YB) sistemleri, son yıllarda birçok bilim insanı tarafından artan bir şekilde çalışılmaktadır (Buttazzo

2008; Kinouchi et al. 2018). Bir makinenin bilinçli olabilmesinin ancak yapay zeka ve biyolojik bilincin; nöroloji, fizyoloji ve mühendislik disiplinlerinin yardımıyla birleştirilmesi ile mümkün olabileceği düşünülmektedir. Yapay ve doğa bilincine sahip sistemleri birleştirmeden önce, belki de gerçekliğin doğası sorunu tartışılmalıdır. Örneğin, bir makine parçaları ve ara bağlantıları ile tanımlanabilir mi? Felsefede iki okul vardır ve bunların konumları ontic ve epistemic olarak adlandırılır. Bu okullarda biri gerçekliğin var olduğuna, diğeri de gerçekleştiğine inanmaktır. Var olma kelimesinin kökeni, materyalizm ile ilişkilendirilirken, gerçekleşme kelimesinin kökeni, gözlemcilere daha önemli bir rol vermektedir (Kak n.d.).

YB, bilinçli robotlar gibi birçok endüstriyel ve tıbbi alanda kullanılabilir. Örneğin, evlerimizde mutfakların bir parçası olan mutfak robotlarının bilinçlenmesi ile diğer mutfak gereçleri ile bağlantılı olarak birçok iş yapılabilir (Yamazaki et al. 2010). Ayrıca gelecekte terapatik ve hemşire robotlar için de yüksek miktarda talep olacağı tahmin edilmektedir (Simon Peter van Rysewyk 2015). Örneğin, akıllı oyuncaklar ve robotik evcil hayvanlar yaşlıları ve çocukları destekleyebilir ve onlara yardım edebilir (Amanda Sharke 2010). Bu robotların, psikolojik konularda stres yönetimi ve danışmanlığa yardımcı olmak için de kullanılabileceği tahmin edilmektedir (Chandra 2017). (Marchetti 2018) çalışmada okuyuculara, her biri belirli bir teknik veya felsefi konuyu ele alan ve "Bilgisayarlar düşünüyor mu, yoksa sadece hesap mı yapıyorlar? İnsanlar düşünüyor mu yoksa sadece hesap mı yapıyorlar? Bilinç insanların ayrıcalığı mıdır? Bilinç, beynin yapıldığı malzemeye bağlı mı yoksa farklı bir donanım kullanılarak kopyalanabilir mi?" gibi birçok soru yöneltilmiştir. Bu sorulara günümüz teknolojisiyle cevap vermek kolay değildir. Çünkü bilgisayar bilimi, nörofizyoloji, felsefe ve hatta din gibi birçok farklı disiplini birlikte analiz etmek gerekecektir. Öte yandan, yakın gelecekte karmaşık hesaplama yapan makinelerde de yapay bilincin ortaya çıkacağına inanılmaktadır. Bu elbette kolay olmayacaktır, çünkü 1015 sinapsı simüle etmek için toplam 4 * 1015 bytes (4 milyon Gigabytes) hafıza gereklidir. Daha sonra, nöron çıktılarını ve diğer durumları depolamak için yardımcı değişkenler dahil olmak üzere tüm insan beynini simüle etmek için yaklaşık 5 milyon Gigabytes hafızaya ihtiyacaç olacaktır. Buda diğer bir soru sormamıza sebep olmaktadır. Böylesi bir hafızaya sahip bilgisayar ne zaman ve nasıl üretilecek?

(Kinouchi and Mackin 2018) çalışmasında, hisseden, düşünen, hareket eden ve öğrenen (bilinçli) insansı bir robotun gerçekleştirilmesine yönelik beyin odaklı bir kontrol sistemi mimarisi önerilmiştir. Nörobilim ve psikoloji disiplinleri, beynin temel operasyonel özelliklerini gerçekleştirmek için birlikte kullanılmaktadır. İnsan davranışının alışılmış davranış ve amaca yönelik davranış olmak üzere iki farklı davranış özelliğinden oluştuğu bilinmektedir (Dezfouli, A., and Balleine 2013). Yapay sinir ağı kullanılarak ilkel bilince dayalı bağımsız öğrenen ve davranış kararları verebilen kavramsal bir kontrol sistemi önerilmiştir. Sistem düzeyinde bir işlev olarak bir bilinç modeli önerilmiş ve optimal davranışın hızlı karar vermesini sağlayan bir yapay sinir ağı modeli sunulmuştur. Sistemin çevreye uyarlanması için model, bir öğretmen veya bir gözetmen olmadan aktör-eleştirmen pekiştirmeli öğrenme yöntemini kullanmayı öğrenmiştir. Ve bu içerikler tüm sistemdeki ilgili fonksiyonel birimleri hızlı bir şekilde takip eden aksiyon kararı için iletilir. Bu faaliyetler, günlük yaşamımızda an be an kararlar verirken zihnimizin nasıl "farkında" olduğunu açıklar. Sistemi kontrol etmek için öğrenme yoluyla bir otonom adaptasyon sistemi kullanılır. Çünkü otonom adaptasyonun zaten hayvan beyninde kullanıldığı bilinmektedir. Kontrol sistemi, ödül ve cezayı dikkate alan, dikkate alınan değere göre kendi kararıyla hareket eden ve eylemin sonuçlarına göre kendini uyarlayan işlevsel bir birim içerir. Örneğin, sistem bir ödül aldığında, değerlendirme birimi hoş bir hal, ceza aldığında ise nahoş bir hal alır. Hoş ve nahoşluk derecesi, ödül ve cezanın derecesine göre değişir. Önerilen sistem modeli, beyin fonksiyonlarına yaklaşabilmek için maksimum verimlilikle çalışılmıştır. Çünkü beynimizin, tasarım olanaklarından maksimum verimi seçen bir tür optimal tasarımdan oluştuğu düşünülmektedir. (Kinouchi et al. 2018) çalışmasında, dinamik denge altında öğrenmeyi gerçekleştiren tekrarlayan sinir ağları (recurrent neural networks-RNN) kullanılarak YSA ile bilinçli bir yapay zeka sistemi önerilmiştir. YSA'lar, (Scellie,B., Bengio 2017) tarafından önerilen dinamik denge fikrine dayalı olarak yeniden tasarlanmıştır. RNN ile önerilen modelin, beynin makroskopik bir modeli olarak daha makul olduğu savunulmuştur. Fenomenal bilincin temelini oluşturan "bilinçli his" hipotezi, aynı zamanda tekrarlayan işleme teorisini (recurrent processing theory -RPT) kapsamlı bir şekilde açıklar ve bilincin küresel nöronal çalışma alanı teorisi (global neuronal workspace theory -GNWT), "sistem düzeyinde öğrenme durumu" ile aynıdır. Bu çalışmada bilinç, bilgi sistemi perspektifinden anlatılmıştır. Bilinçli bir sistemin diğer makine öğrenme yöntemleriyle birleştirilerek deneme yanılma yoluyla dinamik ortamlara uyum sağlayabileceği iddia edilmiştir.

Literatürde biyolojik ve yapay bilinci birbirine bağlamak için başka yöntemler de kullanılmıştır. Bunlardan biri de derin öğrenme yöntemi yerine mantık kullanan sinirsel bilinç akışı (neural consciousness flow-NeuCFlow) yöntemidir. Literatürde derin öğrenme ile mantık sistemleri arasında bir boşluk olduğu düşünülmektedir. Bu boşluğu kapatmak için yazarlar, bilinçsiz akış katmanı, bilinçli akış katmanı ve dikkat akışı katmanından oluşan üç katmanlı mimariye sahip bir hesaplama yöntemi önermişlerdir. Bu model, grafik sinir ağları (graph neural networks -GNN) ve koşullu geçiş matrisleri ile uygulanmaktadır. Önerilen model, bir dizi bilgi tabanı tamamlama (knowledge base completion-KBC) yöntemini çözerek bilgi grafiği

muhakemesi için de test edilebilir. Deneysel sonuçlar NeuCFlow'un gömülü ve path tabanlı sistemleri içeren son teknolojiden daha iyi olduğunu göstermektedir (Xu et al. 2019).

Ses tanıma ve görme gibi insan duyusal algılarını daha iyi kopyalayabilen alanlardaki makine bilincinin özellikleri, makine ve derin öğrenme teknikleri yardımıyla son yıllarda giderek daha fazla geliştirilmektedir (Chandra 2017). Gamez tarafından, makine bilincine ilişkin bir literatür inceleme çalışması sunulmuştur. Bu çalışmada literatür, olağanüstü bilinçli makineler, bilişsel özellikler, insan bilinci ile ilişkili mimari ve dış davranış olarak değerlendirilen dört gruba ayrılmıştır (Gamez 2008). Gamez'den sonra, Reggia tarafından temel konulardaki yinelenen temalara dayalı beş kategoride makine bilinci içeren başka bir alışma da sunulmuştur (J. A. Reggia 2013). Starzyk ve Prasad, makine bilincinin hesaplamalı bir modelini önermişlerdir (J. A. Starzyk 2011). Graziano ise dikkat ve farkındalıkla ilgilenen sosyal ve bilişsel yönlerin bilinci anlamayı desteklediğini iddia etmektedir (M. S. Graziano 2013). Graziano'nun başka bir çalışmasında, farkındalık, dikkat durumunun algısal bir yeniden inşası olarak incelenmiştir. Bu görüşe göre, diğer insanların farkındalığı hakkındaki bilgileri hesaplayan makine, kendi farkındalığımızla ilgili bilgileri hesaplayan makineyle aynı makinedir (M. S. Graziano 2011). Graziano yaptığı bu çalışmalarda "Dikkat şeması olan bir beyin, nasıl öznel farkındalığa sahip olduğu sonucuna varabilir" sorusunu cevaplamaya çalışmıştır (. S. Graziano 2014).

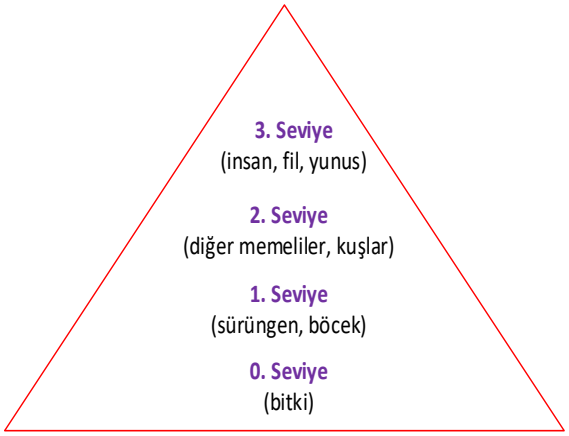
Bilinç, felsefe temeli kullanılarak makinelere dönüştürülebilir. Farkındalığın içsel bir dikkat modeli olduğu, beynin veri işleme mekanizmasının temel bir parçası olarak görülür. Beyin vücut şeması aracılığıyla bir vücut modelini nasıl hesaplar ve bunu vücudun kontrolünde nasıl kullanır? Bu nedenle, dikkat şeması aracılığıyla basitleştirilmiş bir dikkat modelinin dikkati kontrol etmede yararlı olacağı öne sürülmüştür. Lamme, birbirlerinden tamamen ayrılmış görsel dikkat ve farkındalığın tanımlarını sunarak neden karmaşık bir şekilde ilişkili olduklarını açıklamıştır. Dikkat ve farkındalık mekanizmalarından ziyade hafıza ve farkındalık mekanizmaları arasında örtüşme olduğu iddia edilmiştir (V. A. Lamme 2003).

(Chandra 2017) çalışmasında, bazı mevcut bilinç modelleri incelendikten sonra olabildiğince doğal görünen robotik sistemler gibi hisseden hesaplamalı bir model önerilmiştir. Yaratıcılığın, muhakemenin ve öz farkındalığın olmaması, yapay sistemleri veya kısmen bilince sahip robotları insanlardan oldukça farklılaştıracak ve insan işgücüne özel nitelikler kazandıracaktır (P. Dario, E. Guglielmelli 2001). (Dehaene, Lau, and Kouider 2017) çalışmasında, bilincin tanımı ve bilinç ile makine arasındaki ilişki, fizyoloji, psikoloji ve mühendislik perspektifleri ile açıklanmıştır. Bu çalışmada, "bilinç" kelimesi beyindeki iki farklı bilgi işleme hesaplamasını birleştirmektedir. Bunlar: küresel yayın için bilgi seçimi, böylece hesaplama ve rapor için esnek bir şekilde kullanılabilir hale getirilecek (C1, ilk anlamda bilinç) ve sübjektif bir kesinlik veya hata duygusuna yol açan bu hesaplamaların kendi kendine izlenmesidir (C2, ikinci anlamda bilinç). Son zamanlardaki başarılarına rağmen, mevcut makinelerin hala çoğunlukla insan beynindeki bilinçsiz işlemeyi (C0) yansıtan hesaplamalar uyguladıkları iddia edilmiştir. Bilinçsiz (C0) ve bilinçli hesaplamaların (C1 ve C2) psikolojik ve sinir bilimi gözden geçirilerek ve yeni makine mimarilerine nasıl ilham verebilecekleri açıklanmıştır. Bu çalışma basit bir hipoteze dayanmaktadır: "Bilinç" dediğimiz şey, beynin donanımı tarafından fiziksel olarak gerçekleştirilen belirli bilgi işleme hesaplamalarının sonucudur. Hesaplama özelliği açısından diğer teorilerden farklıdır. İşlenen bilginin doğası ve derinliği de dikkate alınmadıkça, salt bilgi teorik niceliklerinin bilinci tanımlamak için yeterli olmadığı düşünülmektedir. C1 ve C2'ye sahip bir makinenin sanki bilinçliymiş gibi davranacağı farz edildiğinde, bir şey gördüğünü bilir, ona güveni ifade eder, başkalarına rapor eder, izleme mekanizmaları bozulduğunda halüsinasyonlara maruz kalabilir ve hatta insanlarla aynı algısal yanılsamaları deneyimleyebilir.

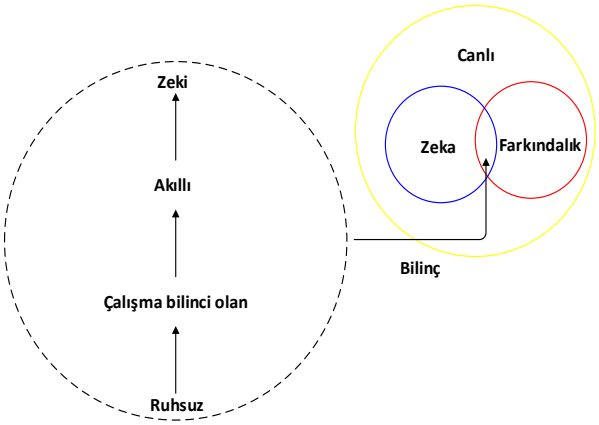
(Pandey 2018) çalışmasında yazarlar, "Yapay zekaya sahip özellikler gerçekten bilinçli olabilir mi?" sorusuna cevap bulmaya çalışmışlardır. Bu çalışmada YB, "zayıf YB" ve "güçlü YB" olmak üzere iki alt alana ayrılmıştır. Zihin ve zekanın, bilinçle yakından ilişkili olduğu ve şuana kadar yapay zekanın, yapay bilince yönelik en umut verici yol olduğu bilinmektedir. Bununla birlikte, literatürde, bilinç, yapay zeka tarafından anlaşılmaya en az yatkın dil olarak kabul edilmiştir. Küresel olarak kabul edilmiş, bilinçli akıllı makine teknolojisi tasarımları ile ilgili yapılan çalışmaların hala yetersiz olduğu bilinmektedir. Bu makalede, zihin, zeka ve yapay bilinç ile ilgili farklı teorik konular analiz edilmeye çalışılmış ve canlıların bilinç piramidinin Şekil 2'deki gibi olduğu düşünülmektedir. Ayrıca bilincin farklı hesaplama yöntemleri tartışılarak makine bilinci üretme olasılığı eleştirel bir dille analiz edilmiş ve aynı zamanda bilinçli bir makinede bulunması gereken özellikler tanımlanmıştır. Yapılan çalışmada ancak zeka ve farkındalık olgularına sahip olan canlıların bilincinin olabileceği ve bilinçli canlıların da ruhsuz, çalışma bilinci olan, akıllı ve zeki şeklinde sınıflandırılabileceği düşünülmüştür (Şekil 3). Yapılan çalışmada ayrıca "Günümüz teknolojisi ile bilinçli bir makine tasarlamak mümkün mü?", "Yapay bilinçli makineler, yapay zeka makinelerin işleyişini aşabilecek mi?" soruları tartışılarak analiz edilmiştir.

Canlı bir hücrenin akla mı yoksa zekaya mı sahip olduğu henüz tartışma konusudur, ancak bu hücrenin bilince sahip olduğu tartışılmazdır. Beynin her bir özelliğini hesaplama yolu ile tanımlamak pek mümkün değildir. Beynin bazı özellikleri bilgisayarınkine benzer olabilir, fakat bütün özellikleri benzer olmadığı bilinmektedir. Günümüzde hala araştırmacılar bilinç ve

beyin arasında bir bağlantı bulmaya çalışmaktadır ancak henüz somut bir şey bulunulmuş değildir (Moravec n.d.).



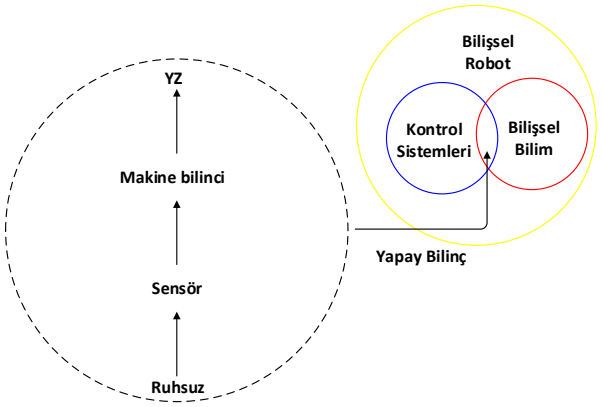
Şekil 2. Bilinç piramidinin farklı seviyelerdeki gösterimi (Pandey 2018) (Representation of the consciousness pyramid at different levels)



Şekil 3. Biyolojik bilincin hiyerarşik gösterimi (Pandey 2018) (Hierarchical representation of biological consciousness)

İlk bakışta YZ ve YB'yi ayırt etmek kolay görünse de, genel olarak YZ ile akıllı bir makine yaratılırken, YB ile bilinçli makineler yaratılmaya çalışılmıştır. Bununla birlikte, bilinç ve zeka konusu oldukça karmaşıktır ve ikisi arasındaki ayrım felsefi temellere dayanmaktadır. Aslında, geleneksel YZ'nin amacı, biyolojik olmayan bileşenler kullanılarak bilincin ilgili özelliklerini yeniden üretmektir. Yapılan çalışmada ancak kontrol ve bilişsel bilim olgularına sahip olan robotların bilincinin olabileceği ve bilinçli robotların da ruhsuz, sensörlü, makine bilinci olan ve yapay zekaya sahip şeklinde sınıflandırılabileceği düşünülmüştür (Şekil 4). Ricardo Sanz'a göre, Şekil 4'te gösterildiği gibi, YB'i modellemek için üç motivasyon vardır (Pandey 2018; Sanz R 2005). Bunlar:

1. Bilinçli makine tasarlamak (bilişsel robotik),
2. Bilinç doğasını anlamak (bilişsel bilim),
3. Bilince uyumlu kontrol sistemleri tasarlama.



Şekil 4. Yapay bilincin hiyerarşik gösterimi (Pandey 2018) (Hierarchical representation of artificial consciousness)

Bazı araştırmacılar, bilincin hesaplama ve algoritma yöntemleri ile tasarlanamayacağına inanmaktadırlar. Buna rağmen daha sonra bazı önemli bilişsel bilinç modellerinin ortaya çıktığı görülmüştür. Bu modellerden bazıları:

a) Moore/Turing kaçınılmaz modeli,
b) Hofstadter–Minsky–McCarthy modeli,
c) Daniel Dennett modeli,
d) Perlis–Sloman modeli,
e) Brian Cantwell Smith modeli.

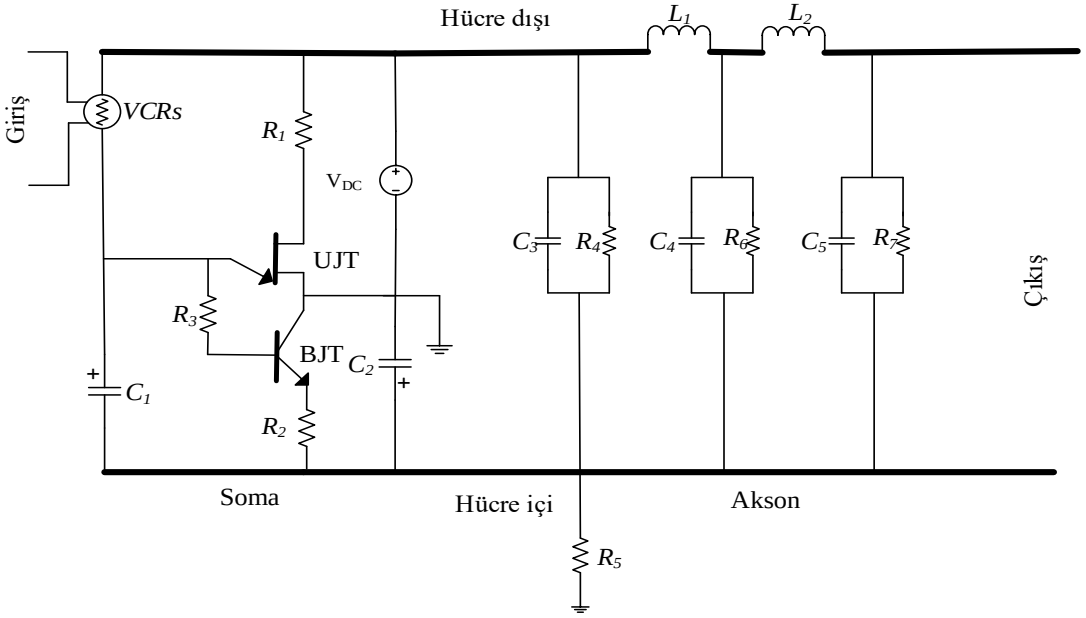
Yapılan çalışmada, beynin hesaplama hiyerarşisini kullanmadaki modülerliğinin YB oluşumunu engelleyebileceğinden bahsedilmiştir. Çünkü insan bilincinin ayrık değil sürekli bir zaman fonksiyonu olduğu bilinmektedir. Üstelik, bu durumu herhangi bir hesaplama teorisi ile açıklamak çok zordur. Çünkü, günümüz teknolojisinde kullanılmakta olan bilgisayarların henüz yeterli deneyime sahip olmadığı düşünülmektedir. Bu durum şöyle bir örnekle açıklanabilir; dijital bir klimayı kontrol sistemi kullanılarak tasarlamak mümkündür, ancak bu sistemi sıcak veya soğuk havayı hissetmesine olanak tanıyacak şekilde tasarlamak (kontrol sistemleri kullanmadan) günümüz teknoloji ile mümkün değildir. Beden, zihin, zeka ve bilinç hem birbirine bağlı hem de birbiriyle ilişkili kavramlardır. Ancak bilinç; zeka, zihin, ve bedenden anlaşılması ve tanımlanması daha zor bir kavramdır (Pandey 2018).

Bilinçli bir nano-haberleşme modelin tasarlanabilmesi için öncelikle biyolojik canlıların ürettiği sinyallerin tasarlanan bu model ile elde edilmesi gerektiği düşünülmektedir. Çünkü canlıların bütün hücrelerinin bu sinyaller (aksiyon potansiyel) aracılığı ile birbiri arasında iletişim kurduklarına inanılmaktadır. Bu sebeple bu çalışma kapsamında, biyolojik hücrelerin ürettiği aksiyon potansiyel sinyaline neredeyse birebir benzer sinyal üreten elektronik bir devre tasarlanmıştır. Daha sonra ise tasarlanan model ile iki hücrenin birbiri ile haberleşmesinden elde edilen veriler incelenmiştir.

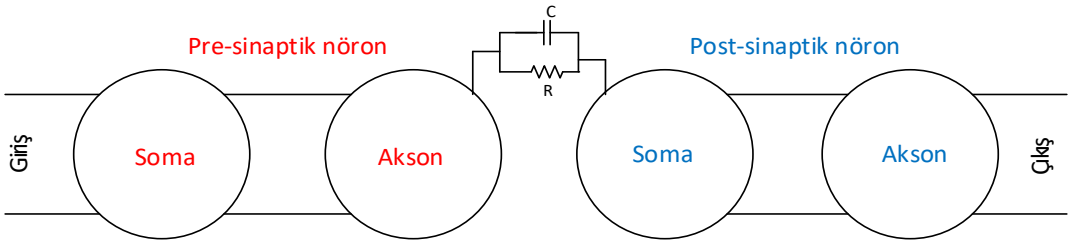
2. Materyal ve Yöntem (Material and Method)

Bu çalışma kapsamında, sinir hücrelerinin haberleşmesini aksiyon potansiyellerin üretim ve iletimi ile gerçekleştirebilen bir model analiz edilmiştir. İlk olarak elektronik tabanlı bir nöro-sinaptik haberleşme modeli tasarlanarak gerçek bir nöron hücresinin oluşturduğu aksiyon potansiyeline en yakın potansiyel elde edilmiş ve elde edilen bu veriler daha sonra analiz edilmiştir.

Başlangıçta, Şekil 5'teki nöron hücresi modeli, aksiyon potansiyeli üretmek için tasarlanmıştır. Unipolar junction transistör (*UJT*) ve diğer elektronik devre elemanları kullanılarak ilk etapta Şekil 5'de görüldüğü gibi elektronik tabanlı bir nöronun soma ve akson modeli oluşturulmuştur. Nöron hücrelerinin dışarıdan gelen uyartılar ile orantılı olarak aksiyon potansiyeli ürettiği bilinmektedir. Eğer dışarıdan gelen uyartı/tepki fazla ise üretilen aksiyon potansiyelinin sıklığı/frekansı da fazla olmaktadır. Böyle hassas bir sistemi önerilen nöron modelinde kurgulayabilmek için soma modelinin girişine dışarıdan gelen uyartıların tepkisini ölçebilmek için voltaj kontrollü bir direnç kaynağı (*VCRs*) bağlanmıştır. Bu direnç değeri giriş geriliminin değeri ile doğru orantılı olarak artıp azalabilmektedir. Dışarıdan gelen uyartının şiddeti fazla ise gerilim değeri artmakta ve dolayısıyla *VCRs*'nin değeri de artmaktadır. Sonuç olarak artan *VCRs* değeri modelden elde edilen aksiyon potansiyelinin frekansını da arttırmaktadır. Önerilen nöron soma modelinde kullanılan *UJT* açık olduğunda, hücre içi ve dışı arasına (nöron zarına) yerleştirildiği varsayılan elektronik model de aktif olur ve zardan geçen akım maksimum olur. Böylece elektronik tabanlı soma modeli darbe üretir. *UJT*'nin açık olması Bipolar junction transistör (*BJT*)'yi de aktif eder ve böylece soma zarındaki iyon kanallarının iyonik kontrolü gibi aksiyon potansiyeli de şekillenmiş olur.



Şekil 5. Tasarlanan nöron modeli (Designed neuron model)



Şekil 6. Pre ve post-sinaptik hücrelerin birleşim bölgesi olan kleftin birbirine paralel bağlı bir direnç ve kapasite ile modellenmesi (Modeling of the kleft, the junction of pre and post-synaptic cells, with a parallel interconnected resistance and capacity)

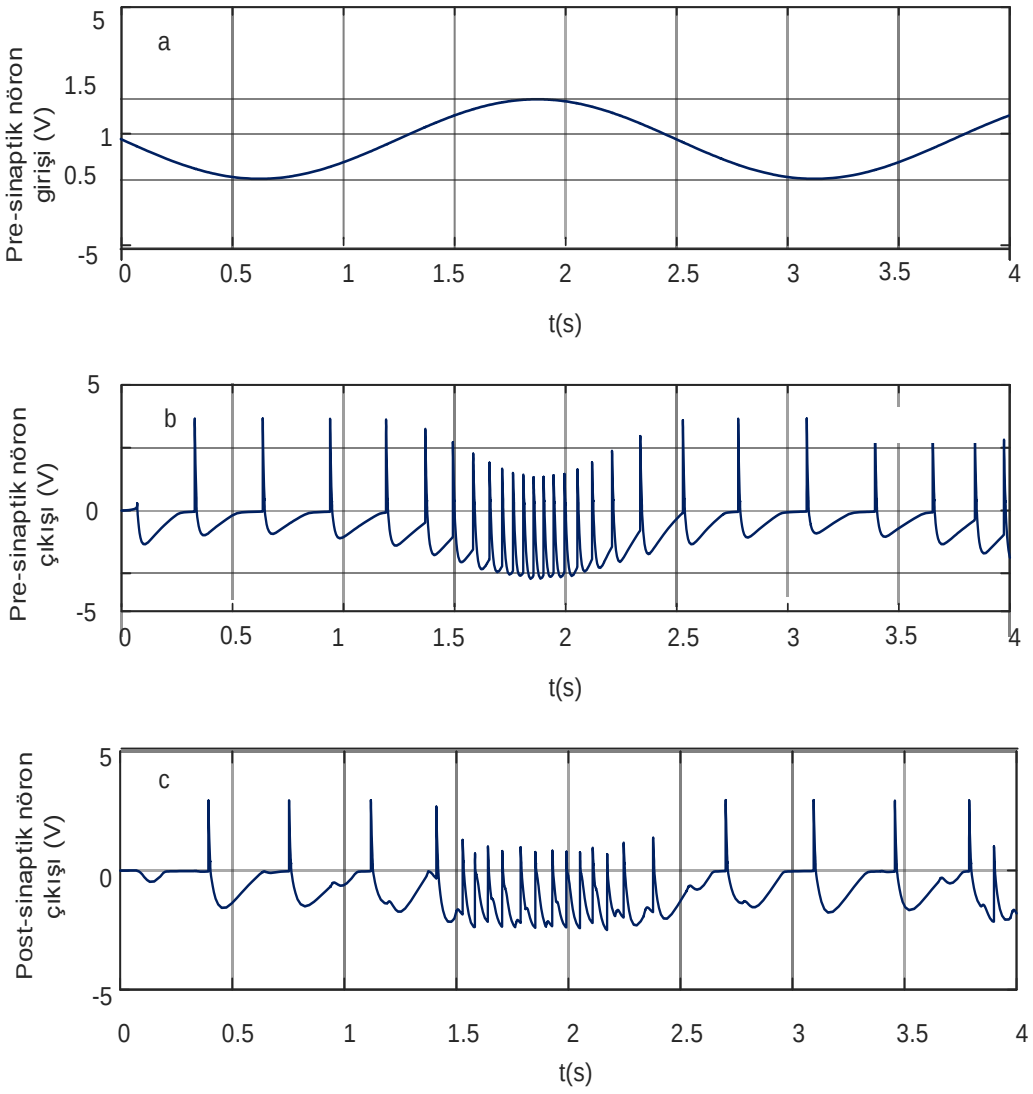
Daha sonra, ise iki nöron hücresinin birbirleri ile (aksiyon potansiyelleri kullanılarak) haberleşebilmesi için Şekil 6'daki gibi bir kleft modeli oluşturulmuştur. Şekil 6'dan görüldüğü gibi önerilen kleft modeli, en basit hali ile bir direnç ve kondansatörden oluşmaktadır. Çünkü iletilen bilgi aynen bir kondansatörün şarj ve deşarj durumu gibidir. Böylece aksiyon potansiyellerinin kleft'teki iletimi bir kondansatörün

şarj ve deşarj durumuna benzetilebilir (Savtchenko, Poo, and Rusakov 2017; Singh and Bal 2017).

3. Bulgular (Results)

Elde edilen nöro-sinaptik haberleşme modeli kullanılarak Şekil 7.a'dan görüldüğü üzere ilk-sinaptik nöronun girişine sinüzoidal bir giriş verilmiş ve Şekil 7.b'de görüldüğü gibi farklı frekanslarda aksiyon potansiyelleri ilk-sinaptik nöronun çıkışında elde edilmiştir. İlk-sinaptik nöronun çıkışı son-sinaptik nöronun girişine bağlanınca çıkışta Şekil 7.c'de görüldüğü üzere küçük bir zaman kayması ve kayıp ile son-sinaptik nöronun akson terminalinde aksiyon potansiyelleri tekrar elde edilmiştir. Girişe 1 V DC eklenmiş genliği 0.5 V olan kare dalga verilince ise Şekil 8'den görüldüğü üzere çıkışta kare dalganın sadece pozitif tepe değerlerinde (eşik değerini geçen değerlerde) eşit frekanslı aksiyon potansiyeller elde edilmiştir.

Bu bağlamda, ilk olarak literatürde bir örneği olmayan elektronik tabanlı soma, akson ve nöro-sinaptik haberleşme modelleri tasarlanarak gerçek bir nöronunun oluşturduğu aksiyon potansiyeline en yakın potansiyel elde edilmiş ve elde edilen bu veriler daha sonra Matlab benzetim programı ile analiz edilmiştir. Bu bakış açısı ile ilk olarak nöron hücrelerinin bilgi transferinde kullandığı aksiyon potansiyeli elektronik tabanlı olarak tasarlanan bir model ile elde edilmiş ve daha sonra bu aksiyon potansiyeli yine elektronik tabanlı olarak tasarlanan modelinin uçlarına başarılı bir şekilde iletilmiştir. Üretilen aksiyon potansiyelin gerçek bir nöron hücresinin ürettiği aksiyon potansiyele yakın olabilmesi için literatürdeki elektronik modellerde kullanılan elemanlar incelendikten sonra böyle bir sistem geliştirilmiş ve bu sistemde kullanılan parametrelerin değeri de yine üretilen aksiyon potansiyelin benzerliğini arttıracak şekilde uzun süren denemeler sonunda hassas bir şekilde tespit edilmiştir.

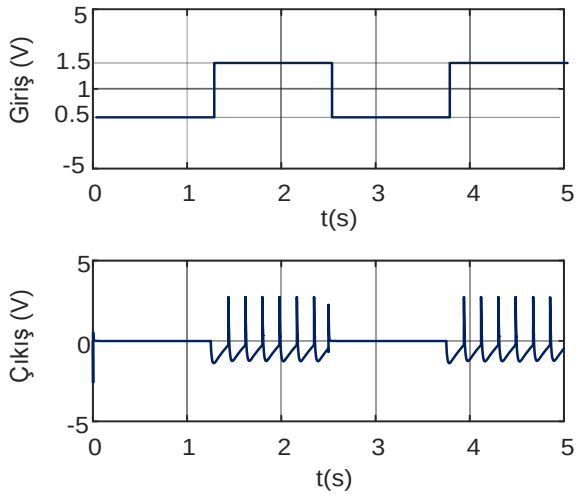


Şekil 7. a) Nöron modelinin girişine 0.5 V genliğe sahip bir sinüzoidal dalga verilince b) pre-sinaptik nöron modelinin çıkışı ve c) post-sinaptik nöron modelinin çıkışı (a) When a sinusoidal wave with an amplitude of 0.5 V is given to the input of the neuron model b) the output of the pre-synaptic neuron model and c) the output of the post-synaptic neuron model)

Önerilen nöro-sinaptik haberleşme modelinde kullanılan sistem parametreleri Tablo 1'de verilmiştir.

Tablo 1. Elektronik tabanlı nöro-spike haberleşme modelinde kullanılan sistem parametreleri (System parameters used in the electronic based neuro-spike communication model)

Parametre	Değeri
R_1	2.2 $k\Omega$
R_2	100 Ω
R_3	100 $k\Omega$
R_4, R_6, R_7, R	1 $k\Omega$
R_5	10 $k\Omega$
C_1	10 μF
C_2	2 μF
C_3, C_4, C_5, C	1 nF
L_1, L_2	10 μH
V_{DC}	10 V
VCR_S	17.2 kΩ ~ 65 kΩ
V_{AC}	1sin(2π0.4)+1



Şekil 8. Nöron modelinin girişine 1 V DC değere 0.5 V genliğe sahip bir kare dalga eklenerek verilince nöron modelinden elde edilen çıkış grafiği (Output graph obtained from the neuron model when given by adding a square wave with 0.5 V amplitude to 1 V DC value at the input of the neuron model)

4. Sonuçlar (Conclusions)

Yapay zeka alanındaki son gelişmeler, insanlar gibi öğrenen ve düşünen sistemler (YB) oluşturmaya olan ilgiyi arttırmıştır. Bu alanda yapılan çalışmalar, eğitilmiş derin sinir ağlarının nesne tanıma, video oyunları ve masa oyunları gibi görevlerde bazı açılardan insanlara eşit hatta onları yenen performansa sahip sistemlerin geliştirilmesi ile günümüzde araştırmacılar tarafından giderek artan bir şekilde çalışılmaktadır. Biyolojik esinli ve performans başarılarına rağmen, bu sistemler birçok yönden insan zekasından farklıdır (Lake et al. 2017). İnsan duyusal algılarını daha iyi kopyalayabilen alanlardaki makine bilincinin özellikleri, makine ve derin öğrenme teknikleri yardımıyla son yıllarda giderek daha fazla geliştirilmektedir. Bilincin, felsefe temeli kullanılarak makinelere dönüştürülebileceğine inanılmaktadır.

Bu bağlamda, elektronik tabanlı olarak tasarlanan nöro-sinaptik haberleşme modelinin bilinçli bir yapay zeka modeli geliştirmek için gerekli aşamaları bu çalışmada sunulmuştur. Önerilen elektronik tabanlı nöron modeli sadece UJT ve BJT gibi basit devre elemanları kullanılarak tasarlanmış ve sonuçta gerçek bir nöronun ürettiği aksiyon potansiyeline çok yakın bir aksiyon potansiyeli üreten devre tasarlanmış ve sonuçları burada verilmiştir. Daha sonraki çalışmalarda ise tasarlanan modelin bilinçli hale getirilmesi üzerine çalışmalar yapılacaktır.

Kaynaklar (References)

Grazian, S., Webb. T. 2014. "A Mechanistic Theory of Consciousness." *International Journal of Machine Consciousness* 6(2).

Lamme, V. A. 2003. "Why Visual Attention and Awareness Are Different." *Trends in Cognitive Sciences* 7(1).

Akan, O. B. n.d. "Icimizdeki Internet Molekuler Haberlesme ve Nanoaglar." Retrieved February 10, 2018 (http://panorama.khas.edu.tr/icimizdeki-internet-molekuler-haberlesme-ve-nanoaglar-154).

Amanda Sharke, Noel Sharkey. 2010. "Granny and the Robots: Ethical Issues in Robot Care for the Elderly." *Ethics and Information Technology* 14.

Banfield, Jeffrey D., and Adrian E. Raftery. 1992. "Ice Floe Identification in Satellite Images Using Mathematical Morphology and Clustering about Principal Curves." *Journal of the American Statistical Association* 87(417):7–16.

Barreto, Guilherme A., Aluizio F. R. Araújo, Christof Dücker, and Helge Ritter. 2002. "A Distributed Robotic Control System Based on a Temporal Self-Organizing Neural Network." *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 32(4):347–57. doi: 10.1109/TSMCC.2002.806067.

Bilal Er, Mehmet, and Ibrahim Berkan Aydilek. 2019. "Music Emotion Recognition by Using Chroma Spectrogram and Deep Visual Features." *International Journal of Computational Intelligence Systems* 12(2):1622–34. doi: 10.2991/ijcis.d.191216.001.

Buttazzo, Giorgio. 2008. "Artificial Consciousness: Hazardous Questions (and Answers)." *Artificial Intelligence in Medicine* 44(2):139–46. doi: 10.1016/j.artmed.2008.07.004.

Chandra, Rohitash. 2017. "Towards an Affective Computational Model for Machine Consciousness." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10638 LNCS:897–907. doi: 10.1007/978-3-319-70139-4_91.

Dehaene, Stanislas, Hakwan Lau, and Sid Kouider. 2017. "What Is Consciousness, and Could Machines Have It?" *Science* 358(6362):486–92. doi: 10.1126/science.aan8871.

Dezfouli, A., and Balleine, B. 2013. "Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized." *PLoS Comput. Biol*.

Eter, P., and Andr As. 2002. "KERNEL-KOHONEN NETWORKS." 12(2):117–35.

Gamez, D. 2008. “Progress in Machine Consciousness.” *Consciousness and Cognition* 17(3).

Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. 2000. *Molecular Cell Biology*. Vol. 29.

J. A. Reggia. 2013. “The Rise of Machine Consciousness: Studying Consciousness with Computational Models.” *Neural Networks* 44.

J. A. Starzyk, D. K. Prasad. 2011. “A Computational Model of Machine Consciousnes.” *International Journal of Machine Consciousness* 3(2).

Kak, Subhash. n.d. “Https://Subhashkak.Medium.Com/Artificial-Intelligence-and-Consciousness-6b5ff2e5b5a.”

Kinouchi, Yasuo, and Kenneth James Mackin. 2018. “A Basic Architecture of an Autonomous Adaptive System with Conscious-like Function for a Humanoid Robot.” *Frontiers Robotics AI* 5(APR). doi: 10.3389/frobt.2018.00030.

Kinouchi, Yasuo, Kenneth James MacKin, and Pitoyo Hartono. 2018. “A Conscious AI System Based on Recurrent Neural Networks Applying Dynamic Information Equilibrium.” *CEUR Workshop Proceedings* 2287(December). doi: 10.29007/2hjj.

Lake, Brenden M., Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2017. “Building Machines That Learn and Think like People.” *Behavioral and Brain Sciences* 40(2012):1–58. doi: 10.1017/S0140525X16001837.

M. S. Graziano. 2013. *Consciousness and the Social Brain*. Oxford University Press.

M. S. Graziano, S. Kastner. 2011. “Human Consciousness and Its Relationship to Social Neuroscience: A Novel Hypothesis.” *Cognitive Neuroscience* 2(2).

Marchetti, Giorgio. 2018. “Consciousness: A Unique Way of Processing Information.” *Cognitive Processing* 4(19).

Moravec, Hans P. n.d. *Mind Children: The Future of Robot and Human Intelligence*. Harvard Un.

Neukart, Florian, Sorin Aurel Moraru, Costin Marius Grigorescu, and Peter Szakacs-Simon. 2012. “Cortical Artificial Neural Networks and Their Evolution - Consciousness-Inspired Data Mining.” *Proceedings of the International Conference on Optimisation of Electrical and Electronic Equipment, OPTIM* 1126–33. doi: 10.1109/OPTIM.2012.6231782.

P. Dario, E. Guglielmelli, C. Laschi. 2001. “Humanoids and Personal Robots: Design and Experiments.” *Journal of Robotic Systems* 18(12).

Pandey, Subhash Chandra. 2018. “Can Artificially Intelligent Agents Really Be Conscious?” *Sadhana - Academy Proceedings in Engineering Sciences* 43(7):1–17. doi: 10.1007/s12046-018-0887-x.

Sanz R. 2005. “Design and Implementation of an Artificial Conscious Machine.” in *IWAC2005, Proceedings of*.

Savtchenko, Leonid P., Mu Ming Poo, and Dmitri A. Rusakov. 2017. “Electrodiffusion Phenomena in Neuroscience : A Neglected Companion.” *Nature Publishing Group* 18(10):598–612. doi: 10.1038/nrn.2017.101.

Scellie,B., Bengio, Y. 2017. “Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation.” *Front. Comput. Neurosci* 11(24).

Simon Peter van Rysewyk, Matthijs Pontier. 2015. *Machine Medical Ethics*. Springer, Cham.

Singh, Sushant, and Naresh C. Bal. 2017. “Membrane Biophysics.” *Introduction to Biomolecular Structure and Biophysics: Basics of Biophysics* 183–204. doi: 10.1007/978-981-10-4968-2_7.

Xu, Xiaoran, Wei Feng, Zhiqing Sun, and Zhi-Hong Deng. 2019. “Neural Consciousness Flow.” 1–30.

Yamazaki, Kimitoshi, Yoshiaki Watanabe, Kotaro Nagahama, Kei Okada, and Masayuki Inaba. 2010. “Recognition and Manipulation Integration for a Daily Assistive Robot Working on Kitchen Environments.” *2010 IEEE International Conference on Robotics and Biomimetics, ROBIO 2010* (October 2016):196–201. doi: 10.1109/ROBIO.2010.5723326.

Yin, Hujun. 2008. “The Self-Organizing Maps: Background, Theories, Extensions and Applications.” *Studies in Computational Intelligence* 115:715–62. doi: 10.1007/978-3-540-78293-3_17.

Zhao, Tong, Yiqian Zhu, Hailiang Tang, Rong Xie, Jianhong Zhu, and John H. Zhang. 2019. “Consciousness: New Concepts and Neural Networks.” *Frontiers in Cellular Neuroscience* 13(July):1–7. doi: 10.3389/fncel.2019.00302.