AITA

2024 Volume 4 Number 1

ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

RESEARCH ARTICLES

https://aitajournal.com

info@aitajournal.com

ISSN: 2757-9778

ISBN: 978-605-69730-2-4

ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

Volume 4 – Issue 1

01.05.2024

e-ISSN: 2757-9778

Publisher: İzmir Bakırçay University

www.aitajournal.com

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

Journal Board

Editorial Board

International Advisory Board

International Indexes

ACADEMIC RESOURCE INDEX: RESEARCH BIBLE

INDEX COPERNICUS INTERNATIONAL

SOBIAD CITATION INDEX

ASIAN SCIENCE CITATION INDEX

INTERNATIONAL INSTITUTE OF ORGANIZED RESEARCH

About Journal

Artificial Intelligence Theory and Applications (AITA) provides coverage of the most significant work on principles of artificial intelligence, broadly interpreted. The scope of research we cover encompasses contributions of lasting value to any area of artificial intelligence. To be accepted, a paper must be judged to be truly outstanding in its field. AITA is interested in work in core artificial intelligence and at the boundaries, both the boundaries of sub-disciplines of artificial intelligence and the boundaries between artificial intelligence and other fields.

Scope

The best indicator of the scope of the journal is provided by the areas covered by its Editorial Board in theoretical (artificial intelligence and computing methodologies) and practical (artificial intelligence applications and applied computing) ways. These areas change from time to time, as the field evolves.

Year : 2024
Volume : 4
Issue : 1

ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

Table of Content — Page

■ RESEARCH ARTICLE

Heart Disease Diagnosis via Web Based Classification Software programmed with Julia Programming Language

Hüseyin Kutlu [a†], **Emek Güldoğan**[b], **Cemil Çolak**[b]

[a]Department of Computer Technologies, Adıyaman University, Adıyaman, Türkiye
[b]Department of Biostatistics and Medical Informatics, Inonu University, Malatya, Türkiye
† hkutlu@adiyaman.edu.tr, corresponding author

RECEIVED JULY 7, 2023
ACCEPTED NOVEMBER 1, 2023

CITATION Kutlu, H., Güldoğan, E., Çolak, C. (2024). Heart disease diagnosis via web based classification software programmed with Julia programming language. *Artificial Intelligence Theory and Applications, 4(*1), 1-10.

Abstract

In recent years, various tools and algorithms have been proposed and continue to be proposed by researchers to develop highly successful medical decision support systems. However, the clinical use of these algorithms is very limited due to various limitations. Making the necessary software installations to run the algorithm and lack of programming knowledge is some of these restrictions. In this study, a web-based classification software developed with the Julia programming language, which can be used by physicians in their medical research and clinical decisions, is introduced. Through this software, coronary artery disease detection was performed with the Cleveland heart disease database, which is a publicly accessible data set. The dataset was classified with eight different classifiers (KNN, SVM, DT, RF, AdaBoost, Gauss Naive Bayes, LDA, LR) supported by the software. The metrics obtained by 10-fold cross-validation of the data set are reported. The SVM classifier achieved the highest classification accuracy with 86.44%. The software proposed in this study may assist clinicians in research and patient identification.

Keywords: Julia; dash; classification; heart disease; web-based artificial intelligence software

1. Introduction

Computerized clinical decision support systems are an important part of healthcare today. A clinical decision support system takes patient information as input. With these inputs, it aims to improve medical decisions and improve health care delivery [1]. Clinical decision support systems support clinicians in their scientific studies and complex decision-making processes [2]. Advances in computer technologies have led to rapid development of clinical decision-making systems since the first use of clinical decision-making systems. Pharmaceutical databases, electronic records of patients, and international open databases have contributed to the development of these systems. Machine learning-based artificial intelligence algorithms are presented as a decision support system to help clinicians predict patient outcomes [3].

It can be said that there are various difficulties in the use of decision support systems that have emerged in recent years by clinicians. Decision support systems are usually

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

connected to a computer system. This situation causes the system to be inaccessible from anywhere at all times. Every algorithm developed cannot always turn into a decision support system. In this case, programming knowledge is needed to code the algorithm. In order to code these programs, various software installations and data to train and test this software are needed. It is also inevitable to have a computer equipped with the equipment where these algorithms can be trained and run. In this study, web-based classification software implemented with the Julia programming language, which is a decision support system that can be run from any platform and which does not need to set up a program and have programming knowledge, is introduced. The software was developed by Inonu University, Department of Biostatistics and Medical Informatics.

In order to benefit from the speed of the Julia programming language, the web-based classification software was implemented in the Julia programming language [4]. Web-based classification software and the Cleveland heart disease database [5], a publicly accessible dataset, were used. With this database, the system has been trained and tested with the aim of detecting coronary artery disease.

Heart disease is often referred to as coronary artery disease. Coronary artery disease is a broad term that can refer to any condition that affects the heart. Coronary artery diseases can be defined as all kinds of disorders such as infections affecting the heart, genetic disorders, vascular disease, heart valve disease. Coronary artery disease is the most common form of cardiovascular disease and is the leading cause of heart attacks [6]. There are many factors that cause coronary artery diseases. While some of these factors cannot be changed (age, gender, family history), some of them can cease to be risk factors by changing their lifestyle (smoking, alcohol, physical activity, etc.). Coronary artery disease can be detected by symptoms of chest pain and fatigue while in some people, it shows no symptoms [7]. Therefore, it is necessary to monitor the symptoms that cause heart disease, and the risk status should be followed up by the physician and the patient. The primary method used to diagnose heart disease is angiography. However, it is very costly and requires technical experience [8]. Apart from this, various techniques such as blood pressure monitoring, echocardiogram, electrocardiography, electrophysiological examinations, myocardial perfusion scans and tilt table test are performed to diagnose heart disease [9].

Machine learning algorithms play a dominant role in diagnosing heart disease. Machine learning algorithms have the advantage of extracting necessary information from large amounts of data. Many studies have been carried out in the field of machine learning with the Cleveland heart disease database. Javeed et al. [10] developed an efficient and less complex model to improve coronary heart disease risk estimation using a random search algorithm and a Random Forest (RF) model. Using the 7-element subset of the features, they achieved an accuracy of 93.33%. The model showed a 3.3% improvement over standard RF. Pasha et al. [11] proposed a new feature reduction (NFR) model for effective heart disease risk estimation in Cleveland, Hungary, Statlog, and Switzerland datasets. They replaced the missing values with the average values of the column. Logistic Regression (LR), Random Forest (RF), Boosted Regression Tree (BRT), Stochastic Gradient Boosting (SGB) and Support Vector Machine (SVM) classification algorithms were used in the study. Classification metric values of the algorithms used were calculated. By comparing the metric values, the algorithm with the highest classification performance was determined. Using LR (9 features) on the Cleveland dataset, they reported an accuracy of 92.53% and an AUC of 0.9268. Saqlain et al. [12] proposed a feature subset selection method to improve cardiovascular risk prediction results using feature selection algorithms (MFSFSA), forward feature selection algorithm (FFSA), and reverse feature selection algorithm (RFSA) based on average fisherman score. They classified the feature subsets with the RBF kernel-based SVM classifier. They tested the proposed model on Cleveland and different data sets. They achieved an accuracy of 81.19% on the Cleveland dataset with seven features. Muhammed et al. [13] proposed an intelligent prediction model for early detection of heart disease by training various machine learning classifiers on the best features of the Cleveland dataset using 10-fold cross validation. The researchers applied four feature selection algorithms: fast correlation-based filter (FCBF), minimum redundancy maximum relevance (mRMR), LASSO, and Relief to obtain key and more relevant features in the study. Researchers have achieved 94.41% accuracy with the Extra Tree classifier in the study. Ali et al. [14] used a 70:30 ratio validation for the training and test datasets to create an autonomous diagnostic system for heart disease identification using an enhanced deep neural network (DNN) and chi-square feature selection for classification in the Cleveland dataset. In the test dataset, they reported the accuracy of the proposed hybrid model as 93.33% and the AUC value as 0.94. Gupta et al. [15] obtained 92.30% classification accuracy using standardized data Logistic Regression in their study by dividing the Cleveland heart dataset for training and testing by a ratio of 70:30. They also obtained the best classification accuracy by testing the KNN classifier with a k value between 2 and 20 and 90.11% at k = 14.

In this study, web-based classification software coded with Julia programming language was used. The software supports eight different classifiers such as K - Nearest neighbors, SVM, Decision tree classifier, Random Forest Classifier, AdaBoost classifier, Gaussian Naive Bayes Classifier, LDA classifier, and Logistic Regression Classifier. The dataset was trained and classified with all the classifiers supported by the software. The software split the data set in a 10-fold cross-validation manner and presented the classification results to the user with tabular metrics.

2. Materials and Methods

Web-based classification software developed with the Julia programming language was used in the study. With the application in question, the heart disease classification model was trained and tested using the Cleveland heart disease database. The application contains eight different classification algorithms. In this section, the software, the dataset and the classifiers are introduced.

2.1. Dataset

The Cleveland heart disease database is an open access database [16]. The database contains 303 observations, 297 of which are complete observations and six observations with incomplete data. The database has 76 features. In studies conducted with the data set in the literature, missing observations are generally removed from the data set. In this way, the data set includes 137 patient (1) and 160 healthy (0) observations. In studies conducted in the literature with the data set, 13 features were generally used to increase the classification performance. In this study, in order to compare the classification performance of the web software with existing studies, missing observations were removed from the dataset and 13 features were used. The features used are described in Table 1.

Table 1. Details of the Cleveland heart disease database

Features		Explanation	State
1.	**Age**	Numeric	input
2.	**Sex**	0: Female, 1: Male	input
3.	**Cp:** chest pain	0: typical angina, 1: atypical angina, 2: non-anginal pain 3: asymptomatic	input
4.	**trestbps:** resting blood pressure (blood pressure in mm Hg at the time of admission to hospital)	Numeric	input
5.	**kol:** Cholesterol value in mg / dl measured with the BMI sensor	Numeric	input
6.	**fbs:** fasting blood sugar	0: < 120 mg / dl 1: > 120 mg / dl	input
7.	**restecg:** resting electrocardiographic results	0: normal 1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression > 0.05 mV) 2: Demonstration of probable or definite left ventricular hypertrophy according to Estes criteria	input
8.	**thalach:** maximum heart rate	Numeric	input
9.	**exang:** angina (compression) due to exercise	Numeric	input
10.	**oldpeak:** Exercise-induced ST depression at rest	Numeric	input
11.	**egim:** hill exercise ST segment slope	1: upsloping 2: straight 3: sloping down	input
12.	**ca:** number of major vessels colored by fluoroscopy (for calcification of vessels)	0: No occluded vessel 1: 1 vessel 2: 2 vessel 3: 3 vessel	input
13.	**thal:** results of nuclear stress test	1: normal; 2: fixed defect; 3: reversible defect	input
14.	**num:** Target variable representing the diagnosis of heart disease (angiographic disease status) in any major vessel	0: < 50% diameter reduction 1: > 50% diameter reduction	output (target)

2.2. Web Based Data Classification Software Programmed with Julia Programming Language

The interface of our data classification software [18], which is under the title of Julia software among Data Science and Artificial Intelligence Based Web Software [17], developed by our department is shown in Figure 1.

Figure 1. Interface of Interactive Web Software

Association Rules Mining, Data Classification Software, Cluster Analysis and Regression Analysis software are available in the Julia Web Software menu. Figure 2 shows the interface of the Classification Software.

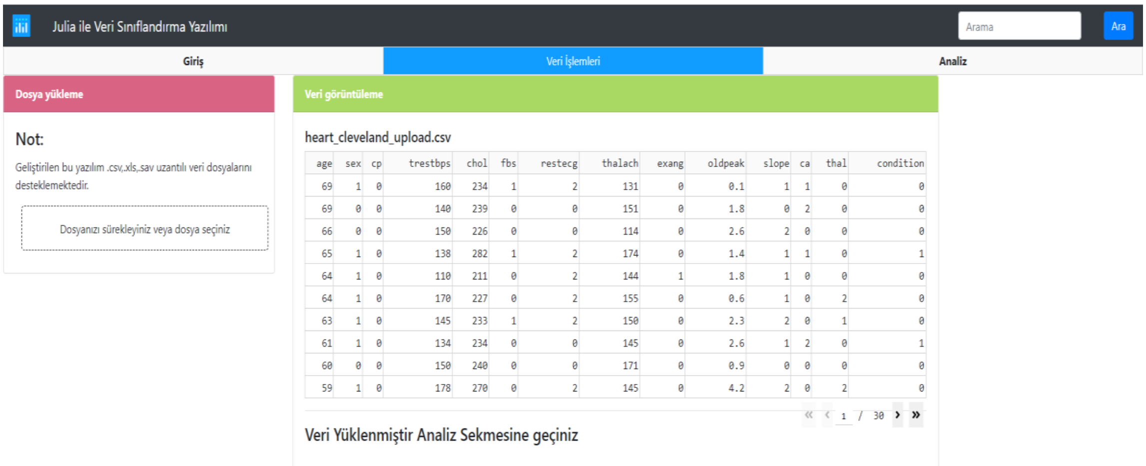

Figure 2. Interface of web-based Data Classification Software programmed with the Julia programming language

The software consists of three menus. The software is introduced in the login menu. File (csv, xlsx, sav) loading and data display operations are performed in the data operations menu. The Analysis menu view of the software is shown in Figure 3.

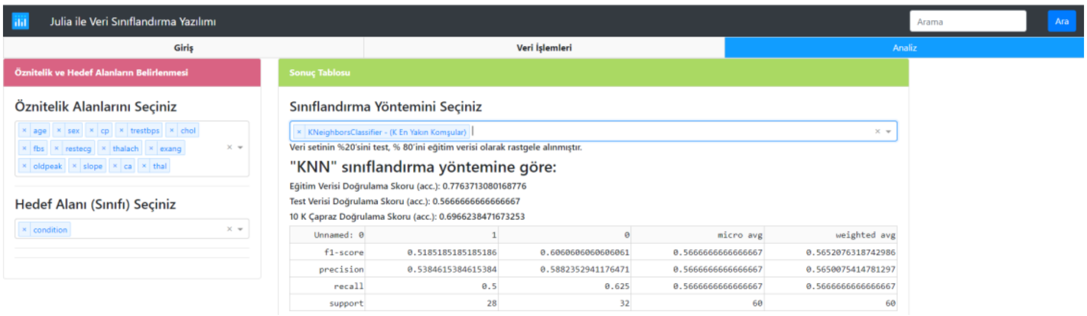

Figure 3. Analysis Menu View of web-based Data Classification Software programmed with Julia programming language

In the analysis menu shown in Figure 3, the feature fields in the data set can be selected. In this way, it can be determined how much which feature affects the model. Besides, the target area can also be selected. From the result table, the classification method (KNN, SVM, DT, RF, AdaBoost, Gauss Naive Bayes, LDA, LR) is selected. Then, classification accuracies of training and test datasets and classification metrics such as F1 Score, Precision, Recall of classes are presented to the user in tabular form.

K-Nearest Neighbours (KNN) Classifiers: KNN is based on estimating the class of the sample based on the information in which class the nearest neighbours of the vector formed by the independent variables are dense. The KNN algorithm makes predictions on two basic parameters; The Distance parameter represents the distance of the point to be estimated from other points. There are different distance calculation algorithms such as Euclid, Minkowski. The K (neighbourhood number) parameter is the parameter that tells how many nearest neighbours will be calculated over [19].

Support Vector Machine (SVM): SVM classifiers classify with the supervised learning method. It relies on drawing a line or hyperplane to separate points placed on a plane. It aims to have this line at the maximum distance for the points of both classes [20].

Decision Tree (DT): DT is one of the tree-based learning algorithms. It is among the supervised learning algorithms. They classify the dataset by dividing it into smaller sets by applying a set of decision rules.

Random Forest (RF) algorithm: RF is an algorithm that produces and classifies multiple decision trees by training each one on a different observation sample. The algorithm creates a decision tree for each sample, and the estimated value result of each decision tree is formed. Voting is performed for each value formed as a result of the prediction. Observation is assigned to the class with the most votes [22].

AdaBoost Classifier: AdaBoost Classifier is one of the Ensemble Learning methods. Boosting is to create a strong learner by bringing together many weak learners and training them cumulatively. In the Adaboost classification model, the training set is first trained with a weak learner. Learners who make incorrect predictions after training are important for the AdaBoost algorithm. In the next training, the incorrectly learned training data in the first estimation is retrained by giving more priority, that is, by increasing the weights. The results are combined by training the weak learner output to be the input to the other learner. In this way, it performs the classification process [23].

Gaussian Naive Bayes Classifier: It makes use of Bayes Theorem during the training phase. According to the conditional independence assumption, each feature is handled independently. In this way, the number of parameters to be estimated is considerably reduced. Probability values are calculated for the algorithm to work. These are the probabilities of each class in the training dataset and the conditional probabilities of each input value given each class value. In the Gaussian Naive Bayes algorithm, in addition to the probabilities of each class, the mean and standard deviation values of each class are also calculated and classification is made.[24]

Linear Discriminant Analysis (LDA): The LDA Classification algorithm is based on developing a probability model per class based on the particular distribution of observations for each input variable. It works by calculating summary statistics for input properties by class label, such as mean and standard deviation. These statistics represent the model learned from the training data. In practice, linear algebra operations are used to efficiently calculate required quantities via matrix decomposition. Estimates are made by estimating the probability of a new instance of each class label based on the values of each input attribute. The class that results most likely is then assigned to the instance. Therefore, LDA can be thought of as a simple application of Bayes' Theorem for classification. LDA assumes that the input variables are numeric and normally distributed and have the same variance (spread). If this is not the case, it may be desirable to transform the data to have a Gaussian distribution and standardize or normalize the data prior to modeling [25].

Logistic Regression Classifier (LR): LR is a data analysis technique that uses mathematical calculations to find relationships between two attributes. LR then classifies using the mathematical relationship it establishes to estimate the value of the target variable.

3. Results

Eight classification algorithms supported by the developed software were trained and tested with the data set. The models were compared with the performance metrics supported by the software. The results are obtained with the 10-fold Cross validation method as default by the software.

3.1. Performance Metrics

Performance metrics supported by the software are explained and listed.

Accuracy (Acc.): Acc. is the ratio of all classification predictions to the number of successfully predicted data. FN and FP represent the number of incorrect predictions of classes with each other. TP and TN represent the number of observations for which classes were predicted correctly. With this information, Accuracy is calculated as shown in Equation (1).

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision (P): P is a probability measure that evaluates the probability of a positive prediction being correct, as shown in Equation (2).

$$P = \frac{TP}{TP + FP} \tag{2}$$

Recall (R): R is the proportion of positively correctly predicted samples belonging to the positive class and the formula is as stated in Equation (3).

$$R = \frac{TP}{TP + FN} \tag{3}$$

F1 Score: Often referred to as the F measure. The F1 score is a measure used to determine the accuracy of a test. Calculates the score taking into account the precision P of the test and the recall R. In order not to make an erroneous model selection in unequally distributed data sets, the F measure is sometimes used instead of Accuracy. The formula for measure F is as stated in Equation (4)

$$f_1 = 2 * \frac{P * R}{P + R} \tag{4}$$

3.2. Experimental Results

The presence of heart disease was estimated in this study. Estimation was carried out with different classification algorithms. The classification metrics of the algorithms were calculated. In Table 2, classification performance metrics are shown together with the classifier parameters.

Table 2. Classification performance metrics

		Metrics		
Methods	**Acc. (%)**	**F1 Score (%)**	**P (%)**	**R (%)**
KNN (K=5)	69,66	56,22	56,33	56,25
SVM (Kernel: Lineer C=0.025)	86,44	74,84	74,91	74,77
DT (MaxDepth:5)	73,37	64,75	64,81	64,73
RF (MaxDepth:5, n_estimators:10, max_features:1)	79,79	71,59	71,57	71,65
AdaBoost	82,31	69,96	70,00	70,08
Gaussian Naive Bayes	85,23	76,74	77,60	77,23
LDA	86,06	73,33	73,66	73,66
LR	86,02	73,30	73,33	73,44

As can be seen in Table 2, the SVM classifier (86.44%) achieved the highest classification accuracy. Figure 4 shows the graph comparing classifier performances.

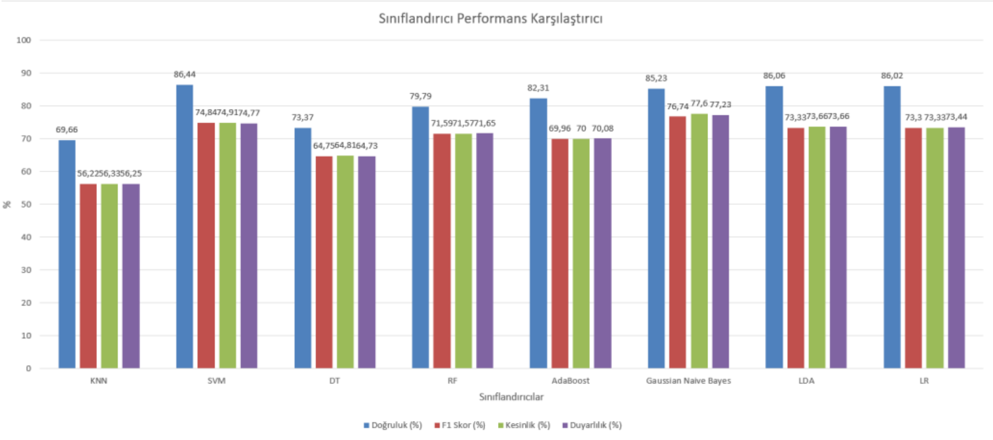

Figure 4 – Classifier performances

4. Conclusion

In this study, a software has been proposed for clinicians to use machine learning methods, which have been successfully presented in the literature, in their research and treatment. The software is a web-based software. It is programmed with the julia language. The software enables the use of classification algorithms without the need for program installation and programming knowledge. It aims to enable clinicians to analyze by simply uploading their data. In the study, the SVM classifier came to the fore with a classification accuracy of 86.44% in the detection of heart disease. The proposed software is under development and in the next versions, 10-fold cross-validation will be user-configurable. In the future, missing and excessive values in the data are planned to be detected by the software and removed from the user-approved data set. In addition, it is considered that the user-approved data set normalization phase will be carried out.

References

[1] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An overview of clinical decision support systems: benefits, risks, and strategies for success,” NPJ Digit Med, vol. 3, no. 1, p. 17, Feb. 2020, doi: 10.1038/s41746-020-0221-y.

[2] I. Sim et al., “Clinical Decision Support Systems for the Practice of Evidence-based Medicine,” Journal of the American Medical Informatics Association, vol. 8, no. 6, pp. 527–534, Nov. 2001, doi: 10.1136/jamia.2001.0080527.

[3] J. Amann et al., “To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems,” PLOS Digital Health, vol. 1, no. 2, p. e0000016, Feb. 2022, doi: 10.1371/journal.pdig.0000016.

[4] K. Gao, G. Mei, F. Piccialli, S. Cuomo, J. Tu, and Z. Huo, “Julia language in machine learning: Algorithms, applications, and open issues,” Comput Sci Rev, vol. 37, p. 100254, Aug. 2020, doi: 10.1016/j.cosrev.2020.100254.

[5] D. W. Aha, “ Heart Disease Data Set,” https://archive.ics.uci.edu/ml/datasets/heart+disease, Jan. 26, 2023.

[6] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” Expert Syst Appl, vol. 36, no. 4, pp. 7675–7680, May 2009, doi: 10.1016/j.eswa.2008.09.013.

[7] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” Expert Syst Appl, vol. 36, no. 4, pp. 7675–7680, May 2009, doi: 10.1016/j.eswa.2008.09.013.

[8] N. Ghadiri Hedeshi and M. Saniee Abadeh, “Coronary Artery Disease Detection Using a Fuzzy-Boosting PSO Approach,” Comput Intell Neurosci, vol. 2014, pp. 1–12, 2014, doi: 10.1155/2014/783734.

[9] Heart tests, “https://www.heartfoundation.org.nz/your-heart/heart-tests,” https://www.heartfoundation.org.nz/your-heart/heart-tests, Jan. 25, 2023.

[10] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, “An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection,” IEEE Access, vol. 7, pp. 180235–180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[11] S. J. Pasha and E. S. Mohamed, “Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction,” IEEE Access, vol. 8, pp. 184087–184108, 2020, doi: 10.1109/ACCESS.2020.3028714.

[12] S. M. Saqlain et al., “Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines,” Knowl Inf Syst, vol. 58, no. 1, pp. 139–167, Jan. 2019, doi: 10.1007/s10115-018-1185-y.

[13] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, “Early and accurate detection and diagnosis of heart disease using intelligent computational model,” Sci Rep, vol. 10, no. 1, p. 19747, Nov. 2020, doi: 10.1038/s41598-020-76635-9.

[14] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, “An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network,” IEEE Access, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[15] C. Gupta, A. Saha, N. v Subba Reddy, and U. Dinesh Acharya, “Cardiac Disease Prediction using Supervised Machine Learning Techniques.,” J Phys Conf Ser, vol. 2161, no. 1, p. 012013, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012013.

[16] Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano, “ Heart Disease Data Set,” https://archive.ics.uci.edu/ml/datasets/heart+disease, Jan. 25, 2023.

[17] “Veri Bilimi ve Yapay Zekâ Tabanlı Web Yazılımları ,” http://biostatapps.inonu.edu.tr/.

[18] “Veri Sınıflandırma Yazılımı,” http://biostatapps.inonu.edu.tr/JVSY/.

[19] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN Model-Based Approach in Classification,” 2003, pp. 986–996. doi: 10.1007/978-3-540-39964-3_62.

[20] Y. I. A. Rejani and S. T. Selvi, “Early Detection of Breast Cancer using SVM Classifier Technique,” Dec. 2009.

[21] C. Kingsford and S. L. Salzberg, “What are decision trees?,” Nat Biotechnol, vol. 26, no. 9, pp. 1011–1013, Sep. 2008, doi: 10.1038/nbt0908-1011.

[22] M. Pal, “Random forest classifier for remote sensing classification,” Int J Remote Sens, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: 10.1080/01431160412331269698.

[23] R. E. Schapire, “Explaining AdaBoost,” in Empirical Inference, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–52. doi: 10.1007/978-3-642-41136-6_5.

[24] H. Kamel, D. Abdulah, and J. M. Al-Tuwaijari, “Cancer Classification Using Gaussian Naive Bayes Algorithm,” in 2019 International Engineering Conference (IEC), Jun. 2019, pp. 165–170. doi: 10.1109/IEC47844.2019.8950650.

[25] Haoshi Zhang, Yaonan Zhao, Fuan Yao, Lisheng Xu, Peng Shang, and Guanglin Li, “An adaptation strategy of using LDA classifier for EMG pattern recognition,” in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2013, pp. 4267–4270. doi: 10.1109/EMBC.2013.6610488.

Acknowledgement

This article was published as a proceeding of 14th Turkish Congress of Medical Informatics Association in 2023.

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

RESEARCH ARTICLE

Comparing the Performance of Ensemble Methods in Predicting Emergency Department Admissions Using Machine Learning Techniques

Murat Emre Yapıcı[a,†], **Abdulkadir Hızıroğlu**[a], **Ali Mert Erdoğan**[a]

[a] Department of Management Information Systems, İzmir Bakırçay University, İzmir, Türkiye,
† m.emreyapici@gmail.com, corresponding author

RECEIVED OCTOBER 30, 2023
ACCEPTED JANUARY 12, 2024

CITATION Yapıcı, M. E., Hızıroğlu, A., & Erdoğan. A. M. (2024). Comparing the performance of ensemble methods in predicting emergency department admissions using machine learning techniques. *Artificial Intelligence Theory and Applications, 4(*1), 11-21.

Abstract

Healthcare data collection, storage, retrieval, and analysis are enabled by various technologies and tools in health information systems. These systems include health information exchanges, telemedicine platforms, clinical decision support systems, and electronic health records. They aim to improve patient outcomes, provider communication, and healthcare workflows. Machine learning is being used in emergency rooms to address challenges such as increasing patient volume, limited resources, and the need for quick decisions. Machine learning algorithms can assist in triage and risk stratification by identifying patients requiring urgent care and predicting the severity of their condition. By analyzing various patient data sources, machine learning can detect patterns and indicators that human clinicians may miss, enabling early intervention and potentially saving lives. However, there is a lack of comparative evaluation of ensemble methods used in analysis. Therefore, this study aims to thoroughly examine and analyze various ensemble methods to understand their efficacy and performance, contributing valuable insights to researchers and practitioners.

Keywords: ensemble methods, logistic regression, prediction, emergency department

1. Introduction

Emergency services are essential healthcare units that provide immediate medical assistance to patients in need. They are categorized based on the urgency and severity of the patient's condition, with red indicating life-threatening emergencies, yellow indicating conditions with a risk of permanent damage, and green indicating mild injuries or illnesses [1]. Information systems play a crucial role in emergency care by providing insights into the workload, patient information, and preliminary assessments in the emergency department. These systems enable informed decision-making for triage and resource allocation, addressing challenges such as overcrowding and improving overall emergency care [2]. Healthcare information systems encompass various technologies, processes, and tools that facilitate the collection, storage, retrieval, and analysis of healthcare data [3]. Electronic health records (EHRs) serve as digital databases of patient information, supporting comprehensive and coordinated care [4]. EHRs aid clinic allergies ending by providing immediate access to vital patient data, alerting healthcare professionals to potential interactions or allergies and suggesting evidence-based treatment options [5]. Clinical decision support systems (CDSS) utilize advanced algorithms and medical knowledge databases to enhance diagnosis accuracy, reduce errors, and improve patient safety [5]. Machine learning is a significant component of information systems, analyzing large volumes of medical data and extracting valuable insights. It improves patient care, optimizes resource allocation, and enhances decision-making processes [8]. In emergency departments (EDs), machine learning algorithms improve triage and risk stratification by accurately predicting the severity of a patient's condition and identifying those in urgent need of care [9]. Furthermore, machine learning can detect patterns and indicators in diverse data sources, enabling early diagnosis and prediction of adverse events that may be missed by human clinicians [10]. This study aims to analyze ED admission rates and develop a predictive model to determine the likelihood of future hospitalization. The objectives include reducing overcrowding, expediting treatment for urgent cases, and increasing employee motivation. By analyzing patient demographics, medical history, and severity of conditions, advanced statistical techniques and machine learning algorithms will be used to develop a reliable framework for predicting hospitalization rates. Implementing the study's findings can improve operational efficiency, patient outcomes, and the work environment in EDs.

2. Literature Review

Table 1. Literature Review

STUDY	TECHNIQUE	EVALUATION	RESULT
(Barak-Corren et al., 2021)	**eXtreme Gradient Boosting**	AUC	AUC **0.90-0.93**
(Lee et al. ,2020)	**Multinomial Logistic Regression,** Neural Network, Support Vector Machine	Accuracy (%95 CI)	Accuracy **MLR = 81.6** Neural Network =81.2 Support Vector Machine=81.4
(Graham et al. ,2018)	Logistic Regression, Decision Trees, **Gradient Boosting**	Accuracy (%95 CI) AUC	Accuracy LR=79.94 Decision Trees=80.06 **GBM=80.31** AUC LR=0.849 Decision Trees=0.824 **GBM = 0.859**
(Peck et al. ,2013)	Logistic Regression	AUC R2	AUC **LR=0.80-0.89** R2 LR=0.58 - 0.90
(Woo Suk Hong et al.,2018)	Logistic Regression, **Gradient Boosting,** **Deep Neural Networks**	AUC	AUC **LR=0.87** **XGBOOST=0.87** **DNN=0.87**
(Sun et al., 2011)	Logistic Regression	ROC Accuracy (%95 CI)	ROC **LR=0.849** Accuracy LR=84.7

Ensemble methods in machine learning have garnered significant attention and demonstrated impressive success rates in various applications [11]-[24],[31]. These techniques aim to improve the performance and robustness of predictive models by combining the predictions of multiple base learners, thereby leveraging the diversity of these learners to achieve better overall results. Despite their widespread adoption and promising outcomes, the existing literature still lacks comprehensive comparative studies that thoroughly evaluate and compare different ensemble methods.

Numerous individual studies in the existing literature showcase the effectiveness of ensemble methods, highlighting their contributions to various tasks. For instance, researchers have demonstrated the benefits of ensemble methods like XGBoost Regression in classification tasks. A study by Barak-Corren et al. (2010) [11] showed that an ensemble of Multinomial Logistic Regression models outperformed individual logistic regression models in predicting customer churn, achieving higher accuracy and better generalization.

However, despite these individual success stories, there is a notable lack of direct comparisons between different ensemble methods in the literature. Few studies conduct head-to-head evaluations to determine which ensemble technique is more suitable for specific scenarios.

For this reason, this study aims to fill this gap by comparing the performance of four prominent ensemble methods: Adaboost, LogitBoost, GentleBoost, and RusBoost.

3. Research Methodology

3.1. Dataset

The dataset used in this study consists of 1267 systematically selected records of adult patients admitted to two emergency departments between October 2016 and September 2017 [25], [32] . Emergency Service. In order to ensure accurate forecasting, certain columns in the dataset needed to be removed, which could potentially impact the accuracy of the predictions. Hence, it is important to highlight the current state of the dataset as it undergoes estimation. The resulting configuration of the dataset is presented below.

These parameters are key indicators used in medical assessments. The mental scale assesses a person's level of consciousness and responsiveness, ranging from alertness to unconsciousness. The Numeric Rating Scale (NRS) for pain measures pain intensity on a numerical scale. Systolic blood pressure (SBP) represents the pressure in arteries when the heart beats, while diastolic blood pressure (DBP) is the pressure when the heart rests between beats. Respiration rate (RR) measures the number of breaths per minute, crucial in evaluating respiratory health. Saturation indicates the oxygen saturation level in the blood, often measured with a pulse oximeter, reflecting the amount of oxygen carried by red blood cells. Together, these parameters provide a comprehensive snapshot of a person's mental state, pain level, cardiovascular health, respiratory function, and oxygenation status, aiding healthcare professionals in making informed decisions about treatment and care.

3.2. Data Preparation

Tasks such as data cleaning, integration, transformation (min, max on all features), and feature selection are involved in this process. One important modification made during data preparation was transforming disposition values into binary categories. This simplification enables easier interpretation and analysis of the dataset, specifically regarding patient outcomes (discharged or admitted).

The Emergency Department categorized and analyzed patients based on variables such as group, sex, age, arrival mode, injury, mental status, and pain. This approach provided insights into patient cohorts, gender patterns, age trends, arrival modes, types of injuries, mental well-being, and discomfort levels. Considering these dimensions allowed for a deeper understanding of the patient population and facilitated in-depth data analysis.

Table 2. Descriptions and Distributions of the variables

Variable	Descriptions and Distributions of the variables
Sex	1: Female (51.8%) / 2: Male (48.2%)
Age	Age (mean: 53.9, std: 18.8)
Patients_number_per_hour	Patients number/hours (mean: 7.5, std: 3.1)
Arrival_mode	1: Walking (6.6%) / 2: 119 use (19.5%) / 3: Private car(61.8%) / 4: Private ambulance (10.7%) / 5: Others (1.4%)
Injury	1: Non-injury (80.9%) / 2: Injury (%19.1)
Mental	1: Alert (95.4%) / 2: Verbal response (2.4%) / 3: Pain response (1.7%) / 4: Unconsciousness (0.5%)
Pain	1: Pain (56.9%) / 2: Non-pain (43.1%)
NRS_pain	Numeric rating scales of pain (between 1-5 (84.2%) / between 6-10(15.8%))
SBP	Systolid blood pressure (mean: 131.6, std: 26.7)
DBP	Diastolic blood pressure (mean: 78.6, std: 14.6)
HR	Heart rate (mean: 82.2, std: 16.4)
RR	Respiration rate (mean: 19.3, std: 1.9)
BT	Body temperature (mean: 36.3, std: 0.7)
Saturation	Saturation to use pulse oximeter (mean: 96.9, std: 4.2)
Disposition	0: Discharge (68.1%) / 1: Admission (31.9%)

To enhance analysis integrity and reliability, missing values and outliers were handled. By identifying and eliminating these problematic data points, the analysis was strengthened in terms of robustness and accuracy. Categorical variables were encoded using One-hot Encoding, representing each unique value as a separate column. This streamlined the dataset and uncovered patterns and correlations. The dataset was divided 80/20 into training and test samples. The training sample was used for model training, while the test sample assessed their performance, ensuring reliable and suitable analysis methods.

3.3. Modelling and Evaluation

Logistic Regression is a popular choice for binary classification tasks due to its simplicity, interpretability, and proven effectiveness. It estimates the probability of an event occurring based on input variables, making it reliable in various domains [26]. To enhance the predictive performance, ensemble methods such as AdaBoost [27], LogitBoost [28], GentleBoost [29], and RUSBoost [30] were employed. These methods combine multiple models to improve accuracy and handle class imbalance. AdaBoost iteratively trains weak classifiers, focusing on misclassified samples, while LogitBoost optimizes Logistic Regression parameters. GentleBoost assigns smaller weights to misclassified samples to reduce sensitivity to outliers, and RusBoost addresses class imbalance by under sampling the majority class. Since Logistic Regression is used very frequently in this field, we added it to the benchmarking study. Lastly, performance metrics such as Sensitivity, Specificity and F1 Score and so on , which are frequently used in Binary Classification, were used to evaluate and compare model performances. The following provides a brief explanation of performance metrics.

Sensitivity: Ratio of true positive (discharged) examples correctly predicted by the model. It is calculated using the formula:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

Specificity: Ratio of true negative (admitted) examples correctly predicted by the model. It is calculated using the formula:

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

Precision: Ratio of true positive examples correctly predicted by the model. It is calculated using the formula:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Negative Predictive Value: Ratio of true negative examples correctly predicted by the model. It is calculated using the formula:

$$Negative\ Predictive\ Value = \frac{TN}{TN + FN} \tag{4}$$

False Positive Rate: Ratio of true negatives incorrectly predicted as positives. It is calculated using the formula:

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{5}$$

False Discovery Rate: Ratio of positives predicted incorrectly as positives. It is calculated using the formula:

$$False\ Discovery\ Rate = \frac{FP}{FP + TP} \tag{6}$$

False Negative Rate: Ratio of true positives incorrectly predicted as negatives. It is calculated using the formula:

$$False\ Negative\ Rate = \frac{FN}{TP + FN} \tag{7}$$

Accuracy: Ratio of correct predictions (both true positive and true negative) to the total number of examples. It is calculated using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

F1 Score: Harmonic mean of precision and sensitivity, providing a balance between the two metrics. It is calculated using the formula:

$$f_1\ Score = 2 * \frac{P * R}{P + R} \quad (9)$$

Matthews Correlation Coefficient: Calculates the correlation coefficient between observed and predicted binary classifications. It takes values between -1 and +1, where +1 indicates perfect predictions, 0 implies no improvement over random guessing, and -1 signifies complete disagreement between prediction and observation.

Area Under the Receiver Operating Characteristic Curve (AUC): Quantifies the performance of a binary classification model across various threshold values. The ROC curve illustrates the relationship between true positive rate and false positive rate for different threshold values. AUC represents the area under this curve.

4. Results

ROC RESULTS

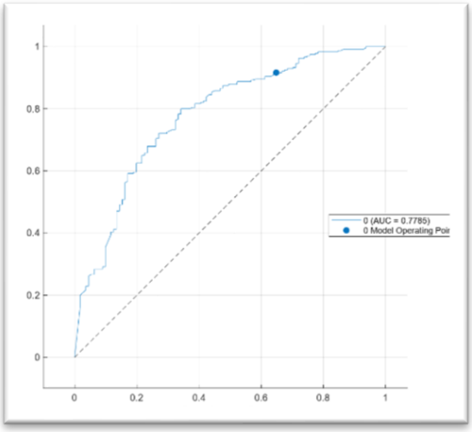

Figure 1. AdaBoost

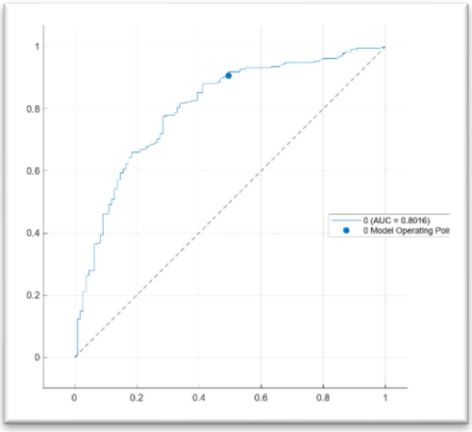

Figure 2. Logistic Reg.

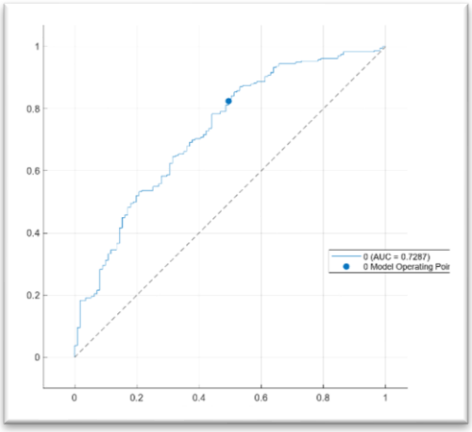

Figure 3. Gentleboost

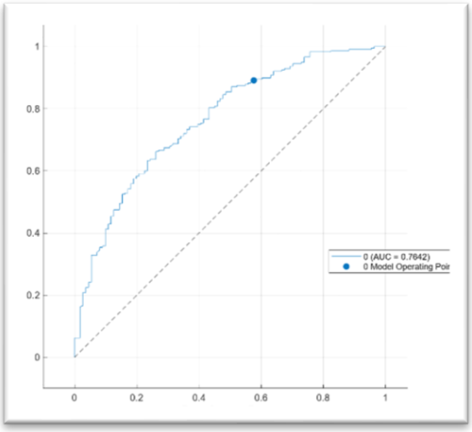

Figure 4. Logitboost

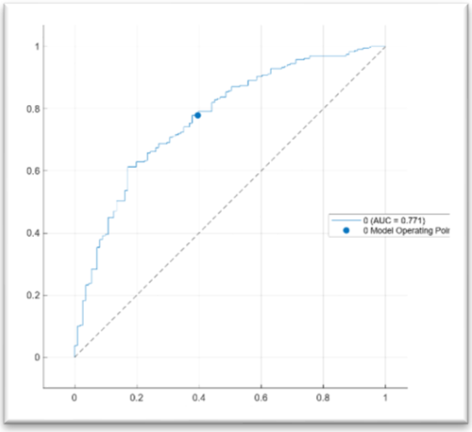

Figure 5. Rusboost

Table 3. Results and Comparison of Methods

	Sensitivity	Specificity	Precision	Negative Predictive Value	False Positive Rate	False Discovery Rate	False Negative Rate	Accuracy	F1 Score	Matthews Correlation Coefficient	AUC
Adaboost	0.753	0.661	0.916	0.351	0.339	0.083	0.246	0.737	0.827	0.333	0.778
Logistic Reg.	0.795	0.717	0.908	0.500	0.282	0.091	0.204	0.778	0.848	0.457	0.801
Gentleboost	0.782	0.571	0.825	0.504	0.428	0.175	0.217	0.723	0.803	0.341	0.728
Logitboost	0.769	0.643	0.891	0.423	0.356	0.108	0.230	0.743	0.826	0.361	0.764
Rusboost	0.809	0.558	0.779	0.603	0.441	0.220	0.190	0.723	0.794	0.375	0.771

The Adaboost model performs reasonably well but has room for improvement. It shows good sensitivity (0.7534) but comparatively lower specificity (0.6610), indicating challenges in accurately identifying negative cases. The model has high precision (0.9167) but a relatively low negative predictive value (0.3514), suggesting a considerable number of incorrect negative predictions. The false positive rate (0.3390) is moderately high, while the false discovery rate (0.0833) is low. The false negative rate (0.2466) can be improved for better performance. The model's accuracy is 0.7379, and the F1 score (0.8271) demonstrates a reasonable balance between precision and sensitivity. The Matthews Correlation Coefficient (0.3333) indicates moderate overall agreement. AUC value of 0.778 indicates that the model has good discrimination ability, distinguishing between classes with moderate accuracy. In summary, while the Adaboost model has some positive aspects, improvements can be made in terms of specificity, negative predictive value, false positive rate, false negative rate, and overall accuracy through optimization and fine-tuning.

The Logistic Regression model performs well with good sensitivity (0.7956) and specificity (0.7179). It has high precision (0.9083) and relatively low false positive rate (0.2821) and false discovery rate (0.0917). The negative predictive value (0.5000) can be improved, indicating room for better identification of negative cases. The false negative rate (0.2044) is relatively low. The model's accuracy is 0.7784, and the F1 score (0.8482) demonstrates a good balance between precision and sensitivity. The Matthews Correlation Coefficient (0.4579) indicates moderate overall agreement. With a AUC score of 0.801, LR has the best result comparing the other models, which means a higher level of accuracy in classifying outcomes based on the model's predictions. In summary, Logistic Regression model shows reliable performance with high sensitivity, specificity, precision, and accuracy. Improvements can be made in the negative predictive value and false negative rate through fine-tuning and optimization efforts.

The Gentleboost model shows mixed results with potential for improvement. It has a sensitivity of 0.7826, correctly identifying a decent proportion of positive cases. However, it struggles with specificity (0.5714) in accurately identifying negative cases. The precision (0.8250) is relatively high, with a majority of positive predictions being correct. The negative predictive value (0.5045) suggests room for improvement in correctly identifying negative cases. The false positive rate (0.4286) is relatively high, indicating a considerable number of negative cases being falsely classified as positive. The false discovery rate (0.1750) is relatively low, suggesting fewer false positive predictions. The false negative rate (0.2174) represents the proportion of positive cases incorrectly classified as negative, which can be further improved. The model's accuracy is 0.7236, and the F1 score (0.8032) demonstrates a reasonable balance between precision and sensitivity. The Matthews Correlation Coefficient (0.3416) indicates a moderate level of overall agreement. Meanwhile, AUC value of 0.728 suggests fair discrimination ability, with some limitations in accurately separating classes. In summary, the Gentleboost model shows a mix of strengths and weaknesses, with room for improvement in specificity, negative predictive value, and false positive rate. Further optimization and fine-tuning efforts are needed to enhance its performance.

The Logitboost model shows reasonable performance. It has a sensitivity of 0.7698, correctly identifying a decent proportion of positive cases, and a specificity of 0.6438, indicating reasonable performance in identifying negative cases. The precision (0.8917) is relatively high, with a majority of positive predictions being correct. However, the negative predictive value (0.4234) suggests room for improvement in correctly identifying negative cases. The false positive rate (0.3562) is moderately high, implying some negative cases being falsely classified as positive. The false discovery rate (0.1083) is relatively low, indicating fewer false positive predictions. The false negative rate (0.2302) represents the proportion of positive cases incorrectly classified as negative, which is moderate. The model's accuracy is 0.7436, and the F1 score (0.8263) demonstrates a reasonable balance between precision and sensitivity. The Matthews Correlation Coefficient (0.3610) indicates a moderate level of overall agreement. AUC score of 0.764 indicates moderately good performance in distinguishing between classes. In summary, the Logitboost model shows moderate performance with good precision and sensitivity. However, improvements can be made in terms of specificity, negative predictive value, false positive rate, and overall accuracy. Further optimization and fine-tuning efforts may enhance its performance.

The Rusboost model shows mixed performance. It has a sensitivity of 0.8095, correctly identifying a relatively high proportion of positive cases, but struggles with specificity (0.5583) in accurately identifying negative cases. The precision (0.7792) is moderate, with a majority of positive predictions being correct. The negative predictive value (0.6036) is relatively high, indicating better performance in correctly identifying negative cases. The false positive rate (0.4417) is relatively high, implying a substantial number of negative cases being falsely classified as positive. The false discovery rate (0.2208) is moderately high, indicating a significant number of false positive predictions. The false negative rate (0.1905) represents the proportion of positive cases incorrectly classified as negative, which is relatively low but can be improved. The model's accuracy is 0.7236, and the F1 score (0.7941) demonstrates a reasonable balance between precision and sensitivity. The Matthews Correlation Coefficient (0.3752) indicates a moderate level of overall agreement. AUC value of 0.771 denotes decent discriminatory power, although slightly lower compared to the other models scores, especially LR, but still indicating a reasonable level of predictive accuracy. In summary, the Rusboost model shows mixed performance with strengths in sensitivity and negative predictive value, but weaknesses in specificity and false positive rate. Further optimization and fine-tuning may be necessary to enhance its overall performance.

Rusboost is a rarely encountered ensemble method that provides a comparative perspective. It outperforms Gentleboost and closely resembles Adaboost, suggesting it as a valuable alternative with similar predictive accuracy. The inclusion of Rusboost expands the knowledge base and promotes exploration of ensemble methodologies. AUC analysis reveals remarkable similarity between Rusboost and Adaboost, showcasing favorable outcomes. This warrants a reevaluation of common approaches and encourages further research on Rusboost's capabilities. Its success enhances analytical outcomes and expands possibilities for future studies.

5. Conclusion

Healthcare data analysis relies on various technologies and systems, including health information exchanges, telemedicine platforms, clinical decision support systems, and electronic health records. Machine learning has revolutionized emergency care by improving triage and risk stratification. Machine learning algorithms accurately identify patients needing urgent care and predict the severity of their conditions, enabling early intervention. Despite a wide range of analysis methods, there is a lack of comparative evaluations of ensemble methods. This study aims to comprehensively examine and analyze ensemble methods for healthcare data analysis. Logistic Regression consistently performs the best, followed closely by Adaboost. Rusboost, an underutilized method, shows promising performance similar to Adaboost. Logitboost also demonstrates comparable results. Gentleboost, however, is the least successful method. These findings highlight the importance of careful selection of ensemble methods for specific prediction studies in the Emergency Department. Researchers can make informed decisions to advance predictive models in emergency care.

Additionally, there are several limitations of the study. One notable limitation is related to the dataset used for analysis, where the availability and quality of the data might introduce inherent biases and limitations in representing the full spectrum of emergency care cases. Moreover, the selection of ensemble methods for analysis might limit the comprehensiveness of the comparison, as other relevant techniques not included could impact the overall conclusions. Furthermore, the study's generalizability might be constrained by the specific context and settings in which the research was conducted, considering different healthcare systems, patient populations, or emergency care protocols.

In future, researchers can carefully select a diverse and representative dataset of emergency care cases, considering different medical conditions, patient demographics, and severity levels. To address privacy concerns, they can work with de-identified or synthetic datasets to ensure compliance with regulations while maintaining the dataset's integrity. Given potential limitations in implementing certain ensemble methods in the healthcare context, the researchers can adopt a focused approach by comparing a subset of ensemble methods that are more suitable for the specific emergency care prediction task. This targeted comparison can ensure the study's relevance and feasibility within the given constraints. To assess the performance of the ensemble methods accurately, researchers can choose appropriate performance metrics aligned with the specific goals of emergency care prediction. Considering the restricted availability of healthcare data, the researchers will utilize techniques like cross-validation and bootstrapping to obtain more reliable estimates of ensemble method performance. These resampling methods will enable them to evaluate the ensemble methods on multiple subsets of the data, yielding more robust and generalizable results.

References

[1] Medicana (n.d.). Tıbbi Servisler ve Acil servis. https://www.medicana.com.tr/tibbi-birimler/acil-servis

[2] Hoot, N. R., & Aronsky, D. (2008). Systematic review of emergency department crowding: causes, effects, and solutions. *Annals of emergency medicine*, 52(2), 126-136.

[3]] Mamlin, B. W., Biondich, P. G., Wolfe, B. A., Fraser, H., Jazayeri, D., Allen, C., ... & Tierney, W. M. (2006). Cooking up an open source EMR for developing countries: OpenMRS–a recipe for successful collaboration. *In AMIA Annual Symposium Proceedings (Vol. 2006, p. 529).* American Medical Informatics Association.

[4] McGinn, C. A., Grenier, S., Duplantie, J., Shaw, N., Sicotte, C., Mathieu, L., ... & Gagnon, M. P. (2011). Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC medicine, 9(1), 1-10.

[5] Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj, 330(7494), 765.

[6] Agency for Healthcare Research and Quality (n.d.). Clinical Decision Support. https://www.ahrq.gov/cpi/about/otherwebsites/clinical-decision-support/index.html#:~:text=Clinical%20decision%20support%20(CDS)%20provides,team%20and%20patient%20to%20consider

[7] HealthIT.gov (n.d.). What is an electronic health record (EHR)?. https://www.healthit.gov/faq/what-electronic-health-record-ehr#:~:text=EHRs%20are%20a%20vital%20part,decisions%20about%20a%20patient's%20care

[8] Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. Critical care medicine, 46(4), 547.

[9] Kunt, M. M. (2021). Emergency Medicine and Artificial Intelligence. https://dergipark.org.tr/en/download/article-file/1985451

[10] Sun, J., Zhang, Y., & Tang, L. (2019). Predicting patient deterioration in the emergency department: A machine learning approach. *Journal of Biomedical Informatics*, 98, 103267.

[11] Barak-Corren, Y., Chaudhari, P., Perniciaro, J., Waltzman, M., Fine, A. M., & Reis, B. Y. (2021). Prediction across healthcare settings: a case study in predicting emergency department disposition. *npj Digital Medicine*, 4(1), 169.

[12] Lee, S. H., Chinnam, R. B., Dalkiran, E., Krupp, S., & Nauss, M. (2020). Prediction of emergency department patient disposition decision for proactive resource allocation for admission. *Health Care Management Science*, *23*(3), 339–359. https://doi.org/10.1007/s10729-019-09496-y

[13] Chen, C., Hsieh, J., Cheng, S., Lin, Y., Lin, P., & Jeng, J. (2020). Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. *International Journal of Medical Informatics*, *139*, 104146. https://doi.org/10.1016/j.ijmedinf.2020.104146

[14] LaMantia, M. A., Platts-Mills, T. F., Biese, K., Khandelwal, C., Forbach, C. R., Cairns, C. B., Busby-Whitehead, J., & Kizer, J. S. (2010). Predicting Hospital Admission and Returns to the Emergency Department for Elderly Patients. *Academic Emergency Medicine*, *17*(3), 252–259. https://doi.org/10.1111/j.1553-2712.2009.00675.x

[15] Parker, C. A., Liu, N., Wu, S. X., Shen, Y., Lam, S. S. W., & Ong, M. E. H. (2019). Predicting hospital admission at the emergency department triage: A novel prediction model. *American Journal of Emergency Medicine*, *37*(8), 1498–1504. https://doi.org/10.1016/j.ajem.2018.10.060

[16] Graham, B., Bond, R., Quinn, M., & Mulvenna, M. (2018). Using data mining to predict hospital admissions from the emergency department. *IEEE Access*, *6*, 10458-10469.

[17] Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Patzer, A., & Schrager, J. D. (2017). Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods of information in medicine*, *56*(05), 377-389.

[18] Peck, J. S., Gaehde, S. A., Nightingale, D. J., Gelman, D. Y., Huckins, D. S., Lemons, M. F., ... & Benneyan, J. C. (2013). Generalizability of a simple approach for predicting hospital admission from an emergency department. Academic Emergency Medicine, 20(11), 1156-1163.

[19] Sun, Y., Heng, B. H., Tay, S. Y., & Seow, E. (2011). Predicting hospital admissions at emergency department triage using routine administrative data. *Academic Emergency Medicine*, *18*(8), 844-850.

[20] Roquette, B. P., Nagano, H., Marujo, E. C., & Maiorano, A. C. (2020). Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. *Neural Networks*, *126*, 170-177.

[21] Mowbray, F., Zargoush, M., Jones, A., de Wit, K., & Costa, A. (2020). Predicting hospital admission for older emergency department patients: Insights from machine learning. *International Journal of Medical Informatics*, *140*, 104163.

[22] Peck, J. S., Benneyan, J. C., Nightingale, D. J., & Gaehde, S. A. (2012). Predicting emergency department inpatient admissions to improve same-day patient flow. *Academic Emergency Medicine*, *19*(9), E1045-E1054.

[23] Leegon, J., Jones, I., Lanaghan, K., & Aronsky, D. (2005). Predicting hospital admission for Emergency Department patients using a Bayesian network. In *AMIA Annual Symposium Proceedings* (Vol. 2005, p. 1022). American Medical Informatics Association.

[24] Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PloS one*, *13*(7), e0201016.

[25] Kaggle (2021, August 28). *Emergency Service - Triage Application*. https://www.kaggle.com/datasets/ilkeryildiz/emergency-service-triage-application

[26] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

[27] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119-139.

[28] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

[29] Freund, Y., & Mason, L. (1999, June). The alternating decision tree learning algorithm. In *icml* (Vol. 99, pp. 124-133).

[30] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2009). RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE transactions on systems, man, and cybernetics-part A: systems and humans*, *40*(1), 185-197.

[31] Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PloS one*, *13*(7), e0201016.

[32] Moon, S. H., Shim, J. L., Park, K. S., & Park, C. S. (2019). Triage accuracy and causes of mistriage using the Korean Triage and Acuity Scale. *PloS one*, *14*(9), e0216972.

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

■ RESEARCH ARTICLE

Efficient and adaptive operator selection in swarm intelligence using machine learning approaches

Mehmet Emin Aydın [a†], **Rafet Durgut** [b]

[a] Computer Science and Creative Technologies, University of the West of England, Bristol, United Kingdom
[b] Engineering and Natural Sciences, Bandırma Onyedi Eylül University, Balıkesir, Türkiye
† mehmet.aydin@uwe.ac.uk, corresponding author

RECEIVED DECEMBER 4, 2023
ACCEPTED JANUARY 10, 2024

CITATION Aydın, M., Durgut, R. (2024). Efficient and adaptive operator selection in swarm intelligence using machine learning approaches. *Artificial Intelligence Theory and Applications, 4(1)*, 22-32.

Abstract

Problem solving has been one of renown artificial intelligence fields attracting research for decades. Swarm intelligence is recognised as the family of the state-of-art approaches in problem solving gained much research attention for the enduring problems. The main challenge appears to be is the speed of algorithmic approximation where many approaches were proposed to accelerate approximation avoiding local optima. Recent research demonstrates that inefficiencies in search procedures can be side-stepped using the experiences gained while search is undergoing utilising machine learning approaches. Machine learning turned to be popular to let approach problems on experience basis. Reinforcement learning becomes a success-proven approach for online learning, especial when training data is not available upfront. In this paper, we overview the usefulness of machine learning and reinforcement learning in performance improvement of artificial bee colony algorithms in solving combinatorial optimisation problems. Furthermore, we demonstrate how supervised and reinforcement learning approaches facilitate swarm intelligence algorithms to gain experience for immediate and later use to build capable and powerful operator selection schemes, which help improve efficiency of swarm intelligence problem solvers.

Keywords: adaptive operator selection; machine learning; reinforcement learning; artificial bee colony; set union knapsack problem

---

1. Introduction

Optimisation involves very hard engineering problems to find out the best solutions within a reasonable time frame. Artificial intelligence offers convenient approaches to alternate the classical problem-solving approaches. Evolutionary computation and swarm intelligence bring many opportunities to researchers and practitioners' attention to handle such hard problems, particularly those problems classified as NP-Hard and NP-Complete problems [1]. However, due to the characteristics of the problems space, difficulties emerge while conducting exploration across the solution space. The difficulties are known as when to intensify search locally and when to diversify towards different regions for better solutions. This is recognised as "Exploration versus Exploitation" (EvE) issue [2]. The dilemma of EvE is applied to all search algorithms

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

including swarm intelligence algorithms such as particle swarm optimisation, artificial bee colonies, and ant colony optimisation variants.

A reasonable success in EvE actions goes through use of different approaches harmonised into problem solving process. However, problem solving requires domain and problem-specific information in order to decide how and when to intensify the search and how and when to diversify it. Research has never stopped to investigate for better problem-solving approaches with lower time complexity. Use of multiple search strategies as well as multiple operators is known an important stream as part of this effort. An important research question in this regard is that how to select an operator / strategy to tackle the temporal state of the problem in the best way and capacity. Is there any generic approach applicable to all types of the problems?

The search issues mentioned above lead to another difficulty on the way, which is how to reach a general approach applicable to every kind of problems. No-Free-Lunch theory [3] has originated from research efforts in this regard which confirms how difficult to come up with an approach applicable to as many problem types as possible. On the other hand, there is a classical dream of artificial intelligence to attain a general problem solver [4], [5].

Recent studies paid attention on the way to develop more generic approaches to reasonably offered solutions for combinatorial optimisation problems [6]–[9]. The author of [10] has surveyed online and off-line machine learning algorithms in efficiency studies of metaheuristics in a wider context, and paid attention to the use of machine learning in building operator selection schemes. Feature analysis has been conducted in the study reported by [9] in order to gauge the impact and predictiveness of each feature identified to characterise the states of operator selection problem. The analysis is carried out with supervised machine learning approaches in that work with a comparative approach. More specific studies reported, especially on utilisation of reinforcement learning approaches for improvement of metaheuristic efficiency [10], [11]. We, the authors of this article, have not come across to any work – to the best of our knowledge – overviewing the use of supervised and reinforcement learning approaches in building efficient adaptive operator selection schemes so as to utilise in swarm intelligence algorithms for boosting performance and efficiency, especially with a generalisation point of view. The aim of this paper is to overview the recent attempts and evaluate them with respect to the level of generalisation.

The rest of this paper is organised as follows: Section 2 introduces how swarm intelligence algorithms are enhanced and made efficient with machine learning referring to all key references and state of the art works in the field, Section 3 summarises the result set for proof-of-concept with relevant discussions and Section 4 concludes the article.

2. Materials and Methods

The main aim of this article is to overview the recent works to develop more generalisable approaches to solve combinatorial optimisation problems, and possibly continuous problems, too. A problem is defined and described with models (e.g., mathematical or simulation models) with which the inter-relations of its components are represented with variables and parameters. Then, the modelled problem requires a solution which best fits in use. The optimisation field has been populated with a variety of heuristic-based approaches alongside traditional search and problem-solving approaches. Among these, swarm intelligence has proven significant success in the field and achieved to be recognised as a state-of the-art paradigm. The following subsections will introduce how recent research has been shaped up around these concepts of the paradigm merged with frontline machine learning principles and approaches.

2.1 Main search algorithm

The optimisation framework always imposes a search algorithm with which the better and the best results fitting to the circumstances is explored. Recent studies pay more attention on heuristic solutions, which are derived from meta-heuristic frameworks such as evolutionary algorithms and swarm intelligence algorithms. Artificial Bee Colony (ABC) optimisation is one of recently developed, success proven, renown swarm intelligence algorithms [12]. The main motivation behind ABC is to imitate the social behaviour of honeybees in nectar search mimicking the way the colonies collaborate to achieve the goals and targets. Search algorithms arrange move from one solution to another via different types of operators, but each operator comes up with limits in approximation, which is believed to be eased with use of other complimentary alternatives. This enforces use of multiple operators /neighbourhood functions by the algorithms [13]. It is believed that ABC is one of bound-free algorithms in approximation due to that there is not much functional constraints systematically impose search instrumented by ABC. Its variants extended with multiple operators have been used for solving many combinatorial [6], [14] and functional [15] optimisation problems, a number of different variants are proposed for efficiency in problem solving.

Another recent swarm intelligence algorithm is called Crow Search Algorithm (CSA), which is recently devised mimicking the social behaviour of crows in handling issues. The algorithm has recently been implemented with multi search strategies [16]. Likewise, particle swarm optimisation [17] and evolutionary, (e.g., differential evolution [18]), algorithms have also been implemented with multiple search strategies, and or operators to diversify search and achieve high efficiency [19], [20]. Grey wolf and whale optimisation algorithm combined with reinforcement learning in [21] and multi-armed bandits [22] used for optimiser selection among Harris Hawks optimiser, differential evolution and whale optimisation algorithm.

2.2 Adaptive operator selection

Operators are neighbourhood functions devised to manage moving from one problem state to another while searching for the best solution with optimisation algorithms. Many algorithms have been designed to use single operator at the beginning, but, adopted utilising multiple operators through the whole process subject to a selection rule or scheme imposed [1]. Obviously, not all schemes proposed are adaptive or systematic. However, recently, especially multi-objective optimisation algorithms have started using multiple operators with adaptive schemes [15, 23, 25].

Operator selection requires choosing an operator from a pool of operators, which is set up to let change the state of the problem in-hand to another state. This is done through the representation vector. The states are represented with sets of numbers; either binary, integer, or real numbers. Evolutionary computation has introduced the concepts of phenotype and genotype [13] with which the problem states are represented in a humanly expression and then into numbers to process and compute, then re-converted back to readable expressions by human. Genotypes are operated with existing move functions, i.e. neighbourhood functions applying the process built in the selected operator.

Operator selection stands out as an issue to tackle with an efficient way. It can be either random or applying a systematic rule. For instance, genetic algorithms imply probabilistic selection of crossover and mutation operators, while variable neighbourhood search requires periodic change of operators so as to regularly condense and relieve the search. On the other hand, there is a sound stream of operator selection studies propose adaptive approaches to impose heuristic intelligence into the selection process for efficiency purposes [18]. Adaptive operator selection with credit assignment and multi-armed bandit approaches has been studied by Fialho as reported in [27]. In general, the logic of how operator selection mechanics work is shown in Figure 1 in which an operator is selected by “Selection Scheme” from “Operator Pool” based on the merits gained and applied to the problem state under consideration, then “Evaluate”d for how productive it was. A “Credit Assignment” mechanism picks up the evaluation results and calculates a credit level to supply to the chosen operator’s merit record for its selectability in the future noting that the popularity of the operators is managed based on merits/credits gained.

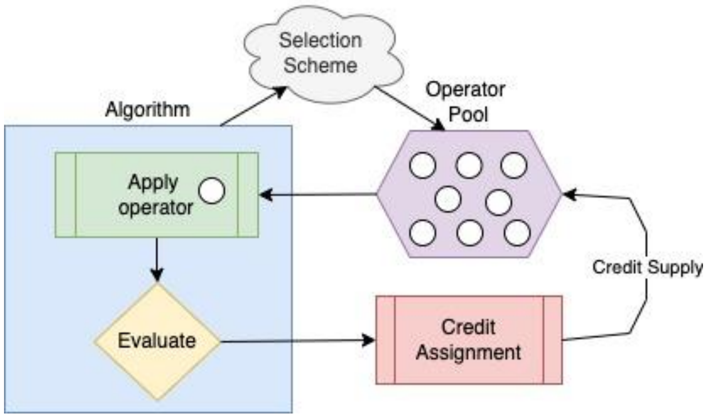

Figure 1. A typical cycle of operator selection mechanism

One of the main concerns of operator selection process is the rule or the scheme to implement for which of the operator from the pool is to be selected for the best move. As mentioned before, the easiest is to do a random selection or a pre-ordered selection, but this would not consider the current circumstances of the search environment and the parametric levels. This fact imposes adaptive approaches. The state-of-the-art approaches are known to be either stochastic or credit-based approaches. As thoroughly discussed in [8] and [12], adaptive approaches include “probability matching” and “adaptive pursuit” are heavily based on credit assignment, while there are a number of other criteria proposed such as fitness-based, diversity-based ones. Many other studies including [15, 18] have been developed on the basis of multi-armed bandit problems.

There are a number of renown operators have been studied and proposed over decades since heuristic optimisation emerged. Operators include random “*flip-flop*”, “*swap*” and “*inverse*” are to assist change a single digit on the representation vector. If it is a binary representation, the operators will change a single bit only to offer a move from current state. There are logic-based operators with which more than one digit of the vector is revised. *AND, OR, NOT* and *XOR* operators have been used to set up simple and complex logic operations to propose new solutions to move to while searching. Furthermore, various genetic operators borrowed from evolutionary algorithms have been implemented and similarities and distance-based operations brought to effect, too. More details can be found in [6, 10, 13].

Recently, machine learning is instrumented to build dynamically functioning credit and reward-based approaches, which enforce to learn from gained experiences either online [6] or off-line [7, 29] to act on the basis of changing environmental circumstances of search process.

2.3 Machine learning for adaptive operator selection

Utilisation of various machine learning techniques into metaheuristics involves many aspects of the search process in order to improve search efficiency in various respects. The main idea behind the use of machine learning in problem solving is to optimise data-driven models to work out with unseen data, which facilitates great opportunities to achieve highly efficient systems. As indicated before, a trend of utilisation from machine learning is observed and ongoing to improve metaheuristic and swarm intelligence optimisation [7, 8, 23, 24, 26]. However, the aim of this paper is to look into the major steps recently taken towards generalisation of problem solving. It is paramount to note that previously developed adaptive operator selection approaches did not take the problem state on board while proposing selection of a particular operator from the pool. The following steps take the problem state on board while making decision to select an operator.

1. The first approach taken on board is the binary representation of the problem states and use of binary operators in managing moves from one state to a neighbouring problem state. The reason behind this step is that any representation approach can be converted into binary and be operated with binary operations regardless of any domain knowledge requirement. This approach has been widely implemented and its efficiency is proven through the development of evolutionary algorithms, especially at the early time. Several recent works have been published using the virtue of binarisation including [6–10]. However, this approach brings the major restriction of scalability, which does not help extensive use of gained experiences. This is due to the fact that every problem state has a specific size of input data and does not support the problems with a different size, which undermines the usability of gained experience over previous problem-solving activities.

2. Transfer learning is a new concept brought forward as part of recent machine learning studies, especially offered by deep learning works. Deep learning offers pre-trained huge neural networks with which any unseen data can be easily handled adapting the pre-existing models into the data and the domain knowledge. However, this remains an important open research issue to enhance the performances further and solve the problems more realistically in a satisfactory level. The idea is to investigate if pre-trained agents can utilise the past experience in solving a completely new problem instance [7]. The study reported in [8] implemented a binarisation based approach to achieve transfer learning built up via reinforcement learning – Q learning – supported with hard-c means clustering algorithm. The learning was conducted across the problem instances in the same size. The results are promising, but the gained experience cannot be applied to solving the problem instances of different sizes.

3. Feature-based problem state representation is another promising approach for generalisation which is brought forward by [28], [29], where prominent analysis is conducted to identify the most relevant features to be selected for this purpose. The problems are better represented with a vector of feature set. The analysis conducted has taken fitness landscape information harvested over the individuals

and the populations and used supervised learning approaches. It sounds very promising to handle problems represented with the features in order to train the models with reinforcement learning and other active learning approaches.

2.4 Reinforcement learning-based adaptive selection scheme

Reinforcement learning is one of hotspot machine learning subjects that attracts so much attention by researchers in the field as well as other disciplines for application purposes. It helps agents /systems learn actively – "active learning" – which facilitates agent training while delivering the tasks. Optimisation and search process is a very dynamically changing environment, which is influenced significantly by the actions taken in the previous stages. It means that an operator selected will change the direction of search and other relevant circumstances hence the next steps will produce taking the output of previous steps into account as the follow-up steps. This works in a sequential way in which one operator is selected and activated, then another is selected accordingly based on the changed circumstances and repeats until the completion. Ultimately, the problem turns into a dynamically built sequence of operators.

The logic embedded in mechanism shown in **Hata! Başvuru kaynağı bulunamadı.** has b een expanded into **Hata! Başvuru kaynağı bulunamadı.** in order to demonstrate how reinforcement is devised and instrumented to let the agent learn and produce credit to the corresponding operator selected and applied. Apparently, "Evaluate" component is the one where this algorithm is embedded. Given that input ***x*** is presented to the "Operator Selection" unit and merged with the output ***o*** generated before and passed to "Cluster Update" unit to label operator ***a*** selected. Once done, operator ***a*** generates a new ***o*** and evaluates the quality of solution with ***F(x,a)***. Meanwhile, ***x*** and ***o*** are passed to the "Reward Generation" unit to produce ***r*** and forwarded to "Cluster Update" to adapt with the new knowledge. This repeats throughout the complete search process.

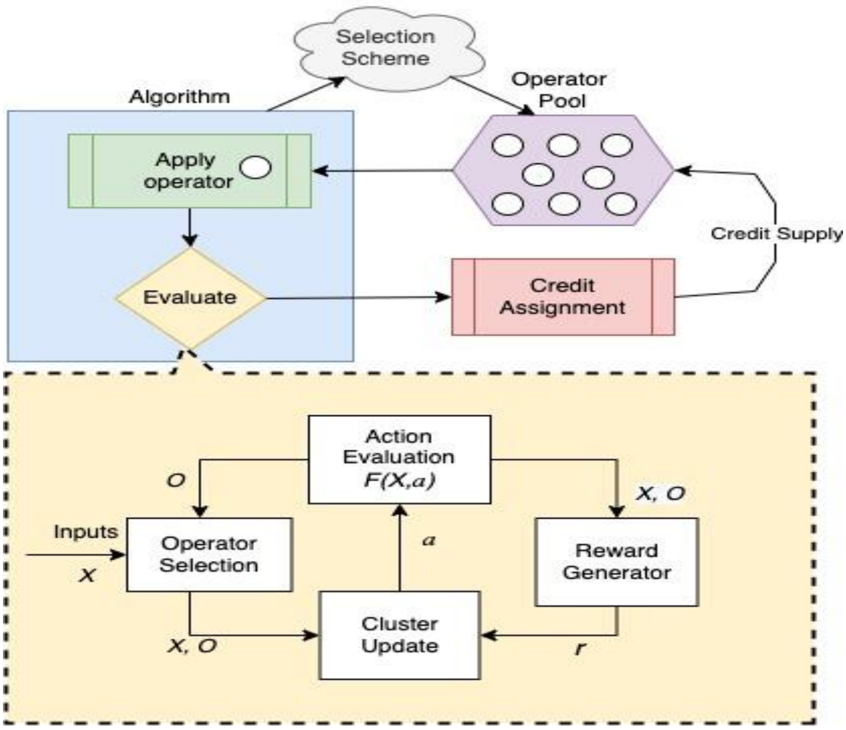

Figure 2. Evaluate function expanded with reinforcement learning components.

3. Results and Discussions

The approaches introduced in the previous subsections have been tested for efficiency with a particular NP-Hard combinatorial problem, namely Set Union Knapsack problem (SUKP). The experimental results are collected running them on a cloud-based high performance compute cluster provided by TUBITAK ULAKBIM. First of all, an ABC framework has been used as a swarm intelligence algorithm with which not much technical constraints are necessarily imposed. A pool of 5 operators is embedded into the ABC implementation, and four variants have been set up to test the above-mentioned ideas to see how helpful the approaches. Similar benchmark instances of SUKP have been solved binary ABC as reported in [30].

Table 2 presents the experimental results of four ABC variants, comparatively, where each is furnished with an embedded pool of operators orchestrated by an operator selection scheme; one of them imposes uniformly randomly selected operators, while the other three variants are bespoken with a reinforcement learning algorithm devised with Q learning and Hard-c-Means algorithms. All four algorithms are compared with respect to mean, standard deviation and the best results (i.e., max value) with respect to the quality of solution measured with fitness function. The tabulated statistics (metrics) have been calculated over 30 trails (runs).

Table 1. Set union knapsack problem (SUKP) benchmark instances.

Problem	*M*	*N*	*W*	*Y*
PI1	400	385	0.15	0.85
PI2	500	485	0.10	0.75
PI3	100	100	0.15	0.85
PI4	400	385	0.10	0.75
PI5	100	100	0.10	0.75
PI6	200	200	0.10	0.75
PI7	200	200	0.15	0.85
PI8	185	200	0.10	0.75
PI9	300	300	0.15	0.85
PI10	200	185	0.15	0.85

Table 2. Comparative results provided by four variants of ABC algorithm for solving 10 instances of SUKP.

	Random			RLABC			RLABC-FL			RLABC-TL		
Problem	**Max**	**Mean**	**STD**	**Max**	**Mean**	**STD**	**Max**	**Mean**	**STD**	**Max**	**Mean**	**STD**
PI1	10168	9997.7	188.3	10168	10027.1	145.2	10168	10055.1	138.0	10175	10123.9	84.4
PI2	11326	11076.9	147.5	11427	11188.1	140.5	11426	11118.6	151.0	11490	11196.1	134.7
PI3	13407	13205.3	206.9	13402	13271.7	100.2	13407	13222.7	149.3	13405	13283.0	67.7
PI4	10852	10512.2	179.2	10877	10647.3	101.1	10994	10616.7	180.0	10831	10626.4	90.5
PI5	13963	13822.4	74.5	14044	13949.0	85.1	14044	13850.6	79.5	14044	13943.2	86.4
PI6	12257	11716.6	255.5	12211	11833.3	178.2	12350	11792.0	253.9	12328	11944.3	201.9
PI7	11800	11491.2	204.6	12019	11652.0	163.0	11821	11550.1	257.2	11821	11627.4	201.4
PI8	13463	13097.6	228.6	13402	13271.7	100.2	13392	13141.2	174.8	13405	13283.0	67.7
PI9	10724	10592.2	177.4	11410	10679.6	176.8	10735	10618.9	118.8	11054	10759.6	144.6
PI10	13671	13230.3	144.7	13609	13352.2	130.0	13671	13376.1	191.4	13609	13399.0	99.7

RLABC is an ABC variant uses binary representation of the problem states and operates the solutions with selecting one operator from a tool of three different operators, training the decision-making agent with reinforcement learning (RL) algorithm – mentioned above – when and how to select each of the operator given the problem state and search circumstances [6]. RLABC-TL uses the same representation and set of operators for transfer learning across different runs of the problem instances utilising the same RL algorithm, but transferring the gained experiences learned previously [7], [8]. On the other hand, RLABC-FL uses a feature-based problems representation, trains the agent with the same RL algorithm to select one of the operators from the pool more efficiently noting that the set of operators in the pool are different from the other three variants. It is open to transfer learning, too.

The results tabulated in Table 2 have been ranked with Wilcoxon sign test, accordingly, to find out the best and the worst performing algorithms. The ranks are presented in Table 3 with respect to 2 metrics; “mean” and the “max” values, where the rank spans from 1 to 4 indicating that 1 is the best and 4 is the worst. The overall performance by each algorithm is averaged at the bottom of the table, where RLABC-TL overperforms the rest in both measures, and the runner up is RLABC in “mean” – with 1.3 versus 1.9 – and RLABC-FL in “max” values – with 1.7 versus 1.9. It is important to note that RLABC-FL may not be properly comparable due to the fact that it uses a different set of operators in the pool. Nevertheless, RLABC-FL demonstrates a clear potential in comparisons of the right-hand-side of the table.

Table 3**.** Comparative results with respect to the ranks collated from both means and maximum results by each of the algorithms.

	Comparisons with Mean in rank				**Comparisons with Max in rank**			
Problem	**Random**	**RLABC**	**RLABC-FL**	**RLABC-TL**	**Random**	**RLABC**	**RLABC-FL**	**RLABC-TL**
PI1	4	3	2	1	2	2	2	1
PI2	4	2	3	1	4	2	3	1
PI3	4	2	3	1	1	4	1	2
PI4	4	1	3	2	3	2	1	3
PI5	4	1	3	2	4	1	1	1
PI6	4	2	3	1	3	4	1	2
PI7	4	1	3	2	4	1	2	2
PI8	4	2	3	1	1	3	4	1
PI9	4	2	3	1	4	1	3	2
PI10	4	3	2	1	1	3	1	2
Mean:	4	1.9	2.8	1.3	2.7	2.3	1.9	1.7

The results collected from the experimentation of RLABC-FL are looked at to realise how contributing is each of the features in the learning process. Figure 1 and Figure 2 show the learning progress of the operator selection agent on which operator has been selected and activated across the whole span of iterations. Apparently, each figure is a multi-plot including 12 plots representing the 12 features found more effective, where each cell – of both figures – indicates the learning progression of the agent by the means of a particular feature. Here, Figure 1 displays the scattered data collected in the first run (i.e. 1st trail of experimentation) per features indicating the selected operators across iterations, while Figure 2 plots the scattered data taken from the last run in the same way to realise the differences in between the approximations through the features. This characterises how good the agent has learned from the first to the last trail / run. All features seem stable except 7th and 11th features, which look not discriminative, sufficiently.

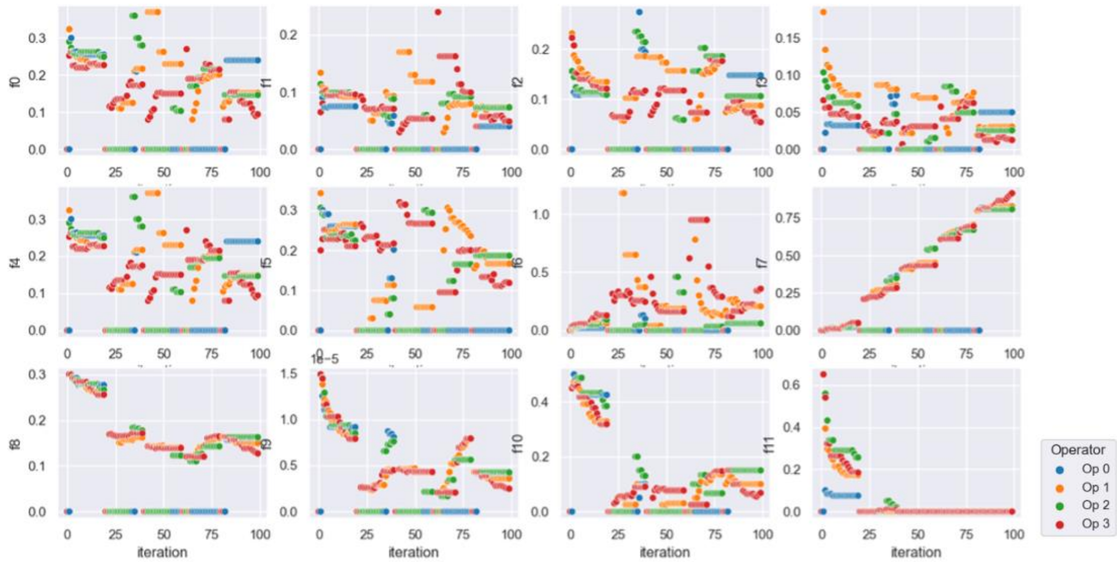

Figure 1. The predictiveness of the features in the first run.

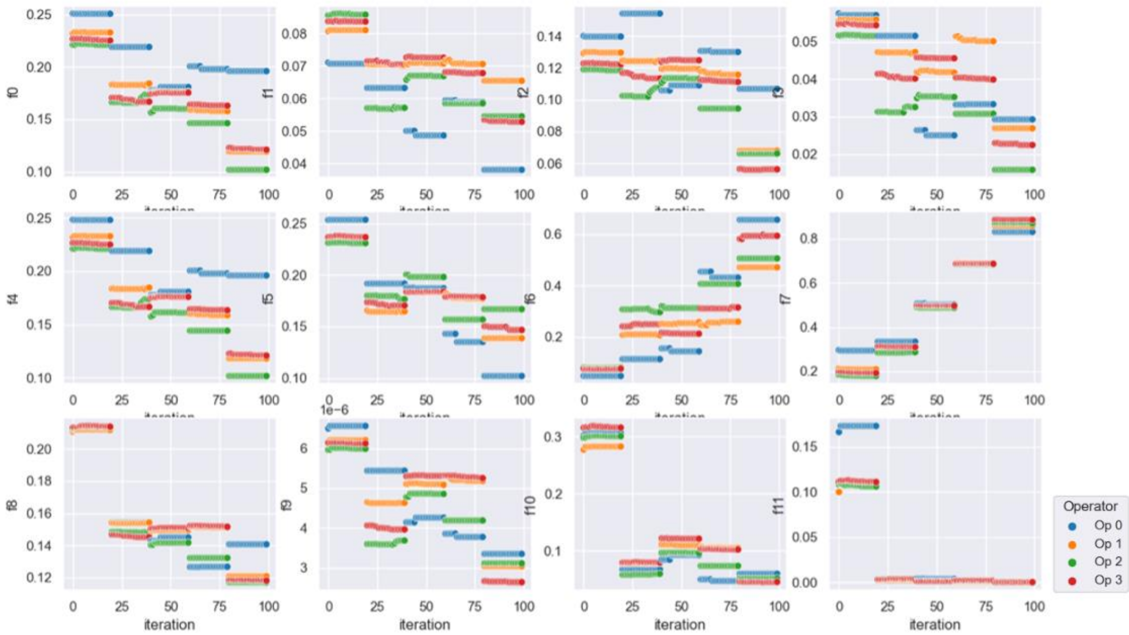

Figure 2. The predictiveness of the features in the last run.

4. Conclusion

This article overviews and discusses use of machine learning, especially reinforcement learning, to improve the efficiency of metaheuristic-based search algorithms with a generalisation point of view. Machine learning has recently been vastly used to model data and help handle real-world problems in a data-driven approach. It offers ways to facilitate domain-awareness in handling the problems. Especially, optimisation problems are normally solved with non-guided search algorithms, which do not use domain knowledge. But with use of machine learning, especially reinforcement learning, the optimisation algorithms can be furnished with data-driven facilities to tackle the problems with domain awareness.

The article focuses on variations of reinforcement learning and data-driven approaches to improve swarm intelligence algorithms, particularly artificial bee colony (ABC) variants. ABC variants are instrumented with adaptive operator selection scheme built with reinforcement learning to solve set union knapsack problem as one of prominent NP-Hard combinatorial optimisation problems. Three variants have been compared initially with a random operator selection scheme and next with one another. The results suggest that reinforcement learning seems promising for building adaptive operators selection scheme, particularly, transfer learning looks a bright way out for this purpose, which needs to be further studied and investigated. Once accomplished, a substantial generalisation would be achieved.

This area of study needs to be extended towards multi-objective optimisation domain with the view that each objective would impose a particular interest in gaining experience and knowledge. Subsequently, there might appear conflicts among the sources of reinforcements, which requires to be resolved. Various aspects of the subject have been tackled so far, but more outstanding issues and aspects need to be studied.

References

[1] G. J. Woeginger, 'Exact Algorithms for NP-Hard Problems: A Survey', in *Combinatorial Optimization — Eureka, You Shrink!: Papers Dedicated to Jack Edmonds 5th International Workshop Aussois, France, March 5–9, 2001 Revised Papers*, G. and R. G. Jünger Michael and Reinelt, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 185–207. doi: 10.1007/3-540-36478-1_17.

[2] M. Črepinšek, S.-H. Liu, and M. Mernik, 'Exploration and exploitation in evolutionary algorithms: A survey', *ACM computing surveys (CSUR)*, vol. 45, no. 3, pp. 1–33, 2013.

[3] D. H. Wolpert and W. G. Macready, 'No free lunch theorems for optimization', *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.

[4] K.-P. L. Proctor Robert W. and Vu, 'Human Information Processing', in *Encyclopedia of the Sciences of Learning*, N. M. Seel, Ed., Boston, MA: Springer US, 2012, pp. 1458–1460. doi: 10.1007/978-1-4419-1428-6_722.

[5] D. B. Leake, 'Problem Solving and Reasoning: Case-based', *International Encyclopedia of the Social & Behavioral Sciences*, pp. 12117–12120, 2001, doi: 10.1016/B0-08-043076-7/00545-3.

[6] R. Durgut and M. E. Aydin, 'Adaptive binary artificial bee colony algorithm', *Appl Soft Comput*, vol. 101, p. 107054, Mar. 2021, doi: 10.1016/J.ASOC.2020.107054.

[7] R. Durgut, M. E. Aydin, and A. Rakib, 'Transfer Learning for Operator Selection: A Reinforcement Learning Approach', *Algorithms*, vol. 15, no. 1, Jan. 2022, doi: 10.3390/A15010024.

[8] R. Durgut, M. E. Aydin, and I. Atli, 'Adaptive operator selection with reinforcement learning', *Inf Sci (N Y)*, vol. 581, pp. 773–790, Dec. 2021, doi: 10.1016/J.INS.2021.10.025.

[9] E.-G. Talbi, 'Machine Learning into Metaheuristics: A Survey and Taxonomy', *ACM Comput. Surv.*, vol. 54, no. 6, Jul. 2021, doi: 10.1145/3459664.

[10] M. Tessari and G. Iacca, 'Reinforcement Learning Based Adaptive Metaheuristics', in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, in GECCO '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 1854–1861. doi: 10.1145/3520304.3533983.

[11] T. Wauters, K. Verbeeck, P. de Causmaecker, and G. vanden Berghe, 'Boosting metaheuristic search using reinforcement learning', in *Hybrid metaheuristics*, Springer, 2013, pp. 433–452.

[12] J. C. Bansal, H. Sharma, and S. S. Jadon, 'Artificial bee colony algorithm: a survey', *International Journal of Advanced Intelligence Paradigms*, vol. 5, no. 1–2, pp. 123–159, Jan. 2013, doi: 10.1504/IJAIP.2013.054681.

[13] C. Ozturk, E. Hancer, and D. Karaboga, 'A novel binary artificial bee colony algorithm based on genetic operators', *Inf Sci (N Y)*, vol. 297, pp. 154–170, Mar. 2015, doi: 10.1016/J.INS.2014.10.060.

[14] R. DURGUT and M. AYDİN, 'Çok boyutlu sırt çantası problemi için adaptif ikili yapay arı kolonisi algoritması (AİYAK)', *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, May 2021, doi: 10.17341/gazimmfd.804858.

[15] M. Düğenci and M. E. Aydin, 'A honeybees-inspired heuristic algorithm for numerical optimisation', *Neural Comput Appl*, vol. 32, no. 16, pp. 12311–12325, Aug. 2020, doi: 10.1007/s00521-019-04533-x.

[16] R. Durgut and M. E. Aydin, 'Multi Strategy Search with Crow Search Algorithm', in *Optimisation Algorithms and Swarm Intelligence*, Prof. N. Vakhania, Ed., Rijeka: IntechOpen, 2022. doi: 10.5772/intechopen.102862.

[17] Z. Cui *et al.*, 'Hybrid many-objective particle swarm optimization algorithm for green coal production problem', *Inf Sci (N Y)*, vol. 518, pp. 256–271, May 2020, doi: 10.1016/j.ins.2020.01.018.

[18] D. Kizilay, M. F. Tasgetiren, H. Oztop, L. Kandiller, and P. N. Suganthan, 'A Differential Evolution Algorithm with Q-Learning for Solving Engineering Design Problems', in *2020 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, Jul. 2020, pp. 1–8. doi: 10.1109/CEC48606.2020.9185743.

[19] W. xiang Wang, K. shun Li, X. zhen Tao, and F. hui Gu, 'An improved MOEA/D algorithm with an adaptive evolutionary strategy', *Inf Sci (N Y)*, vol. 539, pp. 1–15, Oct. 2020, doi: 10.1016/J.INS.2020.05.082.

[20] Q. Lin *et al.*, 'Adaptive composite operator selection and parameter control for multiobjective evolutionary algorithm', *Inf Sci (N Y)*, vol. 339, pp. 332–352, Apr. 2016, doi: 10.1016/J.INS.2015.12.022.

[21] C. Qu, W. Gai, M. Zhong, and J. Zhang, 'A novel reinforcement learning based grey wolf optimizer algorithm for unmanned aerial vehicles (UAVs) path planning', *Appl Soft Comput*, vol. 89, p. 106099, Apr. 2020, doi: 10.1016/J.ASOC.2020.106099.

[22] K. Li, A. Fialho, S. Kwong, and Q. Zhang, 'Adaptive operator selection with bandits for a multiobjective evolutionary algorithm based on decomposition', *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 1, pp. 114–130, 2013.

[23] M. A. Al-Betar, I. A. Doush, A. T. Khader, and M. A. Awadallah, 'Novel selection schemes for harmony search', *Appl Math Comput*, vol. 218, no. 10, pp. 6095–6117, Jan. 2012, doi: 10.1016/J.AMC.2011.11.095.

[24] K. M. Sallam, S. M. Elsayed, R. A. Sarker, and D. L. Essam, 'Landscape-based adaptive operator selection mechanism for differential evolution', *Inf Sci (N Y)*, vol. 418–419, pp. 383–404, Dec. 2017, doi: 10.1016/J.INS.2017.08.028.

[25] W. Khan Mashwani, A. Salhi, O. Yeniay, H. Hussian, and M. A. Jan, 'Hybrid non-dominated sorting genetic algorithm with adaptive operators selection', *Appl Soft Comput*, vol. 56, pp. 1–18, Jul. 2017, doi: 10.1016/J.ASOC.2017.01.056.

[26] D. B. Fogel, 'Phenotypes, genotypes, and operators in evolutionary computation', in *Proceedings of 1995 IEEE International Conference on Evolutionary Computation*, p. 193.

[27] Á. Fialho, 'Adaptive Operator Selection for Optimization', Université Paris Sud - Paris XI, 2010. [Online]. Available: https://tel.archives-ouvertes.fr/tel-00578431

[28] R. Durgut, M. E. Aydin, H. Ihshaish, and R. Abdur, 'Analysing the Predictivity of Features to Characterise the Search Space', in *Artificial Neural Networks and Machine Learning – ICANN 2022*, E. Pimenidis and M. E. Aydin, Eds., Bristol: Springer, Sep. 2022, pp. 1–13.

[29] M. E. Aydin, R. Durgut, A. Rakib, and H. Ihshaish, 'Feature-based search space characterisation for data-driven adaptive operator selection', *Evolving Systems*, Dec. 2023, doi: 10.1007/s12530-023-09560-7.

[30] Y. He, H. Xie, T. L. Wong, and X. Wang, 'A novel binary artificial bee colony algorithm for the set-union knapsack problem', *Future Generation Computer Systems*, vol. 78, pp. 77–86, Jan. 2018, doi: 10.1016/J.FUTURE.2017.05.044.

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

RESEARCH ARTICLE

Prediction of Scoliosis Risk in Adolescents with Machine Learning Models

Murat Ali Çınar [a†], **İbrahim Küçükcan** [b]

[a] Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, Gaziantep, Türkiye
[b] Ütopya Physiotherapy Consultancy Center, Gaziantep, Türkiye
[†] muratali.cinar@hku.edu.tr, corresponding author

RECEIVED MARCH 25, 2024
ACCEPTED APRIL 16, 2024

CITATION Çınar, M.A. & Küçükcan, İ. (2024). Prediction of scoliosis risk in adolescents with machine learning models. *Artificial Intelligence Theory and Applications, 4(*1), 33-42.

Abstract

When considering classifications of scoliosis, 'idiopathic scoliosis' emerges as the most common type. Alongside spinal alterations, individuals with scoliosis undergo changes in stability and gait while standing. Despite existing literature exploring the progression of scoliosis and its impact on foot pressure among those diagnosed with adolescent idiopathic scoliosis, no studies have been found regarding the prediction of scoliosis risk in healthy adolescents. This study aims to develop an machine learning based decision support system capable of forecasting scoliosis risk in adolescents using foot pressure analysis values and machine learning models.

The study encompassed 20 patients diagnosed with adolescent idiopathic scoliosis and 43 healthy adolescents exhibiting similar demographic characteristics, totaling 63 patients. Plantar pressure distributions of all participants were measured both statically and dynamically.

The data collected for all patients comprised: age, sex, percentage of right hindfoot static plantar pressure, percentage of left hindfoot static plantar pressure, percentage of right forefoot static plantar pressure, percentage of left forefoot static plantar pressure, percentage of right foot dynamic plantar pressure, and percentage of left foot dynamic plantar pressure. A dataset including pressure percentages and the presence of scoliosis diagnosis was constructed, consisting of 8 input variables and 1 outcome variable for each patient.

The most effective predictors of adolescent idiopathic scoliosis risk were identified as follows: Subspace KNN (100%), RUS Boosted Trees (100%), Weighted KNN (100%), Bagged Trees (100%), and Fine KNN (100%)

Keywords: adolescent idiopathic scoliosis; plantar pressure distribution; machine learning

1. Introduction

The term "scoliosis" comes from the Greek and means "crooked" or "curved". It was first defined and introduced into the literature by Hippocrates [1-3]. In the Scientific Society on Scoliosis Orthopedic and Rehabilitation Treatment (SOSORT) guide published in 2016, scoliosis was described as a group of conditions that result in various deformities in the shape of the spine, thorax and trunk [1]. Although there are different definitions of scoliosis, they all converge on the fact that it involves a lateral curvature of the spine of more than 10° (Figure 1) [1-4]. There are many classifications of scoliosis; in 1973, the Scoliosis Research Society (SRS) divided scoliosis into two groups: structural and non-

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

structural [4].In non-structural functional scoliosis, spinal curvature develops due to causes outside the spine. There is often shortness of the lower extremities or asymmetry in the tone of the paraspinal muscles. A person with non-structural scoliosis can correct posture. In structural scoliosis, the person has a loss of flexibility and needs treatment to correct the curvature [1-4]. This classification by the SRS is shown in Table-1 [1-4].

Figure 1. Radiological image of an individual with scoliosis [1-4].

Table 1. Scoliosis classification of SRS.

Structural scoliosis	**Non-structural scoliosis (functional scoliosis)**
Idiopathic scoliosis Infantile Juvenile Adolescent	Postural scoliosis
Neuromuscular scoliosis Neuropathic Myopathic	Hysterical
Congenital scoliosis	Caused by nerve root irritation
Neurofibromatosis	Caused by hip contractures
Scoliosis due to connective tissue disorder	Caused by leg length inequality
Osteochondrodystrophy	Inflammatory (appendicitis etc.) related
Due to metabolic disorders	
Traumatic	
Scoliosis caused by tumors or infection	
Scoliosis due to rheumatic diseases	
Scoliosis due to pathologies in the lumbosacral region	

When looking at classifications of scoliosis, "idiopathic scoliosis" appears to be the most common type of scoliosis [1-4]. This term, which was introduced into the literature in 1922, is defined as situations in which no specific disease-causing deformity of the spine can be found [5].

Idiopathic scoliosis is divided into four groups based on the age of onset: infantile (0-3 years), juvenile (3-10 years), adolescent (10-18 years) and adult (18 years and older) [4-6].

Table 2. Classifications of idiopathic scoliosis [4-6]

Chronological	**Angular**		**Topographic**		
Age at diagnosis	**Cobb degrees**		**Apex**		
				from	**to**
Infantile	Low	Up to 20	Cervical	–	Disc C6–7
Juvenile	Moderate	21–35	Cervico-thoracic	C7	T1
Adolescent	Moderate to severe	36–40	Thoracic	Disc T1–2	Disc T11–12
Adult	Severe Severe to very severe Very severe	41–50 51–55 56 or more	Thoraco-lumbar Lumbar	T12	L1 Disc L1–2

Relationship Between Scoliosis and Plantar Pressure

In patients with scoliosis, changes in the spine are accompanied by changes in stability during standing and walking [7]. Because scoliosis affects the biomechanics of the spine in three dimensions, changes in spinal mobility and posture occur, causing movement patterns to change with each step [6-8]. The deformed spine shifts the body's center of mass to help maintain trunk balance, resulting in asymmetry and various gait abnormalities [6-8].

When examining the biomechanics of gait, the literature indicates that the pelvis and spine are intimately involved in the gait process [9]. A study conducted in 2023 highlighted that both static and dynamic plantar pressures are abnormally altered in individuals with adolescent idiopathic scoliosis and require treatment [10]. It has been suggested in the literature that these changes in plantar pressure may aid in the diagnosis of scoliosis, highlighting the need for further research in this area [11-13].

While there have been studies on the progression of scoliosis and its effects on base pressure in individuals diagnosed with adolescent idiopathic scoliosis in the literature, no studies have been encountered predicting the risk of scoliosis in healthy adolescents. The aim of this study is to develop an artificial neural network (ANN)-based decision support system that can predict the risk of scoliosis in adolescents using foot pressure analysis values and machine learning models. In addition, the dataset obtained from this study can serve as a preliminary study for researchers working in the field of scoliosis who wish to conduct research in the field of artificial intelligence.

2. Materials and Methods

The study conducted at Hasan Kalyoncu University Physiotherapy and Rehabilitation Department included 20 patients who were diagnosed with adolescent idiopathic scoliosis (AIS) by a specialist physician and who applied to Gaziantep Utopya Physiotherapy Consultancy Center to receive physiotherapy. 43 healthy adolescent individuals with similar demographic characteristics to the 20 included patients were also included in the study, and a data set was created with the data of a total of 63 participants.
The plantar pressure distributions of all participants were measured using two methods: static and dynamic. Both static and dynamic measurements were performed using the Ottobock Esco Scan device (Germany) and the Presto-Scan, Class I Rule 1, per MDD 93/42/EEC Annex IX, USA software (see Figure 1). The device is approximately 5 mm thick and has a sensor area of 44 x 37 cm with a total of 2288 sensors. It uses resistive sensor technology and can collect pressure and force data up to forty Hertz. Static measurements were taken while the subjects stood in a relaxed position, concentrating on a fixed point in front of them. Percentage values of the total contact area of both feet, including forefoot and hindfoot, were obtained through static evaluation. For dynamic measurements, a two-step protocol was used, utilizing the device's ability to colour code foot pressure points based on pressure percentages.

Figure 2 Plantar pressure analysis (representative image)

The data set for all patients was compiled with the following information: age, sex, percentage of right hindfoot static plantar pressure, percentage of left hindfoot static plantar pressure, percentage of right forefoot static plantar pressure, percentage of left forefoot static plantar pressure, percentage of right foot dynamic plantar pressure, percentage of left foot dynamic plantar pressure, and presence of scoliosis diagnosis (see Table 3). The dataset was randomly divided into two separate sets: 70% of the data was allocated for training the artificial neural network (ANN), while the remaining 30% was reserved for testing the model's performance. Table 3 below presents the dataset, comprising 8 input variables and 1 output variable collected for each patient.

Table 3. Dataset of features collected from patients

Features
Gender
Age
Right hindfoot static plantar pressure percentage
Left hindfoot static plantar pressure percentage
Right forefoot static plantar pressure percentage
Left forefoot static plantar pressure percentage
Right foot dynamic plantar pressure percentage
Left foot dynamic plantar pressure percentage
Outcome- Adolescent idiopathic scoliosis

The study was approved by the Hasan Kalyoncu University Health Sciences Ethics Committee. Informed consent forms were signed and permission to use the data was obtained from all patients included in the study.

2.1 Accuracy

Accuracy is a common metric for assessing the performance of a model, but there are situations where it should be considered. In particular, if there are unbalanced classes (i.e. large differences in sample counts between classes), accuracy may not be a sufficient metric and other metrics (e.g. precision, sensitivity) should also be considered.

In this study, the accuracy metric is used to evaluate the performance of machine learning methods.

2.2 K-fold cross validation

K-fold cross validation is a widely used method for evaluating the performance of a machine learning model. In this method, the data set is randomly divided into k parts (usually 5 or 10). Then one of these k parts is used as the test set, while the other k-1 parts are used as the training set. The model is trained once, and each time a different part is selected as the test set. The results are combined and the overall performance is measured. This method is used to assess how generalizable the model is, as it is tested on different pieces of data. In this way, the overall performance of the model can be more reliably assessed without relying on a single test set. In this study, 3,5 and 10 k were tested.

3. Experimental Results

This study used 25 different machine learning techniques. Each algorithm used different activation functions, optimization algorithms and loss functions, as detailed in Table 4, with or without PCA. All these algorithms were implemented using the machine learning toolbox available in the MATLAB programming language. The numerical results were derived using MATLAB R2021b on an Intel processor running on the Windows 10 platform.

The best performers in predicting the risk of adolescent idiopathic scoliosis were determined to be: Subspace KNN (100%), RUS Boosted Trees (100%), Weighted KNN (100), Bagged Trees (100%), Fine KNN (100%). In addition, the PCA method was used to try different parameter variations and the best results are shown in Table 4. The values of the most and least successful algorithms (confusion matrix) of the dataset are shown in Figures 3 and 4. In this study, the Principal Component Analysis (PCA) method has been used to demonstrate whether there is an improvement in the results from a feature engineering perspective. By applying the PCA method with a ratio of 7/8, it is possible to achieve the same success in the results obtained in the experimental study with the best 7 features out of 8. This was tested to reduce the computational complexity. It was observed that, due to the small number of features, it did not have a positive impact on the performance, as can be seen in Table 4.

Table 4. Machine learning techniques for comparison (accuracy %)

Machine Learning Models	TV		3-Fold CV		5-Fold CV		10-Fold CV	
	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable
Fine Tree	90.3%	87.1%	66.1%	58.1%	72.6%	67.7%	74.2%	67.7%
Medium Tree	90.3%	87.1%	66.1%	58.1%	72.6%	67.7%	74.2%	67.7%
Coarse Tree	87.1%	82.3%	66.1%	58.1%	72.6%	74.2%	72.6%	64.5%
Linear Discriminant	77.4 %	72.6%	66.1%	69.4%	72.6%	71.0%	72.6%	69.4%
Logistic Regression	77.4%	75.8%	67.7%	71.0%	75.8%	71.0%	72.6%	69.4%
Gaussian Naive Bayes	79.0%	79.0%	69.4%	75.8%	75.8%	74.2%	75.8%	77.4%
Kernel Naive Bayes	77.4%	79.0%	66.1%	75.8%	67.7%	67.7%	71.0%	72.6%
Linear SVM	75.8%	74.2%	69.4%	69.4%	71.0%	69.4%	71.0%	69.4%
Quadratic SVM	88.7%	82.3%	**59.7%****	61.3%	64.5%	72.6%	66.1%	66.1%
Cubic SVM	96.8%	93.5%	62.9%	66.1%	62.9%	66.1%	64.5%	69.4%
Fine Gaussian SVM	98.4%	91.9%	69.4%	71.0%	67.7%	71.0%	71.0%	71.0%
Medium Gaussian SVM	83.9%	82.3%	66.1%	64.5%	66.1%	67.7%	72.6%	67.7%
Coarse Gaussian SVM	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%
Fine KNN	**100%***	**100%***	66.1%	67.7%	64.5%	72.6%	69.4%	75.8%
Medium KNN	74.2%	77.4%	69.4%	64.5%	71.0%	64.5%	71.0%	62.9%
Coarse KNN	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%	67.7%
Cosine KNN	77.4%	74.2%	62.9%	67.7%	67.7%	69.4%	69.4%	67.7%
Cubic KNN	79.0%	77.4%	71.0%	69.4%	66.1%	69.4%	72.6%	64.5%
Weighted KNN	**100%***	**100%***	62.9%	67.7%	66.1%	71.0%	74.2%	71.0%
Boosted Trees	67.7%	67.7%	67.7%	67.7%	67.7 %	67.7%	71.0%	67.7%
Bagged Trees	**100%***	**100%***	62.9%	66.1%	74.2 %	69.4%	74.2%	67.7%
Subspace Discriminant	77.4%	75.8%	71.0%	71.0%	**75.8%**	72.6%	71.0%	71.0%
Subspace KNN	**100%**	**100%**	69.4%	67.7%	**69.4%**	**69.4%**	**74.2%**	72.6%
RUS Boosted Trees	**100%**	**100%**	64.5%	58.1%	72.6%	72.6%	77.4%	71.0%

* Best accuracy, ** worst accuracy

In this study, since the total input vector consists of 63 cases, the k-fold cross validation method did not improve the performance. Therefore, this method is not recommended for studies with a small input vector. The experimental results of the study are presented in Table 4 and analyzed in terms of feature engineering techniques. As a result, the traditional value data splitting method is recommended for studies with low input vector.

In order to prevent the model from overlearning, both cross-validation and feature selection methods have been applied and performance degradation has been observed in the results. To overcome this problem, performance can be improved by increasing the amount of data.

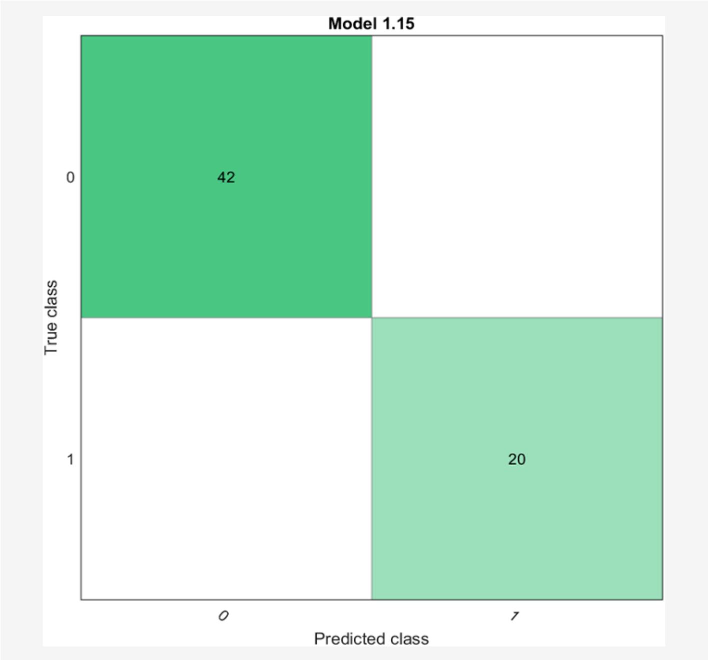

Figure 3. Confusion Matrix for the most successful Fine KNN algorithm.

Figure 4. Confusion Matrix for the worst successful Quadratic SVM algorithm.

The ROC curves of the best-performing Fine KNN algorithm and the worst-performing Quadratic SVM algorithms are shown in Figure 5 and Figure 6, respectively.

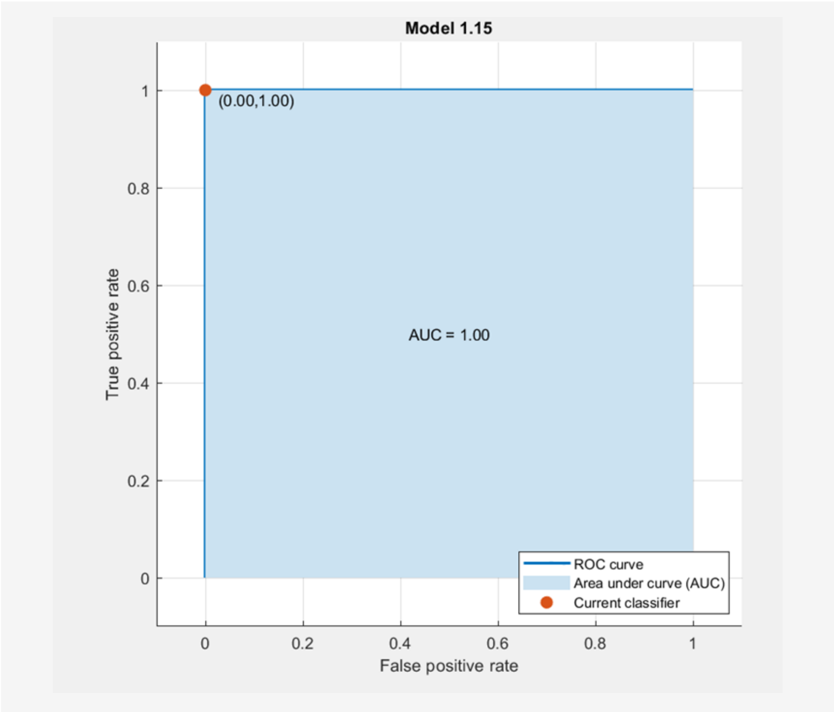

Figure 5. ROC curve for the best performing Fine KNN algorithm.

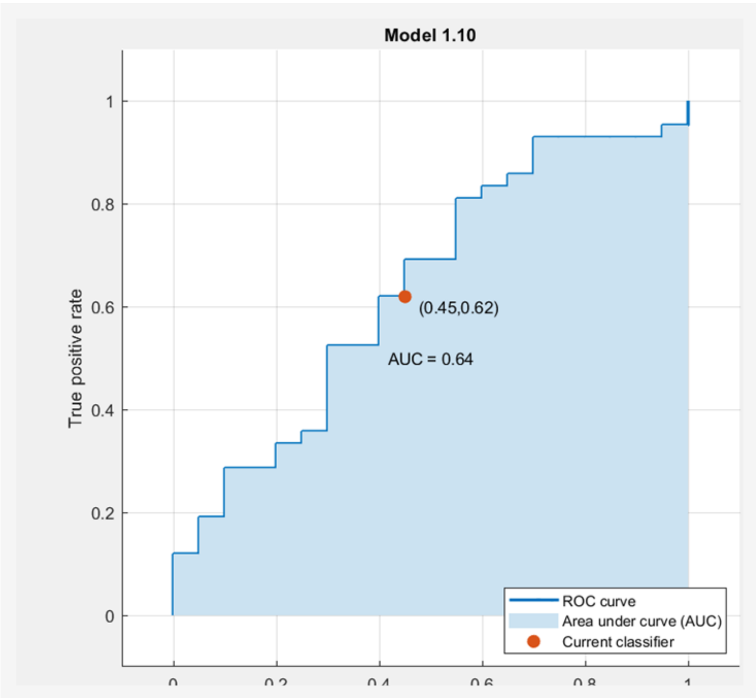

Figure 6. ROC curve for the worst-accuracy performance of Quadratic SVM algorithm.

4. Discussion

In this study, machine learning models were employed for predicting the risk of idiopathic scoliosis in adolescents. The utilized machine learning models incorporated plantar pressure distribution data, revealing models that exhibited 100% performance (see Table 4). Leveraging artificial intelligence for scoliosis risk prediction in this study holds promising prospects. Additionally, we believe that the data obtained from this study serves as a preliminary exploration towards establishing a decision support system based on artificial neural networks (ANN) capable of predicting scoliosis risk in adolescents.
In a 2019 study by XU et al. [14], genetic factors potentially linked to the prognosis of AIS in diagnosed individuals were investigated. They highlighted the potential influence of 10 genetic variants on AIS susceptibility [14]. These variants identified in genetic factor analysis were deemed influential [14], although it's noted that individual testing for each patient may be necessary, posing potential cost challenges [14]. Considering the financial constraints associated with genetic testing for every patient, we propose that plantar pressure analysis coupled with machine learning models offers a cost-effective and rapid alternative for scoliosis risk prediction. However, we acknowledge that plantar pressure analysis alone may not suffice for predicting prognosis in diagnosed individuals. A systematic review published in 2021 emphasized the necessity of developing a patient-specific prediction system for the progression of scoliosis [15]. The review emphasized the insufficiency of relying solely on radiological findings and classification systems [15]. We believe that the dataset obtained from our study holds promise for developing such prediction systems.

In a study by Lv et al. [16], the efficacy of machine learning models for predicting scoliosis risk was assessed. Data including sitting height, biomechanical properties of the lumbar region, pelvis, and shoulder were utilized across five different machine learning models [16]. Radiological imaging was employed in these methodologies, culminating in the creation of a dataset derived from calculations performed on radiological images obtained from patients. Within this dataset, five distinct machine learning models were implemented alongside their respective sets: the Random Forest Model (RFM), Support Vector Machine Model, Artificial Neural Network Model (ANNM), Decision Tree Model (DTM), and Generalized Linear Model (GLM). In our investigation, a total of 25 diverse machine learning models were utilized. We posit that our study holds potential to significantly enrich the existing literature in this domain. Notably, the plantar pressure analysis conducted in our study incurred no costs for either patients or healthy individuals, and the utilized pressure analysis method is devoid of any harmful radiation. A notable strength of our study lies in the absence of adverse effects on patients stemming from the obtained dataset.

5. Conclusion

Radiological evaluations and related algorithms are available for predicting the prognosis of scoliosis in diagnosed individuals. However, predicting the risk of scoliosis in healthy adolescents remains challenging. The machine learning models derived from this study can offer a solution for predicting scoliosis risk in healthy adolescents. Moreover, we contend that incorporating data from plantar pressure analysis into machine learning models in this study will yield significant contributions to the literature. Consequently, predicting scoliosis risk in adolescents using plantar pressure analysis values and machine learning models can provide valuable insights for clinicians in this field. Furthermore, we believe that the dataset obtained from this study can serve as a valuable resource for researchers in the field of scoliosis and those interested in conducting research in artificial intelligence.

References

[1] Negrini, S., Donzelli, S., Aulisa, A. G., Czaprowski, D., Schreiber, S., de Mauroy, J. C., ... & Zaina, F. (2018). 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis and Spinal Disorders, 13, 1-48.
[2] Choudhry, M. N., Ahmad, Z., & Verma, R. (2016). Adolescent Idiopathic Scoliosis. The Open Orthopaedics Journal, 10(1), 143. https://doi.org/10.2174/1874325001610010143
[3] Zagalaz-Anula, N., León-Morillas, F., Andradre-Ortega, J. A., Ibáñez-Vera, A. J., de Oliveira Sousa, S. L., & Lomas-Vega, R. (2020). Case Report: Conservative Treatment of Adolescent Idiopathic Scoliosis Can Alter the Perception of Verticality. A Preliminary Study. Frontiers in Pediatrics, 8, 609555. https://doi.org/10.3389/FPED.2020.609555
[4] Gámiz-Bermúdez, F., Obrero-Gaitán, E., Zagalaz-Anula, N., & Lomas-Vega, R. (2022). Corrective exercise-based therapy for adolescent idiopathic scoliosis: Systematic review and meta-analysis. Clinical Rehabilitation, 36(5), 597-608.
[5] Szopa, A., & Domagalska-Szopa, M. (2017). Correlation between respiratory function and spine and thorax deformity in children with mild scoliosis. Medicine, 96(22).
[6] Negrini, S., Aulisa, A. G., Aulisa, L., Circo, A. B., De Mauroy, J. C., Durmala, J., ... & Maruyama, T. (2012). 2011 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis, 7(1), 1-35.
[7] Daryabor, A., Arazpour, M., Sharifi, G., Bani, M. A., Aboutorabi, A., & Golchin, N. (2017). Gait and energy consumption in adolescent idiopathic scoliosis: A literature review. Ann Phys Rehabil Med., 60(2), 107–116. doi: 10.1016/j.rehab.2016.10.008
[8] Park, H. J., Sim, T., Suh S. W., Yang, J. H., Koo, H., & Mun, J. H. (2016). Analysis of coordination between thoracic and pelvic kinematic movements during gait in adolescents with idiopathic scoliosis. Eur Spine J., 25, 385 –394. doi: 10.1007/s00586-015-3931-0
[9] Hmida, J., Tomschi, F., Strauss, A. C., & Hilberg, T. (2023). Relationship between foot pressure and spinal parameters in healthy adults–A systematic review. Gait & Posture.
[10] Arslan, M., & Görgü, S. Ö. (2023). Effect of short-term spinal orthosis and insoles application on cobb angle, plantar pressure and balance in individuals with adolescent idiopathic scoliosis. Clinical Biomechanics, 110, 106121.
[11] Yağcı, G., & Yakut, Y. (2017). Effects of A Spinal Brace on The Functional Profile of The Feet in Adolescent Idiopathic Scoliosis. Acıbadem Üniversitesi Sağlık Bilimleri Dergisi, (4), 207-213.
[12] Altun, Y., Dilek, B., & Yakut, Y. (2021). Effects of brace on pedobarographic parameters in individuals with adolescent idiopathic scoliosis. Journal of Exercise Therapy and Rehabilitation, 8(1), 44-53.
[13] Gruić, I., Cebović, K., & Medved, V. (2016, November). Comparison of pedobarographic profile in young males with left and right scoliotic posture. In International Congress on Sport Sciences Research and Technology Support (Vol. 2, pp. 89-95). SCITEPRESS.
[14] Xu, L., Wu, Z., Xia, C., Tang, N., Cheng, J. C., Qiu, Y., & Zhu, Z. (2019). A genetic predictive model estimating the risk of developing adolescent idiopathic scoliosis. Current Genomics, 20(4), 246-251.
[15] Lenz, M., Oikonomidis, S., Harland, A., Fürnstahl, P., Farshad, M., Bredow, J., ... & Scheyerer, M. J. (2021). Scoliosis and Prognosis—a systematic review regarding patient-specific and radiological predictive factors for curve progression. European Spine Journal, 30, 1813-1822.
[16] Lv, Z., Lv, W., Wang, L., & Ou, J. (2023). Development and validation of machine learning-based models for prediction of adolescent idiopathic scoliosis: A retrospective study. Medicine, 102(14), e33441.

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

RESEARCH ARTICLE

A Comparative Analysis of Machine Learning Models for Time Prediction in Food Delivery Operations

Elmas Yalçinkaya[a], **Ouranıa Areta Hızıroğlu**[a†]

[a] Department of Management Information Systems, İzmir Bakırçay University, Izmir, Türkiye,
[†] ourania.areta@bakircay.edu.tr, corresponding author

RECEIVED MARCH 26, 2024
ACCEPTED APRIL 29, 2024

CITATION Yalçinkaya, E., & Areta Hızıroğlu, O. (2024). A comparative analysis of machine learning models for time prediction in food delivery operations. *Artificial Intelligence Theory and Applications, 4(*1), 43-56.

Abstract

Accurate time estimation is crucial for ensuring customer satisfaction and operational efficiency in the growing food delivery sector. This paper focuses on comprehensively analyzing factors affecting food delivery times and assessing the effectiveness of machine learning models in forecasting delivery times. For this purpose, authors incorporated a detailed dataset from a food delivery company on the Kaggle platform, encompassing delivery address, order time, delivery time, weather conditions, traffic intensity, and delivery person's profile information. The study evaluated the effectiveness and performance of various machine learning models such as Linear Regression, Decision Trees, Random Forests, XGBRegressor and the k-nearest neighbors (KNN) regression model using metrics like MAE, RMSE, and R^2. The results demonstrate that ensemble methods— XGBRegressor—outperformed the other models in accurately predicting delivery times, achieving an R-squared score of 0.82. Additionally, a thorough analysis of feature importance uncovered the factors influencing delivery time estimation. This study offers insights into leveraging machine learning techniques to optimize food delivery operations and enhance customer satisfaction. The discoveries can assist food delivery platforms in deploying effective time estimation models and emphasizing factors for predictions.

Keywords: machine learning; time estimation; feature importance; food delivery

1. Introduction

In the last few years, food delivery has experienced major expansion with the inception of the online platforms that connect customers, delivery drivers and restaurants. One of the difficulties that customers face is the uncertainty surrounding delivery times that prompts researchers to come up with predictive models that will allow estimations of delivery duration. These models tackle issues of planning, organization, and control to optimize operations of the delivery sector [1]. Moreover, there is a substantial body of research on speeding up the delivery process by using optimization methods that help in the more efficient routing of the vehicles [1] . Studies on delivery platforms require management decisions on issues such as the delivery times and subsidy administration to radically improve profits [2]. In this regard, cutting down on the delivery times is of vital importance for improving the level of customer satisfaction and for holding an edge among competitors. The industry of food delivery is in the process of ever-changing trends, wherein the most attractive features are being created first in both customer service improvement and in the field of delivery process optimization, as well as towards technology exploration to provide quick and prompt service.

There are numerous factors influencing food delivery times, that their impact has been studied, such as distance to the delivery address, restaurant preparation time, traffic conditions, weather conditions, and the experience of the delivery person [3]. Within this framework, the first aim of this study is to explore the key factors influencing food delivery times. These could include variables, such as the distance to the delivery address, restaurant preparation time, traffic conditions, weather conditions, and the experience of the delivery person. The outcome seeks to aid in finding effective solutions for optimizing delivery times, which are crucial for improving operational strategies and customer service.

Machine Learning (ML) is an effective tool for providing ways out of this predicament. Within the framework of the enhancement of food delivery time, academic scholars have explored the application of ML tools. More specifically, ML algorithms can optimize delivery times by learning sophisticated patterns, making predictions based on vast quantities of data. Hence, ML models have been used to detect travel times and incorporate predictors into optimization models to resolve the last-mile delivery problem [4]. This article explores how ML can be used to analyze the elements that determine delivery times and forecast them. This work aims to identify the most effective ML approaches for predicting food delivery times and to make this information usable for improving food delivery services. This will increase customer satisfaction and improve the operational efficiency of businesses. This initiative aims to provide valuable insights to companies in the food delivery industry, enabling them to offer faster and more reliable services.

In order to achieve its objectives, this study will examine the aforementioned factors and analyze how a range of ML models can accurately predict delivery times of food. The models used include Linear Regression, Decision Trees, Random Forests, KNN and XGBoost. The performance of each model will be evaluated using various metrics to determine the most effective one.

The following sections include a literature review of the studies related with ML applied in food sector, with emphasis on the delivery element; presentation of the undertaken methodology followed by the results; and concludes with the discussion and conclusions parts.

2. Literature review

This study aims to understand factors affecting food delivery times and to use these ML models for their prediction. This literature review section presents studies that have dealt with ML applications in the food sector, and particularly the food distribution aspect.

2.1. Relevant Studies

The research encompasses steps such as examining the integration of technological innovations like artificial intelligence and Machine Learning in the food industry, with a focus on the competitive food distribution sector [5]. The necessity for businesses to optimize their processes due to customers' preference for online platforms is highlighted [6]. Various studies have explored ML techniques in different contexts. For instance, an integrated approach combining mechanical modeling and ML was proposed for optimizing the thickness of frozen microwaveable foods, resulting in better heating homogeneity [7]. Liu et al. [7] discussed a framework integrating travel time predictions with order assignment optimization, emphasizing ML applications in food service delivery operations. Yang et al. [8] underscored the potential of ML in efficiently addressing challenges like delivery route decisions, food item demand forecasting, and logistic planning. They proposed a hybrid evolutionary optimization highlighting the performance advantages of ML-based algorithms for food delivery applications, focusing on package service, order selection and delivery route planning. Madani and Alshraideh [9] conducted a study on the vehicle routing problem in food order distributions, assessing the applicability of mathematical modeling and optimization techniques. The research emphasizes the potential of artificial intelligence in routing and timing to enhance the efficiency of logistics and distribution processes. They investigated the use of artificial intelligence and ML techniques to predict consumers' online purchasing decisions, providing significant insights for understanding consumer behavior and personalizing services. In their study, Maluud and Abdulazeez [10] build upon existing research highlighting the environmental, cost, and energy advantages of using e-scooters in postal and package delivery, evaluating the impact of these vehicles on delivery time and energy costs through the application of various ML algorithms. Such information can be utilized to comprehend customer preferences and behaviors in optimizing food delivery processes. The literature review process showed several approaches that were undertaken by scholars, and furthermore, a variety of factors that were explored as shown in the following Table 1.

Table 1. Studies on ML Applications in Food Delivery

Author(s)	**Approach/Model**	**Factors Explored**
Liu et al. [7]	Mechanical modeling and ML	Travel time predictions, order assignment
Zhang et al. [8]	Hybrid evolutionary optimization and ML algorithms for order selection and route planning	Order selection, delivery route planning
Madani and Alshraideh [9]	Mathematical modeling, optimization techniques, and artificial intelligence for routing and timing	Vehicle routing, delivery timing
Maluud and Abdulazeez [10]	ML algorithms for evaluating the impact of e-scooters on delivery time and energy costs, and understanding customer preferences	Vehicle mode (e-scooters), delivery time, energy costs, customer preferences
Moghe et al. [11]	Novel system based on multiple ML algorithms for enhanced delivery time estimates for batched orders	Order batching, delivery time estimates
Hildebrandt and Ulmer [12]	Offline and online-offline estimation approaches using supervised learning to improve meal arrival time estimations in restaurant meal delivery services.	Arrival time estimates, customer selections
Zhu et al. [13]	Utilization of a deep neural network (DNN) incorporating various features to enhance prediction efficacy	Order fulfillment cycle time
Gao et al. [14]	Application of deep learning using a deep network named FDNET, prediction of feasible locations, consideration of factors affecting driver behaviors, and introduction of spatiotemporal information.	Delivery route generation, time prediction, driver behaviors, spatiotemporal information
Liu et al. [15]	Integration of travel-time predictors with order-assignment optimization, reformulations of integrated models for efficient solving, and two simple heuristics for the multiperiod order-assignment problem.	Driver routing behavior, order assignment, travel time predictors
Gao et al. [16]	Applying Deep Learning Based Probabilistic Forecasting to Food Preparation Time for On-Demand Delivery Service	Food preparation time
Hughes et al. [17]	Evaluation of ML methodologies to predict stop delivery times from GPS data	Stop delivery times, duration prediction

Many studies primarily examine operational factors such as travel times, order assignments, and driver behavior, with limited consideration of external factors. Nevertheless, the analysis does not delve deeply into external factors like weather conditions, traffic patterns, and events that may affect delivery times. Moreover, certain studies acknowledge the drawback of not considering real-time factors, indicating the necessity of integrating real-time data (such as traffic updates and weather conditions) into predictive models to enhance the accuracy and adaptability of delivery time estimation. There is also a potential gap in the current research when it comes to incorporating customer preferences and behavior into predictive models. While some studies have touched on this topic, there is room for more extensive exploration of customer-related factors such as order preferences, historical data, and feedback. By incorporating these factors, we can enhance delivery efficiency and ultimately improve customer satisfaction.

This study focuses on optimizing food delivery times through ML approaches, while specifically exploring new factors and methods not previously addressed in the literature. It aims to extend beyond the examination of previously defined factors like traffic intensity, weather conditions, and the delivery person's experience, to delve into less explored dimensions such as the relationship between customer satisfaction and waiting time. Particularly, this research intends to offer strategic improvements in delivery processes by conducting an in-depth analysis of various data features affecting delivery times and comparing different ML methods. This approach expands the existing literature by testing the applicability of advanced algorithms and providing innovative solutions to the complexities and uncertainties of delivery processes. The contributions of this study have the potential to enhance the efficiency and customer satisfaction of food delivery operations, offering directly applicable insights for industry practice.

3. Methodology

This study follows a comprehensive methodology to identify factors affecting food delivery times and predict these times using Machine Learning models. Our methodology includes data collection, preprocessing, modeling, and evaluation, followed by the analysis of the results that Part 4 will present.

3.1. Data Collection

The dataset used in this research is from a food delivery company on the Kaggle platform [18]. It includes various features like delivery address, order time, delivery time, weather conditions, traffic intensity, and delivery person's profile information.

Food delivery adapts rapidly to today's dynamic lifestyle as a courier service. Offered by restaurants, stores, and specialized food delivery companies, customers usually place orders online via a restaurant's website, mobile app, or food ordering services. The delivery process involves various factors from start to end, directly affecting efficiency.

Delivered products may include main dishes, appetizers, beverages, desserts, and grocery items. They need to be transported safely and intact, often in boxes, bags, or thermal carriers. The delivery person's vehicle choice significantly impacts delivery speed and efficiency. Vehicle selection varies with geographic conditions and city structure; agile transport like bicycles or motor scooters in large cities, while cars are more common in wider, open areas.

3.1.1. Dataset Content

The dataset used in this study encompasses extensive data related to food delivery processes. Variables like the delivery person's identity, age, rating scores, coordinates of the restaurant and delivery point, order and delivery times, weather conditions, traffic intensity, condition of the delivery vehicle, and delivery type are analyzed to understand

and predict delivery times. These data are collected to thoroughly examine factors influencing food delivery times and predict them using Machine Learning models.

Table 2. Dataset Content

Variable Name	Data Type	Description
ID	object	Unique identifier for each delivery record.
Delivery_person_ID	object	Identifier for the delivery personnel involved in the delivery.
Delivery_person_Age	object	Age of the delivery personnel, typically a numerical value but listed as an object due to possible non-numeric entries
Delivery_person_Ratings	object	Ratings given to the delivery personnel, typically on a scale, but listed as an object due to possible non-numeric entries.
Restaurant_latitude	float64	Geographical latitude of the restaurant from where the order is dispatched.
Restaurant_longitude	float64	Geographical longitude of the restaurant from where the order is dispatched.
Delivery_location_latitude	float64	Geographical latitude of the delivery location
Delivery_location_longitude	float64	Geographical longitude of the delivery location.
Order_Date	object	The date on which the order was placed.
Time_Orderd	object	The time at which the order was placed.
Time_Order_picked	object	The time at which the order was picked up by the delivery personnel.
Weatherconditions	object	Descriptive information about the weather conditions during delivery.
Road_traffic_density	object	Information about the density of road traffic during delivery.
Vehicle_condition	int64	A numerical rating or categorization of the vehicle's condition used for delivery. This could represent various states of vehicle functionality and may impact delivery efficiency.
Type_of_order	object	Type or category of the order.
Type_of_vehicle	object	Type of vehicle used for the delivery.
multiple_deliveries	object	Indicator of whether the delivery person is handling multiple deliveries simultaneously.
Festival	object	Indicator of whether the delivery occurred during a festival.
City	object	The city in which the delivery took place.
Time_taken(min)	object	The time taken for the delivery, typically a numerical value but listed as an object due to possible non-numeric entries.

3.2. Data Preprocessing

The dataset underwent a data preprocessing process that included filling missing values, converting categorical data into numerical format, and cleaning outlier values. It was transformed into a DataFrame and consists of three CSV files: Sample_Submission, train, and test. Our studies were completed on the train dataset, which has 45,593 rows ranging from 0 to 45,592, indicating its comprehensiveness. There are a total of 20 columns observed, with names and data types including ID, Delivery_person_ID, Age, Ratings, Order_Date, and Time_taken(min), among others. Most columns were of the object type, typically used for categorical and textual information. However, columns like "Age", "Ratings", and "Time_taken(min)" were converted to the correct data types as they contain numerical data. Missing values were observed in some columns, and their correct imputation was crucial as it significantly impacts the analysis results.

3.3. Feature Selection and Engineering

In this study, the feature engineering process was comprehensively addressed. Initially, categorical data such as weather conditions and road traffic density were converted into a numerical format suitable for Machine Learning algorithms through Label Encoding. Subsequently, time-related features that could potentially impact delivery times were extracted from order dates, providing new information like day, month, and year. Additionally, delivery distances were calculated using the coordinates of restaurants and delivery points.

For feature selection, we applied the XGBRegressor model's built-in feature importance calculation method. This method evaluates the importance of each feature based on how much it contributes to the prediction accuracy of the model. The feature importances are calculated using the mean decrease in impurity (MDI) criterion, which measures the decrease in node impurity (weighted impurity decrease) for each feature when it is used for splitting in the decision trees that make up the XGBRegressor model.

All these features played a crucial role in enhancing the model's predictive capability and accuracy in estimating food delivery times. The process of identifying factors critical to the predictive accuracy of ML algorithms is a key step in feature engineering [3]. The detailed and comprehensive approach to feature engineering significantly improved the accuracy of the results and the overall performance of the model.

3.4. Model Development

Various Machine Learning models were developed and compared to predict delivery times. Machine Learning is the area of study that deals with the development of the algorithms and the statistical models that help computers to learn, and make inferences, as well as decisions, based on data without them being explicitly programmed. It is the process of using sophisticated mathematical models to process and interpret data, detect relationships, and make decisions or predictions based on the results. ML models have shown various remarkable results in learning highly intricate patterns from the data and then making future predictions for unseen data [19].

In this study, four different ML models were employed to predict the food delivery times. Each model has distinct characteristics and offers solutions for different types of data structures and complexities. These models include Linear Regression, Decision Trees, Random Forests, XGBoost, and KNN Regression:

1. Linear Regression: A basic and widely used statistical estimation method to model the relationship between independent variables and a dependent variable. It takes a weighted sum of the independent variables to predict a continuous output. This method deals with linear relationships between variables and provides an effective starting point for predicting food delivery times [10].
2. Decision Trees: These work by splitting the dataset into segments using simple decision rules, allowing them to model complex data structures. Decision trees are

non-parametric supervised learning algorithms [20]. Each decision node splits the data set into two or more homogeneous subsets, leading to predictions at the end.

3. Random Forests: Developed by Breiman [21] they combine multiple decision trees to create a 'forest' using ensemble learning techniques. Each tree is built from a random subset of the dataset, and the final prediction is made by averaging the predictions of all trees or by a majority vote. This approach reduces the risk of overfitting and usually produces more reliable and robust predictions than a single decision tree.
4. XGBoost (eXtreme Gradient Boosting): A high-performance gradient boosting library offering powerful algorithms for making predictions on complex datasets. Built on decision trees, XGBoost sequentially constructs trees, each learning from the errors of its predecessors. This model is favored for scalability, speed, and performance and often achieves high success in ML competitions and industrial applications [22].
5. K-Nearest Neighbors (KNN) Regression: A non-parametric algorithm that predicts the value of a data point based on the values of its nearest neighbors in the feature space. KNN regression finds the K closest training examples to the input and averages their output values [23].

Each model was trained using cross-validation and hyperparameters were tuned with GridSearchCV. A critical step in our model optimization process is identifying the most suitable parameters for different machine learning algorithms using GridSearchCV. GridSearchCV evaluates every combination within specified parameter ranges to select those that maximize the model's accuracy score. This procedure is applied to each model, aiming to enhance both its accuracy and generalization capacity. The selection of parameters such as depth (max_depth) for decision trees affects the complexity of the trees and thus the learning capacity of the model, while parameters like the number of trees (n_estimators) used in ensemble algorithms aim to increase the stability and reliability of the predictions. These parameters have been carefully adjusted to enhance our models performance [24].

3.5. Model Evaluation

In this phase, three primary statistical metrics were used to evaluate the prediction performance of the machine learning models regarding food delivery times, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics measure how accurate the model's predictions are and how well they fit the actual data.

1. Mean Absolute Error (MAE):

Mean Absolute Error (MAE) is a measure of how much the predictions of a model deviate from the actual values. For each prediction, the absolute difference from the real value is calculated, and the average of these differences is computed. Mathematically, it can be expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

Here, y_i represents the actual values, and $\hat{y}_i$represents the model's predictions.n is the number of observations. This metric is crucial for understanding the accuracy of the model in predicting the outcomes [25].

2. Root Mean Squared Error (RMSE):

Root Mean Squared Error (RMSE) is a metric used to measure the magnitude of errors in a model's predictions. It calculates the square root of the average of the squares of the prediction errors, giving more weight to larger errors. RMSE offers a comprehensive assessment of model performance. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

In this formula, y_i represents the actual values $\hat{y}_i$ represents the predicted values, and n is the number of observations. RMSE is particularly useful in quantifying the error in terms of the units of the observed data [26].

3. R-squared (R²):

R-squared (R²) is a statistical measure that represents the proportion of variance for the dependent variable that's explained by the independent variables in a regression model. It ranges between 0 and 1 and is used as a measure of how close the model's predictions are to the actual values [27]. A higher R^2 value indicates that the model captures the data well. The formula for R^2 is:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

Here y_i presents the actual values, $\hat{y}_i$ the model's predictions, and $\bar{y}_i$ the average of the actual values. These metrics have been used to evaluate the model's success in predicting food delivery times. Each metric addresses different aspects of the model and, when they are used together, they provide a balanced and comprehensive assessment. It is recommended to use these metrics in conjunction to gain a thorough understanding of the model's predictive capabilities and accuracy [28]. The success of ML algorithms is directly related to how low the values of the performance metrics used are [29].

4. Results

This analysis has identified the features that are most decisive in predicting food delivery times. It provides critical information for developing strategies to more effectively forecast delivery times and enhance delivery processes. These findings offer deeper insight into the factors significantly affecting delivery times, laying the groundwork for future research by providing valuable insights. The analysis of the collected training data thoroughly evaluated how well the models could adapt to different situations and accurately predict outcomes. Parameter tuning, done through GridSearchCV, tested how well the models performed with different settings to identify the most effective ones [20]. Using cross-validation scores, GridSearchCV helped choose the best model setup, preventing overfitting and improving performance. We assessed model performance by examining the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2). These measures assessed how close the models' predictions were to the actual values and their effectiveness in explaining differences in delivery times. Assessing performance was vital for understanding prediction errors and how well the models could handle variations in the dataset. The achieved results are presented below.

a. Calculation of Feature Importances:

After training, the feature importances were determined. These scores reflect the weight each feature has in the model's decision-making process. To obtain a better comprehension of the importance levels, these numbers were visualized in a bar graph form (see Figure 1). Every bar on the graph illustrates a feature, and the height of the bar illustrates the importance of that particular feature. This visualization allows us to instantly identify the most critical features of the model easily, and thus to understand the underlying information on which the model is built [21]. Visualization is a tool that helps us to better understand the intricate decision-making process within Machine Learning models. This can be vividly depicted by the visualization of different features and their effect on the model prediction as illustrated by Breiman [21] during the analysis of 5 million orders at Ele.me where it becomes easier for us to grasp these complicated processes. Moreover, this method not only broadens our understanding of the modeling process but also succinctly and clearly presents how the data affects the model predictions.

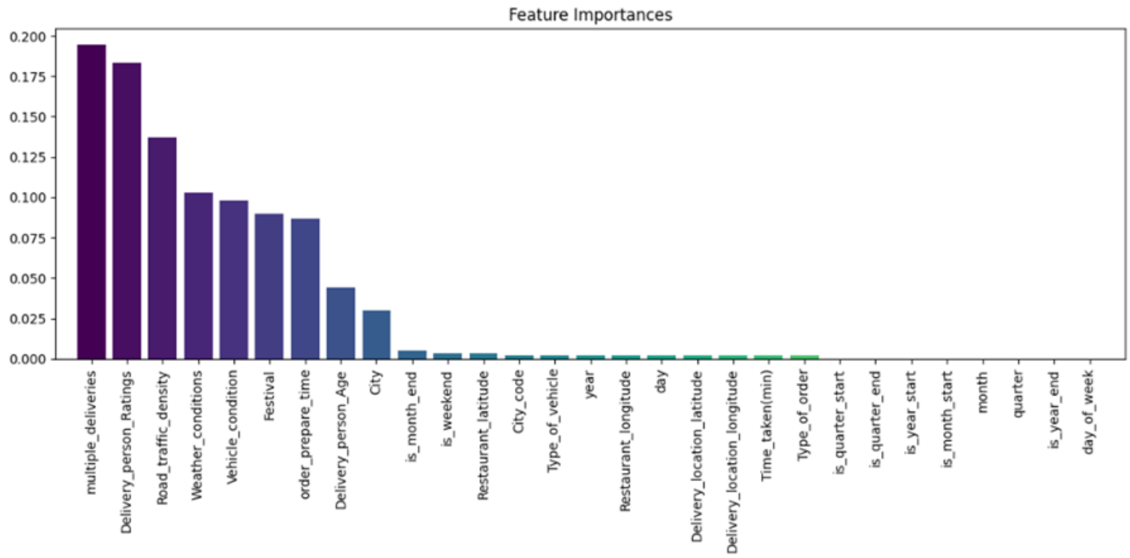

Figure 1. Feature Importance

b. Model Comparison Results:

GridSearchCV was used to find the optimal parameters for each model to enhance performance on the dataset. More specifıbcally, the optimal parameter values chosen were max_depth of 7 for DecisionTreeRegressor, n_estimators of 300 for RandomForestRegressor, and n_estimators of 20 with a max_depth of 9 for XGBRegressor. The best number of neighbors (n_neighbors) for KNeighborsRegressor was set as 7. These parameters control the complexity and learning capacity of the model. The models were trained on the training dataset (X_train, y_train), allowing them to learn patterns and relationships in the dataset.

Once training was complete, predictions were made on the test dataset (y_pred = model.predict(X_test)), and each model's performance was evaluated by comparing it to the actual delivery times (y_test). Table 3 presents the outcomes of this evaluation based on the three primary statistical metrics that were defined in the methodology part, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2).

Table 3. Model Comparison Results

ML Model	MAE	RMSE	R^2
Linear Regression	5.74	7.15	0.417
Decision Trees	4.24	5.58	0.7172
Random Forests	3.21	4.09	0.8126
XGBRegressor	3.16	3.98	0.82
KNN	5.53	7.05	0.43

Linear regression seems to have relatively high errors compared to other models, indicating that it might not capture the underlying patterns in the data well. Decision trees show better performance compared to linear regression, with lower errors and a higher R^2 value, indicating better fitting to the data. Random forests further improve the performance over decision trees, with even lower errors and a higher R^2 value, suggesting a better fit to the data and improved predictive power. Linear Regression and KNN performed relatively poorly compared to other models, with higher errors and a lower R^2 value, suggesting that they may not be the best choice for this dataset. These metrics demonstrate the superiority of the XGBRegressor in predicting food delivery times, having the best performance in all three metrics:

- Mean Absolute Error (MAE): Indicating the average deviation of the model's predictions from the actual values, with an MAE of 3.16, suggesting the model's predictions deviate from the actual values by an average of about 3.16 units.
- Root Mean Squared Error (RMSE): The square root of the MSE, providing the typical magnitude of errors, was calculated to be 3.98.
- R-squared (R^2) Score: Indicating how much of the variance in delivery times the model can explain, an R^2 of 0.82 suggests the model can explain 82% of the variance in the dataset.

Moreover, Figure 2 presents the performance comparison of the different machine learning models based on R-squared scores. The best parameters for each model and the highest R-squared scores achieved are indicated. The high R^2 score for the XGBRegressor indicates it can capture a large part of the variance in the dataset and effectively model the factors affecting delivery times.

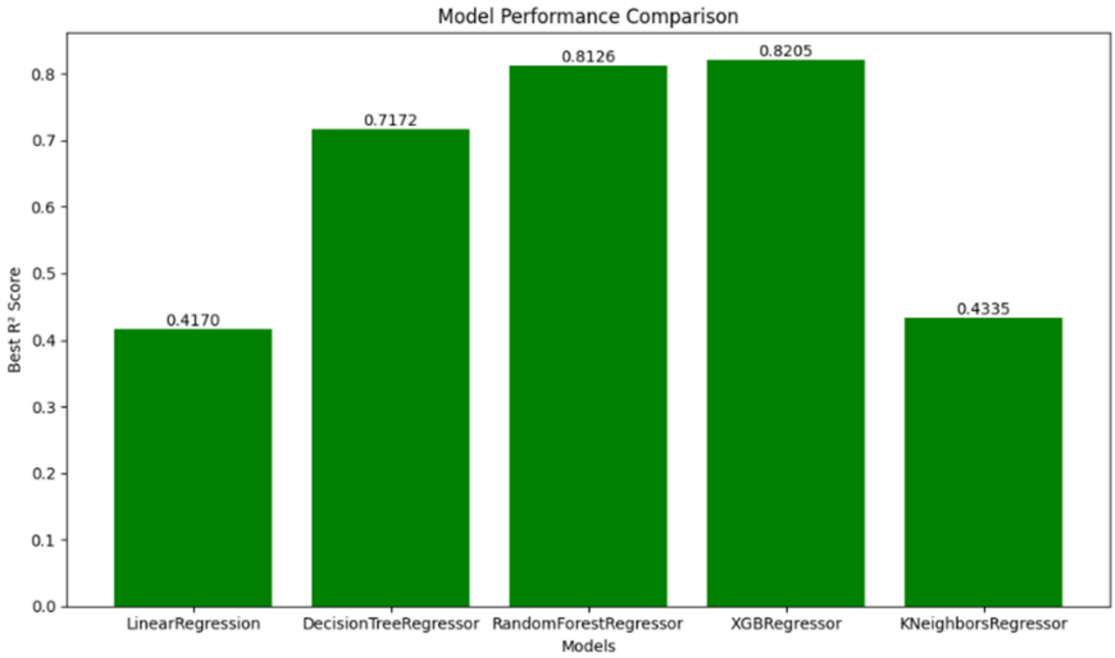

Figure 2. Model Performance Comparison

5. Discussion

This study showed the efficiency of machine learning models, particularly XGBRegressor, in modelling correctly the delivery times of food. The investigation indicates that as a group, ensemble methods are superior to the other models, e.g. Linear Regression, Decision Trees, and Random Forests. The result comes in line with existing literature suggesting that ensemble methods overall outperform in complex prediction assignments owing to their capacity to pick up most intricate patterns and relationships within the data set.

The XGBRegressor model achieved an impressive R-squared score of 0.82, indicating that it can explain 82% of the variance in delivery times. This high explanatory power is a testament to the model's capability in capturing the nuances of the dataset and the factors influencing delivery times. The low values of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) further reinforce the model's accuracy and reliability.

Compared to current state-of-the-art methods, the XGBRegressor model outperformed techniques used in similar studies. For instance, Liu et al. [15] achieved an R-squared of 0.76 for travel time prediction, while the present study attained a higher R-squared of 0.82 for delivery time estimation. Additionally, the proposed approach incorporates a comprehensive set of features, including weather conditions, traffic intensity, and delivery person's profile, which were not considered in some previous works [12, 13].

Feature importances analysis yields very useful information on the factors that are highly correlated to delivery times. This visualization provides for the user an immediate understanding of the most influential features and stakeholders can subsequently focus on optimizing those critical factors. To achieve operational efficiency and customer satisfaction goals, food delivery providers can prioritize the top-ranked features and implement targeted strategies.

It is necessary to keep in mind that the XGBRegressor model gave excellent results but there would be fluctuations in the results depending on the specific dataset used and also the region or operational context in which the model is applied. Consequently, it is suggested to test the model's performance on various datasets and consider the specific features of the target delivery setting.

6. Conclusion

This study aimed to identify the most effective machine learning models for predicting food delivery times and to understand the key factors influencing these times. The results demonstrate that the XGBRegressor model from the XGBoost library outperformed other models, achieving an R-squared score of 0.82 and low values of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). This highlights the model's ability to accurately predict delivery times and capture the complex relationships within the dataset.

As a result, featured importance analysis detected the main factors affecting delivery times. This data is board to a food delivery company to highlight and optimize its operations, thus leading to better customer satisfaction and efficiency of operation.

This study contributes to the already existing body of knowledge on machine learning models for forecasting food delivery times by offering a complete analysis on the critical factors that affect such times. The research outcomes can be exploited by delivery-food companies to boost their businesses and establish a competitive edge in the ever-growing food delivery market. While this study has provided significant insights into the application of machine learning for predicting food delivery times, there are several avenues for future research.

Particularly, the research of the current study was concentrated on a particular group of factors like temperature, traffic, and courier background. Research studies could seek to develop more features which will eventually improve the prediction accuracy of the models. For example, the use of live traffic data, road construction details, or consumer feedback can help improve the services.

Moreover, this research work mainly covered traditional machine learning models. Nevertheless, future analysis can develop deep learning methods like CNNs or RNNs to predict food delivery times. Deep learning models have demonstrated remarkable progress in several domains and could provide additional inputs or higher performance.

Although this research has focused on predicting the delivery times, a future research study can investigate the integration of these predictions with optimization techniques for route planning, resource allocation and operational decision-making. Conjunction of machine learning models with optimization algorithms may eventually result in more efficient and cost-effective food delivery operations.

Finally, within the framework of reinforcing the results and evaluating the implications for the practice, the forthcoming research should include the deployment of the developed models in real-life food delivery operations. Such evaluation will facilitate the assessment of, and refinement of models as based on real life feedback and constraints. Through this future research approach, food delivery optimization can be taken to the next level, resulting in satisfied customers, efficient business operations, and a sustainable food delivery industry.

References

[1] Zhao, Y., & Bacao, F. (2020). What factors determining customer continuingly using food delivery apps during 2019 novel coronavirus pandemic period? International Journal of Hospitality Management, 91, 102683. https://doi.org/10.1016/j.ijhm.2020.102683

[2] Zhao, B., Tan, H., Zhou, C., & Feng, H. (2023). Optimal delivery time and subsidy for IT-enabled food delivery platforms considering negative externality and social welfare. Industrial Management & Data Systems, 123(5), 1336–1358. https://doi.org/10.1108/IMDS-09-2022-0554

[3] Saad, A. T. (2020). Factors affecting online food delivery service in Bangladesh: an empirical study. British Food Journal, 123(2), 535–550. https://doi.org/10.1108/BFJ-05-2020-0449

[4] Chu, H., Zhang, W., Bai, P., & Chen, Y. (2023). Data-driven optimization for last-mile delivery. Complex Intell. Syst., 9(3), 2271–2284. https://doi.org/10.1007/s40747-021-00293-1

[5] Zhuang, X., Lin, L., Zhang, R., Li, J. (Justin), & He, B. (2021). E-service quality perceptions of millennials and non-millennials on O2O delivery applications. British Food Journal, 123(12), 4116–4134. https://doi.org/10.1108/BFJ-01-2021-0049

[6] Yaiprasert, C., & Hidayanto, A. N. (2023). AI-driven ensemble three machine learning to enhance digital marketing strategies in the food delivery business. Intelligent Systems with Applications, 18, 200235. https://doi.org/10.1016/j.iswa.2023.200235

[7] Liu, S., He, L., & Shen, Z.-J. M. (2021). On-Time Last-Mile Delivery: Order Assignment with Travel-Time Predictors. Management Science, 67(7), 4095–4119. https://doi.org/10.1287/mnsc.2020.3741

[8] Zhang, M.-X., Wu, J.-Y., Wu, X., & Zheng, Y.-J. (2022). Hybrid evolutionary optimization for takeaway order selection and delivery path planning utilizing habit data. Complex Intell. Syst., 8(6), 4425–4440. https://doi.org/10.1007/s40747-021-00410-0

[9] Madani, B., & Alshraideh, H. (2021). Predicting Consumer Purchasing Decisions in the Online Food Delivery Industry. In Advances in Machine Learning (pp. 103–117). Academy and Industry Research Collaboration Center (AIRCC). https://doi.org/10.5121/csit.2021.111510

[10] Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. Journal of Applied Science and Technology Trends, 1(4), Article 4. https://doi.org/10.38094/jastt1457

[11] Moghe, R. P., Rathee, S., Nayak, B., & Adusumilli, K. M. (2021, January 2). Machine Learning based Batching Prediction System for Food Delivery. Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD). https://doi.org/10.1145/3430984.3430999

[12] Hildebrandt, F. D., & Ulmer, M. W. (2022, July). Supervised Learning for Arrival Time Estimations in Restaurant Meal Delivery. Transportation Science, 56(4), 1058–1084. https://doi.org/10.1287/trsc.2021.1095

[13] Zhu, L., Yu, W., Zhou, K., Wang, X., Feng, W., Wang, P., Chen, N., & Lee, P. (2020, August 20). Order Fulfillment Cycle Time Estimation for On-Demand Food Delivery. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. https://doi.org/10.1145/3394486.3403307

[14] Gao, C., Zhang, F., Wu, G., Hu, Q., Ru, Q., Hao, J., He, R., & Sun, Z. (2021, August 14). A Deep Learning Method for Route and Time Prediction in Food Delivery Service. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. https://doi.org/10.1145/3447548.3467068

[15] Liu, S., He, L., & Max Shen, Z. J. (2021, July). On-Time Last-Mile Delivery: Order Assignment with Travel-Time Predictors. Management Science, 67(7), 4095–4119. https://doi.org/10.1287/mnsc.2020.3741

[16] Gao, C., Zhang, F., Zhou, Y., Feng, R., Ru, Q., Bian, K., He, R., & Sun, Z. (2022, August 14). Applying Deep Learning Based Probabilistic Forecasting to Food Preparation Time for On-Demand Delivery Service. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/3534678.3539035

[17] Hughes, S., Moreno, S., Yushimito, W. F., & Huerta-Cánepa, G. (2019, December). Evaluation of machine learning methodologies to predict stop delivery times from GPS data. Transportation Research Part C: Emerging Technologies, 109, 289–304. https://doi.org/10.1016/j.trc.2019.10.018

[18] Food Delivery Dataset. (2022, October 1). Kaggle. https://www.kaggle.com/datasets/gauravmalik26/food-delivery-dataset/code

[19] Murdoch, W. J., Singh, C. D., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences, 116(44), 22071-22080. https://doi.org/10.1073/pnas.1900654116

[20] Breiman, L. (2017). Classification and Regression Trees. Routledge. https://doi.org/10.1201/9781315139470

[21] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[22] Yang, N., Korfiatis, N., Zissis, D., & Spanaki, K. (2023). Incorporating topic membership in review rating prediction from unstructured data: a gradient boosting approach. Annals of Operations Research. https://doi.org/10.1007/s10479-023-05336-z

[23] Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons

[24] Pedregosa et al., (2012). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[25] Lin, W., Liu, L., Zhao, G., & Jian, Z. (2023). Developing Hybrid DMO-XGBoost and DMO-RF Models for Estimating the Elastic Modulus of Rock. Mathematics, 11(18), Article 3886. https://doi.org/10.3390/math11183886

[26] Wang, X., Li, W., & Li, Q. (2021). A New Embedded Estimation Model for Soil Temperature Prediction. Scientific Programming, 2021, Article 5881018. https://doi.org/10.1155/2021/5881018

[27] Shahani, N. M., Kamran, M., Zheng, X., & Liu, C. (2022). Predictive modeling of drilling rate index using machine learning approaches: LSTM, simple RNN, and RFA. Petroleum Science and Technology, 40(5), 534–555. https://doi.org/10.1080/10916466.2021.2003386

[28] İnaç, H., Ayözen, Y. E., Atalan, A., & Dönmez, C. Ç. (2022). Estimation of Postal Service Delivery Time and Energy Cost with E-Scooter by Machine Learning Algorithms. Applied Sciences, 12(23), Article 12266. https://doi.org/10.3390/app122312266

[29] Ceylan, Z. (2021). The impact of COVID-19 on the electricity demand: a case study for Turkey. International Journal of Energy Research, 45(9), 13022–13039. https://doi.org/10.1002/er.6631

Acknowledgement

This study has utilized AI Generative tools for translation and text optimization purposes (grammar and syntax).

■ RESEARCH ARTICLE

Happy Student in the Age of Artificial Intelligence

Mustafa Fidan [a†]

[a] Education Academy, Ankara, Türkiye
† gfidanmustafa@gmail.com, corresponding author

RECEIVED MARCH 30, 2024
ACCEPTED APRIL 29, 2024

CITATION Fidan, M. (2024). Happy student in the age of artificial intelligence. *Artificial Intelligence Theory and Applications, 4(*1), 57-75.

Abstract

This paper discusses the potential benefits of integrating artificial intelligence (AI) into education and how it can affect students' well-being and happiness. The study used a phenomenological design and involved 46 teachers who observed their students' experiences with AI and happiness. The findings were organised under five main themes. First, AI provides personalised learning opportunities that empower students. Second, AI enables teachers to focus more on teaching by reducing administrative tasks. Third, AI increases students' engagement and motivation, leading to more effective learning experiences. However, the study also highlighted ethical concerns that need to be addressed to prioritise students' welfare. Finally, the integration of AI into education prepares students for an AI-driven future and positively influences their learning experiences. These findings can benefit students, teachers and administrators by providing valuable insights into how AI can be utilised in education.

Keywords: artificial intelligence; student happiness; educational management; smart board; teacher empowerment

1. Introduction

Artificial intelligence (AI) is a powerful phenomenon of recent times that is widely and rapidly incorporated into every aspect of life. Many examples illustrate how AI is impacting various aspects of human life, such as access to information over the Web, consumption of news and entertainment, surveillance systems that identify individuals, financial market performance, citizens receiving welfare payments, and how drivers and pedestrians navigate [1]. Artificial Intelligence (AI) refers to the ability of machines to perform tasks that often require human intelligence. These tasks include gaming, language translation, expert systems, and robotics. While the idea of machines imitating intelligence dates back centuries, true machine intelligence emerged with digital computers in the 1940s, and as computer processing power improved, artificial intelligence evolved from simpler tasks (such as playing chess) to more complex tasks such as visual pattern recognition and natural language understanding [2]. With its use in every field in recent years, artificial intelligence can be defined as a supercomputer model or an ultra-advanced computer with human-like functions [3].

Intelligence has been defined as the computational part of the ability to achieve goals in the world. Artificial intelligence can be defined as the science and engineering of making

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

intelligent machines, especially intelligent computer programmes. This definition draws attention to the fact that intelligence, whether in humans, animals or machines, is goal-oriented and problem-solving [4,5]. Although there is no common definition, it can be said that each definition brings explanations to the field of artificial intelligence and has an equivalent. When all definitions are analysed, as a general definition, artificial intelligence is expressed as adapting to insufficient information and resources [6]. Artificial intelligence applications come to the fore as the harmony between scarce resources and information that cannot be obtained for numerous reasons and information sources develops.

In recent years, it can be said that there is almost no area where artificial intelligence has not come to the fore. The use of artificial intelligence to increase and accelerate scientific discoveries is one of them. Artificial intelligence helps scientists to generate hypotheses, design experiments, collect and interpret large data sets. Among the breakthroughs of artificial intelligence in the last decade are self-supervised learning, geometric deep learning and generative artificial intelligence methods. Artificial intelligence, which helps scientists in scientific processes with these and similar methods, gives a lot of hope to those interested in what they can do [7].

In the field of education, artificial intelligence transforms the way students learn, interact and relate to information. Studies examining the relationship between human communication and AI emphasise the transforming role of AI in communication processes. Topics such as human-computer interaction and computer-mediated communication, the role of AI as a communicator or mediator, and the tension between human agency and machine agency are addressed. In all these topics, although there are positive and negative effects of artificial intelligence on human communication, there is an optimistic attitude towards the exploration of new concepts, frameworks and topics [8].

The field of AI-Education has evolved over the years and first focussed on intelligent systems replacing traditional classroom teaching. However, in the late 1970s, there was a shift towards the tutoring model where AI-CAI systems aimed to provide individualised tutoring and solve educational problems rather than replace teachers. This led to the emergence of Intelligent Tutoring Systems (ITS). The advent of web-based education has further changed the landscape by offering online learning materials that cater to the needs of students without a teacher. This model provided personalised learning experiences and greater flexibility. AI-supported education benefits students and teachers by providing personalised, flexible and engaging learning opportunities. AI tools enhance the learning process by enabling students to be more independent and improve their learning experience. Moreover, web-based and online educational materials support students' activities such as downloading, studying and completing assignments. Artificial intelligence plays a very important role in education, even in early childhood education [3,9].

Artificial intelligence (AI) has revolutionised the education sector by automating administrative tasks and enhancing the learning experience. By leveraging AI tools, tasks such as scheduling, grading, and attendance tracking can be efficiently automated, giving teachers more time to focus on teaching and providing support to students. Education Management Information Systems (EMIS), powered by AI algorithms, enable education leaders to store, analyse and disseminate data, facilitating data-driven decision-making in various aspects of the education system. A well-designed EMIS can improve efficiency, aid planning and policy-making, and monitor educational outcomes through reliable data analysis and reporting. AI also provides opportunities for personalised learning as it can diagnose student needs, make recommendations and identify students at risk for timely intervention and support. The growing interest in AI in education is leading to further research and evaluation of its pedagogical potential. In general, AI has become an integral component of education, transforming traditional processes and improving educational outcomes [10-12].

It can be said that artificial intelligence automates learning. Modern students are more connected to organizations that want to educate them. In this new era, networks and software can transfer knowledge from educational institutions to learners and continuously improve experiences in real time. Artificial intelligence is attracting great attention in a wide range of areas today. Digital learning, social networks and software-based learning tools have become ubiquitous, leading to increased automation and artificial intelligence transforming every field, including education [10].

Using engaging educational approaches such as game-based learning and mobile learning can improve students' motivation and outcomes in the learning process. Implementing an AI-powered recommendation system can have positive results on students' engagement, motivation, and learning outcomes. Additionally, Artificial Intelligence (AI) has the potential to address fundamental challenges in education, transform teaching and learning practices, and contribute to global education goals. AI technologies can increase student engagement and provide personalized learning experiences by offering instant feedback and enabling self-correction. AI can also support collaborative learning environments that simulate real-world scenarios. Intelligent teaching systems and virtual learning assistants powered by AI encourage active participation, personalized learning, instant feedback and collaboration in education. Overall, AI has the capacity to revolutionize education and improve learning experiences for students [13,14].

Artificial intelligence systems used in education include expert systems, intelligent tutoring systems and dialogue-based systems. Especially during the Covid 19 Pandemic that affected the whole world in 2020, the use of artificial intelligence has become very important with applications such as data mining and student analytics that support online learning. Studies investigating the role of artificial intelligence in special education, all levels of education, educational robots and smart classes are increasing [15,16]. Using artificial intelligence-supported technologies, working with artificial intelligence tools and producing appropriate tools from the first grade of primary school to teacher training in faculties has become very important in recent years [17,18].

In studies questioning the future of artificial intelligence in education, the importance of experts, educators and students working together is emphasized. For this purpose, teachers' collaboration and support from artificial intelligence can play an important role. However, traditional school structures and culture can make this collaboration difficult to achieve. Using artificial intelligence to meet expectations can be beneficial in reducing teachers' workload and providing support to students. However, it is necessary to be aware of the dangers and potential disadvantages that come with the use of artificial intelligence and to address these points. Still, artificial intelligence has the potential to empower students and create change, but it is important to use it responsibly [19].

There are application examples of researchers who want to benefit from the power of artificial intelligence in the field of education. For example, academics specializing in artificial intelligence and education aim to help professors by creating smart campus environments using smart teaching and management techniques. In addition, efforts are being made to improve student learning experiences while increasing teacher efficiency by using artificial intelligence technologies such as images, face recognition and adaptive learning in schools. In education management and decision-making processes, pedagogical quality is tried to be increased by better evaluating educational data with artificial intelligence and big data analysis. In this context, it is thought that artificial intelligence can have an impact on issues such as adaptive learning, teacher evaluation, smart tutoring robots, smart campuses and virtual classrooms and more, thus affecting student curiosity, interest and happiness [20].

The rapid development of AI technologies has had a significant impact on education. These technologies are especially used in the field of artificial intelligence-based education (AIEd). AIEd aims to support the educational process through the use of artificial intelligence technologies. For this purpose, tools such as smart tutoring systems, chatbots, robots, learning analytics dashboards, adaptive learning systems and automatic evaluation are being developed. Chatbots are specifically based on technologies such as natural language processing, machine learning and deep learning. Early chatbots used either keyword matching mechanisms or natural language processing mechanisms. Looking from today to the future, it can be predicted that the use of artificial intelligence technologies in education will witness further advances [21].

Various tools are being developed regarding the use of artificial intelligence (AI) technologies in education. These tools include chatbots, intelligent tutoring systems, robots, adaptive learning systems, learning analytics dashboards, and automated assessment. Chatbots, for example, use natural language processing, machine learning and deep learning technologies. With the advancement of AI, these technologies are expected to become more widespread in education. According to the literature, the AI educational paradigm can be divided into three groups: AI-guided learners, AI-assisted learners, and AI-enhanced learners. The first paradigm, which is the most basic, is that students are recipients of AI services, but AI can also be used to represent knowledge models and guide cognitive learning [22].

The purpose of this study is to question the role of artificial intelligence (AI) in educational processes and its effect on student happiness based on teacher observations. In particular, this study will address the main themes of AI [23], such as adaptive learning, personalization and learning styles, expert systems, and intelligent tutoring systems, and evaluate the potential impact of these themes on student happiness. In this age where technology is rapidly advancing, the aims of this study are to examine how AI can affect the well-being and happiness of students and how this effect can be evaluated in line with teachers' opinions. In this context, the research problem is as follows:

• How do artificial intelligence technologies affect student happiness?

2. Method

The phenomenology pattern was used in this qualitative study, which aimed to explore teacher experiences regarding educational tools and student happiness involving artificial intelligence. Phenomenology is an analytical approach that aims to reveal individuals' personal experiences and their underlying patterns of meaning. This approach involves a comprehensive examination of individual experiences to gain a deeper understanding of a particular phenomenon. Phenomenology was specifically used in this qualitative study involving teacher insights because it enables researchers to comprehensively and nuanced understanding complex human experiences within their contextual frameworks [24].

2.1. Working group

In phenomenological research, purposeful sampling methods can be used to select various individuals or groups. Maximum variation sampling is also one of the purposive sampling methods and aims to capture a wide range of perspectives on the topic you are interested in [25].. The study group consists of 46 teachers who had the opportunity to closely observe their students' experiences with artificial intelligence and happiness. Regarding the opportunities and infrastructure of the teachers working in the city center of Ankara, primary school classroom teachers (19), mathematics teachers (3), guidance teachers (2), English teachers (7), Turkish teachers (5), science teachers (5). teacher (2), music teacher (2), visual arts teacher (2), computer and instructional technologies teacher (3) and physical education teacher (1) voluntarily participated in the semi-structured interview form. Of the teachers who participated in the study with their valuable opinions, 27 were women and 19 were men. The average professional experience is 2.3 years. While 32 of the teachers work in colleges, 14 teachers work in public schools.

It is claimed that artificial intelligence technologies have created a revolution in education [21]. Studies stating that artificial intelligence technologies have left 20 years behind in education are frequently encountered [3]. Although periods such as twenty-thirty years are expressed and the history of artificial intelligence systems does not date back to very old times; There are studies showing that artificial intelligence has developed tremendously and has begun to be widely used in almost all areas of education [16]. When this literature data is considered together, it is predicted that the teachers who participated in the research may also have opinions about artificial intelligence. In addition, the teachers in the study experienced online education during the pandemic period [18]. Teachers with an average professional experience of less than three years are young teachers who experienced the pandemic period while receiving pre-service teacher training and are therefore closer to technology.

2.2. Data Collection and Analysis

This research aims to explore the impact of artificial intelligence on student happiness in teacher experiences. For this purpose, firstly, an in-depth literature review was conducted. This paper is primarily a review of the literature on the phenomena of artificial intelligence and student happiness. The review process involves searching the relevant literature, analysing and comparing the information found and presenting a concrete case. In the in-depth research process before starting the qualitative research part of the paper, the steps taken included identifying the research topic, evaluating and prioritising sources, identifying relationships and main ideas, outlining, writing and finalising the paper. Tips for success in conducting the review are also considered, including maintaining objectivity and balance, avoiding tedious data presentation and avoiding simplistic conclusions. In the literature review phase, it provides conceptual frameworks for understanding the AI era, highlights inconsistencies in the literature, synthesises results and provides an overview of the field [26,27].

In the study, data were collected using a semi-structured interview form. Descriptive and content analysis results were evaluated according to various criteria. The researcher's perspective and experiences were taken into account. Qualitative research processes, data collection, analysis and reporting are interconnected. However, with careful planning, a structured framework for these processes can be created [24]. This framework guides the research process and ensures that the research is systematic and rigorous. The research process provides an in-depth understanding of students' experiences and exploring how these experiences interact with AI. This provides a broader understanding of how AI can impact students' experiences of happiness. Therefore, this research can successfully fulfill the aim of exploring students' experiences of artificial intelligence and happiness by using the qualitative research processes outlined in the literature.

In this study, participants were informed about the purpose of the research and their verbal consent was obtained for voluntary participation. Interviews were recorded for analysis. Data were analyzed using content analysis, which involves grouping similar views based on specific statements and themes. The main purpose of qualitative data analysis is to reveal hidden information in social reality. Each participant's opinion was categorized, themed and coded using the deductive method, and the findings were presented in a report with frequency and percentage values. Although qualitative research processes are generally intuitive and relative, a general framework can be created with planned studies that include data collection, analysis and reporting [24].

2.3. Validity and Reliability

This article focuses on examining artificial intelligence and student happiness with qualitative research method. Reliability and validity are important issues to consider in qualitative research. To ensure reliability, long-term participation, consistent data collection through continuous observation, and the use of triangulation are recommended and taken into account in this study. For transferability, the research context and processes were defined in detail. Steps such as meticulous documentation, audit trails, peer questioning, member checking, and reflexive journaling were followed to increase reliability [28]. In addition, opinions were taken from teachers who interacted with students over a long period of time to gain a deeper understanding of their experiences with artificial intelligence and its impact on their happiness. On the other hand, validation studies were conducted with the participants to ensure that their experiences were accurately represented. Throughout all of these research processes, care was taken to keep a reflexive journal to reflect on researcher biases and influences.

3. Findings

In interviews with teachers, it is seen that artificial intelligence has entered the agenda of schools. All education components, especially students, and naturally teachers are interested in artificial intelligence. Although an artificial intelligence-centered perspective generally prevails, a collaborative interaction led by students and with artificial intelligence tools is gradually being experienced. In the literature, artificial intelligence training paradigms are divided into three groups. These are expressed in three paradigms: the first paradigm in which students guided by artificial intelligence are just recipients, the first paradigm in which students are guided by artificial intelligence, that is, they learn as collaborators, and the other one, empowered by artificial intelligence, in other words, the student as a leader. Although students in the first group, the most basic paradigm, are generally recipients of artificial intelligence services, artificial intelligence can be significantly used to represent knowledge models and guide cognitive learning [22]. Teacher opinions about how artificial intelligence is used in the context of education embody this literature knowledge.

The expressions obtained in teachers' opinions are grouped under five themes and given in table 1. According to teachers, artificial intelligence (AI) in education benefits students and teachers by providing personalized learning opportunities. Thus, it can be said that artificial intelligence helps student empowerment. Second, AI's ability to reduce administrative workload allows teachers to focus more on teaching. Third, AI can increase students' learning engagement and motivation, making the learning experience more effective. Fourth, the potential of AI raises ethical concerns, and these concerns should be addressed with an approach that prioritizes student well-being. Finally, the integration of AI in education positively impacts students' learning experiences and prepares them for an AI-driven future. These five themes provide important information about how AI can be used in education and the potential impacts of this technology on students, teachers, and administrators.

Table 1. Effects of Artificial Intelligence Technologies on Student Happiness

Theme	Description	Teachers' Opinions	Frequency
Personalized Learning and Student Empowerment	Artificial intelligence (AI) in education benefits students and teachers by providing personalized learning opportunities.	39	%86
Reducing Administrative Burden	AI reduces administrative workload and allows teachers to focus more on teaching.	36	%79
Increasing Participation and Motivation	AI can increase students' learning engagement and motivation.	41	%90
Ethical Considerations and Emotional Well-being	The potential of AI also brings ethical concerns. These concerns must be addressed with an approach that prioritizes student well-being.	12	%26
Preparing Students for an AI-Driven Future	The integration of AI in education positively impacts students' learning experiences and prepares them for an AI-driven future.	20	%43

As seen in the table, Artificial Intelligence (AI) can be used in education in various ways, and this use can have various effects on student happiness and well-being. These effects were determined in five basic themes in collaboration with teacher opinions and literature and are explained in the table. According to the table, AI benefits students and teachers by providing personalized learning opportunities. A significant majority (86%) of the teachers participating in the study made statements regarding artificial intelligence's personalized learning and therefore empowerment of students. On the other hand, artificial intelligence regarding its ability to reduce workload accounts for 79% of teachers' opinions. Teachers do their jobs faster and easier with artificial intelligence. On the other hand, there is a very high percentage of teacher opinions (90%) who find artificial intelligence useful in the context of students' motivation. Teacher opinions regarding examples where artificial intelligence can motivate students just by mentioning its name offer remarkable inferences.

In addition to all these positive aspects, the potential of AI also brings ethical concerns. These expressions of concern in teachers' opinions can be explained by an approach that prioritizes student welfare. A significant, albeit low, percentage of teacher opinions (26%) emphasize the need to be careful about artificial intelligence. Beyond all these themes, the fifth theme is explained with the codes of integration of AI into education, its positive effects on students' learning experiences and preparing them for an AI-driven future. Teachers have very hopeful views (43%) about preparing students for the future with an artificial intelligence experience that has been addressed in all its aspects and fulfilled its requirements. Overall, these five themes, when considered together, provide important information about how AI can be used in education, the potential impacts of this technology, and student happiness.

3.1. Personalized Learning and Student Empowerment

The field of Artificial Intelligence-Education has witnessed the emergence of adaptive and intelligent education systems since the 1970s. These systems were originally created to address the limitations of Computer Assisted Instruction (CAI) systems by offering smarter assessment of student knowledge and personalized instruction. These systems, known as Intelligent CAI or AI-CAI, were intended to replace traditional classroom teaching. However, a change took place in the late 1970s with the introduction of the "private lesson" model. Rather than replacing teachers, AI-CAI systems focused on providing individualized tutoring and solving educational problems. This led to the development of Intelligent Tutoring Systems (ITS). Web-based education has also brought about a paradigm shift, with learning materials being delivered online to meet the needs of students without the presence of a teacher or instructor. This model offers personalized learning experiences and greater flexibility for students [9].

Artificial intelligence-supported education benefits students and teachers by providing personalized, flexible and interesting learning opportunities. Artificial intelligence tools are software that can help students' learning processes [29]. Artificial intelligence tools enable students to be more independent and improve their learning experiences because they personalize the learning phenomenon [30]. It also supports students' activities such as downloading, studying and completing assignments by using web-based and online educational materials. In this way, artificial intelligence plays an important role in education, including early childhood education [3].

The views of a classroom teacher (teacher 6) are important in this context: Artificial intelligence is pure happiness. What have we suffered so far? I was able to do vocalization, syllable combining, aloud reading, concretization, and gamification all through the smart board. This is as much as I know, what I can learn, and what I can get from young friends. Who knows what else is out there? I have been teaching primary reading for years and I have seen that these e-books and applications make things much easier. On the other hand, the statements of an English teacher (Teacher 26) who teaches kindergarten and primary school level are as follows: There are teachers I follow on social media. When they find new applications and share them, I use them immediately. I was working hard on material design using the methods I learned at the faculty. Now everything is as easy as a link. Every child participates. Nobody is left out. Attention is at a high level throughout the lesson. Giving homework and tablet etc. It is very easy to reach each student. I can see everything on the school's smart application. Children can see it too.

AI-powered education systems provide an effective and efficient learning experience by allowing students to receive personalized and adaptive instruction. In addition, AI can be used to provide special support to students by increasing awareness of knowledge gaps and helps in real-time evaluation of complex skills and knowledge [12]. In addition, artificial intelligence tools can be used in online training courses to support students in improving their knowledge and skills. Adaptive education systems aim to adapt content and activities according to students' needs, while intelligent education systems can perform tasks such as coaching students and detecting misconceptions. These tools can analyze students' learning needs, adapt learning content, and provide instant feedback [29].

In this context, the views of a mathematics teacher (teacher 21) can provide an important example: Although I do not use the concept of artificial intelligence directly, the tools I frequently use, namely social media, online education tools, videos with educational content, are included in this artificial intelligence. There are dozens of artificial intelligence-supported applications on my computer at home or on my smartphone anywhere. The applications in which I review my notes, plans, and the achievements I have been able to write down, and prepare and share sample videos with my students *are very successful. Also, as a student, I am also studying for the exam. (There is an exam called KPSS to pass to a public school). We plan my exam process with an online institution and my teacher. I have online classes and there is a smart platform that keeps track of what I need to do every week, teaches in the evenings and keeps me informed of the latest developments. Let me tell you the bottom line: there are artificial intelligence tools while learning, teaching, and maybe even while sleeping. You said, "What is its relationship with happiness?" If all this didn't happen, I'd be even more unhappy than this.*

AI-powered adaptive learning platforms analyze student performance data to adapt educational content. These systems create personalized learning paths by identifying individual strengths, weaknesses and learning styles. When students feel that their education addresses their unique needs, they experience a sense of empowerment and ownership over their learning journey [29]. It offers a range of benefits that enhance students' learning experiences. Personalized learning, instant feedback, enhanced collaboration, access to educational resources, intelligent learning analytics and continuous learning support are some of the benefits that AI brings to education. By using AI technologies, educators can enhance students' learning experiences and increase overall educational effectiveness. [14].

Current research shows how AI can assist in improving learning opportunities for students and management systems. Sustainable Development Goal 4 aims to ensure equitable and inclusive education and promote lifelong learning opportunities for all. AI technologies ensure equitable and inclusive access to education by providing appropriate learning opportunities for marginalized people, people with disabilities, refugees, and those living in isolated communities. AI can also personalize learning and create individual learning plans based on students' strengths, weaknesses, preferences, and activities. The use of these technologies is promising for increasing opportunities in terms of quality and access in education [11].

As a result, artificial intelligence (AI)-supported education systems are used to offer personalized and intelligent education to students. AI tools support students' independent learning skills, enhance learning experiences using online materials, and help teachers facilitate teaching. Artificial intelligence also provides benefits such as providing adaptive education and offering tailored support to students by analyzing student performance data.

3.2. Reducing Administrative Burden

Thanks to artificial intelligence tools, many business processes have become digital. Thanks to these automation processes, human labor and management are reduced. For example, arranging lesson times in schools, preparing classroom boards or adapting learning content are no longer entirely a burden on the teacher. Artificial intelligence can perform such tasks faster and more efficiently. While administrative tasks can overwhelm educators and distract them from teaching; AI simplifies administrative processes by automating routine tasks such as grading, attendance tracking, and scheduling. It is conceivable that when teachers have more time to focus on teaching and guiding, students will benefit from a more engaging and supportive learning environment [10].

Education Management Information System (EMIS) is a system used for education leaders to store, analyze and disseminate information. EMIS can make data-driven decisions with artificial intelligence algorithms. The development of EMIS, the generation of AI-powered data in every field, has the potential to support real-time decisions in every aspect of the education sector. A well-designed and functioning EMIS provides useful information to manage and manage the education system more efficiently, develop feasible and cost-effective plans, formulate responsive policies, and monitor and evaluate education outcomes. With data collected reliably and regularly, AI-enhanced EMIS can automatically analyze data to create data dashboards at both the school level and national level [11].

Artificial intelligence (AI) offers various opportunities and advantages in the field of education. AI has the potential to support learning and can be used for tasks such as diagnoses, recommendations and decisions. Additionally, AI-supported education systems are used to analyze classroom dynamics and identify students at risk. In this way, timely intervention is provided. The pedagogical potential of AI is being evaluated by researchers and practitioners and is increasing scientific outcomes. AI plays an important role in the field of education [12]. AI can also enhance collaborative learning and give teachers insight into students' discussions, thus driving student engagement and learning. [11].

With data analysis skills and machine learning algorithms, educators can learn about students' strengths, weaknesses, and learning patterns. Based on this information, teachers can create customized learning paths, set targeted intervention goals, and provide timely feedback. Artificial intelligence also facilitates the creation of various learning materials and media along with technological advancements in education. Teachers can choose from available platforms and applications without needing an in-depth understanding of technology [30]. It can also help teachers free up their time to focus on student guidance and one-on-one communication [11].

In this context, the following statements made by a classroom teacher (teacher 3) may be revealing: After working in public schools for 24 years, I continue to learn with students in a collage. There are smart practices that I learned during the seminar with my young colleagues and that I want to improve myself. It can be downloaded anywhere, in the classroom, at home, and on my smartphone. I'm not saying to make things easier, look, he does things himself. Many websites have countless smart applications. I open a few and show them, and more than half of the class learns them on their own. While they are doing their books and notebooks, I can devote plenty of time to the remaining students. This means being able to make interventions that used to be done outside of class, during extra time, or after a semester's delay, within that class. My workload is less and my mind is more at ease. How happy I am to leave learned children on the way home in the evening! Even though it was difficult to upload photos, videos, a lot of homework files, etc. to the school's smart application, it worked very well. The parent also becomes happy as he keeps track of everything, and it is enough to summarize the situation with the satisfied parent in two words once a month.

In addition, the views of a physical education teacher contain important concretizations in the context of reducing the administrative burden: I prepared children for tournaments for years. These are the ones that come to my mind: chess, tennis, futsal, wrestling (my field, laughter), darts. In the past, preparing documents for these tournaments was more difficult than preparing the children. It's easier now. It's all on the computer. You mentioned what else there is, I'll tell you as soon as I remember: There is filling out the class notebook, for example. I would forget what I achieved in each class. It's been a few years since I learned a practice, it's legendary. Always up to date. Open it, look at it and fill it for two minutes...

It is anticipated that as AI development progresses, the roles of educators will transform. It is suggested that as AI tools take on more analytical tasks, educators will need to focus on “softer” intuitive and empathetic skills [31]. In one sense, the roles of educators are decreasing on the one hand and increasing on the other. Artificial intelligence will also play an important role in administrative tasks and can help streamline processes and reduce costs. However, it is a controversial issue if artificial intelligence completely replaces the teacher or lecturer. While work becomes easier, educators' human touch roles such as guidance, establishing relationships, and encouraging creative thinking come to the fore with increasing importance [30].

As a result, the use of artificial intelligence in education enables the automation of many processes, reducing the burden on teachers and administrators. Artificial intelligence also provides different opportunities for students by offering personalized learning strategies, targeted interventions, and timely feedback. It can analyze classroom dynamics and identify students at risk so timely intervention can be made. However, there is debate about whether artificial intelligence will completely replace teachers or whether their roles will change. Artificial intelligence has the potential to improve education and empower educators to create a more effective learning environment, but educators' skills such as motivation, relationship building, and creative thinking are still important.

3.3. Increasing Participation and Motivation

The future of artificial intelligence in education is exciting. As technology advances, it is clear that we will see more advanced AI-powered tools and platforms. These tools will be able to improve students' learning experience even more than today. Artificial intelligence can enhance the role of educators, personalize learning and facilitate access to information. However, in this process, educators' motivating human touch roles such as mentoring, building relationships, and encouraging creative thinking are increasingly important [30].

Students' participation in the learning process is a strong supporter of motivation. Using engaging educational approaches such as game-based learning and mobile learning can improve learning motivation and outcomes. When an artificial intelligence-supported recommendation system is implemented that affects students' learning engagement, motivation, and outcomes, positive results can be achieved [13].

Artificial intelligence has the potential to tackle some of the biggest challenges in education today, innovate teaching and learning practices, and accelerate progress towards global education goals (UNESCO, 2023). Artificial Intelligence (AI) technologies can increase student engagement and provide personalized learning experiences. AI facilitates students' self-learning by providing immediate-constructive feedback, enabling them to make self-cognitive and behavioral corrections. Additionally, AI can support collaborative learning environments, creating a dynamic learning environment that mimics real-world scenarios. Intelligent teaching systems and virtual learning assistants encourage active participation, personalized learning, instant feedback, and collaborative learning in education through AI-powered tools [14].

A recent study investigated how teacher support may moderate the effects of student mastery on need satisfaction and intrinsic motivation to learn with AI technologies. The results of the study show that both teacher support and student expertise, including self-regulated learning and digital literacy, play an important role in intrinsic motivation and competence to learn with AI chatbots. This highlights the importance of teacher participation and guidance in the effective use of AI technologies in educational environments [33].

Artificial intelligence (AI) is a technology that has the potential to greatly improve education systems. AI is growing rapidly in the education sector, and many innovative companies are creating AI tools to transform learning processes. These tools can create immersive virtual learning environments, break down language barriers, create custom plans for each student, and more. For example, platforms such as Course Hero and Gradescope are tools that demonstrate the potential of AI in education. The application called Course Hero offers homework help to students using artificial intelligence. Duolingo, on the other hand, provides personalized language lessons, ALEKS or MathGPTPro adaptive assessments for math, and personalized learning plans. Doping Memory, on the other hand, personalizes traditional classroom guidance with online tools [34]. Launched on November 30, 2022, ChatGPT attracted attention with its power and ability to perform complex tasks, reaching more than one million users in a week. Although its use in the field of education is controversial, the potential benefits and inherent limitations of ChatGPT in promoting teaching and learning are highlighted [35]. With these and countless similar applications, artificial intelligence provides a variety of opportunities such as individualized education, automatic assessment, data analysis, student collaboration and interaction.

It is possible to concretize this issue with the experiences of a teacher (teacher 29): I myself was appointed thanks to artificial intelligence. I assume you are aware of collage working hours. If you don't know, it's a lot. We are at school from morning to evening and I have no free time. Even if there is, other work and course fillings are done. So, I enrolled in an online course during my college years. Sometimes I worked all night until the morning. The system that identified the subjects I was missing caught my attention at that time. Now I help children. I give a lot of online homework on tablets.

Technological education stakeholders, constantly evolving with significant innovations and changes, have significant impacts on the intellectual happiness and well-being of students. These tools are so fast and widespread that various academic sectors are having great difficulty keeping up with education-based technology trends that can enhance learning and teaching experiences. These education technology trends include broad open online courses or MOOCs, artificial intelligence or AI, augmented reality (AR) and virtual realities (VR), gamification, big data, learning analytics, and many other forms of learning that can be used outside of or in support of the traditional classroom environment. They are available in numerous options [36].

The statements of a classroom teacher (Teacher 12) on this subject can be included: I have been using smart boards etc. for approximately 10 years. At that time it was called smart board, but now it is something completely different. This must be artificial intelligence. Let me put it this way: In the early days, boards were like projectors. Maybe it seemed that way to us, or that we could use it that much. There were things like overhead projectors when we were kids... There was an ongoing technology called projection devices, which were less advanced than overhead projectors, and smart boards. Oh, let me tell you this, the common thing in every period is that these tools always had the motivational power. Technology continues to motivate all children. These tools turn into miracles, especially in the hands of teachers who are knowledgeable (who are interested in technology and can use it correctly in lessons).

Consequently, the use of artificial intelligence (AI) in education holds great promise for improving the learning experience for students. Engaging approaches such as game-based and mobile learning can increase motivation and results, and an AI-powered recommendation system can further increase engagement. However, the role of educators in building relationships, guiding students, and encouraging creative thinking remains important. Teacher support and student expertise in self-regulated learning and digital literacy are also important factors in motivating students to learn with AI technologies.

3.4. Ethical Considerations and Emotional Well-being

The use of AI in education is beneficial for supporting emotional well-being and personalized learning experiences. However, it also raises ethical concerns regarding student privacy, algorithmic bias, transparency, misuse of student data, and potential inequality. AI systems should not undermine student and teacher autonomy, and measures such as transparent communication, ongoing evaluation, and strong security measures are necessary. While AI-powered platforms can analyze students' learning styles and deficiencies to provide personalized support, it is crucial to address privacy concerns and provide fair algorithms. Similarly, AI-powered collaboration tools can improve group discussions and collaboration, but over-reliance on AI can reduce empathy and emotional connection between students. To ensure a balanced approach, it is important to find the right balance between leveraging AI technologies and maintaining ethical principles [37].

The integration of artificial intelligence in education also presents challenges such as privacy and security of student data, ethical issues, and passivation of learning experiences. AI collects and analyzes student data for personalized learning experiences, which means data protection and regulations are important. The use of student data and decision processes raise ethical questions and highlight that AI systems must comply with ethical standards such as fairness and transparency. Additionally, it is possible for AI to deliver personalized learning experiences, but this could lead to passive learning, which carries the risk of overconfidence. Therefore, it is important to strike a balance between human interaction and AI training. Policymakers and educators must address these challenges and manage the role of AI in education in accordance with ethical standards [14].

The views of an exemplary teacher (teacher 38) in the context of establishing a balance between human interaction and artificial intelligence are as follows: We had produced a project to establish a small YouTube channel. The students and I were recording math subjects at home. Watching and interpreting those videos on the smart board in our classroom was a good artificial intelligence-supported activity. We never showed our faces while doing this. Just our hands and a white paper, pencils etc. Most people share everything quite easily. We focus on our lesson. Our goal is not to get likes, etc., but just to learn by telling. It may not be a problem today, but we don't know what will happen in the future. It is better to be cautious. It is my responsibility to protect the rights of children.

AI in education requires carefully approaching issues of privacy, ethics, and striking a balance between AI and human education. The integration of AI in education brings challenges that need to be addressed, such as privacy and ethical issues regarding student data. AI is based on collecting and analyzing personal information to provide personalized experiences. Additionally, over-reliance on AI technologies may pose a risk of a passive learning experience for students. Maintaining a balance between AI and human training is important to sustain meaningful interactions and promote deeper understanding [14].

It is clear that AI and other AI-enabled technologies are here to facilitate human life and contribute to the progress of humanity. However, it is important not to fall into the mindset that technology is good by default and to take a critical approach before fully integrating AI into educational processes. As part of this critical approach, it is important to first establish an ethical policy and clearly define ethical boundaries for how AI will use human-generated data. Additionally, it is important to test AI-enabled training processes and retest automated processes to prevent mechanical learning [23].

In this context, the views of a guidance counselor can be included: It is necessary to think not once but ten times and then share. Everyone has smart devices. Everyone is a videomaker. Everyone is a director. How did it happen? Of course, thanks to artificial intelligence applications. When it comes to easy sharing of effects and music, we are always on the internet. Attention I say attention and attention.

Publicly available generative artificial intelligence (GenAI) tools are rapidly expanding and are therefore outpacing national regulations. This jeopardizes users' data privacy and leaves educational institutions unprepared to use the tools correctly. Therefore, GenAI tools need to be streamlined by adopting a human-centered approach. Key steps such as requiring data privacy protection and setting an age limit for independent conversations with GenAI platforms are suggested. Additionally, it is important to take a human-centered and age-appropriate approach to ethical verification and pedagogical design processes [38].

As a result, artificial intelligence provides many benefits in education and has an important role in supporting emotional well-being and personalized learning experiences. However, it also raises ethical concerns such as student privacy, algorithmic bias, transparency, and misuse of data. Transparent communication, ongoing evaluation, and strong security measures are required to address these concerns. AI-powered platforms have the potential to provide personalized support to students, but privacy concerns and fair algorithms need to be addressed significantly. Policymakers and educators are required to manage artificial intelligence in education in accordance with ethical standards. A critical approach is needed to determine the ethical limits of artificial intelligence use and ensure a human-centered approach.

3.5. Preparing Students for an AI-Driven Future

Artificial intelligence (AI) has various transformational areas in the field of education. Going beyond traditional pedagogical approaches, AI can create educational robots that enrich the learning experience using embedded computer systems. These robots are capable of teaching basic skills such as spelling and pronunciation and can be customized to the student's individual abilities. Additionally, AI supports the use of web-based and online training materials. This facilitates students' activities such as downloading educational materials, studying them, and completing their assignments. In this way, AI plays an important role at every stage of education, from early childhood education to primary school to lifelong learning [3]. This highlights the potential and impact of AI in the education sector. Therefore, applications of AI in education can play an important role in shaping future pedagogical strategies and educational policies. This can help the education sector better understand how it can integrate AI technologies to improve students' learning experiences and achieve educational goals.

The integration of artificial intelligence (AI) in education positively affects students' learning experiences. Thanks to AI, personalized learning, instant feedback and enhanced collaboration can be achieved. Education is important in improving artificial intelligence skills and aims to close the skills gap in this area. Artificial intelligence skills require not only the use of technology but also the rethinking of educational contents and methods. Developing new digital skills is important in a society empowered by artificial intelligence. The aim is to unlock the power of digital competencies that can analyze, use, and decode Artificial Intelligence [11].

I would like to include the opinions of a classroom teacher (Teacher 2) who was very impressed by me while listening: We started primary school online during the pandemic. All my parents and school administration asked, "Teacher, how did you teach these children to read and write remotely?" They said the sentence many times. And I always say, “This is their age. Children are more ready for this age than we are. “I just adapted.” I told. That's it. We are preparing them for the future. Yes, the pandemic left a lot of trauma. We are very sorry. We lost our loved ones. But we had the opportunity to quickly experience the future lives of our children. There was an advertisement in my high school years that said the future will come fast. Now just like that, the future of these children came pretty quickly.

The Information and Communication Technologies Competence Framework for Teachers (ICT-CFT), developed by UNESCO in 2011, outlines the competencies that teachers need to integrate into their practice to develop critical knowledge and awareness in students in the digital age. The framework highlights the role of digital technologies in supporting six key areas of knowledge: Understanding ICT in Education, Curriculum and Assessment, Pedagogy, ICT, Organization and Management and Teacher Professional Learning. It also describes three stages of knowledge acquisition: technology literacy, knowledge deepening, and knowledge creation [34].

It determines the competencies required for teachers to integrate digital technologies into their professional practices and develop critical knowledge and awareness with their students. It also emphasizes that it is not enough for teachers to only have the skills to manage and teach digital technologies, but they also need to support their students in having the ability to collaborate. These skills become a part of citizenship education in the growing technological world, enabling students to participate in the digital society [40].

In this context, the adventure of opening a robotic coding course and the observations of a classroom teacher (teacher 41) can be given as an example: Getting a robotic coding certificate was quite good. Teaching is my main job. My hobby is coding workshop. There must be work to do at school all day long. We even made traditional lessons more fun. Servest activity lessons are written robotically in our syllabus. We changed the name. We switched to Arduino sets this year with the class I started last year. There's been a lot of progress. Even other students at school called me their robotics teacher. We are considering participating in robot tournaments.

The proliferation of artificial intelligence (AI) and digital technologies in the classroom requires teachers to acquire new skills to use these technologies effectively. These capabilities include understanding how AI-powered systems work, interpreting and managing the data provided by these systems. Additionally, teachers need to understand the dangers and opportunities of AI, develop more human capabilities, and equip students with skills that cannot possibly be replaced by machines. Therefore, teacher preparation programs should take these new abilities into account [41]. However, it is not enough for teachers to simply understand and comprehend these new technological possibilities; AI developers also need to collaborate with educators, content designers, and interdisciplinary experts.

Artificial intelligence should be seen as a tool to support educators rather than replace them. Collaboration between artificial intelligence technologies and true natural intelligence teachers is important to create effective learning environments. Digital literacy and artificial intelligence skills should also be encouraged. Students must learn to critically evaluate and use AI-enabled tools and understand ethical rules. In this way, artificial intelligence can improve learning experiences and create positive effects in the field of education [14]. The education world should adapt to technological developments to improve the quality of information and communication technology-focused education by using artificial intelligence systems [30].

The views of a classroom teacher (teacher 18) can be included to support this issue: I also studied computer teaching as a minor so that children would be ready for the future. First of all, I thought I should be ready at the faculty. It's good that! I want to tell you about some of the things I saw in the teachers' room. Many teachers are very eager to adapt and learn new technologies. On the other hand, there are teachers who try to do business only with old stone bath books. Of course, I would like to mention those who are successful in this way, keeping them aside. I think classes whose teachers adapt to artificial intelligence technologies are luckier. So why shouldn't other classes be lucky?

Overall, AI has the potential to greatly improve the quality of education and enhance students' learning experiences. Artificial intelligence can be used to develop educational robots that teach basic skills and can be customized to individual student abilities. Additionally, artificial intelligence supports the use of web-based and online educational materials, providing easier access to learning resources. The integration of artificial intelligence in education enables personalized learning, instant feedback, and improved collaboration between students. Teaching artificial intelligence skills in schools is important to close the skills gap in this rapidly developing field. A happy student in the age of AI is one who feels equipped to navigate the AI environment with confidence.

4. Conclusion

As a branch of computer science that aims to have intelligent behavior and improve human actions, the effects of artificial intelligence on the transformation in the field of education have been dizzying. In education, AI is used for a variety of purposes, such as delivering personalized learning experiences, intelligent training, and data-driven insights through machine learning systems and algorithmic processes. By integrating AI into education, learning processes become more adaptable and personalized. AI, also known as Machine Intelligence, involves the ability to learn and perform various tasks. AI is a branch of science whose visibility increases day by day compared to the previous day, with its algorithmic structure that imitates human intelligence and tries to meet demands by processing user data [23-37].

Collaboration studies in the field of artificial intelligence and education are taking place with increasing momentum day by day. Literature reviews show that artificial intelligence has become increasingly visible in the world of education in recent years. The research results, conducted using the Web of Science database during the literature review phase of this study, show with the VOSviewer visual which concepts artificial intelligence stands out in the field of education.

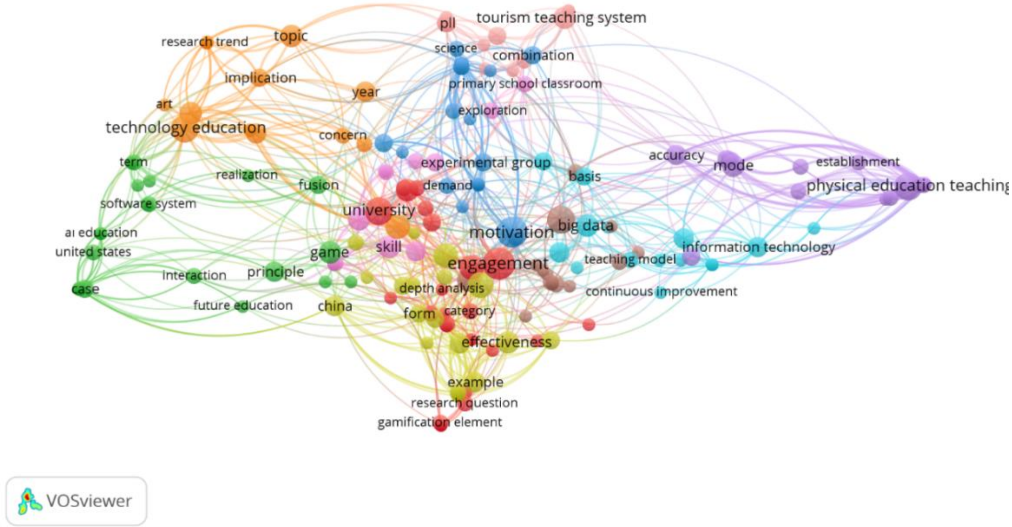

Figure 2 – The impact of artificial intelligence on education

This VOSviewer image shows a word cloud and its relationships analyzing the impact of AI on education. It aims to explain the role of key concepts such as image, technology, education, motivation and effectiveness in the field. Relationships between different concepts are represented by the thickness of the lines, indicating the frequency with which these concepts occur together. Additionally, different colors indicate different groups or related concepts. This image helps viewers understand the main themes and important issues related to artificial intelligence in education. Overall, this image is a valuable tool for identifying concepts that are important in the field of education, especially concepts such as technology, motivation, and effectiveness. Understanding these concepts is extremely important to understanding the impact of artificial intelligence on education.

The role of artificial intelligence (AI) in education is considered in communication processes, human-computer interaction and computer-mediated communication. AI has the potential to transform the way students learn and interact with information. Although AI has both positive and negative impacts on human communication, there is optimism in exploring new concepts and frameworks. AI can automate learning and provide learners with knowledge transfer and continuous improvement in real-time experiences. AI applications in education include creating smart campus environments, increasing teacher efficiency, improving student learning experiences and evaluating educational data. AI is seen as a future component of educational processes such as adaptive learning, personalization, expert systems and intelligent tutoring systems by promoting students' well-being and happiness.

Artificial intelligence cognitive agents, coded learning contents, can perform various cognitive activities such as speaking, hearing, seeing and learning. Artificial intelligence technologies such as vision, natural language and speech have transformed traditional education, transforming information processing and intelligent adaptive learning. This has inspired educational institutions and teachers to rethink their curricula. Teachers and students can make a customized learning plan based on students' needs and existing learning environments by eliminating mismatches through various AI-based methods.

This can provide a more engaging learning experience for students while also increasing their learning capacity and efficiency. Studies addressing the relationship between artificial intelligence and education may benefit teachers, students, and researchers in the future by providing teaching strategies, methods, and techniques, as well as materials for thinking about new directions and applications of artificial intelligence in education [30].

In conclusion, artificial intelligence plays an important role in the field of education and will shape future learning processes. The age of artificial intelligence presents students with both opportunities and challenges. Educators can provide a positive and fulfilling educational experience by focusing on personalized learning, emotional well-being, and ethical issues. It is important to embrace AI as a tool that increases students' happiness and equips them for a dynamic future.

Suggestions

Artificial intelligence (AI) offers a broad perspective with the potential to improve the quality of education and enrich students' learning experiences. AI can be used as a tool to create educational robots that can adapt to students' individual abilities and teach basic skills. Additionally, AI serves as a support mechanism to provide more convenient access to web-based and online educational materials. Incorporating AI into educational processes brings opportunities such as personalized learning, rapid feedback and increased collaboration between students. Teaching AI skills in schools is critical to addressing the skills gap required by this rapidly evolving field. In the age of AI, a student's happiness is directly related to their ability to feel safe in the AI environment.

References

[1] Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

[2] Copeland, B. (2024, April 24). artificial intelligence. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

[3] Chen, L., Chen P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review," in *IEEE Access*, vol. 8, pp. 75264-75278, 2020, https://doi.org/10.1109/ACCESS.2020.2988510

[4] McCarthy, J. (2007). What is artificial intelligence. whatisai.dvi (unl.edu)

[5] Sutton, R. S. 2020. John McCarthy's Definition of Intelligence. Journal of Artificial General Intelligence 11(2):66–67. https://doi.org/10.2478/jagi-2020-0003

[6] Wang, P. (2020). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *11*(2), 73-86. https://doi.org/10.2478/jagi-2020-0003

[7] Wang, H., Fu, T., Du, Y. *et al.* (2023). Scientific discovery in the age of artificial intelligence. *Nature* 620, 47–60. https://doi.org/10.1038/s41586-023-06221-2

[8] Sundar, S.S. & Lee, E. (2022). Rethinking Communication in the Era of Artificial Intelligence, *Human Communication Research*, Volume 48, Issue 3, Pages 379–385, https://doi.org/10.1093/hcr/hqac014

[9] Kahraman, H. T., Sagiroglu, S., & Colak, I. (2010, October). Development of adaptive and intelligent web-based educational systems. In *2010 4th international conference on application of information and communication technologies* (pp. 1-5). IEEE. 10.1109/ICAICT.2010.5612054

[10] Verganti R. Vendraminelli L. Iansiti M. (2020). Innovation and design in the age of artificial intelligence.Journal of Product Innovation Management, 37(3), 212–227. https://doi.org/10.1111/jpim.12523

[11] Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development. https://hdl.handle.net/20.500.12799/6533

[12] Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, *25*(1), 28–47. https://www.jstor.org/stable/48647028

[13] Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. Computers & Education, 194, https://doi.org/10.1016/j.compedu.2022.104684

[14] Potter, K., Abill, R. & Louis, F. (2024). The Impact of Artificial Intelligence on Students' Learning Experience. https://ssrn.com/abstract=4716747 or http://dx.doi.org/10.2139/ssrn.4716747

[15] Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. Learning, Media and Technology, 45(2), 107–114. https://doi.org/10.1080/17439884.2020.1761641

[16] Coşkun, F., & Gülleroğlu, H. D. (2021). Yapay Zekanın Tarih İçindeki Gelişimi ve Eğitimde Kullanılması. Ankara University Journal of Faculty of Educational Sciences (JFES), 54(3), 947-966. https://doi.org/10.30964/auebfd.916220

[17] Fidan, M. (2021). COVID-19 and primary school 1st grade in Turkey: starting primary school in the pandemic based on teachers' views. Temel Eğitim, 3(1), 15-24. https://doi.org/10.52105/temelegitim.3.1.2

[18] Fidan, M. (2021b). Covid-19 pandemisinde öğretmenlik. Maarif Mektepleri Uluslararası Eğitim Bilimleri Dergisi, 5(1), 1-14. https://doi.org/10.46762/mamulebd.789097

[19] Fullan, M., Azorín, C., Harris, A., & Jones, M. (2023). Artificial intelligence and school leadership: challenges, opportunities and implications. School Leadership & Management, 1–8. https://doi.org/10.1080/13632434.2023.2246856

[20] Alam, A. (2022). Employing Adaptive Learning and Intelligent Tutoring Robots for Virtual Classrooms and Smart Campuses: Reforming Education in the Age of Artificial Intelligence. In: Shaw, R.N., Das, S., Piuri, V., Bianchini, M. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Electrical Engineering, vol 914. Springer, Singapore. https://doi.org/10.1007/978-981-19-2980-9_32

[21] Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. Contemporary Educational Technology, 15(3), ep429. https://doi.org/10.30935/cedtech/13152

[22] Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers & Education: Artificial Intelligence, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

[23] Goksel, N. & Bozkurt, A. (2019). Artificial Intelligence in Education: Current Insights and Future Perspectives. In S. Sisman-Ugur & G. Kurubacak (Eds.), *Handbook of Research on Learning in the Age of Transhumanism* (pp. 224-236). IGI Global. https://doi.org/10.4018/978-1-5225-8431-5.ch014

[24] Creswell, J. W., & Poth, C. N. (2016). Qualitative Inquiry and Research Design: Choosing Among Five Approaches. SAGE Publications.

[25] Creswell, J. W., & Guetterman, T. C. (2019). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (6th ed.). https://eric.ed.gov/?id=ED594549

[26] Manisha, B. (2023). A Step-by-Step Guide to Writing a Scientific Review Article, Journal of Breast Imaging, Volume 5, Issue 4, Pages 480–485, https://doi.org/10.1093/jbi/wbad028

[27] Palmatier, R.W., Houston, M.B. & Hulland, J. (2018). Review articles: purpose, process, and structure. J. of the Acad. Mark. Sci. 46, 1–5. https://doi.org/10.1007/s11747-017-0563-4

[28] Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. Evidence-Based Nursing, 18(2), 34-35. https://doi.org/10.1136/eb-2015-102054

[29] Pokrivcakova,S.(2019).Preparing teachers for the application of AI-powered technologies in foreign language education. Journal of Language and Cultural Education,7(3) 135-153. https://doi.org/10.2478/jolace-2019-0025

[30] Pratama, M. P., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. *Klasikal : Journal Of Education, Language Teaching And Science*, *5*(2), 350–357. https://doi.org/10.52208/klasikal.v5i2.877

[31] Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, *21*(2), 155-172. https://doi.org/10.1177/1094670517752459

[32] UNESCO. (2023, April 20). How can artificial intelligence enhance education? Retrieved from https://www.unesco.org/en/articles/how-can-artificial-intelligence-enhance-education

[33] Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. Interactive Learning Environments. https://doi.org/10.1080/10494820.2023.2172044

[34] McFarland, A. (2024, March 1). Eğitim için En İyi 10 Yapay Zeka Aracı. Unite.AI

[35] Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. Journal of AI, 7(1), 52-62. https://doi.org/10.61969/jai.1337500

[36] Pearce, R. (2022). Technology's Impact on Students' Intellectual Happiness, Cognition, and Overall Well-Being. In ICERI2022 Proceedings (pp. 1079-1088). IATED

[37] Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI Ethics, 2, 431–440. https://doi.org/10.1007/s43681-021-00096-7

[38] UNESCO. (2023, September 7). Guidance for generative AI in education and research. Retrieved from https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

[39] UNESCO. (2011). Information and Communication Technologies Competency Framework for Teachers. Paris: UNESCO. Updated in 2018.4. https://unesdoc.unesco.org/ark:/48223/pf0000265721

[40] UNESCO. (2016). ICT Competencies and Standards from the Pedagogical Dimension. Santiago: UNESCO. Erişim adresi: https://unesdoc.unesco.org/ark:/48223/pf0000247122

[41] Luckin, R; Holmes, W; (2016) Intelligence Unleashed: An argument for AI in Education. UCL Knowledge Lab: London, UK. https://discovery.ucl.ac.uk/id/eprint/1475756/