



JOURNAL OF INFORMATION TECHNOLOGIES

BİLİŞİM TEKNOLOJİLERİ DERGİSİ

Volume / Cilt 17 Number / Sayı 2 Year / Yıl 2024 Month / Ay April / Nisan





GAZİ ÜNİVERSİTESİ (GAZİ UNIVERSITY)
BİLİŞİM ENSTİTÜSÜ (INSTITUTE OF INFORMATICS)

BİLİŞİM TEKNOLOJİLERİ DERGİSİ
(JOURNAL OF INFORMATION TECHNOLOGIES)
ISSN: 1307-9697 e-ISSN: 2147-0715



Cilt (Volume): 17

Sayı (Issue): 7

Ekim (October) 2024

Sahibi (Owner)
Dr. Musa YILDIZ
Rektör (Rector)

Genel Yayın Yönetmeni & Baş Editör
(General Publication Director & Editor in Chief)
Dr. Ashhan TÜFEKÇİ
Bilişim Enstitüsü Müdürü
Director of Institute of Informatics

Yardımcı Editör
(Associate Editor)
Dr. Öner BARUT
Bilişim Enstitüsü Müdür Yardımcısı
Vice Director of Institute of Informatics

Editörler (Editors)

❖ Dr. Serdar KULA	Gazi Üniversitesi (Gazi University)
❖ Dr. Hüseyin POLAT	Gazi Üniversitesi (Gazi University)
❖ Dr. Resul DAŞ	Fırat Üniversitesi (Fırat University)
❖ Dr. Muzaffer KAPANOĞLU	Eskişehir Osmangazi Üniversitesi (Eskişehir Osmangazi University)
❖ Dr. Mehmet ŞİMŞEK	Milli Savunma Üniversitesi (National Defence University)
❖ Dr. Mehmet Sıraç ÖZERDEM	Dicle Üniversitesi (Dicle University)
❖ Dr. Mustafa Mahir ÜLGÜ	Sağlık Bakanlığı
❖ Dr. Murat YILMAZ	Gazi Üniversitesi (Gazi University)
❖ Dr. Oktay YILDIZ	Gazi Üniversitesi (Gazi University)
❖ Dr. Olgun DEĞİRMENÇİ	TOBB ETÜ (TOBB Economics and Technology University)
❖ Dr. Recep BENZER	Gazi Üniversitesi (Gazi University)
❖ Dr. Ö. Tolga PUSATLI	Çankaya Üniversitesi (Çankaya University)
❖ Dr. Cihangir TEZCAN	Orta Doğu Teknik Üniversitesi (Middle East Technical University)
❖ Dr. Mehmet SEVRİ	Recep Tayyip Erdoğan Üniversitesi (Recep Tayyip Erdoğan University)
❖ Dr. Muhammed Ali KOŞAN	Kahramanmaraş İstiklal Üniversitesi (Kahramanmaraş Istiklal University)

Yayın Danışma Kurulu (Editorial Advisory Board)

Dr. Ahmet COŞAR <i>Turkish Aeronautical Association University, Turkey</i>	Dr. Aslanbek NAZİEV <i>Ryazan State University, Russia</i>	Dr. Bogdan PATRUT <i>Alexandru Ioan Cuza University of Iasi, Romania</i>
Dr. Deepak GUPTA <i>Maharaja Agrasen Institute of Technology, India</i>	Dr. Jafar A. ALZUBİ <i>Al-Balqa Applied University, Jordan</i>	Dr. Jolanta SABAITYTĖ <i>Vilnius Gediminas Technical University, Lithuania</i>
Dr. Ilya LEVİN <i>Tel Aviv University, Israel</i>	Dr. Pınar KARAGÖZ <i>Middle East Technical University, Turkey</i>	Dr. Ufuk ÇAĞLAYAN <i>Yaşar University, Turkey</i>
Dr. Veysi İŞLER <i>Hasan Kalyoncu University, Turkey</i>	Dr. Victor Hugo Costa DE ALBUQUERQUE <i>Universidade de Fortaleza, Brazil</i>	Dr. Vijender Kumar SOLANKİ <i>CMR Institute of Technology, India</i>
Dr. Ebrahim KHOSRAVI <i>Clayton State University, United States</i>		

Dil Editörü
(Language Editor)
Dr. Çağla Gizem AKKAŞ
*Bilişim Enstitüsü
Institute of Informatics*

Teknik Sorumlu
(Technical Assistant)
Candan TÜMER
*Bilişim Enstitüsü
Institute of Informatics*

Teknik Sorumlu
(Technical Assistant)
Yasemin İÇTÜZER
*Bilişim Enstitüsü
Institute of Informatics*

Sekreterlik
(Secretary)
Bilişim Teknolojileri Dergisi
*Bilişim Enstitüsü
Institute of Informatics*

Bilişim Teknolojileri Dergisi uluslararası hakemli bir dergidir.
Journal of Information Technologies is an international refereed journal.

Yazışma Adresi (Contact Address)
Tunus Cad. No: 35 Kavaklıdere Çankaya/ANKARA
Telefon / Telephone: 0312 202 38 01
Faks / Fax: 0312 212 79 29

Çevrimiçi Değerlendirme Sistemi (Online Evaluation System)
<http://dergipark.gov.tr/gazibtd>
E-posta (e-mail): btd@gazi.edu.tr

Bilişim Teknolojileri Dergisi 3 ayda bir (Ocak, Nisan, Temmuz, Ekim) yayınlanmaktadır.
Journal of Information Technologies is published every 3 months (January, April, July, October).

Algo-AR: Development of an Augmented Reality-Supported Tangible Programming Tool to Improve Algorithmic Thinking Skills

Araştırma Makalesi/Research Article

 Zeynep ÇİPİLOĞLU YILDIZ^{1*},  Süleyman DOĞAN²

¹Bilgisayar Mühendisliği, Manisa Celal Bayar Üniversitesi, Manisa, Türkiye,

² Bilgisayar Mühendisliği, Manisa Celal Bayar Üniversitesi, Manisa, Türkiye

zeynep.cipiloglu@cbu.edu.tr, suleyman@codecosoft.com

(Geliş/Received:08.12.2023; Kabul/Accepted:02.02.2024)

DOI: 10.17671/gazibtd.1398781

Abstract— The main purpose of this study is to develop an educational tool to help children acquire algorithmic thinking skills at an early age while having fun. The methodology combines modern technologies and approaches such as augmented reality (AR), gamification, and tangible user interfaces. In this application, the coding components consist of specially designed tangible command blocks in the form of jigsaw puzzle pieces. The application contains a 3D multi-level game environment, and the user is expected to control the game character by constructing an algorithm with physical command blocks. The constructed algorithm is scanned using a mobile AR application and converted into code that controls the game environment. The major design considerations during the development of this application were simplicity and accessibility. All design decisions were delineated extensively in the paper. In conclusion, an augmented reality-based gamified tangible programming kit is proposed to improve children's algorithmic thinking skills at an early age. The application requires only a smartphone and printable command blocks. Thus, an inexpensive, accessible, and entertaining educational tool is developed.

Keywords— algorithmic thinking, augmented reality (AR), gamification, heuristic evaluation, STEM, tangible user interfaces (TUIs)

Algo-AR: Algoritmik Düşünme Becerilerinin Geliştirilmesi İçin Artırılmış Gerçeklik Destekli bir Somut Programlama Aracı Geliştirilmesi

Özet— Bu çalışmanın temel amacı, çocukların eğlenirken algoritmik düşünme becerilerini erken yaşta kazanmalarına yardımcı olacak bir eğitim aracı geliştirmektir. Metodoloji, artırılmış gerçeklik (AG), oyunlaştırma ve somut kullanıcı arayüzleri gibi modern teknolojileri ve yaklaşımları birleştirmektedir. Bu uygulamada kodlama bileşenleri, yapboz parçaları şeklinde özel olarak tasarlanmış fiziksel komut bloklarından oluşmaktadır. Uygulama, çeşitli seviyeleri olan üç boyutlu bir oyun ortamı içermekte ve kullanıcının fiziksel komut blokları ile bir algoritma oluşturarak oyun karakterini kontrol etmesi beklenmektedir. Oluşturulan algoritma bir mobil artırılmış gerçeklik uygulaması ile taranmakta ve oyun ortamını kontrol eden koda dönüştürülmektedir. Bu uygulamanın geliştirilmesi sırasında tasarımda dikkat edilen başlıca hususlar basitlik ve erişilebilirlik olmuştur. Tüm tasarım kararları makalede kapsamlı bir şekilde açıklanmıştır. Sonuç olarak, erken yaşta çocukların algoritmik becerilerini geliştirmek için artırılmış gerçeklik destekli oyunlaştırılmış somut bir programlama aracı önerilmiştir. Uygulama yalnızca bir akıllı telefona ve yazdırılabilir komut bloklarına ihtiyaç duymaktadır. Böylece ucuz, erişilebilir ve eğlenceli bir eğitim aracı geliştirilmiştir.

Anahtar Kelimeler— algoritmik düşünme, artırılmış gerçeklik (AG), oyunlaştırma, sezgisel değerlendirme, STEM, somut kullanıcı arayüzleri

1. INTRODUCTION

We live in an era of digital transformation where software development and programming skills are essential. Computational thinking (CT), algorithm design, algorithmic thinking and coding skills are extremely important for today and the future. Acquiring these skills becomes more difficult with age. Therefore, activities that develop computational and algorithmic thinking skills must be embraced in preschool and primary education curricula.

Many different approaches and tools have been developed to improve these skills. These approaches can be broadly divided into two categories “plugged” and “unplugged”, in terms of whether they rely on computers. The plugged category includes traditional text-based or visual programming tools. There are also unplugged approaches, which are free of computers. In these applications, problem-solving activities are carried out using real-world objects such as paper and pencil.

Although text-based programming languages and tools are traditionally dominant for programming, visual programming languages are also widely used, especially by children, to gain programming/coding skills. In visual programming languages and tools, programming elements are represented by graphical components rather than textual commands, allowing users to control the program flow by directly manipulating these elements. The most common visual programming tools are Scratch [1], [2], Alice [3], and Blockly [4]. Visual programming languages make the programming process more understandable, concrete and fun because they consist of visual elements and work with the familiar drag-and-drop method. Nevertheless, like traditional programming languages, they still require a computer during the programming phase. However, overuse of computers is known to cause problems such as screen addiction for children.

Alternatively, tangible programming languages and tools have been developed. Such languages allow users to manipulate various physical objects to control virtual objects or physical robots [5]. They can be referred as hybrid programming environments. In many applications of tangible programming such as Algoblocks [6], E-Block [7], and LEGO Mindstorms [8], the control objects contain some electronic and/or mechanical components, which increases the cost and limits the accessibility. Durable and low-cost solutions have also been proposed for the use of tangible programming languages in the classroom environment [9], [10]. These studies generally use low-level image processing techniques, which can lead to computational costs and recognition errors, or some robotic components.

The main objective of this work is to develop simple, comprehensible, accessible, and enjoyable material that will help children acquire algorithmic thinking and programming skills. We also avoid a fully plugged

approach, as the target audience is preschool children. To achieve this, the following contemporary methodologies underpin our approach:

- *Gamification*: It is known that one of the best teaching methods for young children is the gamification technique [11], [12]. The aim of employing gamification is to make the learning process more fun, instructive, and memorable.
- *Tangible programming*: A key concern for children’s physical and mental health is the overuse of computers. Parents generally limit their children’s screen time. The adoption of a tangible programming approach makes the process almost “unplugged”.
- *Augmented reality (AR)*: The benefits of using AR technology are manifold. First, it requires only a smartphone or tablet, making it a low-cost and accessible kit. Second, it enables a flexible tool that is content-renewable, as virtual objects can be placed in the real world. Thirdly, it contributes to entertainment and engagement [13]. Finally, it provides a robust and computationally inexpensive method for detecting code blocks, compared to using low-level image processing operations.

In the proposed application, a maze-like game environment is superimposed on the user’s real world using AR technology, and the game character is controlled by the player’s algorithm constructed using tangible command blocks.

A common problem with tangible AR applications is that there are no well-defined design standards as there are for traditional desktop UIs. Therefore, we take a design-centric view throughout the paper. For the benefit of researchers in the field, the design trade-offs, implementation decisions, heuristic evaluation results, limitations and possible improvements have been detailed.

The rest of the paper is organized as follows: The next section provides a summary of the related studies in the literature. The third section explains the tools and methodologies used to develop the proposed application. In the fourth section, the results of the heuristic evaluation are explained. Comparison of our approach to other similar approaches is performed. The limitations and future work are then described. The last section summarizes and concludes the study.

2. RELATED WORK

The use of AR technology in education is known to make the learning process more enjoyable and memorable [14], [15]. AR technology has been used effectively in many different fields of education [14]–[17]. The use of AR technology for teaching algorithms and programming is relatively new. Experimental studies have shown that AR technology significantly contributes to learning and entertainment factors [13], [18]. Today, tangible programming tools have begun to be combined with AR

technology [19]. For recent literature reviews about different applications of AR, serious games, and tangible interfaces in education, see the papers [20]–[22].

AR Scratch [23] provides features that extend the Scratch programming language to enable the development of simple AR applications. In the ARMaze [24] application, the visual codes on the cube-shaped programming blocks are recognized by an image processing library and the interpreted commands move the character around a maze using AR technology. Similarly, there are applications such as HyperCubes [25] and CodeCubes [18] where programming commands are represented by cubes.

Another example of tangible AR is the ARQuest [26] application, which enables team collaboration using a client-server architecture. In this application, users can design the game environment and present it to other teams. This application appeals to the 9-10 age group and does not contain advanced programming components such as loops and conditions.

The Code Bits [27] application consists of code blocks that are designed to be printed on paper to be affordable and accessible, as in our study. In the CodeBits application, there are no commands such as loops and conditions, and the commands are scanned and transferred to the game environment one at a time. Although this is good for debugging purposes, it can limit usability.

Command blocks in the Code Notes [28] application are in the form of cards and consist of English phrases such as “Turn position to the left, draw a tree” instead of visual codes. Text recognition techniques are used to interpret the commands in the cards. It is stated that the application is mainly aimed at the 12-13 age group. Similarly, the Kart-ON [29] application aims to provide an affordable, extensible, and expressive programming environment using paper programming and relying on text recognition.

The literature review shows that although there are studies on the use of tangible programming tools to teach algorithmic thinking and coding skills to children, these studies have not yet reached saturation point. Most of these tools involve microcontrollers and mechanical components and/or are paid for. This reduces accessibility due to cost. At this point, combining AR technology with tangible programming tools is an alternative. In these studies, programming commands generally consist of blocks in the form of cubes, and visual codes consist of abstract shapes, as they aim to facilitate recognition by the artificial vision system rather than user perception. In addition, current studies generally fail to explain the rationale behind design choices and provide limited insight for further research.

The key contributions of this study and the advantages of the proposed solution can be summarized as follows:

- An affordable and simple educational tool has been developed to improve children's algorithmic thinking skills.
- The primary user group is 5-10 year old children. The application and command blocks have been designed with the target user group in mind.
- The proposed solution offers a compromise between plugged and unplugged approaches to algorithmic thinking activities by combining AR and tangible programming.
- The tool contains no electronic or mechanical components. This makes it cheap and safe.
- The solution does not rely on low-level image processing techniques that can degrade the system's recognition accuracy and interactivity.
- It has a broader command gamut compared to similar tools, including conditions and loops.
- The game environment is presented in 3D and is attached to the real world using AR technology to enhance realism and engagement.
- The results of the usability analysis, design trade-offs, and implementation decisions are explained to help other researchers and designers.

3. METHODOLOGY

This section first describes the use of the proposed kit. It then summarizes the tools and techniques used to develop the proposed system. This is followed by the details of the rationale behind the design and implementation choices, together with possible design trade-offs.

3.1. Overview

In the proposed application, there is a 3D game environment, such as a maze, and the game character must be controlled to move to the target location. The steps performed by the user are shown in Figure 1, and an example view during gameplay is shown in Figure 2.

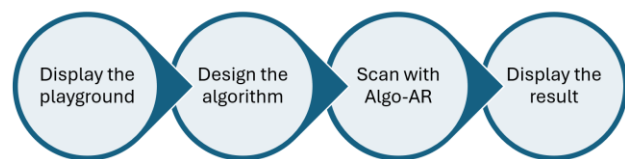


Figure 1. Steps of the usage of the application

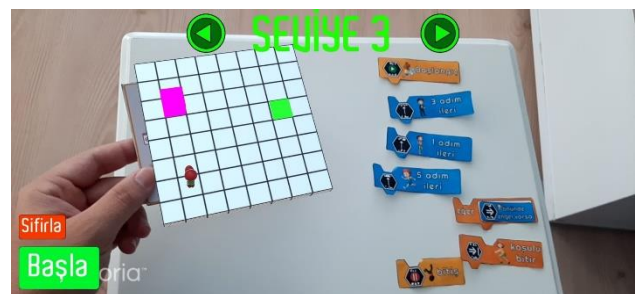


Figure 2. A sample view during gameplay

First, the player inspects the problem on the virtual grid that is displayed on top of the physical gaming platform using AR technology. The player then places specially designed command blocks as physical jigsaw puzzle pieces to create the required algorithm. These physical command blocks are scanned and converted into code using the proposed mobile AR application, and the game character moves in the digital environment according to these codes. With the help of AR technology, the playground is displayed in the user's physical environment, increasing fun, motivation, and engagement.

3.2. Command Blocks

The first step was to determine the commands that would be used to control the character's movements. The most necessary commands have been implemented for this prototype, but it is planned to increase the variety of commands in future developments. The commands in the current prototype are: *Start (Başlangıç)*, *End (bitiş)*, *Move n steps (n adım ileri)*, *Turn right/left (sağa/sola dön)*, *Repeat n times (n defa tekrarla)*, and *If (eğer)*. Therefore, the current prototype involves the most important concepts of programming, including loops and conditions. The users organize the command blocks to build their algorithms. The algorithms must be placed between the *start* and *end* commands. The same background colour is used for the complementary commands (i.e. start-end). The visual codes to be used for the recognition of the blocks were generated using the VuMark tool¹. These command blocks are shown in Figure 3 and printable command blocks are available in the Appendix.



Figure 3. Command blocks

3.3. Mobile AR Application

A mobile AR application called Algo-AR was developed using the Unity² game engine and the Vuforia SDK³. The playground was designed as a 3D grid-like structure in which some cells may have obstacles. The game has several levels of increasing difficulty. Using command blocks, the player is expected to create an algorithm that will transport the game character from its current location to the target location without hitting obstacles. The

installation file of the application and sample videos on how to use the application can be found in the Appendix.

Figure 4 shows an example of the process of converting command blocks into game control. After the users physically create the necessary algorithm using the command blocks, they scan this algorithm with the camera through the Algo-AR application. For the recognition of the command blocks, we use marker-based AR technology. In marker-based AR, where the digital content will be placed is pre-defined to the system with a marker. For this purpose, the image targets of the command blocks and the corresponding text codes were matched using the Vuforia SDK. Target images are recognized by the detector of the Vuforia platform. The recognized text codes corresponding to the image targets are converted by the application into C# functions that provide the necessary parameters for the character's movement, and these functions are executed in the game by the Unity game engine.

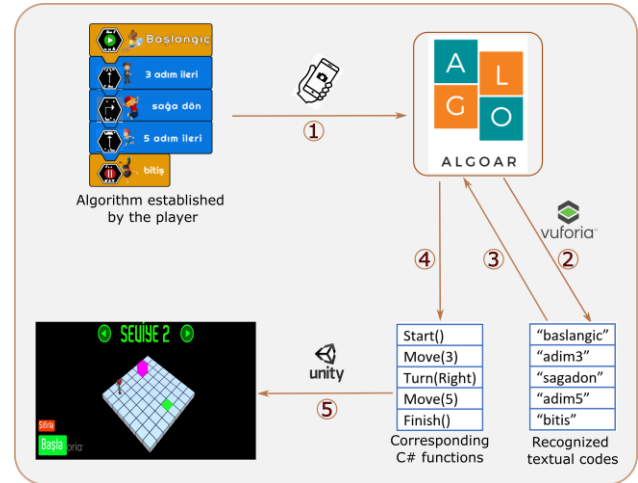


Figure 4. The process of converting the command blocks to game control

Definition of the Game Loop:

```
Start
  L = SelectLevel()
  InspectPuzzle(L)

  Algorithm = ConstructAlgorithm()
  RunAlgorithm(Algorithm)

  DisplaySolution()
End
```

```
Function ConstructAlgorithm:
  Start
    PuzzlePieces = DetectPuzzlePiecesFromCamera(Puzzle)

    Display(PuzzlePieces)

    CommandList = AddCommandToList(PuzzlePieces)

  Return CommandList
End
```

Figure 5. Pseudocode describing the game loop process.

¹ <https://library.vuforia.com/objects/vumarks>

² <https://unity.com/>

³ <https://developer.vuforia.com/>

In Figure 5, a description of the game process is given in a pseudocode format. When the game starts, the player selects the level, inspects the puzzle, constructs the algorithm using the blocks then scans the algorithm with the camera. Detection of the puzzle pieces by the system and the construction of the command list from the detected pieces is also given in the second part (Function ConstructAlgorithm) of the pseudocode. The returned algorithm is executed on the game board and the resulting animation is displayed.

3.4. Design Rationale

This section explains critical implementation issues and major design goals that were considered in the development of the current prototype.

- 1) *Cost*: The proposed tool requires only a smartphone or a tablet and command blocks that can be printed on a piece of paper. This makes it an inexpensive kit.
- 2) *Accessibility*: As a direct consequence of the low cost, this tool is easily accessible to people on low incomes. In addition, since visual codes are used, illiterate children can also use this material. Although the primary audience of the application is children, it is suitable for users of all ages who are new to algorithmic skills.
- 3) *Durability*: The proposed tool does not require complex and fragile mechanical or electrical components as many popular tangible programming tools do. Thus, there is no problem with the breakdown of these components. In the worst-case scenario, if the printed command blocks are torn, they can be reprinted. It is also possible to use durable materials such as wood or plastic for the command blocks. Although this will increase the cost, it will enhance the overall user experience.
- 4) *Safety*: As the target user group is children, special attention must be paid to the safety of the material. As mentioned above, the kit does not contain any electrical components, so it is quite safe to use.
- 5) *Engagement*: Fun is an important factor in increasing people's motivation and learning rate. The more people feel involved in the application, the more they are motivated to learn. To increase fun and motivation, the application is designed as a game with levels of increasing difficulty. Additionally, the game environment and the characters are designed in 3D, which also contributes to engagement. Nevertheless, more visual/audio effects and game levels are required to raise the fun factor of the game. In other respects, the usage of the application requires a focus switch between physical command blocks and the virtual content on the screen, which may damage engagement.
- 6) *Unplugged usage*: Most similar applications aimed at teaching children to program or design algorithms require coding with a computer or mobile phone. Such applications can be referred to as “plugged-in” tools. However, social/mental/physiological problems or digital addiction may arise if too much time is spent with digital devices, especially at a young age. Completely “unplugged” tools are also available to enrich computational thinking without computers [30], [31]. We offer a compromise between plugged and unplugged tools by combining AR and tangible programming. Since the algorithm construction is done using physical command blocks, the plugged usage time is very short. When the game environment is projected on a shared display, parents or teachers can control the application with their smartphones. This can be considered unplugged use for children. This allows for cooperation as well as competition. Therefore, depending on the age of the users, the usage process can be adjusted as plugged or unplugged. This is another flexibility of the proposed tool.
- 7) *Design of the command blocks*: An image target was created to represent each command block. When creating image targets, it is important to have different images for different commands and unique details so that the application can easily recognize the images. However, completely abstract shapes such as QR codes were avoided so as not to complicate children's perceptions. Instead, black-and-white image targets were created using the VuMark tool to identify the commands. Additional visual images were also added for human perception. Although the image targets uniquely identify the commands, text equivalents for each command were also included in the command blocks. Turkish was preferred as the text language for the presentations, but in practice, no modification is required to change the text language. For this reason, command blocks can be used in any language, even without text.
- 8) *Accuracy and efficiency*: Instead of training our models with low-level image processing and computer vision techniques, the marker-based AR method was found to be suitable for recognising the command blocks. This is because in this problem, unlike a standard object recognition problem, the images to be recognized are specified and fixed at the beginning. In addition, due to the variable environmental conditions in mobile environments and the real-time interactivity requirement of the application, a robust solution with low computational cost is needed. Therefore, employing marker-based AR techniques was found to be the most appropriate solution.

3.5. Design Trade-offs

In this section, we summarize the major trade-offs that we encountered during the design and implementation of the proposed tool. In some of these situations, we force the user to meet a specific design goal, while some trade-offs are flexible and left to the user's control.

- 1) *Cost/affordability vs. usability*: Command blocks can be printed on a piece of paper for cost-effectiveness,

but from a usability perspective it is easier to use other firm but more expensive materials such as wood or plastic. At this point, the application does not have a restriction and the user has control over this choice. Another issue related to this trade-off is the use of mobile AR or AR glasses. While AR glasses allow for hands-free use, they are not very accessible due to their cost. For this reason, we have traded off cost against usability in this question.

- 2) *Cost/affordability vs. durability*: As mentioned earlier, the user can choose the material for the command blocks to be printed. This can be a piece of paper (for low cost) or 3D printed using other durable materials.
- 3) *Error prevention (affordance) vs. usability*: The Jigsaw puzzle-like form of the command blocks caters to affordance and hence error prevention. However, this may limit usability somewhat when paper is the material for the command blocks. The tool still works even if the user cannot place the blocks correctly concerning the interlocking mechanism. However, the user is in control of the choice for this compromise and may prefer to use a different material for a better user experience.
- 4) *Simplicity vs. expressiveness*: A wider range of commands, including other programming concepts, would provide a more expressive tool. However, given the age of the target audience, we preferred to keep it simple and understandable. Similarly, the number of command blocks in a particular algorithm is deliberately limited to simplify comprehension and avoid visual clutter on the screen. This was also suggested by the usability experts who evaluated our prototype.
- 5) *Accessibility vs. expressiveness*: Several alternatives were considered for the design of the command blocks. Firstly, a text-based approach was considered to provide an expressive and extensible tool, but this method was abandoned bearing in mind the illiterate children and the method's dependence on the text language. The idea of assigning distinctive images to command blocks was found to be more appropriate and non-abstract visuals were determined for children.
- 6) *Accuracy vs. usability*: For the scanning process of the algorithm, we contemplated three options: *i*) scanning the whole algorithm and playing its result on the game at once, *ii*) scanning and playing the blocks one by one (the idea in [27]), or *iii*) scanning the blocks one by one and playing the whole recognized algorithm at once. Scanning the whole algorithm (option *i*) would be more usable and intuitive, as it was also suggested to us by a heuristic evaluator, but it requires more computation and low-level image processing operations. This could reduce the recognition accuracy and efficiency. Therefore we preferred to scan the blocks sequentially one at a time. Nevertheless, since

the second option limits usability more than the last, the current prototype supports only sequential scanning and full replay (option *iii*). However, we plan to evaluate these options from both usability and efficiency perspectives and update the next prototype of the application accordingly.

4. HEURISTIC EVALUATION

To assess the usability of the proposed system for the target user group and to identify design problems at an early stage, we carried out a heuristic evaluation process, which is a usability testing method in which experts assess the usability of the system against a set of pre-defined criteria. It can be argued that heuristic evaluation is limited in the sense that the evaluators are not part of the target audience. However, it is a very powerful tool because it provides a cost-effective way of detecting design problems early in the development process.

4.1. Evaluation Methodology

Three usability experts independently assessed the proposed application according to Nielsen's usability heuristics [32]. They benefitted from the Heuristic Evaluation Workbook provided by Nielsen Norman Group⁴. They are informed about the purpose and target users of the application. In their evaluations, they were asked to consider children aged 5-10, especially preschoolers, as the target user group. The evaluators first inspected the application freely, then they were shown the main functionality and finally, they played the game levels shown in Figure 6. They used command blocks printed on paper but were also informed about the possibility of using other materials.



Figure 6. Top – sample game levels, and bottom – their solutions

4.2. Findings

The evaluators' findings are explained below according to the predefined usability criteria. These findings are also summarized in Table 1. This table lists the key strengths of the tool and recommended improvements for each usability criterion.

- 1) *Visibility of system status* (“The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time.”)

When the application recognizes a command block, visual and auditory feedback is presented to the user, and the recognized algorithm is displayed online. In Figure 7, the algorithm constructed by the player using physical command blocks is shown on the left. On the right, the recognized commands are displayed virtually as they are scanned.



Figure 7. A screenshot from the gameplay, while scanning the constructed algorithm

- 2) *Match between the system and the real world* (“The system should resemble the experiences that users already had. The design should speak the users' language. Use words, phrases, and concepts familiar to the user, rather than internal jargon. Follow real-world conventions, making information appear in a natural and logical order.”)

The command blocks are designed as interlocking jigsaw puzzle pieces and are used based on the intuitive drag-and-drop metaphor. This is consistent with real-life puzzle games and popular visual programming languages. It also prepares children for block-based programming tools, which are generally the next level in the learning process. The use of the same colour for the complementary command blocks (i.e. start-end) is also a good choice. Although the images on the command blocks generally match universal visual codes such as arrows, play buttons, numbers, etc., the visual code for the *if* command block is somewhat ambiguous.

- 3) *User control and freedom* (“Users should be able to reverse their action if done by mistake.”)

The application provides a visible “Sıfırla (Reset)” button to reset the algorithm. The choice of red colour for this button is considered appropriate as it is a dangerous operation that requires attention. However, an Undo operation should also be provided to allow users to reverse their actions taken in error.

- 4) *Consistency and standard* (“Follow platform and industry conventions. Similar system elements should look similar.”)

AR interfaces and tangible user interfaces are relatively new fields. As a result, there are no widely accepted industry standards as there are for 2D interfaces. On the other hand, the puzzle-like form of the command blocks is consistent with common visual programming conventions.

- 5) *Error prevention* (“Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.”)

Another advantage of the jigsaw puzzle form of the command blocks is that it provides affordance and thus prevents possible syntax errors. Furthermore, the feedback (see item 1) made available to the user when a command block is recognized by the application prevents potential errors.

- 6) *Recognition rather than recall* (“Minimize the user's memory load by making elements, actions, and options visible. The user should not have to remember information from one part of the interface to another. Information required to use the design should be visible or easily retrievable when needed.”)

The application is easy to learn because the command blocks are designed like puzzle pieces that everyone is familiar with from everyday life or common visual programming languages. The simple design of the tool contributes to learnability as well. Expressive and memorable visual codes have generally been assigned to the command blocks. Text codes are also helpful in identifying the meaning of a block, but they are helpless for illiterate children. The control buttons, “Başla (Start)” and “Sıfırla (Reset)”, should also have familiar icons rather than text, taking into account illiteracy. These buttons are designed to be visible, but in the context of AR, background and lighting conditions may affect visibility. Similarly, the nature of AR requires switching focus between real and virtual content, which can generate additional cognitive load.

- 7) *Flexibility and efficiency of use* (“both new and experienced users should be able to efficiently use the system”)

The target user group consists of young children and the application does not generally require shortcuts for experienced users. On the other hand, the developed tool

offers some flexible options that users can customize according to their needs. First, the user can print command blocks on a piece of paper or any other material. Second, the language of the expressions in the command blocks can be changed or completely removed. Third, although the command blocks were originally designed as puzzle pieces, they can also be used flat, as it can be difficult to use puzzle pieces when paper command blocks are preferred.

One evaluator made the following suggestions in addition to the common opinions listed above. The application was designed for mobile AR, which requires holding the mobile device. Adapting it for AR glasses is also suggested to allow hands-free and comfortable use. Scanning the whole algorithm at once could be more efficient and usable.

8) *An aesthetic and minimalist design* (“Interfaces should not contain information that is irrelevant or rarely needed.”)

The interface and the system have been designed in a fairly simple and minimalist manner. However, the game grid and the look and feel of the UI elements could be more aesthetically pleasing. Limiting the overall command gamut is a sound decision, given the age of the target audience and the short-term memory limitations of the human brain. Similarly, the total number of command blocks required to construct a particular algorithm should also be limited. Otherwise, long algorithms will result in visual clutter on the screen, and the scanning of the algorithm will become more difficult as the length of the algorithm increases.

9) *Help users recognize, diagnose, and recover from errors* (“Error messages should be expressed in plain language, precisely indicate the problem, and constructively suggest a solution.”)

Although it is very rare, if the user inadvertently moves the mobile device during the sequential scanning process, the application may fail to recognize some command blocks. In such cases, the user is informed of the situation as the recognized blocks are displayed on the screen. However, the only way to recover from this situation is to reset the algorithm. An Undo operation must also be made available to the user. In addition, the errors in the constructed algorithm could be shown to the user.

10) *Help and documentation:*

More detailed documentation, including the meaning of the blocks, sample usage scenarios, etc. could be provided. Online textual instructions on how to use the system should also be provided as audio messages, taking into account the illiterate population in the target user group.

Table 1. Summary of the heuristic evaluation (C1-C10 refers to the evaluation criteria, in the above order)

	<i>Strengths</i>	<i>Recommendations</i>
C1	Visual/auditory feedback	
C2	Intuitiveness Use of metaphors	Revision of the visual codes
C3	Visible control buttons	Undo operation
C4	Puzzle-like form Drag-and-drop metaphor	
C5	Affordance Feedback	
C6	Easy to learn Familiar elements Simplicity contributes to learnability Memorable visual codes	Audio messages for illiterate children Familiar icons for control buttons
C7	Flexible options for command blocks	Adaptation for AR glasses Scanning the algorithm at once
C8	Simple and minimalist design Limited command gamut	Refining the look-and-feel
C9		Undo operation Showing the algorithm errors
C10		More detailed documentation Audial instructions

5. COMPARISON TO OTHER APPROACHES

In this section, we perform a qualitative comparison of our kit with other alternative tools for cultivating computational thinking. We make this comparison in the context of our target user group (preschoolers) and our main design goals: “unplugged use”, “simplicity”, and “accessibility/affordability”.

In Table 2, interface types and main methodologies of similar studies are listed. The interface type refers to the usage of a tool and we categorize it as *Unplugged*, *Plugged*, and *Hybrid*. Unplugged interfaces rely on traditional methods and do not use any computers, plugged interfaces require a computer and/or electronic components, while hybrid interfaces require computers only for specific stages. Algo-AR falls into the hybrid interface category by combining tangible programming and AR as mentioned before, since a mobile phone is only required for displaying the results. Limiting screen time is one of our crucial goals because of the target users. Compared to totally unplugged activities, AR increases the fun factor.

Table 3 evaluates similar studies in the literature in the context of the design goal of simplicity. Preschool children are the target user group, therefore the design, usage and comprised programming concepts should be simple. As indicated in the table, some of the tools are text-based and not suitable for illiterate children, some contain complex programming concepts, and some are sophisticated for little children. Algo-AR has a simple interface and limits the complexity of algorithms and programming concepts.

Accessibility and affordability are also among our primary design objectives. Therefore, the studies in the literature are also elaborated from this perspective and the summary

is given in Table 4. Unplugged activities can be considered as the most accessible category. There are also free visual programming tools that only require a computer. Some of the tools, such as Algoblocks and E-Block, necessitate special hardware for the command cubes, which limits their accessibility. On the other hand, most of the tools in the hybrid interface category, rely on a mobile device and paper-based command blocks. This provides a low-cost and accessible solution.

Table 2. List of the studies according to interface type and methodologies

Studies	Interface	Tools/Methodologies
[30] [31]	Unplugged	Paper, pen, etc.
Scratch [1], [2] Alice [3] Blockly [4]	Plugged	Visual programming
Algoblocks [6]	Plugged	Tangible programming
E-Block [7]	Plugged	Tangible programming, Microcontrollers
HyperCubes [25]	Hybrid	Tangible programming, AR, Spatial tracking
CodeCubes [18]	Hybrid	Tangible programming, Marker-based AR
ARQuest [26]	Hybrid	Tangible programming, Marker-based AR
Code Bits [27]	Hybrid	Tangible programming, Marker-based AR
Code Notes [28]	Hybrid	Tangible programming, Computer vision, Text recognition
Kart-ON [29]	Hybrid	Tangible programming, AR, Text recognition
Algo-AR	Hybrid	Tangible programming, Marker-based AR

Table 3. List of the studies according to simplicity design goal

Studies	Simplicity
[30] [31]	Simple to use, appeals to all ages of children
Scratch [1], [2] Alice [3] Blockly [4]	Simple drag-and-drop metaphor Not for illiterate children in general Comprises many programming concepts Not suitable for preschoolers
Algoblocks [6]	Appeals to primary and secondary school Includes conditionals, loops, parameters, and basic movement commands Requires command cubes connected with cables which limit usability
E-Block [7]	Appeals to 5-9 age Includes basic movement commands
HyperCubes [25]	Appeals to late elementary and middle school Includes visual and sound inputs Requires constructing command cubes Complex for preschoolers since it includes abstract markers and parameters are adjusted by a menu
CodeCubes [18]	Appeals to 13-14 age Includes basic movement commands Requires constructing command cubes
ARQuest [26]	Appeals to primary school (9-10 age) Includes basic movement commands Designed for collaborative usage
Code Bits [27]	Preschoolers can use Includes basic movement commands Commands are transferred to the application one-by-one
Code Notes [28]	Based on English text Not suitable for illiterate children Includes a wide range of commands

Kart-ON [29]	Based on English text Not suitable for illiterate children Includes a wide range of generalizable commands
Algo-AR	Simple command gamut (loops, conditionals, basic movement) Simple algorithm complexity Simple drag-and-drop metaphor Suitable for illiterate and preschool children

Table 4. List of the studies according to accessibility design goal

Studies	Accessibility/Affordability
[30] [31]	Requires basic materials such as paper, pen So, very cheap and accessible in general
Scratch [1], [2] Alice [3] Blockly [4]	Requires a computer Freely accessible in general
Algoblocks [6]	Requires a computer and special wired command cubes, so not much accessible
E-Block [7]	Requires a computer and command blocks that include microcomputers, infrared transmitters and receivers, batteries, wireless modules, and LEDs. So they are not much accessible
HyperCubes [25]	Requires a mobile device and paper cubes
CodeCubes [18]	Requires a mobile device and paper cubes
ARQuest [26]	Requires a computer, a mobile device and paper tokens
Code Bits [27]	Requires a mobile device and paper tokens
Code Notes [28]	Requires a mobile device and paper tokens
Kart-ON [29]	Requires a mobile device and paper tokens
Algo-AR	Requires a mobile device and paper blocks

6. LIMITATIONS AND FUTURE WORK

Despite its many benefits, the current prototype has some shortfalls and potential for improvement. These improvements can be considered under five main categories: refining the usability aspects, reinforcing the pedagogical aspect, enriching the content, increasing the supportability, and conducting comprehensive evaluations.

Firstly, the user interface and usability of the tool should be improved in light of the results of the heuristic evaluation. These improvements should include refining the visual codes on the command blocks in terms of expressiveness and discoverability, adding an undo operation, considering illiteracy in the design of the UI elements, and improving the documentation.

Second, the current prototype supports programming concepts such as condition and loop that are not found in many similar applications. Considering that people will not develop very complex algorithms during the learning phase, it can be said that the current application has functional competence in terms of targeted acquisition of algorithmic thinking. However, more programming elements such as arithmetic operations, functions and variable definitions can be added. Correct solutions and errors should be displayed on the application. It could also be useful to enable the teacher/parent to track the progress of the students.

Third, more levels should be added to the game to increase the fun and motivational factors. The following improvements can be considered, which will also affect the difficulty level of the game: time limit, types of obstacles, bonus items, and shortest path problems. The application should be made more impressive and customizable by enriching it with more animations, visual effects, sound/music, and virtual environment/character themes. Game themes and difficulty levels should be determined according to age groups, taking into account the views of educators and pedagogues.

Fourth, the current prototype works on Android devices, but it is possible to easily deploy it for other operating systems. Its use with AR glasses, however, should be specifically tested and the application should be adapted to the needs of such use. Since one of the main design goals is affordability, we have not yet considered the use of AR glasses.

Finally, although the usability aspect has been assessed by some experts, the tool has not yet been tested by the target user group. Thus, comprehensive user studies are required to evaluate different aspects of the user experience and to draw generalizable conclusions. The application should be extensively tested with different age groups and its long-term effects in different dimensions (learning, entertainment, motivation, cognition, etc.) should be statistically analysed.

7. CONCLUSION

A prototype application has been developed to improve children's algorithmic thinking skills. The main design goals are simplicity, affordability, entertainment, and unplugged usage. The application has a gamified structure and is intended to be particularly interesting for children. Although the primary audience of the application is children, it is suitable for users of all ages who are just starting to learn algorithm construction. The proposed application has been developed using modern technologies such as AR and tangible programming. A design-oriented view of the work could be of great benefit to the researchers in the field. In summary, a low-cost, flexible, unplugged educational material for the development of algorithmic thinking skills at an early age is proposed and elaborated.

APPENDIX

Materials related to this study can be downloaded from the links obtained by scanning the QR codes in Figure 8. The downloadable materials include sample videos demonstrating the use of the game, the installation file of the Algo-AR application for Android devices, and the printable command blocks.



Figure 8. QR-codes for accessing the sample material of the proposed tool

REFERENCES

- [1] J. Fagerlund, P. Häkkinen, M. Vesisenaho, J. Viiri, "Computational thinking in programming with Scratch in primary schools: A systematic review," *Comput. Appl. Eng. Educ.*, 29(1), 12–28, 2021.
- [2] J. Maloney, L. Burd, Y. Kafai, N. Rusk, B. Silverman, M. Resnick, "Scratch: a sneak preview [education]," **Second International Conference on Creating, Connecting and Collaborating through Computing**, Kyoto, 104–109, 2004.
- [3] M. Conway, S. Audia, T. Burnette, D. Cosgrove, K. Christiansen, "Alice: lessons learned from building a 3D system for novices," **SIGCHI Conference on Human Factors in Computing Systems**, Netherlands, 486–493, 2000.
- [4] N. Fraser, "Ten things we've learned from Blockly," **IEEE Blocks and Beyond Workshop (Blocks and Beyond)**, USA, 49–50, 2015.
- [5] A. Strawhacker, M. U. Bers, "'I want my robot to look for food': Comparing Kindergartner's programming comprehension using tangible, graphic, and hybrid user interfaces," *Int. J. Technol. Des. Educ.*, 25, 293–319, 2015.
- [6] H. Suzuki, H. Kato, "Algoblock: a tangible programming language, a tool for collaborative learning," **4th European Logo Conference**, 297–303, 1993.
- [7] D. Wang, Y. Zhang, S. Chen, "E-block: A tangible programming tool with graphical blocks," *Math. Probl. Eng.*, 2013, doi: 10.1155/2013/598547.
- [8] F. Klassner, S. D. Anderson, "LEGO MindStorms: Not just for K-12 anymore," *IEEE Robotics and Automation Magazine*, 10(2), 12–18, 2003, doi: 10.1109/MRA.2003.1213611.
- [9] M. S. Horn, R. J. K. Jacob, "Designing tangible programming languages for classroom use," **International Conference on Tangible and Embedded Interaction**, Louisiana, 159–162, 2007, doi: 10.1145/1226969.1227003.
- [10] E. Naude, A. Fowler, R. Lemon, C. J. Sutherland, "Kupe's Journey: Building a Low-cost, Screen-free Robotic Programming Environment for Children," **20th International Conference on Ubiquitous Robots (UR)**, USA, 710–715, 2023, doi: 10.1109/UR57808.2023.10202226.
- [11] S. Tobias, J. D. Fletcher, and A. P. Wind, "Game-based learning," *Handb. Res. Educ. Commun. Technol.*, pp. 485–503, 2014.
- [12] G. Lampropoulos, E. Keramopoulos, K. Diamantaras, G. Evangelidis, "Augmented reality and gamification in education: A systematic literature review of research, applications, and empirical studies," *Appl. Sci.*, 12(13), 6809, 2022.

- [13] A. Gardeli, S. Vosinakis, "The effect of tangible augmented reality interfaces on teaching computational thinking: A preliminary study," **21st International Conference on Interactive Collaborative Learning (ICL2018)**, Greece, 673-684, 2020.
- [14] A. Theodoropoulos, G. Lepouras, "Augmented Reality and programming education: A systematic review," *Int. J. Child-Computer Interact.*, 30, 100335, 2021.
- [15] S. A. Hassan, T. Rahim, S. Y. Shin, "ChildAR: an augmented reality-based interactive game for assisting children in their education," *Univers. Access Inf. Soc.*, 21(2), 545-556, 2022.
- [16] Y.-C. Chien, Y.-N. Su, T.-T. Wu, Y.-M. Huang, "Enhancing students' botanical learning by using augmented reality," *Univers. Access Inf. Soc.*, 18, 231-241, 2019.
- [17] Z. Çipiloğlu Yıldız, M. Türker, R. Ak, "Mimari Miras Eğitiminde Artırılmış Gerçeklik ve Fotogrametri Desteği," *Bilişim Teknol. Derg.*, 14(2), 137-149, 2021, doi: 10.17671/gazibtd.792539.
- [18] B. Cleto, C. Sylla, L. Ferreira, J. M. Moura, "CodeCubes: Coding with Augmented Reality," **First international computer programming education conference**, Portugal, 7:1-7:9, 2020, doi: 10.4230/OASICS.ICPEC.2020.7.
- [19] S. Washbrooke, N. Giacaman, "Play, Code, Learn: Fostering Computational Thinking in Primary Aged Learners Through Interactive Play," **IoT, AI, and ICT for Educational Applications: Technologies to Enable Education for All**, Ed.: S. Papadakis, Cham: Springer Nature, Switzerland, 135-162, 2024.
- [20] M. G. Rios, M. Paredes-Velasco, "Using Augmented Reality in programming learning: A systematic mapping study," *IEEE Global Engineering Education Conference (EDUCON)*, Austria, 1635-1641, 2021, doi: 10.1109/EDUCON46332.2021.9454149.
- [21] M. Liang, Y. Li, T. Weber, H. Hussmann, "Tangible interaction for children's creative learning: A review," *Conference on Creativity and Cognition*, 1-14, 2021.
- [22] J. M. Cerqueira, B. Cleto, J. M. Moura, C. Sylla, L. Ferreira, "Potentiating Learning Through Augmented Reality and Serious Games," *Springer Handbook of Augmented Reality*, Eds.: A. Y. C. Nee, S. K. Ong, Cham: Springer International Publishing, 369-390, 2023.
- [23] I. Radu, B. MacIntyre, "Augmented-reality scratch: a tangible programming environment for children," *Conference on Interaction Design for Children*, Italy, 2009.
- [24] Q. Jin, D. Wang, X. Deng, N. Zheng, S. Chiu, "AR-maze: A tangible programming tool for children based on AR technology," *ACM Conference on Interaction Design and Children*, Norway, 611-616, 2018, doi: 10.1145/3202185.3210784.
- [25] A. Fuste, C. Schmandt, "Hypercubes: A playful introduction to computational thinking in augmented reality," *CHI PLAY 2019 - Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play*, Spain, 379-387, 2019, doi: 10.1145/3341215.3356264.
- [26] A. Gardeli, S. Vosinakis, "ARQuest: A tangible augmented reality approach to developing computational thinking skills," *International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, Austria, 1-8, 2019, doi: 10.1109/VS-Games.2019.8864603.
- [27] S. Goyal, R. S. Vijay, C. Monga, P. Kalita, "Code bits: an inexpensive tangible computational thinking toolkit for K-12 curriculum," *International Conference on Tangible, Embedded, and Embodied Interaction*, Netherlands, 441-447, 2016.
- [28] A. Sabuncuoğlu, M. Erkaya, O. T. Buruk, T. Göksun, "Code notes: Designing a low-cost tangible coding tool for/with children," *ACM Conf. Interact. Des. Child.*, Norway, 644-649, 2018, doi: 10.1145/3202185.3210791.
- [29] A. Sabuncuoğlu, T. M. Sezgin, "Kart-ON: An Extensible Paper Programming Strategy for Affordable Early Programming Education," *ACM Human-Computer Interact.*, 6(EICS), 1-18, 2022.
- [30] T. Bell, J. Alexander, I. Freeman, M. Grimley, "Computer science unplugged: School students doing real computing without computers," *New Zeal. J. Appl. Comput. Inf. Technol.*, 13(1), 20-29, 2009.
- [31] A. Juškevičienė, G. Stupuriene, T. Jevsikova, "Computational thinking development through physical computing activities in STEAM education," *Comput. Appl. Eng. Educ.*, 29(1), 175-190, 2021.
- [32] J. Nielsen, "Enhancing the explanatory power of usability heuristics," *SIGCHI conference on Human Factors in Computing Systems*, USA, 152-158, 1994.

Saldırı Tespit Sistemlerinde K-Means Algoritması ve Silhouette Metriği ile Optimum Küme Sayısının Belirlenmesi

Araştırma Makalesi/Research Article

 Fatih TOPALOĞLU

Bilgisayar Mühendisliği, Malatya Turgut Özal Üniversitesi, Malatya, Türkiye
fatih.topaloglu@ozal.edu.tr

(Geliş/Received:31.12.2023; Kabul/Accepted:12.02.2024)

DOI: 10.17671/gazibtd.1412641

Özet—Günümüz internetleri neredeyse yarım milyon farklı ağdan oluşmaktadır. Bir ağ bağlantısında, saldırıları türlerine göre tanımlamak zordur. Çünkü farklı saldırılar çeşitli bağlantılara sahip olabilir ve sayıları birkaç ağ bağlantısından yüzlerce ağ bağlantısına kadar değişebilmektedir. Bu nedenden dolayı saldırı tespiti için kullanılan veri setlerinin doğru sınıflandırılması zorlaşmaktadır. Geçmişte pek çok araştırmacı, farklı yöntemler kullanarak davetsiz misafirleri tespit etmek için saldırı tespit sistemleri geliştirmiştir. Ancak mevcut yöntemlerin tespit doğruluğu ve zaman kaybı açısından bazı dezavantajları bulunmaktadır. Çalışmanın temel motivasyonu, saldırı tespit sistemlerinde yüksek boyutluluğun getirdiği zorlukların üstesinden gelmek ve sınıflandırma performansını geliştirmek, sonuçta izinsiz girişlerin daha doğru ve verimli tespitini sağlamaktır. Çalışmada KDD Cup'99 saldırı tespiti veri setinin k-means kümeleme algoritması ile farklı k değerlerine göre analiz edilmesi ve silhouette metriği ile optimum küme sayısının belirlenmesi amaçlanmıştır. Çalışmada farklı k değerleri için yapılan analizlerde, $k=10$ 'a kadar olası her konfigürasyon için silhouette skoru hesaplanmıştır. Bu metriğe göre en iyi küme sayısı 4 ve silhouette skoru 0.83 olarak bulunmuştur. Ayrıca silhouette grafiği kalınlıkları ile küme boyutları görselleştirilmiştir.

Anahtar Kelimeler— saldırı tespit sistemleri, k-means, silhouette metriği

Determining the Optimum Number of Clusters with K-Means Algorithm and Silhouette Metric in Intrusion Detection Systems

Abstract—Today's internet consists of almost half a million different networks. In a network connection, it is difficult to identify attacks by type. Because different attacks can have various connections and their number can vary from a few network connections to hundreds of network connections. For this reason, it becomes difficult to correctly classify the data sets used for attack detection. The main motivation of the study is to overcome the challenges of high dimensionality in intrusion detection systems and improve classification performance, ultimately providing more accurate and efficient detection of intrusions. In the past, many researchers have developed intrusion detection systems to detect intruders using different methods. However, existing methods have some disadvantages in terms of detection accuracy and time loss. In the study, it was aimed to analyze the KDD Cup'99 attack detection data set according to different k values with the k-means clustering algorithm and to determine the optimum number of clusters with the silhouette metric. In the analysis carried out for different k values in the study, the silhouette score was calculated for each possible configuration up to $k = 10$. According to this metric, the best number of clusters was found to be 4 and the silhouette score was 0.83. Additionally, silhouette graphic thicknesses and cluster sizes are visualized.

Keywords— intrusion detection system, k-means, silhouette metric

1. GİRİŞ (INTRODUCTION)

Bilgisayar ağı sistemlerinde kurumsal veya kişisel bilgi güvenliğini sağlamak amacıyla birçok araç veya yazılım kullanılmaktadır [1]. Bu amaçla kullanılan en önemli araçlardan biri saldırı tespit sistemleridir. Saldırı tespit sistemleri, izinsiz girişleri önlemek ve bilgisayar sistemlerine yasa dışı erişimi engellemek için tasarlanmıştır. Saldırı tespit sistemleri, bir kuruluşun bilgisayar ağındaki içsel ve dışsal saldırıları sınıflandırabilir ve güvenlik ihlali varsa alarmı tetikleyebilir [2]. İzinsiz giriş tespit sistemlerinin temel amacı, şimdiye kadar tanımlanmış veya tanımlanamayan saldırılar olmak üzere izinsiz girişleri tanımak, bu saldırıları keşfetmek, bunlara uyum sağlamak ve izinsiz girişleri hızlı bir şekilde tespit etmektir [3].

Saldırı tespit sistemleri için kesinlik ve kararlılık iki önemli metriktir [4] ve son yıllarda, bu önlemleri geliştirmek için birçok çalışma yapılmıştır [5]. Başlangıçtaki çalışmaların çoğu kural tabanlı uzman sistem veya istatistiksel yaklaşıma odaklanmıştır. Ancak çeşitli performans sonuçları, bu yaklaşımların büyük veri kümelerine uygulandığında doğru ve kesin olmadığını göstermektedir [6].

Bu sorunu çözmek için veri madenciliği yaklaşımları [7,8] ve makine öğrenme teknikleri tanıtıldı [9]. Grafik tabanlı yöntemler [10], Doğrusal Genetik Programlama [11], Bayes Ağı [12], k- NN [13], K- Means kümeleme [14], Gizli Markov Modeli [15] vb. saldırı tespit sistemi mimarisi için araştırılmıştır. Makine öğrenimi [16], veri setinde bulunan özellikler ve sınıflar arasındaki korelasyonu tespit edebilir ve özellik seçimi ve boyut azaltma yoluyla alt kümeleri tanımlayabilir, ardından tahminleri gerçekleştirmek üzere bir model oluşturmak için verileri kullanabilir.

Denetimli öğrenme sınıflandırma ve regresyon ile ilişkilendirilen en yaygın öğrenme türüken, denetimsiz öğrenmede bu ilişkinin önemli bir kısmı kümelemedir. Bu iki öğrenme yöntemi arasındaki temel fark kullanılan veri türüdür. Veriler sınıflandırmada kategorik bir etiketle veya regresyonda sayısal bir değerle etiketlenirken kümelemede etiketlenmez. Bu ise gerçekleştirmeyi ve değerlendirmeyi zor ve karmaşık bir görev yapmaktadır. Bunun için kümeleme algoritmalarının performansının doğru şekilde ölçülmesi çok önemlidir.

Denetimli algoritmalar, doğruluk, R^2 değeri, duyarlılık, özgüllük vb. gibi birçok performans değerlendirme metriğine sahiptir. Bu noktada kümeleme tekniğinin doğruluğunu veya performansını ölçmek için en önemli metrik Silhouette katsayısı veya Silhouette puanıdır. Özellikle yüksek boyutlu veriler analiz edilirken Silhouette puanına ek olarak kümeleme algoritmasının çalışmasını doğrulamak için görselleştirmenin kullanılmasına imkan verir. Yapılan çalışma ile KDD Cup'99 saldırı tespiti veri seti üzerinde k-means kümeleme algoritması ile farklı k

değerlerine göre Silhouette metriği ile optimum küme sayısının belirlenmesi amaçlanmıştır.

Bu çalışmanın ana katkıları şu şekilde özetlenebilir:

- 1) Saldırı tespit sistemleri için K-means algoritması temelli yaklaşım önerilmiştir.
- 2) Saldırı tespit sistemlerindeki gibi büyük boyutlu veri setleri için Silhouette Puanı performans metriği kullanılmış ve test edilmiştir.
- 3) Önerilen metrik optimum küme sayısının belirlenmesi için daha kesin puan ve k sayısını vermektedir.

Bu çalışmanın organizasyonu şu şekildedir: Bölüm 2, ilgili çalışmaları anlatmaktadır. Bölüm 3, veri setini ve Siluet Puanı yöntemini tanıtmaktadır. Bölüm 4, önerilen metod ve deney sonuçlarını açıklamaktadır. Bölüm 5, çalışma özetlenmiştir.

2. LİTERATÜR TARAMASI (LITERATURE REVIEW)

Saldırı tespit sistemleri için:

Arif ve ark. [17] hibrit bir yaklaşım tanıtmıştır. Bu yaklaşımda, düğümün budanması parçacık sürü optimizasyonu tarafından gerçekleştirilir ve ağ tabanlı saldırı tespit sistemlerinde sınıflandırma için budanmış karar ağaçları kullanılmıştır. Ahmed ve ark. [18], boyutluluğun azaltılması için naive bayes özellik alt küme seçici tekniğinin uygulandığı hibrit bir saldırı tespit sistemi oluşturmak için üçlü bir strateji geliştirmiştir. Aykırı değerlerin reddi için optimize edilmiş destek vektör makinesi uygulanırken, sınıflandırıcı olarak öncelikli k-NN uygulanmıştır. Dash ve ark. [19], yerçekimsel arama ile parçacık sürüsü optimizasyon algoritmalarının birleşimi olan (GSPSO) dizisi olan iki yeni hibrit saldırı tespit yöntemi önermiştir. Yao ve ark. [20] saldırı tespit sistemleri için hibrit bir model önermiştir. K-means kümeleme algoritması ve sınıflandırma aşamasında tamamı denetimli öğrenme algoritmaları olan destek vektör makinesi, yapay sinir ağları, karar ağaçları ve rastgele orman farklı parametreler üzerinden karşılaştırılmıştır.

Suad ve Fadl [21], makine öğrenimi algoritmasını büyük veri ortamına uygulayan bir saldırı tespit sistemi modeli tanıttı. Bu çalışmada Spark-Chi-SVM modeli kullanılmaktadır. Ijaz ve ark. [22] vektörlere dayalı bir genetik algoritma tanıtmışlardır. Bu teknikte vektör kromozomları uygulanmıştır. Önerilen yaklaşımın benzersizliği, kromozomları bir vektör olarak ve eğitim verilerini metrik olarak göstermesidir. Alauthaman ve ark. [23], karar ağaçlarına yardımcı olmak üzere ileri beslemeli bir sinir ağı üzerine kurulu, eşler arası bot tespitine yönelik bir yaklaşım önerdi. Venkataraman ve Selvaraj [24] verilerin sınıflandırılması için etkili bir hibrit özellik seçim yapısı önermiştir. Sınıflandırma amacıyla ilgili özellikleri bulmak için simetrik belirsizlik uygulanır. Çalışmada

genetik algoritma alt kümeleri daha yüksek doğrulukla aramak için kullanılmıştır.

Kumar ve Kumar [25] akıllı tabanlı bir hibrit model önermiştir. Bu model daha sonra çok katmanlı algıyı, bulanık mantık denetleyiciyi, uyarlanabilir nöro-bulanık girişim sistemini ve bir nöro-bulanık genetiği entegre etmiştir. Çavuşoğlu ve diğerleri[26] makine öğrenimi teknikleri temelli hibrit bir yaklaşım önermişlerdir. Sınıflandırma amacıyla k- nn ve naive bayes algoritmaları kullanılırken, sınıflandırıcı olarak rastgele orman algoritması kullanılmıştır. Saxena ve ark. [27], yüksek kaliteli özellik alt kümelerini elde etmek amacıyla DBSCAN tabanlı bir hibrit teknik önermişlerdir. Verilerdeki gürültüyü ortadan kaldırmak için DBSCAN, verilerin gruplandırılması için k-means kümeleme kullanılmıştır. Kar ve ark. [28], verilerin sınıflandırılması için uygulamada ID3 adı verilen karar ağaçları algoritmasını kullanır. Sınıf etiketlerini, keşfedilmemiş veri noktasına en yakın noktaya atamak için k- nn yaklaşımı uygulanır. Baykara ve Daş [29] saldırı tespit sistemleri için honeypot tabanlı bir yaklaşım önermektedir. Geliştirilen honeypot sunucu uygulaması sunuculardaki ağ trafiğini gerçek zamanlı animasyonla görsel olarak gösterebilmektedir.

Dutta ve ark. [30] saldırı tespit sistemlerindeki sınıflandırma ölçümlerini geliştirmek için hibrit bir model önermiştir ve sınıflandırma doğruluğunu arttırmak için derin bir sinir ağı uygulanmıştır. Latah ve Toker [31] akış tabanlı çok seviyeli hibrit saldırı tespit sistemi tanıtmıştır. Yazar, sınıflandırma amacıyla k-nn ve aşırı öğrenme makinelerini uygulamış ve özellik seçme yöntemi olarak yazılım tanımlı ağ denetleyicisi kullanılmıştır. Sumaiya Thaseen ve ark. [32] korelasyon tabanlı özellik seçimi, en iyi özellik alt kümelerini seçmek için bir özellik seçme yaklaşımı olarak uygulanırken, yapay sinir ağı bir sınıflandırıcı olarak kullanılır. Safaldin ve ark. [33], saldırı tespit sınıflandırması için bir özellik seçme yöntemi ve destek vektör makinesi olarak geliştirilmiş ikili gri kurt optimizasyonu uygulamıştır. Vallathan ve ark. [34] IoT ortamında derin öğrenme yaklaşımını temel alan şüpheli eylem tespit sistemini önermiştir. Baykara ve Daş [35] web uygulamalarının güvenliği için hibrit bir gerçek zamanlı saldırı ve önleme sistemi yaklaşımı önermişlerdir. Önerilen sistem, kural tabanlı suistimal tespiti ve anormallik tespitini kullanmakta ve veri kaynağı olarak ağ paketlerini kullanmaktadır.

Ishaque ve ark. [36] saldırı tespit sistemi için bulanık mantık, sinir ağları ve genetik algoritma kullanan yeni bir hibrit teknik önermişlerdir. Çalışmada belirsizliğin giderilmesi sürecinde bulanık mantıktan, tahmin amacıyla ise sinir ağlarından yararlanılmıştır. Tahmin sonuçlarının doğruluğunda iyileştirmeler sağlamak amacıyla genetik algoritma kullanılmıştır. Nabi ve Zhou [37] saldırı tespit sistemlerinin geliştirilmesinde makine öğrenimi tekniklerinin karşılaştırıldığı bir çalışma sunulmuştur. Test edilen sınıflandırıcılar arasında J48 ağacı en yüksek doğruluğu sağlamıştır. Sınıflandırıcı performansını geliştirmek için Rastgele Projeksiyon ve PCA projeksiyon

yaklaşımları karşılaştırılmıştır. Aljehane ve ark. [38] ağ güvenliği için derin öğrenme destekli saldırı tespit sistemi (GJOADL-IDSNS) tekniği ile yeni bir altın çakal optimizasyon algoritması önermişlerdir. GJOADL-IDSNS sisteminin ana amacı, ağ güvenliğini sağlamak için izinsiz girişlerin etkili bir şekilde tanınması ve sınıflandırılmasıdır. Fraihat ve ark. [39] büyük ölçekli IoT NetFlow tabanlı ağlar için bir güvenlik önlemi olarak bir ağ saldırı tespit sistemi önermişlerdir. Önerilen NIDS, en uygun özellik kümesini belirlemek için aritmetik optimizasyon algoritmasının değiştirilmiş bir versiyonuyla desteklenen makine öğrenimini kullanmıştır. Seçilen yedi özellik, Rastgele Orman ve Ekstra Ağaçlar da dahil olmak üzere çeşitli ML modellerini eğitmek için kullanılmıştır.

Pramilarani ve Kumari [40] etkili bir saldırı tespit sistemi geliştirmek için maliyet tabanlı rastgele orman sınıflandırıcısı (CRFC) önermişlerdir. CRFC tabanlı sınıflandırma, özellik dengesizliği olsa bile özellikleri bölme sürecini iyileştirmeye yardımcı olan, özelliğin önemine göre hesaplanan maliyet matrisinin dahil edilmesiyle doğaçlama yapmışlardır. Al Nuaimi ve ark. [41] Edge-IIoT-2022 veri kümesini kullanarak veri odaklı saldırı tespit sistemlerini IoT ve endüstriyel IoT ortamlarında 6 adet makine öğrenme algoritması ile değerlendirmişlerdir. Sun ve ark. [42] STS sınıflandırmak ve tespit etmek için parçacık sürüsü optimizasyonunu ve AdaBoost algoritmalarını birleştiren bir sistem önermektedir. Deneysel sonuçlar, PSO-AdaBoost yaklaşımının izinsiz giriş tespitinde üstün doğruluk, kesinlik ve hatırlama sağladığını göstermektedir. Korium ve ark. [43] makine öğrenimine dayalı bir saldırı tespit sistemi önermişlerdir. Modelde veri ön işleme için veri dağıtımını koruyan Z-puanı normalleştirilmesi, model karmaşıklığını basitleştiren ve yürütme süresini azaltan bir regresyon modeli, model seçimi ve eğitimi için rastgele orman, ekstremgradyan artırma, kategorik artırma, hafif gradyan artırma makinesi ve eğitim aşamasındaki davranış kontrol etmek ve aşırı uyumu önlemek için hiperparametre optimizasyonu kullanılmıştır.

3. MATERYAL VE METOT (MATERIAL AND METHOD)

Silhouette metriği, kümeleme performansının değerlendirilmesine yardımcı olan bir ölçümdür. Kümeleme kalitesinin değerlendirilmesi, kümeleme algoritmalarının etkinliğini ve güvenilirliğini belirlemek için önemlidir. Kümeleme denetimsiz bir öğrenme görevi olduğundan kümeleri doğrulamak için net etiketler yoktur. Bu nedenle kümeleme sonuçlarının değerlendirilmesi, Silhouette skoru gibi dahili doğrulama metriklerinin kullanılmasını gerektirir.

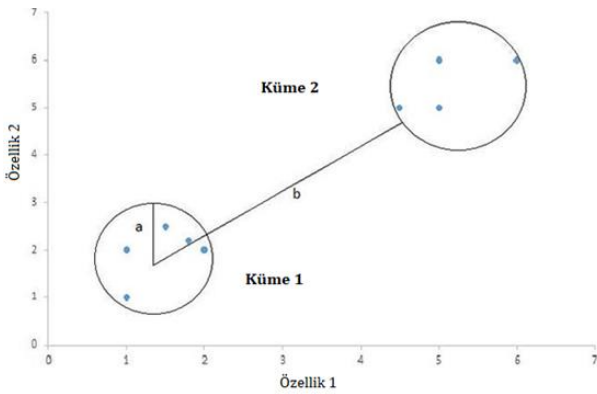
3.1. Veri Seti (Data Set)

Uygulamada KDD Cup'99 veri kümesi kullanılmıştır. KDD Cup'99 veri seti, DARPA98 veri setinden birkaç niteliğin çıkarılmasıyla elde edilmiştir. Veri kümesi 34 süreklili, 7 sembolik olmak üzere 41 öznelik değerli ve 4.898.430 adet veri içermektedir. KDD99 DARPA, saldırı

tespiti için makine öğrenmesi ve veri madenciliği tabanlı çalışmalarda en çok kullanılan veri setidir. Tavallae ve ark. [44], çoğunluğunu index dergilerden oluşan tarama çalışması kapsamında incelediği 276 makalenin % 75'inde KDD99 DARPA veri setinin kullanıldığını tespit edilmiştir.

3.2. Silhouette Metriği (Silhouette Metric)

Silhouette puanı ve silhouette grafiği, kümeler arasındaki ayırma mesafesini ölçmek için kullanılır. Bir kümedeki her bir noktanın komşu kümelerdeki noktalara ne kadar yakın olduğunun bir ölçüsünü görüntüler. Bu ölçüm [-1, 1] aralığına sahiptir ve kümeler içindeki benzerlikleri ve kümeler arasındaki farklılıkları görsel olarak da sunmaktadır. Şekil 1'de silhouette küme gösterimi sunulmuştur.



Şekil 1. Silhouette küme gösterimi
(Silhouette cluster representation)

a: ortalama küme içi mesafe, yani bir küme içindeki her nokta arasındaki ortalama mesafe.

b: kümeler arası ortalama mesafe, yani tüm kümeler arasındaki ortalama mesafe.

Bir veri noktası i için silhouette puanı şu şekilde verilir:

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (1)$$

b_i : bir parçası olması dışında, i veri noktasının en yakın kümesine olan ortalama mesafe olarak tanımlanan kümeler arası mesafedir.

$$b_i = \min_{|C_k|} \sum_{j \in C_k} d(i, j) \quad (2)$$

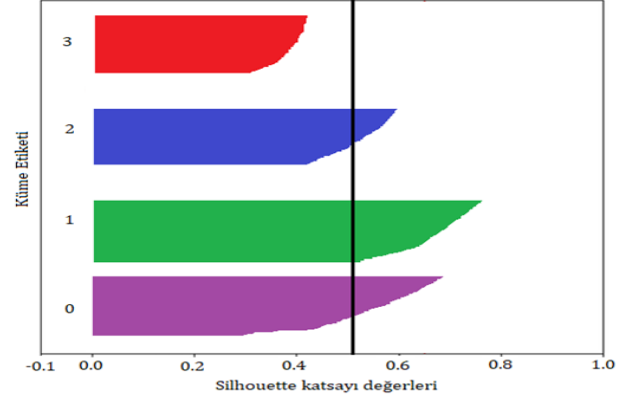
a_i : parçası olduğu kümedeki diğer tüm noktalara olan ortalama mesafe olarak tanımlanan küme içi mesafedir.

$$a_i = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (3)$$

3.3. Grafıksel Özet (Graphical Abstract)

Tüm veri kümesinin genel silhouette puanı, veri kümesindeki tüm veri noktalarının silüet puanının ortalaması olarak hesaplanabilir. Formülden görülebileceği gibi silhouette puanı her zaman [-1, 1] arasında olacaktır.

Şekil 2'de görüldüğü gibi silhouette grafikleri, y ekseninde küme etiketini temsil ederken, x ekseninde gerçek silhouette puanını temsil eder. Silhouette'lerin boyutu/kalınlığı da o küme içindeki örneklerin sayısıyla orantılıdır.

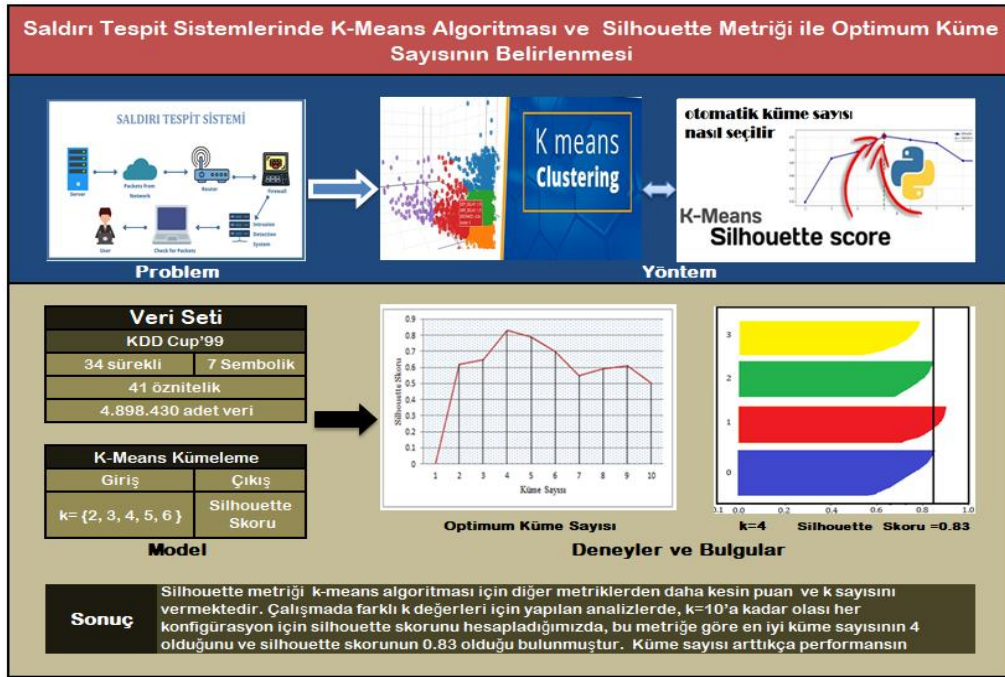


Şekil 2. Silhouette grafik gösterimi
(Silhouette graphic representation)

Silhouette katsayıları +1'e ne kadar yakınsa, kümenin örnekleri komşu küme örneklerinden o kadar uzaktır. 0 değeri, örneklerin iki komşu küme arasındaki karar sınırında veya çok yakınında olduğunu gösterir. Negatif değerler ise bu örneklerin yanlış kümeyle atanmış olabileceğini gösterir. Silhouette katsayılarının ortalamasının alınması tüm kümenin performansını tek bir silhouette puanı ile hesaplanabilmesini sağlar. Tablo 1'de silhouette metriği algoritması sunulmuştur.

Tablo 1. Silhouette metriği algoritması
(Silhouette metric algorithm)

FUNCTION [index] = SILHOUETTE [D]	
1	N=size of D;
2	sum_All=0;
3	foreachclass j=1 to c;
4	M= size of Dj;
5	sum_Class=0;
6	for i=1 to M
7	b=b(i);
8	a=a(i);
9	s=(b-a) / max(b,a);
10	sum_Class+=s;
11	endfor
12	sum_All+=sum_Class / M;
13	endfor
14	index=sum_All / c;

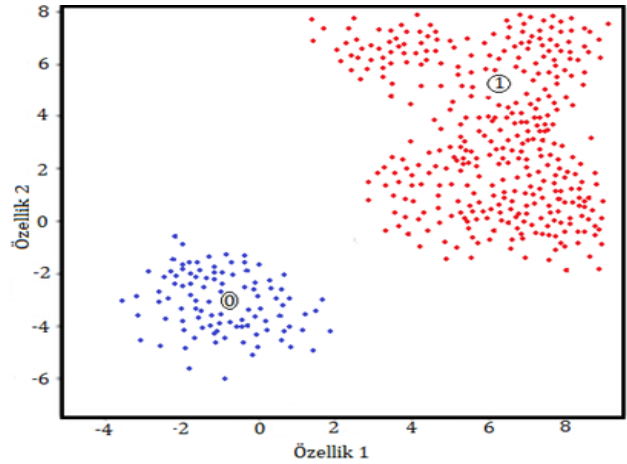
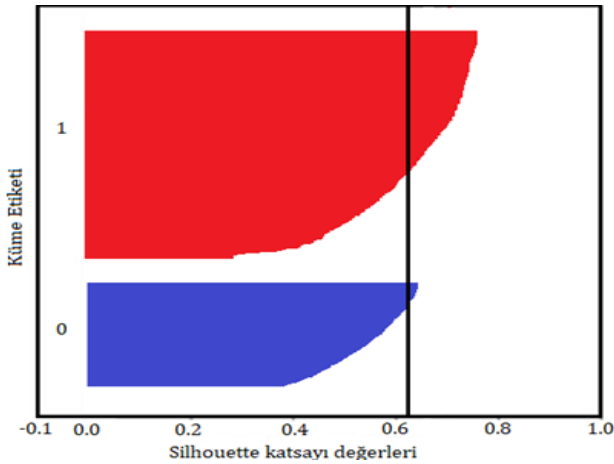


Şekil 3. Grafikselsel Özet
(Graphical Abstract)

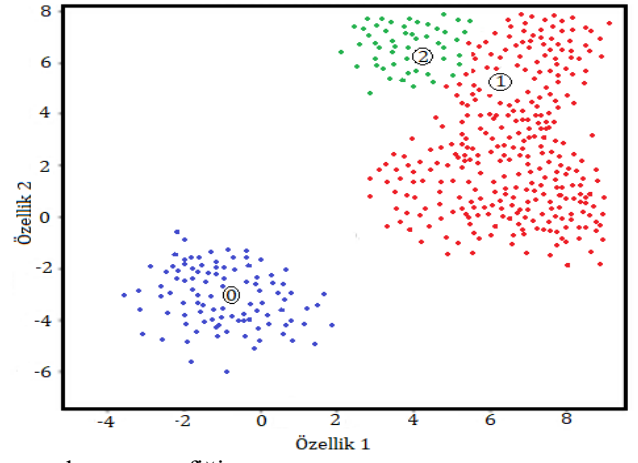
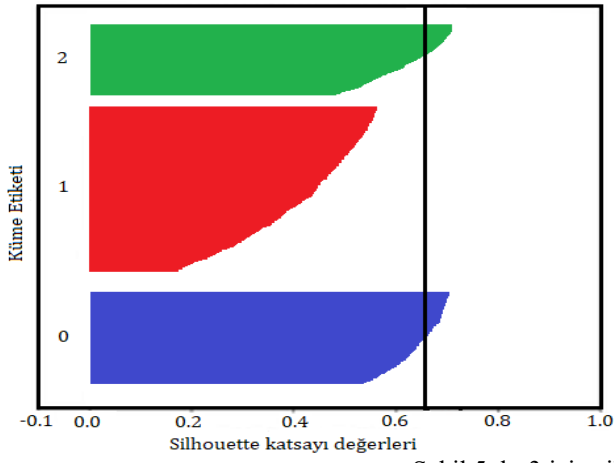
4. UYGULAMA VE SONUÇLARI (APPLICATION AND RESULTS)

Uygulama kapsamında ,KDD Cup'99 saldırı tespiti veri seti üzerinde k-means kümeleme algoritması ile farklı k değerlerine göre silhouette metriği ile optimum küme sayısının belirlenmesi amaçlanmıştır. Bunun için 5 farklı k parametresi sırasıyla 2, 3, 4, 5, 6 değerleri analiz edilmiştir. Her bir konfigürasyonun performansına bakılarak silhouette skoru ve silhouette grafiği ile optimum küme sayısını bulması sağlanmıştır.

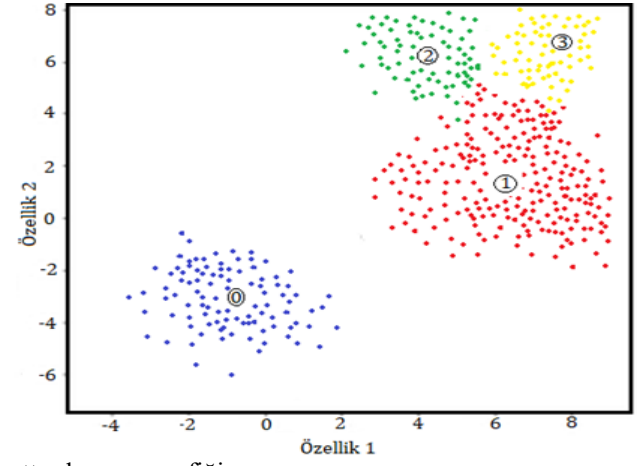
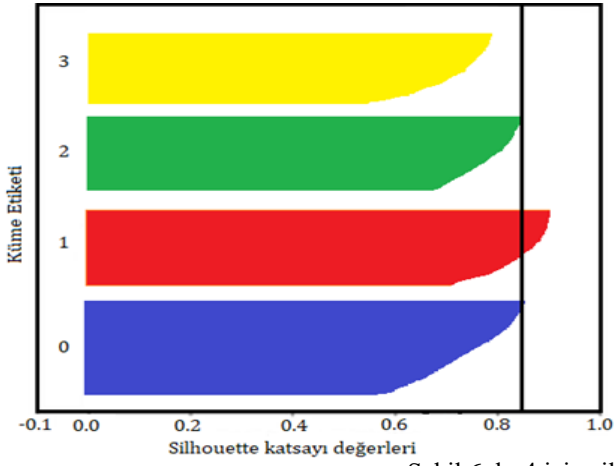
Veriler üzerinde k-means kümeleme k= {2, 3, 4, 5, 6} değerleri için elde edilen silhouette skorları ve silhouette grafikleri Şekil 4, Şekil 5, Şekil 6, Şekil 7 ve Şekil 8'deki gibidir. Her konfigürasyonun performansına bakılarak silhouette grafiği en iyi sayıda kümenin bulunmasını sağlamaktadır. Şekil 4'de popülasyonu ayırmak için iki küme anlamına gelen k=2'yi kullanarak ortalama 0,62 silhouette puanı elde edilmiştir. Küme sayısı üçe çıkarıldığında ortalama silhouette puanı bir miktar artmıştır. Ayrıca her kümede daha az örnek olması nedeniyle küme sayısı arttıkça silhouette grafiğinin kalınlığının azalmaya devam ettiği de görülmektedir.



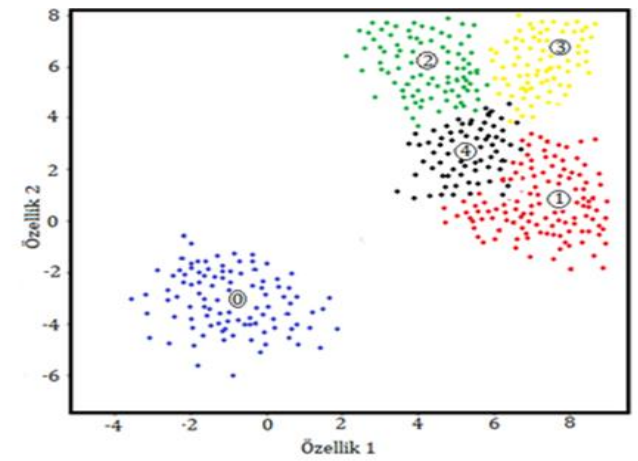
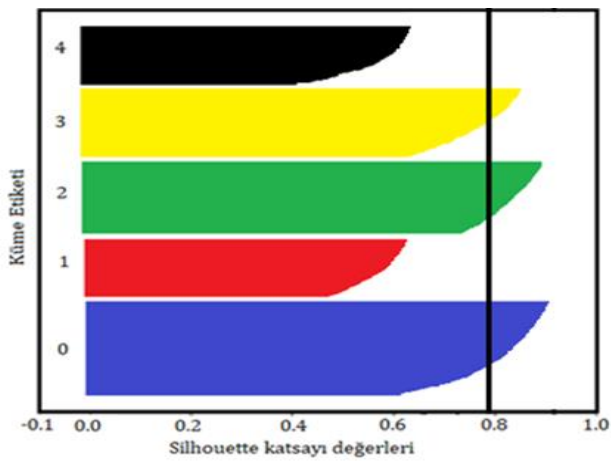
Şekil 4. k=2 için silhouette skoru ve grafiği
(Silhouette score and graph for k=2)



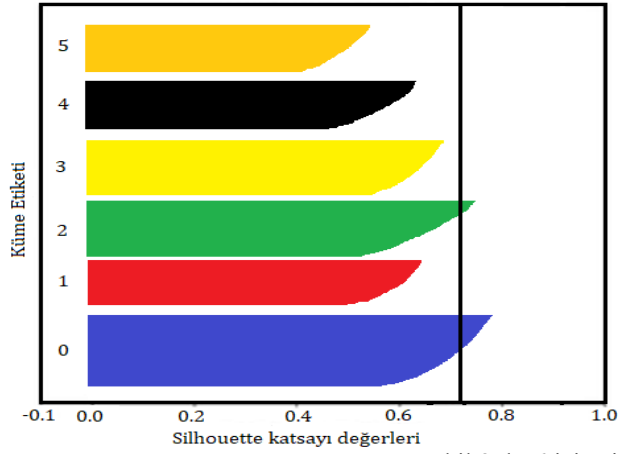
Şekil 5. k=3 için silhouette skoru ve grafiği
(Silhouette score and graph for k=3)



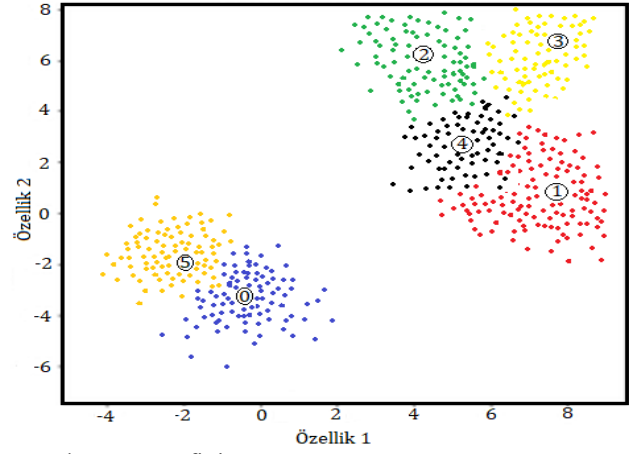
Şekil 6. k=4 için silhouette skoru ve grafiği
(Silhouette score and graph for k=4)



Şekil 7. k=5 için silhouette skoru ve grafiği
(Silhouette score and graph for k=5)



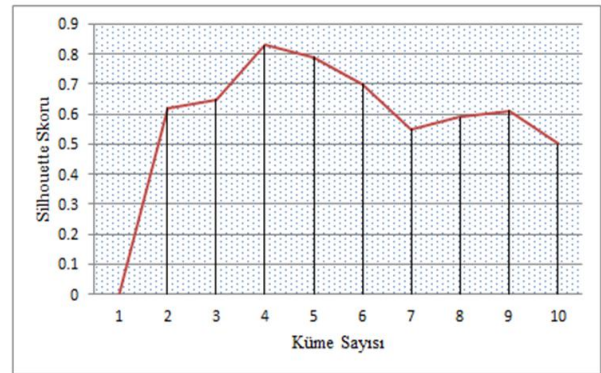
Şekil 8. k=6 için silhouette skoru ve grafiği
(Silhouette score and graph for k=6)



Çalışmada k küme sayısına göre bulunan silhouette puanları Tablo 2’de verilmiştir. Silhouette metriği, en uygun değeri seçmek için kullanılır. Silüet grafiği, ortalamanın altında silüet puanlarına sahip kümelerin varlığı ve ayrıca silüet çizimlerinin boyutunda geniş dalgalanmalar nedeniyle 2, 3 ve 6 değerinin verilen veriler için kötü bir seçim olduğunu 4 ile 5 ‘in karar vermede daha iyi bir seçim olduğunu, en optimum seçimin ise 4 değerinin olduğunu göstermektedir.

Tablo 2. k- means için silhouette analiz sonuçları
(Silhouette analysis results for k- means)

k	Silhouette Skoru
2	0.6223615745161018
3	0.6628246153550858
4	0.8344216225076757
5	0.7938500642344114
6	0.7037404360466380



Şekil 9. Optimal küme sayısı
(Optimal number of clusters)

Çalışmada farklı k değerleri için yapılan analizlerde, k=10’a kadar olası her konfigürasyon için Silhouette Skorunu hesapladığımızda, Şekil 9’da görüldüğü gibi bu metriğe göre en iyi küme sayısının 4 olduğunu ve küme sayısı arttıkça performansın kötüleştiğini görebiliriz. Ayrıca silüet grafiğinin kalınlığından küme boyutu görselleştirilebilir. Küme sayısı arttıkça her kümede daha az örnek olması nedeniyle silhouette kalınlığının azalmaya devam ettiği de fark edilebilir.

5. SONUÇLAR (RESULTS)

Silhouette, veri kümeleri içindeki tutarlılığın yorumlanması ve doğrulanması yöntemini ifade eder. Bu teknik, her bir nesnenin ne kadar iyi sınıflandırıldığına ilişkin kısa ve öz bir grafiksel gösterim sağlar. Çalışmada saldırı tespit sistemleri için en yaygın kullanılan KDD Cup’99 veri seti k-means kümeleme algoritması ile farklı k değerlerine göre analiz edilip silhouette puanlarının ve grafiklerinin bulunması böylelikle optimum küme sayısının belirlenmesi sağlanmıştır. Böylelikle saldırı tespiti için kullanılan veri setlerinin doğru küme değerleri ile sınıflandırılması sağlanmıştır.

K-means kümeleme, basit ve popüler bir denetimsiz makine öğrenme algoritmasıdır. Silhouette metriği bize k-means algoritması için diğer metriklerden daha kesin puan ve k sayısını vermektedir. Çalışmada farklı k değerleri için yapılan analizlerde, k=10’a kadar olası her konfigürasyon için silhouette skorunu hesapladığımızda, bu metriğe göre en iyi küme sayısının 4 olduğunu ve silhouette skorunun 0.83 olduğu bulunmuştur. Küme sayısı arttıkça performansın düştüğü ve silhouette kalınlığının azaldığı görülmüştür.

Kümeleme performans metrikleri olarak, silhouette puanı, rand endeksi, düzeltilmiş rand endeksi, karşılıklı bilgi, Calinski-Harabasz endeksi ve Davies-Bouldin endeksi kullanılmaktadır. Çalışmada silhouette performans metriğine odaklanılmıştır. Gelecekte, kümeleme yönteminden daha doğru sonuçlar elde

edilmesi ve daha geniş bir uygulama kapsamına sahip olması için diğer metriklerin performans analizleri gerçekleştirilebilir. Özellikle yüksek boyutlu veri setlerinin sınıflandırılmasındaki zorluklar nedeni ile çalışma KDD Cup'99 veri kümesi üzerinde sınanmıştır. Ancak meteoroloji, enerji ve sağlık gibi büyük boyutlu veri kümelerinde önerilen yöntemin daha kesin ve başarılı sınıflandırma için kullanılabileceği düşünülmektedir.

KAYNAKLAR (REFERENCES)

- [1] M. Baykara, R. Daş, "SoftSwitch: a centralized honeypot-based security approach using software-defined switching for secure management of VLAN networks," *Turkish Journal of Electrical Engineering and Computer Sciences*, Vol. 27, no. 5, pp. 3309-3325, 2019.
- [2] L. Hung-Jen, C.-h. R. Lin, "Intrusion detection system a comprehensive review", *Journal of network and applications*, vol. 36, no. 1, pp. 16–24, 2013.
- [3] H. L. Motoda, H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, vol. 454, Springer, 1998.
- [4] L. D. S. Silva, A. C. Santos, T. D. Mancilha, J. D. Silva, A. Montes, "Detecting attack signatures in the real network traffic with ANNIDA", *Expert Systems with Applications*, vol. 34, no. 4, pp. 2326–2333, 2008.
- [5] A. Patcha, J. M. Park, "An overview of anomaly detection techniques: existing solutions and latest technological trends", *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.
- [6] C. Manikopoulos, S. Papavassiliou, "Network intrusion and fault detection. A statistical anomaly approach," *IEEE Communications Magazine*, vol. 40, no. 10, pp. 76–82, 2002.
- [7] P. Fournier-Viger, C. W. Lin, A. Gomariz et al., "The SPMF open-source data mining library version 2", *Joint European conference on machine learning and knowledge discovery in databases*, pp. 36–40, Riva del Garda, Italy, 2016.
- [8] P. Fournier-Viger, J. C.-W. Lin, R. U. Kiran, Y. S. Koh, R. Thomas, "A survey of sequential pattern mining", *Data Science and Pattern Recognition*, vol. 1, no. 1, pp. 54–77, 2017.
- [9] A. Smola, S.V.N. Vishwanathan, *Introduction to Machine Learning*, Cambridge University Press, ISBN-10: 0521825830, 2008.
- [10] Z. Xiaojin, *Semi-Supervised Learning Literature Survey*, vol. 2, *Computer Science, University of Wisconsin, Madison*, 2008.
- [11] S. Mukkamala, A. H. Sung, A. Abraham, "Modeling intrusion detection systems using linear genetic programming approach," in *The 17th international conference on industrial & engineering applications of artificial intelligence and expert systems, innovations in applied artificial intelligence*, pp. 633–642, Berlin, Heidelberg, 2004.
- [12] J. Pearl, "Bayesian networks. A model of self-activated memory for evidential reasoning," in *Proceedings of the 7th Conference of the Cognitive Science Society*, University of California, pp. 329–334, Irvine, CA, 2009.
- [13] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression (PDF)," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [14] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, University of California Press, 1967.
- [15] L. E. Baum, T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The annals of mathematical statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.
- [16] M. Mohammed, M. B. Khan, E. B. Bashier, **Machine Learning Algorithms and Applications**, CRC press Taylor and Francis Group, ISBN-10: 1498705383, 2016.
- [17] J. Arif, F. Malik, K. Aslam, "A hybrid technique using binary particle swarm optimization and decision tree pruning for network intrusion detection", *Cluster Computing*, vol. 21, pp. 667–680, 2017.
- [18] I. Ahmed, L. Saleh, M. Fatma, L. Talaat, "A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers", *Artificial Intelligence Review*, vol. 51, pp. 403–443, 2017.
- [19] D. Tirharaj, "A study on intrusion detection using neural networks trained with evolutionary algorithms", *Soft Computing*, vol. 21, pp. 2687–2700, 2017.
- [20] Y. Haipeng, W. Qiyi, "An intrusion detection framework based on hybrid multi-level data mining," *International Journal of Parallel Programming*, vol. 47, pp. 740–758, 2017.
- [21] M. Suad, M. Fadl, "Intrusion detection model using machine learning algorithm on Big Data environment", *Journal of big data*, vol. 5, pp. 1–12, 2018.
- [22] S. Ijaz, F. A. Hashmi, S. Asghar, M. Alam, "Vector based genetic algorithm to optimize predictive analysis in network security", *Applied intelligence*, vol. 48, no. 5, pp. 1086–1096, 2018.
- [23] A. Mohammad, A. Nauman, "A P2P Botnet detection scheme based on decision tree and adaptive multilayer neural networks", *Neural Computing & Applications*, vol. 29, pp. 991–1004, 2018.
- [24] V. Sivakumar, S. Rajalakshmi, "Optimal and novel hybrid feature selection framework for effective data classification," in *Advances in Systems, Control and Automation*, pp. 499–514, Springer, Singapore, 2018.
- [25] K. Neeraj, K. Upendra, "Knowledge computational intelligence in network intrusion detection systems", *Knowledge Computing and Its Applications*, pp. 161–176, Springer, Singapore, 2018.
- [26] C. Unal, "A new hybrid approach for intrusion detection using machine learning methods", *Applied Intelligence*, vol. 49, pp. 2735–2761, 2019.
- [27] S. Akash, S. Khushboo, "Hybrid technique based on DBSCAN for selection of improved features for intrusion detection system", in *Emerging Trends in Expert Applications and Security*, pp. 365–377, Springer, Singapore, 2019.
- [28] P. Kar, S. Banerjee, K. C. Mondal, G. Mahapatra, S. Chattopadhyay, "A hybrid intrusion detection system for hierarchical filtration of anomalies", *Information and Communication Technology for Intelligent Systems*, vol. 106,

pp. 417–426, Springer, Singapore, 2019.

- [29] M. Baykara, R. Daş, "A novel honeypot based security approach for real-time intrusion detection and prevention systems," *Journal of Information Security and Applications (JISA)*, Vol.41, pp. 103-116, 2018.
- [30] V. Dutta, M. Choras, R. Kozik, M. Pawlicki, "Hybrid model for improving the classification effectiveness on network intrusion detection system", in **Conference on Complex, Intelligent, and Software Intensive Systems**, Cham, 2020.
- [31] M. Latah, L. Toker, "An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks", *CCF Transactions on Networking*, vol. 3, pp. 26–271, 2020.
- [32] I. Sumaiya Thaseen, J. Saira Banu, K. Lavanya, M. Rukunuddin Ghalib, K. Abhishek, "An integrated intrusion detection system using correlation-based attribute selection and artificial neural network", *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 2, article e4014, 2021.
- [33] M. Safaldin, M. Qtair, L. Abualigah, "Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks", *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 1559–1576, 2021.
- [34] G. Vallathan, A. John, C. Thirumalai, "Suspicious activity detection using deep learning in secure assisted living IoT environments", *The Journal of Supercomputing*, vol. 77, pp. 3242–3260, 2021.
- [35] M. Baykara, R. Daş, "A Novel Hybrid Approach for Detection of WebBased Attacks in Intrusion Detection Systems," *International Journal of Computer Networks and Applications (IJCNA)*, Vol.4, no. 2, pp. 62-76, 2017.
- [36] M. Ishaque, Md G. Md Johar, A. Khatibi, M. Yamin, "A novel hybrid technique using fuzzy logic, neural networks and genetic algorithm for intrusion detection system," *Measurement: Sensors*, Vol.30, pp. 1-12 ,2023.
- [37] F. Nabi, X. Zhou, "Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security", *Cyber Security and Applications*, Vol.2, pp. 1-8, 2024.
- [38] N. O. Aljehane, H. A. Mengash, M. M. Eltahir, F. A. Alotaibi, S. S. Aljameel, A. Yafoz, R. Alsini, M. Assiri, "Golden jackal optimization algorithm with deep learning assisted intrusion detection system for network security", *Alexandria Engineering Journal*, Vol.86, pp. 415-424, 2024.
- [39] S. Fraihat, S. Makhadmeh, M. Awad, M. A. Al-Betar, A. Al-Redhaei, "Intrusion detection system for large-scale IoT NetFlow networks using machine learning with modified Arithmetic Optimization Algorithm", *Internet of Things*, Vol. 22, pp. 1-22, 2023.
- [40] K. Pramilarani, P. V. Kumari, "Cost based Random Forest Classifier for Intrusion Detection System in Internet of Things", *Applied Soft Computing*, Vol. 151, pp. 1-8, 2024.
- [41] T. Al Nuaimi, S. Al Zaabi, M. Alyilieli, M. AlMaskari, S. Alblooshi, F. Alhabsi, M. F. Bin Yusof, A. Al Badawi, "A comparative evaluation of intrusion detection systems on the edge-IIoT-2022 dataset", *Intelligent Systems with Applications*, Vol.20, pp. 1-10, 2023.
- [42] Z. Sun, G. An, Y. Yang, Y. Liu, "Optimized machine learning enabled intrusion detection 2 system for internet of medical things", *Franklin Open*, Vol.6, pp. 1-11, 2024.
- [43] M. S. Korium, M. Saber, A. Beattie, A. Narayanan, S. Sahoo, P. H.J. Nardelli, "Intrusion detection system for cyberattacks in the Internet of Vehicles environment", *Ad Hoc Networks*, Vol. 153, pp. 1-16, 2024.
- [44] M. Tavallae, N. Stakhanova, A. A. Ghorbani, "Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, no. 5, pp. 516-524, 2010.

ALBERT4Spam: A Novel Approach for Spam Detection on Social Networks

Araştırma Makalesi/Research Article

 Rezan BAKIR*¹,  Hasan Erbay²,  Halit BAKIR³

¹Department of Computer Engineering , Sivas science and technology university, Sivas, Turkey

²Department of Computer Engineering, Ostim Technical University, Ankara, Turkey

³Department of Computer Engineering , Sivas science and technology university, Sivas, Turkey

rezan.bakir@sivas.edu.tr, hasan_erbay@yahoo.com, halit.bakir@sivas.edu.tr

(Geliş/Received:26.01.2024; Kabul/Accepted:04.03.2024)

DOI: 10.17671/gazibtd.1426230

Abstract— Engaging in social media browsing stands out as one of the most prevalent online activities. As social media increasingly integrates into our daily routines, it opens up numerous opportunities for spammers seeking to target individuals through these platforms. Given the concise and sporadic nature of messages exchanged on social networks, they fall within the realm of short text classification challenges. Effectively addressing such issues requires appropriately representing the text to enhance classifier efficiency. Accordingly, this study utilizes robust representations derived from contextualized models as a component of the feature extraction process within our deep neural network model, which is built upon the Bidirectional Long Short-Term Memory neural network (BLSTM). Introducing ALBERT4Spam, the study presents a deep learning methodology aimed at identifying spam on social networking platforms. It harnesses the proven ALBERT model to acquire contextualized word representations, thereby elevating the effectiveness of the suggested neural network framework. The random search method was used to fine-tune the ALBERT4Spam model's hyperparameters, which included the number of BLSTM layers, neuron count, layer count, activation function, weight initializer, learning rate, optimizer, and dropout, in order to obtain optimal performance. The experiments conducted on three benchmark datasets demonstrate that our innovative model surpasses widely used methods in social network spam detection. The precision results stand at 0.98, 0.96, and 0.98 for Twitter, YouTube, and SMS datasets, respectively, showcasing superior performance outcomes.

Keywords—spam detection, word embedding, deep learning, BERT, ALBERT, BLSTM

ALBERT4Spam: Sosyal Ağlarda Spam Tespitinde Yeni Bir Yaklaşım

Özet: Sosyal medyada gezinmek, insanların katıldığı en popüler çevrimiçi etkinliklerden biridir. Sosyal medya, günlük hayatlarımıza daha fazla entegre oldukça, kurbanlarına sosyal ağ siteleri aracılığıyla ulaşmak isteyen spam göndericilere sayısız fırsat sunmaktadır. Sosyal ağlar üzerinden iletilen mesajlar genellikle kısa ve seyrek olduğu için, kısa metin sınıflandırma problemleri ortaya çıkmaktadır. Bu tür sorunların üstesinden gelmek için, sınıflandırıcının etkinliğini artırmak amacıyla metni uygun şekilde temsil etmek önemlidir. Bu amaçla, bu çalışma, sosyal medya platformlarında spam tanımlamak için derin öğrenme yaklaşımı olan ALBERT4Spam'i tanıtmaktadır. ALBERT modelinden gelen bağlamsal kelime temsilleri kullanılarak önerilen Çift Yönlü Uzun Kısa Süreli Bellek (BLSTM) sinir ağı mimarisinin performansı artırılmıştır. Önerilen ALBERT4Spam modelinde kullanılan BLSTM katman sayısı, nöron sayısı, katman sayısı, aktivasyon fonksiyonu, öğrenme oranı, ağırlık başlangıç, optimizasyon ve bırakma gibi hiperparametreler, en iyi performansa ulaşmak için rastgele arama yöntemi kullanılarak optimize edilmiştir. Üç farklı standart veri seti üzerinde yapılan deneysel sonuçlar, önerilen modelin mevcut modellere kıyasla sosyal medya platformlarındaki spam mesajlarını daha başarılı bir şekilde tespit ettiğini göstermektedir. Yapılan deneyler, Twitter, YouTube ve SMS veri setlerinde sırasıyla %98, %96 ve %98 kesinlik sonuçlarıyla daha üstün performans sergilediğini ortaya koymaktadır.

Anahtar Kelimeler— spam tespiti, kelime gömme, derin öğrenme, BERT, ALBERT, BLSTM

1. INTRODUCTION

Even as social networks have become integral to human connection and relationship development, enhancing lives, they have also introduced security concerns that pose serious threats to user privacy and security.

In this discourse, we draw attention to the pervasive issue of spam on social media, a common challenge for users across various networks. Spam messages, defined as unsolicited, low-quality content that disrupts users and wastes their time, encompass a range of nuisances, including advertisements, malicious links, trending hashtags, and images with concealed URLs. Because of the brief and sparse nature of social media text, as well as potential grammar or spelling problems, symbols, emoticons, and various patterns, handling text derived from social media presents particular issues. Addressing the problem of spam messages necessitates finding an effective representation of text to enhance the classifier's ability to discern messages accurately. Various word representation approaches have been proposed, including traditional methods like the Bag of Word model, classical word embedding methods such as word2vec [1], GloVe [2], and fastText [3], and transformer-based methods like BERT [4]. Traditional word embedding models successfully handle regular-length texts, they encounter difficulties with short texts, particularly those sourced from social platforms due to sparsity and text ambiguity. Conventional word embedding approaches, which represent words with dense vectors, address data sparsity. However, despite their restrictions, they show limitations. Notably, the "out of vocabulary" problem arises when models unable to offer vector representations for words that are not in their dictionary. Another challenge is their context independence, where one vector is output for each word irrespective of its position in the sentence. To overcome these limitations, a dynamic word representation is essential. Contextual embedding, prevalent in transformer-based models, emerges as an ideal choice for short texts, generating word embeddings dependent on the surrounding words in a sentence. BERT [4], a prominent transformer-based model, achieves cutting-edge results in NLP. However, its full-sized models pose computational and memory challenges. The Google research team introduced ALBERT [5] (A Lite BERT) as a solution, incorporating parameter-reduction techniques such as factorized embedding parameterization and cross-layer parameter sharing. ALBERT, featuring a self-supervised loss for sentence-order prediction, enhances inter-sentence coherence.

This paper explores the dynamics of spam on social networks, delving into the challenges of short text representation and introducing innovative solutions like ALBERT for improved classification in the realm of spam detection.

Motivation and contribution

Learning natural language representations has proven valuable across various NLP tasks and has become a widespread practice. A notable shift in recent years involves transitioning from pre-training word embeddings—whether standard or contextualized—to comprehensive network pre-training followed by fine-tuning for specific tasks.

The inclusion of contextual word embeddings into neural architectures has improved the performance of cutting-edge NLP tasks dramatically. Given our focus on short text classification challenges, combining contextualized word embeddings with deep learning architectures emerges as an optimal solution for addressing the sparsity and other issues prevalent in social networks' texts. Consequently, we opted to utilize the ALBERT model as a feature extractor, integrating it with deep learning techniques to overcome challenges associated with traditional word representation methods.

The key contribution of our study is that we created a novel hybrid deep neural network model to identify spam on social platforms. This model combines a stacked bidirectional long short-term neural network architecture with contextualized word representations from the ALBERT model to provide more context. To the best of our knowledge, the combination of the bidirectional LSTM and the BERT model has never been applied to the domain of social network spam detection. The research involves an ablation investigation as well as hyperparameter tuning to improve performance. The model's usefulness in handling brief text data was confirmed by testing it on three benchmark datasets.

2. RELATED WORKS

Spam, in various forms, has long been associated with activities causing harm or disruption to users. While the word is not new, having appeared in contexts such as email and SMS spam, the advent of social networks has led to an increase in social spam. Consequently, spam detection on social networks has emerged as a prominent research area over the last decade.

This section provides an overview of significant contributions by various researchers in the realm of spam detection on social networks. Previous studies have employed diverse approaches, relying on features such as content-based features, account-based features, or a combination thereof, along with other innovative attributes. Some studies focused on identifying spam messages, while others aimed at detecting spammers' accounts. For instance, in [6], the authors proposed a methodology to detect spam profiles on Twitter by extracting and analyzing public features, irrespective of the language used in tweets. And in the research study presented in [7], an unsupervised spam detection approach was utilized to differentiate spammers from legitimate users.

In another study [8], enhancements to three classification algorithms for detecting spammers on Twitter were introduced, leveraging features that capture similarities among spam accounts. On a different note, [9] research study presented an unsupervised method for identifying spam messages on Twitter, utilizing a fuzzy string-matching based technique.

Moreover, the paper [10] suggested a strategy for detecting spam accounts that combines text analysis, machine learning, and short link analysis. A dataset was developed to detect user communications, hyperlinks in messages, and spam accounts.

Furthermore, the [11] study investigated Twitter spam detection using a hybrid framework, SMOTE-ENN, which combines SMOTE with Edited Nearest Neighbors. The data is then fed into deep learning classification approaches, and the best framework is found through performance comparisons and simulations.

Addressing the spam issue on YouTube's social network, [12] study introduced a framework employing Support Vector Machine (SVM) and K-Nearest Neighbor (k-NN) for spam comment detection. In [13], similar classifiers, along with a multilayer perceptron, were used to combat spam on YouTube. Authors in [14] study suggested a method for distinguishing spammers from real users based on Twitter-specific features and a Naive Bayes classifier.

On the other hand, the [15] study used Artificial Intelligence algorithms to detect spam in YouTube comments, achieving a 91.65% accuracy rate. According to authors, the study outperforms existing technologies like Google Safe Search and YouTube Bookmaker, highlighting the complexity of spam and the effectiveness of existing AI techniques. Similarly, the [16] study used Naive Bayes classification to identify spam comments on the internet, achieving an accuracy of 92.78%. Furthermore, the [17] study proposes a method for detecting English spam comments on YouTube using a stacked ensemble model and a browser extension to remove spam comments by altering CSS.

Various word representation methods, including TF-IDF vectorizer [12], Bag of Words (BoW) [18], and word2vec [19], have been applied to extract features from text in previous studies [20], [21], [22].

The application of neural networks extends across a myriad of fields, revolutionizing the way we approach complex problems and processes. For example, in the realm of healthcare, neural networks contribute to medical image analysis, disease diagnosis, and drug discovery, enhancing precision and efficiency [23], [24]. And in the realm of agriculture, the application of neural networks is transforming traditional practices and ushering in a new era of precision farming and

resource management. Neural networks are being utilized to analyze vast amounts of agricultural data, ranging from weather patterns and soil conditions to crop health and yield predictions. [19].

Within the realm of cybersecurity, the application of neural networks has emerged as a pivotal area of study and implementation. Researchers and security experts are delving into the intricacies of utilizing artificial intelligence, particularly neural networks, to bolster defenses against evolving cyber threats [26], [27].

More recently, Neural networks have acquired favor in tackling spam on social networking sites. For instance, [28] research paper proposed a tweet-level spam recognition architecture based on convolutional neural networks. In other study [29], on two social network datasets, cost-sensitive ensemble learning approaches with regularized deep neural networks was used in the proposed approach to detect spam messages. Moreover, [30] paper presented a deep learning-based architecture that incorporates knowledge bases such as WordNet and ConceptNet to provide semantic information to word representation.

What distinguishes the suggested approach from earlier researches is the use of the ALBERT model for word representation. This eliminates the need for extra knowledge bases, streamlines the process, and reduces complexity. Furthermore, the model efficiently solves the "out of vocabulary" problem by incorporating previously viewed text during training. This sets it apart from typical word embedding methods.

Moreover, the proposed model, which uses the tokenization strategy in the ALBERT model, is versatile in accepting any text. It was tested on three benchmark datasets, with comparable performance based only on text content.

3. METHODOLOGY

3.1. Used dataset

The experiments utilized three distinct benchmark datasets. The initial dataset, referred to as the Twitter dataset and introduced by Chen et al. in [31], originally comprised 100,000 tweet IDs with corresponding labels. To gather tweet information based on these IDs, a python script-based crawler was implemented. However, only 58,159 tweets were successfully obtained, as other tweets had been deleted either by Twitter or users. Due to the dataset's imbalanced nature, data augmentation was applied to enhance the classifier's performance. Subsequently, the dataset was divided into an 80% training set and a 20% testing set.

The second dataset originated from YouTube comments, as utilized in [18]. This dataset had five different sorts of comments, combining spam and normal comments. Finally, the SMS spam dataset, the third dataset, can be found in the UCI repository. Each dataset is fully described in Table 1.

Table 1. datasets summary

Dataset	The amount of spam	The amount of non-spam	Total number of messages
Twitter	38205	27822	66027
YouTube	941	935	1876
SMS spam	747	4827	5574

3.2. The proposed Model

In this work, we have innovatively combined the Bidirectional Long Short-Term Memory (BLSTM) neural network architecture with the ALBERT pre-trained model to address the challenge of spam detection in social networks. The proposed model comprises two distinct phases: the initial phase involves data preparation and feature extraction, while the subsequent phase focuses on classification. Each of these phases will be comprehensively discussed in the upcoming sub-sections.

3.2.1. Data preparation and feature extraction phase

The pivotal stages in constructing precise deep learning models are data cleaning, processing, and feature selection. Dealing with the condensed text extracted from social networking sites necessitates transforming the text into a more manageable format to enhance the performance of classification algorithms. Accordingly, several pre-processing steps have been implemented on the utilized datasets. These steps involve removing hyperlinks, punctuation, emojis, and special characters, as well as deleting stop phrases and switching the text to lowercase. The execution of data pre-processing was facilitated through the utilization of a purpose-built Python script crafted for this specific objective.

Feature extraction

Feature extraction is a technique that reduces the number of features in a dataset by generating new features from the existing ones. Its purpose is to transform the current features into a lower-dimensional space, thereby decreasing the overall computational cost, training time, and enhancing the model's accuracy. We adopted the ALBERT model as a feature extractor within the recommended deep neural network model in our research.

ALBERT model:

The ALBERT pre-trained model serves as a lightweight version of the renowned BERT model [5]. Given the computational intensity of the BERT base model, consisting of 110 million parameters, there was a demand for a more resource-efficient variant. ALBERT, with 12 million parameters, 768 hidden

layers, and 128 embedding layers, achieves a reduction in parameters while maintaining high performance.

ALBERT leverages two important strategies to achieve this parameter reduction. The first method is cross-layer parameter sharing, in which only the first encoder's parameter is learned and applied evenly across all encoders. This avoids parameter expansion as the network's depth increases. The second technique is factorized embedding parameterization, which uses factorization to lower the embedding layer from 768 to 128 layers. The size of the hidden layers is separated from the size of the vocabulary embedding by dividing the huge vocabulary embedding matrix into two smaller matrices, allowing for an increase in hidden size without considerably growing the parameter size of vocabulary embeddings. Beyond its lightweight nature, ALBERT distinguishes itself from BERT through Sentence Order Prediction (SOP). SOP involves classifying whether two given sentences are in the correct order, in contrast to BERT's next sentence prediction (NSP) during training. NSP entails the model predicting if the second sentence is the next in the original text.

The present study employed the ALBERT pre-trained model for feature extraction because to its tokenization strategy. To handle words that are not in the dictionary, the ALBERT tokenizer uses a data-driven methodology. It is built on a Sentence Piece tokenizer, which is a sub-word level tokenizer. Its byte-level methodology sets it apart from existing tokenization approaches that employ [UNK] or [OOV] tokens by enabling the encoding of complicated characters, like emojis, by representing them as combinations of many bytes. In order to increase the model's efficiency, ALBERT presents parameter reduction strategies. It makes use of cross-layer parameter sharing and shrinks the size of the concealed layers. Let X be the input text, shown as a word sequence. An embedding layer converts every word into a high-dimensional vector. If E represents the embedding layer, then

$$E(X) = [x_1, x_2, \dots, x_n],$$

Where n is the length of the sequence.

ALBERT employs a transformer architecture, consisting of multiple encoder layers. Each encoder layer consists of self-attention mechanisms and feedforward neural networks.

For a given layer l , the output H_l can be represented as

$$H_l = \text{TransformerLayer}(H_{l-1}),$$

where H_{l-1} is the output of the previous layer.

The model is able to assign varying weights to words in a sequence according to how relevant they are to one another thanks to the self-attention mechanism. The self-attention output is passed through a feedforward neural network to capture non-linear relationships.

Each sub-layer output is normalized and passed through a residual connection, aiding in the stable training of deep networks.

3.2.2. Classification phase

The deep learning techniques have been leveraged in this study to execute the classification task within the proposed model, with a specific emphasis on a neural network architecture based on the recurrent neural network (RNN). RNN networks' main advantage is their capacity to remember information from previous calculations and use it in the current one. As such, context dependencies in variable length inputs can be effectively modeled by RNN-based models, allowing for an efficient input composition. Nevertheless, architectures intended to tackle this difficulty have been introduced as a result of the well-documented vanishing gradient problem. Long Short-Term Memory is one such architecture (LSTM). Although LSTMs have demonstrated efficacy, their ability to interpret context in a single direction is limited. Bi-directional Long Short-Term Memory networks, commonly referred to as BLSTM (or Bi-LSTMs), have been developed to overcome this limitation. Unlike traditional LSTM models that process input sequences in only one direction, BLSTMs utilize two separate models during training. The first model learns from the input sequence in its original order, while the second model processes the sequence in reverse. This approach effectively captures contextual information from both the past and the present, ensuring that each segment of an input sequence benefits from comprehensive information.

Overfitting is a common challenge in deep neural networks, manifesting when a model excels on training data but performs poorly on testing data. Essentially, overfitting occurs when a model endeavors to perfectly match the training data, resulting in memorization of data patterns, noise, and random fluctuations. To mitigate overfitting, various techniques can be employed. One straightforward approach involves the utilization of dropout. Dropout serves as a regularization technique designed to prevent neural networks from succumbing to overfitting. The fundamental concept is to randomly deactivate some neurons during training. Standard dropout involves zeroing out a fraction of nodes in every layer before computing the next layer throughout each training iteration. This random dropout prevents the network from becoming overly reliant on certain nodes, allowing for more robust generalization to previously unknown data.

3.2.3. Tuning the parameters of the proposed model

Hyperparameters play a pivotal role in shaping the performance of deep learning models, underscoring the importance of optimizing them to attain optimal results. Various strategies exist to tackle hyperparameter optimization, each offering distinct advantages.

Grid Search, the most straightforward approach, partitions the hyperparameter space into a discrete grid and exhaustively evaluates each combination via cross-validation to gauge performance metrics. Random Search, akin to Grid Search, samples a random subset of grid points, often proving more effective, especially when optimizing with a limited number of hyperparameters.

Bayesian Optimization takes a probabilistic approach by constructing a model that maps hyperparameter values to performance metrics on a validation set. This method iteratively refines its understanding of the hyperparameter space, making informed decisions to explore promising regions efficiently.

Gradient-Based optimization involves computing gradients with respect to hyperparameters and optimizing them using gradient descent. Leveraging gradient information allows for efficient navigation of the hyperparameter space towards regions of improved performance.

Evolutionary Optimization harnesses evolutionary algorithms to explore the hyperparameter space. By emulating principles of natural selection, this method evolves a population of hyperparameter configurations over successive generations, favoring those exhibiting higher performance.

In our study, we opted for the random search method to fine-tune the hyperparameters of our proposed model due to its computational efficiency. Employing this approach, we optimized multiple hyperparameter values and selected the model yielding the most promising results. Detailed insights into the tuned hyperparameters and configurations are elaborated upon in the experimental results section. Figure 1 illustrates the suggested model's block diagram and Algorithm 1 illustrates the detailed process of the proposed method.

Algorithm 1: Spam detection Algorithm

Input: Raw text data R from social networking sites
Output: Probability score indicating the likelihood of spam.
1: For each raw in R do:
 Clean_Text = Remove_URL (raw)
 Clean_Text = Remove_Punctuation (Clean_Text)

```

Clean_Text = Remove_Emoji (Clean_Text)
Clean_Text = Remove_Char (Clean_Text)
Clean_Text = Remove_Stop_Words(Clean_Text)
Clean_Text = Convert_To_Lower_Case (Clean_Text)
End for
2: Function extract_features(Clean_Text):
albert_tokenizer = AlbertTokenizer.from_pretrained('albert-base-v2')
albert_model = AlbertModel.from_pretrained('albert-base-v2')
tokenized_text = albert_tokenizer(Clean_Text)
output = albert_model(tokenized_text)
return output
3: Function hyperparameter_tuning():
hyperparameters = {num_lstm_layers,dropout_existence, dropout_value , learning_rate}
best_hyperparameters = random_search(hyperparameters)
return best_hyperparameters
4:Function Split_dataset (output,Ratio,train,test)
return train,test
5:Train model (model,train, best_hyperparameters)
6:Evaluate_model(model,test, best_hyperparameters)
End

```

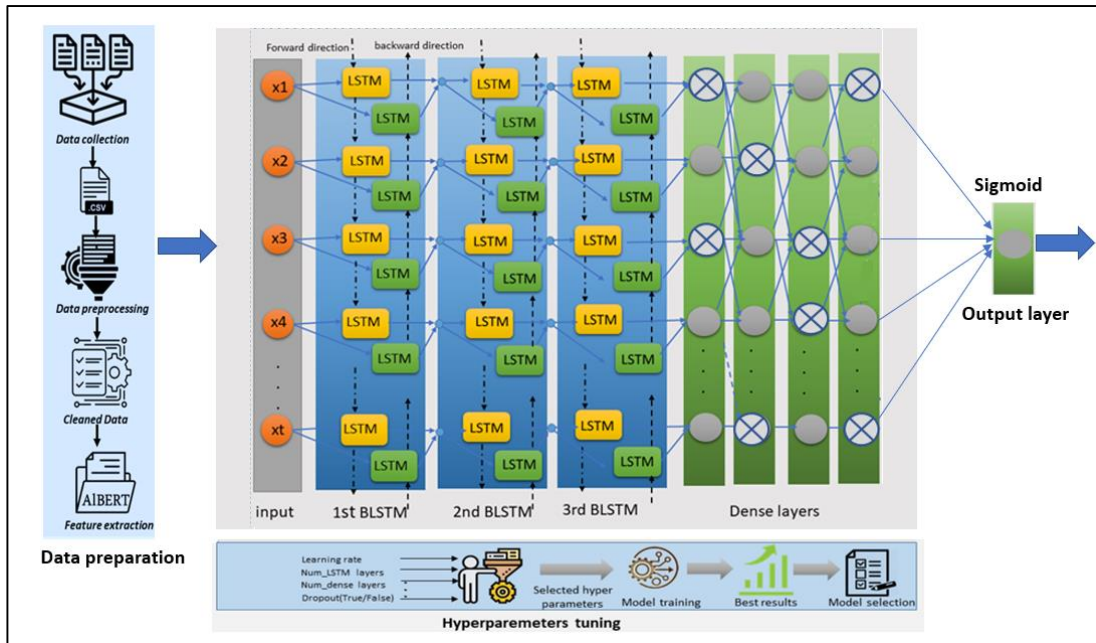


Figure 1. The block diagram of the suggested model

4. Experimental Results

All studies were conducted out with Google Colaboratory, Python3 and GPU runtime.

4.1. Evaluation metrics

The accuracy, Recall, F1-score, and precision metrics have been used as evaluation metrics in this study.

The accuracy metric is the main metric used to evaluate the model's performance. It describes the number of correct predictions over all predictions and can be computed by the following equations:

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \quad (3)$$

Precision is a measure of how many of the positive predictions made are correct (true positives) and can be computed by the following formula

$$Precision (P) = TP/(TP + FP) \quad (4)$$

Recall, expressed in Equation 3, quantifies the classifier's accuracy in correctly identifying positive cases among all positive instances in the dataset.

$$\text{Recall (R)} = \frac{TP}{TP + FN} \quad (5)$$

F1-Score is a measure combining both precision and recall and can be computed by the following formula:

$$\text{F1 - Score} = (2 * P * R) / (P + R) \quad (6)$$

Where TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives. Here, TP stands for the total number of messages correctly classified as spam. TN displays the number of messages that have been correctly classified as non-spam. FP represents the number of non-spam messages that are labeled as spam, while FN represents the number of spam messages that are labeled as non-spam. It is worth mentioning that in the evaluation of our spam detection model, we employed a range of metrics to comprehensively assess its performance across different datasets. While metrics such as accuracy, precision, and recall provide valuable insights into overall model effectiveness, we acknowledge the significance of considering class imbalance, particularly in the context of the SMS spam dataset.

Given the imbalance between spam and non-spam classes in the SMS dataset, we recognized the importance of using evaluation metrics that are robust to such disparities. As such, we employed the F1-score as a primary evaluation metric alongside other standard metrics. The F1-score combines precision and recall, offering a balanced assessment of model performance that is less affected by class imbalance.

4.2. Hyperparameters tuning

Herein this study, we used a random search strategy in order to tune each of the following hyperparameters: the number of BLSTM layers, dropout existence, dropout value, number of dense layers, number of neurons in each of dense layers, learning rate, and weight initializer, and activation functions. The value range of each of these hyperparameters is illustrated in Table 2. The random search algorithm was tuned to do 20 trials with four executions in each trial, and in each execution, the model will be trained for 5 epochs. The validation accuracy has been monitored during the algorithm execution and we have defined our objective function to select the maximum value of the validation accuracy. Table 3 shows the configuration parameters required to adopt the random search technique.

Table 2 Experimented hyperparameteres values during tuning

Hyperparameter	Value range
Number of the BLSTM layers	1-6
Dropout existence	[True, False]
Dropout value	[0.1,0.2,0.3,0.4,0.5]
Number of the Dense layers	1-4
Number of neurons in each the Dense layer	[32, 64, 96, 128]
Learning rate	[0.01, 0.001, 0.0001]
weight initializer	['uniform', 'lecun_uniform', 'normal', 'zero', 'glorot_normal', 'glorot_uniform', 'he_normal', 'he_uniform']
Activation functions	[relu, tanh]

Table 3 Configuration settings used in random search

Number of trials	20
Executions per trial	4
Epochs per trial	5

In the final phase, the best model was identified, achieving optimal validation accuracy across the

three datasets. This selected model comprises three Bidirectional Long Short-Term Memory layers, each with 256 units, followed by four dense layers, each

consisting of 32 units. A dropout is introduced, and the output layer concludes the architecture. The chosen dropout value is 0.2, the weight initializer is set to 'he_uniform', and the Rectified Linear Unit (ReLU) function serves as the activation function in the dense layers.

Figure 2 is a part of the parallel coordinates view illustrating the hyperparameter fine-tuning process using random search. The tuned hyperparameters are shown on the horizontal axes, while the corresponding selected values are shown on the vertical axes. This visual representation illustrates the results of the random search hyperparameter tuning process. Each line in the plot represents a different hyperparameter configuration, with the axes corresponding to the hyperparameters being tuned. The lines show how each hyperparameter configuration performs across multiple evaluation metrics. This visualization allows for a comprehensive understanding of how different combinations of hyperparameters impact model performance. Figure 3 also shows a preview of the random search algorithm's scatter plot matrix view, where each dot represents a trial. This visual depiction provides a comprehensive view of the relationships between pairs of hyperparameters during the random search process. Each point in the scatter plot matrix represents a hyperparameter configuration, with different plots showing the

relationships between different pairs of hyperparameters. This visualization aids in identifying any correlations or patterns between hyperparameters, which can inform further optimization strategies. Figure 4 depicts the accuracy and loss plots for the best trial recorded during hyperparameter tweaking, with the three trials of the best experiment marked by the lines on the charts. This visual representation showcases the results of the hyperparameter tuning experiments conducted specifically on the YouTube dataset. Chart a illustrates the validation accuracy achieved by different hyperparameter configurations, while chart b displays the corresponding validation loss. These charts provide insights into how different hyperparameter settings influence the model's performance on the YouTube dataset, allowing for informed decision-making in selecting the optimal hyperparameter configuration. Table 4 summarizes the results of the random search in detail. The table presents the results of the hyperparameter tuning experiments conducted specifically on the YouTube dataset. Each row represents a different trial ID, with columns detailing the specific hyperparameters explored during the tuning process. The columns include information such as the dropout rate, the number of dense layers, the number of units in each dense layer, the number of BLSTM layers, the learning rate, and metrics such as train accuracy, validation accuracy, train loss, and validation loss.

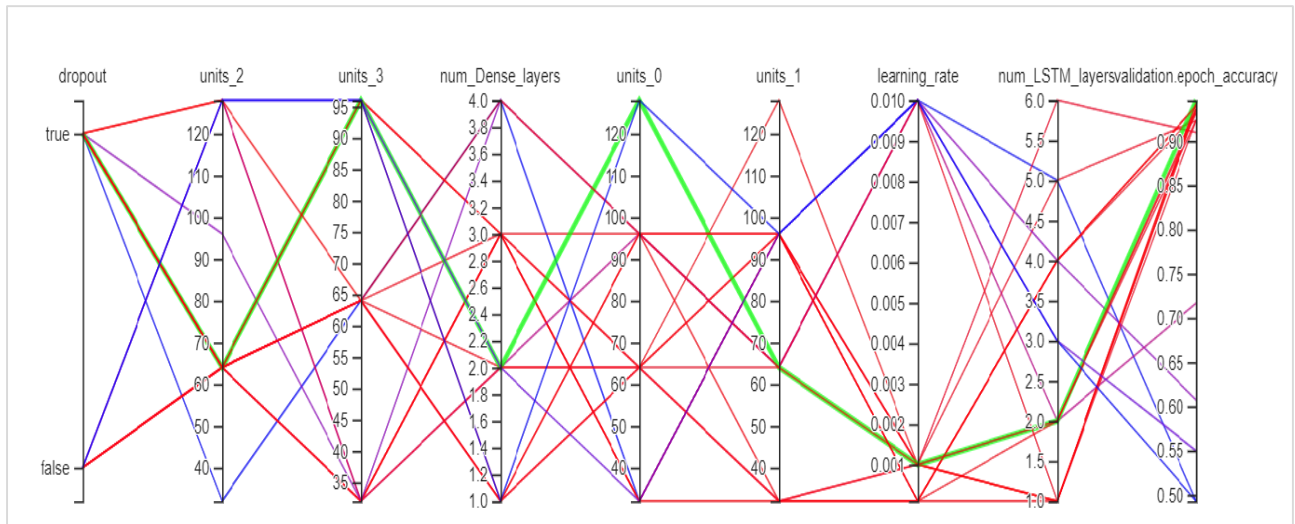


Figure 2. The parallel coordinates view for the conducted random search hyperparameters fine tuning

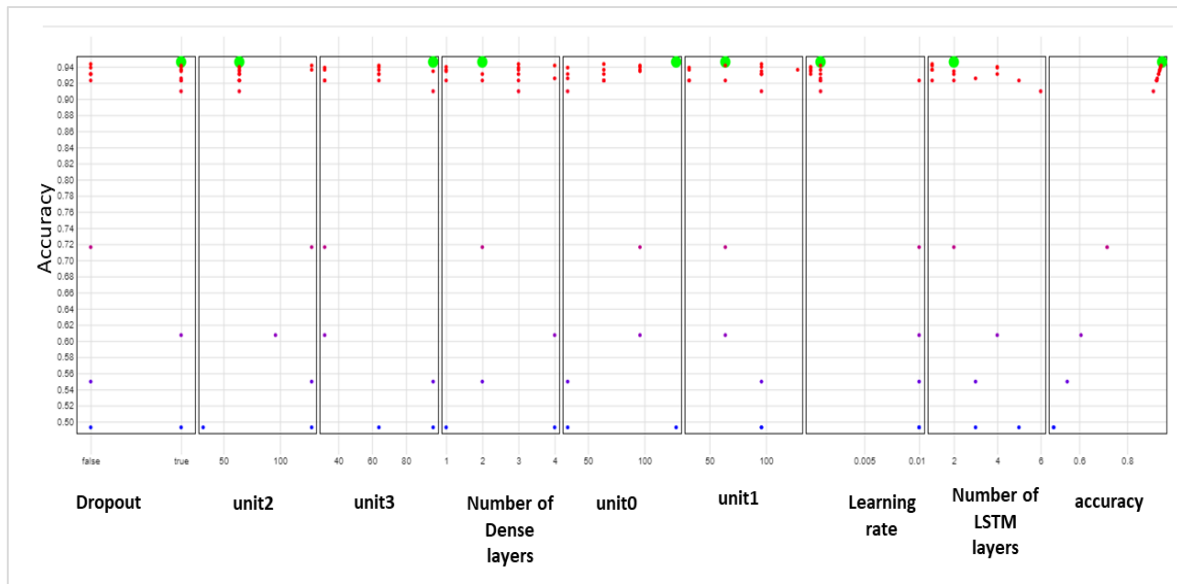


Figure 3. The scatter plot matrix view for the conducted random search

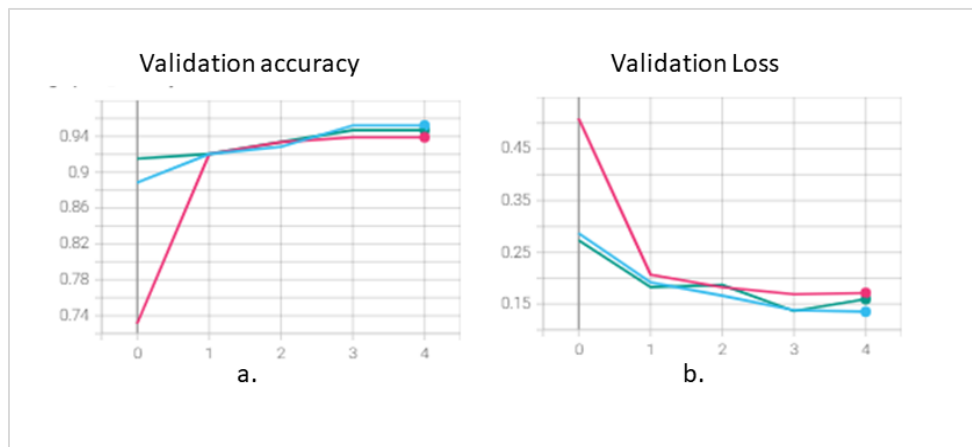


Figure 4 The best hyper parameters tuning experiments' results on YouTube dataset. chart a. represents the validation accuracy and chart b. represents validation loss

Table 4 hyper parameters tuning experiments' results on YouTube dataset

Trial ID	dropout	# of Dense layers	units_0	Unit_1	units_2	units_3	# of BLSTM layers	Learning rate	Train accuracy	validation accuracy	Train loss	Validation Loss
0	true	4	32	64	-	-	3	0.001	0.92516	0.92553	0.19694	0.19296
1	true	4	32	96	32	64	5	0.0001	0.50056	0.49291	0.70626	0.70383
2	true	3	32	96	64	-	6	0.001	0.87653	0.90957	0.34381	0.2838
3	false	3	64	32	64	-	1	0.01	0.95692	0.94326	0.11603	0.18043
4	true	3	64	64	64	-	5	0.01	0.92294	0.92287	0.21103	0.19901
5	false	2	64	64	-	-	4	0.01	0.94559	0.93085	0.15676	0.19206
6	false	3	32	32	64	-	2	0.001	0.94448	0.93085	0.15121	0.19475
7	false	3	32	32	64	-	4	0.01	0.94426	0.93883	0.16144	0.17078
8	true	3	96	64	128	-	1	0.001	0.94848	0.93617	0.14338	0.19642
9	true	1	64	-	-	-	1	0.001	0.94559	0.93617	0.14904	0.17147
10	true	4	96	32	96	32	4	0.01	0.61115	0.60727	0.64572	0.64864
11	false	2	96	96	-	-	2	0.0001	0.69776	0.71631	0.58824	0.58255
12	true	2	128	64	-	-	2	0.001	0.94581	0.94592	0.15504	0.15537
13	true	4	32	32	32	32	3	0.001	0.9853	0.9521	0.0470	0.1944
14	true	1	96	-	-	-	4	0.001	0.94715	0.93972	0.14341	0.19378
15	true	4	96	64	128	64	1	0.001	0.95514	0.94149	0.12171	0.1697
16	true	1	64	-	-	-	2	0.001	0.92449	0.92287	0.26203	0.26102
17	true	1	96	-	-	-	2	0.001	0.95203	0.9344	0.13678	0.18747
18	false	2	32	-	-	96	3	0.01	0.55008	0.54965	0.66753	0.65661
19	false	1	128	-	-	-	3	0.01	0.501	0.49291	0.69329	0.69345

4.3. Results and discussion

After employing the random search approach to determine the optimum hyperparameters that can be used for constructing a deep learning model that can give us the best results for the previously mentioned three datasets, we have gotten the model that has a structure illustrated in table 5. Then, after training the model for 10 epochs, the best results are reached as shown in Table 6. Furthermore, Figure 3 illustrates confusion matrix obtained by applying the proposed model on the used datasets.

According to the results, our proposed approach, which combines contextualized embedding (using ALBERT pretrained model as a feature extractor) with an BLSTM-based deep learning architecture, consistently outperforms the benchmark datasets. As depicted in Table 6, our model achieved impressive accuracies, surpassing 98%, 95%, and 99% on the Twitter, YouTube, and SMS datasets, respectively.

The tuning of hyperparameters proves instrumental in optimizing the performance of our proposed model. In this study, through meticulous hyperparameter tuning, we found that optimal performance is achieved by keeping the number of BLSTM layers at or below 3. Additionally, the optimal learning rate was determined to be 0.001, and a dropout value of 0.2 consistently led to superior results. In the dense layers, the ReLU function was chosen as the activation function during experiments. Given the binary classification nature of our problem, the default activation function in the output layer was the sigmoid function.

Finally, based on the Comparison study results on Table 7, the experimental results unequivocally demonstrated that the ALBERT-BLSTM based model surpassed existing methods from reference studies. These findings underscore the critical role of word representation in significantly improving classification performance, especially when dealing with short texts in social networking sites.

Table 5. The structure of the best model

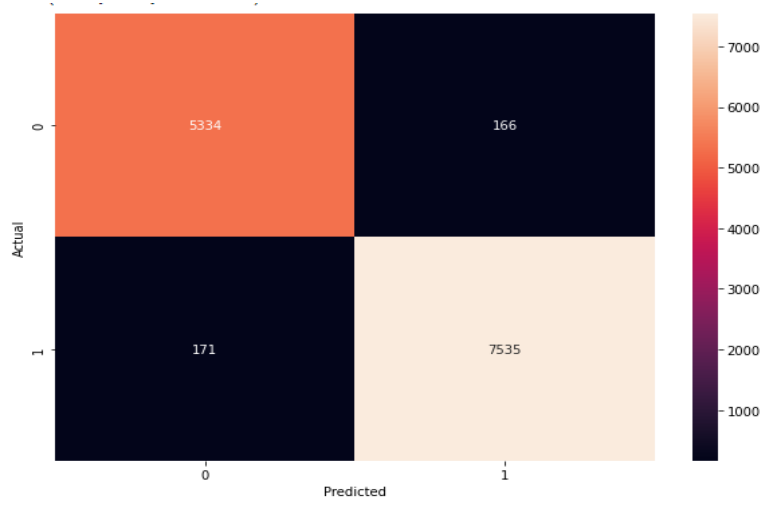
Parameter	Value
# of BLSTM layers	3
# of units in BLSTM layers	256→256→256
# of Dense layers	4
# of units in Dense layers	64→32→32→32
Activation function	ReLU
Weight initializer	he_uniform
Dropout	True
Dropout value	0.2
Output activation function	Sigmoid
Learning rate	0.001
Optimizer	Adam
Loss function	Binary crossentropy
Batch size	128
# of epochs	10

Table 6. Final results on the three datasets

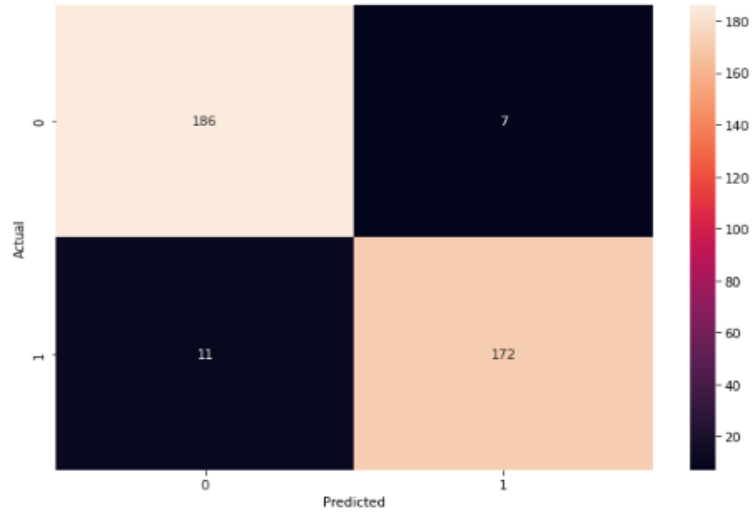
Dataset	Accuracy	Precision	Recall	F1-score
Twitter	0.982	0.98	0.98	0.98
YouTube	0.9548	0.96	0.94	0.95
SMS	0.9912	0.98	0.97	0.97

Table 7. Comparison with other Studies

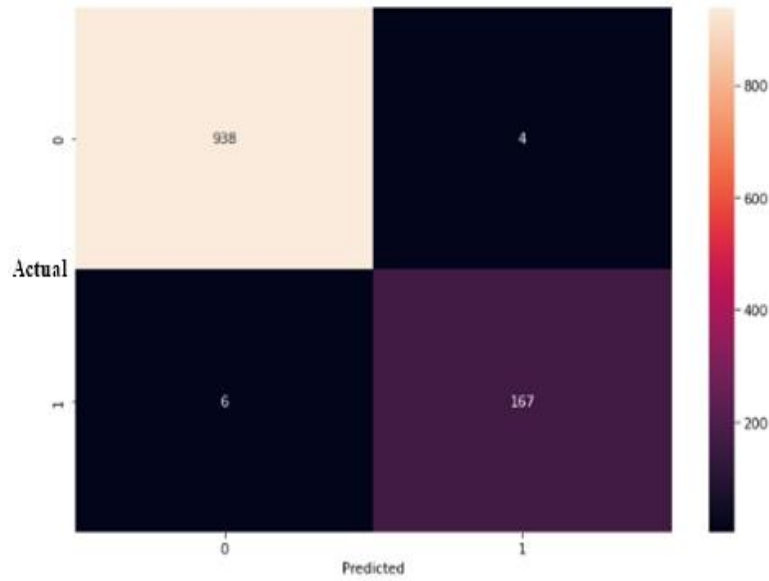
Study	Dataset	Accuracy
[28]	Twitter	0.975
Barushka et. al [29]	Twitter	0.956
[20]	Twitter	0.978
Jain et.al [32]	Twitter	0.9343
Our method	Twitter	0.982
Ghanem and Erbay [20]	SMS	0.99
Jain et.al [32]	SMS	0.978
Our method	SMS	0.99
Nagaraj et.al [33]	YouTube	0.93
[20]	YouTube	0.9521
Our method	YouTube	0.9548



a.



b.



c. **Figure 5.** The confusion matrices obtained by applying the proposed model on the respective datasets: a. The confusion matrix for the Twitter dataset. b. The confusion matrix for the YouTube dataset. c. The confusion matrix for the SMS dataset.

5- CONCLUSION

In conclusion, the pervasive presence of spam messages on social networks poses a significant challenge, particularly in the realm of short text classification. The inherent brevity and ambiguity of text within social media platforms add layers of complexity to the task of spam detection in natural language processing. Acknowledging the pivotal role of robust text representation, our study introduced an innovative approach that integrates contextualized models and a deep learning architecture.

To Avoid the limitations in conventional word embedding models, characterized by struggles with data sparsity, out-of-vocabulary challenges, and context independence, our research proposed a novel methodology for spam detection. Leveraging contextualized word representations derived from the pre-trained ALBERT model, we augmented the performance of a stacked Bidirectional Long Short-

ACKNOWLEDGMENT

The authors are grateful to the UCI repository, Chen et al. [26], Alberto et al. [13], for contributing their datasets.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that there is no conflict of interest.

Term Memory (BLSTM) based neural network architecture. This approach demonstrated efficacy in accurately classifying messages as either spam or non-spam, effectively navigating the intricate nuances of short text in social networks.

Furthermore, hyperparameter tuning played a crucial role in optimizing the proposed model's performance across three benchmark datasets. The utilization of the random search algorithm facilitated the fine-tuning of multiple hyperparameters, including the number of BLSTM layers, dense layers, neurons in each dense layer, dropout presence, activation function, and learning rate.

Looking ahead, the proposed method holds promise for applications in various trending studies such as sentiment analysis, authorship identification, hate speech detection, review spam detection, and more. Future endeavors could delve into these domains to further validate the versatility and effectiveness of the introduced approach.

REFERENCES

- [1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
- [2] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.
- [3] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," arXiv preprint arXiv:1607.01759, 2016.
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
- [5] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," arXiv preprint arXiv:1909.11942, 2019.
- [6] A. Al-Zoubi, J. Alqatawna, H. Faris, and M. A. Hassonah, "Spam profiles detection on social networks using computational intelligence methods: the effect of the lingual context," *J Inf Sci*, vol. 47, no. 1, pp. 58–81, 2021.
- [7] D. Niranjana Kogalahewa, Y. Xu, and E. Foo, "Spam detection in social networks based on peer acceptance," in Proceedings of the Australasian Computer Science Week Multiconference, 2020, pp. 1–7.
- [8] K. S. Adewole, T. Han, W. Wu, H. Song, and A. K. Sangaiah, "Twitter spam account detection based on clustering and classification methods," *J Supercomput*, vol. 76, pp. 4802–4837, 2020.
- [9] A. Kumar, M. Singh, and A. R. Pais, "Fuzzy string matching algorithm for spam detection in Twitter," in Security and Privacy: Second ISEA International Conference, ISEA-ISAP 2018, Jaipur, India, January, 9–11, 2019, Revised Selected Papers 2, Springer, 2019, pp. 289–301.
- [10] O. ÇITLAK, M. DÖRTERLER, and İ. DOGRU, "A hybrid spam detection framework for social networks," *Politeknik Dergisi*, p. 1, 2023.
- [11] C. Kumar, T. S. Bharti, and S. Prakash, "A hybrid Data-Driven framework for Spam detection in Online Social Network," *Procedia Comput Sci*, vol. 218, pp. 124–132, 2023.
- [12] A. Aziz, C. F. M. Foozy, P. Shamala, and Z. Suradi, "YouTube spam comment detection using support vector machine and K-nearest neighbor," *Indones. J. Electr. Eng. Comput. Sci*, vol. 12, no. 2, p. 612, 2018.
- [13] A. Ali and M. Amin, "An Approach for Spam Detection in YouTube Comments Based on Supervised Learning," 2016.
- [14] A. T. Kabakus and R. Kara, "'TwitterSpamDetector': a spam detection framework for Twitter," *International Journal of Knowledge and Systems Science (IKSS)*, vol. 10, no. 3, pp. 1–14, 2019.
- [15] P. Nagaraj, K. M. Sudar, P. Thrived, P. G. K. Reddy, S. B. Babu, and P. S. R. Krishna, "Youtube Comment Spam Detection," in 2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2023, pp. 1–6.
- [16] H. Valpadasu, P. Chakri, P. Harshitha, and P. Tarun, "Machine Learning based Spam Comments Detection on YouTube," in 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 2023, pp. 1234–1239.
- [17] L. Shabadi, P. Srikanth, V. Kumar, and U. Kashyap, "Youtube Spam Detection Scheme Using Stacked Ensemble Machine Learning Model," in 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), IEEE, 2023, pp. 1–7.
- [18] T. C. Alberto, J. V. Lochter, and T. A. Almeida, "Tubespam: Comment spam filtering on youtube," in 2015 IEEE 14th international conference on machine learning and applications (ICMLA), IEEE, 2015, pp. 138–143.
- [19] T. Wu, S. Liu, J. Zhang, and Y. Xiang, "Twitter spam detection based on deep learning," in Proceedings of the australasian computer science week multiconference, 2017, pp. 1–8.
- [20] R. Ghanem and H. Erbay, "Spam detection on social networks using deep contextualized word representation," *Multimed Tools Appl*, vol. 82, no. 3, pp. 3697–3712, 2023.
- [21] R. Ghanem and H. Erbay, "Context-dependent model for spam detection on social networks," *SN Appl Sci*, vol. 2, pp. 1–8, 2020.
- [22] R. Ghanem, H. Erbay, and K. Bakour, "Contents-Based Spam Detection on Social Networks Using RoBERTa Embedding and Stacked BLSTM," *SN Comput Sci*, vol. 4, no. 4, p. 380, 2023.
- [23] H. Bakir and G. Tarihi, "Using Transfer Learning Technique as a Feature Extraction Phase for Diagnosis of Cataract Disease in the Eye."
- [24] H. BAKIR, S. OKTAY, and E. TABARU, "DETECTION OF PNEUMONIA FROM X-RAY IMAGES USING DEEP LEARNING TECHNIQUES," *Journal of Scientific Reports-A*, no. 052, pp. 419–440, Mar. 2023, doi: 10.59313/jsr-a.1219363.
- [25] H. Bakır, "Evaluating the impact of tuned pre-trained architectures' feature maps on deep learning model performance for tomato disease detection," *Multimed Tools Appl*, pp. 1–22, 2023.
- [26] H. Bakır and R. Bakır, "DroidEncoder: Malware detection using auto-encoder based feature extractor and machine learning algorithms," *Computers and Electrical Engineering*, vol. 110, p. 108804, 2023.
- [27] E. Doğan and H. BAKIR, "Hiperparametreleri Ayarlanmış Makine Öğrenmesi Yöntemleri Kullanılarak Ağdaki Saldırıların Tespiti," in *International Conference on Pioneer and Innovative Studies*, 2023, pp. 274–286.

- [28] S. Madisetty and M. S. Desarkar, "A neural network-based ensemble approach for spam detection in Twitter," *IEEE Trans Comput Soc Syst*, vol. 5, no. 4, pp. 973–984, 2018.
- [29] A. Barushka and P. Hajek, "Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks," *Neural Comput Appl*, vol. 32, pp. 4239–4257, 2020.
- [30] G. Jain, M. Sharma, and B. Agarwal, "Spam detection in social media using convolutional and long short term memory neural network," *Ann Math Artif Intell*, vol. 85, no. 1, pp. 21–44, 2019.
- [31] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "A study on real-time low-quality content detection on Twitter from the users' perspective," *PLoS One*, vol. 12, no. 8, p. e0182487, 2017.
- [32] G. Jain, M. Sharma, and B. Agarwal, "Spam detection in social media using convolutional and long short term memory neural network," *Ann Math Artif Intell*, vol. 85, no. 1, pp. 21–44, 2019.
- [33] P. Nagaraj, K. M. Sudar, P. Thrived, P. G. K. Reddy, S. B. Babu, and P. S. R. Krishna, "Youtube Comment Spam Detection," in *2023 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, 2023, pp. 1–6.

A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules

Araştırma Makalesi/Research Article

 Şefik İlkin SERENGİL¹,  Alper ÖZPINAR²

¹ Solution Engineering Department, Vorboss Limited, London, UK

² Mechatronics Engineering Department, Istanbul Ticaret University, Istanbul, Turkey

sefik.serengil@vorboss.com, alper.ozpinar@ticaret.edu.tr

(Geliş/Received:01.12.2023; Kabul/Accepted:29.03.2024)

DOI: 10.17671/gazibtd.1399077

Abstract— Researchers from leading technology companies, prestigious universities worldwide, and the open-source community have made substantial strides in the field of facial recognition studies in recent years. Experiments indicate that facial recognition approaches have not only achieved but surpassed human-level accuracy. A contemporary facial recognition process comprises four key stages: detection, alignment, representation, and verification. Presently, the focus of facial recognition research predominantly centers on the representation stage within the pipelines. This study conducted experiments exploring alternative combinations of nine state-of-the-art facial recognition models, six cutting-edge face detectors, three distance metrics, and two alignment modes. The co-usability performances of implementing and adapting these modules were assessed to precisely gauge the impact of each module on the pipeline. Theoretical and practical findings from the study aim to provide optimal configuration sets for facial recognition pipelines.

Keywords— facial recognition, face verification, deepface, vector models

Yüz Tanıma Üretim Hatlarının ve Modüllerinin Birlikte Kullanımının Karşılaştırılması

Özet— Son yıllarda, büyük teknoloji şirketleri, dünyanın önde gelen üniversiteleri ve açık kaynak topluluğundan gelen araştırmacılar, yüz tanıma alanında önemli ilerlemeler kaydetmişlerdir. Yapılan deneyler, yüz tanıma yaklaşımlarının insan düzeyinde doğruluk sağladığını ve hatta aştığını göstermektedir. Modern bir yüz tanıma süreci genellikle dört aşamadan oluşur: algılama, hizalama, temsil ve doğrulama. Mevcut yüz tanıma çalışmaları genellikle üretim hatlarındaki temsil aşamasına odaklanmıştır. Bu çalışmada, dokuz farklı son teknoloji yüz tanıma modeli, altı son teknoloji yüz dedektörü, üç mesafe ölçümü ve iki hizalama modunun farklı kombinasyonları için çeşitli deneyler gerçekleştirilmiştir. Bu modüllerin uygulanması ve uyumlu hale getirilmesinin genel performansları, her bir modülün üretim hattındaki spesifik etkisini belirlemek amacıyla değerlendirilmiştir. Çalışmanın teorik ve pratik sonucu olarak, yüz tanıma hatları için en iyi konfigürasyon setlerinin paylaşılması hedeflenmektedir.

Anahtar Kelimeler— yüz tanıma, yüz doğrulama, deepface, vektör modeller

1. INTRODUCTION

In recent years, researchers and open-source communities have made substantial advancements in the field of facial recognition. This area of research has garnered attention from major technology companies and top universities worldwide. Experiments indicate that machine learning-based facial recognition approaches developed by these entities have not only achieved but surpassed human-level accuracy.

A widely accepted perspective among facial recognition researchers advocates for the incorporation of four fundamental stages in contemporary facial recognition pipelines [1]: detection, alignment, representation, and verification. Notably, the representation stage stands out as the most crucial phase, tasked with generating vector embeddings from facial images. Presently, state-of-the-art models for the representation module predominantly leverage convolutional neural networks, exemplified by models like FaceNet [2], VGG-Face [3], ArcFace [4], Dlib [5], Sface [6], OpenFace [7], DeepFace [1], and DeepId [8].

Following the representation stage, the verification module calculates distances between vector embeddings of face pairs. Experiments reveal that the distribution of distances for pairs belonging to the same person and different individuals is discrete, particularly in robust models [9]. Various metrics, such as Euclidean, L2 normalized Euclidean form, or cosine similarity, can be employed to compute the distances between vectors.

Prior to the representation stage, both detection and alignment serve as early modules in the pipeline, aiming to provide clear inputs. Substantial progress has been made in recent years in the field of face detection. Noteworthy face detectors include OpenCV Haar Cascade [10], SSD [11], MediaPipe [12] [13], MTCNN [14], Dlib HOG [5], and RetinaFace [15]. Some of these detectors are built on modern convolutional neural networks, such as MTCNN, RetinaFace, and SSD, while others adhere to legacy approaches like OpenCV and Dlib HOG. Additionally, facial landmarks, including the coordinates of the eyes, are provided by these detectors, facilitating the horizontal alignment of faces.

In the realm of facial recognition studies, there is, regrettably, a lack of uniformity as researchers and model creators approach the problem from diverse perspectives. The primary emphasis tends to be on the representation stage within the pipeline, where various combinations of face detectors, alignment procedures, and distance metrics are employed. This variability introduces a challenge in fairly comparing these models, as a facial recognition model may demonstrate superior performance when configured differently, potentially exceeding its initially reported score. The precise influence of detection, alignment, and verification modules on the overall pipeline remains to be fully understood.

This study assesses the efficacy of facial recognition pipelines across four key dimensions: facial recognition models, face detectors, distance metrics, and the impact of enabling or disabling alignment modes. The objective is to identify optimal configuration settings for the pipeline. The paper further seeks to elucidate the tangible influence of each module on the overall pipeline score. Ultimately, the study compares the performance of human subjects to that of different models and configurations, determining which ones surpassed, matched, approached, or fell short of human-level accuracy.

In this context, LightFace [9] [16] emerges as one of the most widely used facial recognition libraries for the Python programming language. Comprehensive in its coverage, it includes various facial recognition models, face detectors, distance metrics, and alignment modes, seamlessly managing all facial recognition procedures in the background. For the purposes of this study, the LightFace package was utilized and implemented to execute a range of experimental setups.

2. DATASET

The Labeled Faces in the Wild (LFW) dataset [17] stands as the predominant dataset in facial recognition studies, comprising 13,233 images belonging to 5,749 identities. For the purpose of this study, the LFW dataset was loaded using scikit-learn [18]. The dataset includes 2,200 pairs for training, 6,000 pairs for 10-fold cross-validation, and 1,000 pairs for testing. Each pair is appropriately labeled as either depicting the same person or different individuals. The original size of images in LFW is (250, 250, 3). Given the focus on pre-trained models in this study, attention is directed solely to the test set.

Half of the pairs of recommended test sets belong to the same person, whereas another half of the pairs correspond to different persons. So, the test set has fairly homogeneous target labels. That is why accuracy scores are mentioned in the performance results. The raw unaligned LFW images were used because detection and alignment impact in the pipeline were evaluated.

There are successful studies in the literature with good results in capturing information from the context such as image background or hairstyle. Neeraj Kumar et al [19] reported the performance of human-beings on LFW dataset. They collected the classification votes of 10 voters for each pair in LFW dataset consisting of 6000 instances via Amazon Mechanical Turk. Besides, voters taken three different tests: verifying original image pairs, tight crop images and inverse crops as shown in Figure 1.



Figure 1. Test Tasks

Humans are able to verify uncropped face pairs with 99.2% accuracy score. They also obtain 94.2% accuracy even for image pairs on same dataset with blurred or blacked facial areas. On the other hand, human-beings have 97.5% accuracy for verifying cropped face pairs. So, people can find out a lot information from the context such as background. Regions outside of the facial area is not strongly recommended to be used not to boost results artificially. That is why, facial recognition researchers always compare built model performances with this score.

3. PIPELINE

A modern facial recognition pipeline consists of 4 common stages: detection, alignment, representation and verification. This section outlines the current procedures in place for these modules.

3.1. Detection and Alignment

Only open source face detectors were used in the detection module. There are five of them: OpenCv, Ssd, MtCnn, Dlib, and RetinaFace. Their role will be to detect faces and discard regions outside of the facial area in order to avoid artificially boosting results.

An alignment step is also performed after detection. The eye locations can then be used to construct a right-angled triangle with two corners at eye locations as shown in Figure 2.



Figure 2. Alignment Procedure

Calculating the cosine of an angle is then done using the cosine rule in Formula 1. Inverse cosine of this term

returns angle in radian. The angle can then be converted to degrees using the formula 2. The image will then be rotated by A degrees until the eyes are horizontal.

$$\cos(A) = \frac{b^2+c^2-a^2}{2bc} \tag{1}$$

$$A^\circ = \arccos\left(\frac{b^2+c^2-a^2}{2bc}\right) \frac{180}{\pi} \tag{2}$$

3.2. Resizing

The outcomes of the detection and alignment modules serve as inputs for the representation module. In facial recognition pipelines, the representation module predominantly relies on convolutional neural networks (CNNs). CNNs typically require inputs to have a fixed shape, yet detected and aligned facial images lack such uniformity. Table 1 provides an overview of the anticipated architectures for the facial recognition models discussed in this study including input shape, embedding dimensions, total parameters (weights and biases) and number of layers. Consequently, resizing the detected and aligned faces becomes imperative before invoking the representation module.

Table 1. Facial Recognition Model's Architectures

Model	Input Shape	Embedding Dimensions	Total Params	Num of Layers
FaceNet-128d	160, 160, 3	128	22M	447
FaceNet-512d	160, 160, 3	512	23M	447
Vgg-Face	224, 224, 3	4096	134M	36
ArcFace	112, 112, 3	512	34M	162
Dlib	150, 150, 3	128	63M	34
SFace	112, 112, 3	128	9M	88
OpenFace	96, 96, 3	128	3M	166
DeepFace	152, 152, 3	4096	102M	9
DeepId	55, 47, 3	160	395K	17

In order to prevent deformation, black padding pixels were introduced to the images instead of resizing them. Figure 3 visually depicts the outcomes of detection and alignment using various detectors, along with the addition of black padding pixels to resize the image to the expected input shape of VGG-Face.



Figure 3. Detection, Alignment and Resizing

3.3. Representation

The representation module of all the facial recognition models discussed in this study is based on convolutional neural networks. In spite of the fact that CNN structures have many output nodes, units were not used for classification tasks. Rather than finding the most dominant output node to make classification, the probabilities of output nodes were used. Figure 4 shows how DeepFace [1] model finding vector embeddings.

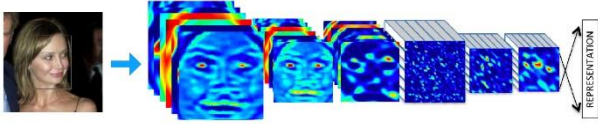


Figure 4. Vector Representation with DeepFace [1]

3.4. Verification

The representation module returns vectors for face pairs, and these vectors are passed to the verification module. The distance between those vectors was calculated in the verification module. Formula 3 shows the euclidean distance formula where p and q are n-dimensional vectors. Similarly, Formula 4 finds the cosine distance where p and q are n-dimensional vectors.

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (3)$$

$$d(p, q) = 1 - \frac{\sum_{i=1}^n p_i q_i}{\sqrt{\sum_{i=1}^n p_i^2} \sqrt{\sum_{i=1}^n q_i^2}} \quad (4)$$

Before feeding the vectors to distance formulas, l_2 were applied for normalization to vectors to have smoother ones as shown in Formula 5 where q is a n-dimensional vector.

$$l_2(q) = \frac{q}{\sqrt{\sum_{i=1}^n q_i^2}} \quad (5)$$

Accordingly, this pair would be classified as the same person if the distance was less than the threshold. Similarly, this pair could be classified as different persons if the distance between them exceeds the threshold value.

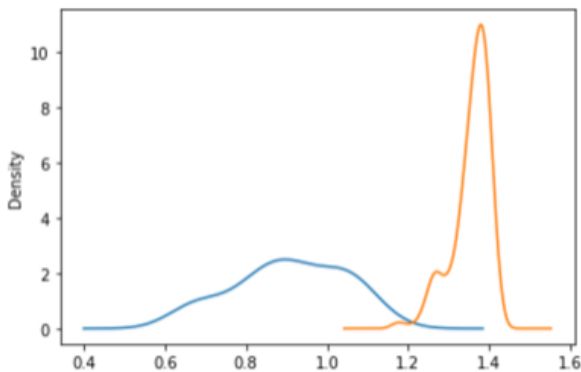


Figure 5. Distance Distribution

The experiments demonstrate that the distance distributions for the same person and different persons classes, as tested with the VGG-Face facial recognition model using L2 Normalized Euclidean distance and enabled alignment mode on unit test items of DeepFace, can be separated, as illustrated in Figure 5. In this context, the C4.5 algorithm [20] was employed to identify the threshold that yields the maximum information gain.

3.4. Handling Many Faces Issue

Some instances of the LFW dataset have some problematic data. For instance, an image may contain many faces and the one targeted by the label may not be known precisely. In this case, all pairs of faces were extracted using the current running detector (e.g. MtCnn), and all candidates were then represented as vectors using the current running facial recognition model (e.g. VGG-Face). Following that, all vector candidates were distanced by using current running similarity algorithms (e.g. cosine and euclidean). As a result, the face pair with the minimum distance was fed into the algorithm. Figure 6 explains how we extract the right face from these problematic samples. The faces of the lady are used in both images for this case since the distance is shorter.

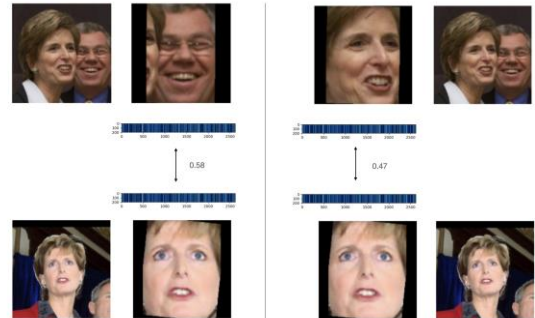


Figure 6. Picking a Face Among Many Faces

4. PRE-TRAINED MODELS

LightFace for python [21] will be used to build facial recognition pipelines. This library basically wraps the following models with pre-trained weights.

4.1. Facial Recognition Models

Researchers from the Visual Geometry Group at Oxford University developed the VGG-Face model. The model's structure was detailed in their paper [3], and they generously provided pre-trained weights for the VGG-Face model in MatLab, Torch, and Caffe on the research group's website [22]. Subsequently, the MatLab weights were converted into Keras, and the VGG-Face model was reconstructed from the ground up using Keras.

Researchers at Google designed the FaceNet model, which outputs 128 dimensions. The model's structure is outlined in their paper [2], but the pre-trained weights are not publicly available. However, David Sandberg undertook the task of re-training FaceNet with both 128 and 512 dimensional outputs and generously shared the pre-trained weights in TensorFlow format [23]. These TensorFlow weights were then converted to Keras format, and a new model in Keras was constructed from the ground up.

ArcFace, a component of the InsightFace project, was developed by the Deep Inside research group [4]. The ArcFace model, along with its pre-trained weights, can be accessed on the group's website for both PyTorch and MXNet. Additionally, Leon D. Garse re-trained the ArcFace model specifically for Keras [24]. The pre-trained weights from this repository were utilized, and a new model was constructed from scratch in Keras.

SFace, primarily developed by Yaoyao Zhong [6], includes shared models and pre-trained weights accessible for both PyTorch and MXNet. The OpenCV community played a role in converting the model to a language-independent ONNX format [25]. Consequently, SFace can be utilized through OpenCV in this context.

Researchers from Carnegie Mellon University developed the OpenFace model. Both the model and its pre-trained weights can be accessed on the research group's website for Lua Torch. Victor Sy Wang took on the task of converting the pre-trained weights of the OpenFace model to Keras [26]. The weights from this re-implementation were utilized, and the model was then constructed from scratch.

Facebook researchers developed the DeepFace model, outlining its structure in their paper [1], although they did not provide the pre-trained weights for the model. Swarup Ghosh took the initiative to re-train the DeepFace model using Keras from scratch and shared the resulting pre-trained weights in his repository [27]. While the original study utilized the Social Face Classification (SFC) [1] dataset with 4030 identities, Ghosh re-trained the DeepFace model using the VGGFace2 [28] dataset with 8631 identities. Consequently, the original DeepFace model produces 4030-dimensional vectors, whereas Ghosh's re-implementation generates 8631-dimensional vectors. The DeepFace model was reconstructed from the ground up, and the pre-trained weights from this re-implementation were utilized.

The DeepId model, crafted by researchers from the Chinese University of Hong Kong, had its model structure detailed in their paper [8]. However, the pre-trained weights were not made publicly available. Roy Ran took on the task of re-training the DeepId model from the ground up and generously shared the resulting pre-trained weights specifically for TensorFlow [29]. To facilitate utilization of these pre-trained weights, a conversion process was employed to adapt them into the Keras format.

Dlib constructs a ResNet model designed for facial recognition purposes [5], with the Dlib library being directly imported for implementation.

4.3. Face Detectors

The MtCnn face detector model was developed by Kaipeng Zhang et al [14]. Iván de Paz Centeno took the initiative to independently re-implement MtCnn from scratch using Keras, presenting it as a project [30]. The MtCnn project has been directly incorporated into the system.

The RetinaFace model, created by researchers from Deep Insight as part of the InsightFace project [4], is provided with both the model and pre-trained weights for PyTorch and MXNet on the research group's website. Additionally, Stanislas Bertrand undertook the task of re-implementing the RetinaFace model using Keras [31]. Following the completion of the RetinaFace project repackaging [32], this new RetinaFace package was directly imported for use.

The SSD (Single Shot Multibox Detector) model, designed by Wei Liu et al [11], had its Caffe model wrapped by the OpenCV community [10]. In practice, SSD was utilized through OpenCV. However, SSD itself does not provide facial landmarks. To align faces identified by SSD, the eye detection module of OpenCV was employed. MediaPipe [12] features a face detection module built on BlazeFace [13], which is structurally derived from SSD but includes facial landmarks. The MediaPipe library was directly imported for use. Additionally, Dlib was employed for face detection using its HOG model [5], and the Haar Cascade module of OpenCV was imported to utilize its face detection functionality.

Re-implementations of certain reference models have been undertaken, and as a consequence, experimental results will be linked to these re-implementations rather than the original studies. It's important to note that even when the same facial recognition model and pre-trained weights were utilized, discrepancies in pre-processing techniques and model implementation may exist compared to those employed in the original study. Specifically, a 2D discarding approach was implemented, discarding regions outside the facial area using commonly available open-source face detectors, as suggested in the literature [19], to prevent the artificial inflation of results.

4. EXPERIMENTS

A total of 378 experiments were performed on the LFW dataset test set, considering combinations of the four dimensions listed in Table 8. This exhaustive approach aimed to identify the optimal configuration set for facial recognition. Additionally, each module in the pipeline was systematically examined to discern its specific impact on the overall results.

Table 8. Different Configuration Sets for Experiments

Configuration	Possible Values
Facial Recognition Models	FaceNet128d, FaceNet512d, Vgg-Face, ArcFace, Dlib, SFace, OpenFace, DeepFace, DeepId
Face Detector Models	OpenCv, Ssd, MtCnn, Dlib, RetinaFace, MediaPipe, No Detection
Distance Metrics	Cosine, Euclidean, L2-Norm Euclidean
Alignment Mode	Enabled, Disabled

Tables 2, 3, 4, 5, 6, and 7 present the performance matrices of facial recognition models and face detectors across various distance metrics and alignment modes. In these tables, columns denote facial recognition models, rows represent face detectors, and the cells display corresponding accuracy values. Additionally, to enhance comprehension of the influence of detection and alignment modules, the scores of facial recognition models without detection and alignment are also provided.

As outlined in Section 3, the output of the representation module consists of a vector pair, and the subsequent step involves measuring the distance between these vectors. A pair is classified as the same person if the distance between them falls below the specified threshold value. Furthermore, ROC curves were generated using distance values and corresponding target labels. This approach is a conventional practice to assess the model's performance from a holistic perspective.

ROC curves for different models were illustrated in Figure 7 for FaceNet512d, Figure 8 for VGG-Face, Figure 9 for Dlib, Figure 10 for FaceNet128d, Figure 11 for ArcFace, and Figure 12 for SFace, focusing specifically on robust detectors (RetinaFace, MtCnn, Dlib, and MediaPipe). It is recommended to view these figures in color for optimal clarity. Additionally, the graphics include area under the curve (AUC) values, and the legends are arranged in descending order based on these values. In this context, a higher AUC value indicates superior model performance.

The ROC curves presented above highlight that FaceNet512d and VGG-Face stand out as the most robust models, regardless of the detector used. However, the performance of other facial recognition models exhibits variations depending on the specific configuration employed.

5. RESULTS

The experiments examined both context-independent and context-dependent conditions. While there is significant importance in running facial recognition pipelines in a context-independent manner, the results obtained under context-dependent conditions may provide a more accurate understanding of the direct influence of face detection on facial recognition pipelines. The study involved comparing these models with human performance, resulting in a context-free accuracy of 97.5% and a non-context-free accuracy of 99.2% [19]. Finally, the study demonstrated the impact of detectors and alignment modules on the pipelines.

Table 9. Comparison of Measured Accuracy Scores in LightFace and Declared Accuracy Scores in Their Original Researches

Model	Measured Accuracy	Declared Accuracy	Original Paper
FaceNet-512d	98.4	99.6	[2]
FaceNet-128d	97.0	99.6	[2]
Dlib	96.8	99.3	[5]
VGG-Face	96.7	98.9	[3]
ArcFace	96.6	99.5	[4]
SFace	93.0	99.5	[6]
OpenFace	78.7	92.9	[7]
DeepFace	68.7	97.3	[1]
DeepId	65.6	97.4	[8]

As shown in Table 9, the measured accuracies of various facial recognition models using the LightFace library offer valuable insights into their performance. A comparison with the accuracy scores declared in their original papers reveals that FaceNet, Dlib, ArcFace, and VGG-Face closely approach the declared performances, albeit falling slightly below. Notably, for models such as VGG-Face and Dlib, the small differences may be a result of the adopted normalization techniques in both the original studies and the LightFace framework. Additionally, it is noteworthy that LightFace employs open-sourced pre-trained weights for their re-implementation instead of using the original ones, as seen in the case of FaceNet and ArcFace. These minor differences are deemed acceptable and do not compromise the overall robustness of these models. Despite the slight variations, the models remain robust and showcase reliable performance. Despite these differences, SFace, while exhibiting a performance lower than declared, shows a manageable gap, suggesting its continued usability.

In contrast, OpenFace, DeepFace, and DeepId exhibit notable disparities from their originally reported accuracies, which can be attributed to the utilization of pre-trained weights retrieved from open-source re-implementations rather than the original studies' pre-trained weights. The original pre-trained weights, as detailed in Section 4, are not publicly available. This distinction highlights the significance of the source of pre-trained weights, as the use of open-source alternatives may lead to differences in model performance. The observed disparities underscore the importance of obtaining and utilizing the original pre-trained weights when aiming to replicate or surpass the accuracies reported in the respective original studies.

5.1. Context Independent Results

FaceNet512d achieved an accuracy of 98.4%, followed by FaceNet128d with 97.0% accuracy. Both models have surpassed human-level accuracy. The most satisfactory configurations are highlighted in bold within the tables.

Dlib achieved an accuracy of 96.8%, ArcFace and VGG-Face demonstrated 96.7% and 96.6% accuracy, respectively. These models achieved accuracy levels comparable to human performance. SFace, with an accuracy of 93.0%, approached human-level accuracy.

OpenFace achieved an accuracy of 78.7%, DeepFace reached 68.7%, and DeepId attained 65.6% accuracy. Consequently, these re-implementations exhibited subpar performance, rendering them unsuitable for deployment in production pipelines. Notably, FaceNet128d, ArcFace, and Dlib displayed significantly reduced performance when the detection stage was omitted. It is imperative to execute the detection stage for these facial recognition models.

OpenCv and Ssd detectors exhibit diminished performance when the alignment mode is enabled. This limitation arises from their inability to pinpoint eye locations with the same precision as RetinaFace or MtCnn. Consequently, it is advisable to omit the alignment stage when using these detectors. RetinaFace, MtCnn, and Dlib, on the other hand, showcase a notable performance enhancement when the alignment mode is activated, establishing them as robust detectors across all considered configurations. The performance of MediaPipe experiences a slight decline with alignment; even with alignment disabled, it ranks below RetinaFace, Dlib, and MtCnn.

Finally, none of the distance metrics demonstrate underperformance, and despite their variances, each contributes significantly to the overall performance.

5.2. Context Dependent Results

In the presence of contextual information, facial recognition models struggle to achieve human-level accuracy. Their performance sees a significant boost when focusing exclusively on the facial region, likely due to their training on context-independent data. FaceNet512d attains a 92% score even with detection disabled, marking the closest performance to human among facial recognition models without detectors. VGG-Face achieves a 90.6% score, while SFace scores 83.4% with detection turned off, both trailing FaceNet-512d. FaceNet512d stands out as the optimal choice for use cases demanding both swift and reliable results, such as facial recognition in crowded environments. In such scenarios, detection and alignment may be skipped, yet the pipeline maintains a commendable score.

5.3. Impact of custom modules in the pipeline

In optimal conditions, the detection impact for ArcFace is 41.8%, FaceNet128d is 30.8%, Dlib is 27.3%, OpenFace is 20.9%, SFace is 9.6%, DeepFace is 7.6%, FaceNet-512d is 6.6%, VGG-Face is 6.1%, and DeepId is 3.6%. Therefore, the inclusion of a detection module is essential for facial recognition pipelines.

Similarly, the alignment impact for SFace is 17.2%, DeepFace is 13.7%, ArcFace is 11.7%, FaceNet128d is 6.5%, VGG-Face is 1.6%, OpenFace is 1.0%, FaceNet-512d is 0.7%, and DeepId is 0.7% in the best-case scenarios. Consequently, alignment serves as a crucial enhancement module for facial recognition pipelines.

Furthermore, identifying the intersections of misclassifications for each facial recognition model across various configuration sets will aid in comprehending the weaknesses inherent in each model. As depicted in Figure 13, FaceNet512d misclassifies 6 pairs, FaceNet128d misclassifies 5 pairs, and both Dlib and ArcFace misclassify 3 pairs in all experiments involving Euclidean, l2 normalized Euclidean, and Cosine distance metrics, as well as RetinaFace, MtCnn, and Dlib face detectors, with alignment mode exclusively enabled.

Predominantly, the common misclassifications for each model involve false negatives. It appears that these models commonly misclassify pairs when individuals wear accessories like sunglasses or caps. Similarly, differences in age and emotion within pairs contribute to misclassifications. Furthermore, the common misclassifications mainly pertain to pairs of the same person.

6. CONCLUSION

In this study, all configuration alternatives were tested in LightFace. To avoid artificially boosting results, the facial area in images was focused and the regions outside of the facial area were discarded.

Based on the results of the experiments, it was determined that some facial recognition models were capable of reaching or exceeding human level accuracy under certain configurations. It can be concluded that the facial recognition model with dependent face detector, distance metric and alignment mode underperform or overperform. Based on their best configuration set, facial recognition models were classified into four categories: passing human level accuracy, reaching human level accuracy, coming close to human level accuracy, and underperforming ones. AUC and ROC curves also provide opinions about the robustness of facial recognition models.

Additionally, this work discusses the contribution of detection and alignment modules to pipeline accuracy. The detection module is an essential component of a pipeline as it can improve performance by up to 40%. While alignment serves as a significant performance booster, improving performance by up to 17%. Results found regarding the robustness of face detectors. OpenCv, SSD and MediaPipe have deteriorated when alignment is enabled, whereas RetinaFace, MtCnn and Dlib are always resilient.

LightFace also features an easy and simple interface. The dimensions mentioned in this study are merely input arguments for a pipeline. Thus, practitioners are able to construct and run facial recognition pipelines with a few lines of code. This study can be used as a guide to assist practitioners in choosing the most appropriate configurations for their facial recognition pipelines according to their individual needs.

Table 2. Accuracy Metric for Euclidean Distance and Disabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	92.8	96.1	95.7	84.1	88.3	78.6	70.8	67.4	64.3
MtCnn	92.5	95.9	95.5	81.8	89.3	76.3	70.9	65.9	63.2
Dlib	89.0	96.0	94.1	82.6	96.3	73.1	75.9	61.8	61.9
MediaPipe	87.1	94.9	93.1	71.1	91.9	73.2	77.6	61.7	62.4
Ssd	94.9	97.2	96.7	83.9	88.6	82.0	69.9	66.7	64.0
OpenCv	90.2	94.1	95.8	89.8	91.2	86.9	71.1	68.4	61.1
None	64.1	92.0	90.6	56.6	69.0	81.4	57.4	60.8	60.7

Table 3. Accuracy Metric for L2 Normalized Euclidean Distance and Disabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	95.9	98.0	95.7	95.7	88.4	90.6	70.8	67.7	64.6
MtCnn	96.2	97.8	95.5	95.9	89.2	91.1	70.9	67.0	64.0
Dlib	89.9	96.5	94.1	93.8	95.6	75.0	75.9	62.6	61.8
MediaPipe	90.0	96.3	93.1	89.3	91.8	74.6	77.6	64.9	61.6
Ssd	97.0	97.9	96.7	96.6	89.4	93.0	69.9	68.7	64.9
OpenCv	92.9	96.2	95.8	93.2	91.5	91.7	71.1	68.3	61.6
None	67.6	91.4	90.6	57.2	69.3	83.4	57.4	62.6	61.6

Table 4. Accuracy Metric for Cosine Distance and Disabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	95.9	98.0	95.7	95.7	88.4	90.6	70.8	67.7	63.7
MtCnn	96.2	97.8	95.5	95.9	89.2	91.1	70.9	67.0	64.0
Dlib	89.9	96.5	94.1	93.8	95.6	75.0	75.9	62.6	61.7
MediaPipe	90.0	96.3	93.1	89.3	91.8	74.6	77.6	64.9	61.6
Ssd	97.0	97.9	96.7	96.6	89.4	93.0	69.9	68.7	63.8
OpenCv	92.9	96.2	95.8	93.2	91.5	91.7	71.1	68.1	61.1
None	67.6	91.4	90.6	54.8	69.3	83.4	57.4	62.6	61.1

Table 5. Accuracy Metric for Euclidean Distance and Enabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	93.5	95.9	95.8	85.2	88.9	80.2	69.4	67.0	65.6
MtCnn	93.8	95.2	95.9	83.7	89.4	77.4	70.2	66.5	63.3
Dlib	90.8	96.0	94.5	88.6	96.8	66.3	75.8	63.4	60.4
MediaPipe	88.6	95.1	92.9	73.2	93.1	72.5	78.7	61.8	62.2
Ssd	85.6	88.9	87.0	75.8	83.1	76.9	66.8	63.4	62.5
OpenCv	84.2	88.2	87.3	73.0	84.4	81.1	66.4	65.5	59.6
None	64.5	92.0	90.6	56.6	69.0	81.4	57.4	60.8	60.7

Table 6. Accuracy Metric for L2 Normalized Euclidean Distance and Enabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	96.4	98.4	95.8	96.6	89.1	92.4	69.4	67.7	64.4
MtCnn	96.8	97.6	95.9	96.0	90.0	90.5	70.2	66.4	64.0
Dlib	92.6	97.0	94.5	95.1	96.4	69.8	75.8	66.5	59.5
MediaPipe	90.6	96.1	92.9	90.3	92.6	75.4	78.7	64.7	63.0
Ssd	87.5	88.7	87.0	86.2	83.3	84.6	66.8	64.1	62.6
OpenCv	84.8	87.6	87.3	84.6	84.0	83.6	66.4	63.8	60.9
None	67.6	91.4	90.6	57.2	69.3	83.4	57.4	62.6	61.6

Table 7. Accuracy Metric for Cosine Distance and Enabled Alignment

	<i>FaceNet128</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	96.4	98.4	95.8	96.6	89.1	92.4	69.4	67.7	64.4
MtCnn	96.8	97.6	95.9	96.0	90.0	90.5	70.2	66.3	63.0
Dlib	92.6	97.0	94.5	95.1	96.4	69.8	75.8	66.5	58.7
MediaPipe	90.6	96.1	92.9	90.3	92.6	75.4	78.7	64.8	63.0
Ssd	87.5	88.7	87.0	86.2	83.3	84.5	66.8	63.8	62.6
OpenCv	84.9	87.6	87.2	84.6	84.0	83.6	66.2	63.7	60.1
None	67.6	91.4	90.6	54.8	69.3	83.4	57.4	62.6	61.1

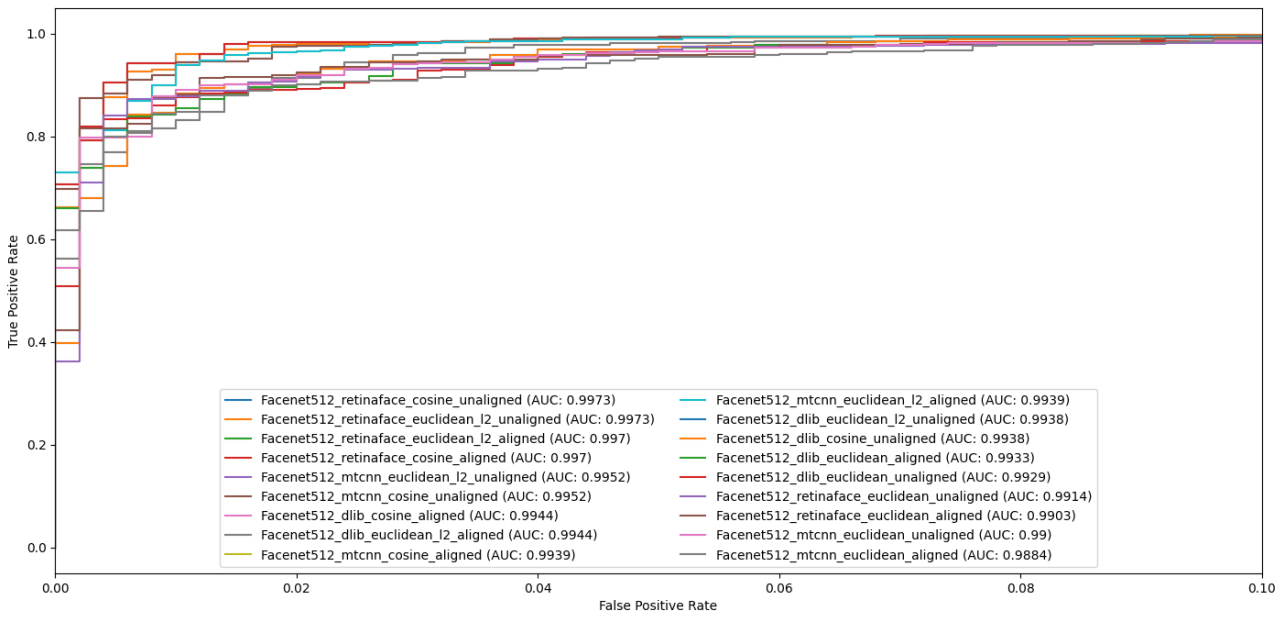


Figure 7. ROC Curves and AUC Scores for FaceNet512d

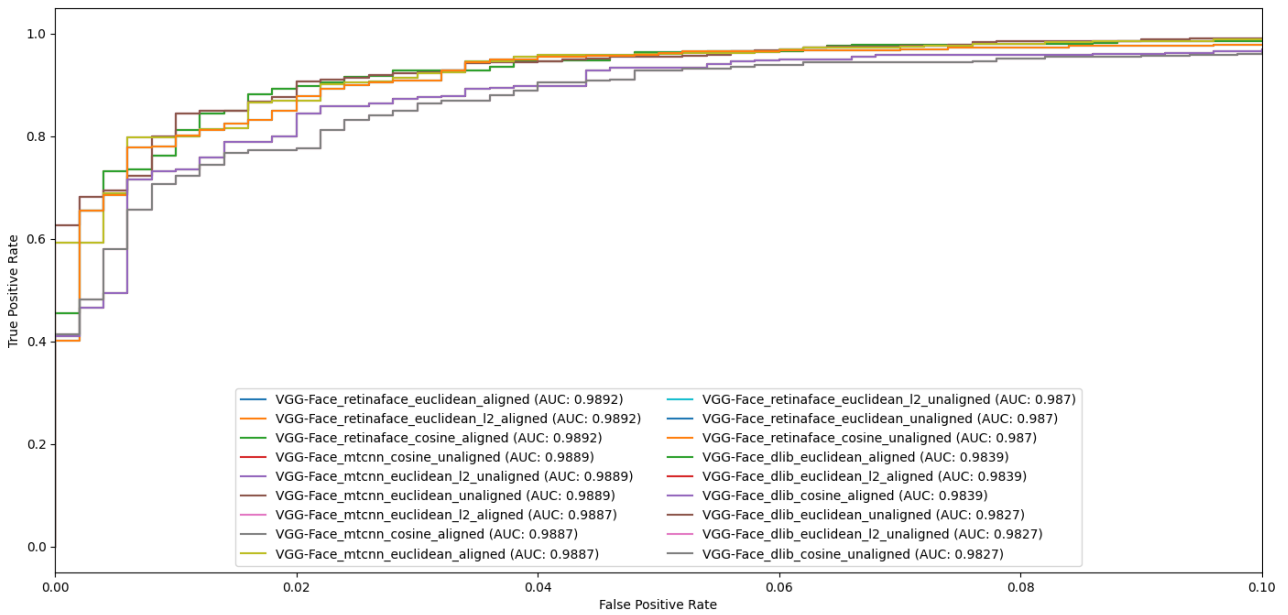


Figure 8. ROC Curves and AUC Scores for VGG-Face

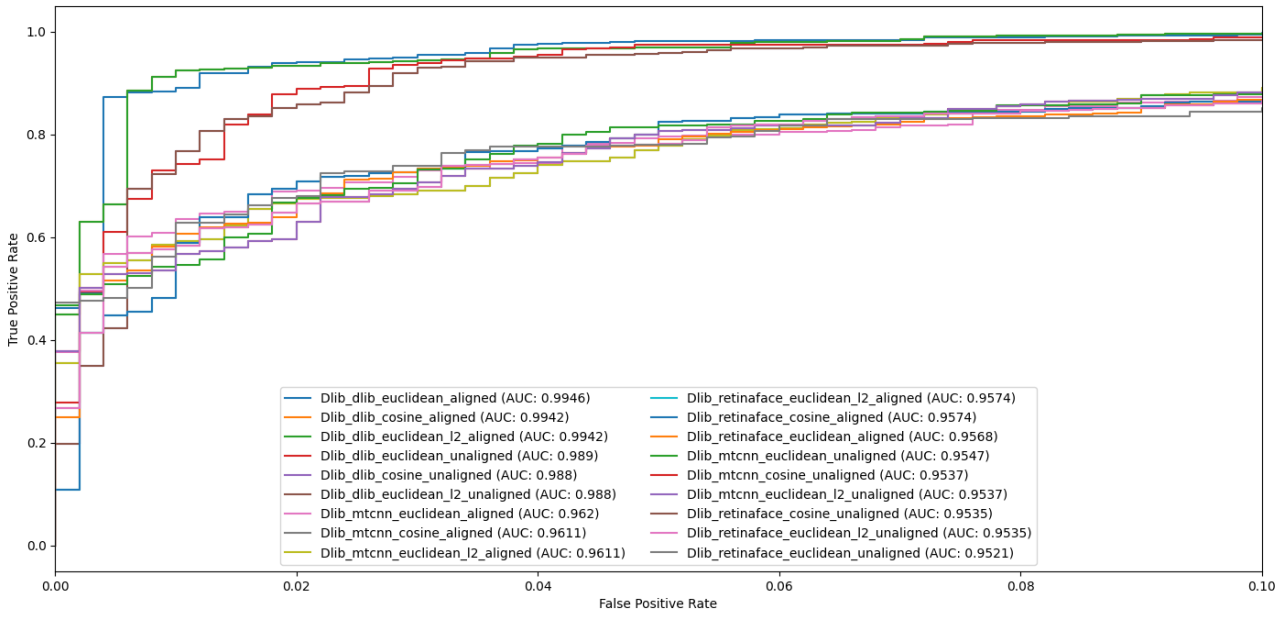


Figure 9. ROC Curves and AUC Scores for Dlib

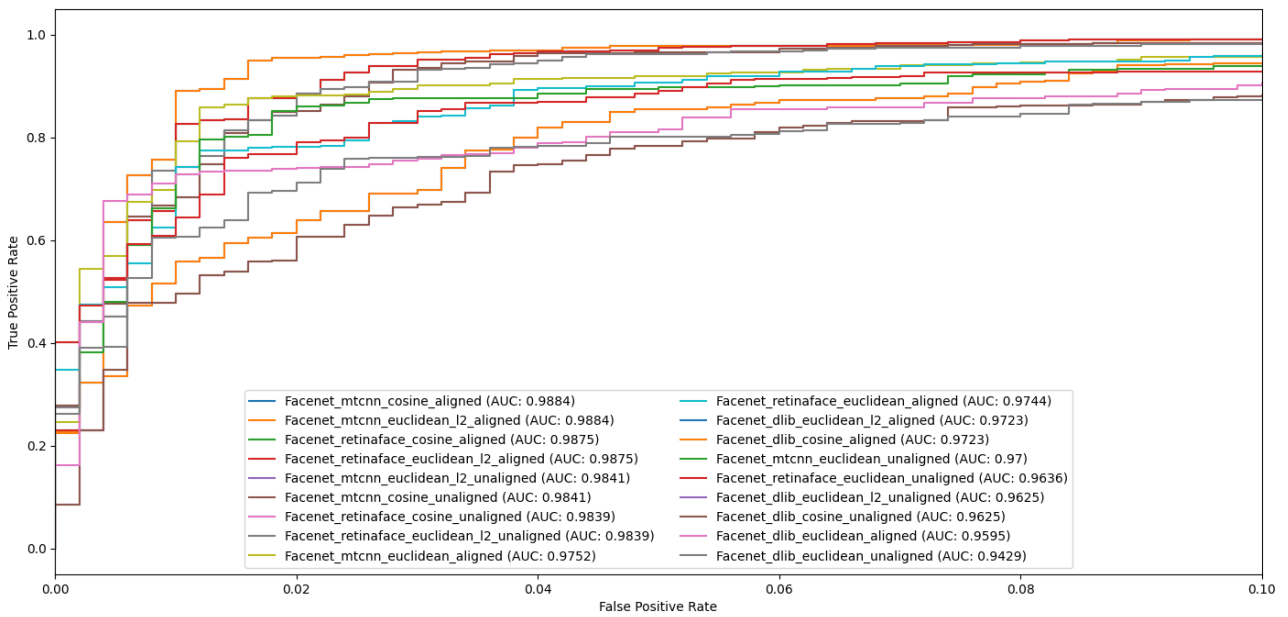


Figure 10. ROC Curves and AUC Scores for FaceNet128d

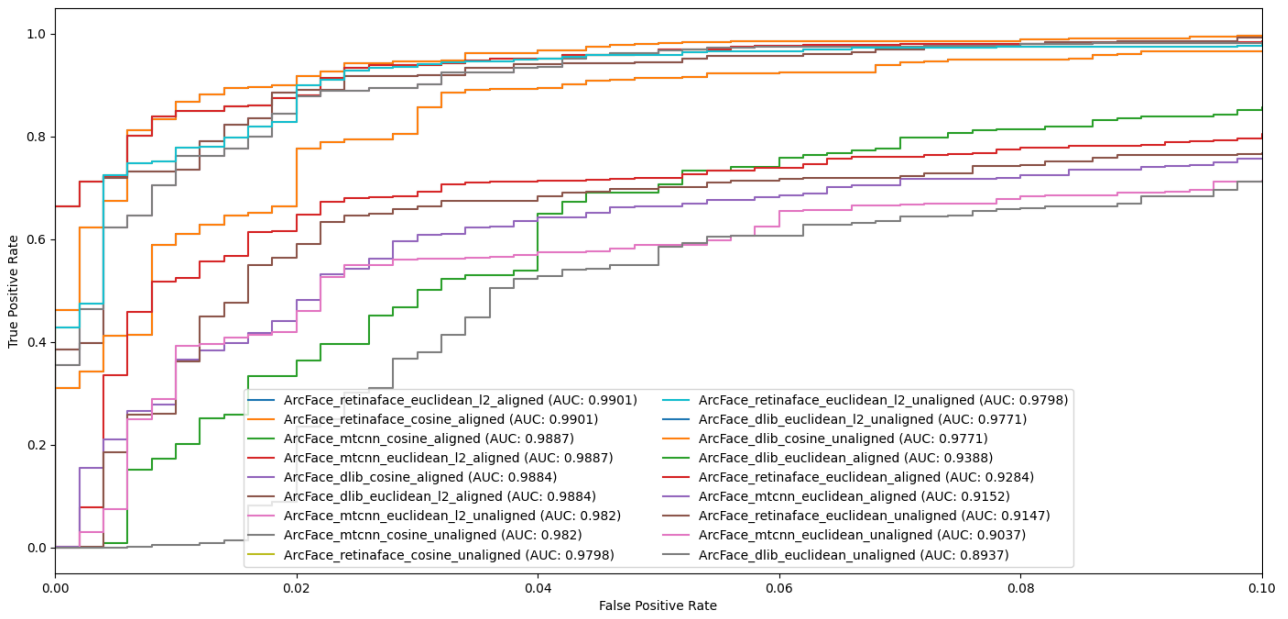


Figure 11. ROC Curves and AUC Scores for ArcFace

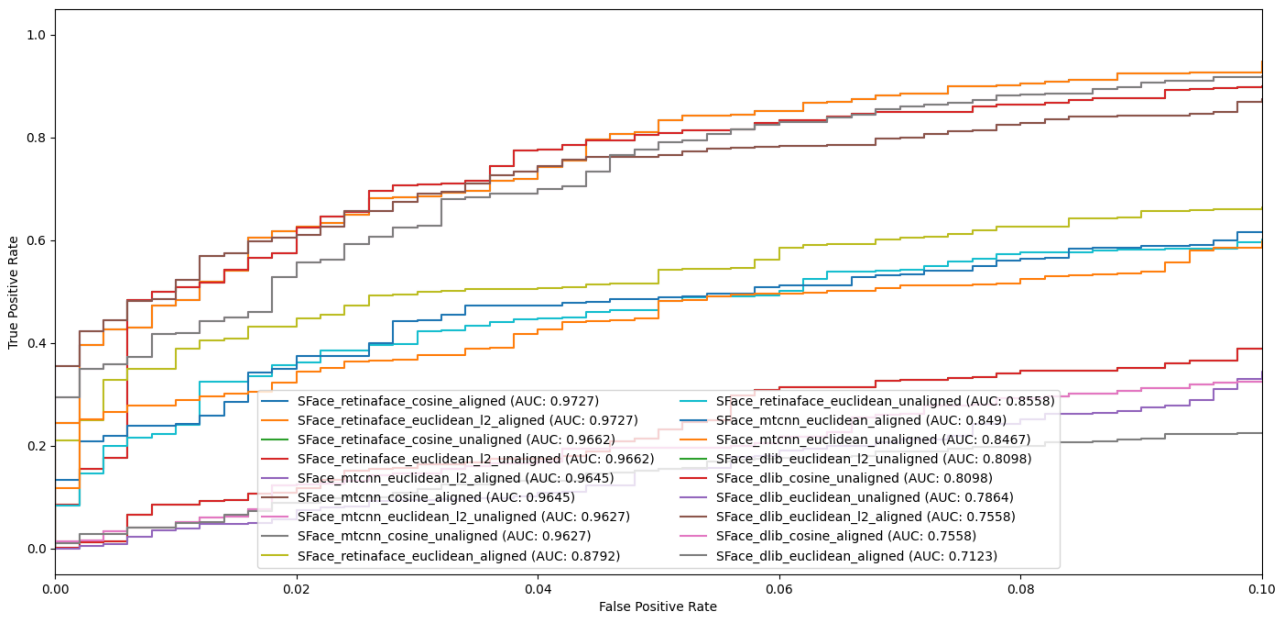


Figure 12. ROC Curves and AUC Scores for SFace

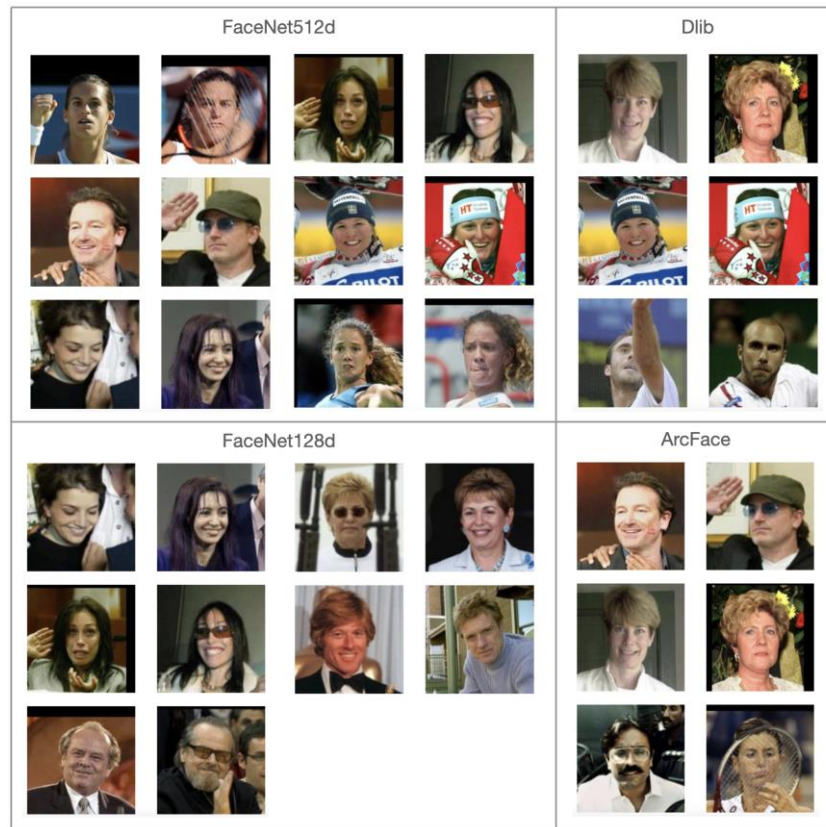


Figure 13. Common Misclassifications

REFERENCES

- [1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deep-face: Closing the gap to human-level performance in face verification", **In Proceedings of the IEEE conference on computer vision and pattern recognition**, 1701–1708, 2014.
- [2] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering", **In Proceedings of the IEEE conference on computer vision and pattern recognition**, 815–823, 2015.
- [3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition", **In British Machine Vision Conference**, 2015.
- [4] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition", **In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**, pages 4690–4699, 2019.
- [5] D. E. King, "Dlib-ml: A machine learning toolkit", *The Journal of Machine Learning Research*, 10, 1755–1758, 2009.
- [6] Y. Zhong, W. Deng, J. Hu, D. Zhao, X. Li, and D. Wen, "Sface: Sigmoid-constrained hypersphere loss for robust face recognition", *IEEE Transactions on Image Processing*, 30:2587–2598, 2021.
- [7] B. Amos, B. Ludwiczuk, M. Satyanarayanan, et al. "Openface: A general-purpose face recognition library with mobile applications", *CMU School of Computer Science*, 6(2):20, 2016.
- [8] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes", **In Proceedings of the IEEE conference on computer vision and pattern recognition**, 1891–1898, 2014.
- [9] S. I. Serengil and A. Ozpinar, "Lightface: A hybrid deep face recognition framework", **In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU)**, 23–27. IEEE, 2020.
- [10] G. Bradski, "The opencv library", *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11):120–123, 2000.
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg. "Ssd: Single shot multi-box detector", **In European conference on computer vision**, 21–37. Springer, 2016.
- [12] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, et al. "Mediapipe: A framework for building perception pipelines", arXiv preprint arXiv:1906.08172, 2019.
- [13] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann. "Blazeface: Sub-millisecond neural face detection on mobile gpus", arXiv preprint arXiv:1907.05047, 2019.
- [14] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. "Joint face detection and alignment using multitask cascaded convolutional networks", *IEEE signal processing letters*, 23(10):1499–1503, 2016.

- [15] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou. “Retinaface: Single-shot multi-level face localisation in the wild”. **In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**, pages 5203–5212, 2020.
- [16] S. I. Serengil and A. Ozpinar. “Hyperextended lightface: A facial attribute analysis framework”, **In 2021 International Conference on Engineering and Emerging Technologies (ICEET)**, 1–4. IEEE, 2021.
- [17] G. B Huang, M. Mattar, T. Berg, and E. L. Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments”, **In Workshop on faces in Real-Life Images: detection, alignment, and recognition**, 2008.
- [18] O. Kramer, “Scikit-learn”, **In Machine learning for evolution strategies**, pages 45–53. Springer, 2016.
- [19] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. “Attribute and simile classifiers for face verification”, **In 2009 IEEE 12th international conference on computer vision**, 365–372. IEEE, 2009.
- [20] J. R. Quinlan, C4. 5: **programs for machine learning**. Elsevier, 2014.
- [21] S. I. Serengil, Deepface: A lightweight face recognition and facial attribute analysis (age, gender, emotion and race) library for python, <https://github.com/serengil/deepface>, 15.04.2024.
- [22] A. Z. Omkar, M. Parkhi, A. Vedaldi, Vgg face descriptor, https://www.robots.ox.ac.uk/~vgg/software/vgg_face/, 15.04.2024.
- [23] D. Sandberg, Facenet: Face recognition using tensorflow, <https://github.com/davidsandberg/facenet>, 15.04.2024.
- [24] L. D Garse, Keras insightface, https://github.com/leongarse/Keras_insightface, 15.04.2024.
- [25] Y. Feng, SFace, https://github.com/opencv/opencv_zoo/tree/main/models/face_recognition_sface, 15.04.2024.
- [26] V. S. Wang, Keras-openface2, <https://github.com/iwantooxxoox/Keras-OpenFace>, 15.04.2024.
- [27] S. Ghosh, Deepface, <https://github.com/swghosh/DeepFace>, 2019. 15.04.2024.
- [28] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “Vggface2: A dataset for recognising faces across pose and age”, **In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)**, 67–74. IEEE, 2018.
- [29] R. Ran, Deepid implementation, <https://github.com/Ruoyiran/DeepID>, 15.04.2024.
- [30] I. P. Centeno, Mtcnn, <https://github.com/ipazc/mtcnn>, 15.04.2024.
- [31] S. Bertrand, Retinaface-tf2, <https://github.com/StanislasBertrand/RetinaFace-tf2>, 15.04.2024.
- [32] S. I. Serengil, Retinaface: Deep face detection library for python, <https://github.com/serengil/retinaface>, 15.04.2024.
- [33] K. Yildiz, E. Gunes, A. Bas, “CNN-based Gender Prediction in Uncontrolled Environments”, *Duzce University Journal of Science & Technology*, 890-898. 2021.
- [34] H. Goze, O. Yildiz, “A New Deep Learning Model for Real-Time Face Recognition and Time Marking in Video Footage”, *Journal of Information Technologies*, 167-175. 2022.
- [35] G. Guodong, N. Zhang, “A survey on deep learning based face recognition”, *Computer Vision and Image Understanding*, 189, 102805, 2019.
- [36] M. Hassaballah, S. Aly, “Face recognition: challenges, achievements and future directions”, *IET Computer Vision*, 9(4), 614-626, 2015.
- [37] D. Heinsohn, E. Villalobos, L. Prieto, D. Mery, “Face recognition in low-quality images using adaptive sparse representations”, *Image and Vision Computing*, 85, 46-58, 2019.
- [38] P. J. Phillips, A. J. O’toole, “Comparison of human and computer performance across face recognition experiments”, *Image and Vision Computing*, 32(1), 74-85, 2014.
- [39] E. G. Ortiz, B. C. Becker, “Face recognition for web-scale datasets”, *Computer Vision and Image Understanding*, 118, 153-170, 2014.

Metin Madenciliği ve Makine Öğrenmesi Teknikleri ile Sağlık Hizmetleri Pazarlamasına Yönelik Twitter Verilerinin Analizi*

Araştırma Makalesi/Research Article

 Büşra SAYLAN¹,  Songül ÇINAROĞLU²

^{1,2}Hacettepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Sağlık Yönetimi, Ankara, Türkiye

bsr.syln23@gmail.com, cinaroglus@hacettepe.edu.tr

(Geliş/Received:24.01.2024; Kabul/Accepted:04.04.2024)

DOI: 10.17671/gazibtd.1424960

Özet— Sağlık hizmeti kullanıcılarının sağlık hizmetleri pazarına ilişkin duygu durumlarının analizi sağlık hizmetleri pazarlamasına yönelik algıların analizinde kritik rol oynamaktadır. Bu çalışmada Twitter kullanıcılarından elde edilen verilerle kişilerin sağlık hizmetleri pazarlamasına yönelik görüşleri duygu analizi kullanılarak değerlendirilmiştir. Twitter'dan API Key ile 1 Ekim 2022 ve 30 Kasım 2022 tarihleri arasında elde edilen 27079 İngilizce dilinde atılan tweet verileri üzerinde yapılan duygu analizi sonucunda Twitter kullanıcılarının sağlık hizmetleri pazarlaması konusundaki görüşlerinin %50,36'sının (n=13638) olumlu, %32,21'inin (n=8721) nötr ve %17,43'ünün (n=4720) ise olumsuz olduğu belirlenmiştir. Oluşturulan kelime bulutlarına göre 'health', 'care', 'market' ve 'insurance' en sık tekrarlanan kelimeler olarak belirlenmiştir. Twitter kullanıcılarının duygu durumlarının en iyi tahmin edicilerini bulmak amacıyla yeni bir veri seti oluşturulmuş ve Yapay Sinir Ağları (Neural Network-NN), Rastgele Orman (Random Forest-RF), Naive Bayes (NB) ve k-en yakın komşu (k-nearest neighbor k-NN) makine öğrenmesi teknikleri ile tahmin edilmiştir. RF (AUC=0,707; CA=0,646) ve NN (AUC=0,706; CA=0,645) diğer makine öğrenmesi teknikleriyle karşılaştırıldığında Twitter kullanıcılarının duygu durumlarının en iyi tahmin edici makine öğrenmesi teknikleri olmuştur. Pisagor ağacı ile Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik duygu durumlarının en iyi tahmin edici değişkeninin 'favorite (beğeni)' kelimesi olduğu ve kullanıcıların sağlık hizmetleri pazarlamasına yönelik atılan tweetlere çoğunlukla katıldığı belirlenmiştir. Gelecekteki yapılacak olan çalışmaların zaman içindeki değişiklikleri daha iyi anlamak ve derin bir perspektif sunmak amacıyla daha uzun bir zaman aralığını kapsayacak şekilde planlanması, ayrıca analizlerin genellenebilirliğini artırmak ve bulguların daha geniş bir kitleye uygulanabilir olmasına katkıda bulunmak için daha büyük bir veri seti kullanılması önerilmektedir.

Anahtar Kelimeler— sağlık hizmetleri pazarlaması, duygu analizi, twitter, makine öğrenmesi, pisagor ağacı

Analysis of Twitter Data on Healthcare Marketing by Text Mining and Machine Learning Techniques*

Abstract— The analysis of healthcare users' moods regarding healthcare market plays a critical role in the analysis of perceptions towards the healthcare marketing. In this study, the views of people on healthcare marketing were evaluated using sentiment analysis with the data obtained from Twitter users. As a result of the sentiment analysis performed on 27079 English-language tweet data obtained from Twitter with API Key between October 1, 2022, and November 30, 2022, 50.36% (n=13638) of Twitter users' opinions on healthcare marketing were positive, 32.21% (n=8721) were neutral and 17.43% (n=4720) were negative. According to the word clouds created, 'health', 'care', 'market', and 'insurance' were determined as the most frequently repeated words. In order to find the best predictors of the mood of Twitter users, a new data set was created, and Neural Networks (NN), Random Forest (RF), Naive Bayes (NB), and k-nearest neighbor (k-NN) were predicted by machine learning techniques. RF (AUC=0.707; CA=0.646), and NN (AUC=0.706; CA=0.645) were the best predictive machine learning techniques of Twitter users' moods when compared to other machine learning techniques. With the Pythagorean tree, it was determined that the best predictor variable of Twitter users' emotional states towards healthcare marketing was the word 'favorite'. Future studies are suggested to be planned to cover a longer time to understand temporal changes better and provide a deeper perspective. Additionally, using a larger dataset is recommended to enhance the generalizability of analyses and contribute to the applicability of findings to a broader audience.

Keywords— health care marketing, sentiment analysis, twitter, machine learning, pythagorean tree

1. GİRİŞ (INTRODUCTION)

Sağlık sektörü tüm dünyada çok hızlı bir şekilde büyüme eğilimindedir ve tüm sektörler arasında kritik bir öneme sahiptir. Sağlık sektörü aynı zamanda ülkelerin sosyo-ekonomik gelişmişlik düzeyinin en önemli göstergelerinden biri olarak kabul edilmektedir [1]. Gelişmekte olan bir ülke kategorisinde yer alan Türkiye’de sağlık sektörüne yönelik son yıllarda önemli yatırımlar yapılmaya başlanmıştır. Türkiye’de 2021 yılında toplam sağlık harcamaları önceki yıla kıyasla %41,6 artış göstererek yaklaşık 354 milyon TL’ye yükselmiştir. Toplam sağlık harcamalarının GSYH’ye oranı ise %4,9 olarak belirlenmiştir [2]. Sadece Türkiye’de değil tüm dünyada sağlık hizmetleri giderek daha fazla ön plana çıkmakta ve diğer sektörler arasında önemli bir paya sahip olmaktadır. Sunulan sağlık hizmetlerinin tüketicileri olan insanların hayatı gibi önemli bir konuyu içermesi, bu hizmetlerin tüketicilerinin çokluğu ve sağlık hizmetlerine özgü özelliklerin varlığı sağlık sektörünün önemini vurgulayan konulardandır [3]. Bu durum aynı zamanda sağlık sektöründe sağlık hizmetlerinin sunumunda pazarlama uygulamalarının kullanımının önemini de belirlemektedir. Sağlık hizmetleri pazarlaması, hedef alınan kitlenin zihinsel ve fiziksel sağlık sorunlarını anlayarak, sorularını yanıtlayarak ve önerilen çözümlerden nasıl yararlanabileceklerini açıklayarak onları etkilemekle ilgili olan pazarlama çabası olarak ifade edilmektedir [4]. Bir başka ifade ile sağlık hizmetleri pazarlaması sağlık hizmetleri sektörünün müşterisi konumunda olan hastaların ihtiyaçlarının tespiti, bu ihtiyaçların karşılanmasında sunulan hizmetlerin belirlenmesi ve hastaların bu hizmetleri kullanması için teşvik edilmesi olarak tanımlanmaktadır [5]. Bu sebeple sağlık hizmeti sunucu ve tedarikçilerinin yenilikleri sürekli takip etmeleri ve değişen koşullara uygun politika ve stratejiler geliştirmeleri oldukça önemlidir [6]. Sağlık hizmetleri pazarlaması hem klasik hem de sosyal pazarlamaya özgü belirli kavram, yöntem ve teknikleri kullanması sebebiyle disiplinler arası bir alan olarak görülmektedir. Bu disiplinler arası alana sahip sağlık hizmetleri pazarlamasının en önemli özelliklerinden biri hizmetlerin ve pazarların olması ancak bu hizmet ve pazarların parasal karşılığının olmamasıdır [7]. Sağlık hizmeti kullanıcılarının artan beklentileri, maliyet artışları, sağlık hizmetleri sunumunda yararlanılan kaynakların verimli kullanılmaması ve hizmet kullanıcılarının giderek bilinçli hale gelmesi sağlık sektöründe pazarlama faaliyetlerinin kullanılmasını zorunlu hale getiren nedenler olarak görülmektedir [8]. Başarılı sağlık hizmeti pazarlamasının en büyük göstergesi ise hizmet sağlayıcıları/sağlık kuruluşları ve hastaları arasında açık bir iletişim hattı geliştirerek yeni hastalara ulaşmak ve hasta bakımını iyileştirmek için pazarlama ve iletişim stratejileri geliştirme süreci olarak görülmektedir [9].

Sağlık hizmetlerinin pazarlamasının temel amacı, hedef pazarın beklenti ve ihtiyaçlarını anlayarak uygun sağlık hizmetleri sunmaktır. Bu hedefe ulaşmak için sağlık çalışanları ile hastalar arasındaki etkileşim adımları belirlenmeli ve yapılan faaliyetlerin hasta ihtiyaçlarına

uygunluğu değerlendirilmelidir. Bu yönetim sürecinin, sağlık hizmeti sunucusu ile tüketicisi olan müşteri arasındaki değer alışverişinde memnuniyeti artırıcı bir katkı sağladığı belirtilmektedir [10]. Müşteri deneyimini geliştirme ve sağlık hizmeti sunucularının müşterilerin ihtiyaçlarını anlamalarının ise, sağlık hizmetleri pazarlamasında kritik bir rol oynadığı savunulmaktadır. Hastaların geri bildirimini, sağlık hizmetlerinin karmaşıklığını ve hastaların değişen ihtiyaçlarını anlama süreçlerini etkiler. Bu tür araştırmalar, sağlık hizmetlerinin verimliliğini artırma ve hastaların deneyimini iyileştirme potansiyeli taşır [11, 12]. Hastaların geri bildirimini, sağlık hizmetleri pazarlamasında müşteri bağlılığının önemini vurgular. İşletmeler için maliyet açısından avantajlı olan müşteri bağlılığı, sürekli değer sunarak duygusal bağlılık oluşturmayı ve aynı sağlık hizmeti sağlayıcısını tercih etmeyi teşvik eder. Bu bağlamda, müşteri bağlılığının işletmelere sağladığı avantajlar arasında satış işlemlerini kolaylaştırma, piyasa istikrarı, fiyat kontrolü, müşteri sadakatini artırma ve güçlü ilişkiler kurma bulunmaktadır [13, 14, 15]. Bu nedenle sağlık hizmetleri, müşteri odaklı strateji ile müşteri memnuniyetini artırmayı, sadakati sağlamayı ve sektörde daha fazla müşteri çekmeyi hedeflemektedir. Bu strateji, sektördeki payın, kalite ve marka değerinin artırılması için kilit stratejilerin belirlenmesine ve uygulanmasına olanak tanımaktadır. Sağlık hizmetleri pazarlaması, sağlık hizmeti sunucularına müşteri ihtiyaçlarını anlama, doğru hizmetleri sunma ve müşteri memnuniyetini artırma konusunda rehberlik ederek sağlık hizmetleri pazarının büyümesine ve hastaların olumlu sonuçlar elde etmelerine katkı sağlamaktadır [12].

Sağlık hizmetleri pazarlarını tipik ürün pazarlarından ayıran en önemli özelliklerinden biri sağlık hizmetlerinin genellikle doğru veya ahlaki bir zorunluluk olarak görülmesidir. Bu durum öncelikle sağlık hizmetlerine erişimi teşvik etmeye yönelik politikalar yoluyla ortaya çıkan sağlık hizmetleri piyasalarını düzenlemede kamu sektörü için büyük bir rol oynamaktadır. Daha açıklayıcı bir ifade ile tüketicilerin en azından kısmi bir hak ile yüksek kaliteli sağlık bakımı kavramlarına sahip olmaları, serbest piyasaların sağlık hizmeti sunumuna ilişkin sosyal hedeflere ulaşmada başarılı olma ihtimalinin düşük olduğu anlamına gelmektedir [16]. İkincisi ise sağlık sektörünün çeşitli tedarikçi bölümlerinin tüketici hizmetleri sunmak ve bunun için ödeme yapmak üzere tedarik zincirinde faaliyet gösterdiği dikey pazar yapısının klasik bir örneği olmasıdır. Yani sağlık hizmeti sunucuları tıbbi cihazları ve farmasötik ürünleri üst düzey sağlayıcılardan tedarik etmekte ve heterojen koşul ve ihtiyaçlara sahip tüketicilere bakım sağlamak için uzmanlıklar düzeyinde çalışmaktadır [17].

Türkiye’de sağlık hizmeti sunan özel kuruluşlar bünyesinde sunulan hizmetleri hakkında bilgi vermek ve potansiyel tüketicileri kendi işletmeleri kapsamına çekebilmek için çeşitli pazarlama stratejileri uygularken kamuda sağlık hizmeti sunan kuruluşlar sadece sosyal pazarlamaya yönelik faaliyetlerde bulunmaktadır [18, 19]. Bununla birlikte koruyucu sağlık hizmetlerinin ön

plana çıkması beraberinde sağlık sektöründe sosyal pazarlama faaliyetlerinin de yaygın olarak kullanılmaya başlanmasına sebep olmuştur [20]. Sağlık sektöründe kullanılan sosyal pazarlama faaliyetleri genel olarak sigara, alkol ve madde bağımlılığının zararları, obezite ile mücadele, salgın hastalıkların önlenmesi, aşı kampanyaları vb. gibi toplum sağlığına yönelik hazırlanmış olan tanıtım kampanyaları ve kamu spotlarını içermektedir [8, 19].

Son yıllarda sosyal medyanın yaygınlaşması kişilerin duygularını çeşitli konularda paylaşmak için ideal bir platform haline gelmiş ve genel kamuoyu üzerinde önemli bir etki bırakmıştır [21]. Özellikle Twitter gibi sosyal medya platformları, kullanıcılara diğer kişilerin düşüncelerini takip etme, yorum yapma ve görüşlerini anlık olarak paylaşma imkânı tanımanın yanı sıra araştırmacıların da yoğun ilgisini çekmektedir [22]. Twitter verileri pazarlama ve birçok alanda müşteri istek ve ihtiyaçlarını belirlemek için geleneksel veri toplama yöntemlerine alternatif olarak kullanılabilir [23]. Sağlık alanında Twitter'dan elde edilen veriler yaygın hastalıkların tanımlanması, kriz yönetimi gibi birçok farklı alan için kullanılabilir [24]. Halk sağlığı alanında, Twitter verilerinin sıkça kullanıldığı ve bu platform üzerinden yapılan bilgi paylaşımının halk sağlığı uzmanlarının ilgilendikleri bir alan olarak dikkat çektiği gözlemlenmiştir [25]. Özellikle sigara kullanımı, diyabet, obezite, şizofreni ve bağışıklama gibi konularda yapılan paylaşımların, halk sağlığında Twitter verilerinin en yoğun kullanıldığı alanlar olduğu görülmüştür [26]. Nitekim Twitter verileri kullanılarak diyabet hastaları üzerinde yapılan duygu analizlerinde tip2 diyabet hastalarının daha çok negatif, tip1 diyabet hastalarının ise pozitif duygu durumu içerisinde buldukları ortaya konulmuştur. Diyabet konulu Twitter paylaşımlarının genel olarak 'nötr' duygu durumunu yansıttıkları görülmüştür [27]. Yine Twitter'ın elektronik sigara konusunda önemli bir pazarlama platformu olduğu ve elektronik sigara konulu Twitter paylaşımları daha çok ticari amaçlı olsa da önemli bir kısmın sigara bırakmaya teşvik ettiği görülmüştür. Bu noktada sigara bırakmaya yönelik Twitter paylaşımlarının halk sağlığı üzerinde olumlu etkilerinin olacağı vurgulanmıştır [28]. Bir başka çalışmada ise Twitter paylaşımlarının halk sağlığı alanında yaygınlıkla kullanıldığı bu sayede toplumun HPV aşısı gibi bağışıklama konusundaki yaklaşımlarını anlamak için Twitter verilerinin önemli bir bilgi kaynağı olduğu savunulmuştur [29]. Ayrıca, halk sağlığını iyileştirmeye yönelik kampanyaların da bu verilerin içinde öne çıkan konular arasında bulunmuştur [20]. Bu sayede halk sağlığı kampanyaları daha geniş kitlelere ulaştırılabilmekte, farkındalık düzeyi yükseltilebilmekte ve bilgi paylaşımı konusunda dezavantajlı gruplara erişimde kolaylıklar sağlanabilmektedir [30, 31, 32]. Türkiye'de ise sağlıkla ilgili mesaj ve bilgi iletiminde Twitter'ın yaygınlıkla kullanıldığı ancak sosyal medya aracılığı ile halk sağlığı alanında güvenilir olmayan bilgi paylaşımlarının önlenmesi için izlenim ve denetimlerin yapılması gerektiği vurgulanmaktadır [33]. Sağlık hizmeti kullanıcılarının en çok hangi konularda Twitter paylaşımı yaptıkları incelendiğinde kullanıcıların daha çok kanıt ihtiyacı

duyulan iddialar ve kişisel deneyimler hakkında Twitter paylaşımlarında buldukları ve sağlıkta hizmet işletmelerinin üretim işletmeleri ile karşılaştırıldığında Twitter paylaşımlarının daha fazla beğenildiği gözlemlenmiştir [34, 35].

Yapılan çalışmalarda sağlıkta Twitter verilerinin paylaşımı konulu çalışmalarda daha çok halk sağlığı perspektifinden sağlıklı yaşam konusunda bilgi paylaşımı, kronik hastalıkların etkin yönetimi ve bağışıklama konularının ön plana çıktığı görülmektedir. Pazarlama sağlıkta yasal ve etik gibi nedenlerle sınırlı olmasına rağmen sosyal medya platformları sağlıkta pazarlama konusunda geniş kitlelere ulaşım imkânı vermekte, halk sağlığını iyileştirme konusunda öncelikli platformlardan birisi olarak ön plana çıkmaktadır. Sağlıkta Twitter verileri kullanılarak halk sağlığını iyileştirmeye yönelik paylaşımlar ve bunların duygu durumu analizleri yapılmış olmasına rağmen sağlık hizmetlerinde pazarlama bakış açısı ile özel bir çalışmanın daha önce yapılmadığı görülmüştür. Bu eksikliğin giderilmesi amacıyla bu çalışmada Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik görüşlerinin metin madenciliği yöntemleri kullanılarak belirlenmesi amaçlanmıştır. Çalışma kapsamında Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik tutumlarını değerlendirmek amacıyla elde edilen tweetler duygu analizi yöntemi kullanılarak olumlu, olumsuz ve nötr ruh hallerine göre incelenerek çalışma kapsamında en sık tekrarlanan kelimeler için kelime bulutları ile analizler görsel olarak desteklenecektir. Çalışmanın bir sonraki adımında analizler kapsamında kullanılan tweet metinleri içerisinde 'health system', 'insurance', 'mental health' kelimelerinin geçip geçmediği ve 'retweet' ve 'favorite (beğeni)' edilip edilmediğine göre kategorik hale getirilip yeni bağımsız değişkenler oluşturulacaktır. Oluşturulan bu bağımsız değişkenler ile Twitter kullanıcılarının sağlık hizmetleri pazarına yönelik olumlu, olumsuz ve nötr duygu durumlarına yönelik Yapay Sinir Ağları (Neural Network-NN), Rastgele Orman (Random Forest-RF), Naive Bayes (NB) ve k-en yakın komşu (k-nearest neighbor k-NN) gibi makine öğrenmesi teknikleri kullanılarak performans değerlendirmeleri yapılacaktır. Son adımda en iyi performans gösteren makine öğrenmesi tekniği belirlendikten sonra Twitter veri kullanıcılarının duygusal durumlarının en iyi tahmin edicilerini belirlemek için Pisagor ağacı kullanılacaktır.

2. MATERYAL VE METOT (MATERIAL AND METHOD)

Bu çalışmada sosyal medya kullanıcılarının sağlık hizmetleri pazarlamasına yönelik görüşlerinin analiz edilmesi amacıyla Twitter'dan toplanan tweetler kullanılmıştır. Bu çalışmada tweet toplamak için Twitter API Key kullanılmıştır. Bu kapsamda RStudio yazılım programında 'rtweet' kütüphanesi kullanılarak 'healthcaremarket' etiketiyle atılmış olan tweetler çalışmaya dahil edilmiştir. Çalışmaya dahil edilen tweetler için herhangi bir lokasyon kısıtlaması yapılmamışken, sadece İngilizce dilinde atılan tweetler kullanılmıştır. Sağlık hizmetleri pazarlaması ile ilgili olarak 1 Ekim 2022 ve 31 Kasım 2022 tarihleri arasında atılan tweetler haftalık

olarak toplanmıştır. Toplamda sekiz haftadan oluşan 27079 tweet çalışmamız kapsamına dahil edilmiştir. Veri setinde yer alan tweetlere Orange programında yapılan ön işlemede ‘transformation’ aşamasında veri setindeki giriş verileri kelimelerin küçük harfe dönüştürülmesi, metindeki vurguların kaldırılması, html etiketlerinin ayrıştırılması, metindeki url bağlantılarının kaldırılması gibi işlem adımları uygulanmıştır. Sonraki adım olan ‘tokenization’ aşamasında Regexp yöntemi kullanılarak veri setindeki aşamaları anlamlı hale dönüştürülmüştür. Son olarak ‘filtering’ aşamasında ise metindeki gereksiz kelime, bağlaç, noktalama işaretleri vs. kaldırılmıştır. Uygulanan bu işlem adımlarından sonra veri setinde en sık kullanılan kelimelerin görselleştirilmesi amacıyla kelime bulutları oluşturulmuştur.

Çalışmamız kapsamındaki veri seti ile Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik duygu durumlarının belirlenmesi amacıyla duygu analizi yapılmıştır. Sosyal medyanın kullanımının yaygınlaşması ile birlikte insanlar bir durum, yer, konu veya olay hakkındaki duygularını, inançlarını veya fikirlerini ifade etmek için Facebook, Twitter ve Instagram gibi platformları sıklıkla kullanılmaktadır [36]. Duygu analizi veya Fikir Madenciliği kişilerin sosyal medya gibi sanal ortamlarda fikir, düşünce, duygu ve tutumlarının yazılan metinlerden analiz eden çalışma alanı olarak tanımlanmaktadır [37]. Duygu Analizi ile kişiler tarafından yazılmış olan metinler girdi olarak kullanılmakta ve bu metinlerin içerisinde saklı olan duyguların pozitif, negatif veya nötr olma durumları kategorize edilmektedir [38, 39]. Son yıllarda sosyal medya platformlarında bulunan yapılandırılmamış biçimde bulunan büyük hacimli verilerden elde edilen duyguları veya fikirleri tespit etmek amacıyla duygu analizi sıklıkla kullanılmaktadır [36]. Siyaset, sosyoloji, ekonomi ve finans gibi farklı sosyolojik alanlarda kullanılan duygu analizi çalışmaları sosyal medya kullanımının yaygınlaşmasıyla birlikte ön plana çıkmıştır [40]. Çalışmamız kapsamında kullanılan veri setinde Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik duygularının analiz edilmesinde duyguların kutuplaşmasını (pozitif/negatif/nötr) ve yoğunluğunu (güç) dikkate alan bir duygu analizi yöntemi olan VADER (Valence Aware Dictionary for Sentiment Reasoning) modeli kullanılmıştır [41]. VADER, kural tabanlı bir duygu analiz aracı ve sosyal medyada duyguları ifade etmek için kullanılan bir sözlüktür [42]. Herhangi bir eğitim verisine ihtiyaç duyulmayan bu yöntem kullanılarak önceden işlenmiş tweetlerin genel puanı hesaplanır ve çıktı olarak bir tweet’in pozitif, negatif ve nötr puanları verilir [43, 44].

Çalışmamızın bir sonraki aşamasında Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik görüşleri duygu analizi ile analiz edildikten sonra elde edilen üç duygu durumunun en iyi tahmin edicilerini belirlemek için yeni bağımsız değişkenler oluşturulmuştur. Bu yeni veri kümesini oluşturmak için sekiz haftalık veri setleri birleştirilerek tek bir veri seti oluşturulmuş ve bu orijinal Twitter metin belgesi ile yeni bir kelime bulutu oluşturulmuştur. Oluşturulan bu yeni kelime bulutunda en

sık tekrarlanan kelimeler gözden geçirilmiş ve sırasıyla “health system, insurance ve mental health” kelimelerinin sıklıklarının yüksek olduğu görülmüştür. Sağlık hizmetlerinde pazarlama konusunda Twitter verilerinin analiz edildiği çalışmaların eksikliği göz önüne alınarak sıkça tekrarlanan bu kelimelerin Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik duygu durumlarını belirleme etkili olabileceği ve bu konuda literatüre katkı sağlanacağı düşünülerek bu kelimeler kullanılarak yeni bağımsız değişkenler oluşturulmuştur. Ayrıca bu kelimelerin yanı sıra kullanılan tweetlerin retweet ve favorite edilip edilmediği gibi özellikle Twitter platformunda yüksek etkileşim getiren değişkenler de yeni veri setine dahil edilmiştir. Metin formatında olan değişkenlerin analize dahil edilebilmesi için ‘health system’, ‘insurance’ ve ‘mental health’ kelimelerinin metin içerisinde geçip geçmemesine (0=hayır, 1=evet) ve ‘retweet’ ve ‘favorite’ edilip edilmemesine göre (0= retweetlenmemiş/beğenilmemiş, 1=retweetlenmiş/beğenilmiş) ikili kategoriye ayrılmış değişkenler ile birlikte toplamda beş bağımsız değişken oluşturulmuştur. Tüm değişkenler aynı forma getirildikten sonra oluşturulan bu bağımsız değişkenlerin veri seti kullanılarak Twitter kullanıcılarının görüşlerinin duygu durumlarının en iyi tahmin edicilerini belirlemek için makine öğrenmesi tekniklerinin performansları karşılaştırılmıştır. Bu çalışma kapsamında NN, RF, NB ve k-NN teknikleri kullanılmış ve Twitter veri kullanıcılarının görüşlerinin duygusal gruplamalarını tahmin etmek için karşılaştırılmıştır.

NN biyolojik sinir ağlarına dayanan bir matematiksel hesaplama modelini temsil etmektedir ve genellikle sınıflandırma problemlerini çözmek için kullanılmaktadır [45, 46]. Bir veri kümesindeki temel ilişkileri anlamaya yönelik algoritmalar dizisi olan NN, insan beyninin işleyişini taklit eden bir süreç aracılığıyla çalışmaktadır [47]. Nöron adı verilen bir dizi basit, bağlantılı işlemciden oluşmakta ve organik veya yapay nöron sistemlerini yansıtarak çeşitli girdilere esnek bir şekilde uyum sağlayabilmektedirler [48, 49]. NN gibi doğrusal olmayan istatistiksel veri modelleme araçlarının kullanılması, girdiler ve çıktılar arasında karmaşık ilişkileri modellemeye yardımcı olarak veri kümelerindeki gizli kalıpları da ortaya çıkarabilmektedir [50]. N girişi olan bir nöronun çıktısını hesaplamak için aşağıdaki formül kullanılır (Eşitlik 1) [51]. Y çıkışı, x_i girişi, w_i ağırlığı, b yanlılığı ve f ise transfer fonksiyonunu temsil eder [52].

$$Y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (1)$$

RF birden fazla karar ağacının kombinasyonundan oluşan bir modeli temsil etmektedir [53]. Bu yöntem birden fazla karar ağacı oluşturmak için genel eğitim kümesinden özellik alt kümesini rastgele seçen ve ardından bireysel karar ağaçlarının ortalama tahmini olarak bilinen model sınıflandırmasını döndüren bir karar ağacı türüdür [54, 55, 56]. Tahmin modelinin başarı seviyesi yeterli öğrenme sağlandıktan sonra doğrulama verilerinin test edilmesiyle belirlenir ve tahmin sonuçlarının doğruluğu ve güvenilirliği karar ağaçları arttıkça artar [53, 57]. Son derece ayarlanabilir ve kullanıcı dostu bir denetimli

öğrenme yaklaşımıdır ve hem regresyon hem de sınıflandırma amacıyla kullanılabilir [58, 59]. RF büyük miktardaki verinin doğru bir biçimde sınıflandırılması ve karmaşık veri yapılarının analizi için uygun bir seçenek olarak görülmektedir. Bu nedenle sınıflandırma algoritmaları arasında diğer modellerle karşılaştırıldığında sürekli olarak en yüksek tahmin doğruluğunu sunmaktadır [60, 61]. Özellik alt kümesi seçimi ve eksik veri ataması için kullanılan bu yöntem oldukça etkilidir ve gelecekteki veri eklemeleri için yüksek performans sunabilir. RF, büyütme, kendini test etme, ağaç birleştirme ve sonlandırma gibi konularda belirli talimatları takip eder ve aşırı uyum konusunda dayanıklıdır. Ayrıca çok yüksek boyutlu parametre uzaylarında ve aykırı değerlerin varlığında diğer ML algoritmalarına göre daha kararlı olduğu söylenmektedir [62]. Karar ağacındaki her düğümün entropisi şu şekilde verilir (Eşitlik 2):

$$ni_j = w_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)}, \quad (2)$$

Burada ni_j , düğüm j 'nin önemini, w_j düğüm j 'ye ulaşan örneklerin ağırlıklı sayısını, C_j düğüm j 'nin karışıklık değerini, $left(j)$, düğüm j 'nin üzerinde sol bölümündeki alt düğümü ve $right(j)$ ise düğüm j 'nin sağ bölümündeki alt düğümü ifade etmektedir.

Bayes Teoremi temel alınarak geliştirilmiş bir istatistiksel sınıflandırma yöntemi olan NB ise basit ve koşullu olasılıklara dayanmaktadır [63]. Hem çok kategorili hem de ikili sınıflandırma için kullanılabilen NB çok sınıflı sınıflandırmanın sınıflandırması için yetersiz akış ve aşırı uyum gibi iki ana sınırlamanın da üstesinden gelir [64, 65, 66]. Bu yöntemde sınıf tahmini diğer özelliklerden bağımsız olarak ele alınır ve bir özelliğin sınıf üzerindeki etkisinin diğer özelliklerden bağımsız olduğu varsayılır [67]. Bağımlı özelliklere sahip sınıfları sınıflandırmak için en uygun olan, güçlü bir koşullu bağımsızlık varsayımına sahip olasılıksal bir sınıflandırıcıdır [68]. Bu yöntem oluşturulan olasılıklar ile modellemedeki belirsizlikleri yakalamaya çalışarak tahmin problemlerinin çözülmesine yardımcı olmaktadır [69]. NB yönteminin uygulanması kolaydır ve genellikle iyi performans değerlerine sahiptir [70]. Çünkü NB, kullanılan her değişken için sonlu olasılıkları hesaplayabilir, çoklu sınıf yoğunluk tabanlı sınıflandırmada sorunları çözmek için sıkça kullanılır ve bu yöntemde belirli bir özelliğin değeri, sınıf değişkenine verilen herhangi bir başka özelliğin değerinden bağımsızdır [71]. Kullanılan NB teoreminin temeli aşağıdaki Bayes formülüdür (Bkz. Eşitlik 3).

$$P(C | X) = \frac{P(X | C)P(C)}{P(X)} \quad (3)$$

$P(C | X)$ son olasılıktır,
 $P(X | C)$ olasılıktır,
 $P(C)$ sınıf öncesi olasılıktır,
 $P(X)$ tahmini ön olasılıktır.

Son olarak en yakın komşuların hesaplanmasına dayanan k-NN ise sınıflandırma ve regresyon problemlerinin çözümünde kullanılan denetimli bir makine öğrenimi tekniğidir [49, 72]. Bu algoritmanın temel prensibi, aynı sınıfa ait en benzer örneklerin yüksek olasılığa sahip olmasıdır. Genellikle, k-NN algoritması bir sorgunun eğitim veri setindeki en yakın k komşusunu belirleyerek, bu k komşudaki ana sınıfla sorguyu tahmin etmektedir [73]. K-NN algoritması, yeni veriler ile önceden tanımlanmış veriler arasındaki benzerliği varsayar ve yeni verileri mevcut kategoriler arasında en uygun olan kategoriye dahil eder [58]. Veri madenciliği, görüntü işleme vb. alanlarda birçok uygulamada kullanılan bu algoritma, sınıflandırılacak nesnenin eğitim verilerine olan en yakın mesafesini hesaplamak için kullanılır. Bu mesafenin Öklid Uzaklığı yöntemi kullanılarak hesaplanması için aşağıdaki formül kullanılır (Eşitlik 4) [74]. Boyutların veya özelliklerin sayısı n 'dir, x_i veri testindeki i özelliği, y_i ise veri dizisindeki i özelliğidir.

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Veri analizinde, en uygun modeli seçmek için literatürde kabul görmüş çeşitli model performansı değerlendirme yöntemleri bulunmaktadır. Sağlık hizmetleri pazarına yönelik durumlarını sınıflandırmak için çeşitli makine öğrenimi tekniklerini karşılaştırmak amacıyla kullanılan performans yöntemleri şu şekildedir:

Doğruluk (Accuracy): Sınıflandırma modelinin doğru tahminlerinin, toplam veri noktalarına oranıyla hesaplanan yaygın bir metriktir ve model tahmininin doğru olma olasılığını hesaplar [75, 76]. Doğruluk hesaplanırken payda olarak karmaşıklık matrisinin tüm girdilerinin toplamı, pay olarak ise Gerçek Pozitif ve Gerçek Negatif ögelerin toplamı alınır [76]. Doğruluk denklemi Eşitlik 5'te gösterilmiştir.

$$Doğruluk = \frac{GP+GN}{GP+YN+YP+GN} \quad (5)$$

Kesinlik (Precision): Kesinlik, pozitif olarak tahmin edilen değerlerin gerçekte kaçının pozitif olduğunu gösterir. Gerçek Pozitifler, model tarafından pozitif olarak etiketlenen öğeleri ifade ederken Yanlış Pozitifler, pozitif olarak sınıflandırılan ancak gerçekte negatif olan öğeleri temsil eder [76, 77]. Doğruluk denklemi Eşitlik 6'da gösterilmiştir.

$$Kesinlik = \frac{GP}{GP+YP} \quad (6)$$

Duyarlılık (Recall): Duyarlılık doğru sınıflandırılmış pozitiflerin gerçek pozitiflere oranıdır. Yani pozitif olarak tahmin edilenlerin, toplam pozitif vakaların ne kadarını kapsadığını göstermektedir. Duyarlılık hesaplanırken Yanlış Negatifler göz önüne alınır. Bu ölçütte temel olan gerçek pozitiflerin test sonuçlarında doğru bir şekilde tahmin edilmesidir [78]. Doğruluk denklemi Eşitlik 7'de gösterilmiştir.

tekniklerinin performans değerleri Tablo 3'te gösterilmektedir. Buna göre Tablo 3'te görüldüğü gibi çalışmada kullanılan makine öğrenmesi tekniklerinin performansları ROC Eğrisi Altındaki Alan (AUC), Sınıflandırma Doğruluğu (CA), F1, Kesinlik ve Duyarlılık değerleri açısından değerlendirilmiştir. Elde edilen bulgular doğrultusunda en yüksek tahmin performansına sahip olan tekniğin RF (AUC=0,707; CA=0,646), ikinci en iyi tahmin performansına sahip olan tekniğin NN (AUC=0,706; CA=0,645) ve üçüncü

en iyi tahmin performansına sahip olan tekniğin ise NB (AUC=0,702; CA=0,646) olduğu belirlenmiştir. Bu makine öğrenmesi teknikleri arasında k-NN (AUC=0,501; CA=0,177) tekniğinin en düşük tahmin performansına sahip olan teknik olduğu görülmüştür. Çalışma bulguları RF ve NN'in diğer makine öğrenmesi teknikleriyle karşılaştırıldığında çok sınıflı Twitter verisi kullanıcılarının duygu durumlarını tahmin etmede daha yüksek sınıflandırma performanslarına sahip olduğunu doğrulamıştır.

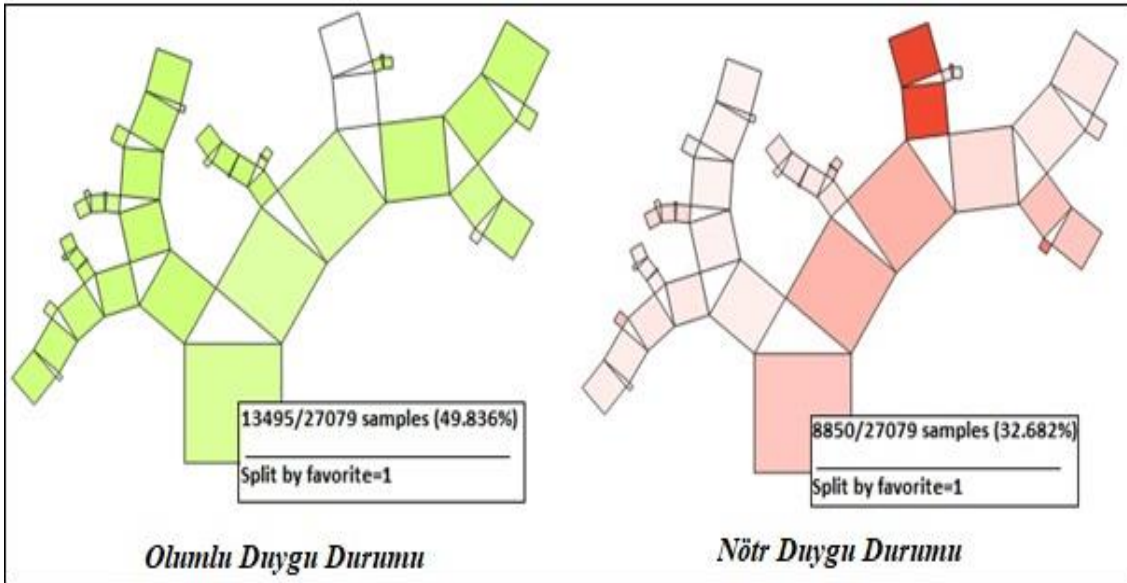
Tablo 3. Makine öğrenmesi tekniklerinin performanslarının karşılaştırılması (Comparison of performance of machine learning techniques)

Model	AUC	CA	F1	Kesinlik	Duyarlılık
Rastgele Orman (Ağaç sayısı=10)	0.736	0.677	0.608	0.627	0.677
Neural Network	0.735	0.677	0.609	0.694	0.677
Naive Bayes	0.730	0.638	0.577	0.530	0.638
k-NN	0.502	0.177	0.060	0.630	0.177

Açıklamalar: AUC: Area Under the ROC Curve-ROC Eğrisi Altındaki Alan; CA: Classification Accuracy-Sınıflandırma Doğruluğu

Çalışmamızın son adımında verilerin görselleştirilmesi ve Twitter kullanıcılarının duygu sınıflarının belirlenmesinde en etkili değişkenin görülmesi için Pisagor ağacı yöntemi kullanılmıştır. Şekil 3 incelendiğinde sağlık hizmetleri pazarlamasına yönelik kullanıcı sayısının en fazla olduğu pozitif (n=13495) ve nötr (n=8850) duygu sınıflarından elde edilip RF sınıflandırma modeli kullanılarak oluşturulan Pisagor ağacına göre en etkili değişkenin 'favorite' değişkeni olduğu görülmüştür. Bu değişkenin her iki duygu sınıfı için Pisagor ağaçlarının ana karar düğümünü oluşturduğu görülmüştür. Bu durum, Twitter kullanıcılarının sağlık hizmetleri pazarlaması

hakkındaki duygu sınıflarının belirlenmesinde 'favorite' değişkeninin önemli bir rol oynadığını göstermektedir. 'Favorite' değişkeninin ana karar düğümü olması, kullanıcıların paylaşımlarında öne çıkan içeriklere odaklanırken, bu içeriklerin pozitif ve nötr duyguları nasıl etkilediğinin anlaşılmasına yardımcı olabilmektedir. Bu bulgular, sağlık hizmetleri pazarlaması stratejilerinin Twitter kullanıcıları üzerinde olumlu etkiler yaratma potansiyeline sahip olduğunu göstermektedir. Diğer değişkenler ele alındığında sırasıyla 'health system', 'insurance', 'retweet', ve 'mental health' değişkenlerinin Twitter kullanıcılarının duygu sınıflarının belirlenmesinde en etkili olan diğer değişkenler olduğu belirlenmiştir.



Şekil 3. Olumlu ve nötr duygu durumunun pisagor ağacı (Pythagorean tree of positive and neutral moods)

4. TARTIŞMA (DISCUSSION)

Sağlık hizmetleri hızlı büyüyen bir sektör olması sebebiyle pazarlama uygulamalarını sıklıkla kullanmakta ve hizmet sunucu, hizmet kullanıcı ve hükümetler açısından önemli bir rol oynamaktadır [19, 83]. Twitter gibi sosyal medya araçları son yıllarda birçok sektörde olduğu sağlık sektöründe de kullanıcıların görüşlerinin, beklentilerinin ve isteklerinin belirlenmesi konusunda sıklıkla başvurulan bir platform haline gelmiştir. Bu nedenle bu çalışma Twitter kullanıcılarının sağlık hizmetleri pazarlamasına ilişkin algılarının incelenmesi amacıyla yapılmıştır. Çalışmamız kapsamında 1 Ekim 2022 ve 30 Kasım 2022 tarihleri arasında ‘healthcaremarket’ etiketiyle elde edilen 27079 adet tweetten elde edilen veriler üzerinde analizler gerçekleştirilmiştir. Bu kapsamda ilk olarak Twitter’den elde edilen tweetler üzerinde duygu analizi yapılarak sağlık hizmetleri pazarlamasına yönelik algıların çoğunlukla olumlu olduğu tespit edilmiştir. Literatüre bakıldığında benzer şekilde Sevim ve Gündük [19]’ün kamu sağlık hizmetleri yöneticilerinin sağlık hizmetlerinde pazarlama faaliyetlerine bakış açılarının değerlendirilmesine yönelik yaptıkları çalışmada sağlık hizmetleri yöneticilerinin kamu sağlık hizmetlerinde pazarlama faaliyetlerine yönelik olumlu bir bakış açısına sahip oldukları tespit edilmiştir. Yine pazarlama alanında yapılan başka bir çalışmada ise müşteri beğeni ve memnuniyetsizliklerinin oluşmasında ön plana çıkan faktörlerin belirlenmesi amacıyla duygu analizi kullanılarak yapılan çalışmalarda negatif yorumların pozitif ve nötr yorumlara kıyasla daha fazla olduğu görülmüştür [23]. Yani değişen ve gelişen dünya ile birlikte diğer hizmet alanlarında olduğu gibi sağlık hizmetleri alanında da pazarlama uygulamalarının olumlu etkileri olduğunu söylemek mümkündür. Literatür incelendiğinde Twitter kullanımının özellikle kronik hastalıklar konusunda bilgi paylaşımı konusunda sıklıkla kullanıldığı görülmektedir. Başka bir çalışmada ise göğüs kanseri hastalarını Twitter üzerinden yaptıkları paylaşımlar duygu durumu analizi ile incelendiğinde hasta tedavisi, hasta desteğinin artırılması ve farkındalığın yaygınlaştırılması konularında olumlu deneyimlerin paylaşıldığı görülmüştür [84].

Ekim ve Kasım ayları için oluşturulan kelime bulutlarına göre en sık tekrarlanan kelimelerin ‘insurance’ ve ‘mental’, ‘care’, ‘drug’ ve ‘system’ gibi sağlık hizmetleri pazarlamasıyla doğrudan ilişkili kelimeler olduğu görülmüştür. Çalışma kapsamında yer alan Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik olumlu, olumsuz ve nötr duygu durumları belirlendikten sonra bu duygu durumlarını tahmin etmede hangi yöntemin daha etkili olduğu sorusuna cevap aranmıştır. Bu noktada aradığımız sorunun cevabını belirlemek için makine öğrenmesi teknikleri kullanılarak performans karşılaştırmaları yapılmıştır. Literatüre bakıldığında Twitter metinleri üzerinde makine öğrenmesi tekniklerinin kullanımının yaygın olduğunu ve Türkçe Twitter mesajları üzerinde duygu analizlerinin yapıldığı bir çalışmada Naive Bayes algoritması, destek vektör makineleri ve lojistik regresyon analizi gibi makine öğrenmesi sınıflama tekniklerinin kullanıldığı görülmüştür [85]. Bu nedenle

çalışma kapsamında beş adet bağımsız kategorik değişkenden oluşan yeni bir veri seti oluşturularak bu yeni veri seti üzerinden makine öğrenmesi tekniklerinin performans karşılaştırmaları yapılmıştır. Yapılan performans karşılaştırmaları sonucunda Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik duygu durumlarını tahmin etmeden en iyi makine öğrenmesi tekniğinin RF ve NN algoritmaları olduğu tespit edilmiştir. Son olarak yeni oluşturulan veri setinde yer alan ‘favorite’ değişkeninin Twitter kullanıcısının duygu durumlarını açıklamada en etkili değişken olduğu belirlenmiştir. Bu durumda Twitter kullanıcıların sağlık hizmetleri pazarlamasına yönelik atılan tweetlere çoğunlukla katıldığı ve olumlu bir yaklaşım sergilediği söylenebilir. Son yıllarda Twitter sağlık alanında Dünya Sağlık Örgütü gibi küresel örgütler tarafından uluslararası kamuoyuyla iletişim kurmak için önemli bir halkla ilişkiler aracı haline gelmiştir. Pandeminin sağlık alanında politika belirleme ve karar verme üzerindeki etkilerinin gözlenmeye devam edildiği 2022 yılı itibariyle DSÖ Twitter paylaşımları üzerinde yapılan bir incelemede, Twitter kullanıcılarının DSÖ’nün Covid-19 ile ilgili paylaşımlarına tepki (favorite-retweet) verdikleri görülmüştür [86].

Çalışmamızda toplumdaki kişilerin sağlık hizmetlerinin pazarlamasına yönelik olumlu duygu durumları belirlenmiş ve bu duygu durumları ileri derecede analiz teknikleri kullanılarak daha detaylı bir şekilde analiz edilmiştir. Çalışmanın sınırlılıkları göz önüne alındığında çalışma kapsamına sadece İngilizce dilinde ve ‘healthcaremarket’ etiketi kullanılarak atılan tweetler dahil edilmiştir. Ayrıca bu çalışmada sosyal medya platformlarından sadece Twitter’ı kullanan kişilerden elde edilen veriler kullanılmıştır. Son olarak çalışmanın bir diğer sınırlılığı da sadece sekiz haftalık bir zaman periyodu için veri setinin oluşturulmasıdır.

5. SONUÇ VE ÖNERİLER (CONCLUSION AND RECOMMENDATIONS)

Başta sağlık olmak üzere birçok alanda sunulan hizmetlere ilişkin kamuoyu görüşleri ve bu görüşlerin en çok paylaşıldığı sosyal medya platformları son dönemde en önemli bilgi kaynaklarından biri haline gelmiştir. Bu çalışma kapsamında da Twitter’den elde edilen verilerle kişilerin sağlık hizmetleri pazarlaması konusundaki görüşleri duygu analizi kullanılarak değerlendirilmiştir. Araştırmamız sonucunda elde edilen bulgular Twitter kullanıcılarının büyük çoğunluğunun sağlık hizmetleri pazarlaması konusunda olumlu bir görüşe sahip olduğunu göstermiştir. Çalışmanın ileri analiz aşamasında Twitter kullanıcılarının duygu durumlarının en iyi tahmin edicilerini bulmak amacıyla oluşturulan yeni veri seti kullanılarak makine öğrenmesi tekniklerinin performansları karşılaştırılmıştır. Bu analiz sonucunda Twitter kullanıcılarının duygu durumlarının en iyi tahmin edici makine öğrenmesi teknikleri RF ve NN olmuştur. Twitter kullanıcılarının sağlık hizmetleri pazarlamasına yönelik duygu durumlarının en iyi tahmin edici değişkenlerinin ise sırasıyla ‘favorite’, ‘health system’,

'insurance', 'retweet', ve 'mental health' olduğu belirlenmiştir. Sosyal medya platformlarından elde edilen güncel bilgilerin politika yapımcılar için önemli bir kaynak olduğu düşünülmektedir. Bundan sonra yapılacak olan çalışmalarda sağlık hizmetleri pazarlamasına yönelik literatürde yer alan sınırlı çalışmalara ek olarak sağlık hizmetleri pazarlamasına yönelik görüşlerin değerlendirilmesinde Twitter'ın yanı sıra farklı sosyal medya platformlarının da çalışma kapsamına dahil edilmesi ve bu platformlardan elde edilen bulguların karşılaştırılması önerilmektedir. Yine ileride yapılacak çalışmaların zaman içindeki değişiklikleri daha iyi anlamak ve derin bir perspektif sunmak amacıyla daha uzun bir zaman aralığını kapsayacak şekilde planlanması, ayrıca analizlerin genellenebilirliğini artırmak ve bulguların daha geniş bir kitleye uygulanabilir olmasına katkıda bulunmak için daha büyük bir veri seti kullanılması önerilmektedir. Ayrıca farklı etiketler kullanılarak daha geniş bir perspektif ile sağlık hizmetleri pazarlamasına yönelik değerlendirmeler ele alınabilir.

KAYNAKLAR (REFERENCES)

- [1] S. Öztürk, O. Uçan, "Türkiye'de Sağlık Harcamalarında Artış Nedenleri: Sağlık Harcamalarında Artış-Büyüme İlişkisi", *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22(1), 139-152, 2017.
- [2] TÜİK, Sağlık Harcamaları İstatistikleri 2021. <https://data.tuik.gov.tr/Bulten/Index?p=Sağlık-Harcamaları-İstatistikleri-202145728#:~:text=Ki%20C5%9F%20ba%20C5%9F%20C4%B1na%20sa%20C4%9F%20C4%B1k%20harcamas%20C4%B1%202020,bin%20206%20TL%20y%20y%20C3%BCKseldi.&text=Toplam%20sa%20C4%9F%20C4%B1k%20harcamas%20C4%B1n%20C4%B1n%20GSYH%20ye,%254%20C6%20olarak%20ger%20C3%A7ekle%20C5%9Fti,2022.>
- [3] A. Ekiyor, A. Çetin, "Sağlık Hizmeti Sunumunda ve Sosyal Pazarlama Kapsamında E-Nabız Uygulamasının Bilinirliği", *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 3(1), 88-103, 2017.
- [4] O. Kitapci, I. T. Dorytol, Z. Yaman, M. Gulmez, "The Paths From Service Quality Dimensions to Customer Loyalty: An Application on Supermarket Customers", *Management Research Review*, 36(3), 239-255, 2013.
- [5] D. F. Biçer, A. Yıldırım, "Sağlık Hizmetleri Pazarlaması Çerçevesinde İç Müşteri Tatmini Etkileyen Faktörler", *Stratejik ve Sosyal Araştırmalar Dergisi*, 4(2), 381-403, 2020.
- [6] B. H. Yeşilkaya, B. Akalın, "Sağlık Hizmetleri Pazarlaması Üzerine Bir Model Önerisi: Kamu ve Özel Hastane Örneği", *Sosyal Araştırmalar ve Yönetim Dergisi*, (1), 105-120, 2022.
- [7] V. L. Purcarea, "The Impact of Marketing Strategies in Healthcare Systems", *Journal of Medicine and Life*, 12(2), 93-96, 2019.
- [8] Ö. Yeşilyurt, "Covid-19 Pandemi Sürecinin Yönetilmesinde Sosyal Pazarlama Aracı Olarak Kamu Spotlarının İncelenmesi: İçerik Analizi", *OPUS International Journal of Society Researches*, 17(Pandemi Özel Sayısı), 3470-3500, 2021.
- [9] J. Knott, Healthcare marketing 2022: Strategies & Tips to Grow Practices. <https://intrepy.com/importance-healthcare-marketing,2022.>
- [10] E. V. L. Purcarea, "The Impact of Marketing Strategies in Healthcare Systems", *Journal of Medicine and Life*, 12(2), 93, 2019.
- [11] C. Bulut, T. Sönel, D. Kolca, "Sağlık Hizmetlerinde Pazarlama Araştırmalarında Uygulanan Teknikler", *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 9(3), 284-294, 2023.
- [12] S. Gök, "Sağlık Hizmetleri Pazarlamasında Güncel Yaklaşımlar", *Sağlık Akademisyenleri Dergisi*, 10(3), 514-522, 2023.
- [13] Ö. Çelen, H. Abuhanoğlu, A. Teke, A. "Memnuniyet, Bağlılık ve Yalnızlık İlişkisi: TSK Ankara Özel Bakım Merkezinde Bir Araştırma", *TAF Preventive Medicine Bulletin*, 15(1), 2016.
- [14] V. Öter, H. D. Südaş, "Algılanan Hizmet Kalitesinin Hasta Bağlılığı Üzerine Etkisi: Devlet Hastanesi Üzerine Bir İnceleme", *Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21(2), 43-57, 2017.
- [15] M. Marangoz, L. Biber, "Kurumsal İmajın ve Kurumsal Ünün Müşteri Bağlılığına Etkileri", *Karamanoğlu Mehmetbey Üniversitesi Sosyal ve Ekonomik Araştırmalar Dergisi*, 2007(2), 173-193, 2017.
- [16] G. B. D. Kıyat, H. Şimşek, B. Özgüleş, "Sağlık Hizmetleri Pazarlamasının Diğer Pazarlama Dalları İçindeki Yeri ve Önemi", *Sağlık Akademisyenleri Dergisi*, 4(3), 147-154, 2017.
- [17] B. Handel, K. Ho, "The Industrial Organization of Health Care Markets", *Handbook Of Industrial Organization*, Cilt 5, Editör: Ho, K., Hortaçsu, A., Lizzeri A., North Holland, Amsterdam, 521-614, 2021.
- [18] K. Rooney, "Consumer-Driven Healthcare Marketing: Using the Web to Get up Close and Personal", *Journal of Healthcare Management*, 54(4), 241-251, 2009.
- [19] E. Sevim, Ö. Güdük, "Kamu Sağlık Hizmetleri Yöneticilerinin Sağlık Hizmetlerinde Pazarlamaya Bakış Açılarının Değerlendirilmesi İstanbul Örneği", *Jaren*, 6(1), 148-156, 2020.
- [20] G. Bayın, Y. Akbulut, "Sağlık Sektöründe Sosyal Pazarlamanın Kullanımı", *Ankara Sağlık Bilimleri Dergisi*, 1(1), 53-72., 2012.
- [21] V. S. Pagolu, K. N. Reddy, G. Panda, B. Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements," **2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)**, IEEE, 1345-1350, 2016.
- [22] J. Leskovec, L. A. Adamic, B. A. Huberman, "The Dynamics of Viral Marketing", *ACM Transactions on the Web (TWEB)*, 1(1), 1-39, 2007.
- [23] K. Ö. Atılgan, H. Yoğurtcu, "Kargo Firması Müşterilerinin Twitter Gönderilerinin Duygu Analizi", *Çağ Üniversitesi Sosyal Bilimler Dergisi*, 18(1), 31-39, 2021.
- [24] A. Onan, "Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi", *Yönetim Bilişim Sistemleri Dergisi*, 3(2), 1-14, 2017.
- [25] C. B. S. Bartlett, R. Wurtz, "Twitter and Public Health", *Journal of Public Health Management and Practice*, 21(4), 375-383, 2015.

- [26] L. Sinnenberg, A. M. Bittenheim, K. Padrez, C. Mancheno, L. Ungar, R. M. Merchant, "Twitter as a Tool for Health Research: A Systematic Review", *American Journal of Public Health*, 107(1), e1-e8, 2017.
- [27] E. Gabarron, E. Dorrnzoro, O. Rivera-Romero, R. Wynn, "Diabetes on Twitter: A Sentiment Analysis", *Journal of Diabetes Science and Technology*, 13(3), 439-444, 2019.
- [28] J. Huang, R. Kornfield, G. Szczypka, S. L. Emery, "A Cross-Sectional Examination of Marketing of Electronic Cigarettes on Twitter", *Tobacco Control*, 23(suppl 3), iii26-iii30, 2014.
- [29] J. Keim-Malpass, E. M. Mitchell, E. Sun, C. Kennedy, "Using Twitter to Understand Public Perceptions Regarding the HPV Vaccine: Opportunities for Public Health Nurses to Engage in Social Marketing", *Public Health Nursing*, 34(4), 316-323, 2017.
- [30] M. Faus, F. Alonso, A. Javadinejad, S. A. Useche, "Are Social Networks Effective in Promoting Healthy Behaviors? A Systematic Review of Evaluations of Public Health Campaigns Broadcast on Twitter", *Frontiers in Public Health*, 10, 1045645, 2022.
- [31] I. Herrera-Peco, B. Jiménez-Gómez, J. J. Peña Deudero, E. Benitez De Gracia, C. Ruiz-Núñez, "Healthcare Professionals' Role in Social Media Public Health Campaigns: Analysis of Spanish Pro Vaccination Campaign on Twitter", *Healthcare*, 9(6), 662, 2021.
- [32] K. McCausland, B. Maycock, T. Leaver, K. Wolf, B. Freeman, J. Jancey, "E-Cigarette Advocates on Twitter: Content Analysis of Vaping-Related Tweets". *JMIR public Health and Surveillance*, 6(4), e17543, 2020.
- [33] A. A. Çobaner, S. Köksoy, "Sağlık Alanında Sosyal Medyanın Kullanımı: Twitter'da Sağlık Mesajları", *Akademik Konferans Bildirileri*, 899, 906, 2014.
- [34] J. L. Lee, M. DeCamp, M. Dredze, M. S. Chisolm, Z. D. Berger, "What are Health-Related Users Tweeting? A Qualitative Content Analysis of Health-Related Users and Their Messages on Twitter", *Journal of Medical Internet Research*, 16(10), e237, 2014.
- [35] S. Leek, D. Houghton, L. Canning, "Twitter and Behavioral Engagement in the Healthcare Sector: An Examination of Product and Service Companies", *Industrial Marketing Management*, 81, 115-129, 2019.
- [36] A. Hasan, S. Moin, A. Karim, S. Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts", *Mathematical and Computational Applications*, 23(1), 11, 2018.
- [37] G. Sarıman, E. Mutaf, "Covid-19 sürecinde Twitter Mesajlarının Duygu Analizi", *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, 7(10), 137-148, 2020.
- [38] B. Akın, U. T. G. Şimşek, "Sosyal Medya Analitiği ile Değer Yaratma: Duygu Analizi ile Geleceğe Yönelim", *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(3), 797-811, 2018.
- [39] B. Özyurt, M. A. Akçayol, "Fikir Madenciliği ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma", *Selçuk Üniversitesi Mühendislik, Bilim ve Teknoloji Dergisi*, 6(4), 668-693, 2018.
- [40] A. B. Eliacik, N. Erdoğan, User-weighted sentiment analysis for financial community on Twitter. **2015 11th International Conference on Innovations in Information Technology (IIT)**, IEEE, 2015.
- [41] R. K. Mishra, S. Urolagin, J. A. A. Jothi, A.S. Neogi, N. Nawaz, "Deep Learning-Based Sentiment Analysis and Topic Modeling on Tourism During Covid-19 Pandemic", *Frontiers in Computer Science*, 3, 1-14, 2021.
- [42] S. Elbagir, J. Yang, "Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment", **In Proceedings of the International Multiconference of Engineers and Computer Scientists**, IMECS, 1-5, 2019.
- [43] S. Vashishtha, S. Susan, "Fuzzy Rule Based Unsupervised Sentiment Analysis from Social Media Posts", *Expert Systems with Applications*, 138, 112834, 2019.
- [44] C. Çılgın, M. Baş, H. Bilgehan, C. Ünal, "Twitter Sentiment Analysis During Covid-19 Outbreak with Vader. *AJIT-e: Academic Journal of Information Technology*, 13(49), 72-89, 2022.
- [45] I.K.M. Jais, A.R. Ismail, S.Q. Nisa, "Adam Optimization Algorithm for Wide and Deep Neural Network", *Knowledge Engineering and Data Science*, 2(1), 41-46., 2019.
- [46] Y. Singh, A. S. Chauhan, "Neural Networks in Data Mining", *Journal of Theoretical & Applied Information Technology*, 5(1), 37-42, 2009.
- [47] X. Qi, G. Chen, Y. Li, X. Cheng, C. Li, "Applying Neural-Network-Based Machine Learning to Additive Manufacturing: Current Applications, Challenges, And Future Perspectives", *Engineering*, 5(4), 721-729, 2019.
- [48] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview", *Neural Networks*, 61, 85-117, 2015.
- [49] B. Mahesh, "Machine Learning Algorithms-A Review", *International Journal of Science and Research (IJSR)*, 9, 381-386, 2020.
- [50] B.G. Marcot, A. M. Hanea, "What is an Optimal Value of k in K-Fold Cross-Validation in Discrete Bayesian Network Analysis?", *Computational Statistics*, 36(3), 2009-2031, 2021.
- [51] M. H. Beale, M. T. Hagan, H. B. Demuth, **Neural Network Toolbox User's Guide. The MathWorks Inc, Natick, A. B. D.**, 2015.
- [52] J. Amita, J. S. Singh, G. P. Kumar, "Prediction of Bus Travel Time Using Artificial Neural Network", *International Journal for Traffic and Transport Engineering*, 5(4), 410-424, 2015.
- [53] C. M. Yeşilkanat, "Spatio-Temporal Estimation of the Daily Cases of COVID-19 in Worldwide Using Random Forest Machine Learning Algorithm", *Chaos, Solitons & Fractals*, 140, 1-8, 2020.
- [54] O. Sevli, "Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması", *Avrupa Bilim ve Teknoloji Dergisi*, (16), 176-185, 2019.
- [55] M. Luo, J. Xie, Y. Yan, Z. Ke, P. Yu, Z. Wang, J. Zhang, "Comparing Machine Learning Algorithms in Predicting Thermal Sensation Using ASHRAE Comfort Database II", *Energy and Buildings*, 210, 1-16, 2020.
- [56] F. E. Ayo, O. Folorunso, F. T. Ibaralu, I. A. Osinuga, "Machine Learning Techniques for Hate Speech Classification of Twitter Data: State-of-the-Art, Future Challenges and Research Directions", *Computer Science Review*, 38, 100311, 1-34, 2020.

- [57] W. Lin, Z. Wu, L. Lin, A. Wen, J. Li, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis", *IEEE Access*, 5, 16568-16575, 2017.
- [58] G. P. A. Mary, M. S. Hema, R. Maheshprabhu, M. N. Guptha, Sentimental Analysis of Twitter Data Using Machine Learning Algorithms. **2021 International Conference on Forensics, Analytics, Big Data, Security (FABS)**, IEEE, 1, 1-5, 2021.
- [59] A. P. Rodrigues, R. Fernandes, A. Shetty, K. Lakshmana, R. M. Shafi, "Real-time Twitter Spam Detection and Sentiment Analysis Using Machine Learning and Deep Learning Techniques", *Computational Intelligence and Neuroscience*, 2022, 1-14, 2022.
- [60] Y. Y. Aung, M. M. Min, An Analysis of Random Forest Algorithm Based Network Intrusion Detection System. **2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)**, IEEE, 127-132, 2017.
- [61] J. L. Speiser, M. E. Miller, J. Tooze, E. Ip, "A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling", *Expert Systems with Applications*, 134, 93-101, 2019.
- [62] A. Sarica, A. Cerasa, A. Quattrone, "Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review", *Frontiers in Aging Neuroscience*, 9, 329, 2017.
- [63] I. B. A. Peling, I. N. Arnawan, I. P. A. Arthawan, I. G. N. Janardana, "Implementation of Data Mining to Predict Period of Students Study Using Naive Bayes Algorithm", *International Journal of Engineering and Emerging Technology*, 2(1), 53-57, 2017.
- [64] M. Bader Alazzam, H. Mansour, M. M. Hammam, S. Alsheikh, A. Bakir, S. Alghamdi, A. S. AlGhamdi, "Machine Learning of Medical Applications Involving Complicated Proteins and Genetic Measurements", *Computational Intelligence and Neuroscience*, 2021, 1-6, 2021.
- [65] B. Chandra, and Manish Gupta, "Robust Approach for Estimating Probabilities in Naive Bayesian Classifier for Gene Expression Data," *Expert Systems with Applications*, 38(3), No. 3, 1293-1298, 2011.
- [66] D. M. Farid, M. M. Rahman, M. A. Al-Mamuny, "Efficient and Scalable Multi-Class Classification Using Naive Bayes Tree", *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, 1-4, 2014.
- [67] K. A. Govindasamy, N. Palanichamy, Depression Detection Using Machine Learning Techniques on Twitter Data. **2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)**, IEEE, 960-966, 2021.
- [68] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, P. Badhani, B. Tech, "Study of Twitter Sentiment Analysis Using Machine Learning Algorithms on Python", *International Journal of Computer Applications*, 165(9), 29-34, 2017.
- [69] H. Parveen, S. Pandey, Sentiment analysis on Twitter Data-Set Using Naive Bayes Algorithm. **2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)**, IEEE, 2016.
- [70] S. Ray, A Quick Review of Machine Learning Algorithms. **2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)**, IEEE, 2019.
- [71] K. A. K. Niazi, W. Akhtar, H. A. Khan, Y. Yang, S. Athar, "Hotspot Diagnosis for Solar Photovoltaic Modules Using A Naive Bayes Classifier", *Solar Energy*, 190, 34-43, 2019.
- [72] I. Triguero, D. García-Gil, J. Mailló, J. Luengo, S. García, F. Herrera, "Transforming Big Data into Smart Data: An Insight on the Use of the K-Nearest Neighbors Algorithm to Obtain Quality Data", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(2), e1289,1-24, 2019.
- [73] S. Zhang, D. Cheng, Z. Deng, M. Zong, X. Deng, X. "A Novel kNN Algorithm with Data-Driven k Parameter Computation", *Pattern Recognition Letters*, 109, 44-54, 2018.
- [74] Y. D. Setiyaningrum, A. F. Herdajanti, C. Supriyanto, Classification of Twitter Contents Using Chi-Square and K-Nearest Neighbour Algorithm, **2019 International Seminar on Application for Technology of Information and Communication**, IEEE, 2019.
- [75] M. Ikonomakis, S. Kotsiantis, V. Tampakas, "Text Classification Using Machine Learning Techniques", *WSEAS Transactions On Computers*, 4(8), 966-974, 2005.
- [76] M. Grandini, E. Bagli, G. Visani, "Metrics for Multi-Class Classification: An Overview", *ArXiv abs/2008.05756*: 1-17, 2020.
- [77] V. Yarğı, S. Postalcioglu, "EEG İşareti Kullanılarak Bağımlılığa Yatkınlığın Makine Öğrenmesi Teknikleri ile Analizi", *El-Cezeri*, 8(1), 142-154, 2021.
- [78] B. Ş. Bozyiğit, Ç. Tarhan, "Makine Öğrenmesi Yöntemleri Kullanılarak Sosyal Medyada Marka İtibar Analizi", *Yönetim Bilişim Sistemleri Dergisi*, 6(2), 57-76, 2020.
- [79] L. Teia, "Special Case of the Three-Dimensional Pythagorean Gear (Australian Senior Mathematics Journal)", *Journal of the Australian Mathematical Society*, 32, 36-49, 2018.
- [80] T. Munz, M. Burch, T. van Benthem, Y. Poels, F. Beck, D. Weiskopf, Overlap-Free Drawing of Generalized Pythagoras Trees for Hierarchy Visualization. **2019 IEEE Visualization Conference (VIS)**, IEEE, 2019.
- [81] E. W. Ambarsari, H. Herlinda, "Pythagoras Tree Applied for Determined Instagram Usage Habit Decision", *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 4(2), 56-61, 2020.
- [82] U. Urbas, D. Zorko, N. Vukašinović, "Machine Learning Based Nominal Root Stress Calculation Model for Gears with A Progressive Curved Path of Contact", *Mechanism and Machine Theory*, 165, 1-14, 2021.
- [83] A. Y. Sarıyıldız, "Sağlık Hizmetleri Pazarlamasında Yeni Yaklaşımlar", *Selçuk Sağlık Dergisi*, 3(2), 166-179, 2022.
- [84] E. M. Clark, T. James, C. A. Jones, A. Alapati, P. Ukandu, C. M. Danforth, P. S. Dodds, "A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter", *ArXiv Preprint ArXiv:1805.09959*, 2018
- [85] A. Onan, "Sentiment Analysis on Twitter Messages Based on Machine Learning Methods", *Yönetim Bilişim Sistemleri Dergisi*, 3 (2), 1-14, 2017.

- [86] B. A. Özoran, “Bir Halkla İlişkiler Aracı Olarak Twitter: Dünya Sağlık Örgütü Paylaşımlarının İçerik Analizi ve Metin Madenciliği ile İncelenmesi”, *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 20(04), 125-14, 202

Konum Tabanlı Yeni Yöntem ile Acil Ulaşım Sorunlarına Karar Destek Sistemleri: Ambulans Örnek Uygulama Canlandırması

(Çam ve Sakura Hastanesi Örneği)

Araştırma Makalesi/Research Article

 Hasan VARLI¹,  Nuri BİNGÖL²

¹Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, Üsküdar Üniversitesi, İstanbul, Türkiye

²Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, Üsküdar Üniversitesi, İstanbul, Türkiye

hasanvarli@hotmail.com.tr, nuri.bingol@uskudar.edu.tr

(Geliş/Received:05.02.2024; Kabul/Accepted:25.04.2024)

DOI: 10.17671/gazibtd.1431518

Özet— Acil yardım hizmetlerinde hayati öneme sahip olan ambulanslar, günün her saati trafikte mahsur kalmaktadır. Henüz tam anlamıyla ambulansa yol veremeye ilgili oturmuş bir yöntemin olmaması, araç kalabalığı, acemi ve duysuz sürücüler nedeniyle ambulansın olay yerinden hastaneye ulaşması güçleşmektedir. Bu çalışma kapsamında, ambulansın en kısa sürede hastaneye varması amaçlanmıştır. Ambulans konumu ve bilgi panelleri konumları arasında Haversine formülü yardımıyla mesafe hesaplanarak bilgi paneli olarak adlandırılan ekranlarla diğer sürücüleri erkenden bilgilendirerek trafiğin akışını bozmadan daha güvenli bir şekilde sol şerit boşaltılacak ve ambulansın geçişini hızlanacak. İki farklı canlandırma yapılmıştır. İlkinde, ambulans logosunu yansıtmaları için ölçülmesi gereken mesafe değerleri 0.4 – 0.9 km aralığındadır. Çıktı olarak hesaplanan değerleri 0.04 – 0.65 km aralığındadır. Ambulans yakındayken ışık ve sesle ikaz zaten etkili bir yöntem olduğundan dolayı faydanın yeterli değildir. İkincisinde, bu mesafe değerleri 1.5 – 2 km aralığında ve gösterim süresi 125 – 200 saniye aralığındadır. Çıktı olarak hesaplanan değerleri ise 0.97 – 1.62 km aralığındadır. Bu mesafeden ambulansın ışık ve sesinin uyarıcılığı zor olacağından bu yöntemde elde edilen fayda artmıştır. Yapılan çalışma sonucunda, uygun değerler verildiğinde çözüm oldukça faydalı olacaktır. Bu çalışma, olay yerine varılması, diğer yolların da kapsanması, diğer acil yardım araçlarının da dâhil edilmesi anlamında gelecekteki çalışmalara ışık tutacaktır.

Anahtar Kelimeler—acil ulaşım, akıllı acil ulaşım planlama, cbs, coğrafi konum, haversine formülü

Decision Support Systems for Emergency Transportation Problems with a New Location-Based Method: Ambulance Sample Application Animation

(Çam and Sakura Hospital Example)

Abstract— Ambulances, crucial for emergency services, often face traffic congestion, hindering their swift arrival at hospitals. The lack of a clear protocol for yielding to ambulances, combined with inexperienced drivers and traffic jams, poses challenges for timely hospital transfers. This study aims to expedite ambulance arrivals by using the Haversine formula to calculate distances between ambulance locations and information panels. These panels inform drivers early, facilitating left lane clearance without traffic disruption, thus accelerating ambulance passage. Two different simulations were conducted. In the first, distance range for logo reflection: 0.4 – 0.9 km. Calculated distance values as output: 0.04 – 0.65 km. When the ambulance is nearby, light and sound insufficient. In the second, distances: 1.5 – 2 km, display time: 125 – 200 seconds. Calculated distance values as output: 0.97 – 1.62 km. It becomes harder to hear the lights and sirens and the benefit obtained in this study increases. Applying appropriate values renders this solution highly beneficial, ensuring swift hospital arrivals. This study paves the way for future research, addressing on-scene arrival, alternative routes, and integration of other emergency vehicles into the system.

Keywords— emergency transportation, geographical location, gis, haversine formula, smart emergency transportation plannin

1. GİRİŞ (INTRODUCTION)

Günümüzde konum bilgisi pek çok ihtiyaç dâhilinde çeşitli amaçlar için kullanılmaktadır [1]. Mevcut yeri bulmak, gidilmek istenen konuma rota oluşturmak, havacılık, araç takip sistemleri, spor uygulamaları ve meteorolojik tahminler, bankacılık uygulamaları, kadastral ölçümler için kullanıldığı gibi alışveriş uygulamalarının civardaki mağazalarla ilgili kampanyaları sunması için de kullanılmaktadır [2].

Konumlar hayati öneme sahip olan sağlık alanında da yoğun olarak kullanılmaktadır. Acil çağrı merkezini arayan birinin otomatik hassas konumunun belirlenmesinde, araçlarda bulunan araç içi acil çağrı sistemlerinde, EGM tarafından sunulan KADES uygulamasında ve arama kurtarma çalışmalarında da konumun bildirilmesi büyük öneme sahiptir. Bazı markaların akıllı saatleri de yine benzer bir çalışma mantığıyla kazayı algılayarak acil durum araması yapabilmekte ve saate ait konumu paylaşabilmektedir.

Konum tabanlı çözümlerin hayati etkisinin önemi açıkça görülmektedir. Özellikle acil ulaşım aracı olan ambulansın hastayla birlikte bir an önce hastaneye varması büyük önem teşkil etmektedir. Bu hedef doğrultusunda, özellikle büyük şehirlerde trafik sorunu ile yüzleşmesi gerekmektedir. Ambulans sürücüsü gerektiğinde ışıklarını ve sireni kullanarak diğer sürücüleri ikaz etmektedir. Buna istinaden birçok duyarlı sürücü ambulansın hızlı yol alabilmesi için yol vermeye çabalamaktadırlar. Trafik yoğun olduğu durumlarda ambulanslar emniyet şeridini de kullanabilmektedir. Ayrıca son zamanlarda, ambulansın araçların ortasından gitmesini sağlayan, sol şeritte bulunan araçların sola yaklaşması, sağ şeritte bulunan araçların ise sağa yaklaşmasıyla ambulansın araçların ortasından geçişini yapması şeklinde açıklanan fermuar yöntemi de mevcuttur [3]. Fakat bu yöntemlerin hiçbiri tek başına ya da birlikte iken tam anlamıyla çözüm sunmamaktadır.

Trafikte çok sayıda duyarlı sürücünün olmasının yanında nadirde olsa daha az duyarlı ve yol vermede geciken sürücüler, ambulansın hastaneye varış hızını kesmektedir. Bununla beraber acemi, çevik hareket edemeyen ya da yüksek sesle müzik dinlemek gibi dış çevre duyarlılığını azaltan etkenlere sahip sürücüler de vardır. Yüksek kapasiteli kentsel yolların her bölümünde emniyet şeridi bulunmayabilir. Bu durumda diğer sürücülerle aynı şeritleri kullanmaları gerekmektedir. Emniyet şeridi olsa dahi bazı diğer sürücülerin gereksiz kullanımı nedeniyle yine zor anlar yaşanmaktadır. Fermuar yöntemi de az bilinmesi ve diğer belirtilen nedenlerle ülkemizde uygulanması kolay olmayan bir yöntemdir.

Bu çalışmanın en temel amacı, konum bilgisinden yararlanılarak ambulansın en kısa sürede hızlı ve güvenli bir şekilde hastaneye varabilmesini desteklemektir.

Yüksek kapasiteli kentsel yolların en sol şeridinin üstüne, bilgi paneli olarak adlandırılan ledlerden oluşan bir

ekranın, L tipinde bir direğe monte edilerek konduğu varsayılmaktadır. Trafikğin hemen her saat olduğu bu yollarda, ambulansın sol şeridi kullanacağı kuralı bilinmektedir. Bilgi panellerine yaklaşmakta olan bir ambulans olduğunda en yakın bilgi paneli tespit edilmektedir. Tespit edilen bu bilgi panelinde, ambulansın kısa bir süre sonra bulunduğu konumdan geçiş yapacağına dair anlam ifade eden ambulans logosunu ekrana yansıtılmaktadır ve diğer sürücüler bilgilendirilmektedir. Böylece sol şeridin önceden boşaltılarak ambulansın hastaneye en kısa sürede ve hızlıca varabilmesi desteklenmektedir. Bu temel amacın yanında dolaylı amaçlar ise, yol verme çabası içerisinde olan diğer sürücüler, bir an önce yol verme telaşına kapılmaması ve trafikğin tehlikeye atılmaması, diğer sürücülerde duyarlılığın artması ve belli bir yol verme yönteminin yerleşmesi şeklinde sıralanmaktadır.

2. GENEL TANIMLAR (GENERAL DEFINITIONS)

2.1. GPS (Global Positioning System)

Uydular üzerinden konum belirleme kavramı işitildiğinde ilk akla gelen uydu sistemidir. ABD tarafından geliştirildiği ilk yıllarda askeri amaç için kullanılmaktaydı. Sonraki yıllarda sivil kullanımlar için de imkân verilmiştir [2]. Böylece Dünya üzerinde çok geniş bir kullanıma sahip olmuştur. Bu geniş kullanım, temelde 31 uydu ile sağlanmaktadır [4].

2.2. TCP/IP (Transmission Control Protocol/Internet Protocol)

TCP/IP, internetin temelini temsil etmektedir ve cihazlar arasındaki iletişimin sağlanması için kullanılmaktadır. TCP kısmı, veriyi paketlere ayırır ve karşı tarafta birleştirilmesini sağlar. IP kısmı, bu paketlerin iletileceği adresi ifade etmektedir. TCP portu 21 numaralı port üzerinden çalışmaktadır [5].

2.3. MQTT Protokolü (Message Queuing Telemetry Transport Protocol)

MQTT protokolünün açılımı, Message Queuing Telemetry Transport olarak bilinen Türkçe karşılığı mesaj kuyruğu telemetri taşıma protokolüdür. TCP/IP üzerinden çalışan mesajlaşma protokolüdür. Broker yapısı üzerinden mesajlar dağıtılmaktadır. Yayıncı(Publisher)-Abone(Subscriber) temelli hafif bir protokol olmasıyla beraber, düşük bant genişliği ve düşük güç tüketimi ile düşük maliyetli ve birçok farklı türdeki düşük donanımlı cihazlarda dahi rahatlıkla çalıştırılabilmektedir [6].

Topic denen konular üzerinden yayıncılar yayın yapmaktadır ve aboneler, adından da anlaşılabilceği üzere bu konulara abone olmaktadır. Abone olunan konularda yayınlanan mesajlar alınmaktadır. Yayınlanan mesajlar payload kavramı üzerinden gönderilmektedir. Herhangi mesaj paylaşıldığında anında mesajları alabilmektedirler [7].

Yayıncının, belirli konular üzerinden brokera mesaj göndermesi yayın yapmak olarak adlandırılmaktadır. Yapılan bu yayınlar, aboneler tarafından alınır ve işlemlere göre değerlendirilmektedir [6].

2.4. JSON (Javascript Object Notation)

JSON, Javascript'e dayanan popüler ve hafif bir veri formatıdır. Verileri, taşıma, depolama ve transfer etmek için kullanılmaktadır. JSON verinin okunması ve yazılması oldukça kolaydır. Genellikle bir isimlendirme ve değer çiftinden oluşmaktadır. Bunun yanında, NoSQL veritabanı sistemlerinde de kullanılmaktadır [8].

2.5. Şifreleme Algoritmaları (Encryption Algorithms)

Günümüzde yaygın olarak pek çok şifreleme algoritması bulunmaktadır. Bu tez kapsamında kullanılan algoritmalarla ilgili alt başlıklar halinde bilgi paylaşılmaktadır.

2.5.1. Rijndael Algoritması (Rijndael Algorithms)

Rijndael algoritması, farklı blok ve anahtar boyutlarıyla kullanılabilir. 128 bitlik olarak ayrılan verilerden oluşan paket bloklarını 128, 192 ve 256 bit anahtar seçenekleriyle şifreleme yapabilen güvenli ve yaygın bir şifreleme algoritmasıdır. Şifrele işlemindeki tur sayısı boyuta göre değişmektedir. Blok boyutu 128bit olduğunda anahtar boyutu 128, 192 ve 256 bit seçildiğinde sırasıyla 10,12 ve 14 turda şifreleme işlemi yapılmış olacaktır. Tur sayısı arttıkça daha güvenli olmaktadır. Fakat daha yüksek hesaplama gücüne ihtiyaç olmaktadır [9].

2.5.2. SHA256 Algoritması (SHA256 Algorithms)

SHA256, şifrelenmek istenen girdinin uzunluğuna bakmaksızın sabit boyutlu olarak 256 bitlik benzersiz özet oluşturmaktadır. En güvenilir algoritmalar arasında yer almaktadır.

2.6. Yeryüzündeki İki Nokta Arasındaki Mesafe Ölçümü (Measurement of Distance Between Two Points on Earth)

Konum ve navigasyon sistemlerinde iki nokta arasındaki mesafeyi doğru bir şekilde hesaplamak oldukça önemlidir. Bu hesaplamalar için Vincenty ve Haversine formülleri sıklıkla kullanılır. Vincenty, özellikle daha uzun mesafelerde, daha doğru sonuçlar verirken Haversine, daha yakın mesafelerde, ortalama olarak %0.5 oranında ufak sapmalarla sonuçlar verir. Haversine formülünde dünya tam bir küre olarak kabul edilir. Haversine, Vincenty formülüne göre daha basit bir hesaplama yöntemine sahip oluşu ve buna bağlı olarak daha az işlem gücü ihtiyacıyla daha hızlı sonuçlar üretebildiği için tercih edilmiştir. [10]. Fonksiyon ve formüllerde kullanılan sembollerin anlamları şu şekildedir. θ , merkez açısı [11], R, dünyanın yarıçapı olan 6371 kilometreyi, D, iki nokta arasındaki mesafeyi ifade etmektedir. Φ_1 , Φ_2 sembolleri iki noktanın enlemlerini ifade ederken, λ_1 , λ_2 sembolleri iki noktanın boylamlarını ifade etmektedir [12]. İki nokta arasındaki merkez açı formülü, Şekil 1'deki gibidir.

$$\theta = \frac{d}{r}$$

Şekil 1. Merkez Açısı Formülü (Central Angle Formula)

Bu formülde, merkez açısı bulabilmek için r değerini bildiğimiz için bunun yanında iki nokta arasındaki mesafeyi de bilmemiz gerekiyor. İki nokta arasındaki mesafeyi bulabilmek için haversine formülünden yararlanılmaktadır [13]. Haversine fonksiyonu Şekil 2'deki gibidir.

$$\text{haversine}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$$

Şekil 2. Haversine Fonksiyonu (Haversine Function)

Haversine formülü Şekil 3'deki gibidir. Hesaplamalarda bu formülden yararlanılmaktadır.

$$\left(\frac{d}{r}\right) = \text{haversine}(\Phi_2 - \Phi_1) + \cos(\Phi_1) \cos(\Phi_2) \text{haversine}(\lambda_2 - \lambda_1)$$

Şekil 3. Haversine Formülü (Haversine Formula)

2.7. Acil Sağlık Hizmetleri (Emergency Health Services)

İnsanlar, herhangi zamanda aniden gelişen sağlık sorunu ya da yaralanmalar yaşayabilmektedir. Bunlar, bazen hayati olabilmektedir.

Acil sağlık hizmetleri, hastaneye ulaşım öncesinde acilen ilk müdahalenin yapılması konusunda gerekli eğitimleri almış kişiler tarafından olay yerine intikalini sağlanması, hastanın en yakın sağlık kuruluşuna ulaştırılması ve sağlık kuruluşlarında verilen tüm sağlık hizmetlerini kapsamaktadır. Aynı zamanda yetkin olmayan kişiler tarafından ilkyardıma doğru yapılamayıp fayda yerine zarara sebep olan kişilerin ve ihmal kaynaklı ölümlerin önüne geçilmesi konularında da sorumluluğu artmaktadır.

3. LİTERATÜR TARAMASI (LITERATURE REVIEW)

"Acil Hizmet Araçlarının GPS Tabanlı Denetim ve Organizasyonu" adlı doktora tezinde, acil hizmet servislerinin denetimi, organizasyonu ve müdahale sorunları üzerine çalışmıştır [14]. Bilinmeyen, tespit edilemeyen, yanlış adresler, gerçek olmayan aramalar, şehirlerin gittikçe büyümesi, hızlı ve çarpık kentleşme, trafik yoğunluğu gibi sebepleri göz önünde bulundurarak yaşanan gecikmeyi vurgulamıştır. Bu sorunların çözümü için ambulansa eklenen GPS destekli bir ünite ile merkezde bulunan sistem ile haberleşmesini sağlayan bir otomasyon düşünülmüştür. GPS tarafından radyo sinyalleri ile merkeze iletilen konum bilgisi, merkezdeki yazılım ile aracın konumu gösterilmektedir. Çağrı yapan kişilerin telefon numaraları bilindiği için numaraya ait geçmiş kayıtlar, gerçek dışı arama tespit edildiyse bu duruma ait bilgiler gösterilmektedir. Aynı zamanda kişiye ait iletişim ve sağlık bilgileri de gösterilmektedir. Bu sayede konum daha hızlı belirlenmekte, ambulans çıkış süreleri azalmaktadır.

“Coğrafi Bilgi Sistemleri ile En Uygun Ambulans Yerlerinin Belirlenmesi” adlı çalışmada acil sağlık hizmetlerinde ambulansın hayatımızdaki önemine, müdahale süresine vurgu yapmaktadır [15]. Ayrıca müdahale aşamasında öncesi ve sonrası ile ilgili hizmet akışı hakkında bilgi verilmiştir. Ambulans istasyonlarının bulunduğu konumlar incelenmiş ve araştırmaya göre literatürde belirlenen 8 dakikalık standart ortalama müdahale süresi göz önünde bulundurularak vakaya müdahale süresine olan etkisi üzerine bilgi paylaşılmıştır. İncelemesi yapılan bölgede mevcut ambulans istasyon konumlarının tam anlamıyla yeterli ve kapsayıcı olmadığı sonucuna ulaşılmıştır. Bu soruna çözüm olarak ambulans istasyonlarının sayısı artırılması gerektiği ve CBS’den faydalanılarak yeni lokasyonların belirlenmesi önerilmiştir. Ayrıca CBS ile gerçeğe en yakın ulaşım verisi elde edilip kurulacak bir merkez üzerinden CBS ve GPS entegrasyonu sağlanarak konumların daha hızlı tespit edilebilmesi ve ambulans için kullanılacak en uygun yolun belirlenerek trafikte daha hızlı yol alınabilmesi görüşü paylaşılmıştır.

“Gaziantep İli 112 Acil Ambulanslarının 3 Yıllık Çalışma Sonuçları” adlı çalışmada, 112 acil ambulanslarında personel sayısı, acil vakalara müdahale sayıları, hastaların ulaşım ve sağlık kurumlarına sevk süreleri, hastalık türlerine ilişkin veriler incelenmiştir [16]. Vakaların cinsiyet ve yaş gruplarına göre dağılımı, istasyon nüfus yoğunlukları ve 112 acil komuta merkezi ile hastaneler arasındaki süreçler araştırılmıştır. Ambulans çıkış sayısının yıllar geçtikçe arttığı belirtilmiştir. Çalışma ile birlikte ambulansların hastaları hastaneye nakil süresinin yıllar geçtikçe rakamsal olarak arttığını ortaya konmuştur. Sonuç olarak, artan vaka sayıları ile ambulansların olay yerine ulaşımı ve hastaneye ulaşımı süresi içerisindeki gecikmeyi azaltmak için planlama yapılması, nakil süreçlerinde yaşanacak olumsuzluklar için hastaneler arası iletişimin kuvvetlenmesinin uygun olacağı paylaşılmıştır.

“Metropolde 112 Acil Sağlık Hizmeti” adlı çalışmada hastaların acil sağlık hizmetleri kapsamında hastane öncesi transport süresi ile elde edilmiş olan demografik veriler üzerinden araştırması yapılmıştır [17]. Üçüncü basamak bir acil serviste 2013 Ocak ayında 112 acil sağlık hizmetlerinde ambulans ile sevk edilen hastalar üzerinde çalışma yapılmıştır. Ambulansın çağrı sonrası çıkış süresi, olay yerine ulaşım süresi ve acil servise ulaşım süresi, trafiğin yoğun olduğu ve olmadığı durumlar açısından belirli saat dilimleri bazında sınıflandırılarak istatistiksel veri olarak verilmiştir. İstatistiksel veriye göre, ambulansın çağrıyı aldıktan sonra olay yerine 10 dakika içinde ulaşma oranı %68.3 olarak hesaplanmıştır. İlk 13 dakikada vakaların %80’ine ulaşılmıştır. Ambulans ile olay yerine gidip sonrasında hastayı ambulans ile hastaneye ulaştırma sürecinde her vakanın doğru değerlendirilemeyip çok da elzem olmayan durumlarda ambulansın ve acil servisin meşgul edildiği görüşü sorun olarak bildirilmiştir. Sonuç olarak, ambulansın hızlı hareketini engelleyecek her türlü durum hasta açısından hayati öneme sahiptir. Doğru vaka tespiti ve hastanın ihtiyacına uygun ekipmanların hastanedeki varlığı gözetilerek merkez tarafından doğru

hastaneye ambulansın yönlendirilmesinin sağlanması tüm hastalar ve ambulanslar için büyük öneme sahiptir.

“IoT Ambulance with Automatic Traffic Light Control” adlı çalışmada hastaya ait bilgilerin gereğinden fazla olması ve ambulansın trafik sinyallerinde yaşanan gecikmeleri ele almıştır. Yapılan çalışma neticesinde hastaya ait sadece temel bilgilerin elde edilmesinin süreçleri kolaylaştırdığı görülmüştür. Aynı zamanda ambulansın trafik sinyallerini kontrol etmesiyle daha hızlı geçiş yapması sağlanmıştır [18].

“Self-Driving Ambulance for Emergency Application” adlı çalışmada ambulansın hızlı, güvenli ve konforlu yol olmasıyla ilgili konulara odaklanılmışlardır. GPRS ve modern tıbbi cihazlarla donatılmış sürücüsüz bir ambulans düşünülmüştür. Bu ambulans, hasta için konforu ön planda tutacak, kaza risklerini azaltacak ve otomatik olarak trafiğin az olduğu yolları tercih ederek en kısa güzergâhtan hastaneye yol alacaktır [19].

İBB iştirak firması olan İSBAK şirketinin ürünü olan Değişken Trafik İşaretleri (DTİ) projesi bulunmaktadır. Maksimum hız limiti, hava durumuna göre güvenli sürüş için sürücülerin daha dikkatli olmaları gibi bilgiler verilmektedir. Ayrıca trafikte geçiş üstünlüğü olan araçlar için kırmızı ışığa takılmasının engellenmesi gibi pilot bir projesi de mevcuttur [20].

Bu pilot çalışması, acil durum araçlarının trafik sinyalizasyon sistemlerinin konumlandırıldığı noktalarda gecikmelerin önüne geçebilmek için yaklaşan bir acil durum aracı olduğunda sinyalizasyon sistemine yeşil ışığı yakması bildirilerek hızlıca yol alması amaçlanmıştır [21].

4. YÖNTEM (METHOD)

Ambulansın olay yerine ulaştıktan sonra en kısa sürede hızlı ve güvenli bir şekilde hastaneye varması üzerine üretilen bu çözümde, ihtiyaç duyulan bilgiler, kısıtlar, bağlılıklar ve varsayım açıkça belirtilmiştir.

Türkiye, İstanbul’da bulunan Başakşehir Çam ve Sakura Hastanesi pilot hedef hastane olarak belirlenmiştir.

Belirlenmiş olan pilot hedef hastaneye gidişini canlandırmak amacıyla 2 örnek rota hazırlanmıştır. Hazırlanan bu rotalar, ambulansın gerçek bir seyrine örnek olacak biçimde belirli zaman aralıkları şeklinde anlık olarak konum bilgisini yayınladığı ve pilot hastaneye yaklaştığı esnada konum yayını otomatik olarak sonlandırdığı varsayılmaktadır.

Trafik bilgisi gözetilmemiştir. Çünkü ambulans sürücüsü tarafından yol seçimi yapılacağı varsayılmıştır. Trafik olsa da olmasa da mecburen yüksek kapasiteli kentsel yolu kullanması gerektiği durumlarda bu sistemden fayda beklenmektedir.

Yüksek kapasiteli kentsel yolları kullanan ambulansların hastanelere daha hızlı ulaşması için yol seçenekleri olarak ambulansın seyahatini canlandırabilmek için yüksek kapasiteli kentsel yollardan “E5 ve “TEM” belirlenmiştir.

Ambulans sürücüsü, bulunduğu konuma göre belirlenen yollardan tercih edeceği yolu seçmelidir. Seçilen yola göre 2 örnek rotadan biri, anında canlandırılmaktadır.

Bilgi panelleri, sunulan yüksek kapasiteli kentsel yol seçenekleri ve örnek rota planlarına uygun olarak yüksek kapasiteli kentsel yolların belirli noktalarına konumlandırıldığı varsayılmaktadır. Bu bilgi panelleri L tipi direğe monte biçimde yüksek kapasiteli kentsel yolların sol şeridinde konumlandırıldığı varsayılmaktadır. Bilgi panelleri ambulans logosunu ekrana yansıtabilmekte ve gösterim süresine göre ekranda kalması sağlamaktadır. Ambulansın seyahat boyunca etkileşimde olması planlanan bilgi panelleri, rota seçimi esnasında iletişim yöneticisi tarafından önceden kayıt altına alınmaktadır.

4.1. Uygulamalar ve İşlevler (Applications and Functions)

Sistem, farklı amaçlara hizmet eden dört uygulamadan oluşmaktadır. Sistemin çalışması için tüm uygulamaların aynı anda çalışır vaziyette olması gerekmektedir.

Broker uygulaması, sistemin kalbi niteliğindedir. Mesajların yayıncıdan alınması ve abonelere dağıtılması görevlerinden sorumludur. Diğer üç uygulama arasında iletişim kurmakta, organize etmekte ve trafiği yönetmektedir.

İletişim yöneticisi, ambulans ve bilgi paneli uygulamasını temsil eden bu üç uygulama için hizmet ettiği özelliğe göre anlaşılabilir kullanıcı adı ve şifreler oluşturulmuştur. Bu kullanıcı adı ve şifreler, MQTT Broker uygulamasına bağlanırken kimlik doğrulaması aşamasında doğrulama yapmak amacıyla kullanılmaktadır.

Sistemin işleyişinin güvenli olması için her uygulamaya ait kullanıcı adı ve şifre bilgisi Tablo 1'deki gibi tutulmaktadır.

Tablo 1. Uygulama Kimlik Bilgileri
(Authentication Identity Information)

Kullanıcı Adı	Şifre
ambulance1	ambulance1
infopanel	infopanel
centerapp	centerapp

Tablo 1'e göre, ambulance1 ambulans, infopanel bilgi paneli yöneticisi, centerapp haberleşme yöneticisi uygulamalarının sisteme bu kimlik bilgileriyle güvenli bir şekilde bağlanıp iletişime geçmesine olanak sağlamaktadır.

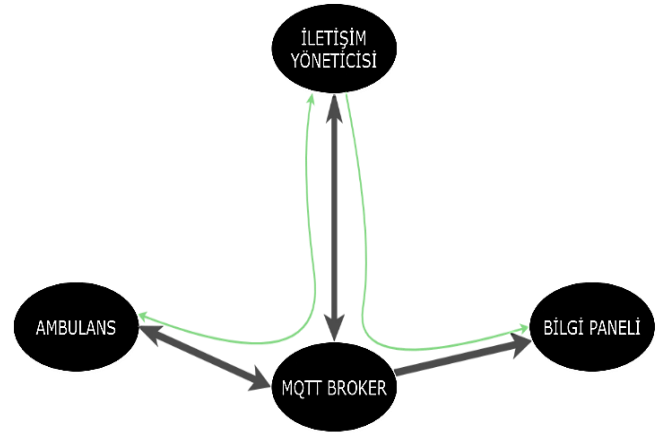
Diğer üç uygulama, MQTT Broker uygulamasıyla aynı cihazda çalışmakta olduğu için diğer üç uygulamanın bağlantı yapabilmesi için cihazı temsil eden ip adresi belirlenmektedir. Kullanılan ip adresi, 127.0.0.1'dir. Kullanılan port, MQTT protokolü için genellikle kullanılan 1884 portudur. Kullanılan port, MQTT protokolü için genellikle kullanılan 1884 portudur.

Ambulans uygulaması, gerçek zamanlı bir seyahati canlandırmak için sürekli olarak belirli zaman aralıklarında, sırayla, otomatik olarak konum bilgisi mesajı yayınlamaktan sorumludur. Yayınladığı konum bilgileri sayesinde bilgi panelleriyle uygun mesafe yakınlığa ulaştığında bilgi panellerinin ambulans logosunu yansıtmasını sağlamaktadır.

İletişim yöneticisi uygulaması, ambulans tarafından seçilen rotaya göre ambulansın etkileşimde olacağı bilgi panellerini belirlemektedir. Ambulans tarafından belirli aralıkla yayınlanan konum bilgisini almakta ve etkileşimde olacağı beklenen bilgi panellerinin konumları ile Haversine formülüyle mesafe ölçümü hesabı yaparak belirli bir yakınlığın sağlanması durumunda ilgili bilgi panelinin ambulans logosunu ekranına yansıtması için bilgi paneline mesaj yayınlamaktadır.

Bilgi paneli uygulaması, bilgi panellerinin ekrana ambulans logosu yansıtması ve ambulans logosunun gösterim süresinin yönetiminden sorumludur. İletişim yöneticisi tarafından belirtilen bilgi paneline ambulans logosunu yansıtması gerektiği mesajı yayımlandığı anda bu mesaj alınır ve ilgili bilgi paneline iletilmektedir. Bu noktadan sonra bilgi paneli ambulans logosunu yansıtmaya başlar ve gösterim süresi boyunca ekranda kalmasını sağlar. Gösterim süresi bittiğinde ambulans logosu yansıtması sonlandırılır.

Aşağıdaki Şekil 4 üzerinden bu dört uygulamasının haberleşmesinin gösterildiği yapı gözlemlenmektedir.



Şekil 4. Uygulamalar arası haberleşme mimarisi
(Inter-application communication architecture)

Şekil 4 üzerindeki haberleşme yapısına göre süreç şu şekilde işler (3, 4 ve 5. adımlar ambulans hastaneye varana kadar tekrar eder.):

1. Ambulans sürücüsü sunulan yüksek kapasiteli yollardan birini seçer ve broker üzerinden iletişim yöneticisinin almasını sağlar.
2. İletişim yöneticisi seçilen rotaya göre etkileşimde olması planlanan bilgi panellerini önceden belirler.
3. Ambulans aracı belirli aralıklarla anlık konum bilgisini paylaşır.

4. İletişim yöneticisi, ambulansın paylaştığı konum ile önceden belirlenen bilgi panellerinin konumları ile mesafe hesaplaması yapar.
5. Ambulans ile bilgi paneli arasında uygun mesafe ölçüldüğünde bilgi paneline ambulans logosunu yansıtması için mesaj gönderilir.
6. Bilgi panelinde ambulans logosu gösterim süresi dolduysa bilgi paneli yansıtmayı sonlandırır.
7. Ambulans hastaneye varır ve süreç tamamlanır.

4.2. İletişim Protokolü, Konular ve Mesaj Paketi (Communication Protocol, Topics and Message Package)

Mesajlar, MQTT protokolü kullanılarak iletilmektedir. Ambulans, Bilgi Paneli ve İletişim Yönetici uygulamaları kendi aralarında bu protokol sayesinde broker üzerinden belirli konular üzerinden mesaj paylaşımı yapmaktadır. Uygulamada kullanılan konular ve kullanım amaçları aşağıda açıklanmıştır.

GetRoutes, ambulans uygulaması tarafından pilot hedef hastaneye gitmesi üzerine seçmesi için iletişim yöneticisinden rota seçeneklerini bu konu üzerinden talep etmektedir.

ListRoutes, iletişim yöneticisi tarafından ambulansın gönderilen rota seçeneklerinin talebiyle ilgili mesaja istinaden bu konu üzerinden ambulans uygulamasına sistemde kayıtlı olan rota seçenekleri iletilmektedir.

SetRoute, ambulansın rota seçeneklerini talep etmesi, iletişim yöneticisinin rota seçeneklerini iletmesi aşamalarından sonra ambulansın yaptığı rota seçimini bu konu üzerinden iletişim yöneticisine göndermesi sağlanmaktadır.

SendAmbulanceLocation, ambulans tarafından rota seçimi yapıldıktan sonra ambulans, belirli zaman aralıklarında anlık konum bilgisini bu konu üzerinden iletişim yöneticisine iletmektedir.

InformDrivers, iletişim yöneticisi tarafından bilgi panelleri ile bu konu üzerinden mesaj paylaşılır. Bu mesajda sadece belirlenen bilgi paneline ambulans logosunu yansıtma mesajı yayınlanmaktadır. Bilgi paneli de bu mesajı alıp ambulans logosunu gösterim süresi boyunca ekrana yansıtılmaktadır.

Uygulamalar arası mesajlar "Payload" kavramı adı altında yayınlanmaktadır. Bu mesajlar uygulamalarda nesnel oluşturularak tutulmaktadır. Bu nesnel uygulamalar arasında transfer edileceği zaman, JSON veri formatına dönüştürülmektedir.

4.3. Verilerin Hazırlanması (Preparation of Data)

Uygulamaların birbiriyle sağlıklı ve güvenli bir şekilde haberleşip amacına uygun hareket etmesi için çeşitli veriler oluşturulmuştur. Bu veriler hazırlarken dikkat edilmesi gereken hususlar ve yöntemler aşağıda alt başlıklar halinde veri setleriyle birlikte açıklanmıştır.

4.3.1. Ambulans Anlık Konum Noktaları Belirleme (Determining Ambulance Current Location Points)

Ambulansın gerçek bir seyahatini canlandırmak için iki adet örnek rota hazırlanmıştır. Bu örnekler, yol seçeneklerine göre canlandırılabilmesi için Google Haritalar servisinden yararlanılarak hazırlanmıştır. Seyahatin başlaması için harita üzerinde bir başlangıç noktası seçilmiştir. Sonrasında bu başlangıç noktasından pilot hedef hastaneye kadar birbiriyle eşit aralıklarda olmayan enlem ve boylam cinsinden noktalar belirleyerek sırasıyla kaydedilmiştir. Bu noktalar gerçek konum gönderiminde olabilecek gecikmeleri de temsil etmek amacıyla eşit zamanlarda gönderilemediği varsayılmıştır. Aynı zamanda her zaman yüksek konum doğruluğunun olmayacağı için bazı noktalar, seyrettiği düşünülen yolun üzerinde veya yola yakın başka bir konumda da olabilmektedir.

Ambulansın gerçek bir seyahatini canlandırması planlanan Rota 1 ve Rota 2 olarak hazırlanan seçeneklere ait bilgilere aşağıda yer verilmiştir.

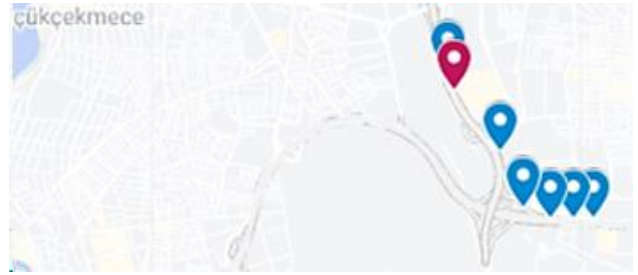
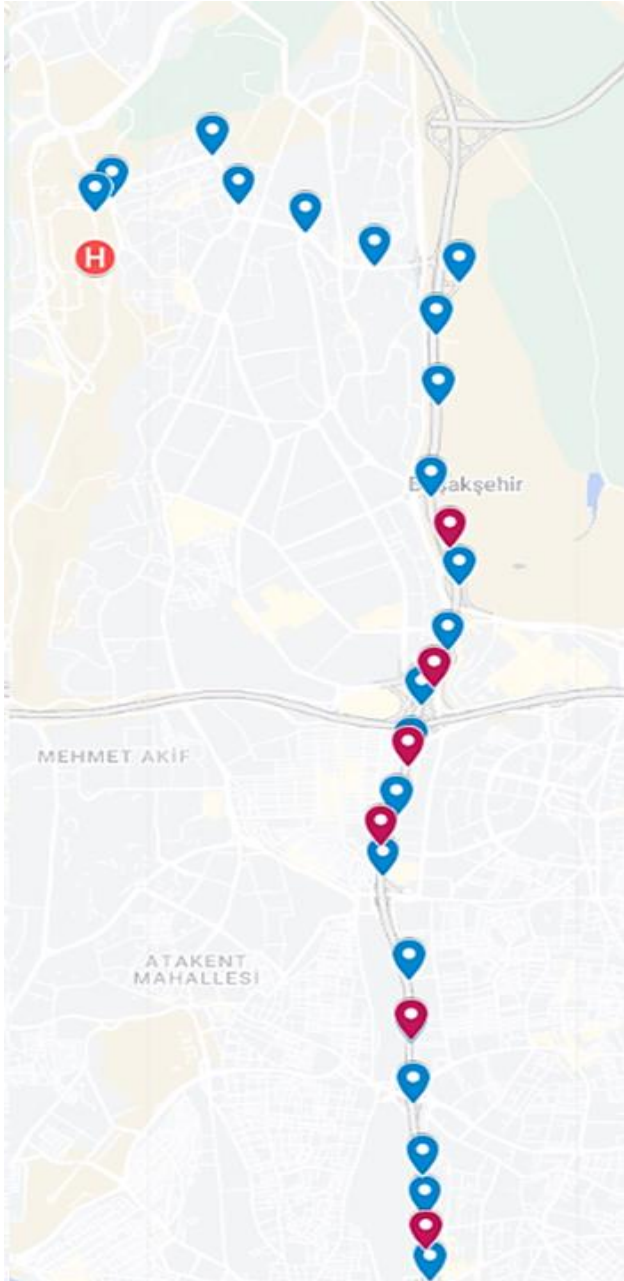
Rota 1 (E5) ve Rota 2 (TEM) olmak üzere ambulansın anlık olarak belirli aralıklarla gönderdiği varsayılan koordinatlar hazırlanmıştır. Rota 1 28 ve Rota 2 25 adet koordinat bilgisinden oluşmaktadır ve seçilen rotaya göre ambulans konum gönderiyormuş gibi canlandırılmaktadır. Aşağıdaki Tablo 2'de Rota 1'e ait koordinat bilgilerine yer verilmiştir.

Tablo 2. Rota 1 Seyahat Koordinatları (Travel coordinates for route 1)

Gönderim Sırası	Enlem	Boylam
1 > Start	40.99298740751887	28.82938386262961
2	40.99295359271643	28.82725833727076
3	40.99301027883759	28.824897993507825
4	40.99393344594995	28.82219432701573
5	40.99816851192638	28.81970523675985
6	41.00437081343914	28.814448107100723
7	41.010119161777624	28.81287096812261
8	41.01600462696554	28.812334526359855
9	41.019857824167474	28.812216509158556
10	41.02645468674724	28.811154354422683
11	41.031869286265064	28.81092904884223
12	41.037736623052545	28.810811031684253
13	41.0472931926247	28.80802153423836
14	41.052859800897785	28.809330452268576
15	41.05831267846455	28.810907591167787
16	41.06290763084206	28.812012661283465
17	41.067898615019836	28.81468414140042
18	41.07396494142018	28.815853584486824
19	41.0820605930251	28.81304262944695
20	41.090454431398584	28.813772190299776
21	41.09694719506897	28.813514698222747
22	41.10188708718531	28.81596087277317

23	41.10323720582183	28.807163227455714
24	41.10643048970727	28.80015729752737
25	41.10883141356347	28.793237198286743
26	41.113462394771865	28.790445537230006
27	41.109622762179185	28.780296057918527
28 > End	41.1081676848614	28.77848288475518

Tablo 2’de ifade edilen koordinat bilgilerinin daha iyi anlaşılabilmesi için Şekil 5 ile görsel olarak harita üzerinde gösterimi sağlanmaktadır. Mavi ile gösterilen noktalar ambulansa ait koordinat bilgilerini ifade ederken, bordo renk ile gösterilen noktalar bilgi panellerinin konumlarını ifade etmektedir.



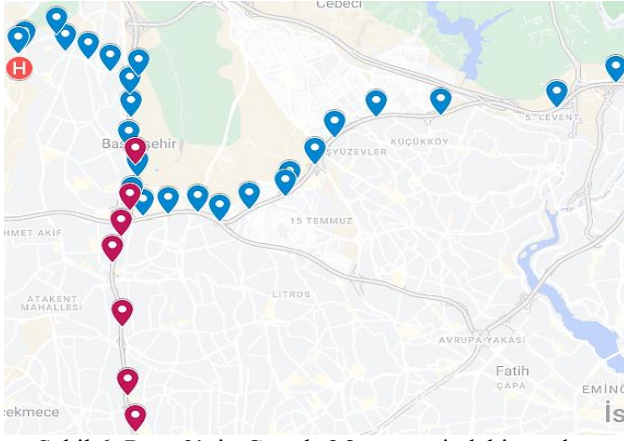
Şekil 5. Rota 1’in Google Maps üzerindeki seyahat noktaları
(Travel points of Route 1 on Google Maps)

Tablo 3’de Rota 2’e ait koordinat bilgilerine yer verilmiştir.

Tablo 3. Rota 2 Seyahat Koordinatları
(Travel coordinates for route 2)

Gönderim Sırası	Enlem	Boylam
1 > Start	41.10012709238443	28.965846795551695
2	41.092959408125715	28.94730583780645
3	41.09100119298792	28.9109297824266
4	41.09056602614201	28.890720862771133
5	41.08490859485699	28.877440715568962
6	41.07729205277947	28.87137803967232
7	41.070762885621534	28.863583170662356
8	41.068586352465985	28.862139676401245
9	41.064886080746724	28.85088042116463
10	41.06146862160465	28.841543809528225
11	41.064162224929696	28.83460301837993
12	41.063623513090604	28.82551874761231
13	41.06300783701372	28.817455181425323
14	41.066317028156156	28.8141889267673
15	41.07396494142018	28.815853584486824
16	41.0820605930251	28.81304262944695
17	41.090454431398584	28.813772190299776
18	41.09694719506897	28.813514698222747
19	41.10188708718531	28.81596087277317
20	41.10323720582183	28.807163227455714
21	41.10643048970727	28.80015729752737
22	41.10883141356347	28.793237198286743
23	41.113462394771865	28.790445537230006
24	41.109622762179185	28.780296057918527
25 > End	41.1081676848614	28.77848288475518

Tablo 3’de ifade edilen koordinat bilgilerinin daha iyi anlaşılabilmesi için Şekil 6 ile görsel olarak harita üzerinde gösterimi sağlanmaktadır. Mavi ile gösterilen noktalar ambulansa ait koordinat bilgilerini ifade ederken, bordo renk ile gösterilen noktalar bilgi panellerinin konumlarını ifade etmektedir.



Şekil 6. Rota 2'nin Google Maps üzerindeki seyahat noktaları

(Travel points of Route 2 on Google Maps)

4.3.2. Bilgi Paneli Sabit Noktaları Belirleme (Determining Fixed Points of Information Panel)

Bilgi panelleri ile ambulansın anlık olarak gönderdiği konum bilgisi arasında yakınlık mesafe ölçümü yapılarak uygun mesafe sağlandığında bilgi paneli ekranına ambulans logosu yansıtılarak diğer sürücülerin bilgilendirilmesi amaçlanmıştır. Google Haritalar servisinden yararlanılarak canlandırma dâhilinde pilot hedef hastaneye giden belirlenmiş yüksek kapasiteli kentsel yollarda, harita üzerinden gözlemlenerek enlem ve boylam cinsinden noktalar belirleyerek birbiriyle mesafeli olarak diğer yolların ana artere katılımından sonra konumlandırılmaktadır.

Konumlandırma esnasında dikkat edilmesi gereken önemli noktalar şunlardır:

- Belirlenen yüksek kapasiteli kentsel yola katılım sağlanan yan yol ya da diğer yollardan sonra konumlandırılması, yola yeni katılan diğer sürücülerin de bilgilendirilmesi açısından önemlidir. Bu nedenle bilgi panelleri konumları belirlenirken yol katılımları dikkate alınmalıdır.
- Birbirleriyle ne çok yakın ne de çok uzak konumlandırılmalıdır.
- Her bir bilgi panelinin ambulans logosunu yansıtması için ambulans tarafından gönderilen anlık konum bilgisi ile kendi konumu arasında yapılan yakınlık mesafe ölçümünde uygun ölçüm mesafenin doğru belirlenmesi gereklidir. Eğer bu mesafe, olması gerekenden daha uzak belirlenmiş ise bilgi panelinde çok erkenden ambulans logosu yansıtacak, eğer yakın ise daha geç yansıtacaktır.
- Ambulans logosunun gösterim süresi ambulans logosunun ekranda kalacağı görüntülenmeye devam edeceği süreyi ifade etmektedir. Bu doğrultuda bu sürenin kendinden bir önceki ve bir sonraki bilgi paneliyle olan mesafesi düşünülerek doğru verilmesi gerekmektedir.

4.3.2.1. Bilgi Panelleri Konum Koordinatları (Info Panel Location Coordinates)

Canlandırmalarda kullanılacak olan yüksek kapasiteli kentsel yollarda belirli aralıklarla konumlandırıldığı

düşünülen bilgi panellerine ait konum bilgilerine Tablo 4 üzerinde enlem ve boylam verilerek yazılmıştır.

Tablo 4. Bilgi Panelleri Konum Koordinatları (Info Panel Location Coordinates)

Adı	Enlem	Boylam
First	41.00291578003355	28.81504449233981
Second	41.0129837869846	28.812550688843977
Third	41.0322686620628	28.810929716571682
Fourth	41.05004337281847	28.807937152376685
Fifth	41.057377569316955	28.810555645858525
Sixth	41.064616934685525	28.81342351987872
Seventh	41.07740139956999	28.814919801976227

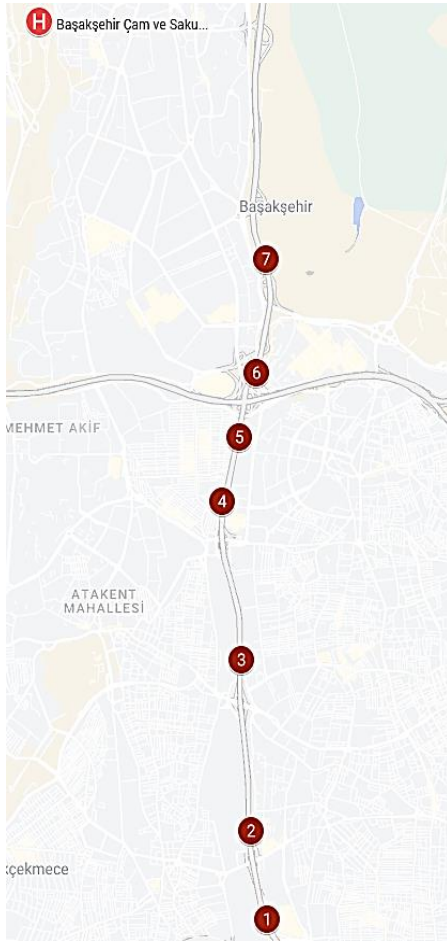
Bilgi paneline ambulans logosunun yansıtılmasıyla ilgili iki farklı yöntem uygulanmıştır. Bunlar yakın ve optimum mesafe yöntemleridir. Yakın mesafe ölçümünde, ambulansın bilgi paneline çok yakın olduğunda ambulans logosunun yansıtması ve bu logonun ekranda kalma süresi azdır. Optimum mesafe ölçümünde ise ambulansın bilgi paneline çok yakın olmadığı ve buna bağlı olarak ambulans logosunun yansıtıldıktan sonra ekranda kalma süresi daha uzundur.

Bu bilgi panellerinin yakın mesafe ölçümlü ve optimum mesafe ölçümlü olarak canlandırılması için ölçüm mesafesi ve gösterim süreleri Tablo 5 üzerinden gösterilmiştir. Bununla birlikte bu bilgi panellerinin geçerli olduğu rotalar, ambulans logosunu yansıtacağı durumla ilgili yakınlık ölçüm mesafesi, ambulans logosu yansıtıldıktan sonra ekranda kalma süresi olan gösterim süresi bilgilerine de yer verilmiştir.

Tablo 5. Bilgi paneli ölçüm bilgileri (Info Panel Measurement Data)

Adı	Geçerli Olduğu Rota	Yakın Mesafe Ölçümlü		Optimum Mesafe Ölçümlü	
		Ölçüm Mesafesi [km]	Yayın Süresi [sn]	Ölçüm Mesafesi [km]	Yayın Süresi [sn]
First	1	0.9	7	1.5	200
Second	1	0.4	7	1.5	130
Third	1	0.5	6	1.5	130
Fourth	1	0.4	6	1.5	125
Fifth	1	0.5	7	2	125
Sixth	1, 2	0.5	7	1.5	180
Seventh	1, 2	0.5	7	2	200

Enlem ve boylam cinsinden tabloda yer verilen bilgi panellerinin harita üzerinde görünümü Şekil 7'deki gibidir.



Şekil 7. Bilgi Panelleri Gösterimi
(Info Panel Display)

4.3.3. Rota Seçenekleri Belirleme (Choosing Route Options)

Google Haritalar servisinden yüksek kapasiteli kentsel yollar gözlemlenmiştir. Bu yollar, pilot hedef hastaneye ulaşım açısından değerlendirilmiş ve yönü tespit edilmiştir. Rota seçenekleri, canlandırmanın başlangıcında, ambulans tarafından seçilmesi istenmektedir.

Ambulansın canlandırmaya başlamadan hemen öncesinde canlandırmanın yapılması istenen Tablo 6'daki rotalardan seçim yapılması beklenir. Seçime göre uygun rota işletilmektedir.

Tablo 6. Rota Seçenekleri
(Route Options)

Adı
E5
TEM

4.4. Verilerin Saklanması (Storing Data)

Veriler, .TXT uzantılı metin belgelerinde satır satır, Tab'lar ile ayrılarak saklanmıştır. Aşağıdaki her bir amaç için ayrı metin belgesi oluşturulmuştur;

- Sisteme bağlanmak isteyen uygulamaların kullanıcı adı ve şifre ile sisteme başarıyla bağlanabilmesi,

- Ambulansın seçmesi için rota öneri seçenekleri
- Broker yapısının yapılandırma bilgisi
- Bilgi panellerinin yüksek kapasiteli kentsel yollar üzerindeki konum bilgileri, kapsadığı yol seçeneği, benzersiz adı, ambulans logosunun yansıtılması için hesaplanması beklenen mesafe değeri, ambulans logosu gösterim süresi
- Ambulansı canlandıran konum bilgileri

4.5 Şifreleme Formülü (Encryption Formula)

Yayınlanan mesajların MQTT protokolü üzerinden TCP/IP ile güvenli, hızlı, açıkça okunamayan ve çözülmesi kolay olmayan biçimde şifreli olarak uygulamalar arasında taşınması ve işlenmesi sağlanmaktadır.

Şifreleme için metin, konu üzerinde yayınlanmadan önce anahtar ile birlikte baytlara ayrılmaktadır. Baytlara ayrılan anahtar SHA256 algoritması ile hashlenerek özetlenmektedir. Sonrasında bu iki değişken ile şifreli bir metni elde etmek için Rijndael şifreleme algoritması kullanılmaktadır. Bu işlem sonrasında şifrelenmiş baytlar base64 formatına dönüştürerek string bir ifade ile üretilmiş olmaktadır.

Şifre çözmek için şifrelenmiş metin, ilk olarak base64 formatından baytlara dönüştürülür. Şifre çözme işleminde kullanılacak olan anahtar da baytlara ayrılmaktadır. Baytlara ayrılmış olan anahtar SHA256 algoritması ile hashlenerek özetlenmektedir. Sonrasında bu iki değişken ile Rijndael şifreleme algoritması kullanılarak şifre çözme işlemi yapılmaktadır. Çözülen şifre bayt array olarak elde edilmektedir. Son işlemde bayt array, string bir ifadeye dönüştürülmektedir.

4.6. Kullanılan Mesafe Ölçüm Formülü (Distance Measurement Formula Used)

Kullanılan mesafe ölçüm formülü, iki nokta arasındaki mesafeyi kuşbakışı yöntemiyle hesaplamak için Haversine formülünden yararlanılmaktadır. Ambulansın seyahat boyunca etkileşimde olması planlanan bilgi panellerinin bulunduğu konum bilgileri ile ambulansın belirli aralıklarla gönderdiği anlık konum bilgisi temel alınarak aralarındaki mesafenin yakınlık ölçümünün yapılarak tespit edilen en yakın bilgi panelinin bulunması amaçlanmıştır.

4.7. Geliştirme Araçları, Teknolojiler ve Kütüphaneler (Development Tools, Technologies and Libraries)

Bu çalışmada 4 uygulamayı geliştirmek için Microsoft Windows 10 işletim sistemi, Pro 22H2 versiyonu üzerinde Microsoft Visual Studio Enterprise IDE geliştirme ortamı, 2022 v17.7.6 versiyonundan yararlanılmıştır. C# programlama dili ile Console uygulamaları olarak kodlanmıştır. .NET 6.0 framework'ünden yararlanılmıştır. Uygulamaların protokol üzerinden haberleşmesi için MQTTnet paketinin 3.0.15 versiyonundan yararlanılmıştır. Verilerin protokol üzerinden gönderilmeden önce json string formatına dönüştürülmesi, protokol üzerinden alındığında json ile class yapısına serialize edilmesi işlemleri için Newtonsoft.Json paketinin 13.0.3

versiyonundan yararlanılmıştır. Protokol üzerinden alınan ve gönderilen veriler için şifreleme yapılmıştır. Bunun için AES simetrik anahtarlı şifreleme algoritması kullanılmıştır.

5. BULGULAR

Yazılımına ait 4 uygulama aynı anda çalıştırılıp 2 örnek rota ile birlikte yakın mesafe ve optimum ölçüm mesafeli bilgi panelleri kullanılmıştır. Her iki rota canlandırmasında da bilgi paneli sayısı ve konumları aynıdır. Ambulans hastaneye en kısa sürede varmasını desteklemesi üzerine bilgi panelleri yardımıyla canlandırmanın çalıştırılması sağlanmıştır.

Rota 1 ve Rota 2 ile yakın ve optimum yakınlık mesafe ölçümlü canlandırmalara ait bulgular sırasıyla Tablo 7 ve Tablo 8, Tablo 9 ve Tablo 10 üzerinde bulunmaktadır. Bu tablolar çeşitli bilgiler içermektedir. Ambulans için gönderdiği konumun sırası ve zaman bilgisi görülmektedir. Bilgi panelleri için değişiklik anı olan ambulans logosunun yansıtılması ve gösterim süresine bağlı olarak gösterimin sonlanması, bilgi panelinin ambulansın gönderdiği konumla arasındaki ölçülen mesafenin değeri ve ambulans logosu gösterim süresine bağlı olarak ambulans logosu yansıtmanın sonlandırılacağı planlanan kapanış zamanı görülmektedir. Ayrıca ambulans gönderdiği konum itibarıyla ilgili bilgi panelinin numarası yazarak "AÇ" ya da "KAPA" komutuyla değiştirildiği an gözlemlenmektedir.

Yakın ve optimum yakınlık mesafe ölçümlü bilgi panelleri bulunmaktadır. Ambulansın gönderdiği konum ile bilgi panelleri arasında yakınlık mesafe ölçümünde uygun ölçüm mesafesine ulaşıldığında bilgi panelleri ambulans logosunu "AÇ" komutuyla ekrana yansıtacaktır.

5.1. Yakın Mesafe Ölçümlü Bilgi Panelleri (Information Panels with Close Range Measurement)

Yakın mesafe ölçümlü bilgi panellerinde Rota 1 ve Rota 2 olarak çalıştırılan canlandırmada sürecin işleyişine ait aşağıdaki sonuçlar elde edilmiştir.

Rota 1 seçilerek canlandırma başlatıldığında canlandırma çıktısı aşağıdaki Tablo 7'deki gibi olmuştur.

Tablo 7. Yakın mesafe ölçümlü rota 1 canlandırma tablosu

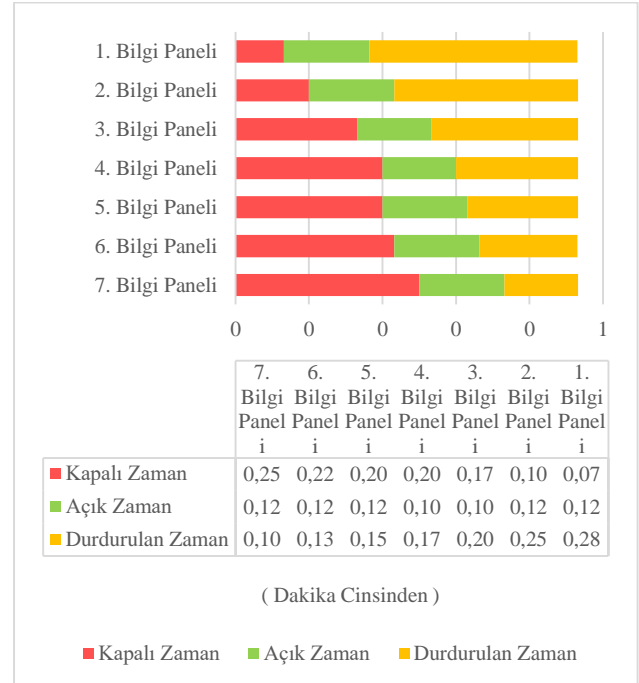
(Route 1 simulation table with close range measurement)

Ambulans Koordinat Gönderim		Bilgi Paneli		
Sıra	Zamanı	Değişiklik Anı	Aktiflik Anında Ölçülen Mesafe [km]	Planlanan Kapanış Zamanı
1	15:15:02			
2	15:15:03			
3	15:15:04			
4	15:15:05			
5	15:15:06	1 - AÇ	0,65698	15:15:13
6	15:15:07			
7	15:15:08	2 - AÇ	0,31966	15:15:15
8	15:15:09			
9	15:15:10			

10	15:15:11			
11	15:15:12	3 - AÇ	0,04441	15:15:17
12	15:15:13	1 - KAPA		
13	15:15:14	4 - AÇ	0,30589	15:15:20
14	15:15:15	2 - KAPA		
15	15:15:16	5 - AÇ	0,10809	15:15:23
16	15:15:17	6 - AÇ	0,22387	15:15:24
17	15:15:18	3 - KAPA		
18	15:15:19	7 - AÇ	0,39005	15:15:26
19	15:15:20	4 - KAPA		
20	15:15:21			
21	15:15:22			
22	15:15:23	5 - KAPA		
23	15:15:24	6 - KAPA		
24	15:15:25			
25	15:15:26	7 - KAPA		
26	15:15:27			
27	15:15:28			
28	15:15:29			

Tablo 7'deki ait sonuçlara göre, ilk bilgi panelinin ambulans logosunu yansıttığı an, 0,65 km ölçüm ile ambulansın beşinci olarak gönderdiği konumdur. Bundan önceki konum gönderimlerinin mesafe ölçümlerinde uygun ölçüm değeri sağlanmadığı için ambulans logosu yansıtılmamıştır. Ambulans logosunu yansıttığı andan itibaren gösterim süresine bağlı olarak planlanan kapanış saati de bu esnada belirlenmektedir. Diğer bilgi panelleri de benzer şekilde ambulans logosunu yansıtmış ve planlanan kapanış saati geldiğinde sonlandırılmıştır.

Aşağıdaki Şekil 8 incelendiğinde bilgi panellerinin birbirinden bağımsız olarak hareket ettiğini göstermek amacıyla daha ayrıntılı grafiksel olarak ele alındı.



Şekil 8. Yakın mesafe ölçümlü rota 1 canlandırma grafiği (Route 1 simulation graphic with close range measurement)

Rota 2 seçilerek canlandırma başlatıldığında canlandırma çıktısı aşağıdaki Tablo 8'deki gibi olmuştur.

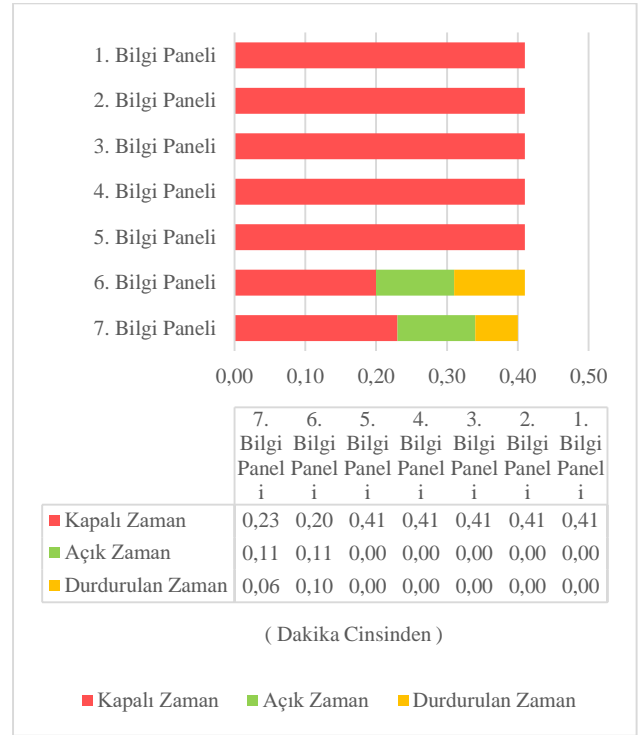
Tablo 8. Yakın mesafe ölçümlü rota 2 canlandırma tablosu

(Route 2 simulation table with close range measurement)

Ambulans Koordinat Gönderim		Bilgi Paneli		
Sıra	Zamanı	Değişiklik Anı	Aktiflik Anında Ölçülen Mesafe [km]	Planlanan Kapanış Zamanı
1	16:25:43			
2	16:25:44			
3	16:25:45			
4	16:25:46			
5	16:25:47			
6	16:25:48			
7	16:25:49			
8	16:25:50			
9	16:25:51			
10	16:25:52			
11	16:25:53			
12	16:25:54			
13	16:25:55	6 - AÇ	0,38244	16:26:02
14	16:25:56			
15	16:25:57	7 - AÇ	0,39005	16:26:04
16	16:25:58			
17	16:25:59			
18	16:26:00			
19	16:26:01			
20	16:26:02	6 - KAPA		
21	16:26:03			
22	16:26:04	7 - KAPA		
23	16:26:05			
24	16:26:06			
25	16:26:07			

Tablo 8'e ait sonuçlarda bakıldığında, ilk bilgi panelinin ambulans logosunu yansıttığı an, altıncı bilgi paneli olup 0.38 km ölçüm ile ambulansın on üçüncü olarak gönderdiği konumdadır. Bundan önceki konum gönderimlerinin mesafe ölçümlerinde uygun ölçüm değeri sağlanmadığı için ambulans logosu yansıtılmamıştır. Ambulans logosunu yansıttığı andan itibaren gösterim süresine bağlı olarak planlanan kapanış saati de bu esnada belirlenmektedir. Altıncı bilgi panelinden sonra gelen yedinci bilgi paneli de benzer şekilde ambulans logosunu yansıtmış ve planlanan kapanış saati geldiğinde sonlandırılmıştır. Altıncıdan önceki bilgi panelleri bu rota kapsamında olmadığı için hiçbir zaman ambulans logosunu yansıtamamışlardır.

Şekil 9 incelendiğinde bilgi panellerinin birbirinden bağımsız olarak hareket ettiğini göstermek amacıyla daha ayrıntılı grafiksel olarak ele alındı.



Şekil 9. Yakın mesafe ölçümlü rota 2 canlandırma grafiği (Route 2 simulation graphic with close range measurement)

Tüm bilgi panelleri birbirinden bağımsız olarak hareket etmektedir. Her bir bilgi paneli kendisine gönderilen ambulans logosunu yansıtmasıyla ilgili mesaj aracılığıyla yansıtma yapmıştır ve gösterim süresine bağlı olarak yansıtmayı sonlandırmıştır. Bir bilgi panelinin ambulans logosunu yansıtıyor ya da yansıtıyor olması diğer bilgi panellerini ilgilendirmemektedir.

5.2. Optimum Mesafe Ölçümlü Bilgi Panelleri (Information Panels with Optimal Measurement)

Optimum mesafe ölçümlü bilgi panellerinde Rota 1 ve Rota 2 olarak çalıştırılan canlandırmada sürecin işleyişine ait sonuçlar elde edilmiştir.

Tablo 9. Optimum mesafe ölçümlü rota 1 canlandırma tablosu

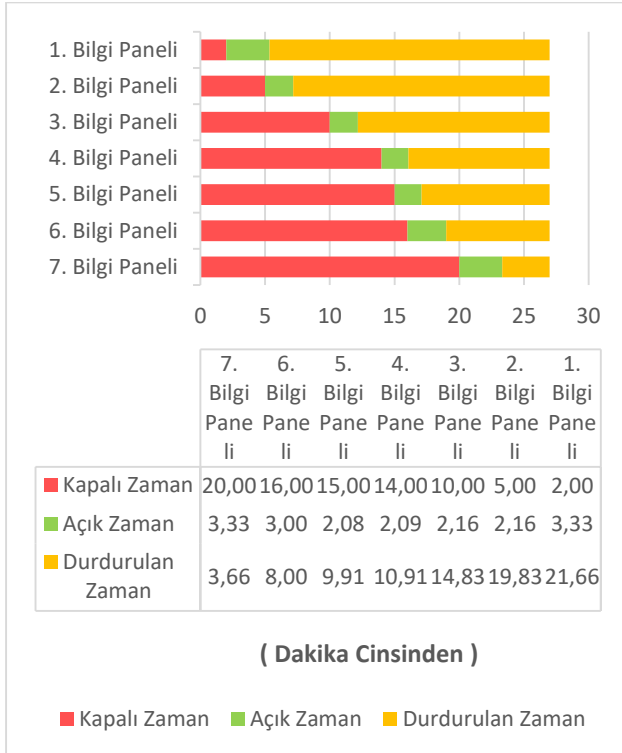
(Route 1 simulation table with optimal measurement)

Ambulans Koordinat Gönderim		Bilgi Paneli		
Sıra	Zamanı	Değişiklik Anı	Aktiflik Anında Ölçülen Mesafe [km]	Planlanan Kapanış Zamanı
1	00:28:06			
2	00:29:06			
3	00:30:06	1 - AÇ	1,37731	00:33:26
4	00:31:06			
5	00:32:06			
6	00:33:06	2 - AÇ	0,97086	00:35:16
7	00:33:26	1 - KAPA		
8	00:34:06			
9	00:35:16	2 - KAPA		
10	00:36:06	3 - AÇ	1,38423	00:38:16
11	00:37:06			
11	00:38:06			

	00:38:16	3 - KAPA		
12	00:39:06	4 - AÇ	1,38951	00:41:11
13	00:40:06	5 - AÇ	1,14129	00:42:11
14	00:41:06	6 - AÇ	1,35163	00:44:06
	00:41:11	4 - KAPA		
15	00:42:06			
	00:42:11	5 - KAPA		
16	00:43:06	7 - AÇ	1,62996	00:46:26
17	00:44:06	6 - KAPA		
18	00:45:06			
19	00:46:06			
	00:46:26	7 - KAPA		
20	00:47:06			
21	00:48:06			
22	00:49:06			
23	00:50:06			
24	00:51:06			
25	00:52:06			
26	00:53:06			
27	00:54:06			
28	00:55:06			

Tablo 9'a ait sonuçlarda bakıldığında, ilk bilgi panelinin ambulans logosunu yansıttığı an, 1.37 km ölçüm ile ambulansın üçüncü olarak gönderdiği konumdur. Bundan önceki konum gönderimlerinin mesafe ölçümlerinde uygun ölçüm değeri sağlanamadığı için ambulans logosu yansıtılmamıştır. Ambulans logosunu yansıttığı andan itibaren gösterim süresine bağlı olarak planlanan kapanış saati de bu esnada belirlenmektedir. Diğer bilgi panelleri de benzer şekilde ambulans logosunu yansıtmış ve planlanan kapanış saati geldiğinde sonlandırılmıştır.

Şekil 10 incelendiğinde bilgi panellerinin birbirinden bağımsız olarak hareket ettiğini göstermek amacıyla daha ayrıntılı grafiksel olarak ele alındı.



Şekil 10. Optimum mesafe ölçümlü rota 1 canlandırma grafiği

(Route 1 simulation graphic with optimal measurement)
Rota 2 seçilerek canlandırma başlatıldığında canlandırma çıktısı aşağıdaki Tablo 10'daki gibi olmuştur.

Tablo 10. Optimum mesafe ölçümlü rota 2 canlandırma tablosu

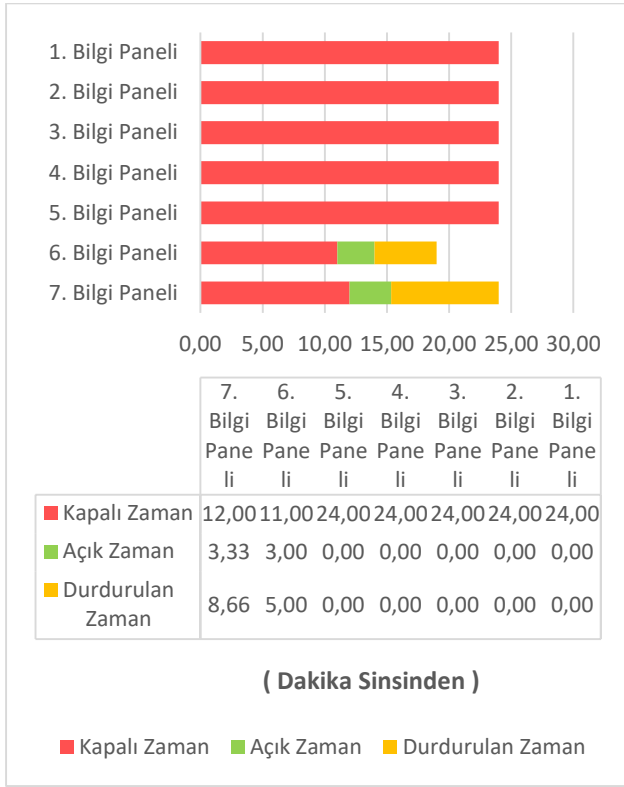
(Route 2 simulation table with optimal measurement)

Ambulans Koordinat Gönderim		Bilgi Paneli		
Sıra	Zamanı	Değişiklik Anı	Aktiflik Anında Ölçülen Mesafe [km]	Planlanan Kapanış Zamanı
1	01:08:19			
2	01:09:19			
3	01:10:19			
4	01:11:19			
5	01:12:19			
6	01:13:19			
7	01:14:19			
8	01:15:19			
9	01:16:19			
10	01:17:19			
11	01:18:19			
12	01:19:19	6 - AÇ	1,02004	01:22:19
13	01:20:19	7 - AÇ	1,61454	01:23:39
14	01:21:19			
15	01:22:19	6 - KAPA		
16	01:23:19			
	01:23:39	7 - KAPA		
17	01:24:19			
18	01:25:19			
19	01:26:19			
20	01:27:19			
21	01:28:19			
22	01:29:19			
23	01:30:19			
24	01:31:19			
25	01:32:19			

Tablo 10'a ait sonuçlarda bakıldığında, ilk olarak ambulans logosunun yansıtıldığı altıncı bilgi panelinin ambulans logosunu yansıttığı an, 1.02 km ölçüm ile ambulansın on ikinci olarak gönderdiği konumdur. Bundan önceki konum gönderimlerinin mesafe ölçümlerinde uygun ölçüm değeri sağlanamadığı için ambulans logosu yansıtılmamıştır.

Ambulans logosunu yansıttığı andan itibaren gösterim süresine bağlı olarak planlanan kapanış saati de bu esnada belirlenmektedir. Altıncı bilgi panelinden sonra gelen yedinci bilgi paneli de benzer şekilde ambulans logosunu yansıtmış ve planlanan kapanış saati geldiğinde sonlandırılmıştır. Altıncıdan önceki bilgi panelleri bu rotada kapsamında olmadığı için hiçbir zaman ambulans logosunu yansıtamamışlardır.

Şekil 11 incelendiğinde bilgi panellerinin birbirinden bağımsız olarak hareket ettiğini göstermek amacıyla daha ayrıntılı grafiksel olarak ele alındı.



Şekil 11. Optimum mesafe ölçümlü rota 2 canlandırma grafiği
(Route 2 simulation graphic with optimal measurement)

6. TARTIŞMA (DISCUSSION)

Ambulansın hayati öneme sahip oluşuna bağlı olarak hastaneye en kısa sürede varmasıyla ilgili çeşitli çalışmalar yapılmıştır. Günden güne artan trafik nedeniyle bir noktadan başka bir noktaya gitmek her geçen gün daha da güç olmaktadır ve trafikte harcanan zaman katlanarak artmaktadır. Özellikle ambulans gibi önemli araçlar için bu durum oldukça kritiktir. Tıbbi müdahaleye ihtiyaç duyan birinin bir an önce hastaneye ulaştırılması büyük önem arz etmektedir. Bu kapsamında ambulansın hızlı yol almasıyla ilgili pek çok çalışma yapılmıştır.

Ambulansın, olay yerinden hastaneye kadar ışık ve sirenin açıp olduğu ve olmadığı durumlara göre hastaneye ulaşım süresi incelenmiştir [22].

Ambulans, ışık ve sirenin açık vaziyette olduğu ve kapalı vaziyette olduğu durumlara göre farklı seyirleri incelenmiştir. Açık olduğu durumda, açık olmadığı duruma göre yüzde 76 oranında hastaneye daha hızlı varıldığı sonucuna varmışlardır. Ambulansın ışık ve siren çözümüyle diğer sürücülere uyararak elde ettiği fayda açıkça görülmektedir. Işık ve siren çözümüne benzer olarak tamamlayıcı nitelikteki bu çalışmada da bilgi panelleri çözümü geliştirilmiştir. Işık ve sireni açma ihtiyacı doğduğunda bu çalışmada yenilikçi ve teknolojik bir çözüm olan bilgi panelleri de devreye alınmalıdır. Böylece diğer sürücüler, uzak mesafelerden gelmekte olan ambulansın ışık ve sireni henüz fark edemiyor iken, bilgi panellerinden çok daha önceden bilgi edinerek ambulans için sol şeridi boşaltıp yol vermiş olacaktırlar. Yapılmış

olan çalışma ile bilgi panelleri çözümünün entegrasyonu ile ambulansın hastaneye hızlı ulaşım oranı daha da artacaktır. Hastanın bir an önce hastanede tam donanımla tedavi altına alınmasının kritik olduğu bu vaziyette, hasta için ciddi bir avantaj sağlanmış olacaktır.

Yapılan diğer bir çalışmaya göre Ankara ilinde, ambulansların günün belirli saatlerindeki faaliyetlerini izlemişlerdir [23]. Gece 12'den sabah 6'ya kadar diğer saat dilimlerine göre daha az vaka bildirilmiştir. Saat 6-18 arasında vaka bildirim sayısı en yüksek olup yanıt süresinin en uzun dilim olduğunu tespit etmişlerdir. Ve sonucunda süreleri iyileştirmek için ambulans sayısının artırılması gerektiği vurgulanmıştır. Fakat bu çözümün uygulanması pek kolay olmayabilir. Ambulansa dönüştürülecek aracın temin edilmesi, yönetmeliklere uygun şekilde teknik ve tıbbi ekipman ile donatılması, araç rutin bakımlarının yapılması gibi ek maddi ve zaman maliyetleri ortaya çıkaracaktır ve çalışmada aktarılan sonucun uygulamasını yavaşlatacaktır. Aynı zamanda ambulans sayısındaki artış, ambulanslarının koordinasyonda kargaşa yaratabileceği gibi trafikte daha çok ambulans olması trafikten kurtulmayı çözmeyecektir. Aksine daha çok ambulans trafiğe dâhil olacaktır.

Teze konu olan ambulansın en kısa sürede hastaneye varmasını destekleyen çözüm olarak bilgi panelleri devreye alındığında tek bir ambulans için katlanılması gereken maddi maliyet yerine bu sistemin kurulmasına harcanacak olan maliyet daha az olacaktır. Bir kez kurulduktan sonra uzunca bir süre ekstra bakım gerektirmeyecektir. Yenilikçi ve teknolojik olan bu çözüm, trafikte daha fazla ambulans katkısı yapmayarak mevcuttaki ambulansların daha efektif olmasını sağlayacak ve görüntü, ses ve karbon emisyonu kirliliğini engelleyecektir. Mevcut sistemdeki ambulans koordinasyon düzeni bozulmayıp trafikten sıyrılarak bir an önce hastaneye varması sağlanacaktır. Böylece ambulans, hastayı hastaneye bıraktıktan sonra ivedi bir şekilde yeni vakaya yönlendirilebilecektir.

Yapılmış olan çalışmalarda temel amaç, ambulansın en kısa zamanda tıbbi yardımı sağlaması, ulaşım sürelerinin gözlemlenmesi yönündedir. Tez kapsamında yapılan ambulansın en kısa sürede hastaneye varmasını sağlayan destekleyici çözümde elde edilen bulgular neticesinde, belirtilen çalışmalarla aynı temel amacın örtüştüğü açıkça görülmektedir.

7. SONUÇ (RESULT)

Bu çalışmada, ambulansın en kısa sürede hastaneye varmaya çalışması esnasında karşılaştığı zorluklara ve bu zorlukların üstesinden gelinmesiyle ilgili çözüme yer verilmiştir.

Acil yardım hizmetlerinin önemi gün geçtikçe artmaktadır. Ambulansın olay yerinden en uygun hastaneye en kısa sürede en hızlı ve güvenli bir biçimde ulaşmasının hasta için ne kadar önemli olduğu bilinmektedir. Ambulanslar, olay yerine ya da hastaneye bir an önce varmayı hedeflemektedir. Fakat bu durumda çoğu zaman

gecikmeler yaşanmaktadır. Ulaşım süresi uzadıkça hasta için ihtiyaç olan tam donanımlı tıbbi tedaviye erişim de o anlamda uzayacaktır. Özellikle nüfusun yıllar geçtikçe arttığı kentlerde ulaşım süreleri, normalden kat ve kat daha fazla olabilmektedir.

Ambulans, hızlı yol almak için siren ve anons ile diğer sürücülerini uyararak hızlı yol almak istemektedir. Fakat bu bir zincir olduğu için her sürücü aynı duyarlılığa sahip olmayabilir, yan ayna açıları doğru olmadığı için görmekte zorlanabilir ya da yüksek sesle müzik dinlemek, telefonla ilgilenmek gibi dikkat dağıtıcı diğer unsurların etkisinde olabilmektedir. Duyarlı sürücülerden de acemi sürücüler olabilir ya da soğukkanlı olamayıp yol vermede gecikebilir ve trafiği de tehlikeye atabilmektedir. Tüm bu durumlarda ambulansa yol vermek için kargaşa içinde gereksiz gazlanma ve frenlemelerle araç parçaları daha çabuk eskimekte, yakıt tüketimi ve karbon emisyonu artmaktadır. Ambulansın hastaneye en kısa sürede varmasıyla ilgili yaşanan bu zorluklar neticesinde ulaşım süresi azaltmak için bir çözüm geliştirilmiştir. Bu çözümle, diğer sürücüler, yolda buldukları konumdan, ambulansın bir süre sonra burada olacağına dair bilgi edinerek daha geniş zaman diliminde, ambulansın sol şeritten rahat ve hızlı bir şekilde ilerleyişi için yol vermiş olacaklar ve trafik tehlikeye atılmayacaktır. Gereksiz gazlanma ve frenleme ile araç parçalarının eskimesi gecikecek, yakıt tüketimi ve karbon emisyonu azaltılmış olacaktır.

Çözümde, 2 farklı rota ile yakın mesafe ölçümlü ve optimum mesafe ölçümlü bilgi panelleri ile canlandırmalar gerçekleştirilmiştir. Her iki yöntemde de bilgi panellerinin konumları aynıdır.

Yakın mesafe ölçümlü iki rota için yapılan bilgi panelleri canlandırmasında, çözümün uygun değerlerle uygulanmaması sebebiyle yeterli olmadığı görülmüştür. Ölçüm mesafeleri birbirine çok yakın ve ambulans logosunun gösterim süresi çok kısadır. Bilgi panellerinin ambulans logosunu yansıtmaya başladığı andaki ölçülen mesafeler 0.04 – 0.65 km aralığında değişmektedir. Bir ambulansın 40 metre kala bu noktadan geçeceği bilgisi çok etkili olmayabilir. Çünkü muhtemelen ambulans zaten siren ve anons ile diğer sürücülerini uyarmıştır. En uzun ölçüm mesafesi olan 650 metre üzerinde baktığımızda, sıkışık bir trafikte bu bilgi yeterli olabilir fakat ambulansın 650 metre mesafede olup bilgi panelinin ambulans logosunun gösterim süresi 7 saniye olduğu için birçok sürücü bu bilgiyi göremeden ambulans logosu yansıtması sonlandırılmış olacaktır.

Optimum mesafe ölçümlü iki rota için yapılan bilgi panelleri canlandırmasında, çözümün uygun değerlerle birlikte yeterli olduğu görülmüştür. Bilgi panellerinin ambulans logosunu ekrana yansıtmaya başladığı andaki ölçülen mesafeler 0.97 – 1.62 km aralığında değişmektedir. Ölçülen en yakın mesafe olan 970 metre kala bilgi panelinde ambulans logosu yansıtacaktır. Ambulans logosunun gösterim süresi ortalama olarak 2 dakika 59 saniyedir. Gösterim süresi de uzun olduğu için diğer sürücüler bu bilgiyi rahatça gözlemleyebilir ve ambulans için sol şeridi sakın ve güvenli bir biçimde boşaltabilir.

Arkadan gelmekte olan trafikteki diğer sürücülerde gösterim süresinin uzun olmasının sağladığı avantaj ile onlar da ambulansın birazdan bu noktadan geçeceği bilgisini görerek sol şeridi boşaltacaklardır. 970 metre mesafeden ambulansın sireni ve anonsunu duymak zor olsa bile bilgi panelleri sayesinde diğer sürücüler aksiyon almaya başlamış olacaklardır.

İki farklı rota ve iki farklı değerlerle yapılan canlandırmalar sonucunda, uygun değerler verildiğinde çözümün oldukça faydalı olacağı düşünülmektedir. Böylece ambulans, hastaneye bir an önce ulaşacaktır.

8. ÖNERİLER (SUGGESTIONS)

Bu çalışmayı desteklemek ve yapılacak çalışmalara ışık tutabilmek amacıyla farklı çözümler geliştirilebilir. Bu çalışmadaki çözüm, olay yerinden hastaneye ulaşım esnasında yüksek kapasiteli kentsel yolların kullanılmasını kapsamaktadır. Bu kapsam genişletilerek çeşitli diğer yollar da dâhil edilebilir. Sadece olay yerinden hastaneye kadar olan sürece ek olarak olay yerine ulaşım da kullanılabilir. Ambulans logosu gösterim süresini sabit vermek yerine trafik bilgisi alınıp ambulansın konumuna bağlı olarak ambulansın bilgi panelini geçtiğinde ambulans logosunun yansıtılmasının otomatik olarak sonlandırılması sağlanabilir. Başka acil ulaşım araçlarının da bu sistemi kullanabilmesi için çalışma yapılabilir. Yolda meydana gelen kaza, yol çalışması gibi durumlar gözetilerek farklı aksiyonlar alınması sağlanabilir.

Çalışmada uygulanan çözüm ve geliştirilecek farklı çözümler olsa da sürücülerin duyarlılığı büyük öneme sahiptir. Gerek ödül sistemi gerekse ceza sistemi devreye alınarak ambulansa yol verme bilinci tüm sürücüler için çok büyük bir önem arz etmelidir. Sürücüler, ambulansın kendi yakınlarını taşıdığını varsayarak bu fikre sahip olabilirler.

KAYNAKLAR (REFERENCES)

- [1] A. C. Uçarlı, F. Demir, S. Erol & R. M. Alkan (2021). Farklı GNSS Uydu Sistemlerinin Hassas Nokta Konumlama (PPP) Tekniğinin Performansına Etkisinin İncelenmesi. *Geomatik*, 6 (3), s.: 247-258. DOI: 10.29128/geomatik.779420
- [2] C. Konukseven (2021). **Aktif Küresel Navigasyon Uydu Sistemleri ile RTK Performansının İncelenmesi**, Doktora Tezi, (Doctoral dissertation, Necmettin Erbakan University (Turkey)).
- [3] İ. Ölmez, & R. Geçen (2018). Acil durumlarda ambulans erişebilirliği: Antakya (Hatay) örneği. *International Journal of Social Science*, 73, s.: 361-375.
- [4] A. Q. Shakor (2022). **Gözlem Aralığının GPS, Glonass, Galileo Ve Beidou Statik PPP Üzerindeki Etkisinin Araştırılması**, Yüksek Lisans Tezi (Doctoral dissertation, Necmettin Erbakan University (Turkey)).
- [5] Ü. Ersöz (2003). **Bilgisayar Ağlarında TCP/IP protokol ailesinin kullanılması ve güvenlik denetimleri**, Yüksek Lisans Tezi (Doctoral dissertation, Sakarya Üniversitesi (Turkey)).

- [6] H. Erdoğan, K. Küçük, & S.A. Khan (2020). Endüstriyel IoT Bulut Uygulamaları için Düşük Maliyetli Modbus/MQTT Ağ Geçidi Tasarımı ve Gerçekleştirilmesi. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi* [Electronic Journal], 7(1), s.: 170-183. <https://doi.org/10.35193/bseufbd.708445>
- [7] O. Mısır, & L. Gökrem (2020). “Nesnelerin İnterneti için MQTT ile Hiyerarşik Haberleşme.” *Journal of New Results in Engineering and Natural Sciences*(12), s.: 1-11.
- [8] T. Lv & P. Yan & W. He (2018). **Survey on JSON Data Modelling. Journal of Physics: Conference Series.** 1069. 012101. 10.1088/1742-6596/1069/1/012101.
- [9] A. Şahin, E. Buluş, & M. T. Sakallı (2014). Modern Blok Şifreleme Algoritmalarının Gücünün İncelenmesi, *Idea*, 8,8.
- [10] Internet: Calculating the Distance Between Two Gps Coordinates with Python (Haversine Formula). <https://nathanrooy.github.io/posts/2016-09-07/haversine-with-python/>. Erişim Tarihi: 25.10.2023
- [11] Internet: Merkez Açık. https://en.wikipedia.org/wiki/Haversine_formula Erişim Tarihi: 25.10.2023
- [12] Internet: Haversine Formula to Find Distance Between Two Points on a Sphere. <https://www.geeksforgeeks.org/haversine-formula-to-find-distance-between-two-points-on-a-sphere/>. Erişim Tarihi: 25.10.2023
- [13] I. G. S. M. Diyasa, D. A. Prasetya, M. Idhom, A. P. Sari, & A. M. Kassim (2022, November). **Implementation of Haversine Algorithm and Geolocation for travel recommendations on Smart Applications for Backpackers in Bali.** In 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS) (s.: . 504-508). IEEE.
- [14] E. Çağlar (2005). **Acil hizmet araçlarının GPS tabanlı denetim ve organizasyonu Yüksek Lisans Tezi (Doctoral dissertation, Bursa Uludag University (Turkey)).**
- [15] S. Ateş, M. Z. Coşkun, & A. Ç. Aydınoglu (2011). Coğrafi bilgi sistemleri ile en uygun ambulans yerlerinin belirlenmesi. TMMOB Harita ve Kadastro Mühendisleri Odası, 13.
- [16] M. Zenginol, B. Al, S. Genç, İ. Deveci, P. Yarbil, D. A. Yılmaz, ... & C. Yildirim (2011). “ Gaziantep İli 112 Acil Ambulanslarının 3 Yıllık Çalışma Sonuçları /3 Yearly Study Results of 112 Emergency Ambulances in the City of Gaziantep”. *Eurasian Journal of Emergency Medicine*, 10(1), 27.
- [17] B. K. Yılmaz, B. Y. Karakuş, E. Çevik, H. Dogan, S. A. M. Mehmet, & A. Kutur (2014). *Metropolde 112 Acil Sağlık Hizmeti. Journal of Istanbul Faculty of Medicine*, 77(3), s.: 37-40.
- [18] P. Manikanta, S. S. Hussian, & R. T. Kodi. (2019, March). Iot ambulance with automatic traffic light control. In 2019 international conference on vision towards emerging trends in communication and networking (ViTECoN) (pp. 1-3). IEEE.
- [19] M. K. Das, & G. Ghosh. (2021, September). Self-Driving Ambulance for Emergency Application. In 2021 5th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (pp. 1-5). IEEE.
- [20] Internet: Değişken Trafik İşaretleri. <https://www.isbak.istanbul/akilli-ulasim-sistemleri/trafik-olcme-bilgilendirme/degisken-trafik-isaretleri-dti/>. Erişim Tarihi: 09.11.2023
- [21] Internet: Acil Durumlarda Trafik Karmaşasına Son. <https://ulastirma.info/pilot-uygulama-istanbul-fatihte-basladi/>. Erişim Tarihi: 20.01.2024
- [22] R. C. Hunt, L. H. Brown, E. S. Cabinum, T. W. Whitley, N. H. Prasad, C. F. Owens Jr, & C. E. Mayo Jr (1995). **Is ambulance transport time with lights and siren faster than that without?.** *Annals of emergency medicine*, 25(4), s.: 507-511.
- [23] K. H. Altintas, & N. Bilir (2001). Ambulance times of Ankara emergency aid and rescue services' *ambulance system. European Journal of Emergency Medicine*, 8(1), s.: 43-50.