



# JOURNAL OF INFORMATION TECHNOLOGIES

BİLİŞİM TEKNOLOJİLERİ DERGİSİ

Volume / Cilt 17 Number / Sayı 1 Year / Yıl 2024 Month / Ay January / Ocak





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. Murat YILMAZ**  
Bilişim Enstitüsü Müdür Yardımcısı  
Vice Director of Institute of Informatics

**Yardımcı Editör**  
(Associate Editor)

**Dr. Öner BARUT**  
Bilişim Enstitüsü Müdür Yardımcısı  
Vice Director of Institute of Informatics

**Editörler (Editors)**

❖ <b>Dr. Ahmet KARAARSLAN</b>	Ankara Yıldırım Beyazıt Üniversitesi (Ankara Yıldırım Beyazıt University)
❖ <b>Dr. Hüseyin POLAT</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Resul DAŞ</b>	Fırat Üniversitesi (Fırat University)
❖ <b>Dr. Mahir DURSUN</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Mehmet ŞİMŞEK</b>	Milli Savunma Üniversitesi (National Defence University)
❖ <b>Dr. M. Hanefi CALP</b>	Ankara Hacı Bayram Veli Üniversitesi (Ankara Hacı Bayram Veli University)
❖ <b>Dr. Mehmet Sıraç ÖZERDEM</b>	Dicle Üniversitesi (Dicle University)
❖ <b>Dr. Oktay YILDIZ</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Olgun DEĞİRMENCİ</b>	TOBB ETÜ (TOBB Economics and Technology University)
❖ <b>Dr. Recep BENZER</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Serdar KULA</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Ö. Tolga PUSATLI</b>	Çankaya Üniversitesi (Çankaya University)
❖ <b>Dr. Cihangir TEZCAN</b>	Orta Doğu Teknik Üniversitesi (Middle East Technical University)
❖ <b>Dr. Uraz YAVANOĞLU</b>	Gazi Üniversitesi (Gazi University)
❖ <b>Dr. Muhammed Ali KOŞAN</b>	Kahramanmaraş İstiklal Üniversitesi (Kahramanmaraş Istiklal University)

**Yayın Danışma Kurulu (Editorial Advisory Board)**

<b>Dr. Ahmet COŞAR</b> <i>Turkish Aeronautical Association University, Turkey</i>	<b>Dr. Aslanbek NAZİEV</b> <i>Ryazan State University, Russia</i>	<b>Dr. Bogdan PATRUT</b> <i>Alexandru Ioan Cuza University of Iasi, Romania</i>
<b>Dr. Deepak GUPTA</b> <i>Maharaja Agrasen Institute of Technology, India</i>	<b>Dr. Jafar A. ALZUBİ</b> <i>Al-Balqa Applied University, Jordan</i>	<b>Dr. Jolanta SABAITYTĖ</b> <i>Vilnius Gediminas Technical University, Lithuania</i>
<b>Dr. Ilya LEVİN</b> <i>Tel Aviv University, Israel</i>	<b>Dr. Pınar KARAGÖZ</b> <i>Middle East Technical University, Turkey</i>	<b>Dr. Ufuk ÇAĞLAYAN</b> <i>Yaşar University, Turkey</i>
<b>Dr. Veysi İŞLER</b> <i>Hasan Kalyoncu University, Turkey</i>	<b>Dr. Victor Hugo Costa DE ALBUQUERQUE</b> <i>Universidade de Fortaleza, Brazil</i>	<b>Dr. Vijender Kumar SOLANKİ</b> <i>CMR Institute of Technology, India</i>
<b>Dr. Ebrahim KHOSRAVI</b> <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) yayımlanmaktadır.**  
Journal of Information Technologies is published every 3 months (January, April, July, October).

# Determination of Optimal Location of Electrical Vehicle Charging Stations in Istanbul with Genetic Algorithm and Geographical Systems

*Araştırma Makalesi/Research Article*

 Kazım YILDIZ<sup>1</sup>,  Onur Yenigün<sup>2</sup>,  Tülin BOYAR<sup>3</sup>

<sup>1</sup>Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul, Türkiye  
<sup>2,3</sup>Marmara Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, İstanbul, Türkiye

[onuryenigun@marun.edu.tr](mailto:onuryenigun@marun.edu.tr), [tulinboyar@marun.edu.tr](mailto:tulinboyar@marun.edu.tr), [kazim.yildiz@marmara.edu.tr](mailto:kazim.yildiz@marmara.edu.tr)

(Geliş/Received:18.07.2023; Kabul/Accepted:31.10.2023)

DOI: 10.17671/gazibtd.1329029

**Abstract**— The use of electric vehicles in Turkey is increasing from one day to the next. This study focuses on determining the efficient locations for the positioning of electric vehicle charging stations to be established in Istanbul. Geographic Information System was used in the study; a location map was created for the traffic density map of Istanbul and the existing electric vehicle charging stations. The genetic algorithm was used to determine the points of highest charging inquiry in the city and to calculate the charging demand satisfaction situation of the existing charging stations. The constraints are that the estimated charging demand is not met to a certain extent at the charging demand points and that the locations are at a maximum distance of 500 metres from the charging demand points. With the solution sets generated according to the genetic algorithm, 10 efficient locations were determined from the 223 Istanbul Parking Lot Operations datasets, representing the locations where the charging station is planned to be installed. The identified locations are the areas with high traffic in Istanbul. In this study, electric vehicle charging network operators can calculate the time to tolerate this cost for the installation of charging stations, which is quite costly; it suggests an approach to location planning so that electric vehicle users can satisfy their charging needs as quickly as possible without creating long queues.

**Keywords**— electric vehicle charging station, electric vehicle charging station positioning, determination of electric vehicle charging station locations

## Genetik Algoritma ve Coğrafi Sistemlerle İstanbul'daki Elektrikli Araç Şarj İstasyonlarının Optimal Konumunun Belirlenmesi

**Özet**— Türkiye'de elektrikli araç kullanımı her geçen gün artmaktadır. Bu çalışmada, İstanbul'da kurulacak elektrikli araç şarj istasyonu konumlandırması için en verimli lokasyonların belirlenmesine odaklanılmıştır. Çalışmada Coğrafi Bilgi Sistemi kullanılmış; İstanbul trafik yoğunluk haritası ve mevcut elektrikli araç şarj istasyonları için lokasyon haritası oluşturulmuştur. Genetik algoritma kullanılarak, şehirde şarj talebinin en yüksek olduğu noktalar belirlenerek mevcut şarj istasyonlarının şarj talebini karşılama durumu hesaplanmıştır. Şarj talep noktalarında tahmini şarj ihtiyacının belirli bir oranda karşılanmaması ve lokasyonların şarj talep noktalarına maksimum 500 metre uzaklıkta olması koşulları aranmıştır. Genetik algoritmaya göre oluşturulan çözüm setleri ile şarj istasyonu kurulması planlanan lokasyonları temsil eden 223 İstanbul Otopark İşletmeciliği veri setinden en verimli 10 lokasyon belirlenmiştir. Belirlenen lokasyonlar İstanbul trafiğinin yoğun olduğu bölgelerdir. Bu çalışmada elektrikli araç kullanıcılarının uzun kuyruklar oluşturmadan şarj taleplerini mümkün olan en kısa sürede karşılayabilmeleri için konum planlamasına yönelik bir yaklaşım önerilmektedir.

**Anahtar Kelimeler**— elektrikli araç şarj istasyonu, elektrikli araç şarj istasyonu konumlandırma, elektrikli araç şarj istasyonu yerlerinin belirlenmesi

## 1. INTRODUCTION

With the support of national policies, it is aimed to reduce the use of fossil fuel vehicles by expanding the use of electric vehicles all over the world in order to reduce environmental pollution. In addition, the fact that fossil fuel resources will be depleted in the near future has accelerated the widespread use of electric vehicles (EV). The rapid increase in the use of electric vehicles necessitates a strong station network throughout the country. The installation of charging stations and the expansion of the charging network are the most important steps in this direction. In our country, studies have been accelerated to expand the network of charging stations. The most important point in these studies is the location of charging stations and site planning. If the correct place cannot be determined and the facility planning is not efficient; various problems will arise both for the charging network operator and for electric vehicle owners, such as not meeting the charging demand and leaving the planned stations empty.

It has been observed in the literature that solutions to the same problem have been found using different approaches. The optimisation problem is defined as a complex combinatorial problem that aims to find the minimum number of critical charging station locations selected to cover the distance between any two nodes in the road network. The required input data is the road network configuration with distances and the assumed autonomy of EVs. The input information is used to generate the graph representing the road infrastructure with nodes as possible locations for charging stations. The application of the presented method is demonstrated on an example road configuration, highlighting its scalability, generality and computational cost [1]. The linear programming approach used in this work has been useful for our study.

Multi-criteria decision-making systems were used in the study carried out to determine the appropriate location in the province of Ankara. In the study, the criteria affecting the location selection were determined; these criteria were weighted and combined with the Fuzzy Analytical Hierarchy Method; a map was created to determine the appropriate location. According to the results of the suitability map, the suggested locations were mostly central districts [2-3].

It was studied on the creation of the infrastructure of EV charging stations. The application model for the study is Balıkesir University Cagis Campus. Vehicle entry and exit data were used and a survey was conducted to determine driver usage habits. With the obtained data, the ideal number for positioning the charging station has been determined by the queuing theory. Parking lots within the campus were determined as candidate locations and how the charging stations would be distributed was determined by multi-criteria decision-making methods. For the charging stations to be located in the province of Istanbul, a study was conducted to determine the locations to be invested from the candidate locations. In the study, a model

was developed in the MATLAB program using the necessary data, and the optimal result is achieved by using YALMIP/Cplex. The model developed in the study yielded results independent of the data size [4-5].

Managing and analyzing spatial data, making calculations with data, etc. Geographic Information Systems (GIS) which have features were used. In the study aimed at positioning electric vehicle charging stations, the choice of location takes into account various criteria. Fuzzy Analytical Hierarchy Process (FAHP) was used to solve the location selection problem by evaluating the determined criteria. The GIS system was used in the analysis process of the electric vehicle positioning problem. In practice, three different districts of Istanbul were selected and the criteria values produced by the multi-attribute decision making method were processed with data layers in GIS and presented as a decision support system. [6,7,25].

The suitability of urban road connections for electric vehicle charging station installation was tried to be estimated. A spatial model consisting of parameters and weights was used. According to the results, transportation centers were determined as the factors that affect the EV charging station positioning of the parking areas in Greece. [8].

The total social cost model was presented, using Ireland as an example, and the operating costs of charging stations under different conditions were calculated. The study evaluated charging station installation costs under sub-headings. A genetic algorithm-based charge position optimization model was constructed to compare the correlation coefficient by dividing Ireland into squares with a rectangle covering. In the study, how long it will take to amortize the cost of a charging station, the power consumption per unit has been calculated and an optimization model has been proposed that minimizes the operating cost in line with the constraints. In addition, sensitivity analysis was conducted to determine which factors were effective on the total costs of charging stations; according to the results of the analysis, it was observed that the daily charging probability and the number of stations were effective [9].

Deep learning model Spatial-Temporal Graph-Informer (STGIN) was applied to predict whether electric vehicle charging stations can be used in the long term. Graph Convolutional Neural Network (GCNN) was used to graphically extract the charging data of urban charging stations according to location and time with the periodic charging data. Electric vehicle charging stations did not work nodes; roads are represented by their edges and a graphical view is obtained. The current charging data was tested with the STGIN model. Compared to other algorithms, it performed better on three different horizons. According to the results of the study, the STGIN model can successfully predict whether the electric vehicle charging station will be used for a long time [10].

They determined the 10 points that would be most efficient in positioning by using the city's traffic flow, existing EV charging points, city road lines and some road features for location selection for charging points to be located in the city of Västerås, Sweden. When the methodology of the study was examined, mapping techniques over GIS were used and optimization was carried out with a genetic algorithm [11-12].

In recent years, efforts to increase the use of electric vehicles have accelerated. With the widespread use of electric vehicles, there are problems in many areas such as range, accessories, and the use of charging stations for electric vehicles. Problems encountered or foreseen to be encountered, and solutions offered to problems; there are many researches and studies in the literature, such as optimization to popularize the use, correct positioning of charging stations, electric vehicle range analysis, and consideration of city electricity infrastructure[13-14].

This study focuses on identifying the characteristics that affect the selection of locations for the installation of electric vehicle charging stations and how to select locations using the identified characteristics. In order to position the EV charging stations at the right location, the locations with high charging demand were determined using Istanbul traffic data; the location information of the EV charging stations in Istanbul was used to select the locations without charging stations. To determine the ideal location, solution spaces were created using the genetic algorithm and the optimisation problem was solved using mixed integer linear programming[14,21,23].

The aim of this study are to identify the factors that will affect the efficient installation of electric vehicle charging stations throughout the province of Istanbul, to collect data on the identified factors, to present efficient installation proposals using genetic algorithms, and to make an innovative contribution to the literature by visualising the results using geographic information systems. This study consists of 3 parts: In the first part, the preliminary research on the subject and previous literature studies are included. In the second part, information about the GIS, the dataset and methods are explained. Finally the results and discussion are presented.

## 2. MATERIAL AND METHODS

Figure 1 Shows the overall design of the system. Traffic flow data and locations of existing EV charging stations were processed and visualised using GIS. The efficient locations among the alternative locations were determined using the genetic algorithm with the obtained data sets. The efficient locations among the alternatives were also visualised using GIS.

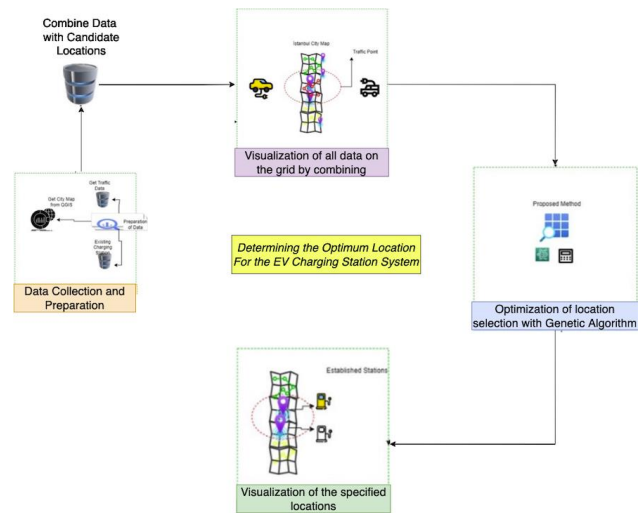


Figure 1. Overall architecture of the system

### 2.1. Datasets

In this study, a grid covering the provincial borders of Istanbul was created and each grid cell was used for the intersection of the data processed on GIS. Istanbul Province Surface area is  $5.170\text{km}^2$ . A cell for the generated grid is  $500\text{m}^2$ . For regions where no traffic flow data is available, the average of neighboring grid cells is used.

In the study, traffic flow of Istanbul, previously located electric vehicle charging stations and location datasets planned to be established were needed. The required datasets were obtained from the Istanbul Metropolitan Municipality website.

Table 1 shows a section from the traffic density dataset for the province of Istanbul. There are 1048519 records in the dataset, which contains the average number of vehicles, maximum-minimum speed and location information for traffic density points [15]. The fields of the records given as examples are presented in the table. Points determined by Longitude and Latitude information are traffic density points. There are minimum speed, maximum speed and average speed data for each traffic density point. Sample data were randomly selected from the regions with high traffic seen in the specified dataset.

Table 1. Traffic density dataset sample for İstanbul

	Longitude	Latitude	Min Speed	Max Speed	Avg Speed
1	28.8446	41.1136	9	78	53
2	29.1522	41.0092	6	64	27
3	28.4711	41.0422	53	96	72
4	29.361	40.817	17	128	73
5	28.8446	40.9708	16	82	55
6	29.295	41.1081	68	152	93
7	28.7128	40.9927	6	91	49
8	28.8995	41.0696	6	46	15
9	28.8226	40.9982	6	114	59
10	28.9764	41.0971E+14	6	77	277

For the locations where charging stations are planned to be installed, the dataset containing the parking lot information of the İstanbul Municipality was used. It is preferred because it is open to access, it represents candidate locations.

Table 2 A section from the dataset used for the parking lots of the İstanbul Municipality is shown. The dataset contains information about 705 parking lots[16]. There are 9 data pertaining to each car park, these data are the district where the car park is located, the name of the car park, working hours, location information, vehicle capacity, car park type. "LONGITUDE" and "LATITUDE" fields are used for the location, and the "PARK\_NAME" field is used to detect the station.

Table 2. A sample from the dataset for the parking lots of the İstanbul

	Park_Name	County_Name	Longitude	Latitude
1	Vali Konağı Caddesi 1	Şişli	28.988	41.0488
2	Şakayık Sokak 1	Şişli	28.9956	41.0495
3	Hüsrev Gerde Sokak 1	Şişli	28.995	41.0484
4	Sezai Selek Sokak 1	Şişli	28.996	41.0534
5	Maçka Caddesi 1	Şişli	28.993	41.0466
6	Maçka Caddesi 6	Şişli	28.9945	41.0478
7	Taşkışla Caddesi 3	Şişli	28.9913	41.0479
8	Vefa Bey Sokak 1	Beşiktaş	29.0066	41.0655
9	Müselles Sokak 2	Şişli	29.0116	41.0688
10	Vali Konağı Caddesi 1	Şişli	28.988	41.0488

The list of electric vehicles charging stations currently in use for the province of İstanbul has been pulled from the İstanbul Municipality website. The current charging station status near the charging stations planned to be located was obtained from this dataset and used to calculate the rate of fulfillment of the charging demand in the optimization modeling.

Table 3 The dataset containing the information about the existing charging stations within the provincial borders of İstanbul is sampled[17]. There are 223 records in the dataset. Each record contains the charging station number, station name, the coordinates of the station, the power of the charging units in the station, and the boundary information for the area of the station. "XCoord" and "YCoord" fields are used to position the charging stations on the grid.

Table 3. The dataset containing the information about the existing charging stations in İstanbul

	StationNo	StationName	Xcoord	Ycoord
1	10000	Eşarj - Akasya AVM, Acıbadem	29.054344	41.000813
2	10001	Eşarj - Akbatı AVM, Esenyurt	28.667032	41.055676
3	10002	Eşarj - ASF Otomotiv, Kartal	29.17877	40.917294
4	10004	SHARZ.NET - Veko Giz Plaza	29.0214538	41.1098236
5	10006	SHARZ.NET - Kemer Petrol	28.9046809	41.1692703
6	10007	SHARZ.NET - Koç Müzesi Deniz Otopark	28.9492851	41.0427988
7	10008	ZES-Migros Ataşehir	29.126572	40.991752
8	10009	ZES-Lukoil Esenyurt	28.676835	41.019524
9	10144	ZES-Pendik	29.254253	40.897597
10	10147	ZES-Novotel	28.979982	41.024665

## 2.2. Generating GIS Data

Driving roads within the provincial borders of İstanbul were created with the Open Street Map plugin over the GIS application. The study base layer was created with the İstanbul city coordinate boundaries. The parameters required for the optimization model were calculated by positioning all the data on the road layer created for the province of İstanbul. Figure 2 shows the ".shp" file created by the roads of the province of İstanbul, which was created through the GIS system. ERS:4326 coordinate reference system is used. Figure 2 shows the İstanbul roadmap was created from the GIS system.



Figure 2. İstanbul Roadmap was created from the GIS system, all data was combined by positioning the information obtained in other datasets on the file created in shapefile type.

Figure 3 shows the borders used for the İstanbul Road Map. In the coordinate system, the points consisting of the borders of the province of İstanbul were determined, and the grid system used in the study was created according to these borders.

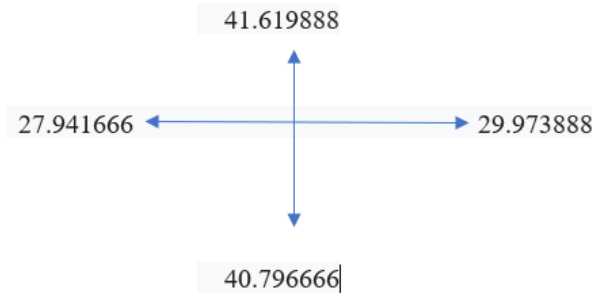


Figure 3. Istanbul provincial boundaries [18]

In Figure 4, the image of the grid to be used for calculations in the optimization model is presented. Each grid cell is  $500m^2$ . In the optimization model, the cell density (charging demand), the current number of charging stations and the planned charging station location are examined separately for grid cells.

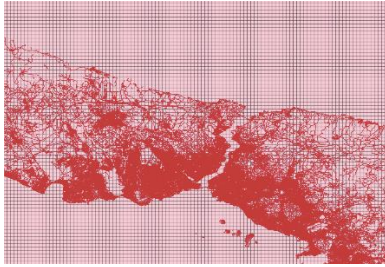


Figure 4 Istanbul province GIS grid layer

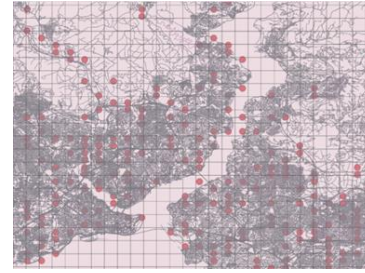
### 2.3. Dataset Preprocessing and Preparation

The datasets used in the study are in “.csv” and “.shp” file types. For the model developed in the GIS system and Python code environment, both file types were used in the study according to the varying needs.

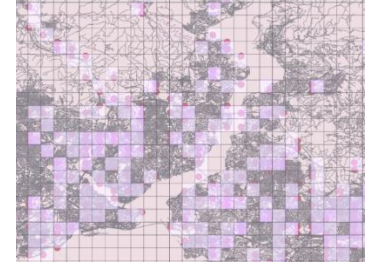
The dataset for the province of Istanbul includes the location of traffic points, the average number of vehicles, and the minimum-maximum speed information. The grid layer divides the Istanbul roadmap into cells. When the Istanbul Traffic Density dataset is added to the grid layer over the GIS system with its current locations, the image in Figure 4-a is obtained. A separate layer was created by selecting traffic points from each grid cell via the GIS system.

In this study, traffic points are considered as places where the charging demand will be intense. Grid cells with high charging demand are selected and used as input in the optimization function.

In Figure 5(a), the traffic density points for the province of Istanbul are shown. Figure 5(b) shows the intersection of the traffic density points with the GIS grid layer created for the province of Istanbul.



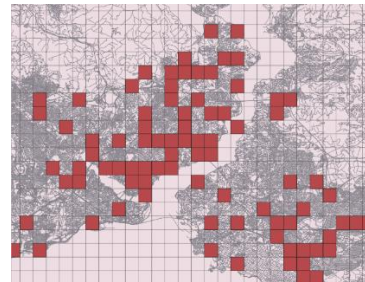
(a)



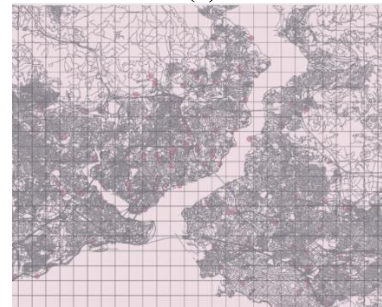
(b)

Figure 5 (a) shows the location of the traffic density points on the map, (b) The intersection of the traffic density points with the grid cells

In the dataset of electric vehicles charging stations already established in Istanbul; there are “Longitude” and “Latitude” information for charging stations, using these areas, charging stations are located on the Istanbul map. How many electric vehicle charging stations are located in each grid cell was calculated via GIS and added as a new field. If the existing electric vehicle charging stations are used in the selection of candidate locations; the charge formed in the grid cell determines the state of meeting the demand. Figure 6(a) shows the location of existing EV charging stations in grid cells, Figure 6(b) shows the locations of charging stations as points.



(a)



(b)

Figure 6 (a) Grid positioning of EV charging stations in Istanbul (b) EV charging station locations in Istanbul



Their locations on the grid were determined by using the location areas in the parking lot dataset of the Istanbul Metropolitan Municipality. The main purpose of this study is to identify 10 sites that will maximise the benefits of the car parks identified as candidate sites. According to the data obtained from other datasets of parking lots; it was evaluated according to the situation of being near the charging demand points. In the parameter used, car parks up to 500m close to the charging demand point are scored as "1", and those over 500m are scored as "0".

After the parking lots were located on the Istanbul map with their location information, it was determined which grid cell they were in on GIS. Figure 7 shows the intersections of the Istanbul Metropolitan Municipality Car Parks with the grid cells.

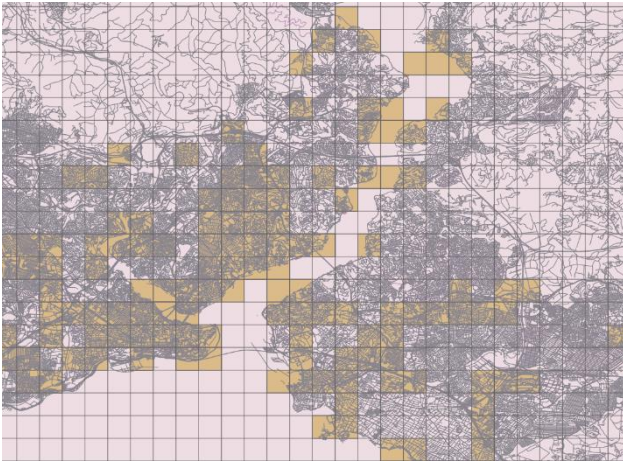


Figure 7. Positioning of İSPARK's locations on the Istanbul city map and intersections with grid cells

#### 2.4. Genetic Algorithm and Optimization Techniques

In the genetic algorithm, which has a very wide application area, all possible solutions are coded as a series. One solution set is selected from the obtained sequence and accepted as the initial population. A fitness value is calculated for each population, and the calculated fitness values show the quality of each population in solution. A group of sequences is randomly selected according to a determined probability value and passed through the replication process. The fitness values of the new individuals emerging by multiplication are also calculated; undergoes crossover and mutation processes. This process flow continues until the predetermined number of generations. The iteration ends when the specified number of generations is reached. As a result, the appropriate sequence is selected according to the objective function[19,20,22,25].

Mixed integer linear programming is the optimization method that are generally preferred by researchers in recent years. It was used first among the classical methods, is a special type of linear programming. Linear programming, on the other hand, can be defined as a technique that minimizes or maximizes a linear cost function. The

equation (1) for a linear optimization problem can be represented as [14,21,22,23]:

$$F(x) = C^T x \quad Ax < b \quad \text{while the input vector is: } x = \{x_1, x_2, x_3 \dots \dots x_n\} \quad (1)$$

A->It is the coefficient matrix for certain values.

B, c-> coefficient vectors for certain values.

A "Mixed Integer Linear Programming" consists of continuous variables, integers and constraints which is shown in equation 2[14,21,22,23]:

$$Z(X) = \min_{(x,y)} \{cx + fy : (x,y) \in X\} \quad (2)$$

The set X is called the set of viable solutions and is defined by linear constraints m, non-negative constraints on variables x,y and integrity constraints on variables y. In matrix representation[14,21,23]:

$$X = \{(x,y) \in R_+^n \times Z_+^p : Ax + By \geq b\} \quad (3)$$

In equation 3, Z(X) shows the efficient objective value when optimization is made over the appropriate set of X. x and y refer to the n (column) vector of the continuous non-negative variables and the p (column) matrix of the integer non-negative variables, respectively.  $c \in R^n$  and  $f \in R^p$  are vectors of objective coefficients.  $b \in R^m$  is the vector of the coefficients (column) on the right side of the constraints m. A and B are constraint matrices with real size coefficients ( $m \times n$ ) and ( $m \times p$ ), respectively.

In the proposed method; for candidate locations, charging station installations are simulated on grid cells that represent solution populations in the genetic algorithm, by keeping the charging station installation cost, daily charging rate, minute charging fee and energy consumption constant. The solution set variables are; it is the charging demand that increases or decreases according to the current charging station and traffic density. From the mutation results obtained, the candidate stations that tolerate the cost the fastest and provide profitability are selected for installation.

Equation 4 was used to construct the objective function for each of the existing charging stations[14,21,22,23]:

$$PrFunc = p[i] * t[i] * q[i] - c[i] \quad (4)$$

p=5-minute charging fee (TL)

t→120 Charging time for an average vehicle (min)

q→10000 Installation cost for a charging station

c→Total charge for charging station

$v_0 = 0.05$  Charging possibility for the charging station to be installed in the cells,  $p_e = 0.17$  kWh Electricity fee,  $l_j = 5$  Maximum charging point to be set up for a station,  $\alpha = 55$  Average battery capacity (kWh),  $N = 10$  Total number of stations planned to be installed. These constraint values used in parameters.

Creating Parameters:  $dr[i] = di[i] - array[i] * m[i]$  if  $dr < 0$   $dr = 0$  (Demand cannot be negative) [14,21,22,23]:

$di = u * fi$  represents the charge demand for cell I,  $u = 0.10$  Daily usage rate is determined as 10%. It is the average of the charging station planned to be established for the  $fi$  cell.

Profitability Account:

$$(cr\_j[j] + ce\_j[j] + ci\_j[j] + 0.1 * ce\_j[j] + 0.1 * ci\_j[j]) * n[j] \setminus + pe * alpha * q[j] \quad (5)$$

$cr\_j = 30$  # Parking fee

$ce\_j = 1100$  # Charge unit price

Creating the Optimization Problem[14,21,22,23]: The problem for which a solution is sought for the research is to maximize the 10 locations to be selected from among the many locations where charging stations are planned to be established. The optimization problem is used to bring the maximized solution.

Parameters for the Optimization Problem[14,21,22,23];  $n$  is the number of charging stations planned to be installed,  $q$  is the number of requests remaining after the locations determined for a fixed number of stations planned to be installed.  $c$  is the total cost for the fee planned to be installed will increase according to the parameters,  $x$  is the usage status of the candidate locations planned to be established (0 or 1) and  $r$  is the distance of the planned charging station to the existing charging station (0 or 1)

All constraints explained as[14,21,22,23]: Constraint 1 ;Remaining below the demand according to the number of daily recharges determined at the charging stations planned to be established

$$q - n * m \leq 0 \quad (6)$$

$m=2$  (Daily charging is fixed as 2)

Constraint 2; The situation where the distance of the candidate locations to the charging demand points is more than 500 meters

$$q[j] - lpSum(r[i][j] * dr[i] \text{ for } i \text{ in demand\_lc}) \leq 0 \quad (7)$$

Constraint 3; Use cases during the determination of candidate locations (whether the location is selected or not)

$$lpSum(x[j] * r[i][j] \text{ for } j \text{ in chg\_lc}) - 1 \leq 0 \quad (8)$$

Constraint 4; Failure to reach the target number after installations

$$probe += n[j] - x[j] \geq 0 \quad (9)$$

Constraint 5; Not reaching the maximum number at the specified location

$$n[j] - l_j * x[j] \leq 0 \quad (10)$$

Constraint 6; Not reaching the number of locations planned to be selected

$$lpSum(x[j] \text{ for } j \text{ in chg\_lc}) == N \quad (11)$$

In the study, firstly, a function that generates populations is created by using the charge demand positions and ISPARK positions with the genetic algorithm. Then, in order to determine the service area for each population, a parameter that takes a value of 0 if the distance of the charging demand point to the potential ISPARK is greater than 500 m, and 1 if it is not, has been added.

A function that includes all parameters and calculates the fulfillment of the charging demand by the existing electric vehicle charging stations for each cell on the grid and rates the charging demand is used. The state of meeting the charging demand was calculated over the GIS system using the existing electric vehicle charging stations located on the grid.

Finally, it is formulated with the Mixed Integer Linear problem consisting of variables and constraints. Optimization function aimed with; is to maximize the profit to be obtained from the stations by subtracting the capital and operating costs required for the installation of the charging station. The situation used as a constraint in the problem is that the station capacity and the charged car are less than the charging demand, the need is met at the charging demand point, and a maximum of 5 charging units are installed in a location.

### 3. RESULTS AND DISCUSSION

As a result of the study, when the traffic data of the province of Istanbul is examined, among the candidate locations for the charging station to be established according to the rate of meeting the charging demand in the regions where the highest charging demand occurs, the parking lots closest to the charging demand point have been determined first. Being close to the charging demand point is limited to 500m.

Table 4 shows 10 locations obtained by optimization problem solving. All the data collected on the grid were solved with the genetic algorithm created and the efficient 10 locations were determined.

Table 4. Ten locations which are obtained by optimization algorithm

	Park Name	Latitude	Longitud e	District
1	Bosna Bulvarı 4	29.0844743 26843	41.024938 615784	Ümraniye
2	Ataşehir Bulvarı 5	29.1253191 232681	40.992844 048405	Ataşehir
3	Ataşehir Bulvarı Taksi Duracağı	29.1267067 463254	40.992571 9800085	Ataşehir
4	Kanlıca Barış Manço Caddesi 1	29.0662089 71994	41.100121 8927195	Beykoz
5	Eski Büyükdere Caddesi 1	29.0062693 981217	41.088122 2691293	Kağıthane
6	Haliç Sosyal Tesisleri Otoparkı	29.0062693 981217	41.088122 2691293	Fatih
7	Avni Akyol Bulvarı	26.6854883 15175	41.070025 875213	Başakşehir
8	Poyrazlı Sokak 1	28.8886640 444539	41.009281 7416347	Güngören
9	Abide-i Hürriyet Caddesi 2	28.9875766 28333	41.063402 364554	Şişli
10	Haznedar Yer Altı Otoparkı	28.8729094 095495	41.008344 2312323	Güngören

Figure 8 shows the optimization function result, 10 efficient locations were determined among 223 İSPARKs in the dataset of candidate locations. According to the results obtained, İSPARKs located in Ümraniye, Ataşehir, Beykoz, Kağıthane, Fatih, Başakşehir, Şişli and Güngören districts of Istanbul were determined as the efficient location.

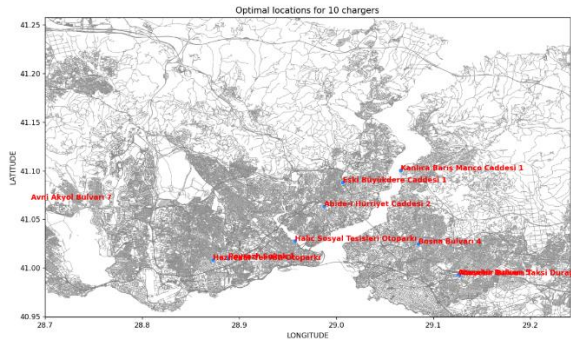


Figure 8. Display of 10 İSPARK points selected according to the results of the optimization function on the map (Python)

#### 4. CONCLUSION

One of the critical issues in increasing the use of electric vehicles is the correct positioning of charging stations. While aiming to increase charging stations; It is very important for both service buyers and service providers to be able to use electric vehicle users without disrupting their travel planning, creating long queues and keeping charging stations empty. For this reason, in this study to determine the efficient location at charging stations, the set of possibilities was examined by genetic algorithm using

existing charging stations, traffic density and candidate location data; with Mixed Integer Linear Programming, which is one of the optimization techniques, the most reasonable solution among the possibilities has been reached. Using the GIS, the province of Istanbul was partitioned with a grid structure and the city was scanned in units of 500 square meters. In the study, the charging demand points of the city were determined on the grid and the rate of meeting this demand was calculated by giving a certain charging rate to the existing charging stations. In future studies, periodic charging time of each station, amount of energy consumption, usage intensity; it is aimed to increase the sensitivity of efficient location determination by using AC-DC comparison data.

#### ACKNOWLEDGMENT

This work has been supported by Üçay Mühendislik A.Ş.



#### REFERENCES

- [1] M.Z. Saeed, Akbari, T. "A Mixed-integer Linear Programming Model for the Plug-in Electric Vehicle Charging Problem in Unbalanced Low Voltage Electrical Distribution Systems Considering Neutral Conductor" *Electric Power Systems Research*, 209, 32-38, 2022.
- [2] K.C. Arayıcı, G. Poyrazoğlu, "Elektrikli Araç Şarj İstasyonu Doğru Konumlandırılması", *Güç Sistemleri Konferansı, Ankara*, 1-4, Kasım, 2018.
- [3] A. Zubiria, A., Menendez, H-J., Grande, P., Meneses, G., Fernandez, "Multi-Criteria Decision-Making Problem for Energy Storage Technology Selection for Different Grid Applications" *Energies*, 1(2), 16, 2022.
- [4] M.S.K, Asnaz., B. Özdemir, "Elektrikli araç şarj istasyonlarının çok kriterli karar verme yöntemleri ile optimal konumlandırılması". *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi*, 4(2), 175-187, 2021.
- [5] S.S., Bilgiliöglü, "Coğrafi bilgi sistemleri ve bulanık analitik hiyerarşi süreci ile elektrikli araç şarj istasyonu yer seçimi", *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 22(1), 165-174, 2022.
- [6] D. Güler, T. Yomralioğlu, "Açık Kaynak Kodlu CBS Yazılımı ve Bulanık Analitik Hiyerarşi Yöntemini İçeren Elektrikli Araç Şarj İstasyonu Yer Seçimi Önerisi", *Harita Dergisi*, 163, 17-28, 2020.
- [7] D Savran, G Tuna, F Macit, R Daş "Coğrafi bilgi sistemleri kullanılarak taşınmaz kültür envanterinin hazırlanması: Kırklareli ili örneği" *Dicle Üniversitesi, Mühendislik Fakültesi, Mühendislik Dergisi*, 8 (1), 147-158
- [8] C., Li, et al., "Robust model of electric vehicle charging station location considering renewable energy and storage equipment". *Energy*, 238, 121713, 2022.
- [9] G., Zhou, Z. Zhu, S. Luo, "Location optimization of electric vehicle charging stations: Based on cost model and genetic algorithm", *Energy*, 247, 123437, 2022.

- [10] R., Luo, et al. "Deep Learning Approach for Long-Term Prediction of Electric Vehicle (EV) Charging Station Availability". **IEEE International Intelligent Transportation Systems Conference (ITSC)**, Indianapolis, 8-15, 2021.
- [11] C., Bian, et al., "Finding the optimal location for public charging stations—a GIS-based MILP approach", *Energy Procedia*, 158, 6582-6588, 2019.
- [12] Q., Zhang, H., Li, L., P.E., Zhu Campana, H., Lu, F., Q., Wallin Sun " Factors influencing the economics of public charging infrastructures for EV—A review", *Renewable and Sustainable Energy Reviews*, 94, 36-38, 2018.
- [13] S., Erdoğan, et al., "Establishing a statewide electric vehicle charging station network in Maryland: A corridor-based station location problem", *Socio-Economic Planning Sciences*, 79, 101127, 2022.
- [14] Internet: O. Sims, Where should we be placing ev charging stations using qgis and pulp, <https://towardsdatascience.com/where-should-we-be-placing-ev-charging-stations-using-qgis-and-pulp-d2bb7b664a55>, 11.05.2022.
- [15] Internet: IBB, Ocak 2020 Trafik Yoğunluk Verisi, <https://data.ibb.gov.tr/dataset/hourly-traffic-density-dataset/resource/db9c7fb3-e7f9-435a-92f4-1b917e357821>, 16.03.2022.
- [16] Internet: IBB, ISPARK Otopark Bilgileri, <https://data.ibb.gov.tr/dataset/ispark-otopark-bilgileri>, 16.05.2022.
- [17] Internet: IBB, Trafik Yoğunluk Haritası, <https://uym.ibb.gov.tr/yharita6/>, 17.03.2022.
- [18] Internet: Wikipedia, İstanbul. <https://en.wikipedia.org/wiki/Istanbul>, 16.06.2022.
- [19] O. Engin, **Akış tipi çizelgeleme problemlerinin genetik algoritma ile çözüm performansının artırılmasında parametre optimizasyonu**, Doktora Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2001.
- [20] R Daş, I Türkoğlu, M Poyraz "Genetik algoritma yöntemiyle internet erişim kayıtlarından bilgi çıkarılması ", *Sakarya Üniversitesi, Fen Bilimleri Enstitüsü Dergisi* 10 (2), 67-72
- [21] J., Islam, et al., "A holistic review on artificial intelligence techniques for well placement optimization problem". *Advances in Engineering Software*, 141, 102767, 2020. Transportation Research Part C: *Emerging Technologies*, 132, 103376, 2021.
- [22] M. Karaköse, H. Yetiş, T. Müezzinoğlu "Optimizasyon Tabanlı Adaptif Görüntü Mozaikleme Algoritması". *Bilişim Teknolojileri Dergisi*, 10(4), 389-400, 2017.
- [23] Y. Pochet, A.W. Laurance, **Production Planning by Mixed Integer Programming**. NY, Springer New York: 77-113, New York, 2006.
- [24] K. Buyukozdemir, K. Yildiz, A. Bas, B.C. Uslu, "A Review of Heuristic Approaches to Vehicle Routing Problems". **IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)**, Australia, 2022.
- [25] Internet Wikipedia, Geographic information system. [https://en.wikipedia.org/wiki/Geographic\\_information\\_system](https://en.wikipedia.org/wiki/Geographic_information_system), 16.06.2023.

# Göğüs Röntgeni Görüntülerinden Akciğer Hastalıklarının Sınıflandırılması için Farklı Derin Öznitelikler ile Beslenen Destek Vektör Makinesi

*Araştırma Makalesi/Research Article*

 Hüseyin ÜZEN\*<sup>1</sup>,  Hüseyin FIRAT<sup>2</sup>

<sup>1</sup>Bilgisayar Mühendisliği, Bingöl Üniversitesi, Bingöl, Türkiye

<sup>2</sup>Bilgisayar Mühendisliği, Dicle Üniversitesi, Diyarbakır, Türkiye

[huzen@bingol.edu.tr](mailto:huzen@bingol.edu.tr), [huseyin.firat@dicle.edu.tr](mailto:huseyin.firat@dicle.edu.tr)

(Geliş/Received:26.09.2023; Kabul/Accepted:30.11.2023)

DOI: 10.17671/gazibtd.1366846

**Özet**— En ölümcül akciğer hastalıklarından üçü olan COVID-19, tüberküloz ve zatürre, rutin olarak göğüs röntgeni (GR) taramaları kullanılarak tespit edilmektedir. Son teknolojik gelişmeler, otomatik teşhis için bilgisayar destekli sistemlerde yeni bir çağ başlatmakta ve önemli faydalar sunmaktadır. Bu çalışma, bu hastalıkları GR'lerden ayırt etmek için tasarlanmış üç aşamalı yeni bir derin öğrenme modeli önermektedir. Modelin ilk aşamasında, derinlemesine ayrılabilir evrişim, geleneksel evrişim ve tam bağlı katmanlar dahil olmak üzere derin özellikleri çıkarmak için bir Evrişimsel Sinir Ağı (ESA) kullanılmaktadır. İkinci aşamada, daha yüksek sınıflandırma başarısı elde etmek için Destek Vektör Makineleri (DVM) sınıflandırıcısı kullanılarak tekrar bir eğitim sürecinden geçirilmektedir. Bu sayede farklı katmanlardan alınan derin özelliklerden daha fazla yararlanılmaktadır. Üçüncü aşamada ise model test edilmektedir. Deneysel çalışmalarda dört sınıftan oluşan GR veri kümesi üzerinde testler gerçekleştirilmiştir. Bu veri kümesi COVID-19, Pnömoni, Normal ve Tüberküloz sınıflarını içermektedir. Kapsamlı deneysel çalışmalar sonucunda önerilen model %99,30 ortalama doğruluk sonucuna ulaşmıştır. Diğer yandan sınıf bazlı sonuçlarda COVID-19 ve Tüberküloz için %100, Normal ve Pnömoni vakaları için ise %98,60 doğruluk oranına ulaşmıştır. Bu sonuçlar COVID-19 ve Tüberküloz sınıflandırması için önerilen modelin çok etkili olduğu görülmektedir. Ayrıca deneysel çalışmaların ikinci bölümünde, önerilen model sonuçları, mevcut modeller ile karşılaştırılmış ve üstün başarılar elde ettiği görülmüştür.

**Anahtar Kelimeler**— COVID-19, pnömoni, tüberküloz, derin öğrenme, destek vektör makinesi

## Support Vector Machine Fed with Different Deep Features for Classification of Lung Diseases from Chest X-Ray Images

**Abstract**— COVID-19, tuberculosis, and pneumonia, three of the most deadly lung diseases, are routinely detected using chest X-ray (CXR) scans. Recent technological advancements are ushering in a new era of computer-assisted systems for automated diagnosis, offering significant benefits. This study proposes a three-stage deep learning model designed to differentiate these diseases from CXRs. In the initial phase of the model, a Convolutional Neural Network (CNN) is used to extract deep features, including depthwise separable convolution, conventional convolution, and fully connected layers. In the second phase, a Support Vector Machine (SVM) classifier is employed for retraining to achieve higher classification accuracy, maximizing the utilization of deep features from different layers. The third stage involves testing the model. Experimental tests were conducted on a CXR dataset comprising four classes: COVID-19, Pneumonia, Normal, and Tuberculosis. Following comprehensive experimental studies, the proposed model achieved an average accuracy of 99.30%. Moreover, in class-specific results, it reached 100% accuracy for COVID-19 and Tuberculosis, and 98.60% accuracy for Normal and Pneumonia cases, indicating the high effectiveness of the proposed model in classifying COVID-19 and Tuberculosis. Furthermore, in the second part of the experimental studies, the outcomes of the proposed model were compared with existing models, demonstrating superior achievements.

**Keywords**— COVID-19, pneumonia, tuberculosis, deep learning, support vector machine

## 1. GİRİŞ (INTRODUCTION)

Her yıl milyonlarca insan COVID-19, Pnömoni ve Tüberküloz (TÜB) gibi akciğer hastalıklarına yakalanmakta ve bu hastalıklar ölümle sonuçlanmaktadır. Bu oranın her yıl artması beklenmektedir [1]. Şiddetli akut solunum sendromu koronavirüs 2 (SARS-CoV-2) virüsünün neden olduğu bir hastalık olan Coronavirus 2019 (COVID-19) ciddi bir salgın haline gelmiştir [2] ve COVID-19 vakalarının katlanarak artması dünya genelinde sağlık sistemi üzerinde oldukça fazla bir baskı oluşturmuştur. Benzer şekilde, Pnömoni ve TÜB de yaşamı tehdit eden hastalıklardır [3][4]. Sonuç olarak, bu tür hastalıkların doğru ve zamanında tespiti, uygun tedavi ve hayat kurtarmak için kritik öneme sahiptir [5].

Pnömoni, TÜB ve COVID-19 dahil olmak üzere birçok bulaşıcı akciğer hastalığı, solunum güçlüğü, öksürük ve ateş gibi benzer semptomları paylaşmaktadır. Çeşitli etkenlerin neden olduğu nefes darlığı, ateş, balgam ve öksürük gibi semptomlara yol açmakta ve hastalık ilerledikçe akciğer opaklaşması görülmektedir. Mevcut tedavilere rağmen, pnömoni özellikle çocuklarda önde gelen bir ölüm nedeni olmaya devam etmektedir. TÜB ayrıca öksürük ve nefes darlığı ile kendini göstermekte ve genellikle pnömoni ve COVID-19 ile örtüşmektedir. İlaça dirençli TÜB giderek artan bir endişe kaynağıdır. Bu nedenle, örtüşen semptomlara sahip bu hastalıkları tespit edip sınıflandırarak doğru tedavi yaklaşımını sağlayacak sağlam bir yapay zeka sistemine ihtiyaç vardır [6].

Akciğer hastalığı teşhisi için standart görüntüleme yöntemleri arasında bilgisayarlı tomografi (BT), göğüs röntgeni (GR) ve manyetik rezonans görüntüleme (MRG) taraması yer almaktadır. MRG ve BT taraması akciğer hastalıklarının değerlendirilmesinde altın standart olmasına rağmen, daha pahalıdır, radyasyona maruz kalmayı gerektirir ve tüm dünyada kolayca bulunamaz [7]. Buna karşılık, GR daha ucuzdur, kolayca temin edilebilir ve kardiyotorasik ve pulmoner bozukluklar için en yaygın tanısal görüntüleme tekniklerinden biridir.

Akciğer hastalığının GR görüntüleri farklılaşma zorlukları sunmakta ve genellikle radyologlar arasında yüksek okuyucu değişkenliğine neden olmaktadır [8]. Pandeminin gelecekteki potansiyel dalgalarıyla birlikte radyologların iş yükü artacak ve radyologların nitel değerlendirmesini geliştirebilecek yeni otomatik görüntü analiz araçlarına acil ihtiyaç duyulacaktır. Bu araçlar, tanısal iş akışını desteklemek için GR'nin bölümlerini sınıflandıracak veya bölümlere ayıracaktır. Karar destek sistemleri, klinik karar verme sürecine yardımcı olmak için tasarlanmıştır ve sağlık hizmetlerinde yükselen bir araştırma trendi olarak kendini kanıtlamıştır [9]. Pandeminin son aylarında, pnömoni veya diğer akciğer hastalıklarının otomatik tespiti, özellikle erken tespiti ve sınıflandırılması, hem klinik hem de yapay zeka araştırmacılarının büyük ilgisini çekmiştir.

Yapay zeka tabanlı tıbbi sistemlerin geliştirilmesi ve bunların tıbbi uygulamaya aktarılması, hastaların

tedavisinde giderek daha önemli bir rol oynamaktadır [10]. Teşhis için kan testi sonuçlarına veya biyobelirteçlere dayanan otomatik yöntemlerin [11] yanı sıra, COVID-19 ve diğer akciğer hastalıklarıyla mücadele etmek için doğru, sağlam ve hızlı tespit teknikleri geliştirmek amacıyla giderek artan sayıda derin öğrenme tabanlı yöntem, özellikle de evrimsel sinir ağı (ESA) tabanlı modeller uygulanmakta ve kullanılmaktadır [12]. GR görüntüleri kullanılarak COVID-19, Pnömoni ve TÜB teşhisi üzerine literatürdeki çalışmalardan bazıları aşağıda verilmiştir.

Mamalakis vd. [6] GR görüntülerine dayalı olarak COVID-19, pnömoni, TÜB veya sağlıklı hastaları teşhis etmek için DenResCov-19 adlı yeni bir derin transfer öğrenme modeli geliştirdiler. Model, mevcut DenseNet-121 ve ResNet-50 ağlarından oluşmaktadır. Deneysel çalışmalar sonucunda dört sınıftan oluşan veri seti ile %75,75 F1-skor değeri elde edilmiştir. Qagos vd. [13] verimli bir ESA mimarisi kullanmıştır. Mimari, 6587 GR görüntüsü kullanılarak eğitilmiştir. 128x128 görüntü boyutu ve 100 epok ile ilerleyen mimari, GR görüntülerini dört sınıfa (Normal, COVID-19, Pnömoni ve TÜB) sınıflandırırken %94,53 doğruluk elde edebilmiştir. Bhandari vd. [14] GR görüntüsünü COVID-19, pnömoni ve TÜB olarak sınıflandırmak için bir açıklama oluşturma çerçevesi ile tamamlanan yeni bir hafif tek ESA modeli önerdiler. Önerilen model, 7132 GR görüntüsünden oluşan kamuya açık veri seti üzerinde test edildi. %94,31 test doğruluk değeri elde edilmiştir. Situla vd. [15], GR görüntülerini sınıflandırmak için VGG16 ile bir dikkat mekanizması kullanmışlardır. Yöntemler üç GR görüntü veri kümesi için değerlendirilmiş ve GR görüntülerini dört sınıfa (COVID-19, Normal, Pnömoni bakteriyel ve Pnömoni viral) sınıflandırırken en yüksek doğruluk %85,43 olarak bildirilmiştir. Bashar vd. [16] GR görüntülerini kullanarak COVID-19 ve pnömoni teşhisi için bir derin öğrenme yaklaşımı önermiştir. Bu bağlamda, Kaggle'dan 21165 GR görüntüsü içeren halka açık bir veri seti elde edilmiştir. Bu çalışmada görüntü iyileştirmeleri ve veri artırma yöntemlerinin yanı sıra VGG16, VGG19, GoogleNet gibi çoklu transfer öğrenme algoritmaları kullanılmıştır. Arttırılmış ve geliştirilmiş veri seti üzerinde VGG16 algoritması kullanılarak en yüksek doğruluk %95,63 olarak elde edilmiştir. Szepesi vd. [17] normal ve pnömoni enfeksiyonlu çocukların GR görüntülerini içeren bir veri seti kullanarak InceptionV3 ve ResNet50 gibi transfer öğrenme modelleri geliştirmiştir. ResNet50 %89,06'lık bir doğruluk elde ederken, InceptionV3 %90,94'lük daha yüksek bir doğruluk elde etmiştir. Hwa vd. [18], GR görüntüleri ve canny kenar algılama görüntüleri kullanarak TÜB'yi tespit etmek için topluluk derin öğrenmesini kullanan bir yaklaşım sunmuştur. İki özellik kümesi çıkarılmış, birincisi GR görüntülerinden, ikinci küme ise canny kenar algılama görüntülerinden çıkarılmıştır. Topluluk modeli tarafından doğruluk açısından elde edilen en iyi sonuçlar %89,77 ve duyarlılık %90,91 olmuştur. Ibrokhimov ve Kang [19], GR görüntülerini kullanarak pnömoniyi tespit etmek için bir derin öğrenme teşhis sistemi geliştirmiştir. VGG19 ve ResNet50 gibi önceden eğitilmiş modeller kullanılarak transfer öğrenimi kullanılmıştır. Veri seti 11956 COVID-19 örneği, 11263 viral veya bakteriyel pnömoni ve 10701 normal örnekten

oluşturmuştur. VGG19 modeli, tüm sınıflarda ortalama %96,6 doğruluk elde ederken, ResNet50 modeli %95,85 doğruluk değeri elde etmiştir.

Bu çalışmanın amacı, GR taramalarına dayalı olarak COVID19, pnömoni, TÜB ve sağlıklı hastalar arasında erken otomatik tespit ve ayırımı fizibilitesini test etmektir. Literatürdeki çalışmalara bakıldığında genellikle önceden eğitilmiş modeller kullanılmıştır. Bu modellerin ortak özellikleri güçlü derin öznitelikler ile softmax sınıflandırıcısını birleştirmektir. Bu çalışmanın ana hedefi, GR görüntü sınıflandırma için güçlü derin öznitelikler elde edilmesi ve bu öznitelikler ile beslenen güçlü bir sınıflandırıcı geliştirmektir. Bu doğrultuda derin özniteliklere dayalı yeni bir model önerilmiştir. Önerilen modelin birinci aşamasında derin öznitelikleri elde etmek için geliştirilen bir ESA ile eğitim yapılmaktadır. Bu ağ mimarisi derinlemesine ayarlanabilir evrişim, geleneksel evrişim ve tam bağlı katmanlardan oluşmaktadır. Önerilen modelin ikinci aşamasında bu derin öznitelikler Destek vektör makineleri (DVM) sınıflandırıcısının girişine verilmektedir. Ayrıca bu aşamada DVM sınıflandırıcı derin öznitelikler kullanılarak eğitilmektedir. Önerilen modelin son aşamasında ise test veri seti ile önerilen sistem test edilmektedir. Kaggle platformundan alınan dört sınıflı GR veri kümesi üzerinde yapılan kapsamlı deneysel çalışmalarda önerilen model ortalama %99,30 doğruluk ve %98,90 F1-skoru ile üstün başarı sağlamıştır. Sınıflandırıcı olarak softmax kullanıldığında elde edilen doğruluk değeri %97,59 iken, önerilen modeldeki gibi DVM kullanıldığında ortalama doğruluk değeri %99,30 olarak elde edilmiştir. Literatürde sıklıkla softmax kullanılmaktadır. Bu çalışmada DVM kullanılarak %1,71'lik bir iyileştirme yapılmıştır.

Bu çalışmanın geri kalanı üç ayrı bölümden oluşmaktadır. Kullanılan veri seti ve önerilen yöntem Bölüm 2'de detaylı bir şekilde ele alınmıştır. Bölüm 3'de gerçekleştirilen deneysel çalışmalar ve bu deneysel çalışmaların sonuçları verilmiştir. Son bölüm olan Bölüm 4 ise çalışmanın genel bir özetinin yer aldığı Sonuçlar bölümüdür.

## 2. MATERYAL VE METOT (MATERIAL AND METHOD)

### 2.1. Göğüs Röntgeni Veri Kümesi (Chest X-ray Dataset)

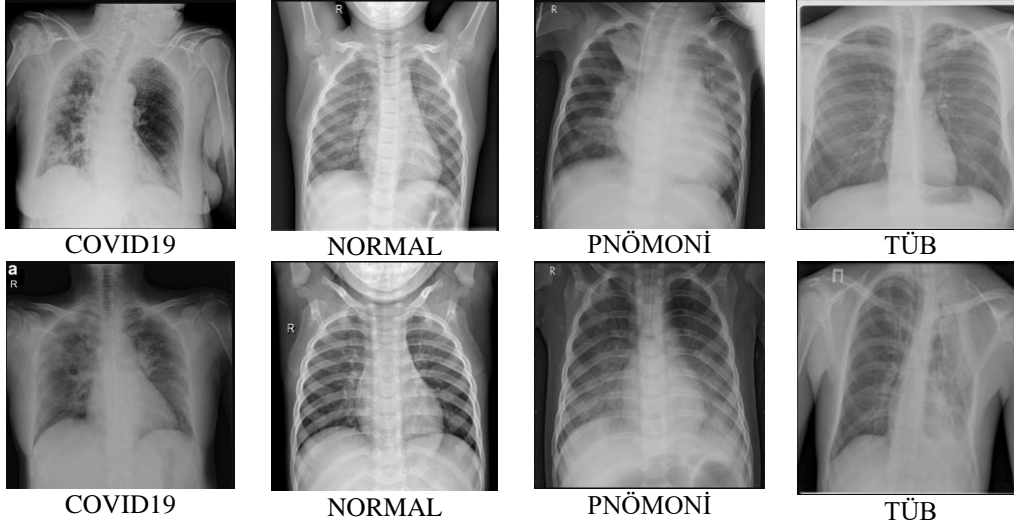
Bu çalışmada, Kaggle platformunda halka açık GR görüntüleri kullanılmıştır [20]. Bu veri kümesi dört sınıf içermektedir. Bu sınıflar: COVID-19, Normal (Sağlıklı), Pnömoni ve Tüberküloz (TÜB). Veri kümesinde, COVID-19, Normal, Pnömoni ve TÜB için sırasıyla 576, 1583, 4273 ve 700 görüntü bulunmaktadır. Veri setinde dört sınıfa ait toplam 7132 GR görüntüsü bulunmaktadır. Çalışmada veri kümesi %80 eğitim, %10 test, %10 doğrulama olarak alınmıştır. Bu durumda, 5706 görüntü eğitim için, 713 görüntü test için ve 713 görüntü doğrulama (validation) için kullanılmıştır. Veri kümesinde eğitim için 461 COVID-19, 1266 Normal, 3419 Pnömoni ve 560 TÜB görüntüsü, test için 58 COVID-19, 158 Normal, 427 Pnömoni ve 70 TÜB görüntüsü bulunmaktadır. Bunun yanı sıra validation için 57 COVID-19, 159 Normal, 427

Pnömoni ve 70 TÜB görüntüsü yer almaktadır. Görüntü boyutları tek tip olmadığından dolayı görüntüler  $224 \times 224 \times 3$  olarak yeniden boyutlandırılmıştır. Görüntüleri  $224 \times 224 \times 3$  gibi standart bir boyuta yeniden boyutlandırmak, görüntü işlemede, özellikle de tek tip giriş boyutlarının gerekli olduğu derin öğrenme gibi görevler için yaygın bir uygulamadır. Literatürde derin öğrenme modelleri için görüntülerin  $224 \times 224 \times 3$  olarak yeniden boyutlandırılması bağlamında çift doğrusal (Bilinear), Biküçük ve En yakın komşu enterpolasyon yöntemleri bulunmaktadır. Bunlarda çift doğrusal ve biküçük enterpolasyon yöntemleri sıklıkla tercih edilmektedir. Çift doğrusal yöntemi, orijinal görüntüdeki en yakın dört pikselin ağırlıklı ortalamasını dikkate alarak yeni piksel değerlerini hesaplamaktadır. Çift doğrusal enterpolasyon hızlıdır ancak görüntü kalitesinden bir miktar ödün verebilmektedir. En yakın komşu enterpolasyonuna kıyasla daha düzgün sonuçlar üretme eğilimindedir. Biküçük enterpolasyon, çift doğrusal enterpolasyondan daha karmaşıktır, yeni piksel değerlerini hesaplamak için daha geniş bir piksel komşuluğunu dikkate almaktadır. Daha yavaş bir işlem hızına sahiptir. Bu çalışma kapsamında, girdi GR görüntülerini  $224 \times 224 \times 3$  boyutuna yeniden boyutlandırmak için çift doğrusal (Bilinear) enterpolasyon ön işleme yönteminin kullanımı tercih edilmiştir. Şekil 1, veri kümesindeki bazı örnek GR görüntülerini göstermektedir.

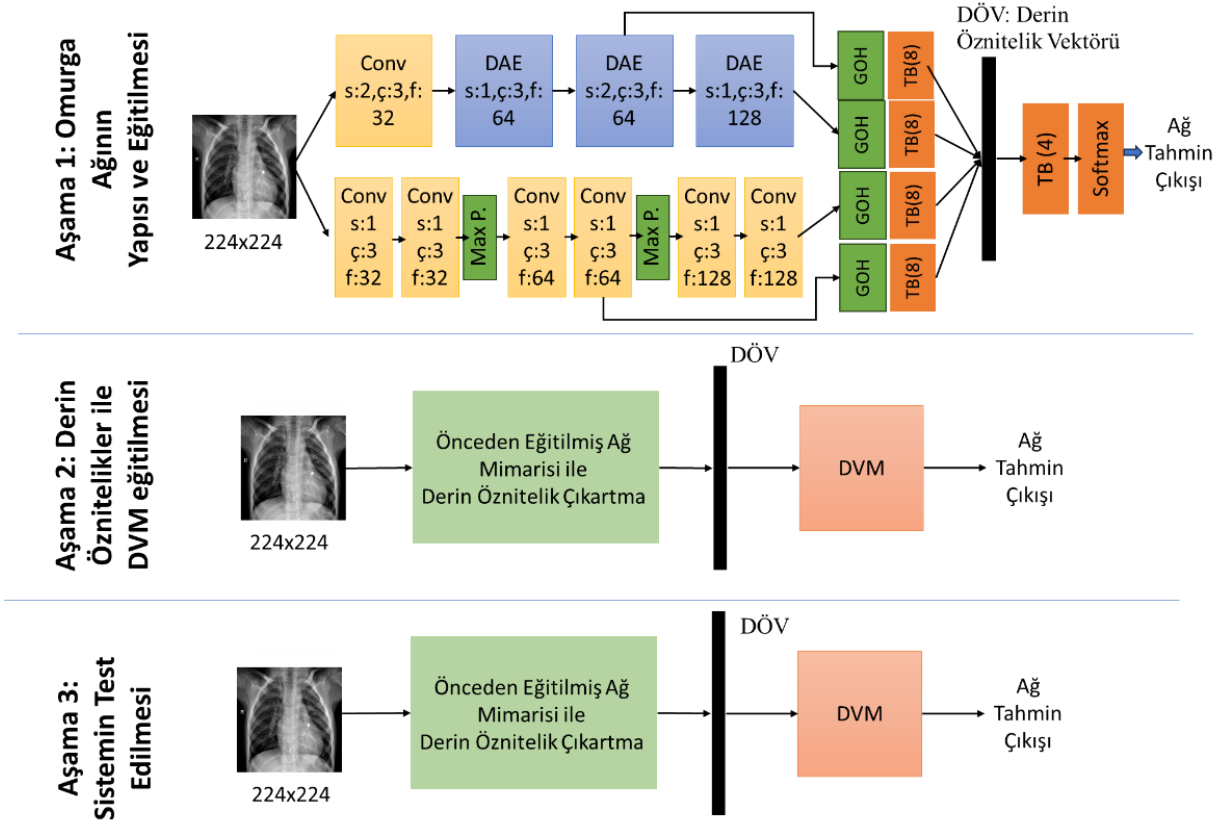
### 2.2. Önerilen Model (Proposed Model)

Bu çalışmada GR görüntü sınıflandırması için 3 aşamalı bir model geliştirilmiştir. Derin özniteliklere dayalı geliştirilen bu model Şekil 2'de verilmiştir.

Önerilen model temel olarak güçlü derin öznitelikler ile beslenen Destek Vektör Makinesi (DVM) sınıflandırıcısıdır. Önerilen model 3 aşamadan oluşmaktadır. Birinci aşamada GR görüntülerinin sınıflandırılma için güçlü derin özniteliklerin elde edilmesini amaçlanmaktadır. Bunun için Derinlemesine Ayrılabilir Evrişim (DAE) [21][22] ve geleneksel evrişim katmanlardan oluşan iki ESA mimarisi paralel olarak giriş görüntüsüne uygulanmaktadır. Daha sonra bu iki mimarinin son katmanlarından ve ara katmanlarından alınan 4 öznitelik haritası birleştirilmektedir. Bu ESA mimarilerinin ayrıntıları Şekil 2'de verilmiştir. Bu yapının temel amacı UNet [23] ve FPNet [24] modellerinde olduğu gibi zengin içerikli öznitelik haritaları elde etmektir. Geleneksel bir ESA mimarisinde ilk katmanlarda kenar, renk ve doku gibi mekânsal öznitelikler içerirken son katmanlarda anlamsal öznitelikler içermektedir [25][26]. Şekil 2'de önerilen bu yapı sayesinde GR görüntülerindeki hastalıkları sınıflandırmak için daha kapsamlı derin öznitelikler elde edilmesi sağlanmıştır. Ayrıca Şekil 2'de görüleceği üzere öznitelik çıkartma katmanları DAE ve geleneksel evrişim katmanları paralel olarak uygulanmıştır. Şekil 3'te geleneksel evrişim katmanı ve DAE katman yapıları verilmiştir.



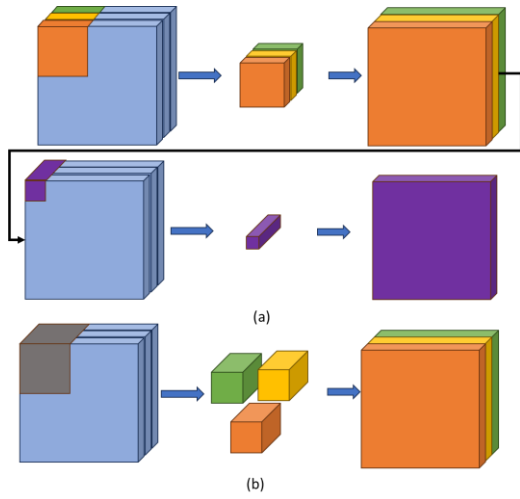
Şekil 1. Dört farklı sınıfa ait örnek GR görüntüleri  
(Sample CXR images of four different classes)



Şekil 2. GR sınıflandırması için önerilen model. Burada verilen s, ç ve f sırası ile stride (atlama) değeri, çekirdek boyutu ve filtre sayısını temsil etmektedir. Diğer yandan GOH ve TB sırası ile global ortalama havuzlama ve tam bağlı katmanları göstermektedir

(Proposed model for CXR classification. Here s, ç and f represent the stride value, kernel size and number of filters, respectively. On the other hand, GAP and FC represent global average pooling and fully connected layers, respectively)





Şekil 3. (a) Derinlemesine ayrılabilir evrişim ve (b) geleneksel evrişim işlemi

(a) Depthwise separable convolution and (b) traditional convolution operation)

DAE katmanı ConvMixer [27] ve MobileNet [28] mimarilerinde kullanılan ve her kanala bağımsız filtre uygulamasını sağlayan etkili bir evrişim bloğudur. Şekil 3(a)'da verilen DAE bloğunda temel olarak girdiye derinlemesine evrişim ve sonrasında noktasal evrişim katmanı uygulanmaktadır. Şekil 3(b)'de görüleceği üzere geleneksel evrişim işleminde girdi bir bütün olarak ele alınmakta ve tüm kanallar ilişkilendirilerek bir evrişim işlemi çıktısı elde edilmektedir. Şekil 3(a)'da verilen derinlemesine evrişim işleminde, girişin her kanalı için bağımsız olarak evrişimler gerçekleştirilen ve her kanalı ayrı ayrı işlerken uzamsal bilgileri korunmaktadır. Bu evrişim ile uzamsal özelliklere uyum sağlanarak daha iyi özellik çıkarımı gerçekleştirilmektedir. Diğer yandan derinlemesine evrişim işlemi sonunda kanallar arasındaki ilişkiyi değerlendirmek içinde bir noktasal evrişim işlemi uygulanmaktadır. Noktasal evrişim, derinlemesine evrişim çıkışından türetilen doğrusal bir kombinasyon aracılığıyla çeşitli kanallardaki ayrıntıları birleştirmek için 1x1 çekirdeklerin kullanılmasını içermektedir. Birincil işlevi, uzamsal boyutları korurken veya potansiyel olarak azaltırken özellik haritasının derinliğini genişletmektir. Sıklıkla 'darboğaz' katmanı olarak adlandırılan noktasal evrişim, kanal sayısını azaltarak hesaplama karmaşıklığının azaltılmasına yardımcı olmaktadır [29]. Şekil 2'de verilen birinci aşamadaki omurga ağında her DAE katman sonrasında ReLU aktivasyon fonksiyonu uygulanmaktadır.

Önerilen modelde tasarlanan omurga ağında geleneksel evrişim ve DAE katman sonunda çeşitli boyutlarda derin öznetelikler elde edilmektedir. Elde edilen bu öznetelikler nihai öznetelik vektörü elde etmek için birleştirilmiştir. Birleştirme işleminde öncelikle her öznetelik haritasına Global Ortalama Havuzlama (GOH) katmanı uygulanmaktadır. Daha sonra 8 nöronlu bir Tam Bağlı (TB) katman uygulanmaktadır. Bu sayede her öznetelik haritası için 8 adet öznetelik elde edilmektedir. Son olarak, bu öznetelikler birleştirilerek 32 adet derin öznetelik elde

edilmektedir. Birleştirme işleminin formülasyonu Denklem (1)'de verilmiştir..

$$K_i = TB^{n_{\text{öron}}=8}(GOH(F_i))$$

$$F_{\text{nihai}} = \text{concat} \begin{bmatrix} K_{DAE1} \\ K_{DAE2} \\ K_{GE1} \\ K_{GE2} \end{bmatrix} \quad (1)$$

Burada verilen  $F_i$  omurga ağından elde edilen öznetelik haritalarıdır. Şekil 2'de gösterildiği gibi  $F_i$  öznetelikleri için 2 adet geleneksel evrişim çıktısı ve 2 adet DAE katman çıktısı kullanılmıştır. Bu öznetelikler öncelikle GOH katmanı ile vektörleştirilmiştir. Daha sonra 8 nöronlu TB katmanı uygulanmıştır. Sonuç olarak her bir öznetelik haritası için 8 değerden oluşan  $K_i$  öznetelik vektörü elde edilmiştir. Birleştirme işleminin son aşamasında tüm vektörler keras kütüphanesinde bulunan birleştirme (concat) katmanı ile birleştirilerek 32 adet değere sahip nihai öznetelik vektörü ( $F_{\text{nihai}}$ ) elde edilmiştir.

Bu öznetelikleri en iyi değere optimize etmek için ağ mimarisinin sonunda softmax sınıflandırıcı uygulanarak ağ mimarisinin parametreleri eğitilmektedir. Sonuç olarak birinci aşamada GR görüntü sınıflandırması için güçlü bir birleştirilmiş Derin Öznetelik Vektörü (DÖV) elde edilmektedir.

Birçok sınıflandırma probleminde derin özneteliklere softmax sınıflandırıcısı uygulanmaktadır [29][30]. Bunun yanında bazı aşırı öğrenme problemi ile başa çıkmak için DVM ve karar ağacı gibi geleneksel makine öğrenme gibi yöntemler uygulanmıştır. Bu sayede daha yüksek başarılar sağlanmıştır [31]. Bu çalışmalara benzer olarak önerilen yaklaşımın ikinci aşamasında birleştirilmiş özneteliklere DVM sınıflandırıcı uygulanmıştır.

DVM, sınıflandırma ve regresyon görevleri için kullanılan güçlü bir denetimli makine öğrenme algoritmasıdır. Birincil hedefi, sınıflar arasındaki mesafeyi maksimuma çıkarırken veri noktalarını farklı sınıflara en iyi şekilde ayıran bir hiper düzlem bulmaktır. DVM algoritmasında hiper düzlem, daha yüksek boyutlarda bir alt uzaydır. DVM, verileri sınıf etiketlerine göre en iyi şekilde ayıran hiper düzlemi bulmayı amaçlamaktadır. Diğer yandan destek vektörleri, hiper düzleme en yakın olan ve hiper düzlemin konumunu ve yönelimini doğrudan etkileyen veri noktalarıdır. Önerilen modelin ikinci aşamasında DVM sınıflandırıcısı eğitim veri kümesi ile eğitilmiştir. Önerilen modelin üçüncü aşamasında geliştirilen sistem test edilmiştir. Yapılan deneysel çalışmalarda DVM sınıflandırıcısı softmax sınıflandırıcısına göre %1,71 daha yüksek doğruluk değeri sağladığı görülmüştür. Bu sonuçlar GR görüntü sınıflandırması için DVM modelinin etkili olduğunu kanıtlamıştır.

### 3. DENEYSEL ÇALIŞMALAR VE SONUÇLARI (EXPERIMENTAL STUDIES AND RESULTS)

Bu bölümde deneysel çalışmalarda kullanılan hiperparametreler, kullanılan değerlendirme ölçütleri, deneysel çalışmalardan elde edilen sınıflandırma sonuçları ve bu sonuçların tartışılması sunulmaktadır.

#### 3.1. Parametre Ayarları (Parameter Settings)

Deneysel çalışmalar 11 GB RTX 3080 ti GPU, 64 GB RAM ve i9-9900KF işlemci ve Ubuntu 20.04 işletim sistemine sahip bir bilgisayarda gerçekleştirilmiştir. Derin öğrenme modeli Keras-TensorFlow kütüphanesi kullanılarak tasarlanmıştır. Deneysel çalışmalarda kullanılan hiperparametrelerden, adım sayısı (epochs) 100, yığın boyutu (batch size) 16 ve öğrenme oranı (learning rate) 0,0001 olarak ayarlanmıştır. Bu değerler yapılan birkaç deneme sonucunda en iyi eğitim senaryosu olarak kabul edilmiştir. Çalışmada veri kümesi %80 eğitim %10 test, %10 doğrulama olarak alınmıştır. Yani; 7132 GR görüntüsünden 5706 görüntü eğitim için, 713 görüntü test için ve 713 görüntü validation (doğrulama) için kullanılmıştır. Buna ek olarak önerilen modelin eğitiminde optimizasyon için Adam modeli kullanılmıştır.

#### 3.2. Değerlendirme ölçütleri (Evaluation Criteria)

Değerlendirme ölçütleri, bir modelin tahminlerinin etkinliğinin nicel ve nesnel bir ölçüsünü sağlamaktadır. Ayrıca sınıflandırma performansını değerlendirmek için de gereklidirler. Bu ölçütler, her biri kendi güçlü yönleri ve sınırlamaları olan modelin performansı hakkında farklı perspektifler sağlamaktadır. Önerilen modelin etkinliği F1-skoru (F1-score), duyarlılık (recall), kesinlik (precision) ve sınıflandırma doğruluğu (accuracy) gibi değerlendirme ölçütleri kullanılarak ölçülmektedir. Önemli bir performans değerlendirme ölçütü olan doğruluk, tıbbi teşhislerde, özellikle dengesiz veri kümelerinde (bir hastalığın diğerlerinden daha yaygın olabileceği durumlarda) genel doğruluğu ölçerken, doğruluk tek başına doğru bir tablo sunmayabilir. Örneğin, Tüberküloz vakalarının COVID-19 veya zatürreden daha yaygın olduğu bir veri setinde, modelin Tüberküloz vakalarını doğru şekilde tanımlarken, potansiyel olarak COVID-19 veya zatürre vakalarını eksik veya yanlış sınıflandırarak yüksek doğruluk elde edilebilmektedir. Bu tıbbi bağlamda kesinlik, belirli bir hastalığın, o hastalık olarak tahmin edilen tüm durumlar arasında doğru şekilde tanımlanmasının önemini vurgulamaktadır. Örneğin, COVID-19 için yüksek hassasiyet, modelin tahmin edilen COVID-19 vakaları arasında gerçek COVID-19 vakalarını doğru bir şekilde tanımlama ve hatalı pozitifleri en aza indirme yeteneğini göstermektedir. Duyarlılık, modelin veri kümesi içindeki belirli bir hastalığın tüm örneklerini yakalama yeteneğine odaklanmaktadır. Tıbbi teşhisler için yüksek duyarlılık, modelin belirli bir hastalık için yüksek oranda gerçek pozitif vakayı başarılı bir şekilde tespit etmesi ve yanlış negatifleri azaltması anlamına gelmektedir. F1 skoru, kesinlik ve duyarlılığı dengelemekte ve her iki ölçümü de eşit şekilde dikkate alan

tek bir ölçüm sunmaktadır. Eşit olmayan bir sınıf dağılımı olduğunda veya hem yanlış pozitiflerin hem de yanlış negatiflerin en aza indirilmesi gerektiğinde özellikle değerlidir.

Doğruluk, yöntem tarafından üretilen doğru tahminlerin yüzdesini ölçmektedir. Doğru tahminlerin sayısının yapılan toplam tahmin sayısına bölünmesiyle belirlenmektedir. Doğruluk değerinin hesaplanması Denklem (2)'de detaylandırıldığı şekilde yapılmaktadır. Önerilen yöntemle yapılan tüm pozitif tahminlerdeki doğru pozitiflerin oranının bir ölçüsü olan kesinlik, doğru pozitiflerin sayısının doğru pozitifler ve yanlış pozitiflerin toplamına bölünmesiyle belirlenmektedir. Kesinlik değerinin hesaplanması Denklem (3)'teki gibidir. Duyarlılık, doğru pozitiflerin sayısının doğru pozitifler ve yanlış negatiflerin toplamına bölünmesiyle belirlenmektedir. Duyarlılık Denklem (4)'teki gibi formüle edilmektedir. Duyarlılık ve kesinliğin harmonik ortalaması olan F1 skoru, özellikle sınıflar dengesiz olduğunda duyarlılık ve kesinliği dengelemek için vazgeçilmez bir ölçüttür. Hem kesinliği hem de duyarlılığı yakalayan tek bir puan sağlayarak genel model performansı değerlendirmesi için güçlü bir ölçüt haline getirmektedir. F1 skoru Denklem (5)'te belirtildiği şekilde hesaplanmaktadır [32][33].

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - \text{skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (5)$$

Tüm bu değerlendirme ölçütleri karışıklık matrisinden çıkarılmaktadır. Bir karışıklık matrisinin açıklayıcı bir gösterimi Şekil 4'te görülebilmektedir. Standart bir karışıklık matrisinde dört ana unsur bulunmaktadır: Gerçek Pozitifler (TP): Bunlar, gerçek kategorinin olumlu olduğu ve modelin bunu doğru bir şekilde olumlu olarak tahmin ettiği durumları ifade etmektedir. Gerçek Negatifler (TN): Bunlar, gerçek kategorinin negatif olduğu ve modelin bunu negatif olarak doğru bir şekilde tahmin ettiği durumları temsil etmektedir. Yanlış Pozitifler (FP): Bu örnekler, gerçek kategori negatif olduğunda ortaya çıkar, ancak model bunu yanlış bir şekilde olumlu olarak tahmin etmektedir. Yanlış Negatifler (FN): Bunlar, gerçek kategorinin olumlu olduğu, ancak modelin hatalı bir şekilde olumsuz olarak tahmin ettiği durumlardan kaynaklanmaktadır [34].

		Gerçek	
		Pozitif (P)	Negatif (N)
Tahmin	Pozitif (P)	TP	FP
	Negatif (N)	FN	TN

Şekil 4. Karışıklık Matrisi (Confusion Matrix)

### 3.3. Deneysel Sonuçlar (Experimental Results)

Önerilen modelin sınıflandırma performansını analiz etmek için dört sınıflı (COVID19, Normal, Pnömoni ve TÜB) GR görüntüleri veri kümesi üzerinde kapsamlı deneysel çalışmalar gerçekleştirilmiştir. Deneysel çalışmalarda önerilen modelin başarısını test etmek için test veri kümesi kullanılmıştır. Önerilen modelin karışıklık matrisi Şekil 5'te verilmiştir. Test veri kümesinde 58 COVID19 GR görüntüsünden tamamının, 158 Normal GR görüntüsünden 151'inin, 427 pnömoni GR görüntüsünden 424'ünün ve son olarak 70 TÜB görüntüsünden tamamının doğru tahmin edildiği görülmektedir. 158 Normal GR görüntüsünden 7'si pnömoni ve 427 pnömoni GR görüntüsünden 3'ü Normal görüntü olarak yanlış sınıflandırılmıştır. Toplamda 713 GR görüntüsünden 703'ünün doğru bir şekilde tahmin edildiği görülmektedir. Karışıklık matrisinden her bir sınıf için TP, TN, FP ve FN değerleri incelendiğinde, COVID-19 için TP:58, TN:645, FP:0 ve FN:0'dır. Bu sonuçlara göre COVID-19 için doğruluk değeri %100'dür. Normal GR görüntülerinin yer aldığı sınıf için TP:151, TN:552, FP:3 ve FN:7'dir. Bu durumda, Normal sınıfı için elde edilen doğruluk değeri %98,60 olmaktadır. Pnömoni sınıfı için TP:424, TN:279, FP:7 ve FN:3'tür. Pnömoni sınıfının doğruluk değeri %98,60'tır. Son olarak TÜB sınıfı için TP: 70, TN:633, FP ve FN ise 0'dır. Bu sonuçlara göre TÜB sınıfının doğruluk değeri %100 olmaktadır. Bu dört sınıfın bazlı doğruluk değerlerinin ortalaması %99,30 olarak elde edilmektedir.

58	0	0	0
0	151	7	0
0	3	424	0
0	0	0	70

Şekil 5. Önerilen modelin karışıklık matrisi  
(Confusion matrix of the proposed model)

Önerilen modelin sınıf bazlı değerlendirme ölçütleri Tablo 1'de verilmiştir. Tablo 1 incelendiğinde önerilen modelin COVID19 ve TÜB görüntülerini %100 doğruluk, kesinlik, duyarlılık ve F1-skoru ile tespit ettiği görülmektedir. Benzer şekilde kullanılan GR veri kümesinde normal görüntüler %98,60 doğruluk, %98,05 kesinlik, %95,57 duyarlılık, %96,79 F1-skoru elde ederken, pnömoni %98,60 doğruluk, %98,38 kesinlik, %99,30 duyarlılık, %98,84 F1-skoru elde etmiştir. Sınıf bazlı sonuçlara bakıldığında, önerilen modelin COVID-19 ve TÜB hastalıklarını başarılı bir şekilde tespit edebildiği görülmektedir.

Önerilen model literatürden son yıllarda yapılmış farklı deneysel çalışmalar ile karşılaştırılmış ve sonuçlar Tablo 2'de verilmiştir. Tablo 2 incelendiğinde, Qagos vd. [13], GR görüntülerinin sınıflandırılması için evrişim, max-pooling, toplu normalleştirme, bırakma (dropout) ve sınıflandırma katmanı olarak da softmax kullanarak yeni bir ESA modeli önermiştir. Önerilen model 3x3 evrişimlerden oluşmaktadır. Bu ESA mimarisinin performansını test etmek için dört sınıftan (TÜB, COVID-19, Normal ve pnömoni) oluşan GR görüntü veri kümesi üzerinde gerçekleştirilen deneysel çalışmalar sonucunda %94,53 doğruluk değeri elde etmiştir. Benzer şekilde GR görüntüleri kullanılarak gerçekleştirilen diğer çalışmaların sonuçları şu şekildedir. Qagos vd. [13], aynı ESA mimarisini kullanarak üç sınıflı (normal, COVID-19 ve pnömoni) GR veri kümesi üzerinde ESA modeli ile %95,02 elde ettiler. Bhandari vd. [14], beş 2 boyutlu evrişim, dört max-pooling, beş dropout katmanı ve softmax sınıflandırıcısından oluşan bir ESA modeli kullanarak GR görüntülerini sınıflandırmıştır. Bu ESA modeli ile dört sınıflı GR veri kümesi üzerinde geliştirilen deneysel çalışmalar sonucunda %94,31 doğruluk sonucu bulunmuştur. Sitaula vd. [15], GR görüntülerinin sınıflandırılması için dikkat tabanlı bir VGG modeli önermiştir. Önerilen model, önceden eğitilmiş DL modeline (VGG-16) ve dikkat modülüne dayanmaktadır. Bu model son katmanında sınıflandırıcı olarak softmax kullanılmaktadır. Önerilen dikkat tabanlı VGG16 modeli ile dört sınıflı GR veri kümesi üzerinde gerçekleştirilen deneysel çalışmalar sonucunda %85,43 doğruluk değeri elde edilmiştir. Liu vd. [35], iki sınıflı (TÜB ve normal) GR veri kümesi üzerinde transfer öğrenme ile AlexNet ve GoogLeNet mimarileri kullanılarak %85,68 doğruluk değeri elde etmişlerdir. Her iki modelde de son sınıflandırma katmanı olarak softmax sınıflandırıcısı kullanılarak %85,68 doğruluk değeri elde edilmiştir. Rahimzadeh vd. [36], GR görüntülerinin sınıflandırılması için bir topluluk öğrenimi modeli oluşturmaktadır. Bu topluluk öğrenimi modeli Xception ve ResNet50 modellerinden oluşmaktadır. Her iki model birleştirildikten sonra son sınıflandırıcı katmanı olarak softmax kullanılmaktadır. Önerilen topluluk öğrenimi modelinin performansını test etmek için üç sınıflı (normal, COVID-19 ve pnömoni) GR veri kümesi kullanılmıştır. Bu topluluk öğrenimi modeli ile %91,40 doğruluk değeri bulunmuştur. Khan vd. [37], dört sınıflı GR veri kümesi üzerinde ESA tabanlı CoroNet kullanılarak sınıflandırma işlemini gerçekleştirdiler. Önerilen CoroNet, dört sınıflı GR görüntülerini sınıflandırmak için derin bir evrişimli sinir ağı modelidir. Bu modelin son sınıflandırıcı katmanında softmax kullanılmaktadır. Önerilen CoroNet modeli ile gerçekleştirilen deneysel çalışmalar sonucunda %89,60 doğruluk değeri bulunmuştur. Shastri vd. [38], açık olarak erişilebilen bir veri kümesi yardımıyla GR dijital görüntülerini kullanarak COVID-19 hastalığını tespit etmeye yönelik bir CheXImageNet modeli tanıttılar. Önerilen model, evrişim, max-pooling, toplu normalleştirme, dropout katmanlarından oluşmaktadır. Diğer ESA modellerinden farklı olarak aktivasyon fonksiyonu olarak Leaky\_ReLU kullanılmıştır. Son sınıflandırıcı katmanında ise softmax kullanılmıştır. Üç

sınıflı (normal, COVID-19, pnömoni) GR veri kümesi ile gerçekleştirilen deneysel çalışmalar sonucunda %93 doğruluk sonucu bulunmuştur. Al-Timemy vd. [39], beş sınıflı GR veri kümesi üzerinde ResNet50 ve alt uzay ayırt edici sınıflandırıcı topluluğu modelini kullanarak sınıflandırma işlemini gerçekleştirmiştir. Derin öznitelik hesaplaması için ResNet50'den ve alt uzay ayırt edici sınıflandırıcı topluluğundan oluşan yeni bir boru hattı, %91,6'lık bir tespit doğruluğu elde ederek beş sınıfın sınıflandırılmasında en iyi performansı gösterdi. Bunun yanı sıra üç sınıftan (COVID-19, normal ve TÜB) oluşan bir veri kümesi üzerinde aynı model kullanılarak %98,6 doğruluk değeri elde ettiler. Asif vd. [40], üç sınıflı GR veri kümesi üzerinde derin ESA ve transfer öğrenmeli InceptionV3 birlikte kullanılarak %93 doğruluk değerini bulmuştur. Önerilen modelin son sınıflandırıcı katmanı diğer çalışmalarda da olduğu gibi softmax'tir. Metwally vd. [41], dört sınıflı 7135 GR görüntüsü içeren veri kümesi

üzerinde önceden eğitilmiş EfficientNetB0, Xception ve NasnetLarge modeli kullanılarak %91,3 doğruluk değeri bulunmuştur. Literatürdeki yöntemler incelendiğinde hepsinin son sınıflandırıcı katmanında softmax kullanıldığı görülmüştür. Bundan yola çıkarak, önerilen modelin son katmanında iki farklı sınıflandırıcı (softmax ve DVM) kullanılarak deneysel çalışmaların sonuçları verilmiştir. Son katmanda softmax sınıflandırıcı kullanılarak dört sınıflı GR veri kümesi üzerinde %97,59 doğruluk değeri elde edilirken, son katmanda DVM kullanılarak yine aynı veri kümesi üzerinde %99,30 doğruluk değeri elde edilmiştir. Bizim önerdiğimiz derin özniteliklere dayalı ESA modelinde son sınıflandırıcı olarak DVM kullanılmaktadır. DVM kullanımının sınıflandırma doğruluğunu %1,71 arttırdığı görülmektedir. Tablo 2'deki tüm modeller incelendiğinde son sınıflandırıcı olarak DVM'nin kullanıldığı önerilen modelimizin açık bir şekilde başarılı olduğu görülmektedir.

Tablo 1. Önerilen modelin sınıf bazlı sonuçları(Class-based results of the proposed model) (%)

Değerlendirme ölçütü	COVID19	NORMAL	PNÖMONİ	TÜB	Ortalama
Doğruluk	100	98,60	98,60	100	99,30
Kesinlik	100	98,05	98,38	100	99,10
Duyarlılık	100	95,57	99,30	100	98,71
F1-skor	100	96,79	98,84	100	98,90

Tablo 2. Önerilen modelin son teknoloji yöntemlerle karşılaştırılması

(Comparison of the proposed model with state-of-the-art methods)

Referans çalışma	Sınıf isimleri ve örnek sayısı	Sınıf sayısı	Model	Doğruluk (%)
Qagos vd. [13]	Normal = 1583, COVID19 = 576 Pnömoni = 4273, TÜB = 155	4	Özel ESA	94,53
Qagos vd. [13]	Normal = 1341, COVID19 = 864 Pnömoni = 1345	3	Özel ESA	95,02
Bhandari vd. [14]	Normal = 1583, COVID19 = 576 Pnömoni = 4273, TÜB = 700	4	Özel ESA	94,31
Sitaula vd. [15]	Normal = 310, Bakteriye Pnömoni = 330, Viral Pnömoni = 327, COVID19 = 284	4	Dikkat tabanlı VGG	85,43
Liu vd. [35]	TÜB = 4248, Normal = 453	2	Transfer öğrenme ile AlexNet ve GoogLeNet	85,68
Rahimzadeh vd. [36]	Normal = 8851, COVID19 = 180 Pnömoni = 6054	3	Topluluk Öğrenimi (Xception ve ResNet50)	91,40
Khan vd. [37]	Normal = 310, Bakteriye Pnömoni = 330, Viral Pnömoni = 327, COVID19 = 284	4	ESA tabanlı CoroNet	89,60
Shastri vd. [38]	Normal = 1341, COVID19 = 864 Pnömoni = 1345	3	CheXImageNet model	93,00
Al-Timemy vd. [39]	Normal = 439, COVID19 = 435 Bakteriye Pnömoni = 439, Viral Pnömoni = 439, TÜB = 434	5	ResNet50 özellikleri ve altuzay ayırt edici sınıflandırıcı topluluğu	91,60
Asif vd. [40]	Normal = 1341, COVID19 = 864 PNEUMONIA = 1345	3	Derin ESA + transfer öğrenmeli Inception V3	93,00
Metwally vd. [41]	Normal = 1583, COVID19 = 576 Pnömoni = 4273, TÜB = 703	4	Önceden eğitilmiş EfficientNetB0, Xception ve NasnetLarge	91,3
Önerilen Model	Normal = 1583, COVID19 = 576 Pnömoni = 4273, TÜB = 700	4	Derin özniteliklere dayalı ESA modeli + Softmax	97,59
Önerilen Model	Normal = 1583, COVID19 = 576 Pnömoni = 4273, TÜB = 700	4	Derin özniteliklere dayalı ESA modeli + DVM	99,30

#### 4. SONUÇ (CONCLUSION)

COVID-19, pnömoni ve TÜB tedavi edilmediğinde veya yanlış yönetildiğinde ölümcül olabilmektedir. GR görüntüleri, bu tür akciğer hastalıklarının teşhisi için en sık kullanılan görüntüleme yöntemidir. Bu görüntülerin yorumlanması için uzmanlara ihtiyaç duyulmaktadır. Hasta sayısı arttıkça, az sayıda olan bu uzmanlar aşırı çalışmak zorunda kalmakta ve bu süreçte tanılarının gecikmesine ve sonuçların zamanında rapor edilmemesine neden olmaktadır. Ayrıca, bu hastalıkların öznel doğası nedeniyle, uzmanların aynı GR görüntüsü yorumu için farklı sonuçlar üretmesi muhtemeldir. Akciğer hastalıklarının tedavisi, özellikle bir durum gözden kaçarsa ve ileride ek sağlık hizmeti maliyetleri gerektirirse pahalı olabilmektedir. Aynı zamanda, bu tür sorunların üstesinden gelmek için, çoklu akciğer hastalıklarının tespiti için otomasyon, teşhis sürecini kolaylaştırmaya yardımcı olabilmekte ve sağlık hizmeti sağlayıcılarının yalnızca uzmanlara güvenmeden tedavi konusunda daha bilinçli kararlar almasını sağlayabilmektedir. Bu, bekleme sürelerinin kısaltılmasına, hasta sonuçlarının iyileştirilmesine ve uzmanların üzerindeki yükün azaltılmasına yardımcı olurken daha hızlı ve daha verimli teşhisler yapılmasını sağlayabilmektedir. Bu doğrultuda son yıllarda bilgisayar destekli teşhis oldukça önemli rol oynamaktadır. Bu çalışmada, enfekte hastaların GR görüntülerini kullanarak pnömoni, COVID19 ve TÜB'ü tespit etmek ve teşhis etmek için yeni bir derin öğrenme modeli önerilmektedir. Bu model, derin özelliklere dayalı yeni bir modeldir ve üç aşamadan oluşmaktadır. Bu modelin ilk aşamasında, eğitim amacıyla derin özellikler elde etmek için özel olarak tasarlanmış bir ESA mimarisi kullanılmaktadır. Bu ESA mimarisi, derinlemesine ayarlanabilir evrişim, geleneksel evrişim ve tamamen bağlantılı katmanları kapsamaktadır. Önerilen modelin bir sonraki aşamasında, bu derin özellikler DVM olarak bilinen bir sınıflandırıcı için girdi olarak kullanılmaktadır. Ayrıca, bu aşamada, DVM sınıflandırıcısı derin özellikler kullanılarak eğitilmektedir. Son olarak, bu yenilikçi modelin sonuç aşamasında, bir test veri kümesi kullanılarak modelin performansının bir değerlendirilmesi yapılmaktadır. Önerilen modelin performans analizi için halka açık bir platform olan Kaggle'da dört sınıf (COVID-19, Pnömoni, normal ve TÜB) ve 7132 GR görüntüsü içeren bir veri kümesi üzerinde deneysel çalışmalar gerçekleştirilmiştir. Önerilen model, GR'lerde belirtilen hastalıkları %99,30'luk ortalama doğruluk, %99,10 kesinlik, %98,71 duyarlılık ve %98,90 F1 skor değerleri ile tespit etmiştir. Ayrıca sınıf bazlı doğruluk değerleri incelendiğinde, önerilen model COVID19 ve TÜB hastalıklarını %100 doğrulukla tespit ederken, pnömoni hastalığını %98,60 doğruluk ile tespit etmiştir. Aynı zamanda önerilen modelde kullanılan DVM sınıflandırıcısının aksine çoğu derin öğrenme modellerinin son katmanında sınıflandırıcı olarak kullanılan softmax sınıflandırıcısı ile de deneysel çalışmalar yapılmıştır. Önerilen modelde DVM yerine, softmax sınıflandırıcısı kullanıldığında %97,59 doğruluk değeri bulunmuştur. DVM ile karşılaştırıldığında, DVM sınıflandırıcısının

%1,71 daha başarılı sonuç ürettiği görülmektedir. Son olarak, literatürdeki benzer son teknoloji çalışmalar ile karşılaştırmalı analizler, önerilen modelin diğer alternatif çalışmalardan daha iyi performans gösterdiğini ortaya koymaktadır. Bu nedenle önerilen model, klinisyenlerin COVID-19, TÜB ve pnömoni vakalarını daha doğru tespit etmelerine yardımcı olabilmektedir. Yakın gelecekteki araştırma doğrultusunda, öncelikle önerilen modelin farklı veri kümelerindeki performansları incelenecektir. Bunun yanı sıra, bu çalışmada bahsedilmeyen diğer akciğer hastalıklarını, önerdiğimiz modele dayanarak tespit etmeyi ve sınıflandırmayı amaçlamaktayız.

#### KAYNAKLAR (REFERENCES)



- [1] D. Visca, C. W. M. Ong, S. Tiberi et al., "Tuberculosis and COVID-19 interaction: A review of biological, clinical and public health effects", *Pulmonology*, 27(2), 151–165, 2021.
- [2] A. E. Gorbalenya, S. C. Baker, R. S. Baric et al., "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2", *Nature Microbiology*, 5(4), 536–544, 2020.
- [3] S. R. Islam, S. P. Maity, A. K. Ray, and M. Mandal, "Deep learning on compressed sensing measurements in pneumonia detection", *International Journal of Imaging Systems and Technology (IMA)*, 32(1), 41–54, 2022.
- [4] A. H. van't Hoog, H. K. Meme, K. F. Laserson et al., "Screening strategies for tuberculosis prevalence surveys: The value of chest radiography and symptoms", *PLoS One*, 7(7), 1–9, 2012.
- [5] H. Abdul, S. Hashmi, and H. M. Asif, "Early detection of COVID-19", *Frontiers in Medicine*, 2020.
- [6] M. Mamalakos, A. J. Swift, B. Vorselaars et al., "DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays", *Computerized Medical Imaging and Graphics*, 94(October), 102008, 2021.
- [7] A. A. Soltan, S. Kouchaki, T. Zhu et al., "Artificial intelligence driven assessment of routinely collected healthcare data is an effective screening test for COVID-19 in patients presenting to hospital", *medRxiv*, 2020.
- [8] G. J. Williams, P. Macaskill, M. Kerr et al., "Variability and accuracy in interpretation of consolidation on chest radiography for diagnosing pneumonia in children under 5 years of age", *Pediatric Pulmonology*, 48(12), 1195–1200, 2013.
- [9] D. Stacey, F. Legare, K. Lewis et al., "Decision aids for people facing health treatment or screening decisions", *Cochrane Database of Systematic Reviews*, 2017(4), 14651858, 2017.
- [10] H. Greenspan, R. San José Estépar, W. J. Niessen, E. Siegel, and M. Nielsen, "Position paper on COVID-19 imaging and AI: From the clinical needs and technological challenges to initial AI solutions at the lab and national level towards a new era for AI in healthcare", *Medical Image Analysis*, 66(April), 101800, 2020.
- [11] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review", *Chaos, Solitons and Fractals*, 139, 110059, 2020.

- [12] M. S. Ahmed, A. Rahman, F. AlGhamdi et al., “Joint Diagnosis of Pneumonia, COVID-19, and Tuberculosis from Chest X-ray Images: A Deep Learning Approach”, *Diagnostics*, 13(15), 2023.
- [13] N. N. Qaqos and O. S. Kareem, “COVID-19 Diagnosis from Chest X-ray Images Using Deep Learning Approach”, **3rd International Conference on Advanced Science and Engineering (ICOASE 2020)**, Duhok, 110–116, 23-24 Aralık, 2020.
- [14] M. Bhandari, T. B. Shahi, B. Siku, and A. Neupane, “Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI”, *Computers in Biology and Medicine*, 150(September), 106156, 2022.
- [15] C. Sitaula and M. B. Hossain, “Attention-based VGG-16 model for COVID-19 chest X-ray image classification”, *Applied Intelligence*, 51(5), 2850–2863, 2021.
- [16] A. Bashar, G. Latif, G. Ben Brahim, N. Mohammad, and J. Alghazo, “COVID-19 pneumonia detection using optimized deep learning techniques”, *Diagnostics*, 11(11), 1–18, 2021.
- [17] P. Szepesi and L. Szilágyi, “Detection of pneumonia using convolutional neural networks and deep learning”, *Biocybernetics and Biomedical Engineering*, 42(3), 1012–1022, 2022.
- [18] S. K. T. Hwa, M. H. A. Hijazi, A. Bade, R. Yaakob, and M. S. Jeffree, “Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images”, *IAES International Journal of Artificial Intelligence*, 8(4), 429–435, 2019.
- [19] B. Ibrokhimov and J.-Y. Kang, “Deep Learning Model for COVID-19-Infected Pneumonia Diagnosis Using Chest Radiography Images”, *BioMedInformatics*, 2(4), 654–670, 2022.
- [20] Internet: JTIPTJ, Chest X-Ray (Pneumonia,Covid-19,Tuberculosis), Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/jtiptj/chest-xray-pneumoniacovid19tuberculosis?select=train>, 20.09.2023.
- [21] H. Üzen, M. Turkoglu, M. Aslan, and D. Hanbay, “Depth-wise Squeeze and Excitation Block-based Efficient-Unet model for surface defect detection”, *The Visual Computer*, 2022.
- [22] H. Firat, M. E. Asker, and D. Hanbay, “Depthwise Separable Convolution Based Residual Network Architecture for Hyperspectral Image Classification”, *Gazi Üniversitesi Fen Bilim. Derg. Part C Tasarım ve Teknoloji*, 10(2), 242–258, 2022.
- [23] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, 234–241, 2015.
- [24] S. Seferbekov, V. Iglovikov, A. Buslaev, and A. Shvets, “Feature pyramid network for multi-class land segmentation”, **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, Salt Lake City, UT, USA, 18-22 June, 272–275, 2018.
- [25] F. Meng, X. Wang, F. Shao, D. Wang, and X. Hua, “Energy-efficient gabor kernels in neural networks with genetic algorithm training method”, *Electronics*, 8(1), 2019.
- [26] A. Mahendran and A. Vedaldi, “Visualizing Deep Convolutional Neural Networks Using Natural Pre-images”, *International Journal of Computer Vision*, 120(3), 233–255, 2016.
- [27] A. Trockman and J. Z. Kolter, “Patches Are All You Need?”, 1–16, 2022.
- [28] A. G. Howard, M. Zhu, B. Chen et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, 2017.
- [29] M. E. Asker, “Hyperspectral image classification method based on squeeze-and-excitation networks, depthwise separable convolution and multibranch feature fusion”, *Earth Science Informatics*, 1427–1448, 2023.
- [30] M. Türkoğlu, K. Hanbay, I. S. Sivrikaya, and D. Hanbay, “Derin Evrişimsel Sinir Ağı Kullanılarak Kayısı Hastalıklarının Sınıflandırılması”, *BEÜ Fen Bilimleri. Dergisi*, 9(1), 334–345, 2020.
- [31] M. F. Özdemir and D. Hanbay, “A Novel Covid-19 Detection System Based on PSO and Hybrid Feature Using Support Vector Machines”, *Journal of Computer Science, IDAP-2022*, 120–129, 2022.
- [32] Y. Ha, Z. Du, and J. Tian, “Fine-grained interactive attention learning for semi-supervised white blood cell classification”, *Biomedical Signal Processing and Control*, 75(September), 103611, 2022.
- [33] A. Naseri and A. Rezaei Nasab, “Automatic identification of minerals in thin sections using image processing”, *Journal of Ambient Intelligence and Humanized Computing*, 2021.
- [34] H. Firat, “Classification of microscopic peripheral blood cell images using multibranch lightweight CNN-based model”, *Neural Computing and Application*, 2023.
- [35] C. Liu, Y. Cao, M. Alcantara et al., “TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network”, **2018 25th IEEE International Conference on Image Processing (ICIP)**, 7-10 Ekim, Atina, 2314–2318, 7-10 Ekim, 2018.
- [36] M. Rahimzadeh and A. Attar, “A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2”, *Informatics in Medicine Unlocked*, 19, 100360, 2020.
- [37] A. I. Khan, J. L. Shah, and M. M. Bhat, “CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images”, *Computer Methods and Programs in Biomedicine*, 196, 105581, 2020.
- [38] S. Shastri, I. Kansal, S. Kumar, K. Singh, R. Popli, and V. Mansotra, “CheXImageNet: a novel architecture for accurate classification of Covid-19 with chest x-ray digital images using deep convolutional neural networks”, *Health and Technology*, 12(1), 193–204, 2022.
- [39] A. H. Al-Timemy, R. N. Khushaba, Z. M. Mosa, and J. Escudero, “An Efficient Mixture of Deep and Machine Learning Models for COVID-19 and Tuberculosis Detection Using X-Ray Images in Resource Limited Settings”, *Studies in System, Decision and Control*, 358(December), 77–100, 2021.
- [40] S. Asif, Y. Wenhui, H. Jin, and S. Jinhai, “Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network”, **2020 IEEE 6th International Conference on Computer and Communications (ICCC)**, Chengdu, 426–433, 11-14 Aralık, 2020.

- [41] M. A. H. Metwally, **Automatic Detection and Multi-Class Classification of COVID-19, Pneumonia, and Tuberculosis Diseases in Chest X-ray Images Using Deep Learning Techniques**, Yüksek Lisans Tezi, University of Victoria, 2022.

# Uyarlanabilir Gezinme Yapılarının E-Sağlık Uygulamalarına Uyumu

## Araştırma Makalesi/Research Article

 Levent Çetinkaya\*<sup>1</sup>,  Hafize keser<sup>2</sup>

<sup>1</sup>Bilgisayar ve Öğretim Teknolojileri Eğitimi, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, Türkiye

<sup>2</sup>Bilgisayar ve Öğretim Teknolojileri Eğitimi, Ankara Üniversitesi, Ankara, Türkiye

[lcetinkaya@comu.edu.tr](mailto:lcetinkaya@comu.edu.tr), [keser@ankara.edu.tr](mailto:keser@ankara.edu.tr)

(Geliş/Received:13.03.2023; Kabul/Accepted:12.01.2024)

DOI: 10.17671/gazibtd.1264449

**Özet**— Çevrimiçi uyarlanabilir gezinme ortamlarına yönelik ilkelerin e-sağlık uygulamalarına uyumu ve düzenlenmesi amacıyla gerçekleştirilen çalışmada, nitel ve nicel araştırma yaklaşımlarının işe koşulduğu yapılandırılmış çevrimiçi Delphi Tekniğinden faydalanılmıştır. Uzmanların çalışmaya katılımı ile ilgili onayları sonrasında başlatılan Delphi sürecinde tek bir soru kağıdında nitel ve nicel veriler eşzamanlı olarak toplanarak veri çeşitlenmesi yapılmıştır. Elde edilen veriler ve bu verilerin değerlendirilmesi ile elde edilen sonuçların uzmanlara sunumu şeklinde tekrarlanan süreç uzmanların uzlaşmaya ulaşmalarıyla son bulmuştur. Çalışmanın ilk turu sonucunda uzmanların gezinmenin uyarlanmasına yönelik ilkeleri yapısal olarak uygun bulduğu belirlenmiştir. Ancak, çevrimiçi öğrenme ortamlarına yönelik ilkelerin e-sağlık uygulamaları çerçevesinde yeniden yorumlanması gerekliliği uzmanlar tarafından bildirilmiştir. Uzman görüşleri doğrultusunda her turda yeniden şekillenen ilkeler üzerinde artık yeni fikirlerin üretilmediği, maddelere ilişkin tüm güçlü-zayıf yanların belirlendiği ve uzmanlardan herhangi bir görüş ya da önerinin gelmediği görülen üçüncü Delphi turunda çalışma sonlandırılmıştır. Çalışma sonucunda e-sağlık uygulamaları tasarımında gezinmenin uyarlanmasına yönelik tüm ilkeler üzerinde uzmanların uzlaşma sağladıkları ve edilen gözlemler arası uyumun (ICC=.930) mükemmel düzeyde olduğu belirlenmiştir.

**Anahtar Kelimeler**— e-sağlık, sağlık hizmetleri, uyarlanabilir sistemler, gezinmenin uyarlanması, uzlaşma, delphi tekniği.

## Integration of Adaptive Navigation Structures Into E-health Applications

**Abstract**— The structured online Delphi Method, which included qualitative and quantitative research methods, was used in the study to integrate and regulate the principles for online adaptive navigation environments to e-health applications. Data triangulation was accomplished in the Delphi process by concurrently gathering qualitative and quantitative data in a single questionnaire, which was started following the experts' approval of their involvement in the study. The experts finally reached an agreement after repeatedly presenting the data acquired and the outcomes generated by the examination of this data to the experts. The first round of the investigation led to the conclusion that the experts deemed the principles for the adaptation of navigation were structurally adequate. Yet, experts reported that the principles for online learning environments should be reinterpreted within the framework of e-health applications. The study was terminated in the third Delphi round when it was observed that no new ideas were produced on the principles that were reshaped in each round in line with the expert opinions, all the strengths and weaknesses related to the items were determined, and no views or suggestions were received from the experts. As a result of the study, it was determined that experts reached a consensus on all the principles for the adaptation of navigation in the design of e-health applications. It was also concluded that the agreement between observers (ICC=.930) was excellent.

**Keywords**— e-health, health services, adaptive systems, adaptive navigation, agreement, delphi technique.



## 1. GİRİŞ (INTRODUCTION)

Sağlık sistemindeki dijitalleşme sürecinin hızlanmasıyla birlikte sağlık sistemini oluşturan paydaşlar ile vatandaşlar arasında dijital bir köprünün oluşum süreci de hız kazanmaya başlamıştır. Dijitalleşme ile birlikte sağlıkta paradigma kaymasının sonucu olarak vatandaşların sağlıkla ilgili süreçlerini yönetebilecekleri hizmetlerin de yaygınlaştığı görülmektedir. Bu noktada Avrupa Birliği ülkeleri ve ABD başta olmak üzere birçok ülkede dijital sağlık hizmetleri önemli ölçüde ivme kazanmaya başlamıştır. Bireyler ve toplumlar için insan sağlığı ile sağlık hizmetlerinin iyileştirilmesinde bilgi ve iletişim teknolojilerinin işe koşulması olarak tanımlanan dijital sağlık [1] ile sunulan hizmetler bireye kendi sağlığında aktif bir katılımcı olma fırsatı sağlamaktadır. Temelde sağlık hizmetlerini anlık olarak ulaşılabilir kılarak, tasarlanan bu sistemlerin insan yaşamının kalitesini doğrudan etkileme potansiyeline sahip olduğunu söylemek mümkün gözükmemektedir. Her ne kadar sağlıkta dijitalleşme sürecinin tarihi eskiye dayanıyor olsa da sağlık sistemi ile vatandaşları bir araya getirebilecek yapılar oldukça yenidir ve sayıları da hızla artmaktadır. Dijitalleşen sağlık sistemi teşhis, tedavi, önleme ve tedavi süreçleri için yeni fırsatlar sunmakla birlikte sağlık profesyonellerinin yanı sıra vatandaşların da sağlık verilerine kolaylıkla ulaşmalarına olanak sağlamaktadır. Bu yapılar aracılığı ile bireylerin mevcut içerikle etkileşim içinde olmaları ve sağlıkla ilgili bilgilerine erişimlerine olanak sağlanarak iletişim sürecinde de aktif olarak katılmalarının desteklenmesi amaçlanmaktadır [2].

Sağlık hizmetlerinden etkin bir şekilde faydalanılması noktasında e-sağlık hizmetlerinin verimli bir şekilde işe koşulması ve bireylerin bu hizmetlere erişebilmelerinin sağlanması oldukça önemlidir. Bu durumda e-sağlık hizmetlerinden faydalanabilecek kitlenin heterojen bir yapıya sahip olduğunun göz ardı edilmemesi gerekmektedir. Nitekim, bu kitlenin demografik özelliklerinin (yaş, cinsiyet, eğitim durumu, vb.) yanı sıra teknoloji ve sağlığa ilişkin okuryazarlık düzeyleri de birbirinden farklılık gösterebilmektedir [bkz., 3, 4, 5, 6, 7]. Yine bireylerin teknoloji ve sağlıkla ilgili yetkinliklerini ifade eden dijital sağlık okuryazarlığı günümüzde önemle üzerinde durulan yetkinliklerden biri olsa da tüm bireylerin bu yeterliliklere sahip olması beklenemez. Bireylerin bu becerilere sahip olmaları sağlıkla ilgili verileri anlamlandırabilmeleri ve süreçlerini yönetebilmeleri için ise hayati önem taşımaktadır. Ancak, bu bir süreç olup yetkinliklerin kazandırılabilmesi için eğitimin yanı sıra zamana da ihtiyaç vardır. Nihayetinde, dijital dönüşümler halk ve nüfus sağlığına fayda sağlamak için muazzam bir potansiyele sahip olsa da mevcut eşitsizlikleri şiddetlendirme konusunda da eşit derecede yeteneklidirler [7]. Bu noktada e-sağlık uygulaması ya da hizmeti geliştiricilerinin tüm kullanıcıların dijital sağlık okuryazarı olmadığı gerçeğini göz önünde bulundurmaları ve bu yapıları tüm paydaşlar açısından kullanılabilir nitelikte tasarlamaları gerekmektedir. Özellikle de istenilen bilgi ve bağlantılara, kullanıcıların kolaylıkla erişebilecekleri gezinme unsurlarını barındıran kullanılabilirliği yüksek

ortamların tasarlanması, temelde sağlık sistemi ile vatandaşları bir araya getirmeyi amaçlayan bu yapıların sürdürülebilirliği açısından da oldukça önemlidir.

Gelişen teknoloji, bilgiyi daha ulaşılabilir hale getirmiş ve sürekli artan bilgi miktarının karmaşaya yol açmadan bireyin ihtiyaçlarına göre sunulmasını gerekli kılmıştır. Bu gerekliliğin yanı sıra bilginin, bireyin özellikleriyle birlikte zamanla değişen ihtiyaçlar doğrultusunda yeniden şekillenen yapılar oluşturmak için uyarlanabilir sistemlerin geliştirilme çalışmaları hız kazanmaya başlamıştır. Bu duruma paralel olarak geleneksel ortamlara yönelik “biri hepsine uyar (one size fits all)” görüşü terk edilmeye [8, 9, 10] ve “biri hepsine uymaz (one size does not fit all)” görüşü [11] yaygın olarak benimsenmeye başlamıştır [12]. Uyarlanabilir çevrimiçi sistemler, yapay zeka, veri madenciliği gibi tekniklerin sürece dahil edilmesi ile birlikte hızlı bir şekilde gelişmeye devam ederek farklı kullanıcı gereksinimlerine ya da kullanıcının zamanla değişen gereksinimlerine uygun olarak değiştirilebilen yapılar haline gelmiştir. Genel olarak kullanıcı tercihlerine ve ihtiyaçlarına göre şekillenen ortamlar sunmayı amaçlayan uyarlanabilir yapıların [10] ön plana çıkmasındaki en önemli unsurların başında kişiselleştirilmiş servisleri sunmak için davranışların gözden geçirilmesi, ilgi ve tercihlerin göz önünde bulundurulması tasarımların gerçekleştirilebilmesi gelmektedir [13]. Sistemdeki mevcut bilgilerin yanı sıra kullanıcının sistem içindeki anlık davranışları dinamik bilgilerin anlamlı bir şekilde yeniden yapılandırılması kritik öneme sahiptir. Bu yapılandırma bireyin sistem içinde gezinme stratejileri ile birlikte istenilen bilgiye erişiminde kilit rol oynamaktadır. Özellikle çevrimiçi ortamlarda kullanıcının istediği bilgiye ulaşmada en kısa yolu bulmasını destekleyecek tasarımların gerçekleştirilmesi önemli bir unsur olarak karşımıza çıkmaktadır. Uyarlanabilir yapılar içerisinde gezinmenin uyarlanması önemli bir basamak olup, kullanıcı modelinin yapılandırılması ve uyarlamaların gerçekleştirilmesi noktasında oldukça önemlidir. Gezinmenin uyarlanması, kullanıcının ortamda izleyeceği en uygun yolu bulması için destek sunulmasıdır [14, 15] ve sistemin sağlıklı işleyebilmesinde kilit rol oynamaktadır. Nitekim çevrimiçi ortamlarda iyi düzenlenmemiş gezinme stratejilerinin başta kaybolma ve aşırı bilişsel yüklenme olmak üzere birçok olumsuzluğa yol açabileceği görülmektedir [bkz., 16, 17, 18, 19]. Bu tür ortamlarda aşırı bilişsel yüklenen ve buna bağlı olarak da kaybolan kullanıcıların performansları düşerek istedikleri bilgilere ulaşmada daha fazla zaman ve çaba harcamaları gerekecektir [20]. Bu noktada aşırı bilişsel yüklenmeyi engelleyeceği belirlenen tasarım ilkelerinin kullanıldığı araştırma sonuçları ise kullanıcıların başarı, hatırlama ve transfer puanlarının arttığını göstermektedir [16].

Genel olarak kullanıcı özelliklerinin çevrimiçi ortamda gezinmelerinde etkili olduğu ve düşük ön bilgi düzeyine sahip bireylerin gezinme ve bağlantı yapılarını anlama konusunda daha çok sorun yaşamaya yatkın oldukları görülmektedir [bkz., 16, 19, 21, 22, 23]. Özellikle yeni kullanıcıların sistem içinde hangi bilgiye nasıl

ulaşabileceklerini kestirebilmeleri çok daha güçtür ve sistemin kullanıcıyı destekleyici nitelikte tasarlanması gerekmektedir. Bu durum hızla yaygınlaşan ve ilk kullanıcı sayısı hızla artmaya devam eden e-sağlık uygulamaları için de geçerlidir. Yaşamın birçok alanına neredeyse durdurulamaz bir şekilde giren dijitalleşme, politikacılar ve sağlık sigortacıları tarafından, genellikle gelecekteki sorunların çözümünün en azından bir parçası olarak görülmektedir [24]. Ancak, diğer taraftan da teknik altyapı, hızla artan veri hacmi ile birlikte sistemle etkileşim, istenilen verilere etkili bir biçimde erişimin sağlanması ve korunması gibi çözümlenmeyi bekleyen birçok sorun da beraberinde gelmektedir. Her ne kadar veri madenciliği ve yapay zeka tekniklerinden faydalanılarak makine temelli çözümler üretilmeye başlanmış olsa da özellikle son kullanıcıların kolaylıkla istenilen bilgiye erişim sağlayabilecekleri sistemlerin tasarlanması bir zorunluluktur. Yine bu doğrultuda kullanıcının sistem içinde gezinirken kaybolması ve aşırı bilişsel yüklenmesi gibi risklerin göz önünde bulundurularak iyi tasarlanmış uyarlanabilir gezinme yapılarının geliştirilmesi gerekmektedir. Bu çerçevede gerçekleştirilen çalışmada, Çetinkaya (2013) tarafından uyarlanabilir öğrenme ortamlarına yönelik geliştirilen tasarımı ilkelerinden [25] hareketle çevrimiçi e-sağlık uygulamalarının son kullanıcılarına yönelik uyarlanabilir gezinme ortamlarının düzenlenmesine ilişkin ilkelerin uzman görüşlerine dayalı olarak belirlenmesi amaçlanmıştır. Bu genel amaç doğrultusunda çalışmada;

- uzman görüşlerine göre Çetinkaya (2013) tarafından geliştirilen uyarlanabilir öğrenme ortamlarına yönelik tasarımı ilkeleri, çevrimiçi e-sağlık uygulamalarının son kullanıcılarına yönelik uyarlanabilir gezinme ortamlarının düzenlenmesinde de benimsenebilir mi?
- uzmanların söz konusu ilkelerin uygunluğuna katılım düzeyleri nedir? sorularına yanıt aranmıştır.

## 2. YÖNTEM (METHOD)

### 2.1. Araştırma modeli (Research model)

Çalışmada nitel ve nicel araştırma becerilerinin işe koşulduğu yapılandırılmış çevrimiçi Delphi tekniğinden faydalanılmıştır. Genel olarak bir grup uzmanın görüşü ile kontrollü geri bildirim sağlanarak, uzlaşma aranabilir bir anket elde etme süreci olarak tanımlanan Delphi tekniği [26] büyük ölçüde amaca hizmet edebilen, deneysel bilgi eksikliği nedeni ile sosyal süreçlerin anlaşılacağı durumlarda kullanılmaktadır [27].

### 2.2. Çalışma Gurubu (Study Group)

Delphi tekniği sürecinde katılımcıların, konular hakkında bilgi ve ilgi sahibi olan bireylerden oluşması gerektiğinden [28, 29], çalışma grubunun belirlenmesinde amaçlı örnekleme yaklaşımı benimsenmiştir. Amaçlı örnekleme yaklaşımında temel amaç, araştırmanın konusunu oluşturan kişi, olay ya da durum hakkında ve belirli bir amaç doğrultusunda derinlemesine bilgi toplamaktır [30]. Gerçekleştirilen bu çalışmada çevrimiçi Delphi uzman

grubu katılımcılarını belirlemek için amaçlı örnekleme kapsamında ölçüt örnekleme yoluna gidilmiştir. Bu çerçevede oluşturulan çalışma grubu doktora eğitimini tamamlamış olmak koşuluyla;

- uyarlanabilir ortamlar, yapay zeka ve öğretimsel uygulamalara yönelik araştırmaları olan (n=7),
- sağlık bilişimi, tıp bilişimi ve tıp eğitimi alanlarında ders vermiş ya da araştırmaları olan (n=8) uzman kişilerden oluşmuştur.

Delphi tekniğinin işe koşulduğu çalışmalarda katılımcı grubun uzmanlık ölçütlerine göre homojenliğinin söz konusu olduğu araştırmalarda 10-15 kişinin yeterli olduğu bildirilmiştir [26]. Bu doğrultuda belirtilen ölçütleri kapsayacak biçimde oluşturulan çalışma grubunu çevrimiçi ortamda ulaşılabilen 15 uzman oluşturmuş ve bu uzmanlar çalışmanın tüm aşamalarında katılım sağlamıştır. Çevrimiçi ortamda gerçekleştirilen Delphi çalışmasının güvenilirliğini sağlamak için veriler anlık olarak kayıt altına alınmıştır.

### 2.2. Verilerin Toplanması (Data Collection)

Yapılandırılmış çevrimiçi Delphi yönteminin benimsendiği çalışmada metodüçi karışım (Intramethod mixing) yaklaşımı işe koşulmuştur. Bu yaklaşım tek bir soru kağıdında eşzamanlı olarak veri toplamak için kullanılan bir karma yöntem stratejisidir [31]. Bu strateji doğrultusunda tekrarlanan turlar içeren yapılandırılmış çevrimiçi Delphi uygulama süreci aşamaları aşağıda belirtilmiştir.

1. İlkelere yönelik görüş belirlenmesi amacıyla önceden belirlenen 22 uzmana çalışmanın amacını belirten bir metin ile birlikte davet mektubu gönderilmiştir. Davete olumlu yanıt veren 15 uzman ile süreç başlatılmıştır.
2. Çevrimiçi öğrenme ortamlarının tasarımında gezinmenin uyarlanması sürecinde Çetinkaya (2013) tarafından geliştirilen ilkeler [25] üzerinde değişiklik yapılmadan çevrimiçi ortama aktarılmış ve uzmanlara gönderilmiştir. İlkelere ilişkin açıklama ve yönergelerin de yer aldığı çevrimiçi anket formu, her bir ilkenin altına ilkeye yönelik uzmanların varsa görüş ve önerilerini belirtmeleri amacıyla serbest metin alanı oluşturularak düzenlenmiştir. Ayrıca, ölçme aracının en sonunda eklemek istedikleri yeni bir ilke önerisi ya da ilkelerin geneline ilişkin varsa görüş ve önerilerini paylaşabilecekleri serbest metin alanı oluşturulmuştur.
3. Uzmanların Çetinkaya (2013) tarafından geliştirilen gezinmenin uyarlanmasına ilişkin ilkeleri [25], e-sağlık uygulamaları çerçevesinde irdeleyerek yaptıkları değerlendirme sonucunda gelen veriler araştırmacılar tarafından incelenmiştir. Uzmanların yorumları ile birlikte ilkelere yönelik değerlendirmeleri gözden geçirilip düzenlenmiştir. Düzenleme sırasında öğrenme ortamlarına yönelik geliştirilen ilkelerin e-sağlık uygulamalarına uyumu ve uygunluğu göz önünde bulundurulmuştur. Bu aşamada ilkelerin uygunluğunun belirlenmesi amacıyla 5'li (Tamamen uygun, Uygun, Kısmen uygun, Uygun değil, Hiç uygun değil) likert tipi çevrimiçi değerlendirme formu oluşturulmuştur.

Form uzmanların her fikri gözden geçirmesi, güçlü ve zayıf yanlarını belirtmesi ve yeni fikirlerini ekleyebilmelerine olanak sağlayabilecek şekilde düzenlenmiştir. Kolaylıkla kullanılabilir ve uzmanların her fikri serbestçe paylaşabilecekleri şekilde tasarlanan ikinci anket formu açıklayıcı yönergelerle birlikte uzmanlara gönderilmiştir.

4. Gerçekleştirilen ikinci Delphi turu sonucunda uzmanlardan gelen değerlendirme sonuçları araştırmacılar tarafından incelenmiştir. Yapılan incelemeler sonucunda ilkelerin ifade ediliş biçimine yönelik uzmanlar tarafından küçük düzeltme önerileri geldiği belirlenmiştir. İlkelerin uygunluğuna yönelik derecelendirme verileri doğrultusunda ise uzmanların uzlaşıya ulaşma dereceleri hesaplanmış ve bir sonraki anket formu hazırlanmaya başlanmıştır. Üçüncü Delphi turu için hazırlanan anket formu; uzmanların ikinci Delphi turundaki önerileri doğrultusunda yeniden düzenlenen ilkelerin yanı sıra tüm uzmanlara ait yanıtların ortalama, standart sapma değerleri ile birlikte uzmanın maddeye kendi verdiği puanında bulunduğu bir formatta yeniden düzenlenerek gönderilmiştir.
5. Uzmanlardan gelen üçüncü Delphi turu verilerinin incelenmesi sonucunda artık yeni fikirlerin üretilmediği, maddelerin tüm güçlü-zayıf yanlarının belirlendiği ve uzmanlardan ilkelere düzenlenmesine yönelik herhangi bir görüş ya da önerinin gelmediği görülmüştür. Uzmanların tüm maddeler üzerinde uzlaşıya vardıkları belirlenen bu turdan sonra uzmanlara katılımları için teşekkür edilerek en son veriler doğrultusunda oluşturulan ilkeler yönergeleri ile birlikte gönderilmiştir.

### 2.3. Verilerin Analiz Edilmesi (Analysis of the Data)

Her turunda verilerin toplandığı Delphi çalışmasının ilk turunda, Çetinkaya (2013) tarafından gezinmenin uyarlanmasına yönelik geliştirilen ilkeler [25] gönderilmiştir. Ayrıca, bu turda gönderilen ilkelere yönelik ek görüş belirtebilecekleri bir serbest yazım alanını içeren veri formu düzenlenerek nitel veriler elde edilmiştir. İkinci ve üçüncü turda ise, hem nitel hem de nicel veriler elde edilirken sürecin son turu olan dördüncü turu sadece nicel verilerden oluşmuştur. Uzmanlardan gelen görüş ve önerilerden oluşan nitel veriler araştırmacılar tarafından incelenerek ilkeler üzerinde gerekli düzenlemeler yeniden uzman görüşüne sunulmak üzere gerçekleştirilmiştir. Uzmanların uzlaşıya ulaşma derecelerinin belirlenmesi için ise; aritmetik ortalama, standart sapma ve ilk çeyrek kayma değerlerinden faydalanılmıştır. Her bir ilkeye ilişkin uzmanların uzlaşma düzeyleri tablo 1’de belirlendiği gibi dört kategoriden oluşmuş [32] ve her etki altında standart sapma ile ortalamaya göre belirlenmiştir.

Tablo 1. Kategori Ölçeklendirme Matrisi

Ortalama Değeri	Standart Sapma Değeri	
	Düşük	Yüksek
	I	II
Yüksek	Büyük Öne Sahip Yüksek Düzeyde Anlaşma	Yüksek Öne Sahip Düşük Düzeyde Anlaşma
	III	IV
Düşük	Alt Öne Sahip	Alt Öne Sahip

	Yüksek Düzeyde Anlaşma	Düşük Düzeyde Anlaşma
--	------------------------	-----------------------

Delphi yönteminin kullanıldığı çalışmalar incelendiğinde aritmetik ortalama, standart sapma ve medyan (ortanca) değerlerinin uzlaşma derecelerinin belirlenmesinde işe koşulduğu görülmektedir. Ancak, kategorik sınıflandırmalarda bu değerlerin aralıklarına yönelik belirli bir standart oluşturulmamıştır. Bu nedenle her bir ölçütün birlikte ele alınarak uzlaşma kategorilerinin oluşturulduğu ve Çetinkaya (2013) tarafından düzenlenen değer aralıkları [25] uzmanların uzlaşıya ulaşma derecelerinin belirlenmesinde işe koşulmuştur (bkz. tablo 2).

Tablo 2. Uzlaşma Aralıkları

$\bar{X}$	Ss	$Q_1$	Yorum	
1.00-1.49	$Ss \geq .75$	$3 \leq Q_1$	Önemsiz uzlaşma olması	1
1.50-2.99			Orta derecede uzlaşma olması	2
3.00-3.74			Ekseriyetle uzlaşma olması	3
3.75-4.49	$.50 \leq Ss < .75$	$3 < Q_1 \leq 4$	Önemli derecede uzlaşma olması	4
4.50-5.00	$Ss < .50$	$4 < Q_1 \leq 5$	Neredeyse mükemmel uzlaşma olması	5

Delphi süreci sonunda belirlenen ilkelerin tamamına yönelik değerleyici güvenilirliğini belirlemek amacı ile ise Sınıf-içi Korelasyon Katsayısı (Intraclass Correlation Coefficient-ICC) hesaplanmıştır. Değerleyici güvenilirliği, değerleyiciler arasındaki uyuma derecesini ölçmek amacıyla işe koşulan Sınıf-içi Korelasyon Katsayısı [33] ile ilgili güvenilirlik aralığı tablo 3’de verilmiştir [34].

Tablo 3. Sınıf-içi Korelasyon Katsayısı Güvenirlik Aralığı

ICC	Yorum
$< .69$	Düşük düzeyde güvenilirlik
$0.70 - 0.79$	Yeterli/Orta düzeyde güvenilirlik
$0.80 - 0.89$	Yüksek düzeyde güvenilirlik
$0.90 - 1.00$	Mükemmel düzeyde güvenilirlik

### 3. BULGULAR VE YORUMLAR (FINDINGS AND INTERPRETATIONS)

Gezinmenin uyarlanması, kullanıcının ortamda izleyebileceği en uygun yolu bulması için destek sunulmasıdır [14, 15]. Çetinkaya (2013) tarafından gerçekleştirilen çalışmada gezinmenin uyarlanmasına yönelik tüm ilkeler [25] üzerinde bağlı önemlilik ölçütlerine göre uzmanların mükemmel uzlaşma sağladığı sonucuna ulaşılmıştır. Yine araştırmacı bulguların sağlamlığı değerlendirme ölçütleri çerçevesinde alanyazma dayalı olarak elde edilen veriler doğrultusunda tüm ilkelerin güçlü araştırma desteğine sahip olduğunu belirlemiştir. Bu noktada alanyazından elde edilen bulgular ile birlikte uzman görüşlerine dayalı olarak ilkelerin geçerliliğinin sağlandığı ve farklı sistemler üzerinde de değerlendirilmesinin faydalı olabileceği belirtilmiştir. Bu bulgulardan hareketle gerçekleştirilen çalışmada hızla yaygınlaşan ve neredeyse tüm toplumun sağlığa ilişkin ihtiyaçlarını karşılamayı hedefleyen e-sağlık uygulamalarının tasarımında da uyarlanabilir gezinme ilkelerinin uyumunun değerlendirilmesi amaçlanmıştır. Bu doğrultuda elde edilen ilkelerle birlikte her bir ilkenin

açıklayıcısı biçiminde düzenlenen yönergeler Tablo 4'te sunulmuştur.

Tablo 4. E-Sağlık Uygulamalarında Gezinmenin Uyarlanmasına Yönelik İlkeler

Gezinmenin Uyarlanması (GU.)		*GU	Bağlı** Önem
<b>GU.1</b>	Kullanıcı modelindeki bilgiler doğrultusunda, bir sonraki bağlantıya ulaşmasını sağlayın.	5/5	5
Kullanıcı modelindeki verilere dayalı olarak kullanıcıların bir sonraki bağlantı ya da içeriğe ulaşmasını sağlanması gerekmektedir. Bu doğrultuda kullanıcılara uyarlanabilir bağlantı yapıları sunularak aşırı bilişsel yüklenme ve kaybolma problemleriyle karşılaşmalarını önlemek için gezinme unsurlarında gerekli düzenlemelerin yapılması gerekmektedir. Bu kapsamda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.93$ , $Ss=0.26$ , $Q1=5.00$ ).			
<b>GU.2</b>	İlgili bağlantılara açıklama/yönerge ekleyin.	5/5	5
Gezinme sırasında bağlantı açıklamaları kullanıcıların sistem içinde kaybolmalarını ve boğulmalarını engelleyerek gezinmelerini destekleme potansiyeli taşımaktadır. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.79$ , $Ss=0.59$ , $Q1=5.00$ ).			
<b>GU.3</b>	Kullanıcı modelindeki bilgiler doğrultusunda dinamik bağlantı yapıları sağlayın.	5/5	5
Kullanıcı modelindeki bilgiler doğrultusunda uygun dinamik bağlantı yapılarının sunulması ve bu doğrultuda ilgililik durumlarına göre bağlantıların gösterilmesi gerekmektedir. Böylece kullanıcılara dinamik bağlantı yapıları sunularak ortamda kaybolmalarını engellenebileceği gibi istenilen bilgi ya da bağlantıya ulaşım kolaylaştırılabilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.80$ , $Ss=0.41$ , $Q1=5.00$ ).			
<b>GU.4</b>	Kullanıcının uygulamanın neresinde olduğunu görmesini sağlamak için kişiye özel geribildirim sağlayın.	5/5	5
Ortamda gezinme sürecinde kullanıcıların mevcut durumları bilgi almaları, buldukları konuma nasıl geldikleri ve bir sonraki ilgili bağlantıya görebilmeleri; kaybolmaları ve aşırı bilişsel yüklenmelerinin engellenmesinde olumlu katkı sağlayabilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.93$ , $Ss=0.26$ , $Q1=5.00$ ).			
<b>GU.5</b>	Kullanıcılara yönelik uyarlanabilir içerik haritası oluşturun.	5/5	5
Kullanıcıların ortamın içeriğini oluşturan unsurların genel yapısını görmelerini sağlayan içerik haritalarının sunulması gezinmeye olumlu katkı sağlayabilir. Bu durum tüm içeriği ve ilgili bağlantıların bir noktada görülmesine olanak sağlayarak sistem içerisinde istenilen içerik ya da bağlantıya kolaylıkla ulaşabilmelerini destekleyebilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.78$ , $Ss=0.36$ , $Q1=5.00$ ).			
<b>GU.6</b>	Kullanıcıların daha önce girmiş oldukları bağlantı, içerik ve sayfalara ilişkin istatistiksel bilgileri görmelerine izin verin.	5/5	4
Kullanıcıların gezinme sürecindeki hareketlerine ilişkin istatistiksel verileri görmeleri, gezinmelerine olumlu katkı sağlayabilir. Böylelikle kullanıcının ortamda yaptığı işlem ve girilen bağlantıların görülebilir olması, geçmişe ait sürecini nicel olarak ve hızlı bir biçimde gözden geçirmesine olanak tanır. Bu durum ortam içerisinde kullanıcıların aşırı			

bilişsel yüklenmesi ve kaybolmaları gibi problemlerin de önüne geçilmesine katkı sağlayabilir.

Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, önemli derecede uzlaşma sağladığı belirlenmiştir ( $X=4.33$ ,  $Ss=0.72$ ,  $Q1=4.00$ ).

<b>GU.7</b>	Kullanıcı modelindeki bilgiler göz önünde bulundurularak işlevsel arama sonuçları sağlayın.	5/5	5
Kullanıcının gereksinim duyduğu bilgiye kolaylıkla ulaşabilecekleri bir sistemin geliştirilmesi, ortam içerisinde gezinmelerine ve istenilen içeriğe ya da bağlantıya kolay ulaşabilmelerine olumlu katkı sağlayacaktır. Arama motorlarının tasarımında uyarlanabilir sistem özelliklerinin kullanılması kişiye özel en ilgili içerik ya da bağlantının sunulmasına olanak sağlayabilir. Bu noktada uyarlanabilir yapıların işlevsel sonuç çıkarma mekanizmalarının ortama entegre edilmesi istenilen kullanıcıların bilgi ya da bağlantıya tek bir yol ile ulaşarak sistem içerisinde kaybolmaları riski azaltılabilir. Yine böylelikle kullanıcının gereksiz içerik ya da bağlantılarla karşılaşmaları engellenerek sistem içinde boğulmalarının da önüne geçilebilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.87$ , $Ss=0.35$ , $Q1=5.00$ ).			

<b>GU.8</b>	Kullanıcının arama performansını geliştirmek için kişiye özel ipuçları sağlayın.	5/5	5
Arama motorları aracılığı ile kullanıcının gereksinim duyduğu bilgi ya da bağlantıya kolaylıkla ulaşabilecekleri bir sistemin geliştirilmesi, ortamda gezinme sürecine olumlu katkı sağlayacaktır. Ancak, arama motorları kullanılarak en uygun içerik ya da bağlantının bulunabilmesi noktasında kullanıcılara doğru arama stratejileri yapabilmeleri için ipuçları sağlanmalıdır. Uyarlanabilir sonuç çıkarma mekanizmalarından kullanıcıların etkili bir biçimde faydalanmaları sağlanmalı ve bu konuda kişiye özel destekleyici ipuçları verilmelidir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.80$ , $Ss=0.41$ , $Q1=5.00$ ).			

<b>GU.9</b>	Kullanıcı modeline bağlı kalınarak kullanıcının içerik tercihlerine göre, yeni eklenen içeriği kolaylıkla takip edebilmesini sağlayın.	5/5	5
Kullanıcıların ilgili içeriğe ya da bağlantıya kolaylıkla erişebilmesi için uygun gezinme araçlarının işe koşulması gerekmektedir. Özellikle e-sağlık uygulamaları ile beraberinde verilen hizmetlerin çeşitlenerek arttığı göz önünde bulundurulduğunda, kullanıcının yeni içerik ya da hizmetleri kolaylıkla görebilmeleri sağlanmalıdır. Bunu sağlayabilecek araçlardan biri olan RSS (Rich Site Summary) yapıları aracılığı ile yeni eklenen içerikler kolaylıkla takip edilebilir. Ayrıca, uyarlanabilir sistemler ile entegre çalışabilen bu yapılar aracılığıyla kullanıcı özelliklerine göre uyarlanabilen ve dinamik olarak güncellenebilen içerik haritaları da sunulabilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ( $X=4.87$ , $Ss=0.35$ , $Q1=5.00$ ).			

<b>GU.10</b>	Kullanıcının önemli gördüğü alanları (bağlantı, metin, konu, sayfa, vb.) işaretlemelerine izin verin.	5/5	5
Kullanıcının kendine göre önemli gördüğü konu, sayfa ya da bölümleri işaretleyebilmelerine olanak tanınması gezinmelerine olumlu katkı sağlayacaktır. Özellikle gezinme sırasında kolaylıkla işaretleme ya da not almalarına olanak sağlayan sistemlerin geliştirilmesi kullanıcının ortama bir sonraki girişinde geçmişte yapmış olduğu işlemler ile birlikte süreç içinde önemli gördükleri unsurların hatırlanmasına da olanak sağlanmış olacaktır. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse "mükemmel uzlaşma" sağladığı belirlenmiştir ( $X=4.93$ , $Ss=0.26$ , $Q1=5.00$ ).			

\*GU: Bağlı önemlilik/Bulguların Sağlamlığı: Gezinmenin uyarlanması [25]

\*\*Bağlı Önem: Gezinmenin uyarlanması ilkelerinin e-sağlık uygulamalarına uyumu

Elde edilen sonuçlar Çetinkaya (2013) tarafından geliştirilen modelin uyarlamaların gerçekleştirilmesi başlığı altında yer alan uyarlanabilir çevrimiçi öğrenme ortamlarında gezinmenin uyarlanmasına yönelik 10 ilke üzerinde elde edilen uzlaşının [25], e-sağlık uygulamalarına yönelik geliştirilen çevrimiçi ortamlar için de sağlandığını göstermektedir. Bu noktada belirlenen 10 ilkenin 9'u üzerinde bağlı önemlilik ölçütlerine göre tüm uzmanların uzlaşmaya vardıkları ve "neredeyse mükemmel uzlaşma (5)" sağlandığı belirlenmiştir. Belirlenen ilkelerden 1'in de ise uzmanların, "önemli derecede uzlaşma (4)" sağladığı belirlenmiştir.

Tablo 5. Delphi Uygulaması Anket Turları Uzlaşma Dereceleri

	Delphi Uygulaması Anket Turları		
	1. Tur	2. Tur	3. Tur
Uzmanlar arasındaki uzlaşma (ICC)	0.799	0.892	0.930
Ortalama değerleri	4.42	4.62	4.79

Gezinmenin uyarlanmasına yönelik ilkelerin e-sağlık uygulamalarına ilişkin uyumu noktasında değerlendiriciler arasındaki uyuma derecesini belirlemek amacı ile hesaplanan sınıf-içi korelasyon katsayısı (ICC= .930), ilkelere ilişkin uzmanlar arasında yüksek düzeyde güvenilirlik sağlandığını göstermektedir. Veriler üzerinde uzlaşının sağlandığını gösteren ve verilerin uyumluluğunu analiz etmek için kullanılan sınıf-içi korelasyon katsayısı değerleri tablo 6'da sunulmuştur.

Tablo 6. İlkeler Üzerinde Delphi Panelistleri Arasındaki Uyuşma

	Sınıf-içi Korelasyon	Güven Aralığı (%95)		F	p
		Alt Sınır	Üst Sınır		
Örneklem Ortalaması	0.930	0.861	0.972	14.237	.000*

Bağıl önemlilik değerlendirme sonuçları, e-sağlık uygulamalarında gezinmenin uyarlanmasına ilişkin belirlenen ilkelere yönelik alan uzmanlarının ortak eğilimlerini göstermektedir.

#### 4. TARTIŞMA, SONUÇ VE ÖNERİLER (CONCLUSION, RESULT AND RECOMMENDATIONS)

Yapılan analizler sonucunda elde edilen sonuçlar alanyazınla birlikte tartışılmış ve bu doğrultuda geliştirilen öneriler başlıklar altında sunulmuştur.

##### 4.1. Tartışma ve Sonuçlar (Conclusion and Results)

Dijital sağlık hizmetlerinin, sağlık profesyonelleri ve vatandaşlar açısından faydalarından bahsetmek elbette mümkün, ancak bu hizmetlerin aynı zamanda kullanılabilirlik açısından da ele alınması oldukça önemlidir. Nitekim hızla artan veri miktarı ile birlikte sistemi kullanan paydaşların istenilen bilgiye kolay erişiminin sağlanması kullanılabilirliği yüksek e-sağlık uygulamalarının geliştirilmesiyle mümkün olabilir. Bu

noktada farklı sistemleri bir araya getirebilecek ve uyum içerisinde çalışabilmelerine olanak sağlayan işlevsel yapıların oluşturulması gerekmektedir. Uyarlanabilir sistemler bu bağlamda önemli bir alternatif olarak karşımıza çıkmaktadır. Uyarlanabilir sistemlerin geliştirilmesinin disiplinler arası araştırma yaklaşımı gerektiriyor olması; çoğu zaman tasarım, metod ve değerlendirmeye uyum sağlayamayan ortamların geliştirilmesine sebep olabilmektedir. Geliştirilen farklı sistemlerin birlikte işlerlik kazanabilmesi noktasında ortak bir dilin oluşturulması ve ortak bir paydada uzlaşılması sürdürülebilir verimli yapıların oluşturulması açısından oldukça önemlidir. Bu durum göz önünde bulundurularak gerçekleştirilen çalışmada gezinmenin uyarlanmasına yönelik önceden geliştirilen ilkelerin e-sağlık uygulamalarına uyumu ve düzenlenmesi hedeflenmiştir. Bu doğrultuda nitel ve nicel araştırma becerilerinin işe koşulduğu yapılandırılmış çevrimiçi Delphi tekniğinden faydalanılan çalışmaya 15 uzman dahil edilmiştir. Metodüçi karışım yaklaşımının benimsendiği veri toplama sürecinde tek bir soru kağıdında nitel ve nicel veriler eşzamanlı olarak toplanarak veri çeşitlemesi yapılmıştır. Uzmanlardan elde edilen veriler ve bu verilerin değerlendirilmesi ile elde edilen sonuçların uzmanlara sunumuyla tekrarlanan bir sürecin izlendiği çalışma, uzmanların uzlaşmaya ulaşmalarıyla son bulmuştur.

Çalışmanın ilk turu sonucunda uzmanların gezinmenin uyarlanmasına yönelik ilkeleri yapısal olarak uygun olduğu belirlenmiştir. Ancak, çevrimiçi öğrenme ortamlarına yönelik ilkelerin e-sağlık uygulamaları çerçevesinde yeniden yorumlanması gerekliliği uzmanlar tarafından bildirilmiştir. Uzman görüşleri doğrultusunda her turda yeniden şekillenen ilkeler üzerinde artık yeni fikirlerin üretilmediği, maddelerin tüm güçlü-zayıf yanlarının belirlendiği ve uzmanlardan ilkelerin düzenlenmesine yönelik herhangi bir görüş ya da önerinin gelmediği görülen üçüncü Delphi turunda çalışma sonuçlandırılmıştır. Uzman görüşleri çerçevesinde uyarlanabilir yapılarla gezinmenin uyarlanmasına yönelik geliştirilen 10 ilke üzerinde tüm uzmanların uzlaşmalarıyla e-sağlık uygulamalarına yönelik uyumu sağlanmıştır (bkz., Tablo 4). Diğer taraftan belirlenen ilkelere ilişkin değerlendiriciler arası uyuma derecesini ölçmek amacı ile kullanılan sınıf-içi korelasyon katsayısı (ICC= .930), mükemmel düzeyde güvenilirlik sağlandığını göstermektedir. Bağıl önemlilik ölçütlerine göre uzlaşma sağlanan ilkelerin tamamı gezinmenin uyarlanmasıyla ilişkili olmakla birlikte sistemin diğer bileşenlerinin işleyişiyle de doğrudan ilişkilidir. Bu noktada gezinmenin uyarlanmasına yönelik tasarımın sistemin sağlıklı bir şekilde işleyebilmesi için kritik öneme sahip olduğunu söylemek mümkün gözükmemektedir. Özellikle bilgi yoğunluğu ve etkileşimin yüksek düzeyde olduğu çevrimiçi ortamların tasarımında kullanıcıların kendi istekleri ve ihtiyaçlarına uygun şekilde sistem içinde gezinmelerine olanak sağlayabilecek şekilde düzenlenmesi gerekmektedir. Aksi takdirde aşırı bilişsel yüklenme ve kaybolma gibi birbirleriyle ilişkili sorunların ortaya çıkmasına neden olunabilir. Özellikle çok fazla bileşen, yol ve araca sahip ortamlardaki seçenekler nedeniyle,

kullanıcıların sistem üzerinde kaybolarak kafalarının karışması sonucu istenilen bilgilere ulaşmada daha fazla zaman ve çaba harcamaları söz konusu olacaktır [18, 20]. Bu durum özellikle sayıları gün geçtikçe hızla artan e-sağlık uygulamaları için de söz konusu olmakla birlikte gezinmeye yönelik unsurlara dikkat edilmemesi durumunda risklerin çeşitlenerek artabileceği de söylenebilir.

Dijital dönüşüm bireylerin hizmetlere erişim ve eşitliğin sağlanmasında muazzam bir potansiyele sahip olsa da tüm bireylerin bu hizmetlerden faydalanabilmelerine olanak sağlayan ortamların geliştirilmesi gerekmektedir. Özellikle de sağlık sisteminin hızla dijitalleşmesi ile birlikte sağlığa erişimde e-sağlık uygulamalarının kullanımı kritik öneme sahip olmaya başlamıştır. Bu noktada e-sağlık uygulamalarının yeni kullanıcıların istenilen bilgiye hızlı ve doğru bir şekilde ulaşmalarının desteklenmesi gerekmektedir. Bunun için ise uyarlanabilir gezinme yapılarıyla birlikte kullanıcıların doğru bir şekilde yönlendirilmeleri sağlanarak sistem içinde kaybolmaları ve boğulmaları engellenmelidir. Özellikle bilginin yapılandırılması ve anlamın şekillenmesinde doğru yolun izlenerek en kısa sürede istenilen içeriğe ulaşılmasının oldukça önemli olduğu göz önünde bulundurulduğunda gezinmeye ilişkin unsurların dinamik olarak uyarlanabilir bir yapıyla tasarlanması kritik rol oynamaktadır. Nitekim bireylerin bilgiye ulaşma stratejileri farklılık gösterebilmekte ve bu durum bilginin edinimi sürecinde bireysel farklılıkların gözetildiği uyarlanabilir ortamların geliştirilmesini zorunlu hale getirmektedir. Bunun yapılabilmesi için ise kullanıcının sistemle ve sistem ile ilişkili tüm bileşenlerle etkileşim sürecinin iyi tasarlanması gerekmektedir. Bir başka ifadeyle kullanıcının sistem üzerindeki tüm hareketleri gözlenerek bir sonraki adımlarında sistem üzerinde kaybolmalarını engelleyecek nitelikte uyarlanabilir gezinme yapıları sunulmalıdır. Aksi takdirde sağlık gibi kritik bir temel hakka erişimde eşitsizliklerin artmasına yönelik risklerin yanı sıra dijital sağlık hizmetlerinin sürdürülebilirliğinin sağlanması noktasında da sorunlarla karşılaşılması söz konusu olabilir.

#### 4.2. Öneriler (Recommendations)

Araştırmanın amacı doğrultusunda elde edilen sonuçlara dayalı olarak geliştirilen öneriler aşağıda sunulmuştur.

1. Çalışma sonucunda gerek alanyazın gerekse alan uzmanlarının ortak eğilimleri ile elde edilen ilkelerin, e-sağlık uygulamalarında gezinme unsurlarının tasarımında işe koşulmasının faydalı olacağı söylenebilir.
2. Uyarlanabilir gezinmeye ilişkin ilkeler çevrimiçi ortamlara yönelik olarak geliştirilmiş olup gerçekleştirilen bu çalışmada ise e-sağlık uygulamalarına uyumu değerlendirilmiştir. Bu noktada çevrimiçi olmayan ya da farklı cihaz türlerinde (tablet, akıllı telefon, vb.) belirlenen ilkelerin geçerliliğine bakılabilir.

3. Hızla yaygınlaşan mobil e-sağlık uygulamaları için de benzer bir çalışmanın planlanması çok daha doğru bir yaklaşım olacaktır.
4. Çalışmada belirlenen ilkeler alanyazın doğrultusunda uzman görüşlerine dayalı olarak belirlenmiştir. Bu belirlenen ilkelerin deneysel etkililiğinin test edilmesine yönelik çalışmaların da yapılması faydalı olacaktır.
5. E-sağlık hizmetlerine yönelik tasarımların, çalışmada belirlenen ilkelere uygunluğunu değerlendirmeye yönelik çalışmalar yapılmalıdır.

#### KAYNAKLAR (REFERENCES)

- [1] P. Kostkova, "Grand challenges in digital health", *Frontiers in Public Health*, 3(134), 1-5, 2015.
- [2] U. H. Bittlingmayer, K. Dadaczynski, D. Sahrai, S. v. Broucke, O. Okan, "Digitale gesundheitskompetenz – konzeptionelle verortung, erfassung und förderung mit fokus auf kinder und jugendliche", *Bundesgesundheitsblatt- Gesundheitsforschung- Gesundheitsschutz*, 63, 176-184, 2020.
- [3] A. Scheerder, A. van Deursen, J. van Dijk, "Determinants of Internet skills, use and outcomes: A systematic review of the second-and third-level digital divide", *Telematics and informatics*, 34(8), 1607-1624, 2017.
- [4] K. Sørensen, J. M. Pelikan, F. Röthlin, K. Ganahl, Z. Slonska, G. Doyle, J. Fullam, B. Kondilis, D. Agrafiotis, E. Uiters, M. Falcon, M. Mensing, K. Tchamov, S. van den Broucke, H. Brand, HLS-EU Consortium, "Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU)", *European journal of public health*, 25(6), 1053-1058, 2015.
- [5] A. Odone, S. Buttigieg, W. Ricciardi, N. Azzopardi-Muscat, A. Staines, "Public health digitalization in Europe", *European journal of public health*, 29(Supplement\_3), 28-35, 2019.
- [6] I. Kickbusch, D. Piselli, A. Agrawal, R. Balicer, O. Banner, M. Adelhardt, E. Capobianco, C. Fabian, A. Singh Gill, D. Lupton, R. P. Medhora, N. Ndili, A. Ryś, N. Sambuli, D. Settle, S. Swaminathan, J. V. Morales, M. Wolpert, A. W. Wyckoff, L. Xue, ... Secretariat of the Lancet and Financial Times Commission, "The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world", *Lancet (London, England)*, 398(10312), 1727-1776, 2021.
- [7] R. van Kessel, R. Hrzic, E. O'Nuallain, E. Weir, B.L.H. Wong, M. Anderson, S. Baron-Cohen, E. Mossialos, "Digital Health Paradox: International Policy Perspectives to Address Increased Health Inequalities for People Living With Disabilities", *J Med Internet Res*. 24(2):e33819, 2022.
- [8] P. Brusilovsky, "Adaptive and intelligent technologies for web-based education", *Künstliche Intelligenz*, 13, 19-25, 1999.
- [9] P. Brusilovsky, "Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation", *British Journal of Educational Technology*, 34(4), 487-497, 2003.
- [10] P. Brusilovsky, C. Peylo, "Adaptive and intelligent web-based educational systems", *International Journal of Artificial Intelligence in Education*, 13(2), 156-169, 2003.

- [11] C. M. Reigeluth, "A new paradigm of ISD?" *Educational Technology & Society*, 36(3), 13-20, 1996.
- [12] S. Somyürek, **Uyarlanabilir Eğitsel Web Ortamlarının Öğrencilerin Akademik Başarısına ve Gezinmesine Etkisi** (Yayınlanmamış Doktora tezi), Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara, 2008.
- [13] C.-M. Chen, "Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance", *Expert Systems with Applications*, 36, 8816–8829, 2009.
- [14] P. Brusilovsky, & L. Pesin, "An intelligent learning environment for CDS/ISIS users", In J.J Levonen and M.T Tukianinen. (Eds.), **Proceedings of the interdisciplinary workshop on complex learning in computer environments (CLCE94)** (29-33). Joensuu, Finland, May 16-19, 1994.
- [15] P. De Bra, "Adaptive Hypermedia on the Web: Methods, techniques and applications, **Proceedings of the AACE WebNet'98** (220-225), Orlando, Florida, USA, November 7-12, 1998.
- [16] E. K. Çakmak, "Çoklu Ortamlarda Dar Boğaz: Aşırı Bilişsel Yükleme", *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 27(2), 1-24, 2007.
- [17] A. E. Kılıç, Ö. G. Ş. Karadeniz, "Hiper Ortamlarda Öğrencilerin Bilişsel Yükleme ve Kaybolma Düzeylerinin Belirlenmesi" *Kuram ve Uygulamada Eğitim Yönetimi*, 40(40), 562-579, 2004.
- [18] T. Murray, "Characteristics and affordances of adaptive hyperbooks", **Proceedings of WebNet 2001**, Orlando, FL, USA, October 23-27, 2001.
- [19] J. Q. Young, J. V. Merrienboer, S. Durning, O. Ten Cate, "Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86", *Medical Teacher*, 36(5), 371-384, 2014.
- [20] S. McDonald, & R. J. Stevenson, "Disorientation in hypertext: the effects of three text structures on navigating performance", *Applied Ergonomics*, 27(1), 61-68, 1996.
- [21] E. Cuddihy, K. A. Mobernd, J. H. Spyridakis, "Web page previews: effect on comprehension, user perceptions, and site exploration", *Journal of Information Science*, 38(2), 103–117, 2012.
- [22] K. A. Lawless, P. G. Schrader, H. J. Mayall, "Acquisition of information online: Knowledge, navigation and learning outcomes", *Journal of Literacy Research*, 39(3), 289–306, 2007.
- [23] L. Salmerón, W. Kintsch, J. J. Cañas, "Reading strategies and prior knowledge in learning from hypertext", *Memory & Cognition*, 34(5), 1157–1171, 2006.
- [24] A. Hansen, M. Herrmann, J. P. Ehlers, T. Mondritzki, K. O. Hensel, H. Truebel, P. Boehme, "Perception of the Progressing Digitization and Transformation of the German Health Care System Among Experts and the Public: Mixed Methods Study", *JMIR public health and surveillance*, 5(4), e14689, 2019.
- [25] L. Çetinkaya, **Uyarlanabilir eğitsel içerikli web ortamlarının tasarım ilkeleri** (Yayınlanmamış doktora tezi), Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara, 2013.
- [26] A. L. Delbecq, A. H. Van de Ven, D. H. Gustafson, **Group techniques for program planning: A guide to nominal group and Delphi processes**, Scott, Foresman & Company, 1975.
- [27] M. Fry, G. Burr, "Using the Delphi technique to design a self reporting triage survey tool", *Accident and Emergency Nursing*, 9(4), 235- 241, 2001.
- [28] W. Alexander, R. Serfass, **Futuring tools for strategic quality planning in education**, Milwaukee, WI: American Society for Quality Press, 1999.
- [29] T. Hatcher, S. Colton, "Using the internet to improve HRD research: The case of the web-based Delphi research technique to achieve content validity of an HRD-oriented measurement", *Journal of European Industrial Training*, 31(7), 570-587, 2007.
- [30] J. A. Maxwell, **Qualitative research design: An interpretative approach**, Thousand Oaks, CA: Sage, 1996.
- [31] A. Tashakkori, C. B. Teddlie, **Handbook of mixed methods in social and behavioral research**, Thousand Oaks, CA: Sage, 2003.
- [32] H. A. Shah, & T. S. Tillman, "An international delphi study to build a foundation for an undergraduate level lean manufacturing curriculum", *International Transaction Journal of Engineering, Management & Applied Sciences & Technologies*, 9(3), 521-530, 2011.
- [33] K. L. Gwet, **Handbook of inter-rater reliability**, Gaithersburg: StatAxis Publishing, 2001.
- [34] J. W. Youdas, J. R. Carey, & T. R. Garrett, "Reliability of measurements of cervical spine range of motion—comparison of three methods", *Physical Therapy*, 71(2). 98-104, 1991.

# Developing Novel Deep Learning Models to Detect Insider Threats and Comparing the Models from Different Perspectives

*Araştırma Makalesi/Research Article*

 Yasin Görmez\*<sup>1</sup>,  Halil Arslan<sup>2</sup>,  Yunus Emre Işık<sup>1</sup>,  Veysel Gündüz<sup>3</sup>

<sup>1</sup>Management Information System, Sivas Cumhuriyet University, Sivas, Türkiye

<sup>2</sup>Computer Engineering, Sivas Cumhuriyet University, Sivas, Türkiye <sup>3</sup>Basis Department, Detaysoft, İstanbul, Türkiye

[yasiingormez@cumhuriyet.edu.tr](mailto:yasiingormez@cumhuriyet.edu.tr), [harslan@cumhuriyet.edu.tr](mailto:harslan@cumhuriyet.edu.tr), [yeisik@cumhuriyet.edu.tr](mailto:yeisik@cumhuriyet.edu.tr), [veysel.gunduz@detaysoft.com](mailto:veysel.gunduz@detaysoft.com)

(Geliş/Received:06.11.2023; Kabul/Accepted:16.01.2024)

DOI: 10.17671/gazibtd.1386734

**Abstract**— Cybersecurity has become an increasingly vital concern for numerous institutions, organizations, and governments. Many studies have been carried out to prevent external attacks, but there are not enough studies to detect insider malicious actions. Given the damage inflicted by attacks from internal threats on corporate reputations and financial situations, the absence of work in this field is considered a significant disadvantage. In this study, several deep learning models using fully connected layer, convolutional neural network and long short-term memory were developed for user and entity behavior analysis. The hyper-parameters of the models were optimized using Bayesian optimization techniques. Experiments analysis were performed using the version 4.2 of Computer Emergency and Response Team Dataset. Two types of features, which are personal information and numerical features, were extracted with respect to daily activities of users. Dataset was divided with respect to user or role and experiment results showed that user based models have better performance than the role based models. In addition to this, the models that developed using long short-term memory were more accurate than the others. Accuracy, detection rate, f1-score, false discovery rate and negative predictive value were used as metrics to compare model performance fairly with state-of-the-art models. According the results of these metrics, our model obtained better scores than the state-of-the-art models and the performance improvements were statistically significant according to the two-tailed Z test. The study is anticipated to significantly contribute to the literature, as the deep learning approaches developed within its scope have not been previously employed in internal threat detection. Moreover, these approaches have demonstrated superior performance compared to previous studies.

**Keywords**— user and entity behavior analysis, machine learning, deep learning, insider threat, cyber security

## İç Tehditlerin Tespit Edilmesi için Özgün Derin Öğrenme Modellerinin Geliştirilmesi ve Modellerin Farklı Perspektiflerde Karşılaştırılması

**Özet**— Siber güvenlik, çok sayıda kurum, kuruluş ve devlet için zamanla hayati öneme sahip bir konu haline gelmiştir. Mevcut çalışmalar incelendiğinde, dış saldırıları önlemek için birçok çalışma yapıldığı, ancak iç tehditleri tespit etmeye yönelik çalışmaların yeterli olmadığı kanısına varılmaktadır. İç tehditlerden gelen saldırıların kurum itibarlarına ve mali durumlarına verdiği zararlarda göz önüne alındığında, bu alanda çalışma eksikliği büyük bir dezavantaj olarak değerlendirilmektedir. Bu çalışmada, kullanıcı ve varlık davranış analizi için tam bağlı katman, evrimsel sinir ağı ve uzun kısa süreli hafıza kullanan çeşitli özgün derin öğrenme modelleri geliştirilmiştir. Modellerin hiper parametreleri Bayesian optimizasyon teknikleri kullanılarak optimize edilerek, analizler, Computer Emergency and Response Team Dataset veri kümesinin 4.2. sürümü kullanılarak yapılmıştır. Kullanıcıların günlük aktivitelerine göre kişisel bilgiler ve sayısal özellikler olmak üzere iki tür özellik çıkarılmıştır. Veri seti kullanıcı veya role göre bölünmüş ve deney sonuçlarına kullanıcı tabanlı modellerin rol tabanlı modellere göre daha iyi performansa sahip olduğunu gözlemlenmiştir. Ayrıca uzun kısa süreli hafızayı kullanarak geliştirilen modellerin diğerlerine göre daha başarılı sonuçlar elde ettiği gözlemlenmiştir. Model performansını literatürdeki çalışmalar ile adil bir şekilde karşılaştırmak için, başarı oranı, tespit oranı, f1 puanı, yanlış keşif oranı ve negatif tahmin değeri metrikleri kullanılmıştır. Bu metriklerin sonuçlarına göre modelimiz, literatürde var olan modellere göre daha iyi performans skorları elde etmiş ve iki kuyruklu Z testine göre performans iyileştirmeleri istatistiksel olarak anlamlı bulunmuştur. Çalışma kapsamında geliştirilmiş olan derin öğrenme yaklaşımlarının daha önce iç tehdit tespitinde kullanılmamış olmasından ve önceki çalışmaların performanslarından yüksek bir performans elde etmesinden dolayı çalışmanın literatüre büyük bir katkı sağlayacağı kanaatine varılmıştır.

**Anahtar Kelimeler**— kullanıcı ve varlık davranış analizi, makine öğrenmesi, derin öğrenme, iç tehdit, siber güvenlik



## 1. INTRODUCTION

The concept of cyber security has become more and more audible, with technology taking place in more areas of our lives. Information is more important for institutions now and the institutions store their data on different servers. Although systems designed to access data from different locations provide convenience for institutions, they cause data security to become more complex. Some of these data such as employee, marketing strategy, application info and project documents have crucial importance for institutions. In this regard, information security practices and institutions' strategy should be compatible and information security applications must secure all systems of institutions [1]. If one of the institution's information system is vulnerable to cyber-attacks, all systems are at risk to attacks. Companies can experience financial losses or loss of reputation because of the information theft. Although most organizations use basic security precautions, the number of security incidents is increasing. Therefore, more resources, which are human, hardware and software, should be allocated for security systems.

Information technology systems of institutions are very resistant to external attacks such as DDOS, malware, phishing and password hacking thanks to advance applications. According to the research, more than 60% of businesses are using technical information security prevention practices such as antivirus program, firewall, intrusion detection system, virtual private networks and anti-spyware software. In addition to this, attacks from inside and outside are increasing [2]. Although these applications are very successful to external attacks, they are not yet sufficient to detect internal attacks. According to a study conducted by the Ponemon Institute in 2017 with 237 companies from 6 countries, it was concluded that insider threats are the most expensive cyber-attack situations and it is foreseen that insider threats will increase in the future [3]. Insider threats are typically perpetrated by existing employees or authorized individuals. In large organizations, user activities and network traffic become highly intricate. The human factor plays a more significant role in the occurrence of internal threats than external threats. However, not every person working in large organizations possesses the same level of experience in information security. Considering these complexities, detecting internal threats emerges as a formidable challenge for institutions.

Especially large companies use security software that collects logs and event data generated by all users, servers, network devices and firewalls to monitor and analyze all security-related events in their infrastructure. This system is called as security information and event management (SIEM). It is possible to make user and entity behavior analysis (UEBA), which is one of the most common approach to detect insider threads, thanks to data collected by SIEM systems.

UEBA aims to determine malicious actions that coming from the insiders. These actions can be done by outsiders

who have impersonated employees or employees who acted maliciously or negligently. Instead of detecting malicious software or antivirus, it is desired to detect anomalies in the behavior of users with UEBA. This situation causes the problem to be difficult because what is an anomaly for one user may not be an anomaly for another user. For example, frequent review of employee information may not be an anomaly for someone who working in the human resources department, but it may be considered as an anomaly for those working in other departments. For this reason, UEBA systems are also developed separately for situations such as department, job description, user or role in the literature [4]. The data used within the scope of UEBA is important for the business processes of the companies and requires confidentiality. Therefore, it is very difficult to find data for academic studies. The Computer Emergency and Response Team Dataset (CERT), which generated synthetically thanks to support of Carnegie Mellon University, is frequently encountered in studies in UEBA field [5], [6].

To date, supervised or unsupervised machine learning algorithms have been used frequently for UEBA. Xiangyu et al. combined one class support machines, recurrent neural networks and isolation forest for UEBA and they reached 91.60%, 93.10% and 100% accuracy, precision and recall respectively [7]. Tuor et al. proposed long short-term memory and deep neural networks based unsupervised deep learning model for UEBA. The proposed model was trained on version 6.2 of CERT dataset and anomaly score for each sample was computed. Experiment results show that, proposed model outperformed the isolation forest, support vector machines and principal component analysis models [8]. Lin et al. applied proposed hybrid machine learning model based on deep belief network and one-class support vector machines on version 4.2 of CERT dataset. They firstly extracted hidden features using deep belief network, subsequently samples were classified using support vector machine model. As a result of experiment analysis, they obtained 87.70%, 81.04% and 12.18% accuracy, detection rate and false positive rate respectively [9]. Yuan et al. extracted user behavior features with user actions and abstracted temporal features using long short-term memory network networks. Afterwards, they converted extracted features to fixed size vector and they obtained 94.49% area under the ROC curve score by training convolutional neural network model [10]. Lo et al. detected the insider threads using distance measurements that were Damerau-Levenshtein, Jaccard and Cosine Distance, and they obtained 39%, 36% and 47% detection rate. They summarized that distance measurement techniques were better than the hidden markov model because they were faster to train and detect [11]. Li and Zincir-Heywood applied both supervised and unsupervised learning models on CERT dataset using numerical and sequential data. As a result of the experiments, they reached 79.75% and 73.10% detection rate with self-organizing maps and C4.5 decision tree respectively for weekly data; 84.60% and 99.87% detection rate with self-organizing maps and C4.5 decision tree respectively for daily data [12]. Igbe and Saadawi

applied artificial immune system algorithm on CERT dataset with 21 extracted features and they reached 83.45%, 4.60% and 89.00% true positive rates, false positive rates and accuracy respectively [13]. Hall et al. applied neural networks, support vector machines, naïve bayes, decision tree, random forest and logistic regression on version 4.2 of CERT dataset. According to the analysis results they obtained 95.8%, 91.3%, 95.4%, 97.5%, 96.1% and 96.5% accuracy for neural networks, naïve bayes, support vector machines, random forest, decision tree and logistic regression respectively and they increased accuracy to 95.8%, 97.2%, 97.2%, 97.2%, 97.2% and 96.8% by applying boosting on these algorithm [14]. Aldairi et al. achieved 93% accuracy and 92% precision on version 4.2 of CERT dataset using isolation forest and one-class support vector machine [15]. Le and Zincir-Heywood extracted features from version 4.2 and 5.2 of CERT dataset using autoencoder, principle component analysis and random projection and they applied logistic regression, artificial neural networks, naïve Bayes and random forest as a classifier. As a result of experiment analysis, they achieved 98.33% accuracy for version 4.2 and 99.04% accuracy for version 5.2 [16]. In other study, Le and Zincir-Heywood achieved 92.93%, 99.54% and 97.13% accuracies using logistic regression, random forest and artificial neural networks respectively with user-session based feature extraction techniques [17]. Nasser et al. obtained 92% accuracy using gated recurrent neural network architecture [18]. Sharma et al. proposed long short-term memory based autoencoder model for user behavior analytics and they achieved 90.17%, 91.03% and 9.84% accuracy, true positive rate and false positive rate respectively [19]. Tian et al. used several deep learning layers, which were long short-term memory, convolutional long short-term memory and multi-layer perceptron, to develop multi-model based system for insider threat detection [20]. Al-Shehari and Alsowail used synthetic minority oversampling technique to overcome imbalance problem of CERT dataset and they obtained 0.79, 1.00, 1.00, 0.84 and 0.99 area under curve score using logistic regression, decision tree, random forest, naïve Bayes and k nearest neighborhood respectively [21]. Nasir et al. achieved 90.60% accuracy, 97% precision and 94% F1 score by using long short-term memory based autoencoder model [22]. Su et al. utilized online version of recurrent neural networks for the version 6.2 of CERT dataset and they achieved 95.3% f-measure [23]. Dosh et al. achieved 94.68% accuracy using three traditional machine learning algorithms [4]. 96% accuracy was obtained by Pantelidis et al. using autoencoder and variational autoencoder based deep learning model [24]. Al-Mhiqani et al. achieved 97% accuracy, 2.88% false positive rate with proposed multilayer framework model [25]. AlSlaiman et al. et al. applied sentiment analysis techniques on HTTP traces, emails, and files of CERT dataset and they combined it with different data representation. For this purpose, they proposed a long short-term memory based deep learning model and they achieved 0.29% false positive rate, 2.27% false negative rate and 97% area under the curve [26]. Li et al. integrated memory-augmented network into autoencoder and they obtained 94.56% area under the curve score [27].

When reviewing the literature, it becomes evident that numerous studies have been conducted on UEBA, resulting in the development of various deep learning models. When examining studies focused on detecting abnormal behavior through the analysis of individual user patterns, it becomes apparent that some studies have achieved an accuracy level of approximately 90% [22]. However, in studies with a success rate of around 98%, the detection rate remained at 67% [16]. In this context, it was concluded that enhancements should be made to the deep learning models used for classifying insider threats. Therefore, our study aimed to develop deep learning models that achieve high accuracy and detection rates while maintaining a low false discovery rate.

It is known that cyber-security is one of the main issue in many area [28]. In this paper, malicious actions from insiders were predicted using CERT datasets. For this purpose, preprocessing steps were applied on CERT dataset and samples were divided into parts based on roles or users. After preprocessing steps, features were extracted and several deep learning models were developed. User behavior is time-dependent, and a malicious action is a combination of a user's activities over time. Therefore, Convolutional neural networks (CNN), long short-term memory (LSTM) and fully connected layer (FLC), which were suitable for time series problem, were used to develop deep learning models. The hyper-parameters are one of the most important factors for the performance of deep learning models. Because of this reason, Bayesian optimization technique, which is faster and more effective than traditional techniques, were used in this study to optimize hyper-parameters [29]–[32]. Bayesian optimization can explore larger spaces. Moreover, unlike grid search, which tests specific values, Bayesian optimization allows for trying any value within the specified range. Consequently, this method can identify more suitable values for hyper-parameters. The performance of the proposed models was measured using five different metrics that are accuracy, detection rate, f1-score, negative predictive value and false discovery rate.

The primary aim of the study is to predict internal threats with higher accuracy compared to the existing literature, owing to the novel approaches and models developed within this research. First innovative aspect of our work is the deep learning approaches used for UEBA were developed by us specifically for this study. In addition to this, to the best of our knowledge, it is first to use Bayesian optimization techniques with deep learning approaches for UEBA on CERT datasets.

The paper is organized as follows: Section II explained the benchmark datasets, feature extraction steps and architectures of proposed models. Section III presents the experimental results, encompassing hyper-parameter optimization, the performance metrics of the trained models, a comparison with state-of-the-art methods, and a two-tailed Z-test. Finally, the paper is concluded and future works were explained in the section IV.

## 2. MATERIALS AND METHODS

### 2.1. Benchmark Dataset

In this study, the Computer Emergency and Response Team (CERT) dataset, which is insider threat benchmark datasets and generated synthetically, were used to train the proposed models and to assess the performance of the proposed model [6]. This dataset consists of device connection, sent or received email, transferred or copied file, http and login activities from 1000 users for 500 days. In addition to this, CERT dataset also contains LDAP files and detailed personal information for each user. Data imbalance has a detrimental impact on model performance in machine learning studies [33]. Ideal conditions for high-performance models involve having a balanced distribution of samples from different classes in the training dataset. In this study, version 4.2 of the CERT dataset was utilized to mitigate the effect of data imbalance on model performance, as it contains the largest number of samples related to malicious actions. Many actions in these datasets are standard actions and also the Computer Emergency and Response Team determined three scenarios to define malicious action. Explanation of these scenarios were given below. Note that, these scenarios were taken directly from the dataset information package [5].

- *“User who did not previously use removable drives or work after hours begins logging in after hours, using a removable drive, and uploading data to wikileaks.org. Leaves the organization shortly thereafter [5].”*
- *“User begins surfing job websites and soliciting employment from a competitor. Before leaving the company, they use a thumb drive (at markedly higher rates than their previous activity) to steal data [5].”*
- *“System administrator becomes disgruntled. Downloads a keylogger and uses a thumb drive to transfer it to his supervisor's machine. The next day, he uses the collected keylogs to log in as his supervisor and send out an alarming mass email, causing panic in the organization. He leaves the organization immediately [5].”*

In other versions of the dataset, these scenarios may change. If an employee has a malicious action, these actions are specified separately in each scenario. The details about this dataset can be found on CERT information package [5].

### 2.2. Feature Extraction

As mentioned earlier, the CERT datasets consisting of raw information from 5 different activities and users' personal information. In order to make UEBA, this data must be combined into a single file in a way that the machine

learning model can understand. Because of this reason, some preprocessing steps were applied on these files then features were extracted. Firstly, each activity type in separate files was combined on a per-user basis and sorted by date. After that, based on one-day activities for all users, features were extracted. In this study a samples represent the daily activities of users. Considering the daily activities, two types of features, which are personal information and numerical features, were extracted. Personal information consists of user role, user functional unit, user team and “o”, “c”, “e”, “a” and “n” results of OCEAN test. The details about OCEAN test can be found at the test information page [34]. The numerical features were computed considering the number of daily activities with respect to activity types. In this context, 24 values, which are listed below, were computed for each sample.

- `n_of_logOnLogOf_work_on_own_pc`: Number of logon-logoff in work on time with user own personal computer.
- `n_of_logOnLogOf_work_off_own_pc`: Number of logon-logoff in work off time with user own personal computer.
- `n_of_logOnLogOf_work_on_other_pc`: Number of logon-logoff in work on time with other user's personal computer.
- `n_of_logOnLogOf_work_off_other_pc`: Number of logon-logoff in work off time with other user's personal computer.
- `n_of_email_work_on_own_pc`: Number of email in work on time with user own personal computer.
- `n_of_email_work_off_own_pc`: Number of email in work off time with user own personal computer.
- `n_of_email_work_on_other_pc`: Number of email in work on time with other user's personal computer.
- `n_of_email_work_off_other_pc`: Number of email in work off time with other user's personal computer.
- `n_of_device_work_on_own_pc`: Number of device in work on time with user own personal computer.
- `n_of_device_work_off_own_pc`: Number of device in work off time with user own personal computer.
- `n_of_device_work_on_other_pc`: Number of device in work on time with other user's personal computer.
- `n_of_device_work_off_other_pc`: Number of device in work off time with other user's personal computer.
- `n_of_file_work_on_own_pc_toMedia`: Number of file copy to temporary device from company's

- computer in work on time with user own personal computer.
- *n\_of\_file\_work\_off\_own\_pc\_toMedia*: Number of file copy to temporary device from company's computer in work off time with user own personal computer.
  - *n\_of\_file\_work\_on\_other\_pc\_toMedia*: Number of file copy to temporary device from company's computer in work on time with other user's personal computer
  - *n\_of\_file\_work\_off\_other\_pc\_toMedia*: Number of file copy to temporary device from company's computer in work off time with other user's personal computer.
  - *n\_of\_file\_work\_on\_own\_pc\_fromMedia*: Number of file transfer to company's computer from temporary device in work on time with user own personal computer.
  - *n\_of\_file\_work\_off\_own\_pc\_fromMedia*: Number of file transfer to company's computer from temporary device in work off time with user own personal computer.
  - *n\_of\_file\_work\_on\_other\_pc\_fromMedia*: Number of file transfer to company's computer from temporary device in work on time with other user's personal computer.
  - *n\_of\_file\_work\_off\_other\_pc\_fromMedia*: Number of file transfer to company's computer from temporary device in work off time with other user's personal computer.
  - *n\_of\_http\_work\_on\_own\_pc*: Number of website activity in work on time with user own personal computer.
  - *n\_of\_http\_work\_off\_own\_pc*: Number of website activity in work off time with user own personal computer.
  - *n\_of\_http\_work\_on\_other\_pc*: Number of website activity in work on time with other user's personal computer.
  - *n\_of\_http\_work\_off\_other\_pc*: Number of website activity in work off time with other user's personal computer

In feature extraction stage, numerical features were calculated considering two important situations: activity time and computer used for activity. It is thought that the increase in the use of computers of other employees and the activities performed in work off time are two important factors in determining the insider malicious actions. In addition to this, the information of whether data is received from the media or data is transferred to the media was also used to compute numerical features for file activities. Considering all these situations, four numerical features were computed for logon, email, device and http activities

separately and eight numerical feature were computed for file activities. Therefore, a total of 24 numerical features were computed using five different activity records.

After feature extraction steps, each sample was labeled as non-malicious (0) or malicious (1) using the answers file of CERT dataset. The answers file consists of malicious event from 5 different activities according to 3 scenarios of CERT datasets. It is assumed that, if any sample contains one of these malicious activities, this sample labeled with "1". Otherwise the sample was labeled with "0". As a result of preprocessing and feature extraction steps, 330452 samples, which consist of 986 malicious and 329466 non-malicious, and 29 features were obtained from 1000 users. Comprehensive details regarding dataset preprocessing and feature extraction can be found in the study proposed by Görmez et al. [35]. The code for these processes is available on the Detaysoft GitHub page [36].

### 2.3. Proposed Models

In this study, two different approaches, which were role based and user based, were used to develop proposed deep learning models. In these approaches, small datasets were generated by dividing the dataset according to role type or user. In the role based models, dataset divided into 42 parts and in the user based models, dataset divided into 1000 parts. In these approaches, models consisting of input layer and hidden layers for each section were generated separately, then combined in a single output layer to create the main model. There are two reasons to use these approach. There are cases where an action that is malicious for one person or group is not malicious action by another person or group. For example, a behavior that is normal for a person working in the IT department may be malicious for the human resources department. Therefore, training the UEBA models separately based on user or role affects the accuracy of model positively. Considering this situation, the input and hidden layers of our models were created user or role based separately. In this case, in order to prevent overfitting situations that may occur in the model and to generalize the model, the models created separately on the basis of role or person were combined in a single output layer. A total six model were proposed using these approaches. One of them was used only for role based, three of them were used only for user based and remaining were used both role and user based dataset. The architectures of these models are shown in figures below (Figure 1-5). The models can be summarized as traditional artificial neural network (MLP), convolutional neural network model (CNN), long short-term memory model in deep two (LSTM\_2), long short-term memory model in deep four (LSTM\_4), combination of convolutional neural network and long short-term memory in deep two (CNN\_LSTM\_2) and combination of convolutional neural network and long short-term memory in deep four (CNN\_LSTM\_4).

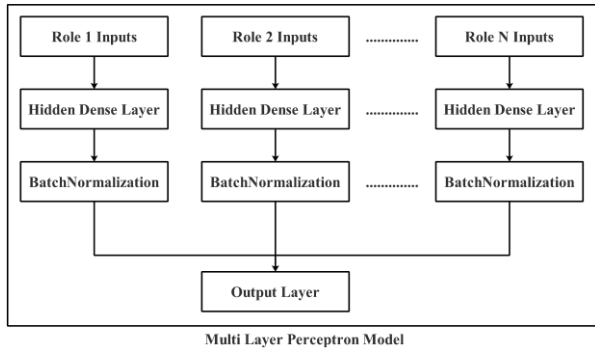


Figure 1. Architecture of role based multi-layer perceptron model

MLP model, which was used only for role based approach, was shown in the figure 1. This model has contained 42 input layer, because there were 42 different roles in the CERT dataset. A fully connected and Batch Normalization layers were added after each input layer. In fully connected layer, relu was used as an activation function and number of neurons were optimized. Note that, all hidden layers are identical which means that the number of neurons were same in each fully connected layer. All of this layers were concatenated at the output layer with softmax function to generate generalized classification layer.

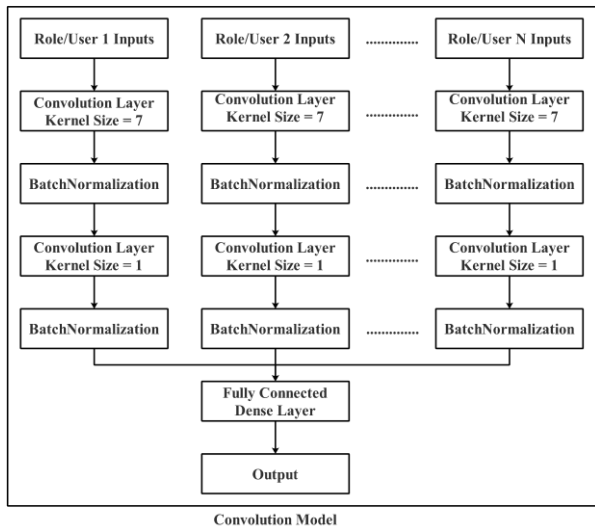


Figure 2. Architecture of role and user based convolution model

Convolution model, which were used for both role or user based approaches, was shown in the figure 2. In this model, 42 input layers were used for role based approach and 1000 input layers were used for user based approach. Note that, two different models were generated for user and role based approaches separately. In this model, each input layer was followed by convolutional layer with kernel size 7, Batch Normalization layer, convolutional layer with kernel size 1 and Batch Normalization layer in sequential order. After that, all of these layers for each user or role were concatenated. Later, followed by the shared fully connected and classification layer and the model was completed. Note that, similar to MLP model, layers created for each user or role are identical. Hyper-parameter

optimization was applied user based and role based models separately.

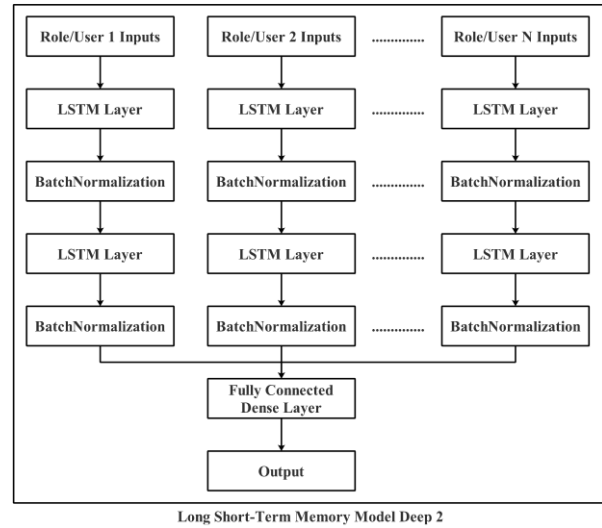


Figure 3. Architecture of role and user based long short-term memory model at depth 2

Long short-term memory model in deep 2, which were used for both role or user based approaches, was shown in the figure 3. Similar to convolution model, also in this model, 42 input layers were used for role based approach, 1000 input layers were used for user based approach and two different models were generated for user and role based approaches separately. In this model, each input layer was followed by one directional long short-term memory layer, Batch Normalization layer, one directional long short-term memory layer and Batch Normalization layer in sequential order. After that, all of these layers for each user or role were concatenated. Later, followed by the shared fully connected and classification layer and the model was completed. Because there was a two long short-term memory layer, this model was named as LSTM\_2. Similar to other models, layers created for each user or role are identical and hyper-parameter optimization was applied user based and role based models separately.

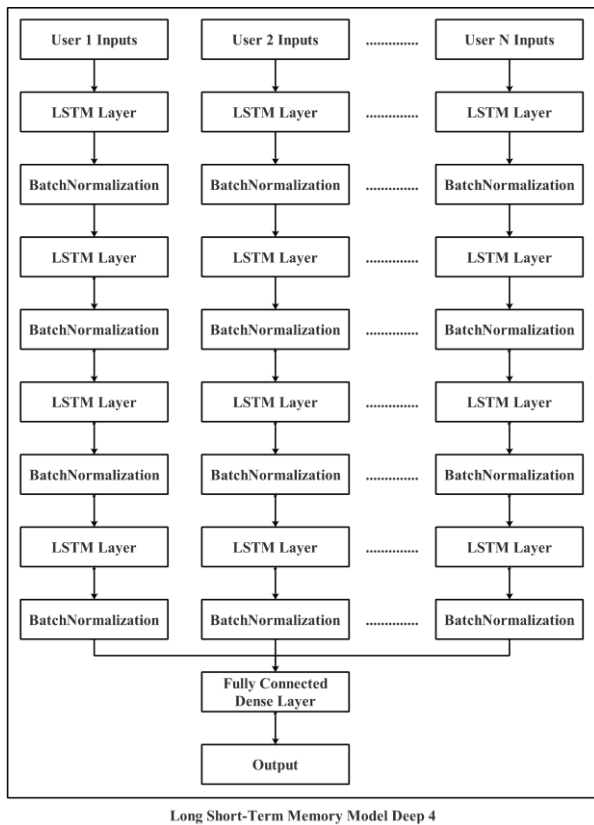


Figure 4. Architecture of user based long short-term memory model at depth 2

Long short-term memory model in deep 4 was shown in the figure 4. This model has contained 1000 input layer, because it was only used for user based approach. The layers of this model was exactly same with the LSTM\_2 model. Because this model contained four one directional long short-term memory layer, only difference from LSTM\_2 model was the depth of the model. For this reason, the model was named as LSTM\_4. This model was used in this study to measure the effect of depth on the performance of model. The depth was only measure for LSTM model, because the initial results showed that the best approach was user based LSTM model for UEBA. Similar to other models, layers created for each user are identical.

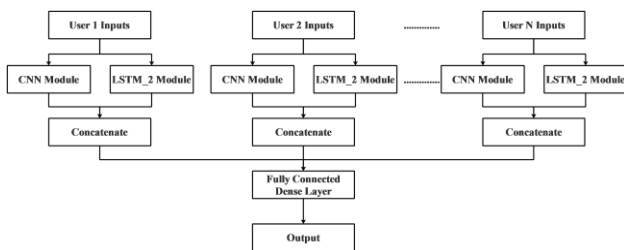


Figure 5. Architecture of user based convolution and long short-term memory model at depth 2

Combination of convolution and long short-term memory model in deep 2, which were used for only user based approach, was shown in figure the figure 5. In this model, input layers, which were generated for each user

separately, were connected in parallel to convolution and LSTM modules. The convolution module was exactly same with the CNN model and the LSTM module was exactly same with the LSTM\_2 model. At the final stage, modules generated for each user separately were concatenated and they followed by fully connected and classification layer. Similar to other models, layers created for each user are identical. This model was called as CNN\_LSTM\_2, because LSTM\_2 was used to generate LSTM module. Combination of convolution and long short-term memory model in deep 4, which were used for only user based approach were also developed. This model was similar to CNN\_LSTM\_2 model where only difference was LSTM\_4 model was used in order to LSTM\_2 model for LSTM modules. Similar to other models, layers created for each user are identical. This model was called as CNN\_LSTM\_4, because LSTM\_4 was used to generate LSTM module.

These models were developed using keras deep learning library of python [37]. In these models, hidden dense layer represents the traditional artificial neural networks. In the convolutional neural network models, two layers, which have different kernel sizes as seven and one, were connected each other serially. Note that, the convolution layers in these models consisted of 1D convolution operation. The horizontal width of the kernel was 29 that is equal to the total number of features. The convolution process was performed between activities performed by the same person on previous days. In the LSTM models return\_sequences parameter was set to True. Similar to convolution process, the LSTM layer also revealed the time series relationship by looking at the previous day activities of the same person. In all layers, relu was used as activation function, glorot\_uniform with seed 1 was used as weight initializers, softmax was used as activation function at classification layer, Adam was used as optimizer and categorical\_crossentropy was used as loss function. During the experiment, the layer specific hyper-parameters of model were optimized using Bayesian optimization techniques. The details about the hyper-parameter optimization process was given in the experiment results section

### 3. EXPERIMENT RESULTS

In the first phase of the experiment, four datasets, which are training, testing, training for optimization (trainingOpt) and testing for optimization (testingOpt), were generated by dividing the main dataset into parts with respect to user and sample date. Firstly, 30% percent of the main dataset (last samples by date for each user) selected to generate testing dataset and training dataset was generated using the rest of main dataset. After this phase, 25% percent of the training dataset (similar to testing dataset, these are the last samples by date of training dataset for each user) selected to generate testingOpt dataset, trainingOpt dataset was generated using the rest of training dataset. TrainingOpt and testingOpt datasets were used for hyper-parameter optimization phase and training dataset was used to train model with optimum hyper-parameter and testing dataset

was used to assess performance of trained model with optimum hyper-parameter.

### 3.1. Hyper-Parameter Optimization

Hyper-parameters are one of the most important factor that affect the performance of the deep learning models. Because number of hyper-parameters are high and hyper-parameter space are large, traditional hyper-parameter optimization techniques are not effective for deep learning models [29]–[32]. Because of these reasons, in this study, hyper-parameters of the models were optimized using Bayesian optimization techniques. This method uses a Gaussian process with two aims: Modeling the surrogate function and optimizing the expected probability which is based on improving the existing best solutions through new trials. It assumes that the function values track a multivariate Gaussian distribution. A Gaussian kernel designates the covariance of the function values among the parameters. In each iteration, the next value of a parameter is selected through the acquisition function over the Gaussian prior.

In this study, Bayesian Optimization technique was implemented using skopt library of python language [37]. From this library, gp\_minimize function was used with hyper-parameter spaces using following parameters settings: n\_cals = 250 and acq\_func = “EI”. This library operates as a unique parameter space type. The model was optimized according to the mean of true negative rate and accuracy. In this context, according to the hyper-parameter type, three different variable types, which were Real, Categorical and Integer, were used. In these types, Real and Integer needs a low and high value and Categorical needs a hyper-parameter space as in grid search technique. For this purpose, these spaces were generated by determining the hyper-parameters to be optimized separately for each model and hyper-parameters of each model were optimized using trainingOpt and testingOpt datasets. Hyper-parameter value type, space, and optimum value were shown in table 1 for each hyper-parameter of role based model and in table 2 for user based model. In this step, learning rate of the model (lr), number of epoch (epoch), batch size (batch), number of hidden unit in dense layer (n\_unit\_dense), number of kernel in convolutional layer (n\_unit\_conv\_1) and dimension of output space of LSTM layers (n\_unit\_lstm\_1-4) were optimized, if model need any of them.

Table 1. Hyper-parameters value type, space and optimum value for role based models

Model Name	Parameter Name	Value Type	Parameter Space	Optimum Value
MLP	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.059841909
	epoch	Integer	low = 10, high = 1500	1458
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	1024
	n_unit_dense	Integer	low = 20, high = 5000	4076
CNN	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.0000004771
	epoch	Integer	low = 10, high = 1500	758
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	2048
	n_unit_dense	Integer	low = 20, high = 5000	77
	n_unit_conv_1	Integer	low = 20, high = 150	101
LSTM_2	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.0000022047
	epoch	Integer	low = 10, high = 1500	780
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	512
	n_unit_dense	Integer	low = 20, high = 5000	2767
	n_unit_lstm_1	Integer	low = 10, high = 100	34
	n_unit_lstm_2	Integer	low = 10, high = 100	28

As evident from the table data, an extensive search was conducted for hyper-parameters, particularly real values. The optimum values obtained could be any value within the specified range, unlike grid search. This is notably

observed in the lr hyper-parameter. Upon examining the table data, it is noted that the epoch and batch parameters generally have high values. The lr parameter varies from model to model, but it is mostly close to the lower limit.

Table 2. Hyper-parameters value type, space and optimum value for user based models

Model Name	Parameter Name	Value Type	Parameter Space	Optimum Value
CNN	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.0000005749
	epoch	Integer	low = 10, high = 1500	1301
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	16

	n_unit_dense	Integer	low = 20, high = 5000	602
	n_unit_conv_1	Integer	low = 20, high = 150	68
LSTM_2	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.00000000143
	epoch	Integer	low = 10, high = 1500	167
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	8
	n_unit_dense	Integer	low = 20, high = 5000	1482
	n_unit_lstm_1	Integer	low = 10, high = 100	51
	n_unit_lstm_2	Integer	low = 10, high = 100	37
	LSTM_4	lr	Real	low = $10^{-10}$ , high = $10^{-1}$
epoch		Integer	low = 10, high = 1500	399
batch		Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	4
n_unit_dense		Integer	low = 20, high = 5000	639
n_unit_lstm_1		Integer	low = 10, high = 100	10
n_unit_lstm_2		Integer	low = 10, high = 100	25
n_unit_lstm_3		Integer	low = 10, high = 100	29
n_unit_lstm_4		Integer	low = 10, high = 100	18
CNN_LSTM_2	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.0001284
	epoch	Integer	low = 10, high = 1500	1065
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	2048
	n_unit_dense	Integer	low = 20, high = 5000	617
	n_unit_conv_1	Integer	low = 20, high = 150	148
	n_unit_lstm_1	Integer	low = 10, high = 100	28
	n_unit_lstm_2	Integer	low = 10, high = 100	57
CNN_LSTM_4	lr	Real	low = $10^{-10}$ , high = $10^{-1}$	0.00000002271
	epoch	Integer	low = 10, high = 1500	1500
	batch	Categorical	{2,4,8,16,32,64,128,256,512,1024,2048}	1024
	n_unit_dense	Integer	low = 20, high = 5000	20
	n_unit_conv_1	Integer	low = 20, high = 150	34
	n_unit_lstm_1	Integer	low = 10, high = 100	16
	n_unit_lstm_2	Integer	low = 10, high = 100	10
	n_unit_lstm_3	Integer	low = 10, high = 100	14
	n_unit_lstm_4	Integer	low = 10, high = 100	26

Based on the results observed in the table, unlike role-based models, hyper-parameters in user-based models may exhibit more variation from model to model. Therefore, it cannot be interpreted as being close to the lower or upper limit for any hyper-parameter value. Particularly when examining hyper-parameter distributions with real values such as lr, the benefits of the Bayesian optimization method become evident once again.

### 3.2. Performance Measures of Model Trained with Optimum Hyper-Parameters

After hyper-parameter optimization process, models were trained using optimum hyper-parameters with training dataset and performance of models were assessed using testing dataset. Two callback functions, which were learning rate callback and early stopping callback, were added to model to improve the performance of our model. In the learning rate callback function, learning rate of the

model was divided by two if there is no improvement on model loss for three epochs. Although number of epoch is optimized, sometimes stopping the model early before reaching the epoch number can affect the model performance positively. Because of this reason, the training process was stopped using early stopping callback, if there is no improvement on model loss for six epochs. For this purpose, testingOpt dataset was used as validation dataset during the training phase. Accuracy is the one of the most common performance metrics to measure the performance of the machine learning models. However, the accuracy score can be very misleading for the unbalanced datasets. Because the number of malicious actions are much less than the number of normal actions, our dataset also considered as unbalanced datasets. Because of these reasons accuracy, detection rate (DR), f1-score, false discovery rate (FDR) and negative predictive value (NPV), which were shown in table 3, were computed to assess the performance of our model. The details about these metrics can be found in the information page [39].



Table 3. Performance measures of proposed model that trained using optimum hyper-parameters

Model Name	Model Base	Accuracy	DR	F1-Score	FDR	NPV
MLP	Role	86.27%	90.67%	92.25%	13.76%	95.09%
CNN	Role	83.07%	86.05%	90.37%	16.95%	89.05%
LSTM_2	Role	79.84%	89.88%	88.40%	20.24%	<b>99.89%</b>
CNN	User	98.86%	97.17%	99.19%	1.122%	95.47%
LSTM_2	User	<b>99.27%</b>	98.07%	<b>99.44%</b>	<b>0.724%</b>	96.86%
LSTM_4	User	99.18%	<b>98.89%</b>	99.38%	0.820%	98.60%
CNN_LSTM_2	User	98.85%	98.21%	99.18%	1.144%	97.56%
CNN_LSTM_4	User	98.94%	98.25%	99.24%	1.057%	97.56%

According to these results, it is desired that all scores except FDR are high and FDR is low. When the experiment results are examined, it is seen that user based LSTM model in deep two obtained the best scores for all metrics except NPV and DR. Role based LSTM model in deep two obtained the best score for NPV metric and user based LSTM model in deep 4 obtained the best score for DR metric. When these results are interpreted, it is understood that solo LSTM models were the most successful predictor for UEBA and the effect of increasing the depth on the performance was negative. In addition to this, user based models have been much more successful than role based models. Considering all these situations, it is thought that the most suitable model for a UEBA platform is the user based LSTM model. The bar chart for each metric is presented in figure 6 to better visualize the performance between models. In this figure, bar charts were shown for each performance metrics separately and different colors were assigned to each model separately. The model shown with the same color in bar charts of all performance metrics.

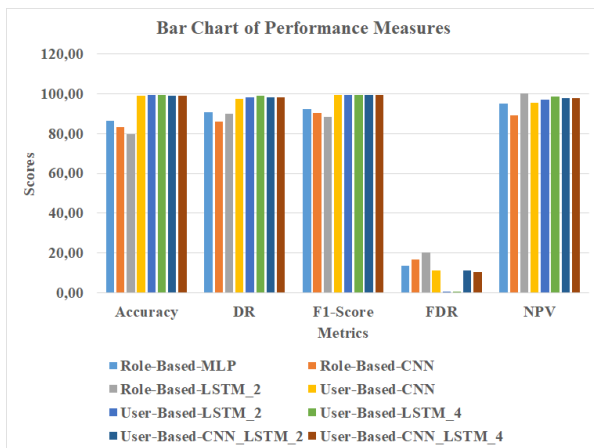


Figure 6. Bar chart of each model based on performance metrics

### 3.3. Comparison with the State-of-the-art

In order to better measure the contribution of the proposed models to the literature, it has great importance to compare them with the existing studies in the literature. Because of this reason, our model was compared with the four models,

which were proposed by Le and Zincir-Heywood [16], Al-Shehari and Alsowail [21], Nasir et al. [22] and Al-Mhiqani et al. [25], from the literature. Table 4 shows the comparison result of these models with our model with respect to accuracy, DR, FDR or F1-score if they exist in state-of-the-art models. For the state-of-the-art comparison our LSTM model with deep 2 was used because it was the model with the best performance. These state-of-the-art models were selected because they were also developed based on user and they trained model on version 4.2 of CERT dataset.

Table 4. Comparison between proposed and state-of-the-art methods on version 4.2 of CERT dataset

Model	Accuracy	DR	F1-Score	FDR
Model of Le and Zincir-Heywood [16]	98.33%	69.51%	-----	<b>0.350%</b>
Model of Al-Shehari and Alsowail [21]	-----	-----	99.00%	-----
Model of Nasir et al. [22]	90.00%	-----	94.00%	-----
Model of Al-Mhiqani et al. [25]	97.00%	-----	84.00%	2.88%
<b>Proposed Model</b>	<b>99.27%</b>	<b>98.07%</b>	<b>99.44%</b>	0.724%

Based on these results, our model outperformed the state-of-the-art model for accuracy, DR and F1-score metrics. However, autoencoder based model proposed by Le and Zincir-Heywood [16] were obtained better FDR score than our model. Considering the proportion taken for FDR, it is seen that both models achieve very low results. Considering that the proposed model achieves much better results especially for the DR score, it is thought that our

model is better than model proposed by Le and Zincir-Heywood [16] according to the average metric results.

### 3.4. Two-tailed Z-test

Based on the experiment results, it is concluded that user based models have better results than the role based models and the proposed model out-performed the state-of-the-art. In this phase, it was checked whether these improvements were significant. For this purpose, two-tailed z test was applied on our experiment results and also our experiment

results were compared with the state-of-the-art based on two-tailed z test score. The details about two-tailed z test can be found at information page [40]. Table 5 shows the p values between the best model (LSTM model in deep two) and the other model and table 6 shows the p values between LSTM\_2 model and state-of-the-art based on comparison metrics. A p-value of less than 0.01 indicates that the proposed model is 99% reliable, according to the two-tailed Z test, and statistically outperforms the models in the literature. Thus, a statistical comparison of the model with the literature can be made by examining the results in Table 6.

Table 5. P-values between LSTM\_2 model and the other model calculated with two-tailed Z test.

Model Name	Model Base	P-Value of Accuracy	P-Value of DR	P-Value of F1-Score	P-Value of FDR	P-Value of NPV
MLP	Role	0.00001	0.00001	0.00001	0.00001	0.00001
CNN	Role	0.00001	0.00001	0.00001	0.00001	0.00001
LSTM_2	Role	0.00001	0.00001	0.00001	0.00001	<b>0.00001</b>
CNN	User	0.00001	0.00001	0.00001	0.00001	0.00001
LSTM_4	User	0.00001	<b>0.00001</b>	0.00001	0.00001	0.00001
CNN_LSTM_2	User	0.00001	0.00001	0.00001	0.00001	0.00001
CNN_LSTM_4	User	0.00001	0.00001	0.00001	0.00001	0.00001

Table 6. P-values between LSTM\_2 model and state-of-the-art model with two-tailed Z test.

Model	Accuracy	DR	F1-Score	FDR
Model of Le and Zincir-Heywood [16]	0.00001	0.00001	-----	<b>0.00001</b>
Model of Al-Shehari and Alsowail [21]	-----	-----	0.00001	-----
Model of Nasir et. al. [22]	0.00001	-----	0.00001	-----
Model of Al-Mhiqani et al. [25]	0.00001	-----	0.00001	0.00001

Considering these results, it is seen that all improvements are statistically significant according to the two-tailed Z test at  $p < 0.01$ . The bold p-values represent the scores where the LSTM\_2 model was worse than the related model. According to the results of two-tailed Z test, the model proposed by Le and Zincir-Heywood obtained statistically significant better results than our model for false discovery rate results. However, when models are considered based on all other metrics, our model seems to be superior. Especially in the detection rate score, it is observed that our model is significantly better than the model of Le and Zincir-Heywood [16].

## 4. CONCLUSION AND FUTURE WORK

In this study, several deep learning models, which were developed using fully connected layer, convolutional neural network and long short-term memory network, were proposed for user and entity behavior analysis. Our models used two different approaches: role based and user based. In user based models, the dataset is split so that each piece contains the data of one user. In role based approach, users with the same role were included in the same partition. Experiment analysis showed that user based models are better than the role based models.

According to the analysis results, the MLP model achieved the highest accuracy among role-based models, with a rate of 86.27%. In addition, its DR, F1-Score, FDR, and NPV scores were measured at 90.67%, 92.25%, 13.76%, and 95.09%, respectively. Surprisingly, the accuracy of this model was even lower than that of the CNN\_LSTM\_2 model, which has the lowest accuracy among user-based models. The CNN\_LSTM\_2 model obtained values of 98.21%, 99.18%, 1.144%, and 97.56% for DR, F1-Score, FDR, and NPV scores, respectively, outperforming the role-based MLP model in all metric scores. Considering these results, it is concluded that developing an insider threat system for an institution using a user-based approach is more effective in terms of performance metric scores. However, it's worth noting that user-based models may face challenges in corporate settings due to potential changes in an individual's authority or department. Such changes can lead to shifts in employee activities, potentially affecting the model's predictive accuracy. To address this issue, fine-tuning of the model is necessary, which may result in a decrease in model efficiency. In light of these considerations, it is concluded that role-based models are more compatible with corporate organizations. Therefore, future studies should focus on improving the accuracy of role-based models.

The best proposed model, which uses two serial connected LSTM layers, outperformed the state-of-the-art models and the improvement was statistically significant according to the two-tailed Z test at  $p < 0.01$ . Although our model obtained 99.27% accuracy on version 4.2 of CERT dataset, it is thought that the model should be improved in other metric scores such as negative predictive value and detection rate. Because of this reason, in the future work, more complex deep learning models will be developed using different approaches such as user, role, department and team based. In addition to this, several feature selection techniques will be applied on CERT dataset. In this way, best deep learning structures for UEBA and the features that can best identify malicious action will be determined. A new UEBA approach will be proposed using this information and a model will be proposed using the data of our institution operating in the field of technology. For this purpose, daily activities of employees will be collected with a SIEM application used in our company. In line with the determined scenarios, these activities will be labeled and a specific dataset will be created for our institution. The proposed deep learning models will be also applied for this dataset and the results will be compared with the performance measures of CERT dataset. After all these analyzes, we are planning to integrate proposed model to our SIEM systems to analyze real time performance of UEBA systems



#### KAYNAKLAR (REFERENCES)

- [1] N. R. Mosteanu, "Artificial Intelligence and Cyber Security – Face To Face With Cyber Attack – A Maltese Case Of Risk Management Approach", *Ecoforum Journal*, 9(2), 2020.
- [2] D. Ghelani, **Cyber Security, Cyber Threats, Implications and Future Perspectives: A Review**, Authorea Preprints, 2022
- [3] Y. Hashem, H. Takabi, R. Dantu, and R. Nielsen, "A Multi-Modal Neuro-Physiological Study of Malicious Insider Threats", **International Workshop on Managing Insider Security Threats**, New York, NY, USA, 33-44, October 2017.
- [4] M. Dosh, "Detecting insider threat within institutions using CERT dataset and different ML techniques", *Periodicals of Engineering and Natural Sciences*, 9(2), 873-884, 2021.
- [5] Insider Threat Test Dataset, [https://kithub.cmu.edu/articles/dataset/Insider\\_Threat\\_Test\\_Dataset/12841247/1](https://kithub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247/1), 21.01.2024.
- [6] W. R. Claycomb and A. Nicoll, "Insider Threats to Cloud Computing: Directions for New Research Challenges", **36th Annual Computer Software and Applications Conference**, İzmir, Turkey, 387,394, July 2012.
- [7] X. Xiangyu et al., "Method and System for Detecting Anomalous User Behaviors: An Ensemble Approach", **30th International Conference on Software Engineering and Knowledge Engineering**, San Francisco, California, USA , 263-307, July 2018.
- [8] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, **Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams**, arXiv, 2017.
- [9] L. Lin, S. Zhong, C. Jia, and K. Chen, "Insider Threat Detection Based on Deep Belief Network Feature Representation", **International Conference on Green Informatics**, Fuzhou, China, 54-59, August 2017.
- [10] F. Yuan, Y. Cao, Y. Shang, Y. Liu, J. Tan, and B. Fang, "Insider Threat Detection with Deep Neural Network", **Computational Science**, Wuzi, China, 43-54, 2018.
- [11] O. Lo, W. J. Buchanan, P. Griffiths, and R. Macfarlane, "Distance Measurement Methods for Improved Insider Threat Detection", *Security and Communication Networks*, 2018(e5906368), 1-18, 2018.
- [12] D. C. Le and A. N. Zincir-Heywood, "Evaluating Insider Threat Detection Workflow Using Supervised and Unsupervised Learning", **IEEE Security and Privacy Workshops**, San Francisco, CA, USA, 270-275, May 2018.
- [13] O. Igbe and T. Saadawi, "Insider Threat Detection using an Artificial Immune system Algorithm", **9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference**, New York, USA, 297-302, November 2018.
- [14] A. J. Hall, N. Pitropakis, W. J. Buchanan, and N. Moradpoor, "Predicting Malicious Insider Threat Scenarios Using Organizational Data and a Heterogeneous Stack-Classifer", **IEEE International Conference on Big Data**, Seattle, WA, USA, 5034-5039, December 2018.
- [15] M. Aldairi, L. Karimi, and J. Joshi, "A Trust Aware Unsupervised Learning Approach for Insider Threat Detection", **IEEE 20th International Conference on Information Reuse and Integration for Data Science**, Los Angeles, California, USA, 89-98, July 2019.
- [16] D. C. Le and N. Zincir-Heywood, "Exploring anomalous behaviour detection and classification for insider threat identification", *International Journal of Network Management*, 31(4), 2021.
- [17] D. C. Le and A. Nur Zincir-Heywood, "Machine learning based Insider Threat Modelling and Detection", IFIP/IEEE Symposium on Integrated Network and Service Management, Washington DC, USA, 1-6, April.
- [18] M. Nasser Al-mhiqani, R. Ahmad, Z. Zainal Abidin, W. Yassin, A. Hassan, and A. Natasha Mohammad, "New insider threat detection method based on recurrent neural networks", *Indonesian Journal of Electrical Engineering and Computer Science*, 17(3), 1474, 2020.
- [19] B. Sharma, P. Pokharel, and B. Joshi, "User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder - Insider Threat Detection", **11th International Conference on Advances in Information Technology**, New York, USA, 1-9, July 2020.
- [20] Z. Tian, C. Luo, H. Lu, S. Su, Y. Sun, and M. Zhang, "User and Entity Behavior Analysis under Urban Big Data", *ACM Transactions on Data Science*, 1(3), 1-16, 2020.
- [21] T. Al-Shehari and R. A. Alsowail, "An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques", *Entropy*, 23(10), no. 10, 2021.
- [22] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, "Behavioral Based Insider Threat Detection Using Deep Learning", *IEEE Access*, 9(1), 143266–143274, 2021.

- [23] D. Sun, M. Liu, M. Li, Z. Shi, P. Liu, and X. Wang, "DeepMIT: A Novel Malicious Insider Threat Detection Framework based on Recurrent Neural Network", **24th International Conference on Computer Supported Cooperative Work in Design**, Dalian, China, 335-341, May 2021.
- [24] E. Pantelidis, G. Bendiab, S. Shiaeles, and N. Kolokotronis, "Insider Threat Detection using Deep Autoencoder and Variational Autoencoder Neural Networks", **IEEE International Conference on Cyber Security and Resilience**, Rhodes, Greece, 129-134, July 2021.
- [25] M. N. Al-Mhiqani et al., "A new intelligent multilayer framework for insider threat detection", *Computers & Electrical Engineering*, 97(1), 107597, January 2022.
- [26] M. Alslaiman, M. I. Salman, M. M. Saleh, and B. Wang, "Enhancing false negative and positive rates for efficient insider threat detection", *Computers & Security*, 126(1), 103066, March 2023.
- [27] D. Li, L. Yang, H. Zhang, X. Wang, and L. Ma, "Memory-Augmented Insider Threat Detection with Temporal-Spatial Fusion", *Security and Communication Networks*, 2022(1), e6418420, 2022.
- [27] T. Karayel, A. Akbıyık, "A Global Perspective of Cybersecurity Research: Publication Trends and Research Directions", *Journal of Information Technologies*, 16(3), 221 – 235, 2023.
- [29] Y. Gormez, Z. Aydin, R. Karademir, and V. C. Gungor, "A deep learning approach with Bayesian optimization and ensemble classifiers for detecting denial of service attacks", *International Journal of Communication Systems*, 33(11), e4401, 2020.
- [30] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms", **Advances in Neural Information Processing Systems**, Nevada, USA, 2012.
- [31] A. Salama, A. E. Hassanien, and A. Fahmy, "Sheep Identification Using a Hybrid Deep Learning and Bayesian Optimization Approach", *IEEE Access*, 7(1), 31681–31687, 2019.
- [32] J. Snoek et al., "Scalable Bayesian Optimization Using Deep Neural Networks," **32nd International Conference on Machine Learning**, Lille, France, 2171-2180, Jun 2015.
- [33] H. Kaur, H. S. Pannu, and A. K. Malhi, "A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions", *ACM Computing Survey*, 52(4), 1-36, August 2019.
- [34] Big Five personality traits: [https://en.wikipedia.org/w/index.php?title=Big\\_Five\\_personality\\_traits&oldid=1114671408](https://en.wikipedia.org/w/index.php?title=Big_Five_personality_traits&oldid=1114671408), 21.01.2024.
- [35] Y. Görmez, H. Arslan, Y. E. Işık, and İ. E. Dadaş, "A User and Entity Behavior Analysis for SIEM Systems: Preprocessing of The Computer Emergency and Response Team Dataset," *Journal Soft Computing*, 4(1), 2023.
- [36] Arge-Preprocessing-CERT: <https://github.com/Detaysoft/Arge-Preprocessing-CERT>, 21.01.2024
- [37] Keras: the Python deep learning API: <https://keras.io/>, 21.01.2024.
- [38] scikit-optimize:<https://scikit-optimize.github.io/stable/>, 21.01.2024.
- [39] Precision and recall: [https://en.wikipedia.org/w/index.php?title=Precision\\_and\\_recall&oldid=1122267443](https://en.wikipedia.org/w/index.php?title=Precision_and_recall&oldid=1122267443), 21.01.2024.
- [40] Z Score Calculator for 2 Poulation Proportions, <https://www.socscistatistics.com/tests/ztest/default2.aspx>, 21.01.2024.

# Hiper Parametre Ayarlama ve Veri Dengelemenin Kalp Hastalığı Tahmini İçin Kullanılan Makine Öğrenimi Algoritmaları Üzerindeki Etkilerinin İncelenmesi

*Araştırma Makalesi/Research Article*

 Fuat Sungur<sup>1\*</sup>,  Halit Bakır<sup>2</sup>

<sup>1</sup>Savunma Teknolojileri ABD, Sivas Bilim ve Teknoloji Üniversitesi, Sivas, Türkiye  
<sup>2</sup>Bilgisayar Mühendislik Bölümü, Sivas Bilim ve Teknoloji Üniversitesi, Sivas, Türkiye

[sungurbey58@hotmail.com](mailto:sungurbey58@hotmail.com), [halit.bakir@sivas.edu.tr](mailto:halit.bakir@sivas.edu.tr)

(Geliş/Received: 18.07.2023; Kabul/Accepted: 31.10.2023)

DOI: 10.17671/gazibtd.1399813

**Özet**— Kalp hastalığı belirtilerinin ihmal edilmesi ciddi rahatsızlıklarla hatta ölümle sonuçlanabilir. Makine öğrenme teknikleri ile ön tanı için bu belirtiler kullanılarak kişide kalp hastalığı olup olmadığına dair tahmin yapılabilmektedir. Bu çalışmada Logistic Regression, Decision Trees, Random Forest, K Nearest Neighbors, Naive Bayes, Gradient Boosting, XGBoost ve Bagging algoritmaları ile kalp hastalığı tahmini yapılmıştır. SMOTE, SMOTETomek, Oversample Minority Class, Undersample Majority Class veri dengeleme yöntemleri ile dört ayrı veri seti oluşturulmuştur. Seçilen tüm makine öğrenme algoritmalarına Random Search ve Bayesian Optimizasyon teknikleriyle hiper parametre optimizasyonu yapılarak sonuçlar karşılaştırılmıştır. Veri dengeleme ve hiper parametre optimizasyonunun kalp hastalığının tahmininde kullanılan makine öğrenme teknikleri performansına etkisi karşılaştırılarak literatüre özgün bir çalışma kazandırılmıştır. Çalışmada Amerika Birleşik Devletleri'nde 319.795 kişi ile yapılan 20 öz nitelikli bir anket olan veri seti kullanılmıştır. Random Forest algoritması SMOTETomek veri dengeleme tekniği kullanılarak ve Bayesian hiper parametre optimizasyonu yapılarak oluşturulan modelde %94 tahmin başarısı elde edilmiştir. Ayrıca, Random Forest algoritması ile Oversample Minority Class veri dengeleme tekniği kullanılarak ve Bayesian hiper parametre optimizasyonu yapılarak %97 sınıflandırma doğruluğu elde edilmiştir.

**Anahtar Kelimeler**— kalp Hastalıkları, rasgele orman, makine öğrenmesi, smote, smotetomek, bayes optimizasyon

## Evaluating The Effects of Hyperparameter Tuning and Data Balancing on Machine Learning Algorithms Used for Heart Disease Prediction

**Abstract**— Neglecting the symptoms of heart disease can result in serious conditions and even death. Machine learning techniques can be used to make predictions about whether a person has heart disease based on these symptoms. In this study, heart disease prediction was performed using Logistic Regression, Decision Trees, Random Forest, K Nearest Neighbors, Naive Bayes, Gradient Boosting, XGBoost, and Bagging machine learning algorithms. Four separate datasets were created using data balancing methods such as SMOTE, SMOTETomek, Oversample Minority Class, and Undersample Majority Class. Hyperparameter optimization was conducted for all selected machine learning algorithms using Random Search and Bayesian Optimization techniques, and the results were compared. By comparing the impact of data balancing and hyperparameter optimization on the performance of machine learning techniques used in predicting heart disease, this study contributes to the literature with an original approach. The study utilized a dataset from a survey of 319,795 individuals in the United States, which included 20 relevant features. The Random Forest algorithm achieved a prediction accuracy of 94% in the model created using the SMOTETomek data balancing technique and Bayesian hyperparameter optimization. Additionally, the Random Forest algorithm, with the Oversample Minority Class data balancing technique and Bayesian hyperparameter optimization, achieved a classification accuracy of 97%.

**Keywords**— heart Diseases, random forest, machine learning, smote, smotetomek, bayesian optimization

## 1. GİRİŞ (INTRODUCTION)

Kalp hastalıkları ölüm nedenleri arasında üst sıralarda yer almaktadır. Birçok insan kalp krizi ve kalp sağlığına bağlı hastalıklar nedeniyle hayatını kaybetmektedir. Dünya Sağlık Örgütü'nün 2019 yılı verilerine göre yaklaşık 17,9 milyon insan kalp sağlığına bağlı hastalıklardan dolayı hayatını kaybetmiştir. Bu sayı dünya üzerindeki ölümlerin yaklaşık %32'sine denk gelmektedir. Kalp hastalıklarına bağlı ölümlerin büyük bir kısmını (%85) ani ölümler oluşturmuştur [1].

Kalp hastalıklarının çeşitli nedenleri olabilir. Bu nedenler genetik faktörler ve çevresel faktörler olarak iki ana başlıkta toplanabilir. Genetik faktörler önceki kuşaklardan aldığımız hastalığa yatkınlık gösteren genlerle alakalı olup henüz bilim bu konunun çözümünde yeterince ilerleme gösterememiştir. Çevresel faktörler ise bebeğin anne rahmine düştüğü andan itibaren, annenin hamilelik sürecindeki alışkanlıkları, beslenmesi ve hatta duyu durumları ile doğumdan sonra bireyin yaşam tarzını kapsamaktadır. Kalp hastalıklarını tetikleyen bireyin yaşam tarzı ile ilgili başlıca göstergeler beslenme alışkanlıkları, zararlı alışkanlıklar ve yaşanılan çevre gelmektedir. Kalabalık, stresli büyük şehirlerde yaşam, sağlıksız beslenme, obezite, hareketsiz, spordan uzak yaşam tarzı başta kalp hastalıkları olmak üzere birçok hastalığa davetiye çıkarmaktadır. Bununla birlikte alkol, sigara ve uyuşturucu madde kullanımı kalp hastalıklarına önemli ölçüde neden olmaktadır.

Bu çalışmada kalp hastalıklarının tahmin edilmesinde makine öğrenme algoritmalarının karşılaştırmalı performans analizinin yapılması amaçlanmıştır. Kalp hastalıklarının erken teşhisi büyük ölçüde hayat kurtarıcı olduğu için bu çalışmanın sağlık sisteminde bir farkındalık yaratması beklenmektedir.

Günümüzde kalp hastalıklarının nedenleri ve kalp hastalıkları ile ilgili birçok çalışma yapılmakta olup bu konuda birçok kaynak bulunmaktadır. Bu kaynaklardan biriside veri setleridir. Veri seti belirli bir konu ile alakalı bilgilerin sistematik ve anlamlı bir şekilde depolandığı yapılarıdır. Bu çalışmada Amerika Birleşik Devletleri Hastalık Kontrol ve Önleme Merkezleri tarafından her yıl rutin olarak yapılan Davranışsal Risk Faktörü Gözetleme Sisteminin bir parçası olan veri setinden yararlanacağız. Veri seti 2020 yılında yaklaşık 320bin kişi ile yapılan anketin 20 özneteliğine dair örnekleri içermektedir.

Veri setlerini işlemenin birçok yolu vardır. Bunlardan birisi de makine öğrenmesi algoritmaları ile veri setlerinin analiz edilmesi ve çıkan anlamlı sonuçların değerlendirilmesidir. Bu çalışmada, makine öğrenmesi sınıflandırma algoritmalarından Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), K Nearest Neighbors (KNN), Naive Bayes (NB), Gradient Boosting (GB), XGBoost (XGB) ve Bagging (BAG) algoritmalarının her biri ayrı ayrı çalışılıp en uygun yöntem ya da yöntemlerin seçilmesi planlanmıştır. En uygun algoritmalar belirlendikten sonra hiper parametre

optimizasyonu (HPO) teknikleri ile en iyi parametreler belirlenerek bu hiper parametreler ile modeller daha verimli hale getirilecektir.

## 2. KAYNAK ARAŞTIRMASI (LITERATURE SURVEY)

Bu kısımda çalışmada yararlanılan kaynakların amaçları, teknikleri, elde edilen sonuçları değerlendirilerek literatüre katkıları ve eksiklikleri gözden geçirilecektir.

Yadav ve ark. [3] kalp hastalıklarını tahmin etmek üzere makine öğrenme tekniklerine dayanan bir arayüz oluşturmayı hedeflediler. Rasgele orman ve karar ağaçlarında %97,08 tespit başarısı, Lojistik Regresyonda %80,52 tespit başarısı elde ettiler. Başka bir çalışmada [4] Bhatt ve ark. kalp hastalıklarının tespit edilmesinde RF, XGB, Multi-Layer Perceptron (MLP), DT makine öğrenme algoritmalarını kullandılar. MLP tekniğinde %87,23 tespit başarısı elde ettiler. Ayrıca bu çalışmada veri temizleme, özellik seçimi ve kümeleme gibi veri düzenleme işlemleri yaptılar. Özdemir [5] kalp hastalıklarının teşhisinde EKG verilerini kullandı. DT, RF, Extra Tree, GB ve Support Vector Machine (SVM) teknikleri kullanılarak yapılan çalışmada Extra Tree algoritmasında %96,14 tespit doğruluğu elde edilmiştir. Başka bir çalışmada [6] kalp hastalıklarının tahmin edilmesinde makine öğrenme algoritmaları kullanıldı. Çalışmada GridSearchCV tekniği ile hiper parametre optimizasyonu yapılarak makine öğrenme algoritmalarının performansını artırmak hedeflenmiştir. % 3-5 aralığında performans artışı elde edilen çalışma neticesinde AdaBoost algoritması ile %95 tespit başarısı elde edilmiştir. Anitha ve ark. [7] kalp hastalıklarının tahmin edilmesine yönelik yaptığı çalışmada KNN, NB ve SVM makine öğrenme algoritmalarını karşılaştırmıştır. Çalışmada kullandığı veri setinde NB algoritmasının %86,6 doğruluk oranı ile en iyi sonuçları verdiği çıkarımında bulunulmuştur. Çifçi M. [8] yapay sinir ağlarını birleştirerek oluşturduğu sistemle kalp hastalıkları tahminini amaçlamıştır. Sınıf Nitelik Bağımlılık Maksimizasyonu, Temel Bileşenler Analizi ve SVM yöntemlerini birleştirilerek kalp hastalıkları ön teşhisinde kullanılabilir bir sistem oluşturmuştur. 270 hastanın 13 öz nitelikli veri setini kullanmıştır. Hibrit sistemin kullanılmadığı sınıflayıcılarda %75,27 doğruluk oranı bulunurken, kendi hibrit sisteminde %89,17 başarı elde etmiştir. Kamat P. [9] çalışmasında J48, NB ve SVM makine öğrenme algoritmalarını k-kat çapraz doğrulama ile birlikte kullanmıştır. Kullanılan yöntemlerin kalp hastalığı tahmininde kullanılabilirliği sonucuna ulaşmış ancak model sonuçları ile ilgili paylaşım yapılmamıştır. Çalışmanın bu yönü ile eksik kaldığı söylenebilir. Rajdhan ve ark. [10] kullandıkları veri seti ile kalp hastalıklarını tahmin etmeyi amaçladılar. NB, DT, LR, RF tekniklerini karşılaştırıp RF algoritmasının %90,16 ile diğerlerine kıyasla en yüksek doğruluk oranına sahip olduğunu elde ettiler. Görgün M. [11] 165'i kalp hastası olan 303 kişinin bilgilerinden oluşan veri setini kullanmıştır. Cinsiyet, yaş, kolesterol, diyabet, göğüs ağrısı gibi belirtileri kalp hastalığı tahmininde kullanmıştır. Makine öğrenmesi algoritmalarından KNN, SVM, NB, LR, DT, RF, XGB, LightGBM ve BAG algoritmalarını karşılaştırmış, RF

kullanarak %90,16 doğruluk oranı elde etmiştir. Çalışmada %81,97 tahmin değeri ile en düşük doğruluk oranına sahip LightGBM algoritması göze çarpmaktadır. Repaka ve ark. [12] mobil uygulama yolu ile kişiden aldıkları bilgileri veri setinde işleyerek kişinin kalp hastalığı olup olmadığını tahmin etmek istedikler. NB algoritmasını kullandıkları uygulamada %89,77 doğruluk oranına ulaştılar. Köse O. [13] kullandığı veri setinde kalp hastalığına etki eden 38 adet faktör üzerinde çalışmıştır. DT algoritmalarından Chi-Square Automatic Interaction Detection ve CART tekniklerini kullanarak karşılaştırmalı sonuçlar elde etmiştir. Veri setinde %70'i eğitim örnekleme, %30 'u test örnekleme aldığı çalışmanın daha başarılı olduğu sonucuna ulaşmıştır. CART ve Chi-Square Automatic Interaction Detection algoritmalarının farklı ağaç oluşturma yöntemleri benimsediği, Chi-Square Automatic Interaction Detection algoritması ile detaylı sonuçlar elde edilirken CART algoritması ile genel sonuçlar elde edildiğini sonucuna ulaşmıştır. Ramalingam ve ark. [14] 2019 yılında makine öğrenme algoritmalarını kullanarak kalp hastalıklarını tahmin etmeyi amaçlamışlar. DT modellerinin Temel Bileşen Analizi ile kullandığında performansın iyi olmasına rağmen uyum problemleri ile karşılaşıldığı sonucuna ulaşmıştır. RF algoritmasında birden çok DT kullanarak uyum problemlerini çözebildikleri çıkarımında bulundular. NB modellerinin hızlı olduğu ve iyi sonuçlar (%83,49) gösterdiğini belirttiler. SVM yönteminin ise doğruluk oranının çok yüksek olduğu (%92,1) sonucuna ulaştılar. Sharma ve ark. [15] makine öğrenme tekniklerinin karşılaştırmalı tahlilini yapmıştır. KNN modelin SVM modelden daha verimli çalıştığı çıkarımında bulunmuştur. J48 modelinin aşırı uyum konusunda DT modelinden daha iyi olduğu, yapay sinir ağlarının çerim için eğitimde daha iyi sonuçlar verdiği çıkarımında bulunmuştur. Kâmil [16] araştırmasında RF ve Kmean kavramlarına dayalı Parçacık Sürü Optimizasyonuna tekniklerini sentezleyerek en iyi K ortalama tekniğini ve RF'deki ağaç sayısını bulmayı amaçlamıştır. Kullandığı yöntem ile geleneksel yöntemlerden daha kullanışlı işlevsel ve hızlı sonuçlar elde etmiştir. Salman I. [17] 787 hasta ve 24 değişken içeren veri setini kullandığı çalışmada NB, DT ile güçlendirilmiş Naive Bayes (TAN) ve TAN ve Chow-Liu (TANI) karşılaştırması yaparak kalp hastalıkları tahmini yapmayı amaçlamıştır. Sonuçta TANI metodunun diğerlerine kıyasla en yüksek doğruluk oranına sahip olduğu sonucuna ulaşmıştır. Konda ve ark. [18] kullandıkları veri setindeki 15 ayrı niteliği kullanarak kalp hastalığı ön tanısı için makine öğrenme modellemesi yapmayı amaçladılar. DT, NB ve Yapay Sinir Ağları algoritmalarını kullandıkları çalışmada %89 doğruluk oranı ile DT algoritmasının en uygun yöntem olduğu çıkarımında bulundular. Tarawneh ve ark. [19] 303 örnekleme 14 öznelik içeren araştırmalarında SVM, KNN, RF, J48 algoritma modellerini kullanarak %89,2 doğruluk oranını ile kalp hastalıklarının tahmini ile ilgili çalışma gerçekleştirdiler. Venkatesh ve ark. [2] çalışmalarında Naive Bayes algoritma modelini kullanmıştır. Kalp hastalıklarının tahmininde NB modelinin %97,1 doğruluk oranına sahip olduğunu ortaya koydular. Çil E. [20] 2022 yılında yaptığı çalışmada Yapay Sinir Ağları, SVM ve KNN, LR, DT, RF, NB algoritmalarını karşılaştırmalı bir

şekilde kullanarak kalp hastalıklarını tespit etmeyi amaçlamıştır. LR (%90,77), SVM (%90,52) ve Yapay Sinir Ağları (%90,54) algoritmalarında en anlamlı değerlere ulaşılmıştır. Kesinlik skorlarında LR ve Yapay Sinir Ağları, Duyarlılık skorunda NB, F1 skorlarında ise DT ve NB'nin en başarılı modeller olduğu sonucuna ulaşılmıştır. Sonuç olarak çalışmada LR ve Yapay Sinir Ağları modellerinin en uygun modeller olduğuna karar verilmiştir.

Literatürde kalp hastalıklarının tespit edilmesi ile ilgili çalışmaların daha çok tahmin modelleri oluşturmaya yönelik olduğu görülmüştür. Araştırılan çalışmaların hiç birisinde veri dengelemenin model performansı üzerindeki etkisi incelenmemiştir. HPO yapılan çalışmalarda HPO tekniklerinin karşılaştırması yapılmamıştır. Araştırılan kaynaklar içerisinde en verimli model % 97,1 tahmin başarısı ile NB algoritmasına [2] aittir. Veri dengeleme ve hiper parametre optimizasyonunun kalp hastalığının tahmin edilmesinde kullanılan makine öğrenme algoritmalarının performansı üzerindeki etkisinin literatürdeki çalışmalardan farklı ve araştırılması gereken bir alan olduğu düşüncesi bizi bu çalışmaya motive etmiştir. Bizim çalışmamız kalp hastalıklarının tahmin edilmesinde %97 tahmin başarısının yanında veri dengeleme ve HPO'yu içeren detaylı uygulamaları ile benzer çalışmaların önüne geçmektedir.

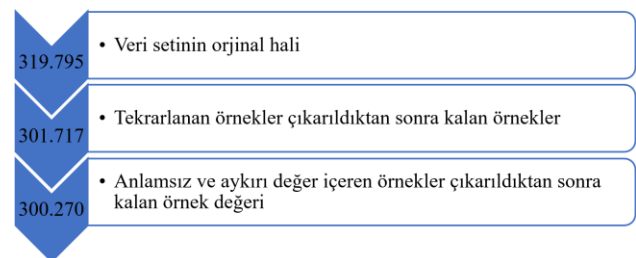
### 3. MATERYAL ve YÖNTEM (MATERIAL and METHOD)

#### 3.1. Veri Seti (Dataset)

Çalışmada kullanılan en önemli materyal veri setidir. Bu çalışmada kullanılan veri seti, Amerika Birleşik Devletleri Hastalık Kontrol ve Önleme Merkezleri tarafından her yıl rutin olarak yapılan Davranışsal Risk Faktörü Gözetleme Sisteminin bir parçasıdır. 2020 yılında ait veriler kalp rahatsızlıkları ile ilgili yaklaşık 320000 kişi ile yapılan anket araştırması sonuçlarını kapsamaktadır.

#### 3.2. Veri Setinin Hazırlanması (Preparation of the Data Set)

Çalışma sırasında karışıklığı önlemek için öznelik isimleri daha anlaşılabilir isim kısaltmaları ile değiştirilmiştir. Veri setinde ilk durumda 319.795 olan örnek, tekrarlayan satırlar, anlamsız aykırı değerler çıkarıldıktan yani veri seti temizlendikten sonra 300,270 örnek olarak güncellenmiştir. Temizleme işlemi aşamaları Şekil 1'de gösterilmiştir.



Şekil 1. Veri setinin temizlenme aşamaları (Stages of cleaning the data set)

Veri setinin orijinal halinde 319.795 örnek, 18 öznitelik bulunmaktadır. “Race” öznitelığının her bir değeri On-Hot-Encoding tekniği kullanılarak birer öznitelige dönüştürülmüştür. Bu işlemden sonra öznitelik sayısı 23 olmuştur. Öznitelik değerleri üzerinde istatistiksel işlemler yapılabilmesi önce numerik dönüşüm yapılmıştır. Daha sonra her değer kategorik olarak [0,1] arasına ölçeklendirilmiştir. Bu işlem sonrasında korelasyon analizi ile özniteliklerin çıktı ile ilişkisi incelenmiştir. Kaggle

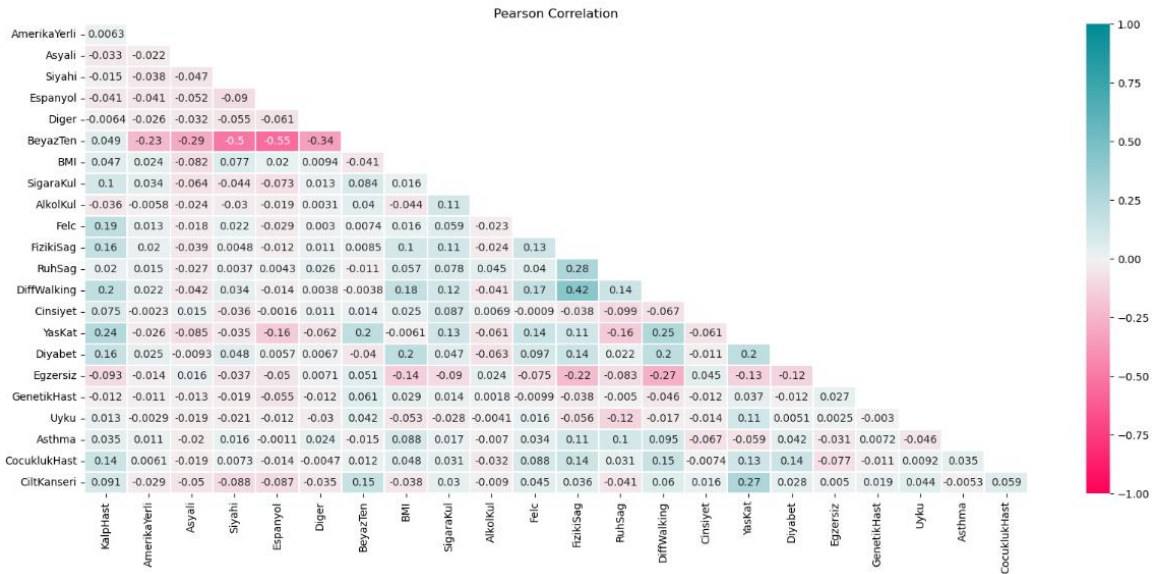
sitesinden alınan veri seti 279 soruluk anketin 18 soruya indirgenmiş halidir [21]. Öznitelikler ışığında kişide kalp hastalığı olup olmadığını değerlendirmeye yönelik işlem yapılması amaçlanmaktadır. Bu nedenle kalp hastalığı olup olmadığı ile ilgili öznitelik “HeartDisease” çıktı indeksi olarak belirlenmiştir. Veri setinin %20’si test verisi, %80’i ise eğitim verisi olarak kullanılmıştır. Veri setinde bulunan öznitelikler ve açıklamaları Tablo 1 de verilmiştir.

Tablo 1. Veri seti öznitelik bilgileri (Dataset variable information)

Öznitelik Adı	Türü	Aldığı Değerler	Açıklama
HeartDisease	Object	(Yes, No)	Kişide kalp hastalığı olup olmadığını gösterir.
BMI	Float	(12,02<<94,85)	Vücut kitle endeksini (VKİ) gösterir.
Smoking	Object	(Yes, No)	Kişinin sigara kullanıp kullanmadığını gösterir.
AlcoholDrinking	Object	(Yes, No)	Kişinin alkol kullanıp kullanmadığını gösterir.
Stroke	Object	(Yes, No)	Kişinin felç geçirmediğini gösterir.
PhysicalHealth	Float	(0<<30)	Kişinin fiziksel sağlığının son 30 günde kaç gün iyi olmadığını gösterir.
MentalHealth	Float	(0<<30)	Kişinin ruhsal sağlığının son 30 günde kaç gün iyi olmadığını gösterir.
DiffWalking	Object	(Yes, No)	Yürürken veya merdiven çıkarken zorluk yaşayıp yaşamadığını gösterir.
Sex	Object	(Male, Female)	Kişinin cinsiyetini gösterir.
AgeCategory	Object	18-24,25-29, ...,75-79,80+	13 Yaş kategorisi
Race	Object	(White, Hispanic, Black, Other, Asian, Native)	Kişinin ırkını gösterir. 6 ayrı kategoride ırk bulunur.
Diabetic	Object	No, Yes, No borderline diabetes, Yes (during pregnancy)	Kişinin diyabet hastası olup olmadığını gösterir. 4 ayrı kategoride diyabet verisi bulunur.
PhysicalActivity	Object	(Yes, No)	Kişinin egzersiz yapıp yapmadığını gösterir.
GenHealth	Object	(Yes, No)	Kişinin genetik hastalığı olup olmadığını gösterir.
SleepTime	Float	7<<21	Kişinin günlük uyuduğu süreyi gösterir.
Asthma	Object	(Yes, No)	Kişinin astım hastası olup olmadığını gösterir.
KidneyDisease	Object	(Yes, No)	Kişinin böbrek hastalıkları olup olmadığını gösterir.
SkinCancer	Object	(Yes, No)	Kişinin cilt kanseri olup olmadığını gösterir.

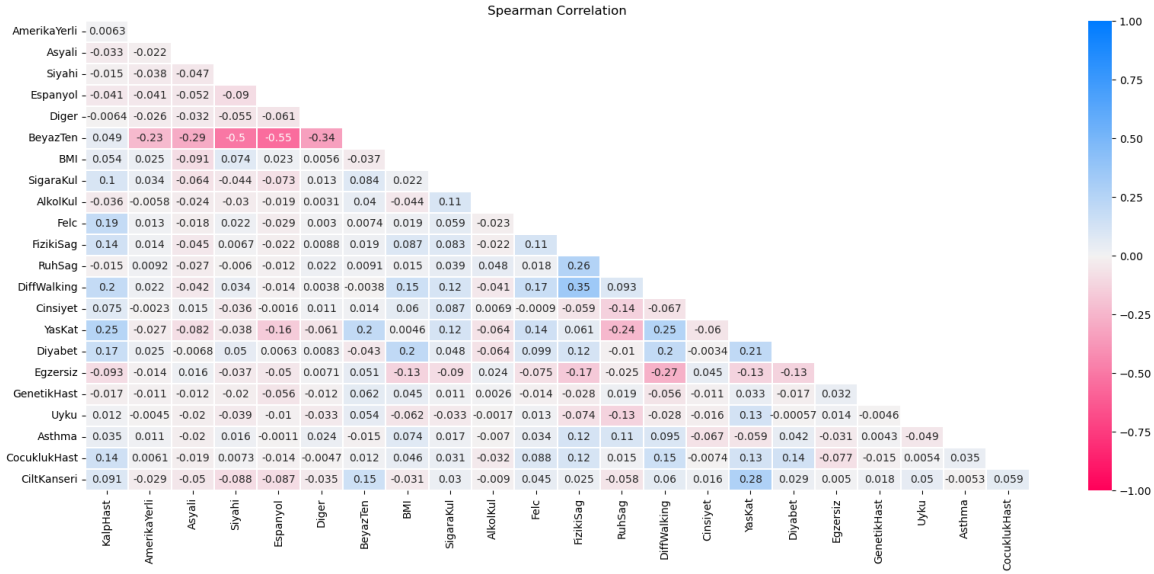
Korelasyon ilişkisi [-1,1] arasında herhangi bir değer olabilir. Korelasyon -1’e yaklaşıyorsa negatif ilişki vardır. Korelasyon +1’e yaklaşıyorsa pozitif ilişki vardır. 0 değeri ise değişkenler arasında ilişki olmadığını gösterir. Pearson korelasyon katsayısı iki değişken arasındaki lineer ilişkiyi incelerken, Spearman korelasyon katsayısı monoton ilişkiye odaklanır. Şekil 2’te Pearson korelasyon analizi ve Şekil 3’te Spearman korelasyon analizi sonuçları

gösterilmiştir. Buna göre “Amerikan yerli ırkı”, “Uyku süresi” ve “Genetik hastalıklar” özniteliklerinin çıktımız olan “HeartDisease” ile en düşük ilişkiye sahip olduğu anlaşılmıştır. Sırasıyla “Yaş kategorisi”, “Yürüyüş ve merdiven çıkmada zorluk” ve “Felç geçirme” özniteliklerinin “HeartDisease” ile en yüksek ilişkiye sahip olduğu anlaşılmıştır.





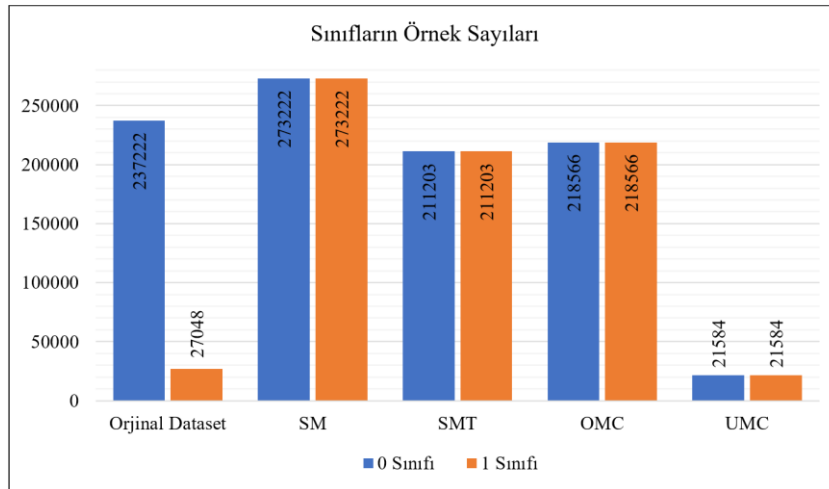
Şekil 1. Özniteliklerin Pearson korelasyon ilişkisi (Pearson correlation relationship of attributes)



Şekil 2. Özniteliklerin Spearman korelasyon ilişkisi (Spearman correlation relationship of attributes)

Korelasyon ilişkisine bakıldıktan sonra veri setinin dengeleme işlemine geçildi. Oversample Minority Class (OMC), Undersample Majority Class (UMC), SMOTE (SM), SMOTETomek (SMT) veri dengeleme yöntemleri ile oluşturulan dört ayrı veri seti veri dengelemenin model performansına etkisini görmek üzere ayrı ayrı kaydedilmiştir. Dengeleme işlemleri sonucunda veri

setindeki örnek sayıları dengeleme yöntemine bağlı olarak değişmiştir. Kalp hastalığı sınıfları 0 ve 1 değerlerinin dengeleme yöntemlerine bağlı değişimi Şekil 4'te gösterilmiştir. SM veri dengeleme yöntemi ile oluşturulan veri seti 546444 örneklidir. UMC veri dengeleme yöntemi ile oluşturulan veri seti 43168 örneklidir.



Şekil 3. Veri dengeleme uygulamaları (Data balancing applications)

### 3.3. Veri Setinin Dengelenmesi (Balancing the Dataset)

#### 3.3.1. SMOTE

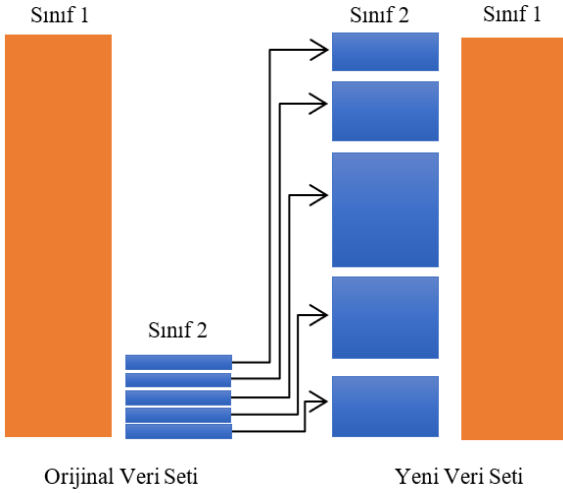
SMOTE yönteminde azınlık sınıfı oluşturulurken her azınlık örneğinden bir numune alınarak KNN sınıfının komşusundan rasgele seçilen parçaları birleştirilir ve bir düzlem boyunca numuneler tanıtılarak örneklem oluşturulur. Gereken örneklem büyüklüğüne bağlı olarak K en yakın komşudan alınacak örneklem de değişir [22].

#### 3.3.2. SMOTETomek

SMOTETomek tekniğinde ilk adımda azınlık örneklerini çoğaltmak için SMOTE yöntemi kullanılır. İkinci adımda ise çoğaltılmış azınlıklar ile oluşturulan yeni veri setinde Tomek Link eşleri Tomek Link tekniği kullanılarak kaldırılır. Tomek Link yönteminde çoğunluk sınıfı verilerinden azınlık sınıfı verilerine en yakın veriler kaldırılır [23].

### 3.3.3. Azınlık Sınıfını Yüksek Örnekleme (Oversample Minority Class)

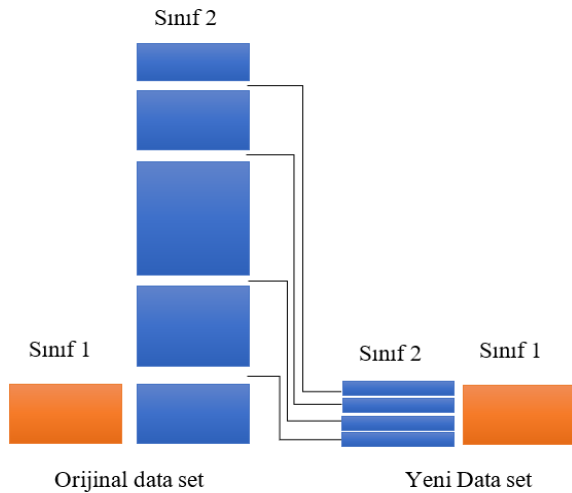
Bu aşırı örneklemme tekniği veri setinin dengesizliğini gidermek için azınlık sınıfı çoğunluk sınıfına eşitlenene kadar azınlık sınıfına yeni örnekler eklenir. Rastgele çoğaltım işlemi Şekil 5’de görüldüğü gibi azınlık sınıftan rastgele örnekler alınarak azınlık sınıfına eklenir. SMOTE gibi sentetik aşırı örneklemme yönteminde ise yapay örnekler oluşturularak azınlık sınıfı çoğaltılır [24].



Şekil 4. Azınlık sınıfını yüksek örneklemme (Oversample minority class)

### 3.3.5. Çoğunluk Sınıfı Azaltma (Undersample Majority Class)

Çoğunluk sınıfını azınlık sınıfına indirgeme yönteminde çoğunluk sınıfına ait verilerden rasgele veya istatistiksel örnekler alınarak azınlık sınıfına Şekil 6’daki gibi eşitlenir. Çoğunluk ve azınlık sınıfı farkı fazla olduğunda veri kaybı fazla olacağı için zayıf örneklem sunar. Ancak bazı durumlarda çoğunluk sınıfı verilerini temizlediği için iyi örneklem sunabilir [24].



Şekil 5. Çoğunluk sınıfı azaltma (Undersample majority class)

### 3.4. Makine Öğrenme Algoritmaları (Machine Learning Algorithms)

#### 3.4.1. Lojistik regresyon (Logistic regression)

Lojistik Regresyon ikili bağımsız olmayan değişkenlerin olasılıksal tahmininde kullanılan bir yöntemdir. Amaç çıktı değişkenleri ile bağımsız değişkenler arasında anlamlı bir olasılığın varlığını sorgulamaktır. Temel olarak 0,5 değerinin altındaki olasılıklara 0, 0,5 değerinden fazla olasılıklara 1 değeri atanır [20].

#### 3.4.2. Karar Ağaçları (Decision Trees)

Karar ağaçları algoritması akış şeması benzeri hiyerarşik bir düzenle çalışan tahmin algoritmasıdır. DT yönteminde tahmin sınıfı etiketi dallanmanın başladığı yerden yani ağaç kökünden beslenir. Değer, öznelik ile karşılaştırılır ve o değere ait dala kadar takip edilerek sonraki bağlantı noktasına sıçrama yapılır. DT algoritmasında her bağlantı noktası bir sınıf etiketine sahiptir. Düğümlere bağlanan dallar öznelikleri etkileyen tüm verilerin sınıflandırılması bitene kadar devam eder. Sınıflama, şekillendirme ve analizi diğer yöntemlere göre daha kolay olduğu söylenebilir [10].

#### 3.4.3. Rasgele Orman (Random Forest)

Rasgele orman algoritması küçük boyutlu ve büyük boyutlu problemlerinde başarılı ve hızlı çalışabilen bir tekniktir. Rasgele orman algoritmasının en büyük avantajı regresyon ve sınıflandırma problemlerinde az parametre ile kullanılabilmesidir. Rasgele orman algoritmasında ağaç sayısının atırılmasından optimal ağaç sayısının belirlenmesi önemlidir. Sınıflandırma yaklaşımı açısından rasgele orman algoritması oluşturmak için düğümü durduracak bir karar gelene kadar her düğüm en az iki dala ayrılır. Düğümler bir eşik tarafından veri varyansı minimum olacak şekilde seçilir. Tüm bu dallar ve düğümler içinde ilerleyerek bir yerde tahmin yapılır [25].

#### 3.4.4. K En Yakın Komşu (K Nearest Neighbors)

K En Yakın Komşu algoritması regresyon ve sınıflandırma problemlerinde de kullanılabilir. Seçilen örneklem verilerinin en yakın komşularını etiketlemek suretiyle analiz yapar. Etiketleme yapıldıktan sonra girdilerin etiket değerlerine göre tahminde bulunur. KNN yönteminde girdinin eğitim verilerine uzaklığı çeşitli hesaplama yöntemleri ile yapılır. Ayrıca kullanıcı tarafından komşu verilerin sayısı da seçilebilmektedir. Henüz sınıflandırılmamış veri, veri seti içerisinde kendi konumunun diğer veriler ile mesafesine göre sınıflandırma işleminin yapılması fikrine dayanır. Yani test verisi hangi veri veya veriler yakın ise o veri veya veri gruplarının en uygun sınıfına yerleştirilir. KNN algoritmasında test verileri bulunur ancak eğitim verileri bulunmaz. Eğitim verileri sınıflandırıcı kabul edilerek değerlendirilir. Bu durum tüm yeni verilerin tekrar tekrar taranmasına tarama sürecinin uzamasına neden olur [26].

### 3.4.5. Naive Bayes

Naive Bayes yönteminde veriler olasılık değerlerine göre sınıflandırılır. Yöntem, verilerin olasılıksal değerlerini tek tek hesaplar ve bu değerlerin en yüksek olduğu duruma göre sınıflara atamalar yapar. Eğitim verisi çoğaldıkça isabetli sonuçlar alınma oranı da atmaktadır. Bu durum küçük veri setlerinde çalışmasına engel değildir. Dengesiz veri setlerinde rahat çalışabilmektedir [11]. NB sınıflandırmasında verilerin bir sınıfa dahil olduğu olasılığı ile hareket edilir. Bu olasılık hesaplandığında verinin hangi sınıfa dahil olduğu veya olabileceği anlaşılır. NB yöntemi pratik ve kolay anlaşılabilir yöntemdir. Uygulamada verilerin tek bir defa taranması yeterli olacaktır. Bu nedenle hızlı bir yöntem olduğu da söylenebilir [20].

### 3.4.6. Gradyan Artırıcı Sınıflandırıcı (Gradient Boosting)

Regresyon ve sınıflandırma problemlerinde kullanılabilen Gradyan artırıcı zayıf tahmin kalıplarının birleştirilerek karar ağaçlarının oluşturulduğu bir yöntemdir. Zayıf modeller hesaplanarak genel modelin hatasını azaltır [27]. Gradyan artırıcı genel olarak karar ağaçları gibi tahmin modellerinin geliştirilmesi üzerine kuruludur. Karar ağaçları düğümlerindeki tahminlere göre çıktılar veren düğüm ve dallardan oluşur. Daha stabil sonuçlar veren düğümleri dallandırarak sonuca ulaşır [28]. Karar düğümlerine dayalı olarak tahmine dayalı sonuçlar veren düğümlerden ve yapraklardan oluşur. Regresyon ağaçları, bireysel olarak zayıf modellerdir, ancak bir bütün olarak bakıldığında doğrulukları çok gelişmiştir. Bu nedenle, topluluklar kademeli olarak artan bir şekilde inşa edilir, öyle ki her topluluk önceki topluluktaki hatayı matematiksel olarak düzeltir.

### 3.4.7. XGBoost

XGBoost neredeyse tüm yapay zekâ problemlerinde kullanılabilen güçlü bir algoritmadır. Teknik aynı koşullarda hali hazırda kullanılan tüm algoritmalarından daha hızlı çalışır ve çok fazla örneği ölçkeleyebilir. XGBoost algoritmasının birçok avantajı vardır. Sık olmayan verileri analiz etmek için yeni bir algoritma mantığı geliştirmiştir. Paralel ve birbirinden ayrık bilgi sentezi yolu ile daha hızlı tahminlere yol açar. En önemli avantajlarından sayılan harici çekirdek hesaplamalarını kullanmasıdır. Bu sayede bir tek algoritma ile milyonlarca örneğin işlenmesinin yolu açılmış olmaktadır [29].

### 3.4.8. Torbalama (Bagging)

Bagging algoritması regresyon ve sınıflandırma problemlerinde kullanılabilir bir tekniktir. Aslen kullanılan yöntemlerin kararlılığını ve tahminlerin doğruluğunu artırmak için geliştirilmiştir. Sonuca ulaşmak için rastgele oluşturulan set sınıflarını birleştirir. Bagging algoritması basit mantığı ve yüksek doğru tahminleri ile öne çıkmaktadır. Bu yöntemin temel ilkesi zayıf dalların bir araya gelerek güçlü ağaçlar oluşturmasıdır. Eğitim setindeki tüm ağaçların her birine tekrar tekrar gidilir ve

her bir ağaç bir sınıfa yönelir. En yüksek yönelim alan sınıf sonuç olarak değerlendirilir [30].

## 3.5. Hiper Parametre Optimizasyonu (Hyperparameter Optimization)

Makine öğrenmesinde parametreler ve hiper parametreler ayrı kavramlardır ve birbiri ile karıştırmamak gerekir. Parametreler algoritma işleyişi içinde öğrenilir, tahmin edilir ve bu değerler makine öğrenmesi eğitimi devam ederken güncellenir. Algoritma eğitimi bitiğinde ise parametreler modelin bir parçasını oluşturur. Tahmin modeline ait parametreleri bulmak için ise hiper parametre değerleri kullanılır. Hiper parametreler algoritmaya özgü verilerdir, bu nedenle bu değerleri verilerden yola çıkarak hesaplanamaz. Hiper parametreler algoritma eğitimi içinde tahmin edilemeyen veya öğrenilemeyen parametrelerdir. Farklı veri setlerinde farklı hiper parametreler elde edilebilir. Hiper parametre ayarları makine öğrenimini kontrollü bir şekilde yapmanın gerekli bir bölümüdür. Doğru biçimde ayarlanmayan parametreler modelimizin yetersiz sonuçlar üretmesine neden olacaktır.

### 3.5.1. Izgara Arama (Grid Search)

Izgara arama belirlenen alt ve üst sınırlarda ve belirli steplerde kapsamlı parametre belirleyicidir. Belirlenen adımlarda tüm kombinasyonlar değerlendirilir ve değerlendirme sonucuna göre en verimli parametreler elde edilmeye çalışılır. Izgara arama uygulanması kolay bir yöntemdir ancak kapsamlı tarama yaptığı için büyük veri setlerinde maliyet artar.

### 3.5.2. Yarım Izgara Arama (Halving Grid Search)

Yarım izgara arama çalışma mantığı Izgara Arama tekniğinden esinlenir ancak Izgara Arama'ya göre çok daha verimli çalışan bir parametre belirleme yöntemidir. İyi performans gösteremeyen veriler filtrelenerek daha iyi adaylar üzerinde çalışan ve değerlendirme yapan tekniktir. Yarım izgara arama özetle minimal seviyede kaynakla aramaya başlayıp en yüksek düzeyde yineleme yaparak en doğru verileri seçimini yapar.

### 3.5.3. Rasgele Arama (Random Search)

Rasgele arama, oluşturulan kılavuzda rasgele değerlerde arama yaparak performansı en çok artıran değerleri belirlemeye çalışır. Her tekrarda hiper parametrelerin bir kombinasyonunu rasgele belirler ve modelin verimliliğini tutar. Yapılan tekrarlar neticesinde en iyi sonucu veren parametre grubunu belirler.

### 3.5.4. Bayes Optimizasyonu (Bayesian Optimization)

Bayes optimizasyon tekniği rasgele ve grid arama tekniklerinin aksine belirli bir standartla modeli sonuçlandırmak yerine performansı en çok artıran değerlerin bulunduğu alana yoğunlaşarak doğruluğu yüksek olasılıklı modeller oluşturmayı hedefler. Her veri

performansı bir sonraki veri girdisinin düzeyini güncelleyerek devam eder. Bu sayede daha hızlı ve verimli çalışmaktadır.

### 3.6. Skor Açıklamaları (Score Descriptions)

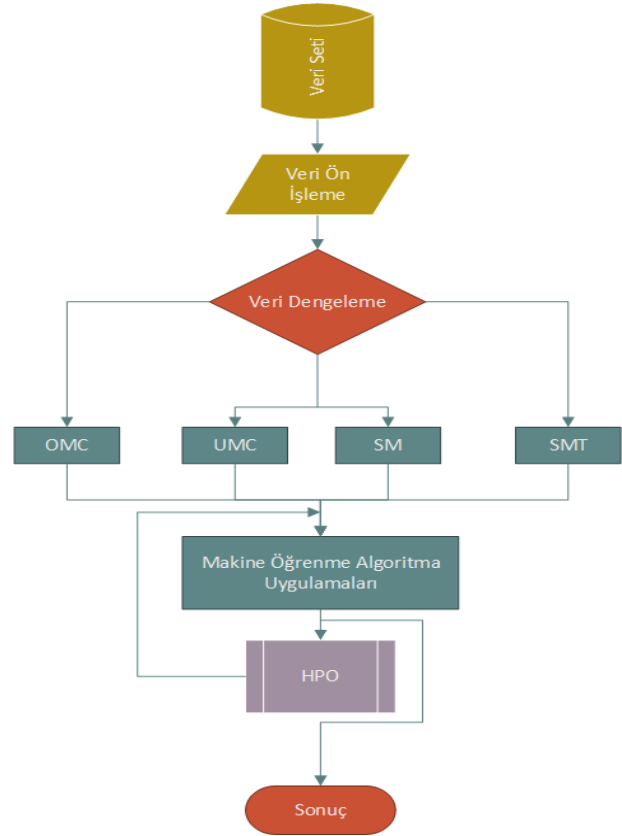
Tablo 2’te algoritma değerlendirmelerinde kullanılan metrikler hakkında bilgi verilmiştir.

Tablo 2. Değerlendirme metrikleri (Evaluation metrics)

Skor	Açıklama
Accuracy	Doğruluk skoru, doğru sınıflandırılmış verinin toplam veriye oranını temsil eder.
Precision	Kesinlik skoru, pozitif tahminlerin ne kadarının gerçek pozitif olduğunu gösterir. Doğru pozitif tahminlerin, doğru pozitif ve yanlış pozitif tahmin toplamına oranıdır.
Recall	Geri çağırma skoru, tüm pozitif veriler içinden ne kadar doğru pozitif tahmin edildiğini gösterir. Doğru pozitif tahminlerin doğru pozitif ve yanlış negatif tahmin toplamına oranıdır.
F1 Score	F1 skoru kesinlik ve geri çağırma skorlarını birleştiren iyi bir ölçüttür. Genel olarak doğruluk skorundan daha güvenilir bir sonuçtur. F1 skoru Precision ve Recall skorlarının harmonik ortalamasıdır.
Cross Validation (CV)	Çapraz doğrulama skoru uygulanan diğer skorları test etmenin, doğruluğunun hesaplanmasının bir yoludur. Çapraz doğrulamada test ve eğitim setleri yeniden örneklenecek farklı veri seti oluşturulur ve bu veri setlerinde test edilen skor tekrar uygulanır. Bu çalışmada çapraz doğrulama skoru F1 skorunu test etmek için kullanıldı.

## 4. BULGULAR ve TARTIŞMA (FINDINGS and DISCUSSION)

Çalışmanın yapıldığı bilgisayar Intel Core i7-9750H 2,60GHz CPU, 16GB Ram, NVIDIA GeForce GTX 1650 (4GB), 512 GB SSD donanım özelliklerine sahiptir. Bilgisayarda Windows 10 Pro 64 Bit işletim sistemi bulunmaktadır. Şekil 7’de verilen şema tüm çalışmanın özeti niteliğindedir.



Şekil 6. Tüm çalışmanın akış şeması (Flow chart of the whole study)

Çalışmada makine öğrenmesi algoritmalarını çalıştırmak için Python dili tercih edilmiştir. Algoritmalar Anaconda3 Navigatör Jupyter derleyici ile derlenmiştir. Çalışmada veri kümeleri ile çalışmak için Pandas kütüphanesi, diziler ile çalışabilmek için Numpy kütüphanesi, etkileşimli görseller oluşturmak için Matplotlib kütüphanesi ve istatistiksel grafikler için Seaborn kütüphanesi, istatistiksel modelleme ve makina öğreniminde kullanılan birçok aracı barındıran Sklearn kütüphanesi, veri eşitleme yöntemleri barındıran Imblearn kütüphanesi kullanılmıştır.

Çalışmada bazı kısaltmalar kullanılmıştır. Bu kısaltmalar Tablo 3’te verilmiştir.

Tablo 3. Kısaltmalar (Abbreviations)

Kısaltma	Tanımı	Grubu
OMC	Oversample Minority Class	Dengeleme algoritmaları
SM	SMOTE	
SMT	SMOTETomek	
UMC	Undersample Majority Class	
BAG	Bagging	Yapay zekâ algoritmaları
DT	Decision Trees	
GB	Gradient Boosting	
KNN	K Nearest Neighbors	
LR	Logistic Regression	
NB	Naive Bayes	
SVM	Destek Vektör Makinası	
RF	Random Forest	
XGB	XGBoost	Hiper parametre optimizasyonu algoritmaları
HPO	Hiper parametre Optimizasyonu	
RS	Random Search	
BO	Bayesian Optimization	

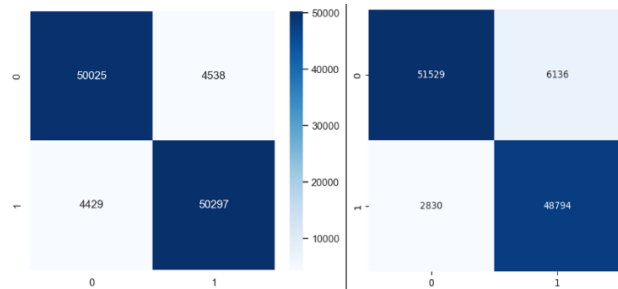
Makine öğrenme algoritmaları veri dengeleme teknikleri ile dengelenerek oluşturulan veri setleri ile eğitilmiş ve sonuçlar bu bölümde verilmiştir. HPO uygulamaları bu bölüm başlığı altında incelenmiştir.

#### 4.1. SM ile Dengelenmiş Veri Seti Algoritma Sonuçları (SM Balanced Dataset Algorithm Results)

Tablo 4'te SM ile dengelenmiş veri setine uygulanan tahmin algoritmalarına ait sonuçlar gösterilmiştir. F1 skorlarında en iyi sonuçların %92 değeri ile RF ve XGB algoritmalarının olduğu görüldü. RF algoritmasında tüm skorlar yakın değerler alırken XGB Recall skoru diğer skorlardan daha düşük olduğu görülmektedir. Bu durum karışıklık matrisinde daha açık görülmektedir. Şekil 8'de görüleceği üzere XGB algoritmasında tüm pozitif veriler içinde doğru tahmin edilen değerler düşük, yanlış tahmin edilen negatifler daha yüksek çıkmıştır.

Tablo 4. SM uygulanan veri setine ait algoritma skorları (Algorithm scores of the SM applied data set)

Algoritma	Accuracy	Precision	Recall	F1-Score	CV Mean
LR	%75	%74	%78	%76	%75
DT	%89	%89	%89	%89	%75
<b>RF</b>	<b>%92</b>	<b>%92</b>	<b>%92</b>	<b>%92</b>	<b>%92</b>
KNN	%86	%81	%94	%87	%85
NB	%71	%74	%66	%70	%71
GB	%87	%88	%87	%87	%87
<b>XGB</b>	<b>%92</b>	<b>%95</b>	<b>%89</b>	<b>%92</b>	<b>%92</b>
BAG(DT)	%91	%93	%89	%91	%91



Şekil 7. SM veri seti RF Karışıklık Matrisi (solda) (SM dataset RF Confusion Matrix (on the left)), SM veri seti XGB Karışıklık matrisi (sağda) (SM dataset XGB Confusion Matrix (on the right))

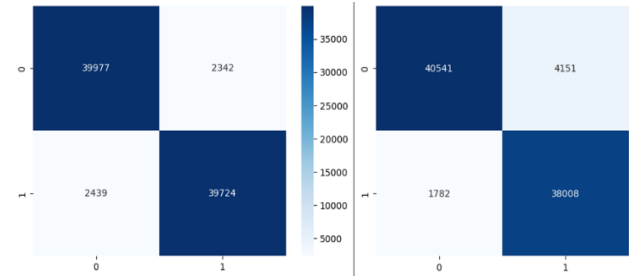
#### 4.2. SMT ile Dengelenmiş Veri Seti Algoritma Sonuçları (SMT Balanced Dataset Algorithm Results)

Tablo 5'te verilen skor değerlerine göre F1 skorlarında %94 değeri ile RF ve %92 değeri ile XGB'nin en iyi tahmin başarısı elde ettiği söylenebilir. SMT veri dengeleme yöntemi ile oluşturulan veri setinde RF algoritmasında tüm skorlar yakın değerler alırken XGB Recall skoru diğer skorlardan daha düşük olduğu görülmektedir. Şekil 9'da RF algoritmasının yanlış tahmin edilen 0 ve 1 sınıflarının sayılarının yakın olduğu görülmektedir. RF algoritmasının daha dengeli tahmin yaptığı çıkarımında bulunabiliriz. Şekil 9'da XGB'nin 0 sınıfı olarak tahmin edilen ancak 1 sınıfına ait olan 4151 değerinin yanlış tahmin edildiğini görülmüyor. Buna karşılık 1 sınıfı olarak tahmin edilen 0 sınıfına ait değerlerin 1782

olduğu görülüyor. Bu durumda SMT dengeleme sisteminin çoğaltma işlemi sonucunda çoğalttığı 1 sınıfını 0 sınıfına çok yaklaştırdığı ve yanlış tahminlere neden olduğu söylenebilir.

Tablo 5. SMT uygulanan veri setine ait algoritma skorları (Algorithm scores of the SMT applied data set)

Algoritma	Accuracy	Precision	Recall	F1-Score	CV mean
LR	%76	%74	%79	%77	%76
DT	%91	%91	%91	%91	%90
<b>RF</b>	<b>%94</b>	<b>%94</b>	<b>%94</b>	<b>%94</b>	<b>%94</b>
KNN	%87	%82	%95	%88	%86
NB	%72	%74	%68	%71	%72
GB	%88	%89	%87	%88	%88
<b>XGB</b>	<b>%93</b>	<b>%96</b>	<b>%90</b>	<b>%93</b>	<b>%93</b>
BAG(DT)	%93	%95	%91	%93	%92



Şekil 8. SMT veri seti RF Karışıklık Matrisi (solda) (SMT dataset RF Confusion Matrix (on the left)), SMT veri seti XGB Karışıklık matrisi (sağda) (SMT dataset XGB Confusion Matrix (on the right))

SMOTE tabanlı dengelenmiş (SM ve SMT) veri setlerinde XGB algoritmasının dengelenmeden önce çoğunluk sınıfı olan negatif (0) sınıfını daha iyi sınıflandırdığı, dengelenmeden önce azınlık sınıfı olup yapay olarak çoğaltılmış pozitif (1) sınıfını (0) sınıfına göre daha kötü sınıflandırdığı görülmüştür. RF algoritmasının ise azınlık veya çoğunluk sınıfı farkı olmaksızın negatif ve pozitif sınıflarını eşit doğrulukta sınıflandırdığı dikkat çekmektedir. XGB algoritmasında kalp hastası olup hasta değil olarak sınıflandırılan kişi sayısı daha fazladır. Kalp hastası olduğu halde kalp hastası olmadığı yönünde yapılan sınıflandırmanın insan hayatı açısından daha riskli bir durum oluşturduğu için SM ve SMT dengelenmiş veri setlerinde RF algoritması XGB algoritmasına daha göre öncelikli tercih sebebidir.

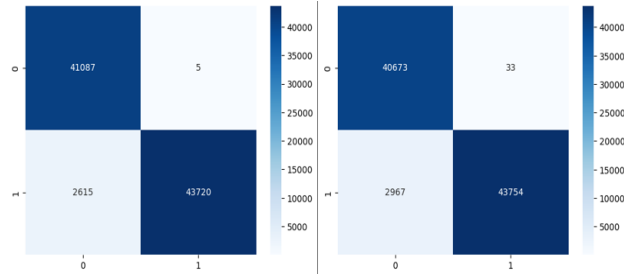
#### 4.3. OMC ile Dengelenmiş Veri Seti Algoritma Sonuçları (OMC Balanced Dataset Algorithm Results)

Tablo 6'da F1 skorlarında en iyi tahmin olasılıklarını %97 değeri ile RF ve %97 değeri ile BAG algoritmalarının verdiği görülmüştür. BAG algoritması tahmin edici(base\_estimator) değeri varsayılan DT algoritması olarak ayarlanmıştır. Recall skorlarının %100'e çok yakın olması algoritmanın aşırı öğrenme yapmış olabileceğini akla getirmektedir. Çünkü OMC dengeleme sisteminde azınlık sınıfı olan '1' sınıfının rasgele kopyalanarak çoğaltılması söz konusudur. Veri setinin dengelenmeden önceki halinde '0' sınıfı 273.222 örnek içerirken '1' sınıfı 27.048 örnek içeriyor. OMC uygulandıktan sonra her iki

sınıf 218.566 örnekte dengelenmiştir. Bu durumda azınlık sınıfı yaklaşık 8 katı kadar çoğaltılmıştır. Dengesizlik oranı yüksek olduğu için veri setinin test kısmı hangi oranda seçilirse seçilsin eğitim seti içerisinde test seti verileri olacaktır. Bu nedenle algoritmaların öğrenme yapmak yerine aşırı öğrenme yapmış olabileceğini söyleyebiliriz. Şekil 10'da karışıklık matrisi verileri bu savı destekler niteliktedir. Hem RF hem de BAG algoritması için SMOTE tabanlı dengeleme sistemlerinin aksine OMC dengeleme sisteminde rasgele çoğaltılan "1" sınıfının hatalı tahminlerinin az, aksine "0" sınıfının yanlış tahminlerinin daha fazla olduğu Şekil 10'da görülmektedir.

Tablo 6. OMC uygulanan veri setine ait algoritma skorları (Algorithm scores of the OMC applied data set)

Algoritma	Accuracy	Precision	Recall	F1-Score	CV mean
LR	%75	%74	%76	%75	%75
DT	%95	%91	%100	%95	%94
<b>RF</b>	<b>%97</b>	<b>%94</b>	<b>%100</b>	<b>%97</b>	<b>%96</b>
KNN	%89	%82	%99	%90	%87
NB	%70	%74	%64	%69	%70
GB	%76	%74	%81	%77	%76
XGB	%79	%76	%83	%80	%78
<b>BAG(DT)</b>	<b>%97</b>	<b>%94</b>	<b>%100</b>	<b>%97</b>	<b>%96</b>



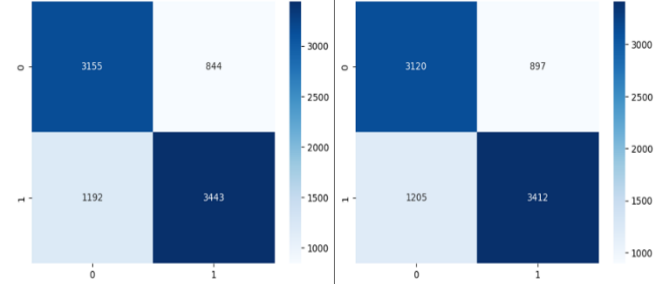
Şekil 9. OMC veri seti RF Karışıklık Matrisi (solda) (OMC dataset RF Confusion Matrix (on the left)), OMC veri seti BAG Karışıklık matrisi (sağda) (OMC dataset BAG Confusion Matrix (on the right))

#### 4.4. UMC ile Dengelenmiş Veri Seti Algoritma Sonuçları (UMC Balanced Dataset Algorithm Results)

UMC ile dengelenmiş veri setinde dikkat çeken ilk ayrıntı Tablo 7'de görüldüğü gibi diğer veri dengeleme teknikleri ile oluşturulan veri setlerine göre algoritma uygulamalarında düşük skorlar elde edilmiş olmasıdır. UMC veri setinde yaşanan veri kaybının düşük skorlara neden olduğu düşünülmektedir. Zira UMC uygulanmadan önce 300.270 satır örnek barındıran orijinal veri seti UMC uygulandıktan sonra 43.168 örneğe düşmüştür. Bu durum kalp hastalığı tahmininde önemli ölçüde performans kaybına neden olmuştur. Tablo 7'de görüldüğü gibi F1 skorlarında en iyi tahmin başarısı gösteren algoritmalar %77 ile GB ve %76 ile XGB olmuştur. Şekil 11'de UMC setine ait en iyi algoritmanın karışıklık matrisi verilmiştir. Kalp hastası olduğu halde kalp hastası değil olarak etiketlenen Yanlış negatif değerlerinin toplam veri içindeki oranı yüksek olduğu görülmektedir.

Tablo 7. UMC uygulanan veri setine ait algoritma skorları (Algorithm scores of the UMC applied data set)

Algoritma	Accuracy	Precision	Recall	F1-Score	CV mean
LR	%75	%75	%76	%76	%75
DT	%68	%68	%67	%67	%67
RF	%75	%74	%77	%76	%75
KNN	%72	%72	%73	%72	%72
NB	%70	%73	%63	%68	%71
<b>GB</b>	<b>%76</b>	<b>%74</b>	<b>%80</b>	<b>%77</b>	<b>%76</b>
<b>XGB</b>	<b>%76</b>	<b>%74</b>	<b>%79</b>	<b>%76</b>	<b>%75</b>
BAG(DT)	%73	%74	%70	%72	%72



Şekil 10. UMC veri seti GB Karışıklık Matrisi (solda) (UMC dataset GB Confusion Matrix (on the left)), UMC veri seti XGB Karışıklık matrisi (sağda) (UMC dataset XGB Confusion Matrix (on the right))

Tüm veri dengeleme teknikleri (SM, SMT, OMC, UMC) ile oluşturulan veri setleri ile eğitilen makine öğrenme algoritmaları değerlendirildikten sonra her dengelenmiş veri seti için en yüksek skoru veren iki ayrı algoritma seçilmiştir. Bu algoritmalara Hiper parametre optimizasyonu uygulanmıştır. Buna göre SM dengeleme için RF ve XGB, SMT dengeleme için RF ve XGB, OMC dengeleme için RF ve BAG (DT tabanlı), UMC dengeleme için XGB ve GB algoritmaları seçilmiştir.

Hiper parametre optimizasyonu işlemi için iki ayrı optimizasyon tekniği seçilmiştir. Bu tekniklerden diğer arama tekniklerine göre daha hızlı arama yaparak maliyeti azalttığı için Random Search (RS) ve öğrenme tabanlı ve verimli çalıştığı için Bayesian Optimization (BO) tercih edilmiştir. Algoritmaların hiper parametre yapılandırma değerleri Tablo 8'de gösterildiği gibi belirlenerek her iki arama tekniği için uygulama yapılmıştır. Tablo 9'da Hiper parametre optimizasyonu uygulama sonucunda elde edilen optimum hiper parametreler ve bu hiper parametreler ile makine öğrenme algoritmalarının tekrar uygulama sonuçları gösterilmiştir. Buna göre SM ve SMT veri setlerinde RF ve XGB algoritmaları için BO'un daha iyi sonuçları verdiği için bu teknik tercih edilmiştir. Ayrıca, OMC veri setinde RF algoritması için BO, BAG algoritması için RS tercih edilmiş ve UMC veri setinde BO tercih edilmiştir. Bu işlem sonucunda, BO tekniğinin RS tekniğine göre daha iyi sonuçlar verdiği görülmüştür.

Tablo 8. Hiper parametre yapılandırma değerleri (Recommended hyperparameter values)

	Hiper parametreler	Önerilen hiper parametre değerleri
RF	max_depth	[10, 20, 30, 40, 50]
	min_samples_leaf	[1, 2, 4, 6]
	min_samples_split	[2, 5, 10, 15]
	n_estimators	[200, 400, 600, 800]
XGB	colsample_bytree	[0,3,0,4,0,5,0,7],

	learning_rate	[0,05,0,10,0,15,0,20,0,25,0,30],
	max_depth	[3,4,5,6,8,10,12,15],
	min_child_weight	[1,3,5,7],
	gamma	[0,0,0,1,0,2,0,3,0,4]
BAG	bootstrap	[True, False]
	bootstrap_features	[True, False]
	n_estimators	[5, 10, 15],

	max_samples	[0,6, 0,8, 1,0]
GB	n_estimators	[5,50,150,250,500],
	max_depth	[1,3,5,7,9],
	learning_rate	[0,0001, 0,001, 0,01, 0,1, 1]

Tablo 9. Optimize edilmiş parametrelerle uygulama sonuçları (Application results with optimized parameters)

Dengeleme Sistemi	Algoritma	HPO Tekniği	En iyi Parametreler	Skor
SM	RF	BO	max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=800	Train Score: %98 Test Score: %92
		RS	max_depth=40, min_samples_leaf=1, min_samples_split=10, n_estimators=600	Train Score: %97 Test Score: %92
	XGB	BO	colsample_bytree=0,7, gamma=0, max_depth=15, learning_rate=0,3, min_child_weight=1	Train Score: %97 Test Score: %94
		RS	colsample_bytree=0,3, gamma=0,1, max_depth=15, learning_rate=0,15, min_child_weight=5	Train Score: %93 Test Score: %92
SMT	RF	BO	max_depth=40, min_samples_leaf=1, _samples_split=2, n_estimators=800	Train Score: %100 Test Score: %94
		RS	max_depth=40, min_samples_leaf=1, min_samples_split=2, n_estimators=800	Train Score: %100 Test Score: %94
	XGB	BO	colsample_bytree=0,7, gamma=0,2, max_depth=15, learning_rate=0,3, min_child_weight=1	Train Score: %98 Test Score: %94
		RS	min_child_weight=1, max_depth=15, gamma=0,4, learning_rate=0,1, colsample_bytree=0,4	Train Score: %95 Test Score: %93
OMC	RF	BO	max_depth=50, min_samples_leaf=1, min_samples_split=2, n_estimators=600	Train Score: %100 Test Score: %97
		RS	max_depth=50, min_samples_leaf=2, min_samples_split=5, n_estimators=600	Train Score: %98 Test Score: %94
	BAG	BO	bootstrap=False, bootstrap_features=True, _samples=1,0, n_estimators=15	Train Score: %98 Test Score: %96
		RS	bootstrap=False, bootstrap_features=True, _samples=1,0, n_estimators=10	Train Score: %99 Test Score: %98
UMC	XGB	BO	colsample_bytree=0,4, gamma=0,1, max_depth=4, learning_rate=0,15, min_child_weight=1	Train Score: %77 Test Score: %76
		RS	min_child_weight=5, max_depth=5, gamma=0,3, learning_rate=0,1, colsample_bytree=0,3	Train Score: %77 Test Score: %76
	GB	BO	learning_rate=0,1, max_depth=3, n_estimators=150	Train Score: %77 Test Score: %76
		RS	learning_rate=1, max_depth=1, n_estimators=500	Train Score: %76 Test Score: %76

Hiper parametre optimizasyon teknikleri uygulanıp en iyi hiper parametreler bulunduğundan sonra tüm veri setlerine makine öğrenme algoritmaları en iyi hiper parametre ile tekrar uygulandı ve sonuçlar Tablo 10'da gösterildi. Çıkan sonuçlar karşılaştırılarak bu çalışma için en iyi algoritma ve tekniklere karar verildi. SM ile dengelenmiş veri setinde HPO'dan sonra modellerin performansının arttığı görülmektedir. SMT veri dengeleme tekniği ile oluşturulan veri setinde ise RF algoritmasının

performansı değişmezken XGB algoritmasının F1 skorda performansı artmıştır. OMC veri dengeleme tekniği ile oluşturulan veri setinde model skorlarının HPO ile değişmediği görülmüştür. UMC veri dengeleme tekniği ile oluşturulan veri setinde yine HPO'nun anlamlı bir fark yaratmadığı görülmüştür. Bu değerler skorların tam kısımları olmakla birlikte ondalık kısmında küçük farkların olduğu gözlenmiştir. Genel olarak HPO ile model performanslarının arttığı sonucuna ulaşılmıştır.

Tablo 10. En iyi hiper parametreler ile tekrar uygulama sonuçları (Reapplication results with the best hyperparameters)

Dengeleme Sistemi	Skor	RF		XGB		Dengeleme Sistemi	Skor	RF		XGB	
		A	B	A	B			A	B	A	B
SM	Train (%)	100	98	92	97	SMT	Train (%)	100	100	93	98
	Test (%)	92	93	92	94		Test (%)	94	94	93	94
	CV (%)	92	92	92	93		CV (%)	94	94	93	94
	Accuracy (%)	92	93	92	94		Accuracy (%)	94	94	93	94
	F1 (%)	92	93	92	94		F1 (%)	94	94	93	94
Dengeleme Sistemi	Skor	RF		BAG		Dengeleme Sistemi	Skor	XGB		GB	
		A	B	A	B			A	B	A	B

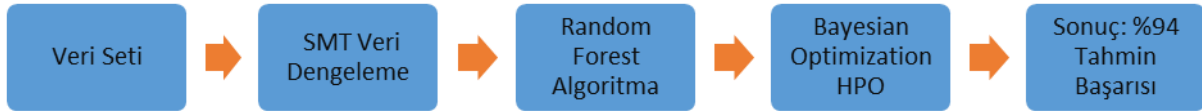
OMC	Train (%)	100	100	100	99	UMC	Train (%)	81	77	77	77
	Test (%)	97	97	97	97		Test (%)	76	76	76	76
	CV (%)	96	96	96	97		CV (%)	75	76	76	76
	Accuracy (%)	97	97	97	97		Accuracy (%)	76	76	76	76
	F1 (%)	97	97	97	97		F1 (%)	77	77	76	76

A: HPO'den önceki skor (Score before HPO)

B: HPO'den sonraki skor (Score after HPO)

SM veri dengeleme yöntemi ile oluşturulan veri setinde XGB %94, RF %93 tahmin, SMT veri dengeleme yöntemi ile oluşturulan veri setinde XGB ve RF %94 tahmin, OMC veri dengeleme yöntemi ile oluşturulan veri setinde RF ve BAG %97 tahmin ve UMC veri dengeleme yöntemi ile oluşturulan veri setinde XGB %77 tahmin başarıları elde edilmiştir. Buradan UMC ile dengelenmiş veri setinde yaşanan kayıplardan dolayı performans düşüklüğü

yaşanmış ve bu veri seti elenmiştir. OMC veri setinde yaşanan aşırı öğrenme nedeniyle bu veri seti elenmiştir. SM ve SMT veri setlerinde XGB algoritmasının Recall skor sonuçlarının tatmin edici olmaması nedeniyle XGB algoritması elenmiştir. Bu çalışmada RF makine öğrenme algoritmasının en güvenilir skoru verdiği sonucuna ulaşılmıştır. Sonuca ulaşmak için yapılan işlem Şekil 12'de gösterilmiştir.



Şekil 11. Bu çalışmada en verimli tahmin algoritması için iş akışı (This study workflow for the best prediction algorithm)

#### 4.5. Benzer Çalışma Sonuçları ile Karşılaştırma (Comparison with Other Study Results)

Bu bölümde kalp hastalıklarının tahmini ile ilgili benzer çalışmalara ilişki karşılaştırma yapılmıştır. Diğer araştırmacıların en iyi sonucu veren algoritmaları ile bu çalışmadaki aynı algoritmanın en iyi skoru karşılaştırılarak Tablo 11'de gösterilmiştir. Bu çalışma ile kalp hastalıklarının tahmin edilmesinde verimli sonuçlar veren modeller elde edilmiştir. Bu çalışmanın diğer çalışmalara göre üstünlükleri de bulunmaktadır. Model skorları karşılaştırıldığında bu çalışmanın benzer çalışmalara göre üstün olduğu göze çarpmaktadır. Tablo 11'deki veriler de bu bilgiyi desteklemektedir. Model skorları ile birlikte bu çalışmada model performanslarını etkileyen veri dengeleme teknikleri ve hiper parametre optimizasyonu teknikleri incelenmiştir. Model performanslarını etkileyen yöntemleri karşılaştırmalı olarak inceleyerek benzer çalışmalardan farklı özgün bir kaynak ortaya koymuş olduk.

Tablo 11. Benzer çalışmaları ile karşılaştırma tablosu  
(Table of comparison with the work of others)

Çalışmalar	RF-OMC	NB	DT	LR
<b>Bu Çalışmadaki Sonuçlar</b>	%97	%71	%95	%77
Anitha ve ark. [7]	-	%86,6	-	-
Coşar ve ark.[31]	%88	-	-	-
Rajdhan ve ark. [10]	%90,2	-	-	-
Mertcan Görgün [11]	%90,2	-	-	-
Repaka ve ark. [12]	%89,8	-	-	-
Ramalingam ve ark. [14]	%83,5	-	-	-
Kamil [16]	%98	-	-	-
Salman [32]	-	%86	-	-
Konda ve ark. [18]	-	-	%89	-
Tarawneh ve ark [19]	-	%86	-	-
Venkatesh ve ark. [2]	-	%97	-	-
Çil [20]	-	-	-	%90,8
Sağlain ve ark. [33]	-	%86,7	-	-
Taşçı ve ark. [34]	-	%88	-	-
Ekrem ve ark. [32]	%87	-	-	-
Gündoğdu [35]	%89,7	-	-	-

#### 5. SINIRLILIKLAR (LIMITATIONS)

Veri seti Amerika Birleşik Devletinde yapıldığı için hem yaşam biçimi hem de bölgesel farklılık gösterebilecek bazı özneliklerin çalışmayı kısıtladığı söylenebilir. "İrk" özneliği çok fazla çeşitlendirilmediği için Amerika kıtası dışında kullanımı bu özneliğin çalışmayı kısıtladığı söylenebilir. Veri setinin büyük boyutlu olması bazı algoritmaların uygulanmasını zorlaştırmaktadır. Destek Vektör Makinası veri setine uygulanmış ancak günlerce süren işlemlerden sonuç alınamadığı için bu çalışmadan çıkarılmıştır.

#### 6. SONUÇLAR (CONCLUSION)

Bu çalışma, farklı makine öğrenmesi algoritmalarının verimini artırmak üzere uygulanan veri düzenleme ve hiper parametre optimizasyonu tekniklerinin karşılaştırılarak kalp hastalığı tahmininde en verimli modelin ortaya konduğu bir çerçeve sunmaktadır. Çalışmadaki asıl amaç mevcut veri seti üzerinden kalp hastalığı tahmini yapan en verimli algoritmanın belirlenmesidir. Bununla birlikte veri dengeleme ve hiper parametre optimizasyonunun kalp hastalığında kullanılan makine öğrenme algoritmaları üzerindeki etkisi ikinci amaç olarak belirlenmiştir. Bunun için mevcut veri seti temizlenerek, dört farklı veri dengeleme tekniği ile (SM, SMT, OMC, UMC) oluşturulmuş her veri setine ayrı ayrı tahmin algoritmaları uygulanmıştır. Hiper parametre optimizasyonu ile en iyi hiper parametreler bulunarak bu parametreler ile tahmin algoritmaları tekrar eğitilmiştir. XGB algoritması SM ve SMT dengelenmiş veri setlerinde %94 tahmin başarıları elde etmiştir. SM ve SMT veri dengeleme yöntemleri ile oluşturulan veri setlerinde XGB algoritması ile yapılan tahminlerde kalp hastası olduğu halde kalp hastası değil olarak sınıflandırılan örneklerin daha fazla olduğu ve bu durumun insan hayatı açısından riskli bir durum olduğu değerlendirilmiştir. Bu nedenle XGB algoritması bu çalışma için en iyi algoritma olmadığı yönünde



değerlendirme yapılmıştır. OMC algoritması ile dengelenen veri setinde RF ve XGB algoritmaları çoğaltılan azınlık sınıfını orantısız bir şekilde, yüksek doğrulukta tahmin ettiği görülmüştür. Veri setinin orijinal hali çok dengesiz olduğu ve OMC dengeleme tekniğinin rasgele kopyalama yolu dengeleme yaptığı için bu veri seti ile eğitilen tahmin algoritmalarının aşırı öğrenme yaptığı çıkarımında bulunulmuştur. UMC ile dengelenmiş veri setinde veri kaybı çok fazla olduğu için algoritma skorları yüksek çıkmamıştır. UMC dengeleme setinin büyük ve dengesiz veri setlerinde başarılı olmadığını sonucuna ulaşılmıştır. RF algoritması SM dengelenmiş veri setinde %93, SMT dengelenmiş veri setinde %94 ve OMC dengelenmiş veri setinde %97 başarı elde etmiştir. RF algoritmasının her ne kadar OMC ile dengelenmiş veri setinde aşırı öğrenmeye yöneldiği çıkarımı yapılsa da RF algoritmasının tüm veri setlerinde diğer tahmin algoritmalarına oranla yüksek tahmin başarısı gösterdiği sonucuna ulaşılmıştır. SM ve SMT veri dengeleme teknikleri arasında çok fazla fark olmamakla birlikte SMT tekniği ile oluşturulan veri setleri ile eğitilen modellerin daha verimli sonuçlar çıkardığı görülmüştür. RF ve XGB algoritmalarının LR, DT, KNN, NB, GB ve BAG algoritmalarından daha iyi performans gösterdiği görülmüştür. Bu çalışma farklı veri setleri ve HPO sonuçlarına göre için en iyi sonuç veren algoritmanın RF algoritması olduğuna karar verilmiştir. Model performansları göz önüne alınarak en sağlıklı sonucu SMT veri dengeleme tekniği ile oluşturulan veri setine uygulanan RF algoritmasının %94 tahmin başarısı ile verdiği söylenebilir. Bu çalışma kalp hastalıklarının tahmin edilmesinde makine öğrenme algoritmalarının kullanılabilirliğini ve çeşitli iyileştirme kombinasyonları ile tahmin performansının artırılabilirliğini ortaya koymuştur. Bu çalışmanın bundan sonraki çalışmalara ışık tutarak makine öğrenme tekniklerinin kalp hastalıklarının teşhisinde ön tanı kriteri olarak kullanılabilirliğini göstermiştir.

Yapılan bu çalışma veri dengeleme teknikleri ve HPO'nun kalp hastalığının tahmin edilmesinde kullanılan makine öğrenme tekniklerinin performansını etkilediğini ortaya koymuştur. Farklı tekniklerin bir arada kullanılması ile daha verimli modellerin oluşturulabileceğini göstermiştir. Bu çalışma sonraki çalışmalar için kaynak niteliğinde olup geliştirilmeye de açıktır. Farklı veri setleri ile çalışılabilirliği gibi bu çalışmada kullanılan veri seti için birtakım iyileştirmeler yapılabilir. Genetik algoritması gibi sezgisel algoritmalar kullanılarak farklı hiper parametre optimizasyonu yöntemleri denenebilir. Öznitelik seçimi gibi yöntemlerle maliyet azaltılabilir. Ayrıca derin öğrenme teknikleri ile kalp hastalıklarının tespitinde farklı modeller oluşturulabilir.

*Etik Kurul Onayı ve Çıkar Çatışması Beyanı (Ethics Committee Approval and Conflict of Interest Statement)*

Hazırlanan makalede etik kurul izni alınmasına gerek yoktur. Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

### *Yazarların Katkıları (Authors' Contributions)*

SUNGUR, Araştırma yapmış, deneyleri gerçekleştirip yorumlamış, alıntı olmayan görsel, tablo ve şekilleri oluşturmuş, makalenin yazım işlemini yapmıştır.

BAKIR, Sonuçları yorumlamış, analiz yapmış, gerekli düzenleme, şekillendirme işlemlerini yaparak makalenin yazım işleminde müşavirlik yapmıştır.

### **KAYNAKLAR (REFERENCES)**

- [1] "Kardiyovasküler Hastalıklar." Jan. 2021. [Online]. Available: <https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [2] R. Venkatesh, C. Balasubramanian, and M. Kaliappan, "Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique," J Med Syst, vol. 43, no. 8, Jan. 2019, doi: 10.1007/s10916-019-1398-y.
- [3] A. L. Yadav, K. Soni, and S. Khare, "Heart Diseases Prediction using Machine Learning," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023, pp. 1-7. doi: 10.1109/ICCCNT56998.2023.10306469.
- [4] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, 2023, doi: 10.3390/a16020088.
- [5] A. Özdemir, "Makine Öğrenmesi Algoritmaları ile Aritmilerin Sınıflandırılması," Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi, vol. 39, no. 3, pp. 394-402, 2023.
- [6] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," Processes, vol. 11, no. 4, 2023, doi: 10.3390/pr11041210.
- [7] S. Anitha and N. Sridevi, "Heart Disease Prediction Using Data Mining Techniques," Journal of Analysis and Computation, vol. 13, no. 2, 2019, [Online]. Available: [www.ijaonline.com](http://www.ijaonline.com).
- [8] M. E. Çifci, "Kalp Hastalıklarında Kullanılan Yapay Zekâ Teknikleri Ve Uygulamaları." 2019.
- [9] P. Kamat and M. C. Beyene, "Survey on prediction and analysis the occurrence of heart disease using data mining techniques," International Journal of Pure and Applied Mathematics, vol. 18, no. 8, 2018, [Online]. Available: <https://www.researchgate.net/publication/323277772>
- [10] A. Rajdhan, A. Agarwal, and M. Sai, "Heart Disease Prediction using Machine Learning," IJERT Journal International Journal of Engineering Research & Technology. 2020. [Online]. Available: [www.ijert.org](http://www.ijert.org)
- [11] M. Görgün, "Makine Öğrenmesi Yöntemleri ile Kalp Hastalığının Teşhis Edilmesi," Yüksek Lisans Tezi, Lisansüstü Eğitim Enstitüsü, İstanbul, 2020.
- [12] A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, "Design and implementing heart disease prediction using naives Bayesian," in Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 292-297. doi: 10.1109/icoei.2019.8862604.
- [13] O. Köse, "Sınıflama ve Regresyon Ağaçları Tekniği İle Kalp Hastalıklarına Etki Eden Bazı Faktörlerin Belirlenmesi." 2018.

- [14] V. V Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: A survey," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 2.8 Special Issue 8, pp. 684–687, 2018, doi: 10.14419/ijet.v7i2.8.10557.
- [15] H. Sharma and M. A. Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms," *national Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 8, 2017.
- [16] K. H. Kamil, "Artificial Neural Network Approach For Heart Disease Classification." p. 58, 2020.
- [17] I. Salman, "Heart attack mortality prediction: An application of machine learning methods," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 6, pp. 4378–4389, 2019, doi: 10.3906/ELK-1811-4.
- [18] S. Konda, A. Govardhan, and G. R. Rao, "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques," K. Srinavas, Ed., 2020, pp. 1953–1957.
- [19] M. Tarawneh and O. Embarak, "Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 29. Springer Science and Business Media Deutschland GmbH, pp. 447–454, 2019. doi: 10.1007/978-3-030-12839-5\_41.
- [20] E. Çil, "Makine Öğrenmesi Algoritmalarıyla Kalp Hastalıklarının Tespit Edilmesine Yönelik Performans Analizi." 2022.
- [21] P. Kamil, "Personal Key Indicators of Heart Disease," <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>. Jan. 2022.
- [22] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16. pp. 321–357, 2002.
- [23] Z. Wang, C. Wu, K. Zheng, X. Niu, and X. Wang, "SMOTETomek-Based Resampling for Personality Recognition," *IEEE Access*, vol. 7, pp. 129678–129689, 2019, doi: 10.1109/ACCESS.2019.2940061.
- [24] Miss. M. S. Shelke1, Dr. P. R. Deshmukh2, and Prof. V. K. Shandilya, "A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique," *International Journal of Recent Trends in Engineering and Research*, vol. 3, no. 4, pp. 444–449, Jan. 2017, doi: 10.23883/ijrter.2017.3168.0uwxm.
- [25] A. Sekulić, M. Kilibarda, G. B. M. Heuvelink, M. Nikolić, and B. Bajat, "Random forest spatial interpolation," *Remote Sens (Basel)*, vol. 12, no. 10, Jan. 2020, doi: 10.3390/rs12101687.
- [26] E. Deniz, "Yapay sinir ağları ve K-en yakın komşu algoritması ile toprak çeşitliliğinin belirlenmesi." p. 69, 2021.
- [27] E. Akca, "Satış Tahminlemede Hibrit Bir Yaklaşım: PESTEL Rfm, Gradient Boosting." Jan. 2022.
- [28] A. Abraham, Paramartha, D. Jyotsna, K. Mandal, A. Bhattacharya, and S. Dutta, *Advances in Intelligent Systems and Computing 813 Emerging Technologies in Data Mining and Information Security*. [Online]. Available: <http://www.springer.com/series/11156>
- [29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery*, Jan. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [30] M. Zareapoor and P. Shamsolmoali, "Application of credit card fraud detection: Based on bagging ensemble classifier," in *Procedia Computer Science, Elsevier B.V.*, 2015, pp. 679–685. doi: 10.1016/j.procs.2015.04.201.
- [31] M. COŞAR and E. DENİZ, "Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi," *European Journal of Science and Technology*, Jan. 2021, doi: 10.31590/ejosat.1012986.
- [32] Ö. Ekrem, O. K. M. Salman, B. Aksoy, and S. A. İnan, "Yapay Zekâ Yöntemleri Kullanılarak Kalp Hastalığının Tespiti," *Mühendislik Bilimleri ve Tasarım Dergisi*, vol. 8, no. 5, pp. 241–254, Jan. 2020, doi: 10.21923/jesd.824703.
- [33] M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, "Identification of Heart Failure by Using Unstructured Data of Cardiac Patients," in *Proceedings of the International Conference on Parallel Processing Workshops, Institute of Electrical and Electronics Engineers Inc.*, Jan. 2016, pp. 426–431. doi: 10.1109/ICPPW.2016.66.
- [34] M. E. TAŞÇI and R. ŞAMLI, "Veri Madenciliği İle Kalp Hastalığı Teşhisi," *European Journal of Science and Technology*, pp. 88–95, Jan. 2020, doi: 10.31590/ejosat.araconf12.
- [35] S. GÜNDOĞDU, "Kalp hastalık risk tahmini için Python aracılığıyla sınıflandırıcı algoritmalarının performans değerlendirilmesi," *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, vol. 23, no. 69, pp. 1005–1013, 2021, doi: 10.21205/deufmd.2021236926.