REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

A NEW POWER LINE FAULT DETECTION USING ENHANCED MACHINE LEARNING TECHNIQUES

Nawar Majid Dhahir ALDHAHER

Master's Thesis

Supervisor

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Istanbul, 2023

A NEW POWER LINE FAULT DETECTION USING ENHANCED MACHINE LEARNING TECHNIQUES

Nawar Majid Dhahir ALDHAHER

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2023

The thesis titied A NEW POWER LINE FAULT DETECTION USING ENHANCED MACHINE LEARNING TECHNIQUES prepared by NAWAR MAJİD DHAHIR and submitted on 12/04/2023 has been **accepted unanimously** for the degree Master of Science in Electrical and Computer Engineering

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr. Abdullahi Abdu İBRAHİM	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr. Zeynep Altan	Department of Software Engineering Beykent University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master`s Thesis.

Submission date of the thesis to the Institute of Graduate Studies: ____/____/____

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Nawar ALDHAHER

Signature

DEDICATION

I would like to thank my supervisor Asst. Prof. Dr. AYÇA KURNAZ TÜRKBEN and my all family.

ABSTRACT

A NEW POWER LINE FAULT DETECTION USING ENHANCED MACHINE LEARNING TECHNIQUES

ALDHAHER, Nawar

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Date: April / 2023

Pages: 56

Power lines are one of the most common engineering systems designed to transmit large amounts of electricity from one corner of the country to as far away in other directions. The distribution of lines over different landscapes and geographical locations makes them more vulnerable to various weather disasters, often resulting in line outages. The faulty line should be removed as soon as possible in order to limit the unnecessary massive generation of electricity from the fault point and, at the same time, to restore the stability of the system as soon as possible to resume the normal operation of the energy flow. This is where the importance of having a robust error identification, classification, and containment algorithm that can successfully manage and maintain a digital transition system comes in.

In this study, new method-based machine learning techniques applied to detect faults in the transmission lines. The proposed method based GWO and PSO which are used to train the AdaBoost and presented 99.64% accuracy.

Keywords: GWO, PSO, KNN, AdaBoost, Power line transmission.

TABLE OF CONTENT

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

SVM	:	Support Vector Machine
NN	:	Neural Network
ANN	:	Artificial-Neural-Network
CNN	:	Convolutional Neural Network
PSO	:	Particle Swarm Optimization
GWO	:	Grey Wolf Optimizer
AdaBoost	:	Adaptive Boosting
UPS	:	Uninterruptible Power Supply
5G	:	Fifth Generation
PLC	:	Power-Line Communication
HV	:	High Voltage
LV	:	Low Voltage
MV	:	Medium Voltage
NB-PLC	:	Narrow Band Power Line Communication
UNB-PLC	:	Ultra-Narrowband Powerline Communication
BB-PLC	:	Broadband Power Line Communication
QB-PLC	:	Quasi-Band Powerline Communication
MHz/KHz	:	Megahertz / Kilohertz
TWh	:	Terawatt Hour
AMR	:	Automated Meter Reading
AMI	:	Advanced Metering Infrastructure
CIN	:	Cervical Intraepithelial Neoplasia
Mbps	:	Megabits Per Second
BTS	:	Base Transceiver System
FCC	:	Federal Communications Commission
ARIB	:	Association Of Radio Industries and Businesses
CENELEC	:	European Committee for Electrotechnical Standardization
ART	:	Assisted Reproductive Technology
IVF	:	In Vitro Fertilization
dB	:	Decibel

MRI	:	Magnetic Resonance Imaging
FCM	:	Fuzzy C-Means Clustering Algorithm
ICSB	:	International Society for Computational Biology
SCB	:	Stochastic Contextual Bandit
SER	:	Sequential Regression
RLS	:	Recursive Least Squares
ERAS	:	Enhanced Recovery After Surgery
ADHD	:	Attention Deficit Hyperactivity Disorder

1. INTRODUCTION

It is estimated that worldwide electrical energy consumption will increase by 60% in 2030 to approximately 37000 TWh. This increase in demand is expected to be seen mainly in world capitals. According to the United Nations data, the world population will double in the next 40 years, and the population, which was 3.4 billion in 2009, will reach 6.9 billion by 2050. According to the International Energy Agency, with this population growth, capital cities in the world are expected to have a share in two-thirds of the said energy demand increase. Considering the concepts of meeting energy demand, global warming, carbon emissions and fossil fuel use, it should be managed in an environmentally friendly way. This will significantly increase the dependence on renewable energy sources. But the problem is that today's electricity grids are not designed to meet the growing energy demand or to support electricity generation from renewable energy sources. For this reason, the need for updating the existing electricity grid arises. Updating the existing grid is to reveal the concept of "Smart Grid" [1]. The starting point of the emergence of the Smart Grid concept is the increasing demand for electrical energy, but there are other parameters that support the creation of the smart grid concept. In line with today's technological developments and consumer and service provider demands, the need for service quality increase is the most prominent among these parameters. Smart grids, together with the communication technologies they use, offer innovations that can be listed as remote instantaneous monitoring, control and intervention, flexibility, reliability, elimination of errors that may occur due to human influence, instant data transmission in electricity generation, transmission and distribution processes. Distributed production facilities, electric vehicles, smart home and smart city systems, energy storage systems, generation, transmission, distribution control systems, remote meter reading systems, together with the communication methods they are used in technical sense, constitute the whole of smart grids. The concept of smart grid provides the opportunity for distribution control and communication and measurement related to grid operations, with smart technologies, by using digital information and control technology to increase the effect, safety and efficiency of the electricity grid. It provides the basis for the integration and development of energy efficiency and demand-side resources by providing the integration of generation and distribution resources, including renewable energy resources. It provides dynamic

optimization of grid operations with the development of communication standards for the interoperability and communication of devices and equipment connected to the electrical grid. In addition to all these, smart networks offer consumers and service providers the opportunity to monitor and monitor at any time or in real-time. Although the reasons for the outages in the base stations vary, the most common cause of the interruption is the power outage. For this reason, telecommunication companies have turned to various investments such as batteries, UPS and fixed-portable generators in base stations. In this way, it is aimed to minimize the network interruption due to the power outage. With the transition of mobile communication technology to 4.5G in our country, the number of network equipment installed at stations and therefore the need for energy has also increased. With the transition to the next target, 5G, this need is expected to rise to a higher level. Most of the energy consumption will be realized by base stations and Access Point devices. Therefore, efficient energy consumption is of great importance. conventional electricity grid. It has been designed with the aim of maintaining electricity generation, transmission and distribution services in an uninterrupted and safe manner. Over time, the sole aim of providing security and continuity in the operation of the traditional electricity grid has not been able to meet the new needs and optimization requirements. 20th century electricity networks were designed as local scale, small scale networks. In the framework of developing economic needs and increasing demand, the additions over time have led to the transformation of existing networks into complex networks that are difficult to control. By the end of the 1960s, almost every home and workplace in the developed world had access to electricity. Towards the end of the century, this confusion started to show itself as increasing energy losses and control/follow-up difficulties, especially in crowded cities and highly industrialized regions. When today's modern needs cannot be met by the existing network infrastructure in many areas, new searches have started. These searches are among the reasons that reveal the concept of smart grid as the future of traditional grids [2]. Three basic strategies have been determined to meet the increasing energy demand with the use of high frequencies. These; Resource Allocation Strategies have been named as Interference Management and Energy Efficiency Transplantation. One of the most important parts of the base station is the base station's transceiver system, known as the BTS (Base Transceiver System). The elements that determine the performance of BTS are not only its own features, but also the power supply of the base station and environmental factors. This situation makes

the detection and solution of the malfunction more complicated. One of the reasons for the outages in the base stations is bad weather conditions. A station must be in communication with another base station or central station in order to serve. In order to achieve this, communication is made via air communication, communication via satellite or via fiber cable, which is a more reliable but more costly method. The most widely used method is the communication made by providing transmission over the airway, since the investment cost is low and possible fault resolution is easier. To achieve this, the receiving and transmitting antennas of the two base stations must be facing each other and there must be no physical obstacles between them. Thus, communication is provided over the airway via electromagnetic waves. The most important factor that negatively affects airway communication is bad weather conditions. Rain, snow, hail…etc. weather events, such as the transmission path, have a disruptive and quality-reducing effect. In some cases, there are situations called loss of vision that lead to interruptions. In order to prevent this situation, telecommunications companies use weather forecasts to make some forecasting studies in order to predict and prevent disruptions. For this, it will be sufficient to correlate the base station outage history records with the weather history records based on location and date. Another working method used for estimation is artificial neural networks studies conducted under the science of artificial intelligence. With the development of technology, computers have become much smarter and studies in the field of artificial neural networks have increased. In general, artificial neural networks are a method that allows computers to think like human brains by examining the working principles of the human brain and transferring them to computers, thus enabling the expansion of studies in various fields. Although these fields are generally seen as biology and mathematics, artificial neural networks are also used in the field of forward prediction. Artificial neural networks as content; have the ability to learn, memorize and make decisions by establishing relationships. By using examples, he/she creates a unique experience history and has the ability to make similar decisions on other similar issues [3]. There are various artificial neural network models in the literature. Scientists have worked to transfer the thinking, remembering and problem-solving features of our brain to computers. Some researchers have also succeeded in creating ANN models that can perform various functions of our brain. Studies have been carried out in the field of weather forecasting using ANN techniques. By determining the input parameters such as water vapor pressure, relative humidity, wind speed, and air pressure, an ANN was created

on the temperature estimation as the output parameter [4]. With the help of this ANN, the output parameter, that is, the temperature, was estimated by using the input parameters. With this and similar applications, meteorology, stock market, economy, etc. forecasts are also made.

As in the traditional electricity grid, the basic components of the smart grid are generation, transmission and distribution. These components are detailed within themselves. In particular, distribution services come to the fore at this point [5].

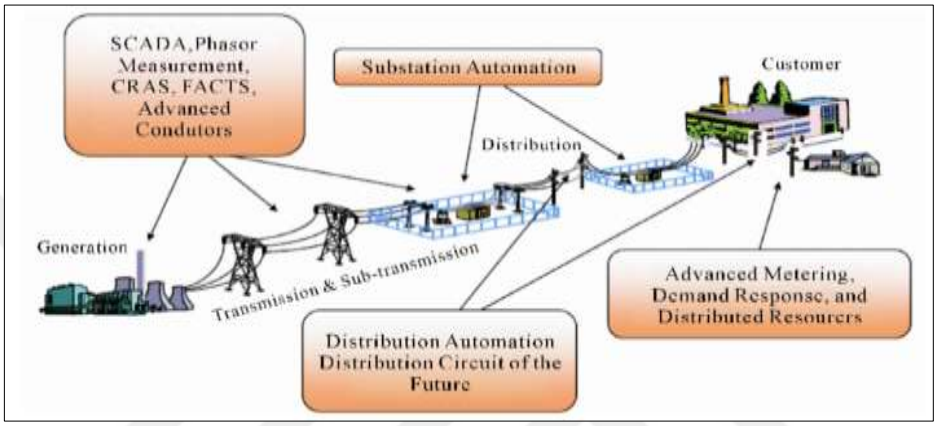

Figure 1.1: Smart Grid.

Electricity generation, transmission, distribution services and other components have to be connected to each other through a communication channel. Communication is the most essential component of the smart grid concept. Business and remote monitoring are components in which both consumers and service providers play a role. Service providers are generally considered as components that provide electricity distribution service and communication service. Another key component is consumers. Consumers can follow the instant meter reading results and carry out their energy efficiency transactions and choose the most suitable tariff according to the tariff structure offered by the retail service provider. Electricity distribution companies, electricity retail companies and electricity generation companies that make up the electricity market, direct the electricity market by trying to make the most accurate energy demand and forecasts by monitoring the instantaneous consumption of consumption points and generation points with remote meter reading systems, which is one of the smart grid operations [6][7].

2. OVERVIEW

2.1 RELATED WORKS

Communication with PLC is widely used within the scope of smart network, especially for remote smart meter reading applications. Basic parameters of communication type with PLC; communication types and standards, modulation techniques form the basis of this type of communication. In smart meter applications based on communication between transmitter and receiver two points, programmable development boards to investigate the suitability of PLC use with different scenarios and remote smart meter reading field model where real data are obtained come to the fore.

2.2 COMMUNICATION OVER POWER LINES

PLC technology is defined as the realization of data communication while carrying out generation, transmission and distribution operations over the existing electricity network. With the development and gaining importance of high frequency applications, PLC has started to be used for remote control and network monitoring in electricity generation transmission and distribution processes [8]. With its end-to-end communication capability, PLC technology has proven to be sensitive to the new generation electricity grid, and more importantly, PLC standards and certifications have been created with the technological handling of PLC communication type, and as a result, PLC has attracted attention in terms of communication and remote control of the electrical network (Sharma and Saini, 2017). PLC communication is a two-way communication method between the smart grid object in the production, transmission and distribution lines and the remote-control center. It is used as a data communication medium for applications such as remote meter reading, load control and protection signaling in production, transmission, distribution lines and electrical networks with PLC technology. While signaling is used on high voltage lines, data such as meter reading results are communicated over MV and LV lines. In today's applications, it is seen that the communication method with PLC is widely used for remote meter reading, which is one of the electricity distributions processes. This is because there is already a wired communication method ground for data communication, with the existing distribution network existing from the transformer to the meter. From the point of view of the electrical network, the benefits of using PLC as a communication method are listed as the fact that it

uses existing power lines and will communicate with the devices connected to these lines, so there is no need for new site configuration, no coverage problem with the ability to communicate in difficult field conditions, no need for antenna use, while the disadvantages are listed as follows: It can be listed as the power lines are not designed for communication purposes, communication impedance and communication channel conditions are constantly changing, excessive switching number in the network, signal weakening and exposure to interference. Although communication with PLC is possible in every layer of the electrical network, the most common usage is observed in the LV electricity distribution network. At the points where electricity generation, transmission and distribution processes are carried out, communication can be provided over the existing electricity network with a PLC-enabled modem.

2.3 BASIC PARAMETERS OF POWER LINE

Communication PLC channel, which carries data between two transmitter and receiver points over the power line, encounters problems such as signal attenuation, noise, signal interference, variable and irregular load change and switching, coupling circuit compatibility, electrical network dynamic structure and differences. Various models have been developed in the time and frequency domain in order to overcome these problems in order to provide communication using the PLC method. Time plane-based models are known as multipath propagation model while frequency plane-based models are known as transmission line model. Statistical and deterministic approaches are used for the PLC channel model. Due to the dynamic and variable nature of the PLC channel, the statistical approach method created with the data obtained from the network comes to the fore. In recent studies, it has been observed that models in which both statistical and deterministic approaches are used together are tried to be created [9]. For the use of the PLC method as a communication type, three basic parameters must be taken into account: the impedance value of the PLC channel, the noise that the PLC channel will encounter in the network, and the signal attenuation that may be experienced during end-to-end data transmission.

2.4 IMPEDANCE VALUES

The impedance value, which has a time-varying character according to the frequency and the condition of the loads in the network, is expressed as the input impedance or the channel

impedance for the NB-PLC channel. In the operating frequency range of the NB-PLC channel, low value input impedances can be a problem in the point of power transmission in PLC modems that work together with the counters as integrated or separate. The signal amplitudes specified by the globally accepted PLC standards must also be valid for the low channel impedance values of the line that is aimed to communicate with the PLC. This is possible by determining the lowest impedance value and determining the maximum power required at the transmitter point according to this value. If the impedance value of the receiver/transmitter circuit is the same as the channel impedance where data communication is provided, the power of the signal that will carry data between two points will be transmitted to the line at the highest level. In PLC technology, coupling circuits provide the connection of carrier signals with the power line. With capacitive coupling circuits designed especially using capacity, bidirectional communication is created between transceiver points. The main task of the coupling circuit is to equalize the impedance values of the transmitting and receiving circuits on the modem with the power line impedance where the communication is provided. This is achieved by changing the lower and upper cutoff frequency values of the coupling circuit according to the line impedance parameters [10].

2.5 NOISE

Noise ‹is one of the basic parameters that can change the operation of the PLC communication channel negatively with the disruptive effect it creates. The noises observed in the power line, the characteristics of the loads in the network, the connection status of these loads on the line, the switching on and off situations, the frequency occurring in the network. The noises, with the characteristics they create in the network, reflect the frequency and amplitude values of the carrier signals that provide data communication between two points positioned as transmitter and receiver. The parameters of the noise, frequency, duration, amplitude and power density spectrum, are the main values used in noise analysis. For the NB-PLC level, the data obtained as a result of the measurements made with the spectrum analyzer, oscilloscope, data collection cards, analytical studies and according to the network loads. Noises are classified under three main headings as impact noise, continuous noises and emission-induced noises. Impact noise is the most emphasized and modeled noise type for PLC technology. The source of impulse noise is loads and equipment in the electrical network. Pulse noises are high amplitude noises that occur in the form of

regular or irregular pulses. Periodic and synchronous pulse noises with the network, noises caused by the switching of semiconductor elements, pulse noises that are compatible with the network but asynchronous in terms of frequency, noises caused by the operation of switched power supplies, non-periodic noises are caused by the network frequency in the temporary states that occur during the switch-on and output moments of the devices in the network. can be exemplified as independently occurring noises. Continuous noises are noises that occur in a certain part of the line due to the network frequency in the PLC channel and the harmonics of this frequency value in the same period and due to the loads in the network. Transmission noise is the noise that occurs due to cable length [11].

2.6 SIGNAL ATTENUATION

In PLC technology, the two most important parameters for signal attenuation are the length of the power line where data communication is made and the number of attachments on this line. It is expected that the long distance between the transmitter and receiver points and the excessive presence of additional points on this line will cause signal attenuation. In addition to these situations, another parameter to be considered is the loads on the communication line and the variability in the line impedance due to the variable situations that occur as a result of these loads being switched on and off. These variations will indirectly affect the signal attenuation as they will differentiate the signal amplitude. For this reason, different signal attenuation values can be seen on lines with the same conductor length and the same number of joints. As a result of a study conducted in our country, signal attenuation values were measured at 30 dB in the industrial region, 20 dB in the urban region and 19 dB in the rural region, and it was determined that these signal attenuation levels were not excessively higher than the signal attenuation values of wireless communication technologies [12].

2.7 TYPES AND STANDARDS OF COMMUNICATION OVER POWER LINES

PLC signal was first used for New York City Street lighting control up to 500 Hz. Afterwards, with the progress of PLC technology, it was applied for reliable communication over the network line in data transmission processes and this communication method was first established on the electricity distribution network. In smart meter applications, its efficiency has been proven by the widespread use of PLC method and it has formed the communication infrastructure of smart meter applications around the world. Due to this

situation, world-wide PLC standards and modulation types have been created based primarily on energy measurement processes with smart meters. Communication types with PLC are divided into four basic titles as UNB-PLC, NB-PLC, Half Band (QB)-PLC and BB-PLC, considering data transfer rates and operating frequency ranges. The created PLC standards are clustered in relation to NB-PLC and BB-PLC. 3 kHz - 500 kHz frequency range and 1.2 kbps - 1Mbps data communication rate range are seen for NB-PLC, while 1.8 MHz - 100 MHz frequency range and 9.8 Mbps - 1.8 Gbps data communication speed range are seen for BB-PLC. range is displayed. G3-PLC and PRIME standards of NB-PLC technology are widely studied and widely used standards. In addition, it can be said that the use of Home Plug AV and IEEE P1901-2010 standards for BB-PLC technology is high. NB-PLC is known as a PLC type with intensive processing volume, which is used to obtain smart meter data remotely, especially in the electricity network. In this direction, organizations named CENELEC of European origin, FCC of American origin and ARIB of Japanese origin have determined the frequency bands that will establish their own standards for the NB-PLC level. Table 3.1 shows these frequency ranges. CENELEC, which regulates European standards, has divided the 3 kHz – 148.5 kHz operating frequency range into four sub-frequency bands to ensure efficient and trouble-free operation, taking into account the diversity of operations. In the 9 kHz – 95 kHz frequency range, the CENELEC-A band is reserved exclusively for distribution and generation service providers. In the 95 kHz – 125 kHz frequency range, the CENELEC-B band is reserved for consumer general applications. CENELEC-C band in the frequency range of 125 kHz – 140 kHz is the frequency band regulated for indoor control communication. In the frequency range of 140 kHz – 148.5 kHz, CENELEC-D band is the band created for the use of alarm and security systems of consumers.

2.8 POWER LINE COMMUNICATION IN SMART GRIDS

Although the transformation of traditional grids into smart grids has come to the fore at the electricity distribution level, this transformation also takes place at the electricity generation and transmission levels. For this reason, the need for a communication system, which is one of the components of the smart grid, is valid for every layer of the electrical grid and every voltage level. PLC communication type is widely used at the LV level of the electricity distribution network. As a result of this situation, most of the PLC standards created in the

electrical network are concentrated at the NB-PLC level. In addition to this, it is possible to use NB-PLC technology for transmission and generation operations at the MV and HV level of the network, although it is not sufficiently implemented and standardized. The place where PLC technology is used the most in terms of transaction volume is at the LV electricity distribution level. NB-PLC type and connected communication standards provide the communication needs of AMR and AMI systems, which are the most basic applications of the smart grid, in order to remotely obtain the data of millions of electricity meters at consumption points in the network. In addition, it is seen that PLC communication technology at the AG level plays a role in eliminating the need for remote monitoring and monitoring among the triangle of electric vehicles, charging stations and electricity distribution service providers, which will be one of the most active components of the next generation electricity grid, the use of which is starting to increase. PLC technology used for the communication of indoor energy management and indoor control systems, known as smart building, is the PLC application seen in another network layer. The energy data produced by the production facilities, especially the distributed generation facilities established with renewable energy sources such as PV and Wind, can be read remotely via PLC within the scope of AMR and AMI applications over the bidirectional meter. Controlling the devices at the MV level and providing status information with a communication system is extremely important for the development of the smart grid28 and PLC technology is a communication technology that can be used in this context. In the MV network, fault detection, power quality measuring devices, and transmission of power quality parameters of places such as transformer centers and distribution centers to central software with PLC are the main operations. The use of PLC for remotely detecting oil temperatures and secondary winding voltage levels of transformers providing HV/MV conversion is also possible at MV level. PLC is a communication technology that can be used for position information and remote control of MV level switching equipment such as breakers and disconnectors. In addition to these, it is possible to see islanding, which is an undesirable situation, as a result of malfunctions in the network established as MV level distributed generation. It is possible to detect and prevent this unwanted situation by injecting PLC signal into the MV network [13].

3. MATERIAL AND METHODS

3.1 ARTIFICIAL INTELLIGENCE (AI)

It is to gain the abilities of human intelligence, such as reasoning, making sense, generalizing, and learning from past experiences, to machines or computers. Since the main purpose of artificial intelligence consists of the idea of analyzing the human thinking algorithm, it can be said that its history dates back to Aristotle. In historical sources, Aristotle mentioned the difficulties about this issue, where he worked on the algorithm of thought, but could not achieve success. With the invention of the first computer in the second world war, computer software started and artificial intelligence emerged in a real technological sense. Can Alan Turing, who changed the course of the war with the invention of the computer, think of machines in 1950? He is the first person to come up with the idea of artificial intelligence with the question. Turing later introduced the Turing test to distinguish humans from machines [14]. In this test he developed, he aimed to evaluate his performance by pairing a human with another human or a computer. Although it is not possible to distinguish between the two, the test was considered successful. The term artificial intelligence was first mentioned by Minsky and McCarthy at a conference in 1956. There is no single consensus diagnosis on AI, multiple approaches include:

a. Processing information; Cognitive capacities to achieve our human, social or professional goals,
b. The ability to act towards a goal, think logically, and relate beneficially to one's environment,
c. The ability to recognize, understand and adapt to new situations, and find solutions to difficulties encountered, adapting to any problems and other living things.

When these multi-approach logics are considered, it is seen that during evolution, there was a parallel behavior consciousness with living things in nature. For example, there is an interaction between plants and tree roots, or the individual and social behavior of dolphins, chimpanzees, bees, birds, ants, dogs and many other species. With this logic, artificial intelligence facilitates human life by using computational power and instant calculations in many areas, and with the ability to evaluate new data in real time according to previously processed data [15]. Today, artificial intelligence has been integrated into the lives of people from all walks of life, in many areas such as personal assistants, automatic public

transportation, aviation, computer games. Recently, new developments in artificial intelligence have adapted to achieve greater accuracy in patient diagnosis and improve patient care. Many areas such as radiological images and pathology are evaluated with machine learning and they increase the abilities of healthcare personnel by helping the diagnosis and treatment process of patients. Health care is a basic need for every person. For the realization of physical existence, health is at the basic level of Maslow's pyramid of needs. Health services are an integrated service branch that can be diversified in both socio-cultural and economic contexts. This branch of service includes the stages of diagnosis, treatment, health improvement and rehabilitation, and the processes form integrity with each other. A good health system provides an increase in health literacy by preventing disease before it occurs, diagnosing the existing disease before it progresses, applying the necessary treatment, creating post-treatment rehabilitation and raising public awareness. The quality, accessibility and cost of health services are among the branches that are developing today. Developments in technology always affect the quality of life positively. When the developments in the industry were adapted to the field of health, the concepts of Health 5 1.0, Health 2.0, Health 3.0 and Health 4.0 emerged. Health includes vaccine developments in 1.0, the development of drug production with mass production in 2.0, the emergence of new antibiotics, developments in the field of radiology as a result of computer and information technologies in 3.0, and artificial intelligence-based developments in 4.0. With the emergence of artificial intelligence that it can be used to specifically solve or clarify complex biomedical problems, interest in its potential has grown exponentially, and as a result it has begun to enter the field of health. Warner et al. A study was conducted on establishing an automatic diagnosis system by obtaining data from 1035 patients who were referred and analyzed for cardiac catheterization as a result of congenital heart disease in 1961. Following this initial research on the diagnosis, in addition to diagnosing the disease, another study was conducted with a computer program developed at Stanford University in 1970, in order to prescribe antibiotics and determine the causative bacteria and determine the dose adjustment according to body weight. The program named GUIDON, which was derived by intelligent computer-aided training of artificial intelligence, was developed and used in late 1970 to teach medical students the diagnosis of infectious diseases. Other artificial intelligence systems have been developed and progress has been made on the basis of diagnosis and comparisons of different systems have been made [16]. In the following

studies, studies were conducted comparing the diagnostic capabilities of physicians with the capabilities of artificial intelligence. According to the results of these studies in which artificial intelligence systems were evaluated, it was reported that the systems helped physicians diagnose and strengthen the diagnosis with an objective approach. These studies, which were carried out in the early stages of the history of artificial intelligence, shed light on many studies that are now applied. The first step in the development of artificial intelligence includes the collection of an unlimited amount of organized data, the second the emergence of the internet, which creates an instant communication network, and the third, the instant automatic processing of the data. Artificial intelligence is undergoing an exponential development, including its transformation in medical applications. Parallel to these developments, there are many new technology branches such as robots, quantum computers, and virtual reality [17]. With reference to global health research, artificial intelligence-driven health studies consist of four categories. These are diagnosis, analysis of the patient's morbidity or mortality risk, epidemic forecasting and monitoring, health policies and planning. AI-driven health interventions are thought to lead to better health outcomes in low- and middle-income countries. Artificial intelligence is rapidly evolving in healthcare because of its potential to unlock the power of big data and gain insights to support evidence-based clinical decision-making and achieve value-based care. In the fields of artificial intelligence in health, there are research branches such as disease diagnosis, classification, detection, prediction, diagnosis, determination of risk levels, monitoring and prognosis follow-up. It also includes various fields such as the processing of medical images, health services management, and determination of physician opinions. It is critical for healthcare leaders to understand the status of AI technologies and how these technologies can be used to improve the efficiency, safety and accessibility of healthcare by supporting the digital transformation of healthcare. Artificial intelligence has been used in many studies on women's health. Positive results have been achieved in many studies for the diagnosis of gynecological cancer. Endometrial cancer is one of the gynecological cancer types that increases over time and is common in many parts of the world. Although it is not a commonly used screening test, early diagnosis is very important. According to research done. In the light of the information obtained from hysteroscopy images, an artificial intelligence-based system was established. This study included 177 patients, 60 images of normal endometrium, uterine myoma, polyps, atypical endometrial hyperplasia and endometrial cancer. As a result

of the study, it is seen that the diagnostic accuracy has increased from 80% to 90% and above. The study appears to be a tool to facilitate the diagnosis of endometrial cancer. A different study was also conducted with the aim of accurately detecting endometrial lesions and endometrial lesion type with hysteroscopic images. In this study, the importance of artificial intelligence in self-diagnosis is emphasized because it is successful with higher sensitivity compared to gynecologists in the pre-cancer period. As with endometrial cancer, there is no screening test for ovarian cancer and the rate of recurrence is high because it is diagnosed at a late stage. Early diagnosis leads to positive changes in many areas from the treatment process to the surgical process. In an artificial intelligence study for the diagnosis of ovarian cancer, blood test results from 202 patients were created using five different deep learning 7 methods, along with information obtained from the patient's history, imaging tests and pre-operative examination. As a result of the study, it was concluded that the pathological diagnosis of ovarian cancer may play a role in estimating from preoperative examinations. Cervical cancer ranks second in terms of mortality from gynecological cancers. Cervical intraepithelial neoplasia (CIN) is closely associated with HPV-infected patients with persistent infection. HPV screening periods have been increased in developed countries. There is no method other than periodic observation for women with positive screening programs [18][19]. A study was conducted using artificial intelligence technology to determine whether the HPV genotype would predict the risk of cervical dysplasia persistence and recurrence before treatment. As a result of the study, it was revealed that some types of HPV may pose a risk in cases of recurrence and cancer. For this reason, it has been concluded that artificial intelligence technologies are an advantageous technology for personalized treatment and additional vaccination programs. It is known that the peracetic acid test and the post acetic acid test are a method used for the diagnosis of CIN with colposcopy images. According to a study, it has been reported that these images are combined with artificial intelligence technology, resulting in very high estimates of specificity, accuracy and sensitivity for the diagnosis of CIN. In a study conducted with deep learning methods developed for the prediction of survival of 797 patients with uterine adenosarcoma, it was reported that highly accurate predictions were made in personalized prognoses. A review of the development of artificial intelligence in gynecological cancers addresses the accuracy of diagnosis and prediction of early detection. However, since artificial intelligence has not been fully humanized, it is emphasized that there will be many

moral problems such as ethics and health insurance in the future. Breast cancer screening has been shown to significantly reduce the death rate in women. The increasing use of screening exams has led to increasing demands for fast and accurate diagnostic reporting. With the developing screening tests, artificial intelligence technology has taken a very important way in the diagnosis of breast cancer [20]. In the last five years, systems developed by artificial intelligence technology, deep learning and neural networks, and digital mammography images have been used for the diagnosis of breast cancer, which is the most studied subject in the field of gynecology. Commercial products have emerged in artificial intelligence technology developed for breast cancer diagnosis. According to the studies obtained from these developing technologies, it has been concluded that the performance of artificial intelligence technology is equal when compared with experienced radiologists. In the light of many studies with positive results, it has become clear that artificial intelligence technology will have an important role in the diagnosis of breast cancer in the future [21]. With the onset of menopause, artificial intelligence technologies have also been developed for osteoporosis, which deeply affects women's health. This technology is able to calculate and evaluate the risk of osteoporosis. In the artificial intelligence technology study, which was studied to identify patients at risk for bone fractures, it was observed that osteoporosis fractures were determined beforehand using different deep learning methods. ART has become an indispensable issue in the treatment of infertility. The decision-making phase in ART practices is clinician-centered, and the decision-making process is based on the experience of the physician and evidence-based practices. The reason for applying ART is based on many factors. In vitro fertilization (IVF) is a popular method of treating complications such as endometriosis, poor egg quality, a genetic disease of the mother or father, ovulation problems, antibody problems that damage sperm or eggs, inability of sperm to penetrate or survive. IVF application is a very laborious process as it is both costly and uncertain. Recurrent reproductive failure, such as recurrent pregnancy loss and recurrent implantation failure, is characterized by complex etiologies and is particularly associated with various maternal factors. Currently, it is believed that recurrent reproductive failure is closely related to the maternal environment, which in turn is influenced by complex immune factors. Without the use of automated tools, it is often difficult to assess the interaction and synergistic effects of various immune factors on pregnancy outcome. For such reasons, artificial intelligence technology has begun to be researched in the field of assisted

reproductive techniques. According to a study; Different data panels were created to predict pregnancy outcomes at four different gestational nodes, including biochemical pregnancy, clinical pregnancy, ongoing pregnancy, and live birth. Parameters used for diagnosis, hormone levels, autoantibodies, peripheral immunology, endometrial immunology and embryo parameters and these data included 64 variables. According to the results of the study; It has been concluded that it can serve as a basis to assist in more precise and personalized diagnosis and treatment planning for patients with recurrent pregnancy loss, thus providing a light for clinicians to treat more accurately. According to another study, parallel results were obtained and positive results were obtained in terms of diagnosis, treatment and live birth prediction of artificial intelligence technology in IVF treatment. In line with these positive results related to gynecological diseases and fertility, artificial intelligence technology has also been used in assisted reproductive techniques. In addition, in a study evaluating perinatal outcome predictions by using artificial intelligence technologies together with amniotic fluid taken from asymptomatic pregnant women with short cervical length. Six different deep learning machine techniques were compared and evaluated. In this study, which had a limited patient population, it was concluded that some deep learning methods were better than others, and in general, artificial intelligence technology accurately predicted perinatal outcomes [22]. It has also been concluded that with the artificial intelligence technology developed for fetal heart rate scans, it can accurately predict fetal asphyxia. The nursing profession is a practice profession based on the knowledge discipline consisting of ideal values as a whole of human health and well-being. After the invention of the computer, with the beginning of the technology era, robot technology and artificial intelligence systems entered our lives. Artificial intelligence technology has started to be used in the field of nursing as well as in all occupational groups. In this way, the patient is better known for nursing care and patient safety increases. Contrary to this situation, Watson et al. think that the use of various technological networks or devices will drive nurses away from the patient. In another view, it is argued that it would be correct to use it as a partner in nursing care in order to contribute to nursing care. In this idea, it is stated that the task sharing of the nursing profession duties together with artificial intelligence can be divided in a method suitable for relative competence [23]. The first entry of artificial intelligence into the nursing profession was with the developed robots. Serious muscle strength is required when positioning the patient in the bed for nursing care.

Reminding medication times, helping muscle strength, dealing with various paperwork actually increases the time the nurse will allocate for care. In addition, it is thought that the risk of disability and diseases in nurses will be minimized thanks to the developed robots. Da Vinci surgeon-controlled robots, which eliminate the anatomical limitations of the hands during surgery, reduce the responsibility of the operating nurse and also increase the surgeon's operating efficiency. It has been used to guide gamma nail surgery with X-ray evaluation images in intertrochanteric femur fracture surgeries with an artificial intelligence-based surgery system. It has been concluded that the use of an artificial intelligence-based surgery system has better nursing efficiency compared to the control group, and shortens the recovery period of the patient. In another similar study, the postoperative nursing effect of artificial intelligence robot-assisted thoracic surgery was evaluated. It has been seen that the surgeries performed with the help of artificial intelligence robot have the advantages of less trauma, less pain, faster recovery, safer and more comprehensive lymph node dissection. In line with these results, nursing care processes are more efficient in artificial intelligence-based surgeries [24]. Robots named TUG, developed for nurses to save time, deliver materials such as laboratory samples, patient food and blankets to the designated destinations [25]. In addition, the robot named 'Ro-bear', which can fulfill the commands such as putting the patient on a chair, lifting or putting the patient in a wheelchair, which can provide the physical disability of the nurse, also helps mobilization. In addition, by integrating artificial intelligence into robot technology, robots that can think and act like humans are being developed. For example, robots named PARO have been developed to positively affect the mood of dementia patients and to change the level of pain. Global aging is becoming more and more serious and the problems arising from nursing care of the elderly will increase in the future. In parallel with this problem, studies are being carried out to develop more functional robots by using convolutional neural network structure for the prediction of three-dimensional human joints in robots developed. In a study examining nurses' views on artificial intelligence and robots, 86% stated that they believe that robots will replace nurses in the future. However, many nurses also stated that their workload will decrease. One of the branches of nursing where the workload is very intense is emergency nursing. New studies are needed to increase the work efficiency of health personnel due to the large number of patients admitted to emergency services and the simultaneous critical interventions. In a study, the last three years' disease data of 10 emergency patient groups such as sudden heart

disease, stroke, and shock, which are the patient groups that frequently apply to the emergency department, were used. In the artificial intelligence algorithm developed urgently to examine the first aid nursing management system, it was observed that the triage time and triage accuracy increased significantly within 3 years. In another study, a visual artificial intelligence system set based on medical information mining was designed for hospital emergency care management. In this visually developed artificial intelligence system, it predicts that the needs of the emergency first aid nursing management system will be analyzed and will greatly facilitate clinical studies. Nursing observations are made every 15 minutes or every hour at night in order to control breathing in patients hospitalized in the acute mental illness unit. Nursing observation is to ensure patient safety, but the patient's sleep is interrupted and, as a result, recovery times are prolonged. In order to prevent this situation, a sensor with artificial intelligence software that monitors micro movements and color changes in the body and provides signal transmission is placed in the patient rooms. As a result of the study, it was concluded that the patients in the experimental group had a shorter hospital stay and woke up less at night than the experimental group (Barrera et al. 2020). Another usage area of artificial intelligence is nursing profession education. Nurses are an integral part of the assessment and management of pain in many care settings that require extensive communication and clinical skills. A study to evaluate the application of simulated virtual learning to acute pain on undergraduate nursing students in Brazil was applied to 14 undergraduate students. It was observed that pain assessment revealed positive results in learning and filled the thematic gaps. In another similar study, computer simulation was used for pain education to undergraduate nursing students and positive results were obtained in terms of participation in education and evaluation. In another study using three-dimensional simulation, it was concluded that it has positive contributions to education. In another study conducted to improve the communication skills of nursing students, the system consisting of four different case scenarios was developed in the form of a three-dimensional avatar. It was concluded that the outcomes of the study would help nursing students to develop their self-efficacy and effective communication skills [26]. Accurate diagnosis with the development of radiological images increases the quality of nursing care. For example, for the magnetic resonance imaging (MRI) features of patients with ovarian endometriosis, artificial intelligence FCM algorithm has been applied in the clinical diagnosis of ovarian endometriosis MRI. Artificial intelligence FCM algorithm plays an important role in

confirming the diagnosis. The fact that the clinical diagnosis was correct enabled the nursing intervention to be comprehensive, and it was concluded that the nursing satisfaction of the patients was higher than the control group [27]. In a similar study, an artificial intelligence-based system was developed for childhood asthma patients to increase the quality of nursing care. The study group of another study using the FCM algorithm was patients with diabetic nephropathy. Early diagnosis of diagnoses made with artificial intelligence technology developed based on MRI images shortens the healing process12 and increases nursing service satisfaction. In another similar study, based on Deming's cycle theorem in children with mycoplasma pneumonia; Nurses were trained for care by using artificial intelligence technology with planning, application, examination and processing steps. As in other studies, patient satisfaction is higher in the experimental group of the study, and the recovery period and hospital stay are shorter. The same is true for ultrasound imaging studies. In a study comparing various artificial intelligence algorithms on ultrasound images for better diagnosis of pelvic floor dysfunction in 120 patients who underwent laparoscopic total hysterectomy for benign uterine diseases, it was concluded that the ISCB algorithm was better than the SCB, SER and RLS algorithms. In addition to the artificial intelligence technology applied on both the experimental and control groups, nursing care was given to the experimental group within the scope of the accelerated recovery after surgery (ERAS) protocol. It was concluded that the ERAS protocol is better in the recovery process than classical nursing care, and patient satisfaction is higher. Another important part of the nursing profession is patient education. In a randomized study of 447 patients with chronic obstructive pulmonary disease, artificial intelligence application was used to assess quality of life after 4 and 12 months. In the study, a 9-month web-based knowledge exercise was applied. It was concluded that the quality of life, emotional and psychological conditions of those in the experimental group were better than the control group after 12 months [28]. Attention deficit hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders in children. It is manifested by inattention, hyperactivity, impulsivity and other symptoms that are not suitable for the level of development, as well as impairment in social, academic and occupational functioning in different situations. The treatment of children with ADHD is mainly based on drug therapy and psychological nursing intervention. Therefore, the actual efficacy assessment of this treatment regimen is crucial. Neural networks are widely used in artificial intelligence. This study combines artificial

intelligence with the evaluation of clinical treatment effects of children with ADHD, and an intelligent model based on neural networks is designed to evaluate the clinical effectiveness of drug therapy and psychological nursing intervention in children with ADHD. In this way, it is believed that nursing care will be more efficient [29].

3.2 DATA MINING

At the conference on knowledge discovery in databases, which was held for the first time in 1995, the data stacks created by information technologies were expressed as follows [30]. “It is estimated that the amount of information in the world doubles every 20 months. What needs to be done with these chunks of raw data. Computers cause data dump” . According to the study carried out by Winter Corporation to find the largest database by determining the volume of databases used in businesses, Sears, Roebuck and Co. Its decision support database reached 4630 gigabytes of data in 1998. This rapid increase in data sizes in database systems has pushed data experts to seek new methods. From this point of view, knowledge discovery, data warehouse, data mining etc. concepts and technologies have emerged [31]. Although the data take up large volumes of space, their use values are not high. The inferences obtained as a result of certain transformations applied to the data are called information. Although information occupies less space than data, its use value is high [32]. Terabytes of data are kept in database systems developed for the realization of business operations by today's technologies. Senior management in businesses need important information hidden in the data stack to determine future projections and strategies. Extracting important information and patterns from large volumes of data using various algorithms, techniques and processes is called data mining. According to Piatetsky-Shapiro, data mining is the extraction of important and key information from large chunks of data [33]. Data mining is simply the acquisition of information and pattern extraction by various learning algorithms and analysis, which can be obtained by human and conventional techniques by spending great effort and working time [34]. In other words, data mining is the search for correlations and rules that will enable us to predict the future from large amounts of data [35]. Data mining is the production of predictions by inferring the patterns and relationships contained in the data stacks subject to analysis by using various analysis techniques and tools. In data mining, which aims to identify unknown information in large data piles, the control is in the data. The analyst has no chance to direct the analysis according

to his own thoughts and expectations [36]. Data mining aims to extract important information from data that does not make sense in its bare form. In other words, it is the extraction of important information for the business from a large data pile. In short, data mining is the process of analyzing corporate data obtained from different sources, determining the relationships between the data and trying to make predictions about the future. Data mining is important for decision support systems as it provides strategically important information about the future of the organization [37].

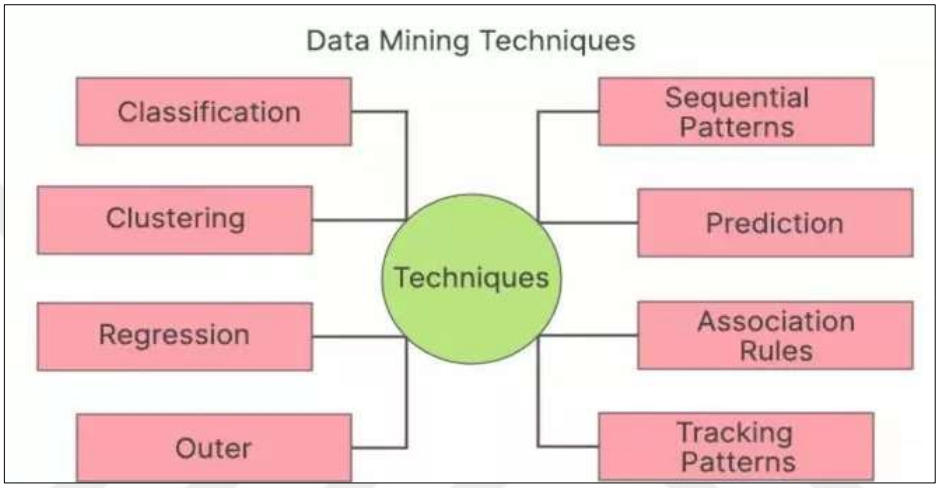

Figure 3.1: Data Mining.

3.3 HISTORICAL DISCOVERY AND DEVELOPMENT OF DATA MINING

a. Computer started to be used in 1950s.
b. In the 1960s, data collections were created by storing data in flat and hierarchical files.
c. In today's database applications, relational database systems, in which we store data, started to be used in the 1970s.
d. In the 1980s, the use of database applications began to increase. Businesses have started to develop automation projects that will fulfill their jobs.
e. In the 1990s, it started to be questioned whether information can be extracted from the data stacks produced by database applications and kept in storage environments, which will contribute to the strategic decisions of businesses.
f. In 1989, the first meeting of the knowledge discovery group was held on the study of extracting knowledge from large data piles.

g. Piatetsky-Shapiro focused on the basic concepts and definitions of decision support systems in his study in 1991 [38].
h. First data mining application was developed in 1992.
i. An international conference on data mining and information extraction was held in 1995.
j. The use of data warehouse and data mining applications in data analysis projects has become widespread.

Data mining emerged conceptually in the 1960s with the use of computers in solving data analysis problems. In these periods, it was accepted that the discovery of knowledge from data would be the result of long searches due to the limited computation and storage capabilities of computers and databases. Therefore, the concepts of data scanning or data capture have been used in data mining [39]. In the 1990s, the name data mining began to be used in data analysis, with the use of algorithms instead of traditional statistical methods. Moving from this development, scientists have been working on statistics, data science, machine learning, etc. They tried to do data mining by developing applications using techniques [40]. With the use of computers in data analysis, statistical methods that could not be done before were developed and started to be used in data analysis. Thus, in the process of data discovery, statistics and data mining started to be used in a strong cooperation [41]. Today, with the increase in the use of artificial intelligence in industry and many fields with Industry 4.0, machine learning has gained great importance. Statistics and machine learning concepts started to be used in the field of data mining with the production of algorithms for special learning problems by using the structure of artificial intelligence algorithms that imitate human learning.

3.4 IMPORTANCE OF DATA IN DATA MINING

Data mining includes algorithms and techniques from various disciplines, including statistics, databases, artificial neural networks, pattern recognition, machine learning, data analysis and visualization. These disciplines and data mining are intertwined and complement each other as a process [68]. The data to be used in data mining must meet the criteria described below:

a. Availability of Data: The data related to the analysis subject are converted into a suitable format for analysis.

b. Does the Data Provide Relevant Attributes? By examining the obtained variables, the unrecorded variables are tried to be determined.
c. Data Noisy? It is the case that the data contains incorrect data. The more noise the data contains, the lower the reliability of the analysis results.
d. Adequacy of Data: In data mining, the adequacy of data is determined not by the size of the data, but by how many qualities it contains.
e. Expert Information on Data: It is the taking of expert opinion according to the area of expertise of the data to be analyzed.

In data mining, unlike statistics, the data stack to be analyzed contains large and many variables. When the number of variables used in the analysis model increases, the computational complexity and the degree of mathematical equations increase. This creates a structure that cannot be handled with conventional statistical methods. Accessing large datasets is difficult due to parameter sizes such as rows, number of columns, and volume, and often because these data are mixed in different data sources. In addition to this, big data creates problems in terms of incomplete, dirty and erroneous data as well as data collection processes. In the field of data mining, the size of the data subject to analysis and the raw data that are not normalized according to the mathematical model established in terms of analysis bring different problems to the agenda compared to conventional statistical techniques [42]. In data mining, after the model and variables to be used in data analysis are determined, the raw data is passed through processes such as data cleaning, normalization, and incomplete data completion in order to make it suitable for the model. Such preprocessing and data cleaning processes take up up to 80% of the time spent on analysis [43]. In data pre-processing, the data collected for use in analysis is classified, and its suitability for analysis is determined, and the data selection process is run. The selected data is combined into a data structure established according to the determined model variables. Merged data can be incomplete data completion, normalization, etc. It is passed through processes and made suitable for analysis [44]. In the selection of the data, the characteristics of the raw data and the criteria of suitability of the data size for analysis are taken into consideration.

3.5 DATA WAREHOUSE

Special databases in which the data collected for data mining applications are stored after a certain format is called data warehouse. In data warehouses, data in different formats

collected from many sources are combined and stored under a common roof in a way that data mining applications can interpret. An example data warehouse architecture can be seen in the figure below [45].

Data warehouses are database structures created for reporting purposes, separate from the basic transaction database of the institution, to provide important information to the strategic decision processes of the institution. Data warehouses provide summary executive reports on the state of the business from large chunks of data. The characteristics of data warehouses, which form the basis of strategic decisions about the future of the enterprise, are listed below [46]:

a. Purpose-Oriented: The data related to the subject to be analyzed is examined and modeled.
b. Consolidated: Data from different sources are combined and kept in data warehouses. The consistency of data from different data sources is ensured by the use of various data cleaning and merging techniques.
c. Time Variable: Since analysis and estimations will be made in data warehouses, long-term historical data are kept.
d. Not Variable: Since data is taken to a different database at regular intervals in data warehouses, it is not affected by instantaneous changes in the live system. Two methods can be used to extract data from systems containing different data sources. These are unified databases and data warehouse.

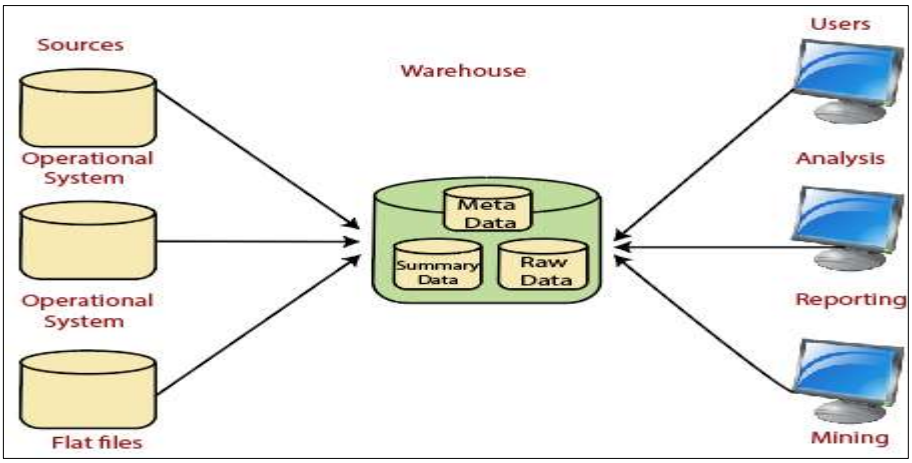

Figure 3.2: Data Warehouses [47].

In merged databases where databases are merged, the query is split into subqueries and run on each database and the results are combined. In the data warehouse solution, the data is combined and stored in the data warehouse for later use, according to the demands and requirements

3.6 DATA PREPARATION

At this stage, data is generated for the analysis modeling process. The data is transformed into a format that can be used by data mining tools by subjecting it to various cleaning and transformations. At this stage of the data preparation process, data stacks are tried to be converted into a table format to be stored in the database. With the features selected according to the model variables planned to be established, records are created and stored in tables. The data preparation process is of great importance in model building. Running the model established for the solution of the problem with incomplete, incorrect and unrelated data will cause the expected results not to be obtained. This will require efforts to continually return and re-evaluate the data. If the data preparation process is evaluated as the analysis stage of data mining, it is necessary to spend 85% of the entire effort at this stage in order to produce the right model and accordingly accurate results [48].

Knowledge Discovery Although data mining is a phase of knowledge discovery in theory, it is used synonymously with data mining in application development. For knowledge discovery, the determined model for analysis is applied to data with data mining techniques. Thus, patterns in the data are tried to be found. Data mining aims to find hidden, important, previously unknown, useful information from big data sources. The stages of knowledge discovery in data mining are listed below:

a. Data Cleaning: It is the process of extracting data that is not related to the model and that conflicts with the general trends of the data in terms of consistency.
b. Data Integration: Data from different data sources are combined.
c. Data Selection (Reduction): Data related to the analysis are determined and selected. After checking the significance of the data for the model, the data describing only specific, non-recurring events that are not compatible with the general trends contained in the data subject to the analysis are excluded from the model data.
d. Data Transformation: Data is made analyzable with data mining techniques.
e. Data Mining: Data mining algorithms are applied to the data and analyzes are made.

f. Pattern Evaluation: Patterns representing the information obtained through data analysis and measurements are developed.
g. Presentation and Evaluation: It is the presentation of the information obtained by data mining methods to the users

As seen in the Crisp-DM stages, raw data is subjected to data discovery and data preparation processes before model operations are started in the data mining process. In Pyle's study, it is claimed that 60% of the total time spent developing a data mining project is devoted to data preparation, while only 5% is devoted to modeling [49]. In data mining, data quality is very important for the reliability of the analysis. Because analyzes to be made with erroneous data will produce erroneous results.

a. Data pre-processing,
b. Correcting data errors that will hinder the analysis of the data,
c. Determining the meanings contained in the data in order to perform data analysis for the solution of the problem,
d. Transforming the data stack subject to analysis into a model compatible and meaningful for the model. In data mining applications, it is necessary to use various data preprocessing techniques according to the problem to be solved, the data collected for analysis and the established model. Therefore, it is important to use correct data preprocessing techniques for the model to produce meaningful results [50]. Data preprocessing techniques in CRISP-DM processes, consisting of 4 processes, are listed as,
e. Data Cleaning.
f. Data Merging.
g. Data Transformation.
h. Data Reduction.

The increase in the data quality of the analyzes made after the application of data preprocessing techniques has increased both the quality of the analysis results and the time spent in the data mining process. Generally, raw data in databases has several problems before it is preprocessed. These:

a. Areas that are not significant for the model.
b. Missing values.
c. Values that conflict with the general trends contained in the data.

d. Irrational values.
e. Fields that are not compatible with the model are meaningless values.

3.7 MISSING VALUES

Databases open to user input and data warehouses created using data from these databases may contain incomplete, inconsistent and noisy data. In the literature, such data should be referred to as dirty data and should be subjected to data cleaning processes. Kim et al. A large taxonomy of dirty data was prepared in the study of [51]. Whatever the reason, erroneous data should be corrected during data preprocessing. If all fields of a record consist of missing data, the record is excluded from analysis. If there is data in some areas of the record and data is missing in other areas, the data is completed with various missing data completion methods, according to the model to be developed and the statistical trends contained in the data. These:

a. Assigning a fixed value instead of missing data.
b. Completing the missing value with the arithmetic mean of the same type of data.
c. Replacing it with a random value produced from a statistical distribution.
d. Operations.

In order for the analysis algorithm used by the model to produce accurate and consistent results, missing data must be completed in accordance with the model and existing data. Replacing the missing values with methods such as regression and decision tree will allow more accurate and consistent analysis results to be obtained, as it will provide standardized data to the model.

3.8 MEANINGLESS VALUES

In systems where data is collected by user data entry, the user's incorrect data entry, application errors, etc. For reasons, meaningless values may occur in the data. By calculating the variance of the values contained in the data feature, meaningless and noisy values can be determined according to the deviation rate. Cluster analysis, regression, histogram etc. For the model using methods, meaningless data can be determined.

3.9 DATA REDUCTION

Data reduction is the creation of a small, concise copy of the data to be used in the analysis, without losing its meaning. Reduced data enables data mining analyzes to produce better results, thanks to benefits such as reducing the application computational complexity of data mining models and focusing on the problem with data on the main topic. Data reduction methods listed below are used in data mining:

a. Data merging or creating a data cube.
b. Size reduction.
c. Compressing the data.
d. Converting the data to discrete format.

3.10 DATA MINING MODELS

The model used in data mining can be functionally listed in three different groups. These: [52]

a. Classification and regression models.
b. Clustering models.
c. Association rules.

Two different learning model approaches are used in data mining techniques [53]:

a. In the supervised learning model, the actual result value dataset is divided into learning and test data. The model is trained with the learning dataset whose input and output values are known. After the training of the model is completed, predictions are generated with the test dataset. Error calculations are made by comparing the predicted values produced for the test dataset and the actual output values with various measurement metrics. The smaller the calculated error values, the more accurate estimates are considered to be produced.
b. In unsupervised learning, since the class to guide the learning process is not the data with known output value, it is tried to be obtained by examining the patterns in the data and creating information to guide the learning process. Classification and regression models are defined as supervised learning models, and clustering and association rules are defined as unsupervised models. In supervised models, the class of data samples in the analysis data is known; This information guides the learning and prediction process. In unsupervised learning, there is no such information. The learning algorithm used in the

analysis operates the self-learning process by examining the patterns, relationships and dependencies in the data. The classification of data mining methods is shown schematically below.

3.11 CLASSIFICATION AND REGRESSION

Models Learning in data mining is attempted by classifying the collected data in a group. Classification and regression algorithms are defined as supervised learning models. In supervised learning models, the learning algorithm trained with one part of the data set is expected to predict the result value with the other part. Regression and decision tree algorithms are known as the most important algorithms among supervised learning models. Important methods known as classification and regression algorithms are listed below [54].

3.12 DECISION TREES

In decision trees, which consist of a series of rules from root to leaves, learning is carried out by branching in the tree, according to the values of the variables in the learning model. The following features can be listed as the strengths of decision trees in the flowchart structure [55], New rules can be easily derived from existing rules in decision trees.

a. Capable of classifying data easily.
b. It can be applied to models with discrete and continuous variables.
c. The model has the ability to choose which variables are important. In decision trees, nodes show the test of features, node branches show the test results, and tree leaves show the labels of the tree classification. In decision tree models, the creation of the tree involves two processes [56].
d. Tree creation: According to the property values of the data samples, the tree is formed by forming branches and nodes from root to leaf.
e. Tree pruning: The data cleaning process of data mining is run on the tree. Values that are not important to the model, outliers and illogical values, and values that describe one-off events are deleted from the tree. In classification with a decision tree, the classification is made by determining the features that best divide the sample data. The correct determination of the features that best classify the data is important for the model to make the right decisions. The features that best divide the samples can be determined by using

the goodness function. The goodness functions used according to the data mining model classification algorithms are listed below:

f. Information gain: ID3, C4.5.

g. Gini index: Each attribute is divided by two and all possible two divisions are tested.

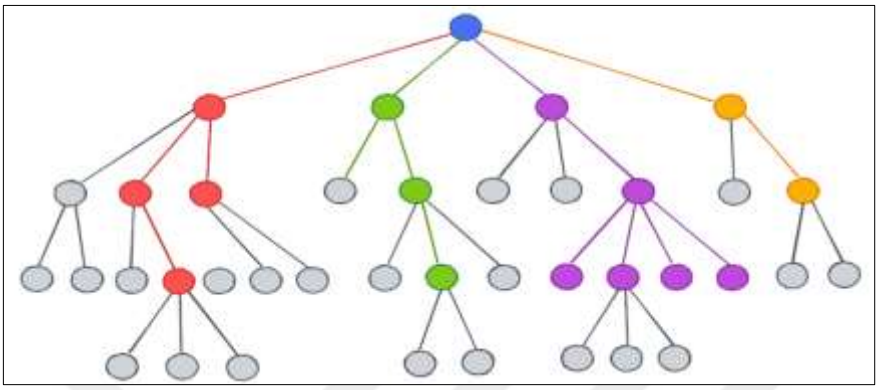

Figure 3.3: Decision tree

ID3 Algorithm, The ID3 algorithm offered by Quinlan works with categorical data [83]. With the decision tree created in the algorithm, the multidimensional data is divided into 33 sections with a determined quality, and the actions to be taken according to the features are determined at each step. The complexity of decision tree algorithms is determined by the number of nodes and leaves they contain. Therefore, techniques are used to reduce the number of nodes and leaves in the ID3 algorithm. It is important to correctly determine the nature of the decision trees to start partitioning. Failure to start partitioning from the appropriate attribute causes too many nodes and leaves in the decision tree. In the specification process for branching, the following steps are followed [57].

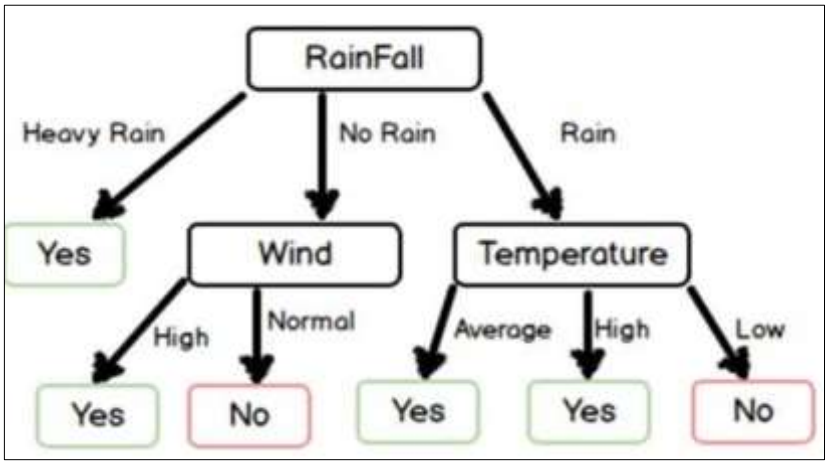

Figure 3.4: ID3 Algorithm .

3.13 K-NEAREST NEIGHBOR

(K-Nearest Neighbor) Algorithm In the k-nearest neighbor algorithm, known as the supervised learning model algorithm, the class value is estimated for the newly added observation values to the model after the learning process is run with the datasets whose class, that is, the result value, is known. In this classification technique, the distances of the observation records added to the data set are calculated, and the k closest observation records are selected as adjacent to each other and classified in the same group.

The steps of the k-nearest neighbor algorithm are listed below:

a. First of all, the number of nearest neighbors (k) parameter that will form a class is selected in the algorithm.
b. Distance functions are used in the algorithm to find the nearest neighbors according to a point to be referenced. The distance between the dataset records and the reference point is calculated; those closest in distance are grouped in the same class as neighbors.
c. The values calculated with the distance function are ordered from smallest to largest and the smallest k value is chosen as the solution.
d. As a solution, the class of the selected values is determined and the most duplicate class is selected.
e. The determined class is chosen as the estimated class value of the data observation sample used for the prediction.

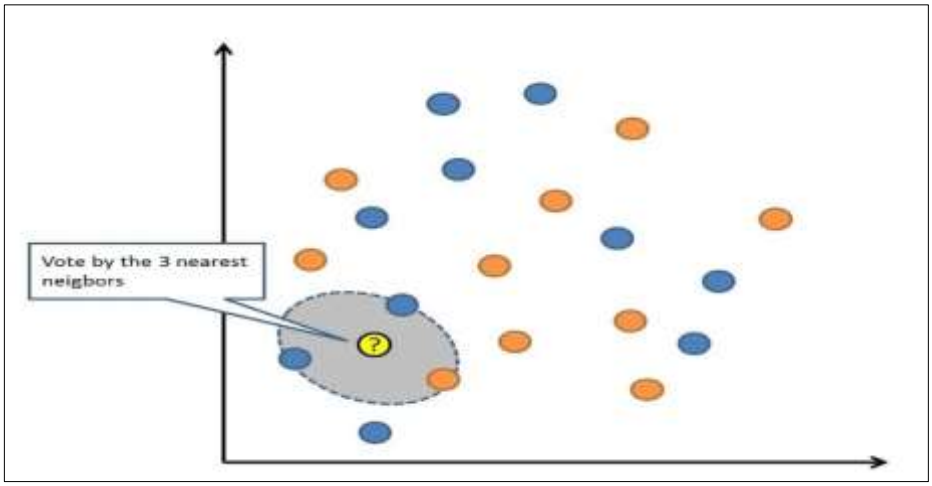

Figure 3.5: K-Nearest Neighbour Algorithm.

3.14 RANDOM FOREST ALGORITHM

Random forest algorithm is a tree-based decision-making algorithm, and the learning process is operated by creating a decision tree with randomly selected elements from the data set [58]. In the random forest algorithm, the final decision tree is obtained by summing the values obtained from the different sub-decision trees created from the data set.

Each subtree that is a member of the random forest has one vote in choosing the popular class for the input value x. These votes are used in the formation of the final result [59]. The features and advantages of the random forest algorithm are listed below:

a. It is the learning algorithm that produces the best-known result. It produces a very high accuracy classifier.
b. It works very effectively in large databases.
c. Handles thousands of input data variables without deleting the variable.
d. It gives predictions about which variable is important in the classification process.
e. While generating a forest of decision trees, it estimates the generalization error without inherent bias.
f. Although it has an effective method in estimating missing data, it maintains the accuracy of the estimation even in case of missing data to a large extent. The disadvantages of the random forest algorithm are listed below.
g. Random forest algorithm has been found to be subject to overfitting for some datasets, with noisy classification and regression tasks.
h. In datasets containing categorical variables at different levels, the random forest algorithm has been found to produce results in favor of these categorical variables at many levels.

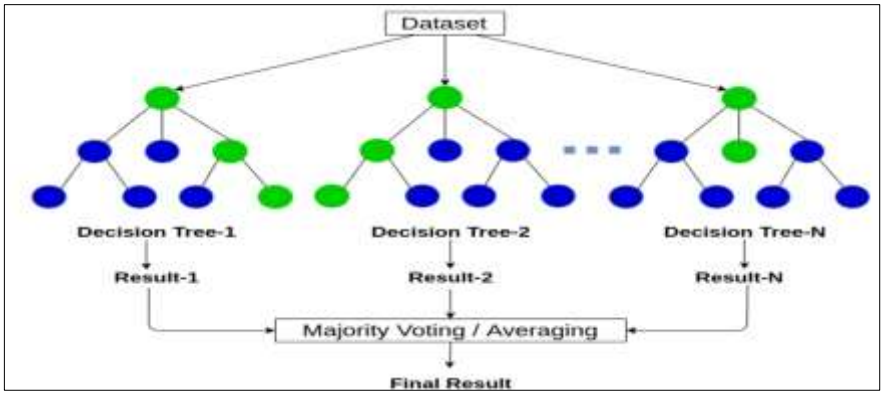

Figure 3.6: Random Forest Algorithm.

3.15 STOCHASTIC GRADIENT DESCENT ALGORITHM

Stochastic gradient descent algorithm is a popular algorithm in machine learning and forms the basis of artificial neural networks [60]. Gradient is the slope of a function. It is calculated by looking at how much a change in one of the variables affects the other. Gradient descent is mathematically a convex function whose output value is the partial derivative of the input values. The stages of the algorithm are listed below:

a. The slope of the objective function is tried to be found according to each parameter and feature. In short, gradients are tried to be found.
b. Random initial values are selected for the parameters.
c. The gradient function is refreshed by replacing the parameter values.
d. The step size for each feature is calculated by the expression (Step size = gradient * learning rate).
e. The new parameters are calculated by the expression (New parameters = old parameters - step size).
f. Repeat steps 3-5 until the gradient is 0.

The learning rate is an important parameter for the gradient descent algorithm. Choosing a high value of the learning rate causes the algorithm to advance in large steps along the slope and miss the point that receives the minimum value. For this reason, choosing a small value such as 0.01 is important for the algorithm to reach the correct values. The gradient descent algorithm moves in large steps along the slope when starting the algorithm from a high value; As it gets closer to the target, it moves with very small values in order not to miss the target value. The gradient descent algorithm works very slowly on large data sizes because it has to deal with all of the data points for each feature. Stochastic gradient descent algorithm shows high performance in data mining due to its feature of working with random data. In the stochastic gradient descent algorithm, one data is randomly selected from the data set at each stage. This greatly reduces the burden of data calculations. Instead of one data at each stage, a small sample of data can also be selected. This method, which is called mini-heap, ensures that the gradient descent produces both good predictions and the speed of the algorithm.

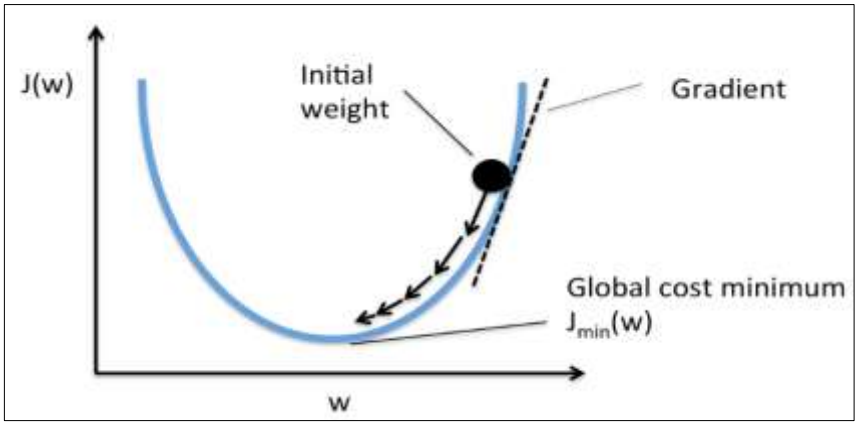

Figure 3.7: Stochastic Gradient Descent Algorithm [61].

3.16 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are a system that takes biological neurons as an example. In these learning systems, the way the human brain works is taken as an example. There are 100 million interconnected nerve cells in the human brain that communicate with each other via electromechanical signals. The connections that connect nerve cells to each other are called synapses. Each neuron is connected to thousands of neurons and receives signals from these neurons. When the sum of these signals exceeds a certain threshold value, the nerve cell responds via axons and learning takes place [62]. In this system, which takes the human brain as an example, it is not possible to construct neuron communication with the complexity of the human brain. For this reason, it is not yet possible to design artificial systems that show learning abilities in the power of the human brain.

Input values to artificial neural networks are given from neurons in the input layer shown in pink. In this system, neurons are connected to each other by connections we call synapses. Each link has a weight and a function, and these functions are applied to the input values and are propagated to the output layers. As a result, output values are taken from the layers shown in green and learning is realized. Perceptron learning algorithm was discovered by Rosenblatt in 1943. Perceptron is the mathematical function that models a neuron in the simplest way. The perceptron consists of one or more input values, processor and output values, and Figure 3.8 shows the schematic representation of a simple perceptron

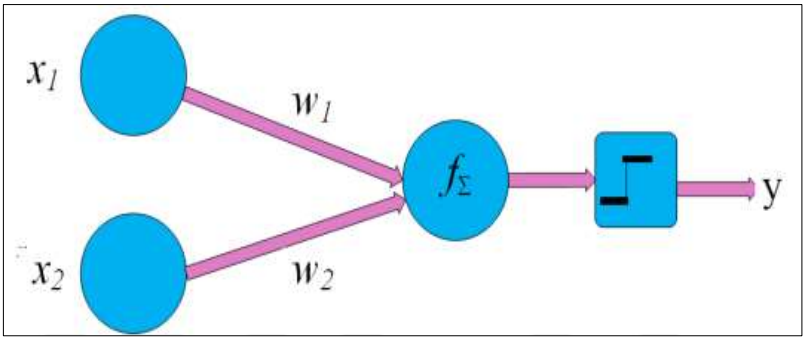

Figure 3.8: Neural Network

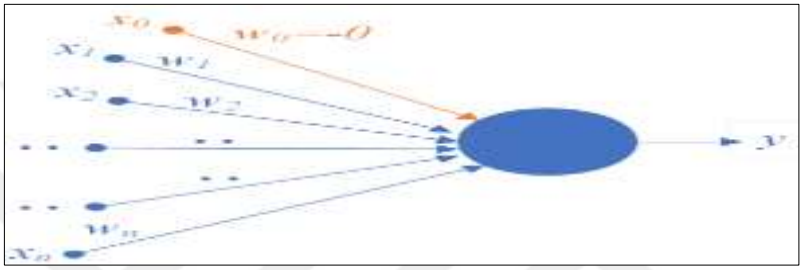

Figure 3.9: Activation Function.

3.17 CLUSTERING MODELS

In clustering models, which are an unsupervised learning model, the data set consists of data elements with similar characteristics. Data elements in the same cluster are highly similar to each other, while data elements in different clusters show little similarity [62]. Figure 3.10 in clustering models. As can be seen, it is aimed to group data elements in dissimilar clusters according to their similarities. The similarity of the data elements that make up the dataset can be calculated using distance functions. According to the distance function value, close elements are grouped in the same cluster, while distant elements are grouped in different clusters.

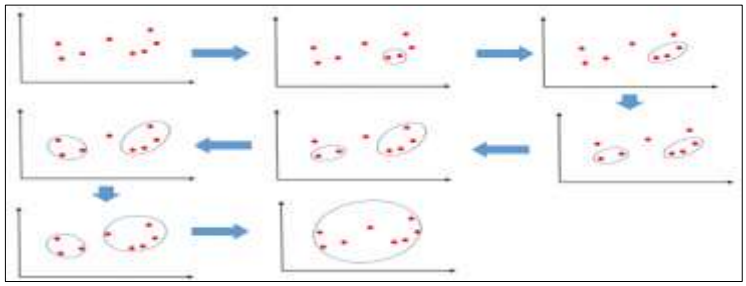

Figure 3.10: Clustering.

3.18 HIERARCHICAL CLUSTERING

In the hierarchical clustering method, unlike the k-means clustering method, the number of clusters does not need to be certain; but the algorithm needs a stop condition. The operation steps of the hierarchical clustering method, which was found by Kaufmann and Rousseau in 1990 and shown schematically in Figure 3.8, are listed below [63]:

a. The algorithm starts working by accepting each of the data elements as a cluster.
b. Distances between clusters are calculated; The clusters closest to each other are merged.
c. The distances of the clusters to each other are calculated using the single link method.
d. The process continues to run until all the data elements are collected in a single cluster.

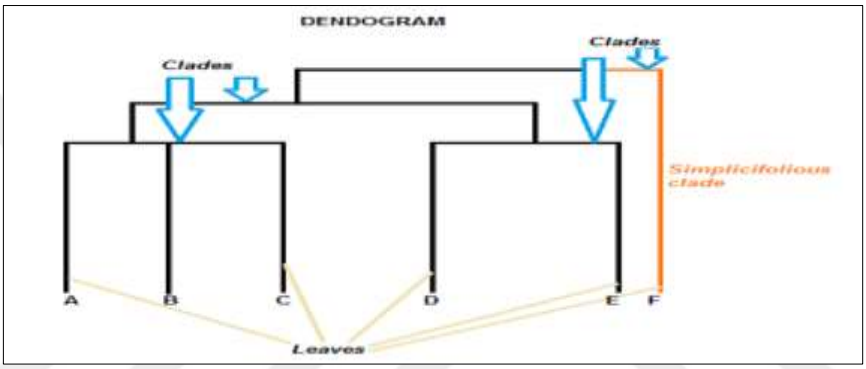

Figure 3.11: Hierarchical Clustering.

3.19 ASSOCIATION MODELS

Association models try to analyze the simultaneous occurrence of more than one event in relation to each other. In data mining with association models, the analysis of various knowledge areas is important. For example, by determining the shopping habits of the customers, it can be determined which products they prefer to buy together and increase the sales figures. Association rules and sequential time patterns are used in market basket analyzes based on data mining to determine product preference habits of customers in their shopping [64]. Some examples of association patterns used to describe simultaneous, interrelated events are listed below: 1. A customer buying a Coke from the convenience store has a 75% probability of buying potato chips. 2. The customer who buys nonfat yogurt and low-fat cheese has 85% probability of buying 0% fat milk. Sequential time patterns examine events that have a relationship, as in association rules; but events do not happen simultaneously, they happen sequentially. The probability of the next event occurring

depends on the occurrence of the previous event. Some examples of sequential time patterns can be seen below. 1. A patient who has undergone X surgery can have a Y infection with a 45% probability for up to 15 days. 2. In case of a decrease in the ISE index, if there is an increase of more than 15% in stock A, there will be an increase in stock B with a probability of 60% up to 3 working days. As can be seen from the examples, the events follow each other; takes place sequentially.

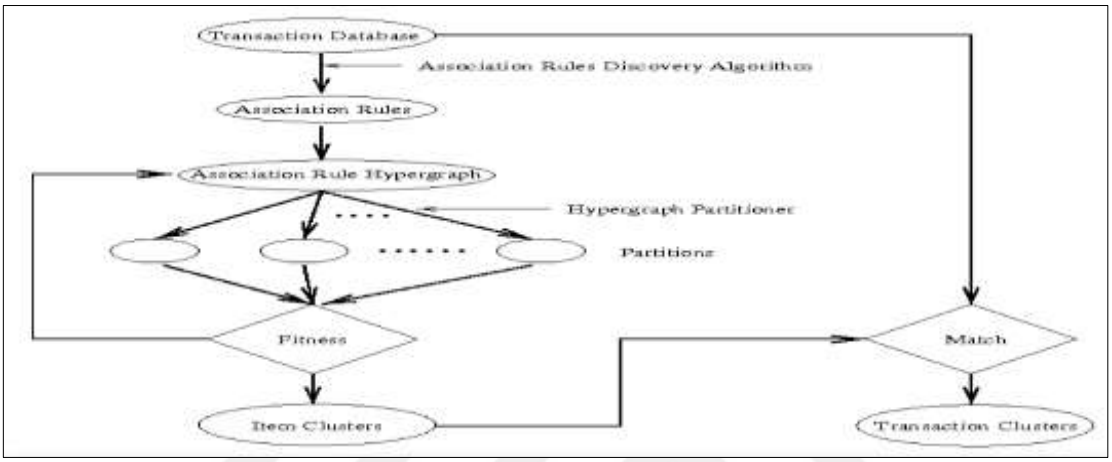

Figure 3.12 Association Models Clustering [64].

3.20 MODEL ERROR EVALUATION

In supervised learning models, the data obtained by processing the raw data with data preprocessing processes is divided into two sets as learning and test data. The algorithm trained with the learning dataset is expected to produce predictions with the test data. In the test data used in supervised learning models, the actual result value is known against the input values. The result value estimated by the learning algorithm is compared with the actual result value using various measurement metrics and the accuracy of the model is determined. The most primitive method that can be used to determine the accuracy of a learning model is the simple validation method. In a supervised learning model, where 5% to 33% of the dataset is reserved as test data and the other part as learning data, predictions are produced with test data after training with learning data. Correct and incorrectly classified events are determined by comparing the prediction results obtained with the test data with the actual results. Using these parameters, the error rate is calculated by dividing the number of incorrectly predicted events by the number of all events, and the accuracy rate by dividing the incorrectly predicted events by the number of all events. Based on the fact

that the estimated value has two states, true or false, the sum of the accuracy and error rate is 1, as a requirement of probability theory.

$$\text{Correct rate} = 1\text{- error rateI.} \tag{3.1}$$

Cross-validation methods may yield better results if the amount of data is limited. In this method, the dataset is randomly divided into two datasets, a and b, with 44 equal elements Correct rate = 1- error rate (3.1). First of all, the learning model trained with a learning dataset is expected to produce predictions with b dataset. In the second stage, the learning and test datasets are swapped and the same operations are repeated. The error rate is calculated by taking the average of the error values calculated in both stages. In datasets consisting of several thousand rows of data, if the dataset can be divided into n groups, model evaluation can be performed with the n-fold cross-validation method. In the n-fold cross-validation method, the data is divided into n data sets. Supervised learning processes are carried out by selecting one of the datasets, the test and the others learning set. This process continues in n phases, with each phase selecting one of the test datasets from the other datasets. The final error rate is calculated by taking the average of the error rates calculated at each stage. Bootstrapping is preferred when the dataset size is small. This method is determined by randomly selecting the training dataset one by one from the dataset. Randomly selected data is added back to the original dataset each time. For this reason, a selected data may be selected again, as well as not selected at all. After the learning process is run with the training dataset, the remaining data forms the test set. In order for the model to produce correct results, the samples selected as the training and test dataset should describe the model well. In the model, sampling is done by working n times for n samples, and learning and testing stages are run for each sample. The final error rate is determined by calculating the average of the calculated error rates.

3.21 FEATURE EXTRACTION

Information Extraction Accessing summary information from the document stack obtained from many data sources is called information extraction. By comparing information from web pages, information can be extracted from large-scale documents [65]. The information extraction method extracts elements and entities from the document; determines the relationships between them. Information extraction techniques infer information about the

entities of the text and their relationships by using propositions that give information about the entities in the analyzed text. Information extraction methods are important in terms of highlighting the terms and relations related to the subject of the text. In text analysis with these methods, if the purpose is to reveal unknown terms and relationships, knowledge discovery methods are used. Knowledge discovery methods reinforce the information obtained from the text with other sources. In this method, by going further than terms and relations between terms, a set of relations that are dependent on each other is created with special structures and functions. In such systems, structural data can be used together with textual data. Evaluation of the data obtained in information extraction systems is done with accuracy and sensitivity criteria as in information retrieval systems. Estimates are used as a measure of information extraction. While sensitivity is calculated by the ratio of correct predictions to all predictions, accuracy is calculated by dividing correct predictions by the number of assets found [66].

3.22 PROPOSED METHODOLOGY

In this section, new method based GWO and PSO applied to detect faults in transmission lines.

The data obtained from Kaggle and read using Python scripts. Then, the proposed method divided into two groups train and test. The data in the first stage, classified using AdaBoost to classify the datasets to two labels defect and normal classes. The PSO and GWO are applied to enhance the performance of the AdaBoost. The flowchart of the proposed method presented in the figure 3.13.

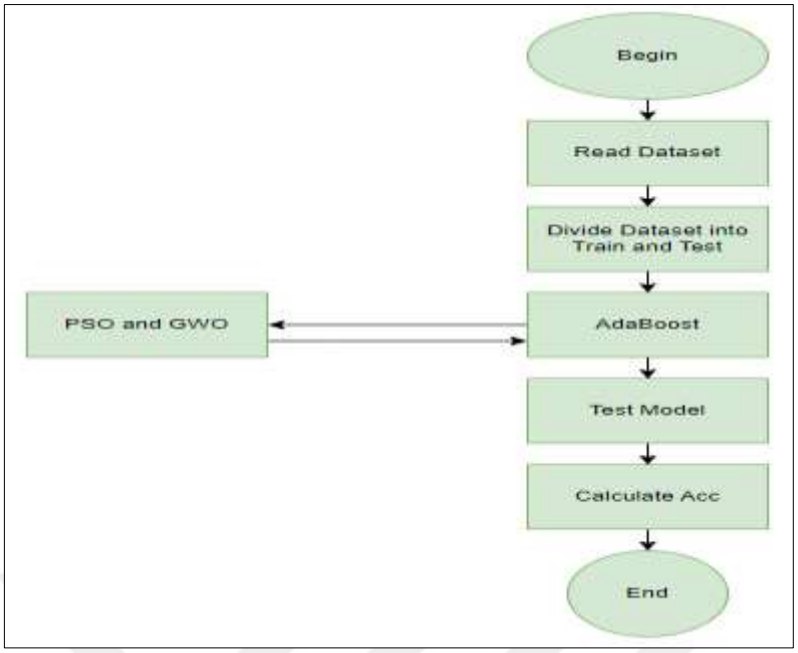

Figure 3.13 Proposed Method.

4. SIMULATION RESULTS

4.1 VALIDATION TECHNIQUES

In the system, it is necessary to verify the operations performed after the learning algorithm is run on the data. For this reason, the data are divided into training and test data. With this validation data, which is also selected from the training data, predictions are made on the trained model. In the hold-out method, which is the simplest of the validation methods, the analysis dataset consists of training data used for learning and test data used to test the accuracy of the trained model. In the K-layered cross-validation algorithm, k of the data in the training set. If it is applied to the training process once, the algorithm $(k-1)$ is used as the validation set at once. If the value of k is large, a decrease in the variance value is encountered. In one-output validation, k is the number of data points in the dataset n. The algorithm is trained with data other than a point inside the n point. The improvement function is tested with the data on the relevant point. While evaluating these methods, they are matched using the square error method.

4.2 MACHINE LEARNING ALGORITHM EVALUATION

Applying a machine learning algorithm to a dataset and obtaining results does not mean that the solution to the problem has been accurately defined. Therefore, it is necessary to determine the effectiveness of the solution by various methods. The algorithm can be evaluated with many metrics in which the dataset used in machine learning is decisive. Another important measure in determining the efficiency of the machine learning model is the data to be used. Using the training dataset as a dataset in performance measurement will cause bias in the model. During the training, the model preparing itself according to the training data will produce good results if the training data are used as test data. For this reason, it would be more accurate to divide the learning dataset into two as training and test data and to make predictions with the test data after the learning process is run with the training dataset [132][133]. Accuracy and precision criteria are important in the evaluation of machine learning algorithms. In consecutive estimations, the fact that the actual values are close to the predicted value, but the estimated values are far from each other, indicates high estimation accuracy and low sensitivity. If the prediction values are close to each other but far from the true value, the precision of the prediction values will be high and the

accuracy will be low [132][134]. Errors encountered in machine learning algorithms can generally be grouped into two categories:

a. Systematic errors: These are the errors that occur within a pattern. They point to a systemic problem.
b. Random errors: Errors that do not occur within a pattern.

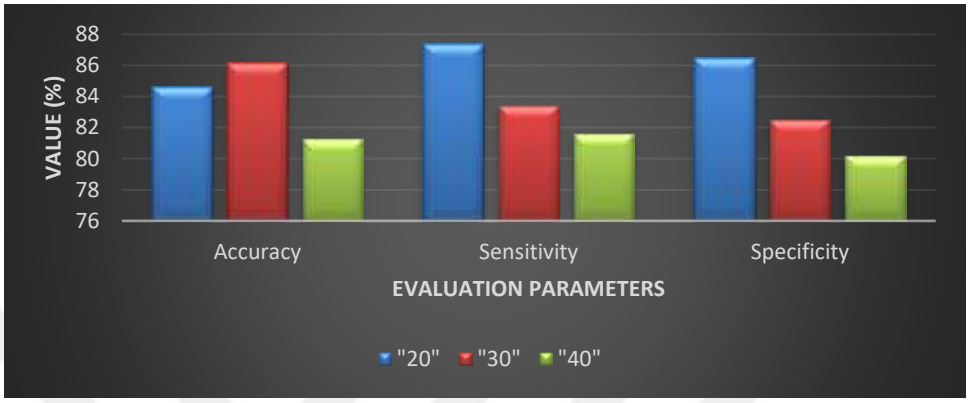

Figure 4.1: Results of KNN [59].

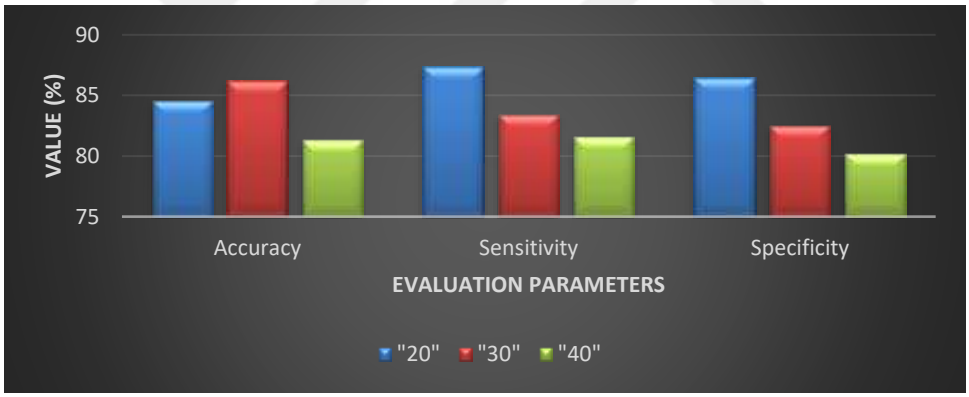

Figure 4.2: Results of DT [59].

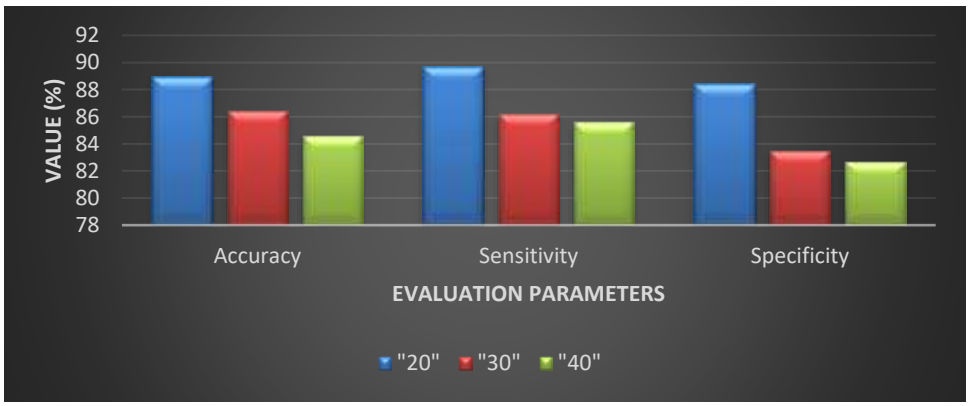

Figure 4.3: Results of RF [59].

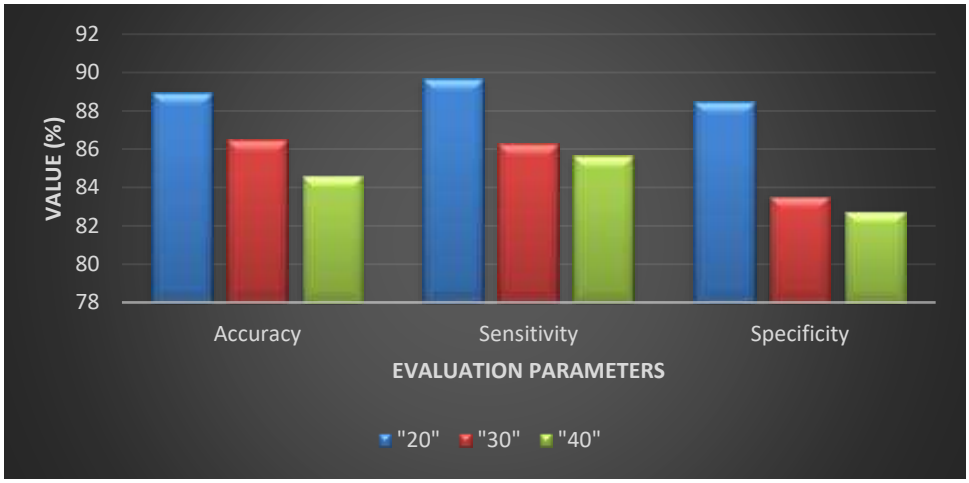

Figure 4.4: Results of ANN [59].

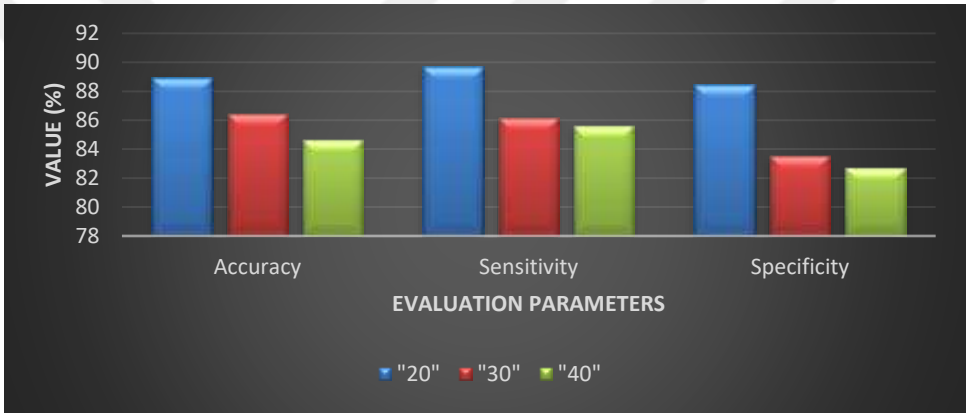

Figure 4.5: Results of AdaBoost [59].

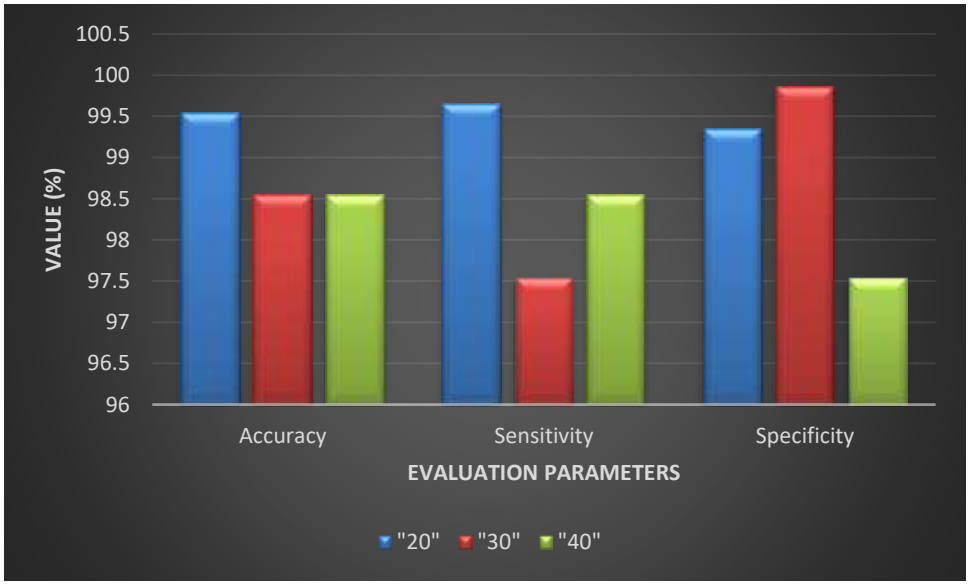

Figure 4.6: Results of PSO and GWO based AdaBoost [59].

5. CONCLUSIONS

Electricity generation, transmission, distribution services and other components have to be connected to each other through a communication channel. Communication is the most essential component of the smart grid concept. Business and remote monitoring are components in which both consumers and service providers play a role. Service providers are generally considered as components that provide electricity distribution service and communication service. Another key component is consumers. Consumers can follow the instant meter reading results and carry out their energy efficiency transactions and choose the most suitable tariff according to the tariff structure offered by the retail service provider. Electricity distribution companies, electricity retail companies and electricity generation companies that make up the electricity market, direct the electricity market by trying to make the most accurate energy demand and forecasts by monitoring the instantaneous consumption of consumption points and generation points with remote meter reading systems, which is one of the smart grid operations.

Since the algorithms applied during the implementation of the learning algorithms are supervised learning algorithms, the dataset was divided into two, half of it was used for learning, the other half was used for estimation and the jobs of the future were tried to be predicted with the data set at hand.

In this study, the proposed method applied GWO based on PSO and AdaBoost to detect faults in the transmission lines. The proposed method presented 99.64% accuracy which is high when compared with several traditional techniques.

REFERENCES

[1] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, "Cost-based modeling for fraud and intrusion detection: Results from the JAM project," in Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00, vol. 2, pp. 130–144,2000.

[2] M. Alkasassbeh, G. Al-Naymat, A. B. A. Hassanat, and M. Almseidin, "Detecting distributed denial of service attacks using data mining techniques," International Journal of Advanced Computer Science and Applications, vol. 7, no. 1, 2016.

[3] A. Abraham, C. Grosan, and C. Martin-Vide, "Evolutionary design of intrusion detection programs.," Int. J. Netw. Secur., vol. 4, no. 3, pp. 328–339, 2007.

[4] H. Liu, Y. Sun, and M. S. Kim, "Fine-grained DDoS detection scheme based on bidirectional count sketch," in 2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN), pp. 1–6, 2011.

[5] C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," Comput Secur, vol.70, pp. 255–277, 2017.

[6] A. Abraham and J. Thomas, "Distributed intrusion detection systems: a computational intelligence approach," in Applications of information systems to homeland security and defense, IGI Global,pp. 107–137,2006.

[7] N. Sharma and S. Mukherjee, "A novel multi-classifier layered approach to improve minority attack detection in IDS," Procedia Technology, vol. 6, pp. 913–921, 2012.

[8] H. Alshemmari, A. E. Al-Shareedah, S. Rajagopalan, L. A. Talebi, and M. Hajeyah, "Pesticides driven pollution in Kuwait: The first evidence of environmental exposure to pesticides in soils and human health risk assessment," Chemosphere, vol.273, 2021.

[9] N. Patani and R. Patel, "A mechanism for prevention of flooding-based DDoS attack," International Journal of Computational Intelligence Research, vol. 13, no. 1, pp. 101–111, 2017.

[10] Y. Feng, R. Guo, D. Wang, and B. Zhang, "Research on the active DDoS filtering algorithm based on IP flow," in 2009 Fifth International Conference on Natural Computation, vol. 4, pp. 628–632,2009.

[11] A. Meek, "DDoS attacks are getting much more powerful and the Pentagon is scrambling for solutions." BGR, 2015.

[12] J. Mirkovic and P. Reiher, “A taxonomy of DDoS attack and DDoS defense mechanisms,” ACM SIGCOMM Computer Communication Review, vol. 34, no. 2, pp. 39–53, 2004.

[13] S. T. Zargar, J. Joshi, and D. Tipper, “A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks,” IEEE communications surveys & tutorials, vol.15, no.4, pp. 2046–2069, 2013.

[14] T. T. Oo and T. Phyu, “Analysis of DDoS detection system based on anomaly detection system,” in international conference on advances in engineering and technology (ICAET’2014). Singapore, 2014.

[15] S. Sinha and M. Sharma, “Simulation and Analysis of DDoS Attacks by Specialized Simulator using Virtualization,” International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol. 3, no. 2, 2015.

[16] B. Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, “Deep learning for classification of malware system call sequences,” in AI 2016: Advances in Artificial Intelligence: 29th Australasian Joint Conference, Hobart, TAS, Australia, December 5-8, 2016, Proceedings 29, pp. 137–149, 2016.

[17] B. Hang, R. Hu, and W. Shi, “An enhanced SYN cookie defense method for TCP DDoS attack,” Journal of Networks, vol. 6, no. 8, 2011.

[18] D. Wang and J. Jie, “A multi-core-based DDoS detection method,” in 2010 3rd international conference on computer science and information technology, vol. 4, pp. 115–118,2010.

[19] N. Hoque, H. Kashyap, and D. K. Bhattacharyya, “Real-time DDoS attack detection using FPGA,” Comput Commun, vol. 110, pp. 48–58, 2017.

[20] J. Singh, M. Sachdeva, and K. Kumar, “Detection of DDOS attacks using source IP based entropy,” vol. 3, pp. 201–210, 2013.

[21] S. M. Lee, D. S. Kim, J. H. Lee, and J. S. Park, “Detection of DDoS attacks using optimized traffic matrix,” Computers & Mathematics with Applications, vol. 63, no. 2, pp. 501–510, 2012.

[22] H. Rahmani, N. Sahli, and F. Kamoun, “DDoS flooding attack detection scheme based on F-divergence,” Comput Commun, vol. 35, no. 11, pp. 1380–1391, 2012.

[23] G. B. Senirkentli *et al.*, “Proton therapy for mandibula plate phantom,” in healthcare, vol. 9, no. 2, pp. 167,2021.

[24] V. P. Nigam and D. Graupe, “A neural-network-based detection of epilepsy,” Neurol Res, vol. 26, no. 1, pp. 55–60, 2004.

[25] A. Abraham and C. Grosan, “Evolving intrusion detection systems,” Genetic Systems Programming: Theory and Experiences, pp. 57–79, 2006.

[26] L. Deng and D. Yu, “Deep learning: methods and applications,” Foundations and trends® in signal processing, vol. 7, no. 3–4, pp. 197–387, 2014.

[27] K. A. K. Niazi, S. A. Khan, A. Shaukat, and M. Akhtar, “Identifying best feature subset for cardiac arrhythmia classification,” in 2015 Science and Information Conference (SAI), pp. 494–499,2015.

[28] A. Shenfield, D. Day, and A. Ayesh, “Intelligent intrusion detection systems using artificial neural networks,” Ict Express, vol. 4, no. 2, pp. 95–99, 2018.

[29] A. Kaushik, H. Gupta, and D. S. Latwal, “Impact of feature selection and engineering in the classification of handwritten text,” in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 2598–2601,2016.

[30] F. Grisoni et al., “Combining generative artificial intelligence and on-chip synthesis for de novo drug design,” *Sci Adv*, vol. 7, no. 24, 2021.

[31] S. Ibrahim, R. Djemal, and A. Alsuwailem, “Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis,” Biocybern Biomed Eng, vol. 38, no. 1, pp. 16–26, 2018.

[32] N. Kohli, N. K. Verma, and A. Roy, “SVM based methods for arrhythmia classification in ECG,” in 2010 international conference on computer and communication technology (ICCCT), pp. 486–490,2010.

[33] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, “Efficient processing of deep neural networks: A tutorial and survey,” Proceedings of the IEEE, vol. 105, no. 12, pp. 2295–2329, 2017.

[34] D. Petkovic *et al.*, “Setap: Software engineering teamwork assessment and prediction using machine learning,” in 2014 IEEE frontiers in education conference (FIE) proceedings, pp. 1–8,2014.

[35] C. Sobie, C. Freitas, and M. Nicolai, “Simulation-driven machine learning: Bearing fault classification,” Mech Syst Signal Process, vol. 99, pp. 403–419, 2018.

[36] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, “Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks,” Neural Networks, vol. 129, pp. 313–322, 2020.

[37] J. H. Lee, J. Shin, and M. J. Realff, “Machine learning: Overview of the recent progresses and implications for the process systems engineering field,” Comput Chem Eng, vol. 114, pp. 111–121, 2018.

[38] K. Nigam, A. McCallum, and T. M. Mitchell, “Semi-Supervised Text Classification Using EM.” 2006.

[39] S. Ray, S. Bansal, A. Gupta, D. Gupta, and F. Shaikh, “Understanding Support Vector Machine algorithm from examples (along with code),” Analytics Vidhya, vol. 13, pp. 19, 2017.

[40] A. M. Karim, Ö. Karal, and F. v Çelebi, “A new automatic epilepsy serious detection method by using deep learning based on discrete wavelet transform,” vol. 4, pp. 15–18, 2018.

[41] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. v Çelebi, “A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing,” Biocybern Biomed Eng, vol. 39, no. 1, pp. 148–159, 2019.

[42] S. M. H. Bamakan, H. Wang, and Y. Shi, “Ramp loss K-Support Vector Classification-Regression; a robust and sparse multi-class approach to the intrusion detection problem,” Knowl Based Syst, vol. 126, pp. 113–126, 2017.

[43] A. M. Karim, F. V. Çelebi, and A. S. Mohammed, “Software Development for Blood Disease Expert System,” Lecture Notes on Software Engineering, vol. 4, no. 3, pp. 179, 2016.

[44] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. v Çelebi, “A new generalized deep learning framework combining sparse autoencoder and Taguchi method for novel data classification and processing,” Math Probl Eng, vol. 2018, 2018.

[45] L. Wang *et al.*, “Parameter optimization of a four-legged robot to improve motion trajectory accuracy using signal-to-noise ratio theory,” Robot Comput Integr Manuf, vol. 51, pp. 85–96, 2018.

[46] A. Karim, “Development of secure Internet of Vehicle Things (IoVT) for smart transportation system,” Computers and Electrical Engineering, vol. 102, p. 108101, 2022.

[47] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” Physica D, vol. 404, 2020.

[48] A. M. Karim, H. Kaya, V. Alcan, B. Sen, and I. A. Hadimlioglu, “New optimized deep learning application for COVID-19 detection in chest X-ray images,” Symmetry (Basel), vol. 14, no. 5, p. 1003, 2022.

[49] G. Shen, W. Zhou, W. Zhang, N. Liu, Z. Liu, and X. Kong, “Bidirectional Spatial-Temporal Traffic Data Imputation via Graph Attention Recurrent Neural Network,” Neurocomputing, 2023.

[50] Y. Chen, “Voltages prediction algorithm based on LSTM recurrent neural network,” Optik (Stuttg), vol. 220, 2020.

[51] N. Andrei, “A Dai–Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization,” Appl Math Lett, vol. 21, no. 2, pp. 165–171, 2008.

[52] A. M. Karim and A. Mishra, “Novel COVID-19 Recognition Framework Based on Conic Functions Classifier,” Healthcare Informatics for Fighting COVID-19 and Future Epidemics, pp. 1–10, 2022.

[53] A. Jentzen and P. von Wurstemberger, “Lower error bounds for the stochastic gradient descent optimization algorithm: Sharp convergence rates for slowly and fast decaying learning rates,” J Complex, vol. 57, 2020.

[54] R. Ye and Q. Dai, “Implementing transfer learning across different datasets for time series forecasting,” Pattern Recognit, vol. 109, 2021.

[55] Y. Liu, A. Li, X.-M. Zhao, and M. Wang, “DeepTL-Ubi: a novel deep transfer learning method for effectively predicting ubiquitination sites of multiple species,” Methods, vol. 192, pp. 103–111, 2021.

[56] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, “Few-shot transfer learning for intelligent fault diagnosis of machine,” Measurement, vol. 166, 2020.

[57] A. Karim, “Development of secure Internet of Vehicle Things (IoVT) for smart transportation system,” Computers and Electrical Engineering, vol. 102, 2022.

[58] X. Yang, Y. Zhang, W. Lv, and D. Wang, “Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier,” Renew Energy, vol. 163, pp. 386–397, 2021.

[59] C. Li, S. Zhang, Y. Qin, and E. Estupinan, “A systematic review of deep transfer learning for machinery fault diagnosis,” Neurocomputing, vol. 407, pp. 121–135, 2020.

[60] Z. Qiu, S. Zhao, X. Feng, and Y. He, “Transfer learning method for plastic pollution evaluation in soil using NIR sensor,” Science of The Total Environment, vol. 740, 2020.

[61] A. M. Karim, H. Kaya, V. Alcan, B. Sen, and I. A. Hadimlioglu, “New optimized deep learning application for COVID-19 detection in chest X-ray images,” Symmetry (Basel), vol. 14, no. 5, 2022.

[62] S. Kim, Y.-K. Noh, and F. C. Park, “Efficient neural network compression via transfer learning for machine vision inspection,” Neurocomputing, vol. 413, pp. 294–304, 2020.