REPUBLIC OF TÜRKİYE

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

HYBRID METHOD BASED DEEP BELIEF NETWORKS FOR MALWARE DETECTION IN IOTS

Ranya Ihsan Ali ALLABAN

Master's Thesis

Supervisor

Asst. Prof. Dr. Oguz KARAN

Istanbul, 2022

HYBRID METHOD BASED DEEP BELIEF NETWORKS FOR MALWARE DETECTION IN IOTS

Ranya Ihsan Ali ALLABAN

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled HYBRID METHOD BASED DEEP BELIEF NETWORKS FOR MALWARE DETECTION IN IOTS prepared by RANYA IHSAN ALI ALLABANAND submitted on 12/12/2022 has been accepted unanimously for the degree of Master of Science in Information Technologies.

____

Asst. Prof. Dr. Oguz KARAN

Supervisor

Thesis Defense Jury Members:

Asst. Prof. Dr. Oguz KARAN	Faculty of Engineering and Architecture, Altinbas University	__________________
Asst. Prof. Dr. Sefer KURNAZ	Faculty of Engineering and Architecture, Altinbas University	__________________
Assoc. Prof. ADİL DENİZ DURU	Faculty Of Sport Sciences, Sport Coaching, Marmara 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 Graduate Education Institute: ____/____/____

DEDICATION

To the first supporter my husband who made this dream possible ..

To the candle who lighted my life my sweet child murad

To my dear father and mother and all my family who have encouraged me to succeed..

Your presence , love , support, attention, understanding makes every thing better.

ABSTRACT

HYBRID METHOD BASED DEEP BELIEF NETWORKS FOR MALWARE DETECTION IN IOTS

ALLABAN Ranya Ihsan Ali

MSc, Electrical and Computer Engineering, information tecnology Istanbul Altınbaş University,

Supervisor: Asst. Prof. Dr. Oguz Karan

Date: December , 2022

Pages: 71

Malware is any software deliberately intended to reason disturbance to a computer, attendant, customer, or computer network, escape secluded information, improvement illegal admission to info or schemes, divest users access to info or which naively delays with the user's processer security and privacy.

In this study, new method based DBN applied to detect malware attacks in IoTs. The presented method combined DBN with LDA which used to reduce the size of input features by selecting important features. then, the selected features wired to the DBN that trained to classify these features to normal and abnormal labels...

Keywords: LDA, DBN, Machine Learning.

TABLE OF CONTENT

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

RNN	:	Recurrent Neural Network
KNN	:	Nearest neighbor algorithm
DL	:	Deep Learning
AI	:	Artificial Intelligence

1. INTRODUCTION

Cybersecurity attacks have become increasingly important under the current conditions for both humanity and the states of the world. Although there are many attack methods and scopes, their transformation into a threat has evolved into a broad perspective, ranging from individual damage to cyber warfare, and has become a threat. To counter these threats, many defense and security measures are taken or attempted to be taken [1].

IoT is rapidly evolving as a new approach to the development of modern wireless communications. The main idea of this concept is that common objects or objects in our environment interact with each other, cooperate with neighbors and use radio frequency identification (RFID), sensors, mobile phones, etc [2]. It interacts with unique addressing schemes such as The purpose of the common Internet of Things is to connect and integrate the physical world and cyberspace. Internal structures that communicate with each other through protocols. From a conceptual perspective, the IoT is based on three capabilities of smart objects:

i. Being identifiable,

ii. communicating, and

iii. Interacting – with each other, with interconnected objects, with end users or other entities in the network. The properties of smart objects are listed as follows:

a) Having a physical shape and physical properties (eg size, shape, etc.).

b) They should have minimal communication functions such as being found when called, accepting and responding to incoming messages.

c) Have a unique identifier.

d) Correlated with at minimum one big name and one direct. The given name is a social-legible explanation of the point and can be applied for rational reasons. The address is a robot-readable loop used to connect with the body [3].

e) Have a few simple processing skills. This can vary from the capability of an external communication to match a particular use case (as with passive RFIDs) to the ability to perform highly complex calculations, including service discovery and network management tasks [4].

f) They may have means of sensing natural experiences (eg heat, light, electromagnetic energy point) or initiating activities that involve natural life (actuators). After a structures-level viewpoint, the Internet of Things can be deemed as a very active and fundamentally dispersed interacted system comprising of a larger amount of smart items that create and use data.

The benefits from IoT applications are enormous for both people and companies. Approximately of the talented requests are: advanced worldwide source cable logistics management, creation discovery, manufacture mechanization, keen homelands and applications, e-government (electric authorized permits and currency), progressive combined fleet organization, health organization and e-health organization (enduring nursing and enduring annals) requests [5]. The Internet of Things creates a number of advantages over the application areas we briefly mentioned above when compared to traditional information systems. These advantages are; It can be counted as inclusion of different types of devices in the network, measuring activities, widespread data exchange with wireless networks in the vicinity, energy saving, monitoring capabilities, self-organization ability, joint data management between information systems, integrated security and privacy protection mechanisms. The biggest change in the lifestyle of mankind was the agricultural revolution by adopting a settled life 10,000 years ago. Food production improved, which supported population growthThe agricultural revolution that began in the late eighteenth century was followed by a series of industrial revolutions. This process caused a shift from muscular strength to mechanical strength. The first industrial revolution. Around 1840, the construction of the railways and the introduction of the steam engine led to the production of machinery. The second industrial revolution. Mass production began in the 19th century with the help of power lines and assembly in the late 20th century. The third industrial revolution. It started in the 1960s and developed under the influence of semiconductors, mainframe computers (the 1960s), personal computers (the 1970s and 1980s) and the Internet (the 1990s), also called the computer revolution or the digital revolution. According to the various definitions and scientific publications used to describe the initial and third industrial revolutions, we are on the cusp of the fourth industrial revolution. This revolution started in this century and built on the digital revolution. Ohira. It's cheaper and cheaper to use ubiquitous mobile internet, and it features machine learning with more powerful sensors and artificial intelligence. Due to the convergence of physical and digital things in the era of the Fourth Industrial Revolution, production based on instant and dynamic data processing continues without restrictions in the application development process in many areas [6,7]. There are 9 technological elements that led to the fourth industrial revolution. These elements are:

i. Big data analysis,

ii. Internet of things,

iii. Smart robots,

iv. Simulation,

v. Additive manufacturing (e.g. 3D printing),

vi. Horizontal/vertical software integration,

vii. Cloud computing,

viii. Augmented reality,

ix. Cyber security.

The Internet of Things is the key bridge that will enable the fourth industrial revolution to bring physical and digital applications closer together. In the perspective of the fourth industrial revolution, thanks to connected technologies and different platforms, the Internet of Things is being rewritten in terms of the relationship between people and things (goods, services, places, etc.). This technological development has made the Internet of Things one of the most important factors that have led to the Fourth Industrial Revolution, also known as Industry 4.0. The terms "Industry 4.0" and "Internet of Things" are widely used in the literature. RFID (Radio Frequency Identification) is a short range communication technology in which RFID tags use radio frequency electromagnetic fields to interact with RFID readers. Physically, RFID tags are similar to self-adhesive tags [8]. It is a small electronic chip with an antenna, commonly used to receive signals from readers and to transmit identification tags. Dimensions may be too small. Hitachi designed a poster measuring 0.4mm x 0.4mm x 0.15mm. Three types of RFID tags are used. Passive RFID tags allow you to read tags based on RF energy to identify tags and objects. No battery or electricity. These applications are used for supply chain, passport issuance, electronic tax collection and product tracking. Active RFID tags have their own battery/power supply and can communicate with the reader. Smart tags can be connected to external sensors to monitor temperature, pressure, chemicals and other conditions. Active RFID tags are used for wallet management in manufacturing, in hospital laboratories and in remote sensing computer systems. Semi-passive RFID tags, on the other hand, use energy to power the chip and increase read range during communication. Active and semi-manual RFID tags are more expensive than manual tags. Tags can contain many different types of data, but the most common type of data used in IoT applications is electronic product code (EPC). EPC is a globally unique object identifier. These unique identifiers ensure that objects tagged with RFID tags have unique identifiers. The benefits of RFID asset tracking for the Internet of Things are well known. The technology has applications in logistics, supply chain management, aviation, food safety, retail, utilities and more. Walmart, the US Department of Defense and many other organizations actively use RFID technology. However, the tracking capabilities provided by RFID mean that IoT is commonly implemented in applications, and the advantage of RFID means that data can be used remotely over the Internet. Scalable RFID is not a new technology being developed for the Internet of Things. It is known to have been used after World War II [9,10].

1.1 PURPOSE OF THIS STUDY

The main purpose of this research is to evaluate experienced or virtual cyber attack scenarios, to increase the awareness of institutions or individuals about the harm that may arise, to increase the cyber security measures based on the severity of the situation, to get help from professional experts, to use the necessary security software and to keep them up-to-date. For this purpose, it has been tried to create awareness about the following stages.

i. What is the concept of cyber attack?

ii. For what purposes are cyber attacks organized?

iii. To what extent can personal damages reach as a result of cyber attack?

iv. To what extent can corporate losses reach as a result of cyber attacks?

v. To what extent can the damages that may occur in the states and globally as a result of the cyber attack reach?

vi. What are the precautions to be taken by evaluating cyber attack scenarios?

vii. Can the damages that may occur as a result of cyber attacks be prevented by raising awareness before they arise?

viii. How can we developed new methods based machine learning techniques to detect cybersecurity attacks?

ix. What are the differences in the methods and purposes used by instant attacks and timed attacks?

2. RELATED WORKS

2.1 INTERNET OF THINGS CONCEPTS

In this learning, a machine or system performs a specific task in order to maximize its output from its inputs. Reinforcement learning is defined as a type of learning that updates metrics to increase reward and decrease punishment. In reinforcement learning, inputs come from the external environment and outputs are produced in the form of actions. The goal of this learning is to find behavioral processes that interact with machines and systems, get feedback from the environment, and maximize rewards. Rewards or penalties are provided in the environment and settings are updated to get the best possible rewards and avoid penalties. With the development of IoT technology and the gradual development and application of IoT technology, new information and communication systems have been developed. The goal of new information and communication systems is to connect the physical world with networks such as the Internet. Technologies such as the Internet of Things are important technologies for building information and communication systems. The Internet of Things includes a wide range of elements used in our daily lives, such as consumer electronics, smart home systems, and medical devices such as modems and telephones. Thanks to this network, communication between these objects occurs in a loop without human intervention, and the whole process can be clearly defined at the object level. The concept of the Internet of Things (IoT), referred to in the literature as the "Internet of Things", defines a new technological model that allows different and similar devices to connect to the Internet through networks in different ways. give. Internet connection via the Internet with additional activation of actuators such as sensors and controllers. According to another definition, the concept of the Internet of Things is defined simply as a link between global infrastructures that enables communication between different devices. Radio Frequency Identification (RFID) technology is one of the key elements of how IoT technology works and has been extensively studied in the literature, but it is important for IoT technology to distinguish between RF devices and people. This is a component. The use of sensors and RFID in IoT is an example of how business owners and IT administrators use data for everything (IoT in that sense). After a brief introduction, it makes sense to talk about the development technology of the Internet of Things and modern RFID technology [11,12,13].

Figure 2.1: IoTs

2.2 IOT CONCEPT, PURPOSE AND DEFINITION

Internet of Things technology is a technological infrastructure that people and businesses have used for a long time but have not yet adopted. Today, the Internet of Things is a concept used when viewed through a phone, internet-connected modem, or smart device. However, it can be said that the scope and definition of this technology have changed significantly in recent years due to the rapid development and utilization of this technology. IoT devices include RFID tags, sensors, cell phones, refrigerators, barcodes, and many other devices that communicate wirelessly. In this context, the Internet of Things generally refers to an open standards-based technology infrastructure that enables interactions between devices on the Internet using proprietary addressing schemes [14,15]. However, there are other uses and standards for the Internet of Things, especially in the business world. The use of the Internet of Things raises the question of business scale. Businesses today use legacy technologies such as barcodes, QR codes, modems, and sensors. However, these techniques are designed for monitoring purposes, not real-time data mining. The main functions of IoT goals are real-time data flow, information flow and processing capabilities. In this context, it is important to know what the Internet of Things is and understand its scope. The Internet of Things runs on a variety of technologies and platforms. The Internet of Things describes a technological infrastructure that allows different technologies to interact with each other and exchange data. IoT devices include smartphones, toys, heavy equipment, consumer electronics, and almost anything that connects to the Internet. Most importantly, all things can connect to the Internet in an electronic environment through this medium and exchange data with other things. To support this statement, the Internet of Things is defined as a model of connecting things to the Internet using wireless or wireless technologies to achieve desired goals. This is because the Internet of Things is defined as a combination of pervasive computing, the Internet, sensors, communication technologies, and embedded devices. One of the main goals of IoT technology is to reduce the distance between physical objects and online platforms. The concept of objects in a network infrastructure represents real or virtual actors involved, such as real objects, people, virtual data, intelligent software agents, and so on. Therefore, another goal of the Internet of Things is to create an environment where basic information about heterogeneous connected objects can be effectively shared with others in real time [16,17].

An IoT connection between devices is similar to a connection between a computer and the Internet. Communication between IoT devices is managed by a central node and logging is required as each device connects. Besides these connectivity similarities, the Internet of Things offers just as many possibilities as the Internet. What you can do with the Internet of Things depends on the hardware and technology you use. IoT devices can detect, track, automate, and control things. In addition to all these functions, there are many industrial functions and applications such as IoT, security, monitoring and control, payment, remote management and maintenance, metering, etc. For IoT devices, the possibilities and fields are enormous. From a business perspective, IoT has many industries and applications, such as product monitoring, smart home managers managing smart homes, and doctors monitoring patients remotely [18,19]. After defining what the concept of IoT is and its definition, some definitions made for this technology in the literature are as follows;

2.3 THE DEVELOPMENT OF IOT TECHNOLOGY AND ITS RELATION TO RADIO FREQUENCY IDENTIFICATION (RFID)

The term "internet of things" first became popular in 1999 when the Automatic Identification Center at the Massachusetts Institute of Technology (MIT) began developing and implementing an enterprise-grade radio frequency identification (RFID) infrastructure. However, the focus is: the business is not about the Internet of Things, but about technologies such as electronic encryption and automatic identification. With the development of these technologies, the Internet of Things has also gradually developed. Gartner, Inc. It is a global company that provides advice, tools, and information as a service to business leaders, including information technology, finance, general sales and marketing. At the top of the list published in 2014, IoT ranked as the most advanced technology and potential. The Internet of Things connects things, providing fast connectivity and data exchange between computing devices, applications, and other devices. It allows companies and individuals to communicate in constant interaction. These factors bring many benefits to businesses, including: b. Communication and information exchange that reduces effort and costs. The Internet of Things (IoT), which provides communication between people and things through the rapid development and use of the Internet in various fields, depends on the availability, security, stable structure and adaptability of wireless devices in the network. It uses Transmission Control Protocol (TCP) and Internet Protocol (IP). It has become a new technological phenomenon. In this sense, the Internet of Things, which facilitates communication between objects, can connect these intelligent objects with technologies such as sensors. The development of the Internet of Things is accelerating due to the introduction and development of technologies such as RFID and Wireless Sensor Networks (WSN), which allow and simplify remote access to the greatest extent. These technologies make it possible to connect things, people, things, things, and things anytime and anywhere [20,21,22].

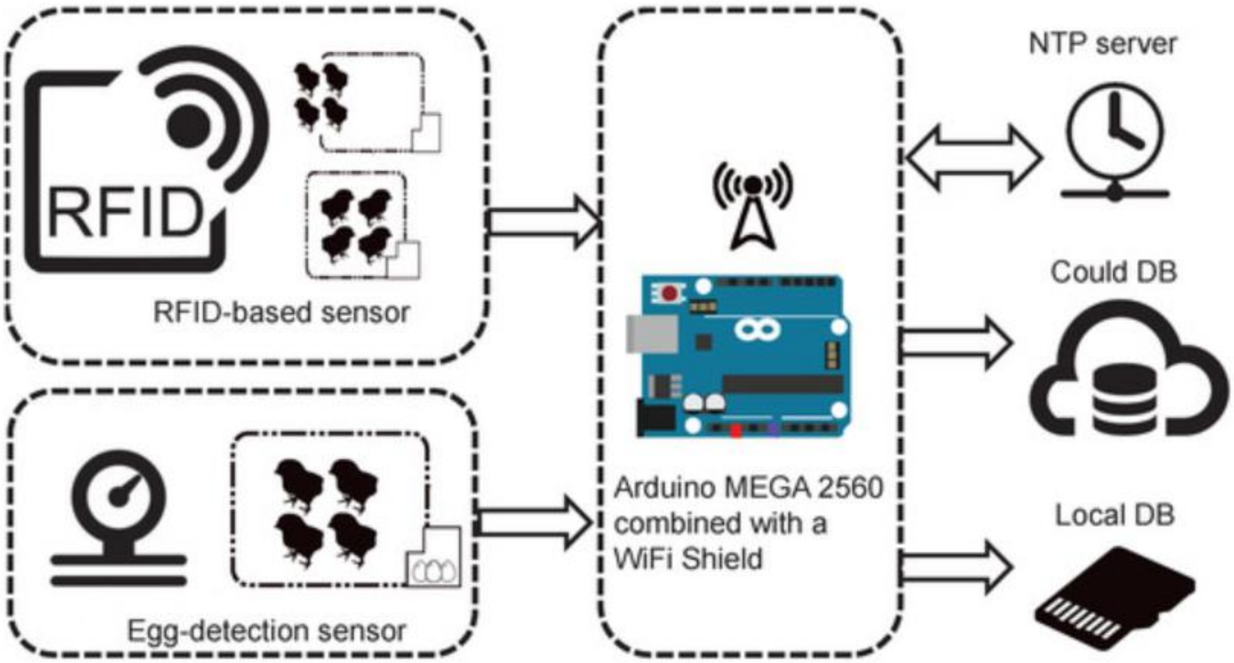


Figure 2.2: IoTs and RFID

2.4 IOT USAGE AREAS AND IOT USAGE IN SUPPLY CHAIN

IoT has a wide area in terms of usage area. The spectrum ranges from IT to the industrial sector, from medical systems to smart home systems. However, since the subject of this study is the industrial sector and therefore the supply chain, this literature will focus on this. In this context, Figure 2.3 briefly presents an industrial IoT application case followed by the actual application of IoT in the food supply chain [24].

In the industrial realm, the Internet of Things can be used for any type of industrial activity, including commercial or financial transactions between companies, institutions, and other organizations. Related to this are logistics, production, process management, services, banking and brokerage. The first relevant application of the Industrial Internet of Things is logistics and supply chain management. Many materials and products such as food, clothing, furniture and equipment can be identified and tracked using technologies such as RFID tags and blockchain. These methods can be used to effectively manage inventory and retail sales, reduce inventory errors and reduce inventory availability in local logistics, monitor/diagnose product, item and product damage for production and product management. Store processes, retail sales, inventory purchasing, quick check, animal tracking, certified animal trade tracking, product or feed tracking for irrigated farms, business records management, real-time vehicle diagnosis, collection, shipping management, payment by process, mobile tracking entries optimize. The Internet of Things uses objects or devices to enable interaction between devices and users. The internal processes of the Internet of Things allow objects and devices to connect to larger computer networks and exchange data through internal servers. The Internet of Things is a heterogeneous technological group that can interact with more than one network. The inventors and developers of IoT technology laid the foundation for IoT on the interoperability of multiple network infrastructures connected to everyday personal and industrial devices. In this context, it can be said that IoT devices can connect to the internet and include objects that users commonly use in their daily and business lives [25,26]. To support these claims, it covers a wide range of products and systems, from devices and systems such as IoT devices, smart devices and heaters to scientific research.

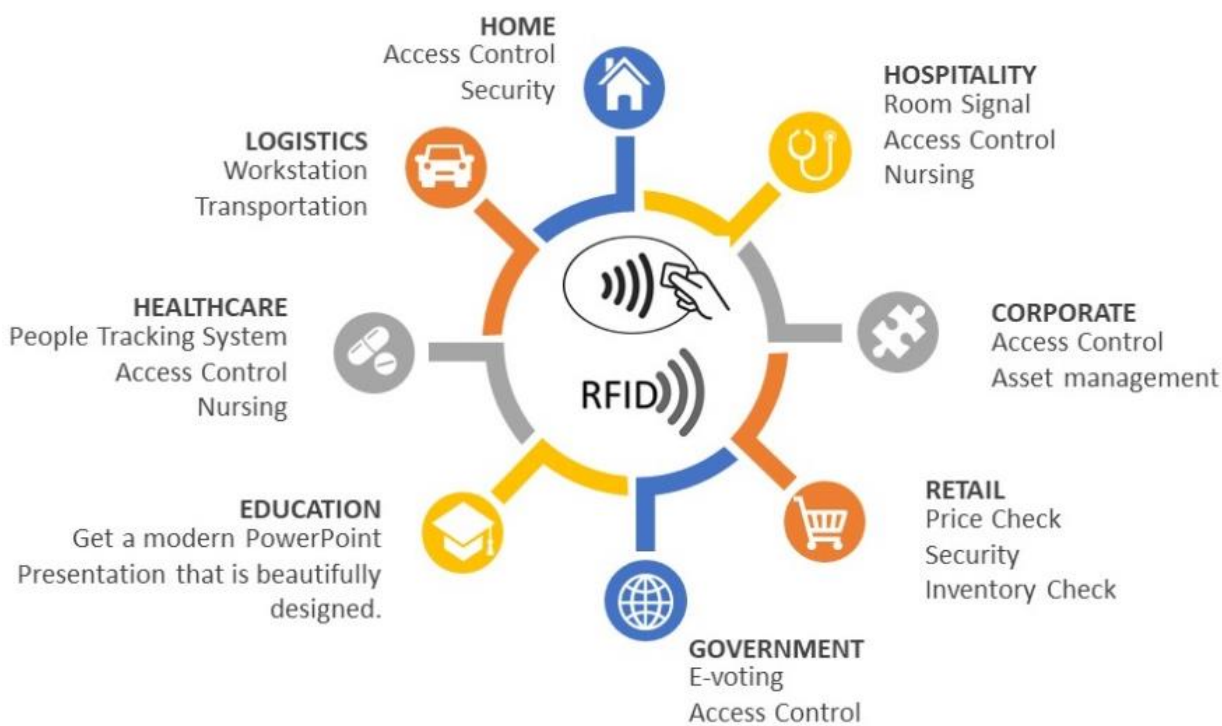


Figure 2.3: Applications of IoTs

As you can see, the food industry chain includes a wide range of processes such as agriculture and food processing, from the production chain to the final consumer and finally the disposal of waste. Each of these areas has its own characteristics, requirements, and responsibilities. Technologies like the Internet of Things can now solve a variety of problems, from tracing the history of food to tracing the supply chain, and from designing food plants to designing food safety and waste management mechanisms. Complex supply chains leverage our vast technology network to streamline processes, solve problems and enable effective collaboration between the various actors in the supply chain [28]. Through technologies such as the Internet of Things, it provides a coherent communication and control mechanism in supply chain management across a wide range of technology networks. It also enables the virtualization of consumer and waste supply chains, use/consumption by manufacturers, manufacturing, packaging sales, marketing, distribution and logistics, and distribution groups. In the literature, the "virtual object perspective", often referred to as the "virtual object perspective", is the digital representation of physical objects that provide these communication and control mechanisms and generally link related objects. The best example of this is the technology of the Internet of Things. In this context, the use of these technologies in the food supply chain and the virtualization of the supply chain are of great importance for factors such as product quality and shelf life. Although virtualization of certain elements of the food supply chain alone is not sufficient to ensure food safety and quality, information and control mechanisms of the food situation are very important for good governance in this sector [29,30].

As you can see, SCM is primarily focused on the operational integration of business processes. This means that participants coordinate the governance of business processes to increase the value they deliver to their end customers. In conclusion, we note that the main concept of the supply chain system is the business processes, controls and coordination carried out by the corporate network. As shown in the figure. 2.4, these operations, control and coordination can be supported by technologies such as the Internet of Things.

Figure 2.4: Control using IoTs

2.5 THE STRUCTURE OF THE INTERNET OF THINGS

IoTs is designed to easily connect different objects by bringing them together at a common point over networks. Objects do not have to be of the same type to communicate with each other. Different types or objects of the same type can meet and communicate on a common channel via the internet. In this context, the IoT structure can be expressed as a decentralized and heterogeneous ecosystem. There are different classifications in the literature in terms of the structure or architecture of IoT. In general, from the perspective of the service-oriented IoT structure, it is widely believed that it consists of three layers as shown in Figure 2.4: the detection layer, which is responsible for data collection and device control, the network layer, which is responsible for transmitting data, and the application layer, which is responsible for data processing. However, these classifications have also been discussed more broadly. For example, in another study, the application layer was not discussed, but the service layer to meet user needs and provide services, and the interface layer to interact with other applications and users were discussed an IoT structure with four layers: physical layer, transport layer, middleware and application layer [31,32].

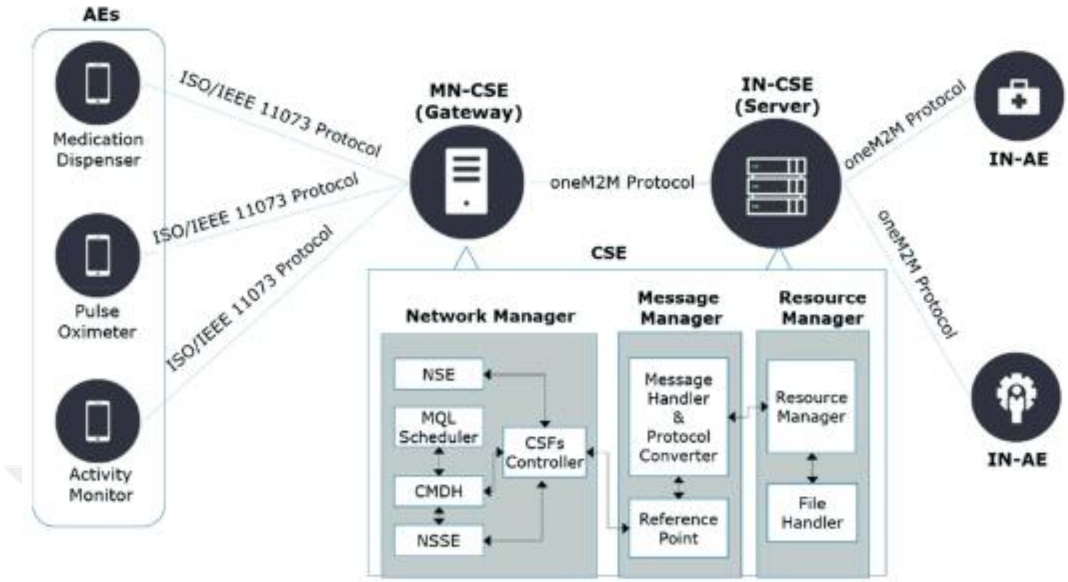


Figure 2.5: Structure of IoTs

Data access layer: data throughout the product lifecycle, including time, location, sensor detection data, RFID data, video recording, core content B: support as a service, intelligent data processing traffic, food control, industrial monitoring, and media Middleware Technology Cloud Computing Internet Services Application Layer Support Lower Layer 2.5G/3G/4G WLAN GPS IP Network Layer WSN Interim Maintenance Technology Maintenance Technology Setup RFID Middleware Sensor UWB NFC Wi-Fi Camera Discovery Colli Application Layer Support Layer Service Platform Data Communication Layer In this layer, data access and processing levels are predefined. The whole life cycle of agri-food generally includes the agricultural period, production period, logistics period, storage period and sales period [33,34].

i. Data transfer layer: This layer transmits the data received from the access layer to the application service support platform server, and uses network technologies such as WiFi and Ethernet to access the cloud from the data platform. From there, the bandwidth is specially extracted. mechanism.

ii. Application Services Support Base: This is a multi-part platform that provides various services, such as statistical analysis services for various devices and data processing devices.

iii. Application Modules Layer: This layer provides a set of application modules. Users can select one or more application modules to meet the needs of today's farming/processing business. For example, users can use the nutrition facts module to get information about specific foods.

iv. User Application Layer: User applications are made up of target user application modules. The target audience can be divided into four types: producers, buyers, sellers, and listeners. Each user application can be configured in multiple application unit.

Components of the Internet of Things The Internet of Things has benefited from many technologies in all these processes in order to both maintain its development and carry out its activities. Many internet and technology-based elements are required for objects to meet and interact in a common network. In this respect, IoT has incorporated quite different components throughout its development period. However, if we need to classify these components basically, there are three IoT components that provide seamless interaction [35,36].

i. WSN: WSN, called Wireless Sensor Networks, is a system of RF transceivers, sensors, microcontrollers, and power supplies. Through organization, configuration, diagnostics and refinement, these sensor networks have evolved to solve problems and run applications that traditional technologies cannot solve.

ii. Addressing schemes: Addressing schemes required for communication on an industrial network have a mechanism responsible for transferring the appropriate code or packets from one location to another. The ability to uniquely identify objects is critical to IoT success. Moreover, it allows not only identification but also remote management of devices over the Internet.

iii. Visualization: Provides abstract analysis and presentation of heterogeneous data using data visualization, infographics, and interactive technologies. To get the most out of the IoT, you need to use easy-to-understand visualizations. It can be said that it is mainly 2D and 3D screens that are very useful for accessing data.

iv. Data collection and analysis. With the increasing speed of digitization and access to data, the amount of data received has become enormous. The storage, classification and analysis of the data obtained have become very important. At this stage, it is very important to collect useful data and understand data analysis. This data plays a crucial role in the functioning of factors such as forecasting, storage management and decision making at the points of delivery and production [37].The interaction of the IoT with its components in the development process and the realization of the IoT as a result of this interaction are shown in Figure 2.6.

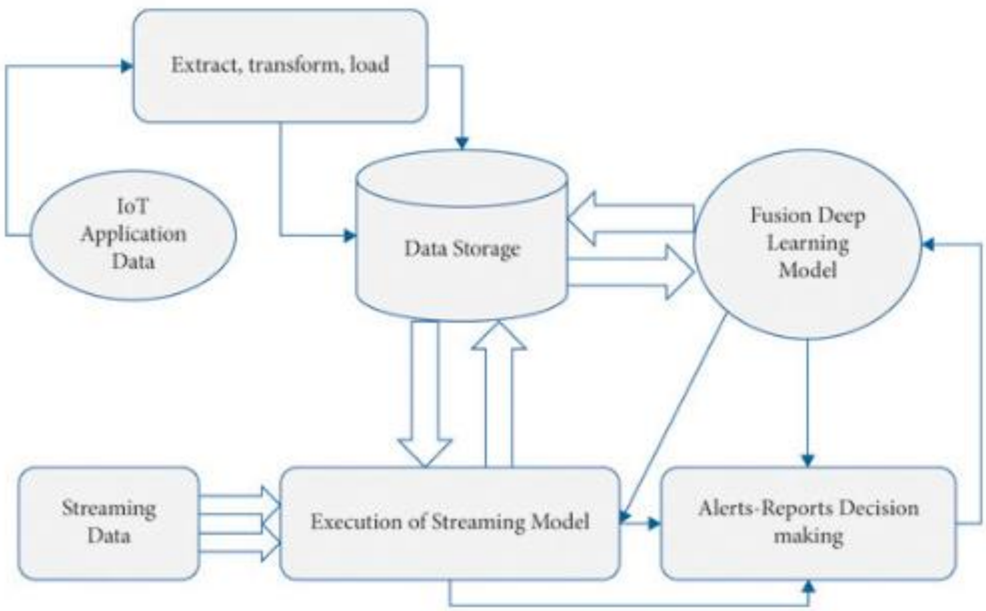


Figure 2.6: Data collection using IoTs

2.6 RISKS OF THE INTERNET OF THINGS

There are still question marks about the reliability of IoT devices. Published reports show that 70% of IoT devices are likely to be hacked and therefore have the potential to put users at risk. The main reasons for this are the vulnerabilities in the devices used in IoT production and other devices connected via IoT, especially the possibility of security attacks due to accessible vulnerabilities in the network layers of the devices. In this context, it is especially important for businesses that use devices and information technology experts of product manufacturers to reduce security vulnerabilities in devices and to identify harmful applications RFID 1980- WSN 1990- IoT 2009- -Automatically identify and track -Smart sensor network -Health monitoring -Industrial monitoring -Environmental monitoring - An important task falls on the interconnectedness of objects - Computing everywhere - Cyber-Physical system. In addition to the complexity of the security of the devices, since the use of such technologies requires multiple layers and different configurations, in addition to the IoT risks, the question of how compatible these innovative technologies can be with existing technologies due to their multiple layers and structures is also important [38,39]. Some of the 'main' challenges faced by the use of IoT in large industrial applications can be listed as follows;

i. Energy Efficiency

ii. Difficulties with communication and data (connectivity, latency, throughput, standardization)

iii. Scalability (network size, interoperability)

iv. Security and safety (reliability, privacy protection) Although there are different discussions and opinions about the risks in using IoT, especially the new technologies.

There are risks such as difficulties in adopting these technologies, the fact that other members of the supply chain have not adopted these technologies when a business uses these technologies, or cyber attacks, which are today's reality. However, when we look at the literature, it can be stated that the benefits of IoT applications are more than the harms, especially in terms of the food supply chain [40].

2.7 BENEFITS OF THE INTERNET OF THINGS

In this section, besides the general benefits of IoT technology, its benefit on food supply chains will be discussed. IoT-based food supply chain applications cover the entire chain from production to consumer. These applications include production, food processing, packaging, sales and marketing, logistics and distribution, retail, consumer interaction and even waste management. As can be seen, it can be stated in general terms that these application areas consist of five stages;

i. Producing

ii. Storing

iii. Transporting

iv. Selling

v. Consuming

vi. Controlling post-consumption (such as waste management) Various factors within businesses and in their supply chains facilitate the research, adoption, implementation and development of IoT-based applications in the food supply chain [41,42]. Some of them can be listed as follows;

vii. On issues such as traceability and good management of the process regarding the variability and perishability of products.

viii. In some food safety accidents.

ix. With the globalization, the problems that occur due to the significant increase in the distance covered by food products, especially in terms of delivery.

x. Lack of feedback mechanism in businesses or activities, unpredictable supply diversification with the emerging market and variability in consumer behavior.

xi. Requirements for sustainability or managing these needs within the scope of the waste abundance caused by globalization and excessive consumption.

xii. With technological developments and intensive use of technology, the management of data collected from business individuals, integration and smart support weaknesses can be supported".

xiii. Needs for information systems, information exchange and instant operation, which cover not only individual but also all supply chains of enterprises.

xiv. The need to provide richer information about products across the food supply chain, through enhanced sensor data processing and high-level information fusion.

xv. The need to provide qualitative and reliable information to final consumers as well as transparency in all or partial processes. As can be seen, IoT can offer some solutions to existing or potential problems not only in the food supply chain, but also in the supply chains in general, through technologies that support IoT and other technologies. In this context, it can be stated that these technologies are not only problem-based, but also have effects that can improve the functioning of businesses [43,44]. IoT applications usually demonstrate their activities by means of a code or the installation of smart devices capable of communicating with each other. In order for IoT-based devices to perform their functions, there are some requirements and features they must have. These requirements represent both the mechanism needed for devices to work and the benefit provided by obtaining this mechanism [45,46].

Figure 2.7: Supply chain using IoTs

3. MATERIAL AND METHODS

3.1 ARTIFICIAL INTELLIGENCE

Although various branches of science were influential in the formation and development process of artificial intelligence, the first spark was thrown in the 1950s to start researches, concrete examples and serious research on the subject. Many artificial intelligence and artificial intelligence technologies have been developed to date [47,48]. It is still being developed today. In this period 1950-1960 period, which is called by various names such as the spring period of artificial intelligence, the first buds, the first building blocks of artificial intelligence for application began to be laid. In his article published in 1950, Claude E. Shannon wrote that computers can play chess. Shannon in his article; He stated that chess is ideal for machines that can think, because satisfactory results can be obtained by trial and error in chess, chess usually requires an action of "thinking", the rules of playing chess pieces and the purpose of the game are precise, and the structure of chess is compatible with the digital nature of modern computers. Same year , Alan Turing published his article on whether machines can think or not in Mind magazine. In this article, Turing put forward the "Imitation Game" (Touring test, as it is called) in order to prove whether machines can think or not. In this test of Turing; a woman, a man and three interrogators (regardless of their gender). The interrogator must correctly guess the gender of the people in the other room by asking questions in a room separate from the other two people - the answers must be given in writing via a computer in order for the test to be carried out in the most ideal way. While one of the respondents is telling the truth, the other person is trying to mislead the questioner [49]. For example, when the interrogator asks the length of the hair of the parties; A male person will say that their hair is short and the longest strand is 9 inches (22.86 cm). Turing stated that in this test, if a machine is substituted for the male person, and if the machine misleads the interrogator in this game, which is played many times, it should be concluded that machines can think. He stated that the adult human brain can be obtained.62 During this period, three important meetings were held in Los Angeles for the development of learning machines and artificial intelligence: the Learning Machines Session in 1955, the Artificial Intelligence Summer Research Project at Dartmouth College in 1956, and the Mechanization of Thought Processes in 1958. . In the Learning Machines Session held in 1955, some pattern recognition experiments with networks consisting of nerve cell-like elements, information techniques that can be used in image processing, and programming of computers that can play chess were discussed. At this meeting, the foundations were laid for many studies that will try to give machines the ability to see in the future [50]. Under the leadership of John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Cluade E. Shannon in the summer of 1956 at Dartmouth College for two months with a team of ten people. A workshop was organized within the scope of the Research Project. In this workshop, various topics such as automatic computers, how a computer should be programmed to use language, neuronal networks, the machine's self-improvement and becoming a more intelligent machine, and the creativity of the machine were discussed. In the Mechanization of Thought Processes symposium held in 1958, it was aimed to examine the topics of artificial thinking, character and pattern recognition, learning, automatic language translation, biology, automatic programming, industrial planning, mechanization of editorial work, and also focused on topics such as intuitive programming methods, common sense programs66. John McCarthy wrote one of the first artificial intelligence programming languages called LISP (List Processing) in 1957. The difference of McCarthy's program from other programming languages is that it can work on situations such as sentences and rules.68 In this period, programs that provide symbolic integration and programs used to distinguish similar geometric shapes applied in intelligence tests contributed to the development of expert systems [51]. 1960-1970 period Until the first half of the 1960s, many researches and projects were carried out on artificial intelligence. The period between 1965 and 1970 is known as the dark period or the winter of artificial intelligence. Because, in the process up to 1965, scientists had an overly optimistic approach to the development of artificial intelligence and believed that the construction of intelligent computers was simple70. When scientists' expectations and predictions about the future (for example, that a computer could become a chess champion until 196871) did not turn out as they wanted, studies on artificial intelligence slowed down in this period. The idea of building intelligent machines72 by simply uploading data has led to an "explosion of combinations". In other words, even in order to answer very simple questions, the desired success could not be achieved by applying the "try all possibilities and choose the best result among them" mechanism. Although there was a slowdown in this period, research on artificial intelligence continued. A mobile vending machine construction project was initiated by the Advanced Research Projects Agency (ARPA) of the United States of America (USA) in 1964 within the scope of the vending working statement, with little or no outside intervention. In 1966, project leader Peter Hart and his team (Nils Nilsson, Mike Wiber, Claude Fennema, Helen Chan

Wolf and Tom Garvey) started work. Within the scope of the project, which lasted between 1966 and 1972, the first mobile robot, named Shakey, with the ability to sense and judge its environment was developed. There is a structure called "three-layer hierarchy" within Shakey. In this context, the lowest level programs run all the motors and collect sensory information. Intermediate programs control simple actions such as going to a predetermined location and process images from the television camera. The work of planning more complex actions is done by top-level programs. In 1966, the artificial intelligence program ELIZA was created by Joseph Weizenbaum. In this program, it is aimed that humans and machines can communicate and chat with each other through natural language processing. To explain the working procedure of the program simply, keywords are searched among the entries entered in order to chat with the program. When such a keyword is found, the sentence is converted into a rule associated with the keyword. In this context, keywords have a rating or priority number, and the natural language processing process works sensitively against such numbers [53]. There are several examples of chats with ELIZA. In some of these conversations, it was seen that ELIZA fell into repetition, and in some of them, she deviated from the topic of conversation. ELIZA; It has basic problems such as not having a mechanism to respond intelligently if the input text does not contain a keyword, and not defining the most important keyword in the input message. 1970-1990 period In the period between 1950-1965, there was an explosion of projects, articles and research on artificial intelligence; However, overly optimistic predictions and expectations of scientists could not be realized in the following years (1970-1980). Even though millions of dollars were invested in this period, the expected developments from artificial intelligence technology could not be realized and as a result, interest in research on artificial intelligence decreased; It also caused the lack of sufficient financial resources for artificial intelligence. States have also reduced their funding for artificial intelligence. Therefore, this period is called the winter of artificial intelligence. Advocates of the old-fashioned or classical artificial intelligence model say that imitating nerve cells in the human brain is a waste of time, that an intelligent computer can perform every cognitive operation, and that icons representing high-level concepts are sufficient for artificial intelligence. Seymour Papert and Marvin Minsky, one of the leading names in artificial intelligence, wrote a book called "Perceptrons" on this subject, saying that artificial neural networks cannot produce solutions to some problems, have no scientific value, and that they cannot learn even some simple functions when these structures are arranged in a certain way [54,55]. In particular, Papert and Minsky's authorities in the field of artificial intelligence, their leading names prevented the support on artificial neural networks, and the development of artificial neural networks, which is one of the applications of artificial intelligence, is a problem. Although investments and research on artificial intelligence have decreased, it has been understood that other fields of science such as psychology, linguistics, and biology can also be used on artificial intelligence in this period. Therefore, this period is also called the partnership period. In 1975, the mass spectrum results found by the program called MetaDendral, which has the ability to learn, constituted the first example of publishing the results found by a computer in scientific journals.87 Invented by James L. Adams in 1960 and followed by The Stanford Cart, developed by a number of graduate students over a period of years, culminated in 1979 with Hans Moravec. Moravec enabled his Stanford car to navigate autonomously in a room full of chairs.Due to the negative reasons experienced in the 1970s and 1980s, studies were focused on expert systems with knowledge of a particular job or subject, instead of general artificial intelligences that could do anything [56,57]. In the mid-1970s, MYCIN, one of the leading experts in expert systems, was developed at Stanford University. MYCIN is a program that includes the diagnosis and treatment information of infectious disease specialists and uses the “if-then” rules that physicians follow when diagnosing the disease, and works by using a “precision multiplier” to qualify these rules. Artificial intelligence with a history of ups and downs. One of the studies that caused it to gain momentum again is the project initiated by the Ministry of International Trade and Industry of Japan in 1981 to create the fifth generation computer systems. Within the scope of this project, it is aimed to enable artificial intelligence to make inferences from large databases and to produce computers that can communicate with people, interpret pictures and translate with natural language processing. Japan's announcement about the fifth-generation computer systems had a worldwide impact. He pioneered the initiation of similar projects in the USA and Europe [58,59]. In this context, a committee was established in England in 1982 to provide consultancy in the field of information technology; In 1983, the European Strategic Program of Research in Information Technology (ESPRIT) was initiated by the European Economic Community. Between 1987 and 1993, artificial intelligence studies started to slow down again due to reasons such as the understanding that the Japanese projects on fifth-generation computer systems could not be completed within ten years, deficiencies in computer systems, and insufficient access to data, and artificial intelligence's winter came into question again. Especially with the opening of the internet to the public in 1998, developments in the field of artificial intelligence, as in every field, gained momentum and many successes were achieved in this field. Examples of artificial intelligence that can play chess and beat professional chess players have been encountered in the historical process; However, the desired results could not be obtained from these trials. For many years, efforts have been made by the International Business Machines Corporation (IBM) to produce computers that can play chess. IBM's computer named “Deep Blue” defeated world chess champion Gary Kasparov. There are two versions of Deep Blue, the first of which was a chess match with Kasparov in 1996; but he lost this game. The next version of Deep Blue faced Kasparov again in 1997 and won two of the six-set games, lost one and drew all three to win against Kasparov.93 Deep Blue's defeat of Kasparov; It is stated that this victory of Deep Blue is based on various factors such as being a highly parallel system with multiple levels of parallelism, powerful search extensions, complex evaluation function, and efficient use of the Grandmaster game database. had an impact on processing. Especially in solving complex problems, the limits of parallel processing have been understood to a great extent. This research gave the idea that a computer could be designed to solve complex problems in other fields, using deep knowledge to analyze a large number of possible solutions. Architecture used in Deep Blue; It has been used in various fields such as market trends, risk analysis, data mining, creation of new drugs.95 The National Aeronautics and Space Administration (NASA) launched the spaceship “Deep Space 1” on October 24, 1998. The mission of this spaceship is to evaluate whether twelve new advanced technologies, including the artificial intelligence robot system called "autonomous remote agent", are working successfully in space. The purpose of this robot is to plan and execute the activities of the spacecraft.96 The most important feature of this robot, and its difference from other conventional spaceship commands, is that the operators on the ground communicate over targets instead of detailed commands with specified time. This robot determines an action plan to achieve the goals given to it and puts the plan into effect by giving commands to the spacecraft. The USA established ARPA after the Russians sent the spacecraft called Sputnik into space during the cold war, and later renamed DARPA (Defense Advanced Research Projects Agency). changed to. DARPA organized a competition called “The Grand Challenge” on March 13, 2004 in order to develop autonomous vehicle technology. According to the rules of the competition, the given race route must be completed autonomously, without any manual intervention, in less than 10 hours. Although 15 competitors participated in the 1 million dollar prize race, no vehicle could complete the difficult desert route. The following year, on October 8, 2005, the same race was repeated and this time a prize of 2 million dollars was determined. The given race route must be completed autonomously, without interference. Although 15 competitors participated in the 1 million dollar prize race, no vehicle could complete the difficult desert route. The following year, on October 8, 2005, the same race was repeated and this time a prize of 2 million dollars was determined. A total of 5 vehicles were able to complete this race and the Stanford Racing Team won with their vehicle named Stanley for 6 hours and 53 minutes. These races have helped create a mindset and research community that will almost certainly make fleets of autonomous vehicles and other land vehicles a decade later for the first quarter of the 21st century. Autonomous vehicle technology is being developed day by day. Today, companies such as Tesla have started to produce autonomous vehicles. Studies and legal regulations regarding the inclusion of autonomous vehicles in traffic have begun in various countries. In 2007, IBM started to work on creating a computer good enough to compete with the champions in the Jeopardy game. In 2011, he developed a computer with artificial intelligence named "Watson", which can answer questions asked in natural language, which includes technologies such as natural language processing, information, information representation, reasoning and machine learning. To test this computer, jeopardy was competed with Brad Rutter and Ken Jennings, one of the game's most successful players of all time. Ken Jennings won $300,000, Brad Rutter $200,000, and Watson $1,000,000. The contest in question is the first example of man's encounter with a machine in the history of jeopardy, and the machine won the contest.99 Although Watson won the contest, Watson could not understand what words meant one by one. Rather, he understands the peculiarities of the language used by humans. In this context, it is able to understand the content of the text under changing conditions.100 Since 1986, Honda has started working to produce an assistant robot that supports people's living spaces and robots that can replace humans in dangerous or inaccessible areas, beneficial to humans and society [60]. It is stated that it has shown the world how effective deep learning and advanced search techniques are in artificial intelligence technology. In 2017, a higher version of AlphaGo was developed and named AlphaZero. The difference of AlphaZero from AlphaGo is that only the knowledge of the game's moves and winning rules was taught, no other human knowledge was needed. AlphaZero developed itself with a 40-day training and, thanks to this training, it beat AlphaGo.106 Hanson Robotics company developed a humanoid robot named Sophia in 2016. Although Sophia is known for her human-like nature, it was modeled after the famous Hollywood actress Audrey Hepburn's face. Having gestures and facial expressions, SOPHIA can communicate effectively, easily answer questions asked and chat with people. Sophia is based on various artificial intelligence technologies such as artificial neural networks, natural language processing, machine perception that enables her to understand human faces, gestures, and emotions. However, thanks to the underlying artificial intelligence components, Sophia's reactions can be specific to any situation or interaction [61]. In 2017, Sophia became the first robot to acquire the citizenship of any country by obtaining the citizenship of Saudi Arabia. Hanson Robotics and Patrick Tresset collaborated with Sophia to teach portrait drawing. This ability of Sophia was shown to the world for the first time on July 17, 2019.108 Today, many artificial intelligence technologies such as robots that clean and sweep floors, hearing aids that adapt to ambient sound levels and prevent hum, traffic sign recognition systems, and intelligent speed adaptation have been developed. Apple's Siri, Amazon's Alexa, and virtual assistants called Amelia109, which can understand the questions people ask and how they feel, have also been developed. Artificial Intelligence Applications Although the working mechanism of the brains of living things, especially humans, is complex, how it works has not been fully resolved yet. Various applications have been developed over time in order to transfer such a complex structure to computers or machines. Robotics Czech playwright Čapek's 1920 "R.U.R. - The concept of robot, which took its place in the literature with the theater play "Rossum's Universal Robots", is one of the applications of artificial intelligence. In the doctrine, the definition of a robot shows both physical and mental activity; but they are made as created systems that are not biologically alive. According to this definition of Richards, robots are understood as machines that can physically move and make rational decisions about the tasks they need to do. Although this definition seems to be able to be adopted by legal systems, it remains in a very narrow scope. Because according to this definition, machines that assemble the parts of vehicles or carry heavy parts or facilitate the work of the employees in the factory are not considered as robots according to this definition. on the other hand, they are mechanical entities that can act semi-autonomously or fully autonomously with direct human control. Therefore, it cannot be said that every robot has artificial intelligence [62]. In order for a robot to have artificial intelligence or to be qualified as an intelligent robot, it must have perceptual or sensory characteristics. With this feature, it can respond appropriately to changes in its environment. The feature that distinguishes artificial intelligence robots from other robots is that the reaction to be given by the robot is not automatic, but based on previous experience.

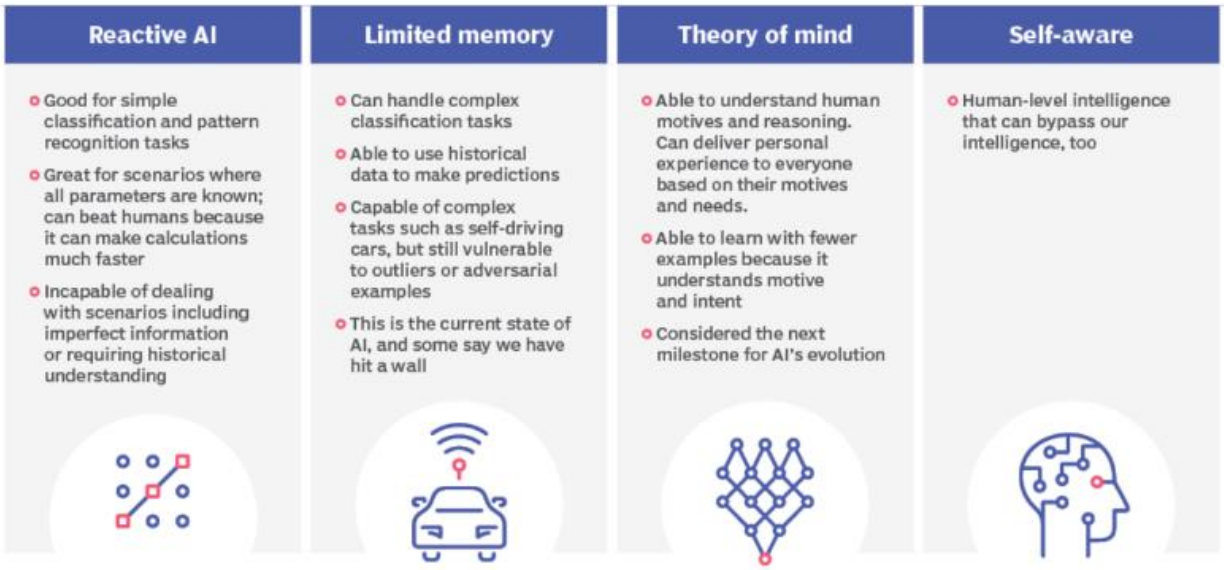


Figure 3.1: AI Types

3.1.1 Expert Systems

Another application that has been created for the purpose of transferring human intelligence to a computer or machine is expert systems. Systems that produce solutions based on information compiled from real people in solving a problem that requires expertise are called expert systems. Expert systems are computer software used to solve problems related to data. With this software; Many data such as solutions and experiences of various problems that require expertise are stored in the data warehouse, and it is aimed to make inferences using the stored data when the same or similar problem is encountered again in the future. In general, expert systems; It consists of a knowledge base and a logical inference and control mechanism that provides creative and correct answers and solutions that enable it to make separate inferences from this knowledge base. This working mechanism of expert systems is different from other databases designed to obtain efficiency and functionality from information systems. Because database systems provide ease of access to the data desired to be obtained from the data stacks stored in their warehouses. Expert systems, on the other hand, aim to obtain original and qualified knowledge applied by experts in the relevant field. It also gives information about how and why it arrived. The inference mechanism of expert systems performs its evaluation using various methods such as Modus Ponens, solving [63]. The system that provides the transition between data processing and information processing steps is called an expert system. In data processing, databases are used depending on an algorithm with a traditional programming method. Information processing, on the other hand, is the reasoning of information obtained from rules and facts based on experiences, symbolic rules and relationships, graphs and models without relying on algorithms. In this context, expert systems; in order to exhibit a hierarchical approach, to obtain intermediate results before reaching the end of the problem solving, to explain how they reached a certain result, to reach more than one result when necessary, to work with certain or incomplete information, to use the data obtained from the previous problems in solving the problems they will encounter in the future [64]. In addition to all this, expert systems can also obtain new data as a result of their search and reasoning.

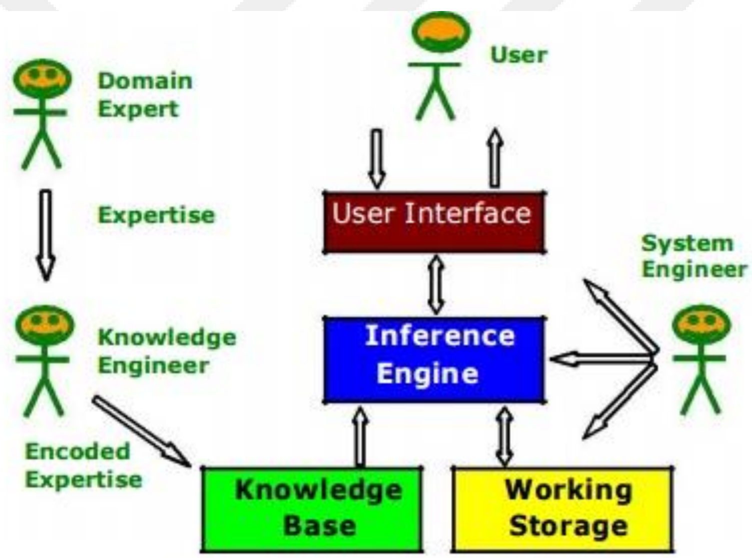


Figure 3.2: Expert system structure

3.1.2 Natural Language Processing

Natural language processing, which is basically intended to communicate with humans and machines, is carried out by processing both sound and text. Voice communication is an inevitable part of life[65,66]. While a person can overhear the conversation of the people next to him and join them, a computer today does not yet have that competence. In the same way, people do not have any trouble in distinguishing the person they are talking to from other voices in the environment they are in, while they are communicating with voice, they pause when speaking and add emotion to their expression. This situation is not easy for computers. In this context, computers primarily convert the sounds they hear into a written document. Then it goes into the process of processing these written documents. In other words, the processing of sound by artificial intelligence consists of two stages. The first stage is voice recognition, which is the correct perception of the voice and its conversion into text; The second stage is to understand the speech, which is the analysis of the created text and capturing the meaning in the text. Natural language processing software analyzes the sentences that are directed from the outside in voice or written form, according to grammatical rules; Then, it divides the sentences into parts, word by word, and puts them into commands in a way that the computer can understand. At the stage of capturing meaning, it usually starts from the predicate and compares it with the keywords in the application dictionary [67,68]. In the process of recognizing the voice and capturing the meaning, the background sound is separated with the speaker, where the words or sentences start and end during the speech, the meaning is missed due to homonyms or the same thing is expressed in different ways. (for example, "Set an alarm for 8 a.m.!" to set an alarm for the smart assistant Siri, Can you set an alarm for 8 a.m.?", "Can you wake me up at 8 a.m.?" or "Let me ask for an alarm. Expressions such as ." can be used.). In order to overcome the difficulties in natural language processing, the morphological analysis method for the automatic analysis of the structure of a word by computers; the names (person, place, institution, etc.) and numeric entities (time, time, etc.) in the sentences. currency, date, etc.), asset name recognition methods, etc. methods are used. Developing tools and equipment to assist writing in the process of translating sounds into writing and directly in the processing of written documents, developing the processes of reading a printed text, summarizing the text, understanding the text, reading and writing in foreign languages, and developing various auxiliary tools and materials in this context, language translations, correcting spelling mistakes. many applications are within the scope of natural language processing, such as editing, verbal or written communication, and answering questions [69]. Classifying text among disparate datasets (eg positive or negative customer reviews); text parsing, analysis of emotion or idea in a text; detecting entities, events, dates, times and concepts in a text; recognition of entity names such as persons, places, institutions in a text; temporal relation inference in terms of determining the sequence of events; Subjects such as the detection of word types such as nouns, adjectives, prepositions and predicates in the text are also within the scope of natural language processing, and methods such as convolutional neural networks, deep neural networks, deep belief networks, which are deep learning methods related to these subjects, are used.

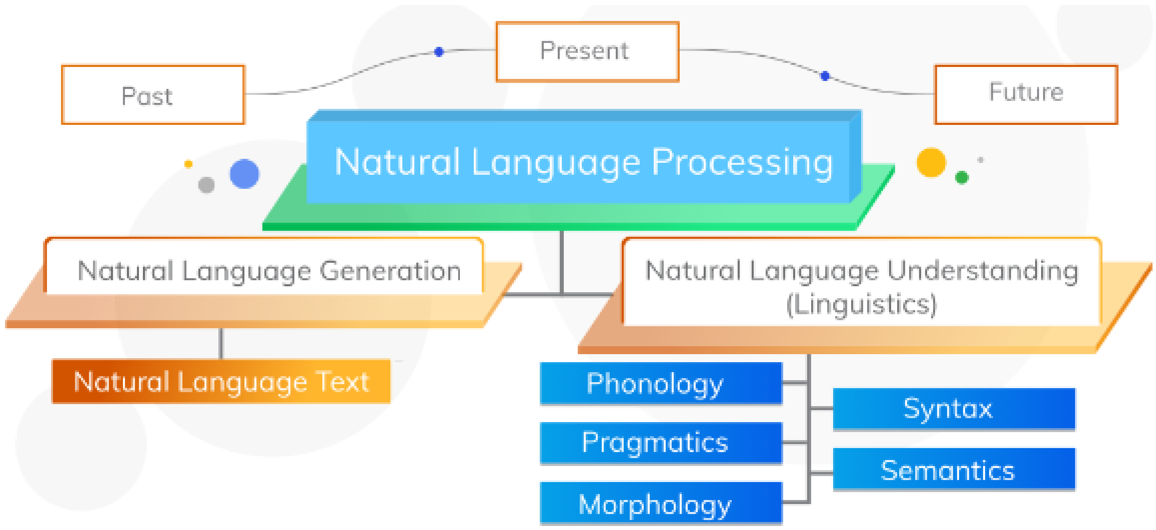


Figure 3.3: NLP

3.1.3 Pattern Recognition

The process of analyzing an image, sound, electronic signal or any data set and placing it into certain categories is called pattern processing. This function is to perform classification together with the symbolic information obtained from the results of the examination with the feature extraction mechanism by collecting information with its sensors and with the help of identification schemes. Exists [70]. The pattern recognition process generally consists of two phases. First of all, the feature vector is created. In this context, a data structure is created in order to perform the recognition process and each pattern is defined with a vector. In the next step, considering various features of a pattern, it is placed in the class that has the closest features to this pattern. In order for an image to be recognized, first of all, the image must be converted into a form that the machine can understand. Accordingly, the image is divided into pixels and then the numerical values corresponding to these pixels are placed. In voice recognition, the pattern recognition process is performed after the audio signal is received digitally in the computer environment and the received raw sound is passed through various filtering and windowing methods. In this way, many patterns are displayed to the computer. and the computer is trained with this data by machine learning methods. After a successful training, he can identify the patterns that come before him [71,72].

The phone application called "Seeing AI" developed by Microsoft is an example of pattern recognition technology. With this application, it is aimed to facilitate the lives of visually impaired people by carrying out activities such as identifying coins, distinguishing and defining objects, telling product contents, describing the environment and describing photos with sound[73,74]. There are various problems such as distortion of the pattern, noise, the pattern not appearing for any reason or the formation of redundancy, and due to these problems, uncertainties and errors occur in pattern recognition.

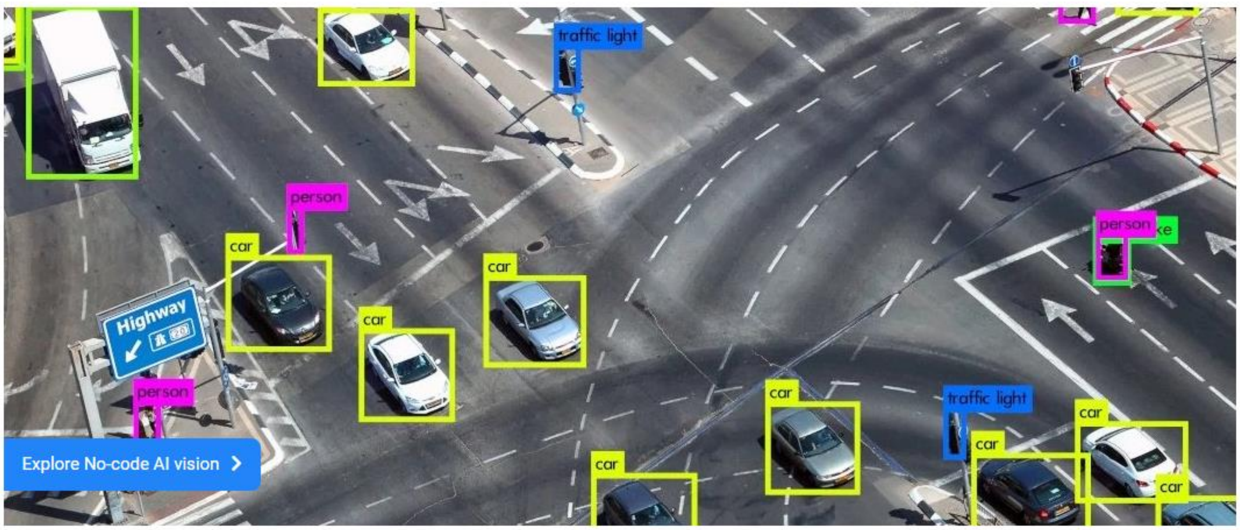


Figure 3.4: Pattern Recognition

3.2 MACHINE LEARNING

In machine learning, which is a sub-branch of artificial intelligence, it is aimed to use the knowledge and experience obtained from previously encountered problems in computers or machines when faced with the same or similar problem or a problem that has never been encountered before in the future. In this context, instead of using strict and unchangeable classical coding methods specific to each problem -the coding needs to change as the problems to be solved change-, machine learning is applied to algorithms that enable it to be applied to similar or different problems without the need to change any code -by editing a general formula and improving the formula through training process [75,76]. In this context, unlike the coding of a regular computer program, in the coding process of ordinary computer programs, the inputs previously given to the system are processed sequentially and the desired result is achieved. When the system encounters a situation that has not been coded before, it may make an error or abort the process. - When artificial intelligence encounters a situation that has not been encountered before, the system can solve the problems that come before it without making any mistakes, based on its own knowledge and experience. To summarize the steps in machine learning, the first step is to solve the problem. definition comes. In this context, it is necessary to answer what the problem is to the machine or what the machine needs to learn. Afterwards, the data to be applied to the problem needs to be analyzed and transformed into information, in other words, revealing the hidden relationships between the raw data. Accordingly, the qualities that are thought to have an effect on the results in the data sets should be investigated and as a result of these researches, the structure of the data sets that will be presented to the machine as experience should be created. Afterwards, a model needs to be established in order to determine suitable algorithms for machine learning, and this model is tested with various performance evaluation methods (such as holdout, repeated holdout, cross validation), and the model with the best performance is started to be applied. produced. The first rule was established with Hebb's book published in 1949. According to this rule known as Hebb's Rule, it was revealed that the nerves were stimulating each other jointly and as a result of these stimulations, the connection strengths between them - in other words, their weights - would increase in proportion to the product of their own activities. ; Hopfield learning rule that if the expected result and the calculated result are the same, the connections between cells should be strengthened and if the results are different, these connections should be weakened; Various learning rules have been produced, such as the Kohonen learning rule, which states that the cell that produces the best result should be rewarded and that the connections of its neighbors should be strengthened. Learning rules introduced after Hebb and Hebb have been shaped over time. Accordingly, machine learning methods are basically divided into three groups as supervised learning, unsupervised learning and reinforcement learning [78].

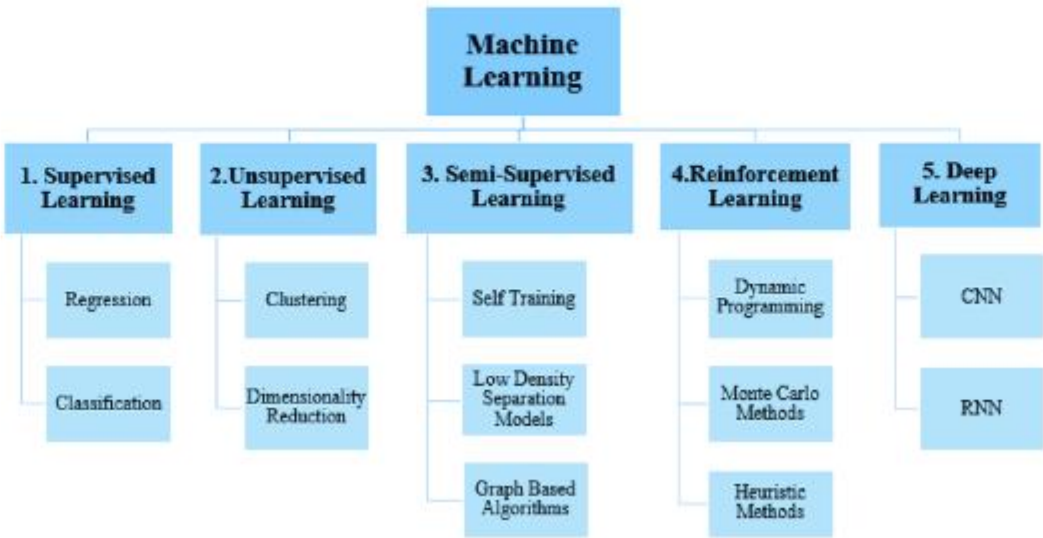


Figure 3.5: Machine Learning Types

3.2.1 Supervised learning

In supervised learning (it is also called supervised learning, supervised learning or teacher learning). A group of input data and their output values are given together. In this learning method, it is expected from the machine to comprehend the relationship between the input data and their output values and to obtain the output values closest to the target values (the result values that are expected to correspond to the input values are called the target values). In supervised learning, it is important to choose the training data that will be suitable for the problem and to have a homogeneous problem space. Because otherwise, the machine will not be able to produce suitable solutions to problems it has never encountered before. . The machine is expected to achieve a high degree of success during this test process. It is stated that the person who created such a machine has an important role in achieving this success. Supervised learning; classified into two groups: classification and regression [79]. In the classification methods, in the training process of artificial intelligence, the outputs that the inputs should come out are divided into various categories/classes. Algorithms used to place the inputs in the correct output classes are called classifiers. Classification methods are concerned with which class the inputs of unknown class should be included in. For example, a machine is trained with pictures of flowers such as roses, tulips, daisies, and flower types are taught and classifications are made according to this training. The machine, which is trained with classification methods, then determines which classes the flowers in front of it belong to, places them in the relevant class, and reaches the result. An example can be given to a machine that is trained so that customers are placed in the relevant class as a result of the evaluation of various data such as their age, income status, and gender, and that it can make a risk assessment of new customers who want to get credit in the future and place them in the relevant class. With regression methods, one or more inputs, in other words, independent variables ; It is aimed to understand the causality relationship and effect between the output and the dependent variable, in other words. In regression methods, a value is estimated instead of determining a class by using explanatory variables. A machine that can predict the prices of the houses in front of it after training a machine with various data such as square meters, number of rooms, building age, location, price, etc., and a machine developed for the estimation of the amount of precipitation in a region can be given as examples of machines produced by regression methods. Many algorithms such as K-nearest neighbor algorithm, Bayes Naive classification, support vector machines, logistic regression analysis, decision trees (such as ID 3, J48, CART, CHAID) are used in classification. In regression, many algorithms such as simple linear regression, multivariate linear regression, polynomial linear regression, nonlinear regression, random forest regression, support vector machine, decision tree are used. The more data is used in the training of artificial intelligence algorithms, the higher the performance; however, the amount of data used is not the only factor affecting performance. For example, when creating decision trees, it is necessary to prevent the machine from over-learning. Accordingly, no matter how big the machine's database is, if it creates an impure subset as a result of the tests, any effort to divide this subset will cause the machine to over-learn and prevent the machine from obtaining the desired performance. role is of great importance.

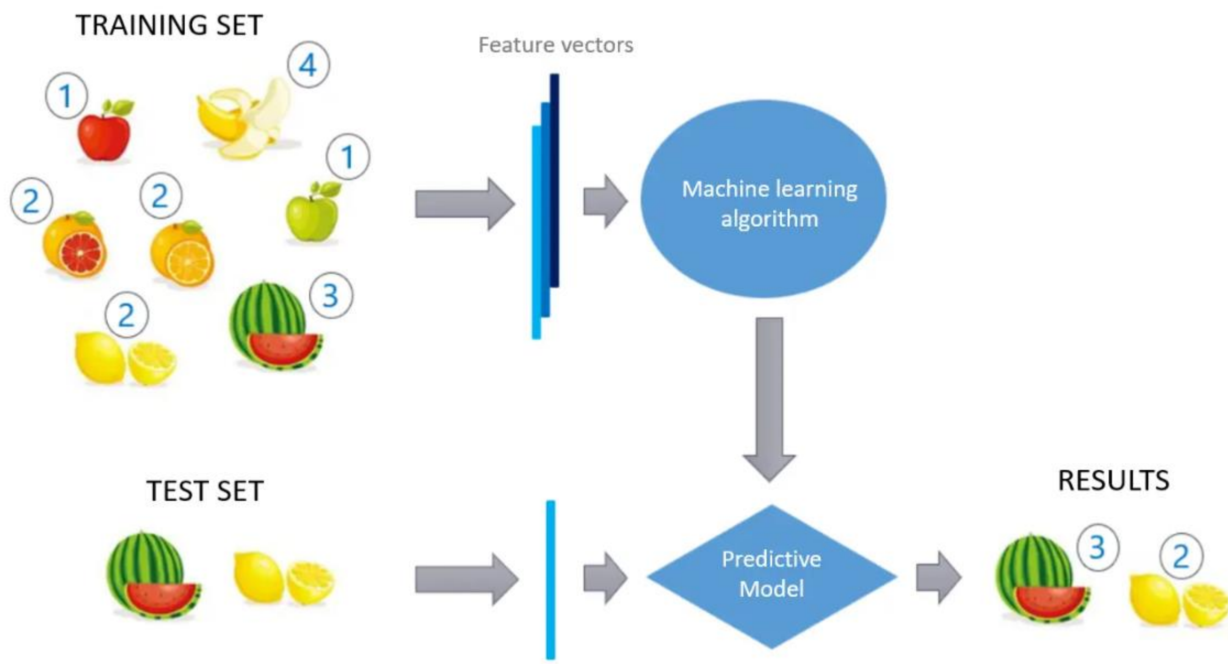


Figure 3.6: Supervised Learning

3.2.2 Unsupervised learning

In machine learning, it is aimed to reveal the characteristics of the data in the data sets; However, unsupervised learning methods are used in cases where there is no or incomplete information about the characteristics of the data sets. Because unsupervised learning aims to have more information about these data, to determine the distribution among the data and the basic structure in the data, based on similar examples among the data. In unsupervised learning, it is not known which classes are in the data sets or how many types of classes [80]. In this context, unsupervised learning algorithms explore and process issues such as data structure and the relations of data with each other. In other words, in unsupervised learning, inputs are added to the machine; however, labeling is not done about which output value these inputs should have or which output class they will be included in. This process is expected to be done by the machine. This method is called unsupervised learning (unsupervised learning, unsupervised learning, unsupervised learning) because there is no outside supervision about which output the inputs should be directed to. Unsupervised learning is divided into two as clustering and size reduction. After the data sets are entered into the machine, the machine evaluates the relationship between these data, which data is close to each other, and collects the data that are closely related to each other and creates special clusters. This process is called clustering. Algorithms used for clustering are called clustering algorithms. Today, which is called the information age, there are many large data sets. These data sets make it more difficult to perform various operations that require high computational power, such as vectors and matrices, and cause the performance of the system to decrease. causes trouble.) Useful information is preserved while the data sets are converted to a lower-dimensional space (one-dimensional, two- or three-dimensional space) with the size reduction method. the use of the machine created by the learning method; the machine created with the aim of grouping certain features of the customers shopping from a store and increasing the spending of the people in this group by offering special ads or campaigns to these groups can be shown as an example of the machines created with the unsupervised learning methods.

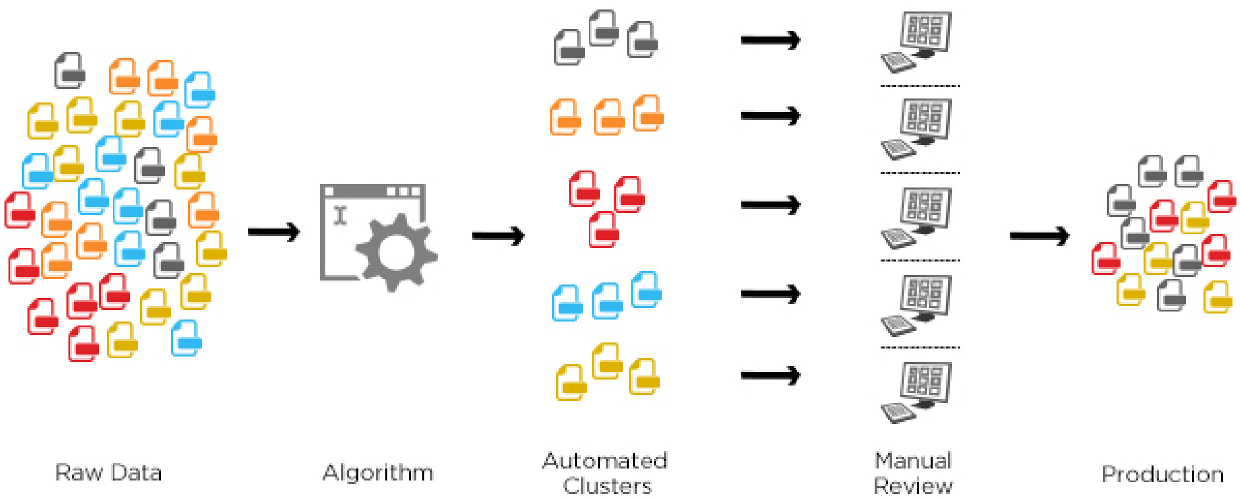


Figure 3.7: Unsupervised Learning

3.2.3 Reinforcement learning

In reinforcement learning (reinforced learning, learning with criticism), only inputs are entered into the system, as in unsupervised learning, and no value is shown about the outputs of these inputs. In this context, the machine learns by trial and error method in a dynamic environment. In the reinforcement learning method, a signal is sent by the consultant who supports the learning in order to provide feedback on whether the output values reached by the machine are correct or incorrect; but it is not stated why the result reached in this signal is wrong [81]. In this context, the signal sent by the consultant indicates how accurate the output value reached by the machine is and continues the learning process that will enable the machine to reach the correct result. . In order to achieve this goal, the agent inside the computer or machine moves in the environment (the environment in which the agent moves) by trial and error method. The agent, who is likened to a player trying to overcome the obstacles around him in order to reach the highest score in a computer game, calculates the output values for every right and wrong move he makes and stores these values in the data table [82]. As the agent learns from the actions it has made, it updates the values it has stored, and this process is constantly repeated to calculate future probabilities. Reinforcement learning method; Self-driving vehicle technology is used to build machines that can play games like chess or Go.

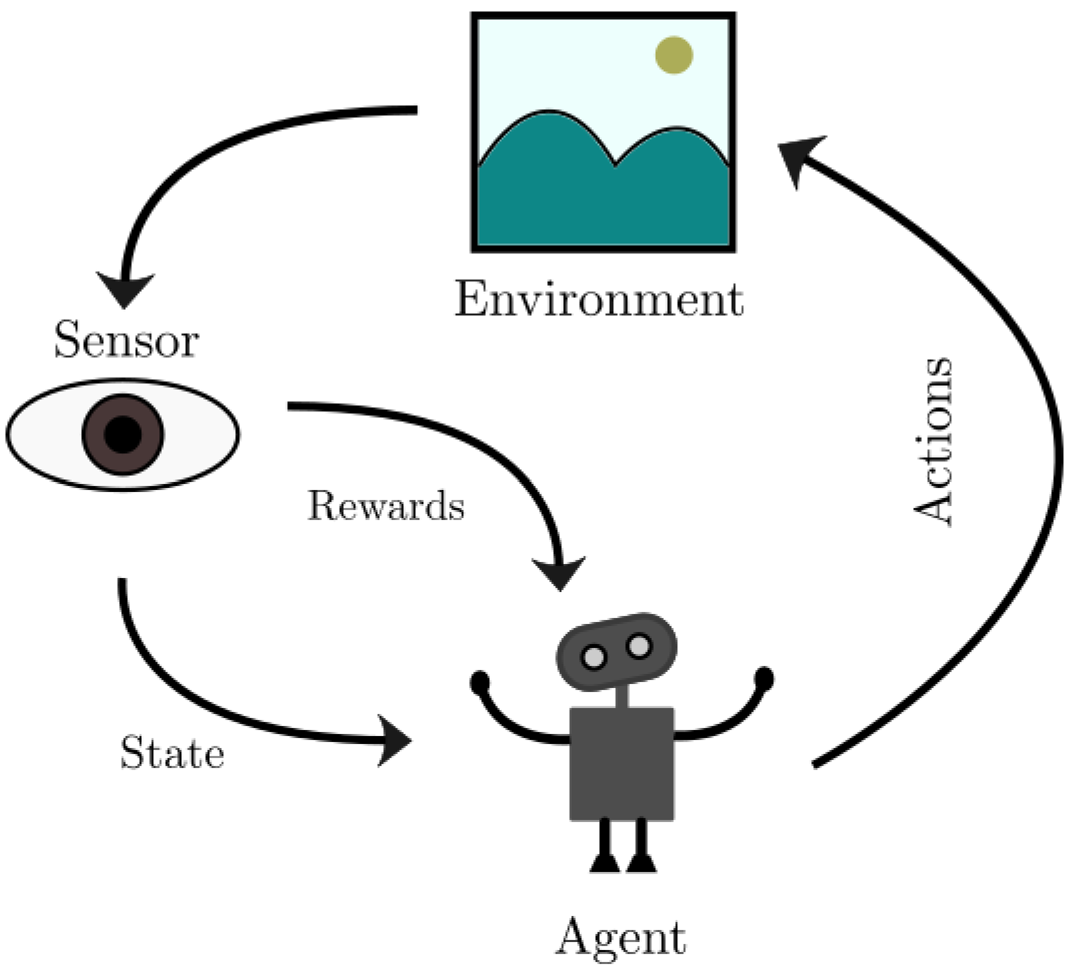


Figure 3.8: Reinforcement Learning

3.2.4 Deep learning

Although deep learning has made its name recently, it is actually an old term. One of the biggest problems experienced since the beginning of artificial neural networks is the lack of sufficient data; but nowadays it has become easier to access data sets called "Big Data". With the development of hardware technologies such as CPU (Central Process Unit) and GPU (Graphics Processing Unit), it has become possible to train and develop artificial neural networks. Today, it has been one of the first terms that come to mind when it comes to artificial intelligence or machine learning. The term deep learning was first used in 1986 by Rina Dechter in her work titled “Learning While Searching in Constraint-Satisfaction-Problems”. The word “deep” in deep learning comes from the large number of layers in the network, in other words, the layers in the network being deeper and wider. Deep learning is used to learn from pictures, video, audio, and video, instead of learning with coded rules as in traditional machine learning methods. and learn from text data [83]. In other words, it performs the learning process on raw data.184 Deep learning is a learning based on the representation of data. In this context, the characteristics of the data are divided into various representation groups and transformed into a hierarchical structure. In this hierarchical structure, the features in the upper layer are obtained by deriving the features in the lower layer. For example, raw data in image processing; corners, parts of objects, regional shapes, etc. are processed by separating them into representation groups. To give a more concrete example, in the processing of human images in deep neural networks, edges and simple shapes are processed in the first layer and transferred to the middle layer. In the intermediate layers, more complex structures such as hands, feet, arms, and heads are processed to form a hierarchy of the data [84].

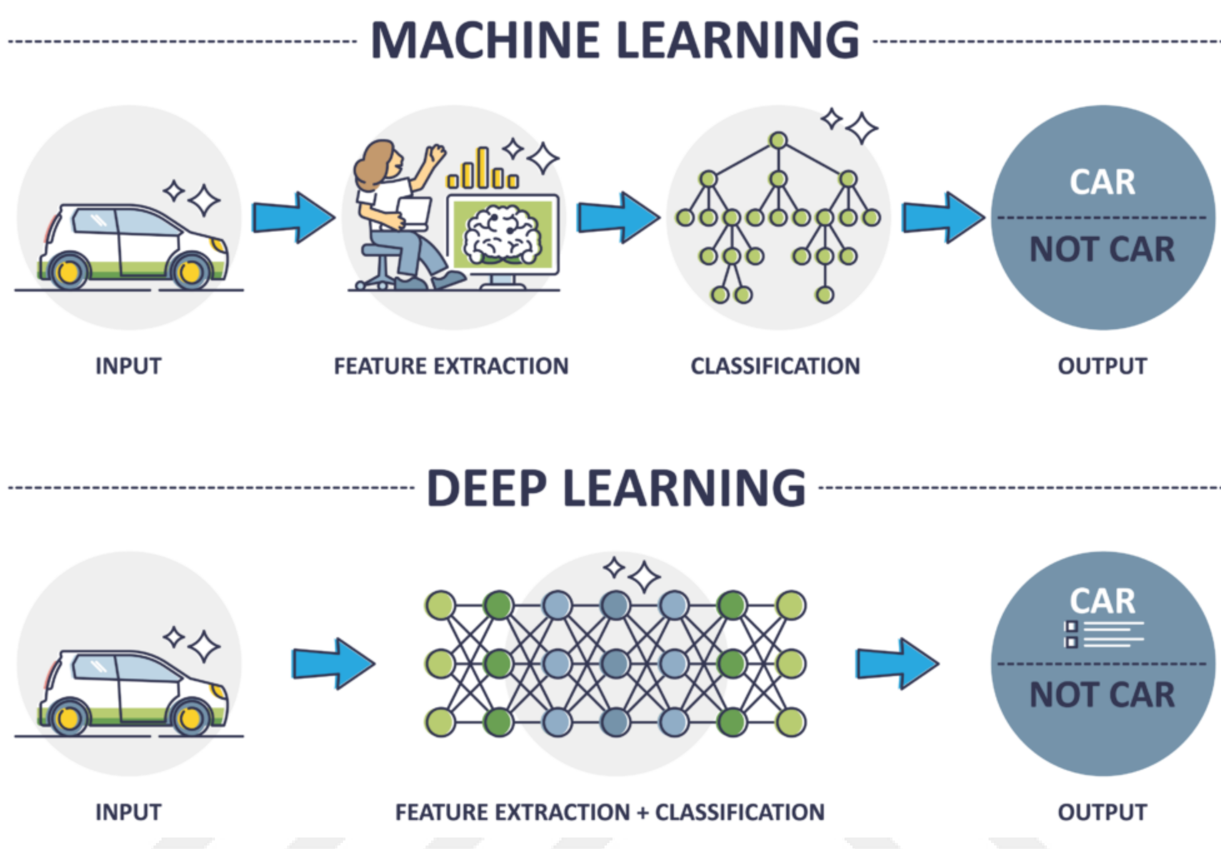


Figure 3.9: Deep learning vs Machine learning

3.2.5 Deep Belief Network (DBN)

Artificial neural networks, another application of artificial intelligence whose main purpose is to imitate the intelligence of living things, is a system created by imitating the nervous system of humans. Artificial neural networks are designed to simulate some functions and learning mechanism of the brain. In this context, when an information is sent to the artificial neural networks, this information primarily reaches the input layer, which performs the same task as the dendrites in the nerve cell [85]. This information coming from the input layer is then transferred to the hidden layer (it performs the body function in the nerve cell). Then it is sent from the hidden layer to the output layer (it performs the task of the axons in the nerve cell).144 Artificial neural networks are divided into two groups according to the number of layers they contain. Accordingly, artificial neural networks consisting of only input and output layers are single-layer neural networks; Artificial neural networks with intermediate layers together with input and output layers are called multilayer neural networks. According to the working structure of the network, artificial neural networks are divided into feed forward artificial neural networks (only forward between the layers of the data flow) and feedback / propagation artificial neural networks (not only forwards between the layers of the data flow). Threshold values are obtained by multiplying the inputs from the outside world with the weight values in the input layer. Then, with the threshold values obtained, operations are performed between various summation and activation functions and these are processed into information. After all these processes, an output is obtained to be transferred to the outside world or to another nerve cell. The basic working structure of artificial neural networks is like this, and although there are differences according to the learning algorithms used in these systems, the number of hidden layers and the number of neurons in the hidden layer may vary in order to provide the best performance that can be obtained from the system. Neural doctor Warren McCulloch and mathematician Walter Pitts; Inspired by the computational ability of the human brain, they created the first artificial neural network model with electrical circuits in 1943. In 1949, Donald Hebb wrote the book "Organization of Behavior" describing the learning process. This process, known as "Hebb's Rule", stated that for artificial neural networks with the ability to learn and adapt, the learning function will be realized by changing the number of connections of the artificial neural network. Hebb's rule has been a source of inspiration for many studies. The “Perceptron” model was developed by Frank Rosentblatt in 1957 [86]. This model; It is a single layer learning neural network consisting of a single nerve with a single output.148 John Von Neuman, on the other hand, suggested simulating simple nerve functions using vacuum tubes and telegraph relays to assist in this work. In 1959, Bernand Widrow and Marcian Hoff “Adaline and developed an artificial neural network called Madaline. In this model, Madaline acts as a filter to prevent echoes on telephone lines. This artificial neural network model is the first artificial neural network to be used in practice and has been the source of many studies. It has a variety of applications such as control [87,88].

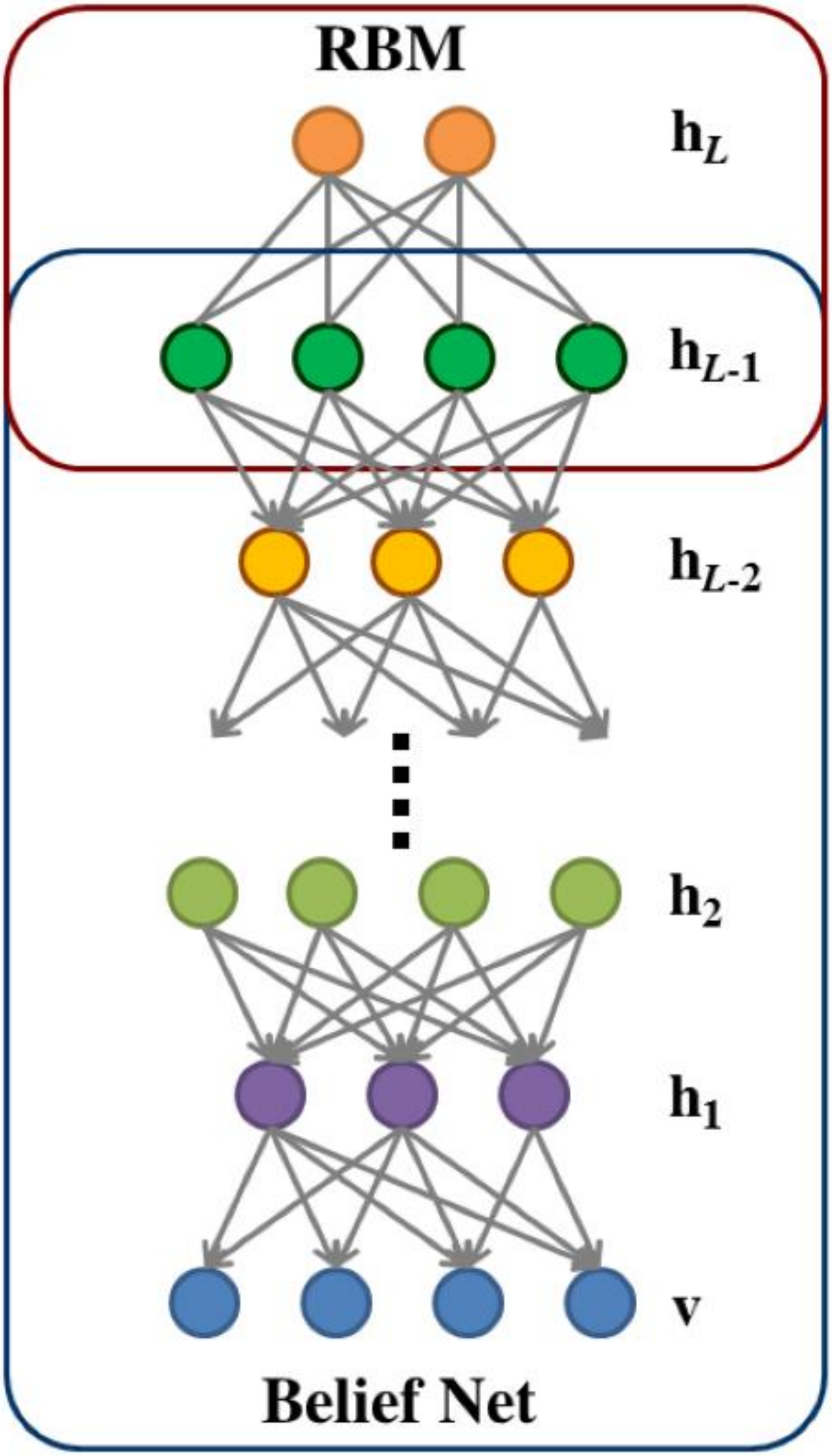


Figure 3.10: DBN Structure

3.3 A NEW MALWARE DETECTION SYSTEM BASED DBN

LDA is feature selection method applied to analysis the malware dataset. The LDA analysis the dataset that consist from 513 input features and reduce this size to 10% by selecting only the important features. then, the output of this step wired to the DBN. The purpose of a DBN is to extract high-level attributes and sensitive attributes. In this article, the DBN consists of two RBMs, and these RBMs are functions that are incrementally extracted, which reduces the size of the input. This method can be easily formulated to handle data of any size by increasing and decreasing the number of RBMs in a DBN. The first step in this process is to extract the functionality from the input. The extracted function is always connected to the planar neural network found in the last step of each procedure. A planar neural network used to classify features tracked, trained, and extracted from previous RBM layers. The flowchart of the proposed method presented in the figure 3.11.

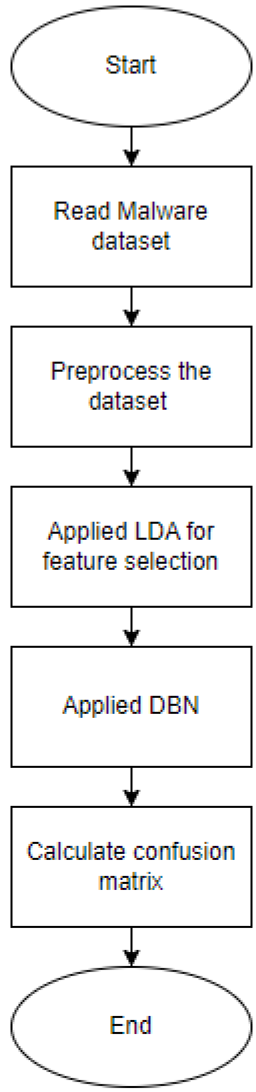


Figure 3.11: DBN Structure

4. EXPERIMENTS AND DISSCUSION

In this section, the results of the proposed method presented and explained in details. In first execution. DT used to classify the ECG signals and the results presented in the figure 4.1.

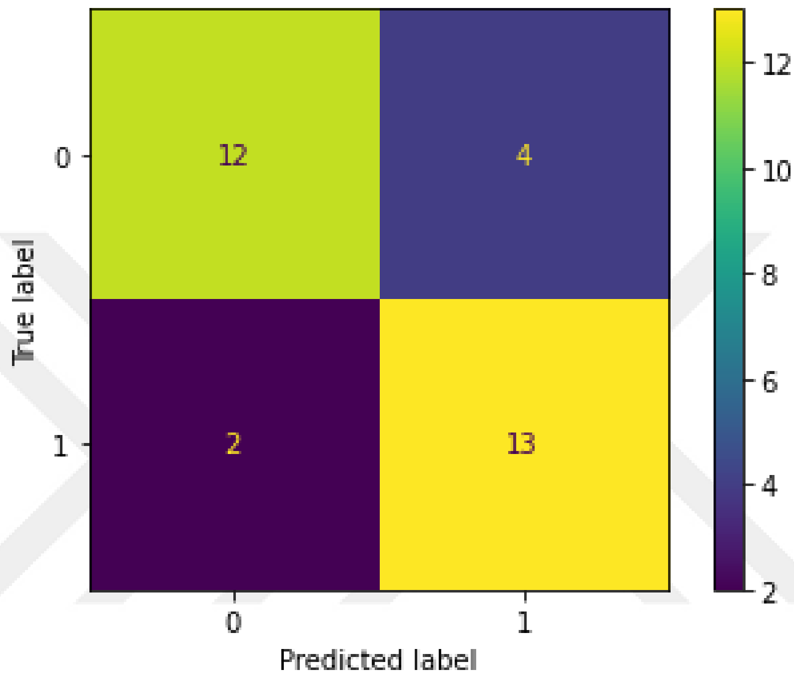
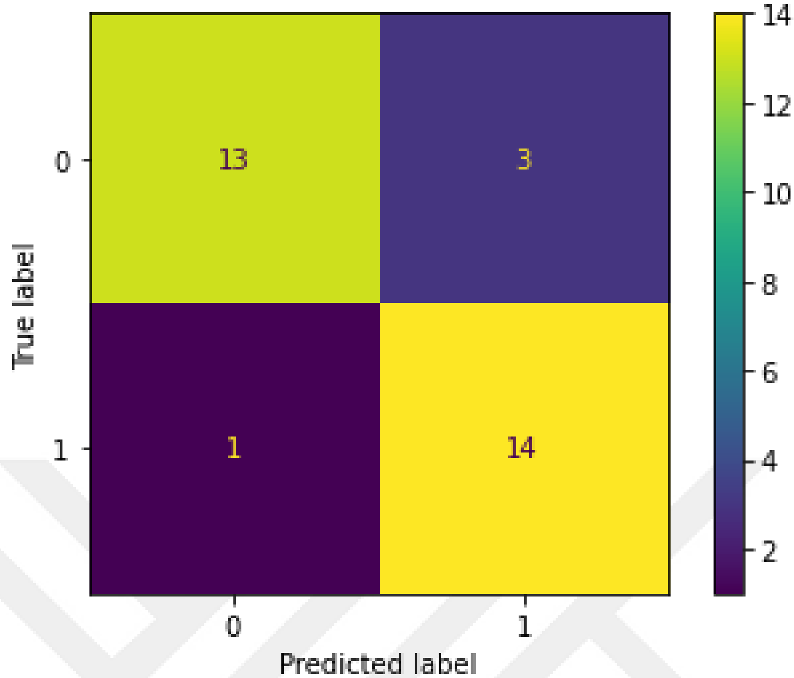


Figure 4.1: DT Results

Then, in the second experiment the results of the random forest presented in the Figure 4.4.



. **Figure 4.2:** Random Forest Results

Furthermore, the SVM results presented in the figure 4.3.

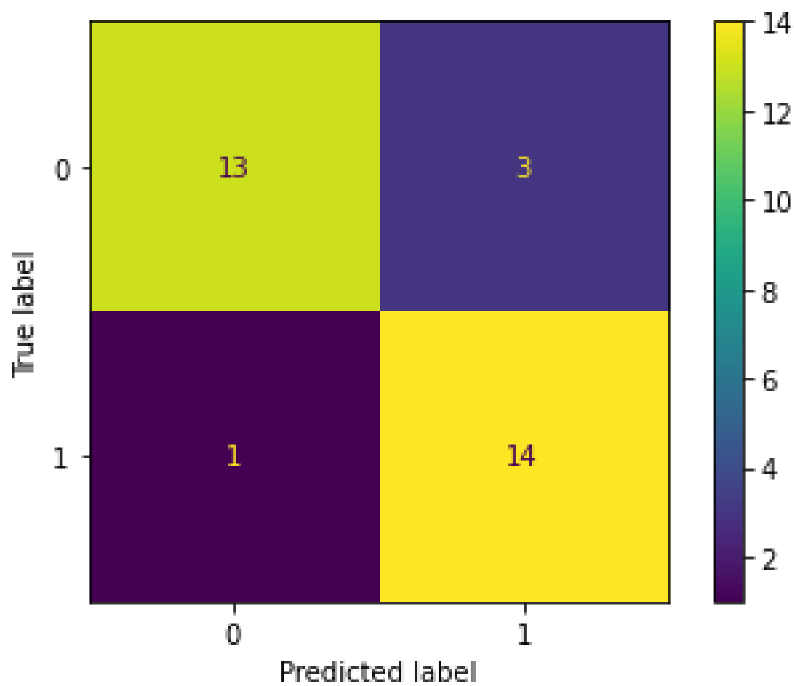


Figure 4.3: SVM Results

Moreover, the results that presented by the KNN shown in the Figure 4.4.

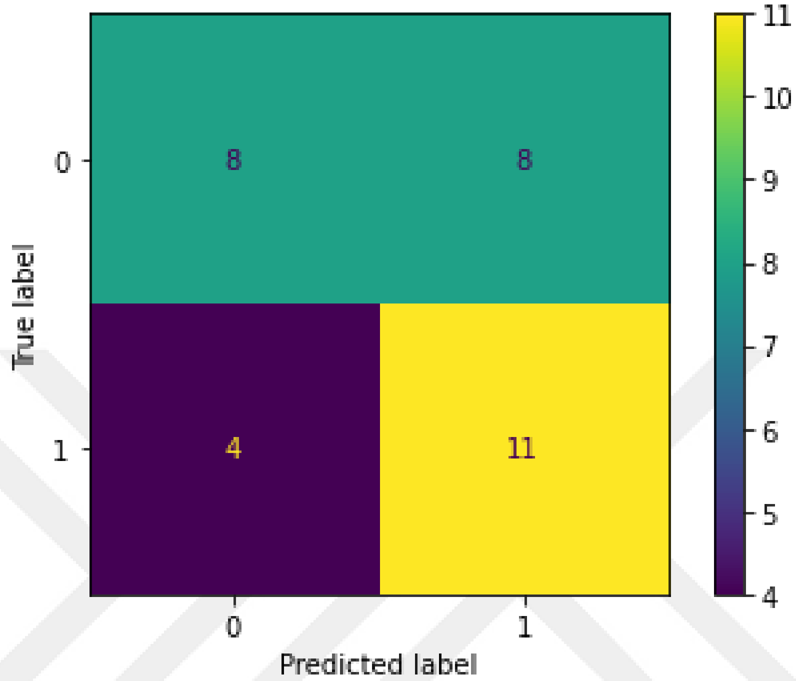


Figure 4.4: KNN Results

Finally, in figure 4.5 the AdaBoost applied to classify the heart signals to the normal and abnormal as shown in the figure 4.5.

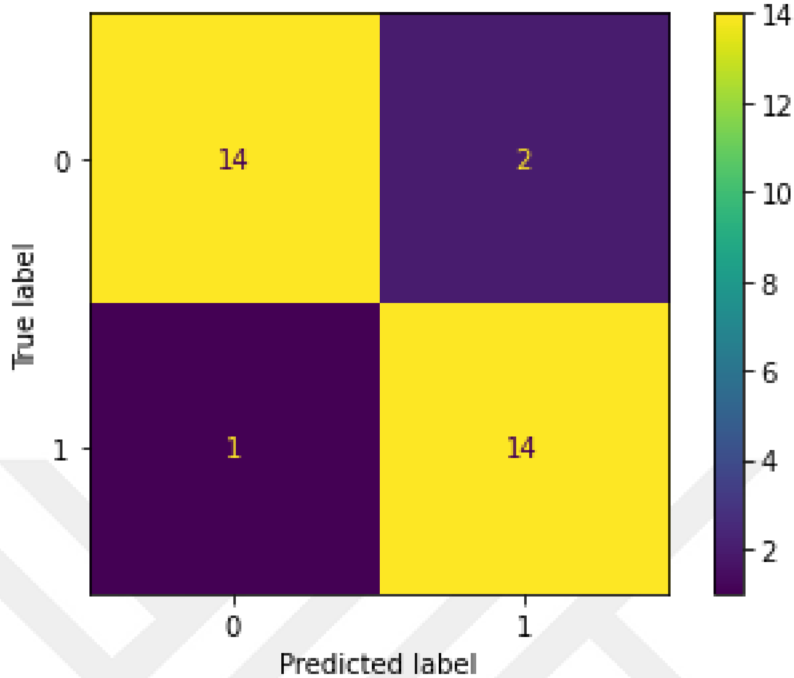


Figure 4.5: Proposed Method Results

5. CONCLUSIONS

Recently, cyber attacks have become much more serious problems for individuals, institutions and governments. Attackers use various types of cyber threats to infiltrate and attack information systems. Cyber threats; Along with cyber-attacks that create new crimes that have emerged by making use of the possibilities of information and communication technologies; It also includes versions of existing crimes transferred to cyberspace, in which the means of information and communication technologies are used. Cyber security is of critical importance because cyber attacks cause irreparable economic losses to individuals, institutions and governments, as well as create serious problems in terms of reputation and image.

The Malicious Code is; virus, worm, trojan horse, Bots/Botnets as the common name. This type of malicious code/software first infects a computer on the network and then spreads through that infected computer to other computers on the network. It is designed to damage computer systems and networks, render them inoperable, and negatively affect their functioning. In addition to malicious software, the Unwanted Program also poses a threat to computers by unwanted software, programs such as browser parasites, adware, spyware, and so on. These programs are secretly installed on the computer, although the user does not give any permission, and usually performs operations such as advertising, collecting and analyzing the user's internet traffic, and stealing information. Unwanted e-mail, more commonly known as spam messages; Although most of them are sent via e-mail, other electronic communication channels/services such as SMS, MMS, instant message, search engines, blogs, and VoIP may also encounter such messages. Using a more general definition, OECD defined it as electronic messages, junk mail, junk news posts, and mass posts generally related to commercial activities (

In this study, new method presented to detect the malware attacks in IoTs by using deep belief network based LDA. The aim of LDA in this study is to reduce the size of input malware features. The malware dataset is very critical and sensitive dataset which the dataset consist features that analyzing of these features assist to detect that there are malware or not. The LDA reduce the size of input features carefully which mean extracted sensitive features by deleting unaffected features from input dataset.

The presented method validated using dataset obtained from Kaggle and the results in most tests be with 100% accuracy. The applying LDA really assist the DBN to enhance its performance. The DBN show also remarkable performance when combined with LDA.

In the future works, the researcher can apply other feature selection techniques such as: PCA, K-mean, ICA to combine with deep learning frameworks to present best performance. Furthermore, there are several deep learning techniques that can applied instead of deep belief network such as recurrent neural network and long short term memory.

REFERENCES

[1] V. Mitra, H. Franco, M. Graciarena and D. Vergyri, "Medium-duration modulation cepstral feature for robust speech recognition," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 1749-1753, doi: 10.1109/ICASSP.2014.6853898.

[2] O. Hazrati, S. Ghaffarzadegan and J. H. L. Hansen, "Leveraging automatic speech recognition in cochlear implants for improved speech intelligibility under reverberation," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, QLD, 2015, pp. 5093-5097, doi: 10.1109/ICASSP.2015.7178941.

[3] R. Lotfidereshgi and P. Gournay, "Biologically inspired speech emotion recognition," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 5135-5139, doi: 10.1109/ICASSP.2017.7953135.

[4] Y. Suh *et al.*, "Development of distant multi-channel speech and noise databases for speech recognition by in-door conversational robots," *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*, Seoul, 2017, pp. 1-4, doi: 10.1109/ICSDA.2017.8384419.

[5] P. S. Nidadavolu, V. Iglesias, J. Villalba and N. Dehak, "Investigation on Neural Bandwidth Extension of Telephone Speech for Improved Speaker Recognition," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 6111-6115, doi: 10.1109/ICASSP.2019.8682992.

[6] L. Li, W. Xu, J. Wu, S. He and X. Li, "The Hokkien isolated word recognition system based on FPGA," *2014 International Conference on Anti-Counterfeiting, Security and Identification (ASID)*, Macao, 2014, pp. 1-5, doi: 10.1109/ICASID.2014.7064971.

[7] W. Shu-Guang, Z. Xiang-Yang and W. Qiang, "Isolated word recognition in reverberant environments," *2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Xi'an, 2011, pp. 1-4, doi: 10.1109/ICSPCC.2011.6061575.

[8] S. Sawant and M. Deshpande, "Isolated Spoken Marathi Words Recognition Using HMM," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-4, doi: 10.1109/ICCUBEA.2018.8697457.

[9] F. Ang, H. Tsutsui and Y. Miyanaga, "Incorporation of time-varying LP cepstral features in HMM-based isolated word speech recognition," *2015 International Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, 2015, pp. 1-4, doi: 10.1109/ISSCS.2015.7204030.

[10] S. C. Sajjan and C. Vijaya, "Comparison of DTW and HMM for isolated word recognition," *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, Salem, Tamilnadu, 2012, pp. 466-470, doi: 10.1109/ICPRIME.2012.6208391.

[11] S. Masood, M. Mehta, Namrata and D. R. Rizvi, "Isolated word recognition using neural network," *2015 Annual IEEE India Conference (INDICON)*, New Delhi, 2015, pp. 1-5, doi: 10.1109/INDICON.2015.7443697.

[12] M. Sher, N. Ahmad and M. Sher, "TESPAR feature based isolated word speaker recognition system," *18th International Conference on Automation and Computing (ICAC)*, Loughborough, 2012, pp. 1-4.

[13] C. Wei and Y. Yang, "Mandarin isolated words recognition method based on pitch contour," *2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, Hangzhou, 2012, pp. 143-147, doi: 10.1109/CCIS.2012.6664385.

[14] G. Wang, Y. Zhang, M. Sun, X. Wang and Y. Zhang, "Speech signal feature parameters extraction algorithm based on PCNN for isolated word recognition," *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, Shanghai, 2016, pp. 679-682, doi: 10.1109/ICALIP.2016.7846618.

[15] M. Raczynski, "Speech processing algorithm for isolated words recognition," *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Swinoujście, 2018, pp. 27-31, doi: 10.1109/IIPHDW.2018.8388238.

[16] Preeti Saini, Parneet Kaur, Mohit Dua, “Hindi Automatic Speech Recognition Using HTK,” International Journal of Engineering Trends and Technology (IJETT) – Volume4 Issue6- June 2013.

[17] Gaurav, Devanesamoni Shakina Devi, Gopal Krishna Sharma, Mahua Bhattacharya, “Development of Application Specific Continuous Speech Recognition System in Hindi”, Journal of Signal and Information Processing, 2012, 3, 394-401.

[18] M. K. Linga Murthy, G.L.N. Murthy,"Isolated Word Recognition Using LPC & Vector Quantization", International Journal of Research in Engineering and Technology(IJRET)- Volume1 Issue3- Nov2012.

[19] L. Xu and M. Ke, "Research on isolated word recognition with DTW-based," *2012 7th International Conference on Computer Science & Education (ICCSE)*, Melbourne, VIC, 2012, pp. 139-141, doi: 10.1109/ICCSE.2012.6295044.

[20] Hall T, Beecham S, Bowes D, Gray D, Counsell S. “A systematic literature review on fault prediction performance in software engineering”. IEEE Transactions on Software Engineering, 38(6), 1276-1304, 2012.

[21] Catal C, Diri B. “A systematic review of software fault prediction studies”. Expert Systems with Applications, 36(4), 7346-7354, 2009.

[22] Pal B, Hasan A, Aktar M, Shahdat N. “Cluster ensemble and probabilistic neural network modeling of class ımbalance learning in software defect prediction”. Artificial Intelligence and Applications, In Press.

[23] Shirabad S, Menzies TJ. School of Information Technology and Engineering, University of Ottawa. “The PROMISE repository of software engineering databases”. http://promise.site.uottawa.ca/SERepository (01.10.2017).

[24] Koru A, Liu H. “Building effective defect-prediction models in practice”. IEEE Software, 22(6), 23-29, 2005..

[25] Pelayo L, Dick S. “Applying novel resampling strategies to software defect prediction”. Fuzzy Information Processing Society NAFIPS ’07, San Diego, USA, 24-27 June, 2007.

[26] Frank E, Witten IH. "Generating accurate rule sets without global optimization". 15th International Conference on Machine Learning, Wisconsin, USA, 24-27 July 1998.

[27] Menzies T, Greenwald J, Frank A. "Data mining static code attributes to learn defect predictors". IEEE Transactions on Software Engineering, 33(1), 2-13, 2007.

[28] Lessmann S, Baesens B, Mues C, Pietsch S. "Benchmarking classification models for software defect prediction: A proposed framework and novel findings". IEEE Transactions on Software Engineering, 34(4), 485-496, 2008.

[29] Japkowicz N, Stephen S. "The class imbalance problem: A systematic study". Intelligent Data Analysis, 6(5), 429-449, 2002.

[30] Batista G, Prati R, Monard M, "A Study of the Behavior of several methods for balancing machine learning training data". ACM SIGKDD Explorations Special issue on learning from imbalanced datasets, 6(1), 20-29, 2004.

[31] He H, Bai Y, Garcia EA, Li S. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning". 2008 IEEE International Joint Conference on Neural Networks, Hong Kong, China, 1-8 June 2008.

[32] Xiangwen Wang, Xianghong Lin, Xiaochao Dang, Supervised learning in spiking neural networks: A review of algorithms and evaluations, Neural Networks, Volume 125, 2020, Pages 258-280, ISSN 0893-6080.

[33] Khalil Moshkbar-Bakhshayesh, Performance study of bayesian regularization based multilayer feed-forward neural network for estimation of the uranium price in comparison with the different supervised learning algorithms, Progress in Nuclear Energy, Volume 127, 2020, 103439, ISSN 0149-1970.

[34] Jiabin Zhang, Hu Su, Wei Zou, Xinyi Gong, Zhengtao Zhang, Fei Shen, CADN: A weakly supervised learning-based category-aware object detection network for surface defect detection, Pattern Recognition, Volume 109, 2021, 107571, ISSN 0031-3203.

[35]] Lu Han, Wenjun Li, Zhi Su, An assertive reasoning method for emergency response management based on knowledge elements C4.5 decision tree, Expert Systems with

Applications, Volume 122, 2019, Pages 65-74, ISSN 0957-4174.

[36] X. Wang, C. Zhou, X. Xu, Application of C4.5 decision tree for scholarship evaluations, Procedia Computer Science, Volume 151, 2019, Pages 179-184, ISSN 1877-0509.

[37] Sunanda Gamage, Jagath Samarabandu, Deep learning methods in network intrusion detection: A survey and an objective comparison, Journal of Network and Computer Applications, Volume 169, 2020, 102767, ISSN 1084-8045.

[38] Fatsuma Jauro, Haruna Chiroma, Abdulsalam Y. Gital, Mubarak Almutairi, Shafi'i M. Abdulhamid, Jemal H. Abawajy, Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend, Applied Soft Computing, Volume 96, 2020, 106582, ISSN 1568-4946.

[39] Y. Li, B. Sixou, F. Peyrin, A review of the deep learning methods for medical images super resolution problems, IRBM, 2020, ISSN 1959-0318.

[40] Xiyun Yang, Yanfeng Zhang, Wei Lv, Dong Wang, Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier, Renewable Energy, 2020, ISSN 0960-1481.

[41] Shaolong Sun, Shouyang Wang, Yunjie Wei, A new ensemble deep learning approach for exchange rates forecasting and trading, Advanced Engineering Informatics, Volume 46, 2020, 101160, ISSN 1474-0346.

[42] S. Hernández, Juan L. López, Uncertainty quantification for plant disease detection using Bayesian deep learning, Applied Soft Computing, Volume 96, 2020, 106597, ISSN 1568-4946.

[43] Hui Yin, Yuanhao Gong, Guoping Qiu, Fast and efficient implementation of image filtering using a side window convolutional neural network, Signal Processing, Volume 176, 2020, 107717, ISSN 0165-1684.

[44] Samira Zare, Moosa Ayati, Simultaneous fault diagnosis of wind turbine using multichannel convolutional neural networks, ISA Transactions, 2020, ISSN 0019-0578.

[45] Qi Li, Pengfei Li, Kezhi Mao, Edmond Yat-Man Lo, Improving convolutional neural network for text classification by recursive data pruning, Neurocomputing, Volume 414, 2020, Pages 143-152, ISSN 0925-2312.

[46] JongCheol Pyo, Lan Joo Park, Yakov Pachepsky, Sang-Soo Baek, Kyunghyun Kim, Kyung Hwa Cho, Using convolutional neural network for predicting cyanobacteria concentrations in river water, Water Research, Volume 186, 2020, 116349, ISSN 0043-1354.

[47] He Li, Xiaohui Li, Feng Yuan, Simon M. Jowitt, Mingming Zhang, Jie Zhou, Taofa Zhou, Xiangling Li, Can Ge, Bangcai Wu, Convolutional neural network and transfer learning based mineral prospectivity modeling for geochemical exploration of Au mineralization within the Guandian–Zhangbaling area, Anhui Province, China, Applied Geochemistry, 2020, 104747, ISSN 0883-2927.

[48] Marcus Stoffel, Franz Bamer, Bernd Markert, Deep convolutional neural networks in structural dynamics under consideration of viscoplastic material behaviour, Mechanics Research Communications, Volume 108, 2020, 103565, ISSN 0093-6413.

[49] Amin Khatami, Asef Nazari, Abbas Khosravi, Chee Peng Lim, Saeid Nahavandi, A weight perturbation-based regularisation technique for convolutional neural networks and the application in medical imaging, Expert Systems with Applications, Volume 149, 2020, 113196, ISSN 0957-4174,.

[50] Kaiyu Zhang, Jinglong Chen, Tianci Zhang, Zitong Zhou, A Compact Convolutional Neural Network Augmented with Multiscale Feature Extraction of Acquired Monitoring Data for Mechanical Intelligent Fault Diagnosis, Journal of Manufacturing Systems, Volume 55, 2020, Pages 273-284, ISSN 0278-6125.

[51] Jialin Tang, Qinglang Su, Binghua Su, Simon Fong, Wei Cao, Xueyuan Gong, Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition, Computer Methods and Programs in Biomedicine, Volume 197, 2020, 105622, ISSN 0169-2607.

[52] Zhi-jun Lu, Qi Qin, Hong-yin Shi, Hao Huang, SAR moving target imaging based on convolutional neural network, Digital Signal Processing, Volume 106, 2020, 102832, ISSN

1051-2004,.

[53] M.V. Valueva, N.N. Nagornov, P.A. Lyakhov, G.V. Valuev, N.I. Chervyakov, Application of the residue number system to reduce hardware costs of the convolutional neural network implementation, Mathematics and Computers in Simulation, Volume 177, 2020, Pages 232-243, ISSN 0378-4754.

[54] Zhiying Fang, Han Feng, Shuo Huang, Ding-Xuan Zhou, Theory of deep convolutional neural networks II: Spherical analysis, Neural Networks, Volume 131, 2020, Pages 154-162, ISSN 0893-6080.

[55] Rachna Jain, Preeti Nagrath, Gaurav Kataria, V. Sirish Kaushik, D. Jude Hemanth, Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning, Measurement, Volume 165, 2020, 108046, ISSN 0263-2241.

[56] Priyadarshiny Dhar, Saibal Dutta, Vivekananda Mukherjee, Cross-wavelet assisted convolution neural network (AlexNet) approach for phonocardiogram signals classification, Biomedical Signal Processing and Control, Volume 63, 2021, 102142, ISSN 1746-8094.

[57] Xuliang Liu, Daolun Li, Jinghai Yang, Wenshu Zha, Ziqi Zhou, Liping Gao, Jiahang Han, Automatic well test interpretation based on convolutional neural network for infinite reservoir, Journal of Petroleum Science and Engineering, Volume 195, 2020, 107618, ISSN 0920-4105.

[58] Adil Moghar, Mhamed Hamiche, Stock Market Prediction Using LSTM Recurrent Neural Network, Procedia Computer Science, Volume 170, 2020, Pages 1168-1173, ISSN 1877-0509.

[59] Refat Khan Pathan, Munmun Biswas, Mayeen Uddin Khandaker, Time series prediction of COVID-19 by mutation rate analysis using recurrent neural network-based LSTM model, Chaos, Solitons & Fractals, Volume 138, 2020, 110018, ISSN 0960-0779.

[60] Alex Sherstinsky, Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network, Physica D: Nonlinear Phenomena, Volume 404, 2020, 132306, ISSN 0167-2789.

[61] Pritika Bahad, Preeti Saxena, Raj Kamal, Fake News Detection using Bi-directional LSTM-Recurrent Neural Network, Procedia Computer Science, Volume 165, 2019, Pages 74-82, ISSN 1877-0509.

[62] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, Yinhai Wang, Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values, Transportation Research Part C: Emerging Technologies, Volume 118, 2020, 102674, ISSN 0968-090X.

[63] Ying Chen, Voltages prediction algorithm based on LSTM recurrent neural network, Optik, Volume 220, 2020, 164869, ISSN 0030-4026.

[64] Neculai Andrei, A Dai–Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization, Applied Mathematics Letters, Volume 21, Issue 2, 2008, Pages 165-171, ISSN 0893-9659.

[65] Muhammed Maruf Öztürk, İbrahim Arda Cankaya, Deniz İpekçi, Optimizing echo state network through a novel fisher maximization based stochastic gradient descent, Neurocomputing, Volume 415, 2020, Pages 215-224, ISSN 0925-2312.

[66] Andreea Koreanschi, Oliviu Sugar Gabor, Joran Acotto, Guillaume Brianchon, Gregoire Portier, Ruxandra Mihaela Botez, Mahmoud Mamou, Youssef Mebarki, Optimization and design of an aircraft's morphing wing-tip demonstrator for drag reduction at low speed, Part I – Aerodynamic optimization using genetic, bee colony and gradient descent algorithms, Chinese Journal of Aeronautics, Volume 30, Issue 1, 2017, Pages 149-163, ISSN 1000-9361.

[67] Arnulf Jentzen, Philippe von Wurstemberger, Lower error bounds for the stochastic gradient descent optimization algorithm: Sharp convergence rates for slowly and fast decaying learning rates, Journal of Complexity, Volume 57, 2020, 101438, ISSN 0885-064X, https://doi.org/10.1016/j.jco.2019.101438..

[68] Xuemei Dong, Ding-Xuan Zhou, Learning gradients by a gradient descent algorithm, Journal of Mathematical Analysis and Applications, Volume 341, Issue 2, 2008, Pages 1018-1027, ISSN 0022-247X, https://doi.org/10.1016/j.jmaa.2007.10.044..

[69] Rui Ye, Qun Dai, Implementing transfer learning across different datasets for time series forecasting, Pattern Recognition, Volume 109, 2021, 107617, ISSN 0031-3203,.

[70] Yu Liu, Ao Li, Xing-Ming Zhao, Minghui Wang, DeepTL-Ubi: A novel deep transfer learning method for effectively predicting ubiquitination sites of multiple species, Methods, 2020, ISSN 1046-2023,..

[71] Jingyao Wu, Zhibin Zhao, Chuang Sun, Ruqiang Yan, Xuefeng Chen, Few-shot transfer learning for intelligent fault diagnosis of machine, Measurement, Volume 166, 2020, 108202, ISSN 0263-2241,..

[72] Shanshan Wang, Lei Zhang, Jingru Fu, Adversarial transfer learning for cross-domain visual recognition, Knowledge-Based Systems, Volume 204, 2020, 106258, ISSN 0950-7051,..

[73] Xiyun Yang, Yanfeng Zhang, Wei Lv, Dong Wang, Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier, Renewable Energy, 2020, ISSN 0960-1481,..

[74] Chuan Li, Shaohui Zhang, Yi Qin, Edgar Estupinan, A systematic review of deep transfer learning for machinery fault diagnosis, Neurocomputing, Volume 407, 2020, Pages 121-135, ISSN 0925-2312,..

[75] Yue Wang, Yuting Liu, Wei Chen, Zhi-Ming Ma, Tie-Yan Liu, Target transfer Q-learning and its convergence analysis, Neurocomputing, Volume 392, 2020, Pages 11-22, ISSN 0925-2312,..

[76] Zhengjun Qiu, Shutao Zhao, Xuping Feng, Yong He, Transfer learning method for plastic pollution evaluation in soil using NIR sensor, Science of The Total Environment, Volume 740, 2020, 140118, ISSN 0048-9697,..

[77] Santi Kumari Behera, Amiya Kumar Rath, Prabira Kumar Sethy, Maturity status classification of papaya fruits based on machine learning and transfer learning approach, Information Processing in Agriculture, 2020, ISSN 2214-3173,..

[78] Seunghyeon Kim, Yung-Kyun Noh, Frank C. Park, Efficient neural network compression

via transfer learning for machine vision inspection, Neurocomputing, Volume 413, 2020, Pages 294-304, ISSN 0925-2312,..

[79] Xiang Li, Wei Zhang, Hui Ma, Zhong Luo, Xu Li, Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks, Neural Networks, Volume 129, 2020, Pages 313-322, ISSN 0893-6080,..

[80] Piyush Kant, Shahedul Haque Laskar, Jupitara Hazarika, Rupesh Mahamune, CWT Based Transfer Learning for Motor Imagery Classification for Brain computer Interfaces, Journal of Neuroscience Methods, Volume 345, 2020, 108886, ISSN 0165-0270,..

[81] Xinghua Li, Zhongyuan Hu, Mengfan Xu, Yunwei Wang, Jianfeng Ma, Transfer learning based intrusion detection scheme for Internet of vehicles, Information Sciences, Volume 547, 2021, Pages 119-135, ISSN 0020-0255,..

[82] Sheikh Saud, Basharat Jamil, Yogesh Upadhyay, Kashif Irshad, Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach, Sustainable Energy Technologies and Assessments, Volume 40, 2020, 100768, ISSN 2213-1388,..

[83] Jie Wei, Hui Chen, Determining the number of factors in approximate factor models by twice K-fold cross validation, Economics Letters, Volume 191, 2020, 109149, ISSN 0165-1765,..

[84] Tzu-Tsung Wong, Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, Pattern Recognition, Volume 48, Issue 9, 2015, Pages 2839-2846, ISSN 0031-3203,..

[85] Gaoxia Jiang, Wenjian Wang, Error estimation based on variance analysis of k-fold cross-validation, Pattern Recognition, Volume 69, 2017, Pages 94-106, ISSN 0031-320..

[86] Nima Shiri Harzevili, Sasan H. Alizadeh, Analysis and modeling conditional mutual dependency of metrics in software defect prediction using latent variables, Neurocomputing, Volume 460, 2021, Pages 309-330, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.05.043..

[87] Chao Ni, Xiang Chen, Fangfang Wu, Yuxiang Shen, Qing Gu, An empirical study on pareto based multi-objective feature selection for software defect prediction, Journal of Systems and Software, Volume 152, 2019, Pages 215-238, ISSN 0164-1212.

[88] Ruchika Malhotra, Shine Kamal, An empirical study to investigate oversampling methods for improving software defect prediction using imbalanced data, Neurocomputing, Volume 343, 2019, Pages 120-140, ISSN 0925-2312.