REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technologies

A NEW METHOD FOR DEVELOPING NEW INTRUSION DETECTION SYSTEM

Shahad Jasim Hasan GBURI

Master's Thesis

Supervisor

Asst. Prof. Dr. Oğuz KARAN

Istanbul, 2022

A NEW METHOD FOR DEVELOPING NEW INTRUSION DETECTION SYSTEM

Shahad Jasim Hasan GBURI

Information Technologies

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled A NEW METHOD FOR DEVELOPING NEW INTRUSION DETECTION SYSTEM prepared by SHAHAD JASIM HASAN GBURI and submitted on 21/12/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technologies.

Asst.Prof. Dr. Oğuz KARAN

Supervisor

Thesis Defense Committee Members:

Asst.Prof. Dr. Oğuz KARAN	Department of Software Engineering, Altınbaş University	_____
Asst.Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altınbaş University	_____
Assoc. Prof. Dr. Adil Deniz DURU	Department of Trainer Education, 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 Institute of Graduate Studies: ___/___/___

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

Shahad Jasim Hasan GBURI

Signature

DEDICATION

To my father who was taken from me (may he rest in peace)The first person who believed in my abilities and with love encouraged me to go into the experience of postgraduate studies, to my generous mother, the first teacher who taught me to challenge myself and tame difficulties (may god prolong her life). My brother and sisters, you are the stars that decorate my sky, My friends and relatives; who have helped me through my study. To my professors who helped write this thesis, I dedicate my appreciation and this scientific effort to you.

ABSTRACT

A NEW METHOD FOR DEVELOPING NEW INTRUSION DETECTION SYSTEM

GBURI , Shahad Jasim Hasan

M.Sc., Information Technology, Altınbaş University,

Supervisor: Asst.Prof.Dr. Oğuz Karan

Date: December /2022

Pages: 68

After the increased use of digital devices connected to the Internet, which in turn leads to users sharing their personal information, the adoption of intergovernmental institutions, including civil ones, in their work of archiving, management, marketing, etc. . On networked distributed systems, security experts have made the focus of their efforts to provide the necessary cybersecurity, which is to prevent interference by unauthorized persons from gaining access to databases and to protect networks from malicious attacks that cause losses of funds or assets connected to digital networks. Network and in our research, we tried to address the problem of (DDoS) through machine learning technology, where the (Extra Trees Classifier) algorithm was used, which reduces the size of the input data with a low computational cost and low variance, and the Random Forest algorithm with a low bias to choose the best available results from the ExtraTreesClassifier and train The model using the (Dense Neural Network) for learning and testing (IDS) model achieved an accuracy of 99.85%.

Keywords: Distributed Denial of Service Attack, Artificial Intelligence, IDs, Deep Learning, DNN, Cyber Security, Extra Trees Classifier, Random Forest.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF CHARTS

Page

ABBREVIATIONS

DDoS	:	Distributed Denial-0f-Service Attack
IDS	:	Intrusion Detection System
DNN	:	Deep Neural Network
SVM	:	Support Vector Machine
LR	:	Linear Regression
ANN	:	Artificial Neural Network
DSL	:	Digital subscriber line
SBS	:	signature-based systems
ABS	:	Abnormal Intrusion Detection Systems
TCP	:	Transmission Control Protocol
IP	:	Internet Protocol address
ICCC	:	International Conference on Computer Communications
ARPANET	:	Advanced Research Projects Agency
OS	:	Operating System
RNN	:	recurrent neural networks
ELM	:	Extreme Learning Machine
NIDS	:	network-based intrusion detection systems
PIDS	:	Protocol Based Intrusion Detection System
APIDS	:	Application Protocol-Based Intrusion Detection System
ReLU	:	Rectified Linear Units

ML : MACHINE LEARNING

1. INTRODUCTION

In the past few years, cyber attacks on networked systems have increased, as stated by Derek Mankey expert in global security at Fortinet, a company specializing in marketing cyber security products, “Every minute we witness about half a million attack attempts occurring in cyberspace” [1], which Some global organizations may cost financial losses not insignificant, and these attacks varied to include data theft, fabrication of information, disruption of service or digital systems of individuals or organizations, and some attacks threatened assets that deal directly or indirectly with the Internet space[2][3]. During the last years of the twentieth century, communication technology has developed* and passed on several stages until the present time, in which the global network of the Internet has become an important part of all aspects of the life of the inhabitants of the planet, which are managed by networked applications and systems, including education, financial transactions, bank accounts, home monitoring systems, institutions, messaging applications and social communication Navigation, aviation, army, and others. Communication technology and data transmission began in the past through telephone wires with the digital subscriber line (DSL) technology provided via copper wires, which is a distance-sensitive technology, as the longer the connection, the lower the quality of Services. Because telecommunications services are received by cables and the high cost of expanding wired local networks It makes it necessary to develop wireless data transmission technology and use it in the applications field[4][5]. In 1972, the Communications Conference (ICCC) was held, at which the first wide area network was announced by the Advanced Research Projects Agency (ARPANET) and it is also the first network for packet switching with distributed control in which the TCP / IP protocol Which are considered the basis of wireless communication technology. Wireless networks began in the 1970s with a project ALOHANET Al-Research, a radiofrequency-based research project at the University of Hawaii, and in 2008 Wi-Fi was certified by the Wi-Fi Alliance. Wi-Fi technology is widely used in the telecom industry and continues to evolve to this day the main reason for the development of WiFi is that it is easy to install, cheaper than a wired network, and data can be shared over long distances. But like any technology in which some vulnerabilities are exploited by vandals, in this research, we address the problem of Distributed Denial of Service (DDoS) attacks, in which the attacker exploits the bandwidth characteristic of the target network. Network bandwidth can be

defined as a standard for evaluating the quality of a wireless network, as it represents the amount of data that is transferred between two parties during a certain period. It is consumed when downloading large files or flowing high-quality video content, and this slows down connections on other devices on the same network [5]. In October 2020, Google's Threat Analysis Group (TAG) posted to the site Their update shows the development of hackers in the style of threat and change their tactics in the US elections for the year 2020, and the most important observations that were mentioned is the monitoring of security engineers, a UDP amplification attack exported by Chinese Internet service providers, and it is considered the largest bandwidth attack that lasted for six months by three service providers Chinese and at a peak of 2.5 terabytes per second! The target was thousands of Google IP addresses. Damien Mincher, Security Reliability Engineer at Google writes: The attacker used several networks to spoof 167 Mpps (millions of packets per second) to the exposed 180,000 CLDAP, DNS, and SMTP servers, which would then send large responses to us. This shows the volumes that a well-resourced attacker could achieve: this was four times greater than The attack that broke the record 623 Gbps Mirai botnet the previous year". Attacks avoid identifying their source by using spoofed IP addresses and the DDoS attack is Sneaking to a web server, to send many requests to serve the page, the goal is to make it crash because of the number of intense requests, and disrupt the database receiving too many requests, the result is Internet bandwidth capacity overflow and unity Central processing and RAM and that is lead to causing a denial of service during the attack. One of the most prominent DDoS attacks in recent years is the Wannacry virus, known as global ransomware, due to which network services have been suspended in about 156 countries[6]. It is difficult to distinguish between legitimate packets (requests for service who are authorized to access the server) and illegal packets (DDoS attacks), which required the need to create systems that detect illegal packets before they reach the server, and they are known as Intrusion Detection Systems (IDS). There are two types of these systems (IDS), which are signature-based systems (SBS) and they contain a database that includes signatures of previously known attacks and works on a comparison between the signatures and the data received by the network, and if a match is found, an attack warning is issued[7]. Systems based on network traffic anomalies (ABS) work by machine learning technology by creating a statistical model that represents normal network traffic, then identifying any behavior that significantly deviates from the network's normal model and considering it an attack.[8]. Intrusion detection systems based on

anomalies in wireless network traffic are considered the best way to avoid new attacks because of their ability to distinguish illegal packets at the moment of their attack, unlike signature-based systems where they need to constantly update databases and therefore cannot detect attacks that do not have their signature, which It should be noted that a downside to ABS is the false alert, which results in the identification of traffic being legitimate as an attack[9]. Machine learning technology, deep learning, and neural networks play a major role in the design of anomaly detection systems in network traffic, as many studies that relied on machine learning showed positive results in detecting the sequence of intruders on network servers. There are three main types of machine learning methods and they are learning Supervised, semi-supervised, and unsupervised learning. Numerous machine learning algorithms are designed including Naive Bayesian Classifier, Deep Neural Network (DNN), Support Vector Machine (SVM), Regression, K-Near Neighbor (KNN), Linear Regression, Artificial Neural Networks (ANN), etc [10].

1.1 MOTIVATIONS

The increase in the use of the wireless network in the last years lead to several security problems occurring in this technology. Then, the researchers in the last year tried to present effective and novel studies and techniques to deal with these problems. furthermore, several machine learning and deep learning techniques applied to various security problems in the last years. The aim of using deep learning techniques is because of the remarkable results presented by these studies. Then, This study presents a learning model consisting of machine learning using selection algorithms with low computational cost and deep learning using dense neural network algorithms that are efficient in detecting DDoS attacks in the wireless network at the lowest cost and highest accuracy.

1.2 CONTRIBUTION

Machine learning technology has achieved success in solving modern digital problems such as face recognition, image analysis, translation, and others. It is no secret to many that the problem of cyber-attacks has emerged, which has led to the removal of networks from service. In previous studies, artificial intelligence and deep learning had a significant impact on the design of intrusion detection systems, and in our research, we contributed to the creation of a deep learning model. The data was analyzed and appropriate features were selected from the database to train the model taking advantage of the advantages of the

algorithm extra trees classıfer in terms of low variance. And the advantages of the algorithm are random forest, in terms of low bias, and then the training phase using a dense deep learning algorithm that is characterized by the speed of learning, which led to an increase in the accuracy of the model work and a decrease in false alarms.

2. BACKGROUND AND LITERATURE REVIEW

2.1 RELATED WORK

Kanimozhi [11] et al 2019 in their research that is published in the International Journal of Recent Technology and Engineering they are used the Anaconda 3 work environment was used with the Python 3 programming language, using the Jupyter Notebook tool and its Python Machine Learning Sklearn kit library.

Where the unwanted or redundant features were removed through the Sklearn library, and then the desired features were passed to the Classifier _Forest _Random algorithm, and the best four built-in features are ct_dst_src_ltm, s-bytes, Sttl, and s-load, where the results appeared with an accuracy of 89% [11].

Sarika [12] et al in 2020 In their research using deep learning technology in the Internet of Things used twenty hidden layers of the neural network (DNN) and divided the UNSW-N15 data into 70% training, 15% validation, 15% testing, and used the attributes of 8 data as inputs, they are 16,15,10,9,8,7,4,2 and showed accuracy in intrusion detection of 99.2%. The experiment was conducted by MatLab 2016b with 8GB of RAM and running on Mac OSX[12].

ALEESA [13] et al in 2021 they ar use the UNSW-NB15 data for deep learning prepared by filling in the empty fields with a value of zero and also converting the text data into numbers. Then the values of the numeric columns in the data set were changed to a common scale, without distorting the differences in Value ranges this procedure includes only the features of the different ranges and this processing is known as normalization. Then, the data is divided into three groups, 70% for training and 30% for verification and testing, followed by the fourth stage, which is to propose a system to detect attacks, which depends on the method of deep learning using models (ANN and DNN and RNN), and the results were as follows, the binary classification of ANN technology was with an accuracy of 99.26%, and the result of the DNN technique was with an accuracy of 99.22%, concerning the classification accuracy of RNN it was 85.42%.

and It was observed that the efficiency of the artificial neural network in the binary classification results is much better than the multiclass classification results. Where the accuracy was 99.26% for the binary classification and the accuracy was 97.89% for the multiclass classification. While the loss ratio for the binary rating was 1.51% and the loss

rate was 5.27% for the multi-class rating [13].

Sharma [14] et all in 2019 in their research about Creatıon a multi-layer system that detects intrusion by using ExtraTrees selection to select the best features from unsw_n, they are To identify the important features of each attack type individually for each (ELM) and learning model was built using an extreme learning machine ensemble, Then all ELMs results are combined using a softmax layer to improve results and increase accuracy even further. An accuracy of 98.24% was determined, and a group of ELMs was implemented in parallel with the GPUs for real-time intrusion detection [14].

Rani [15] et all in 2022 in their research about Effective network penetration detection by addressing class imbalance using deep neural networks multimedia tools and applications To successfully address this issue, a deep neural network method based on classifier-level imbalance is suggested. The min-max normalization method is used after the network data has first undergone data transformation preprocessing. Following that, the neural network receives the measured data and modifies the cross-entropy function to address the issue of category imbalance. Weighting the categories during classifier training achieves this. Numerous tests are run on two difficult data sets, NSL-KDD and UNSW-NB15, to show the superiority of the suggested methodology. It covers comparisons with currently available works as well as widely utilized imbalance tactics including under-sampling, over-sampling, and packing. In the NSL-KDD and UNSW-NB15 datasets, the suggested method achieves classification accuracy of 85.56% and 90.76%, respectively[15].

Lopez[16] et all in 2022 in their research about Intrusion detection based on analyzing network traffic behavior The models were taught using the CICIDS2017 database, and they were processed and executed from the null and text values, and the most important features were selected using (RFE) and the importance of the selected features was also confirmed during the operation of the classification algorithm. The results were as follows: Logistic Regression Model (MLR) with an accuracy of 88% And the KNN model, with an accuracy of 95% Random Forest (RF) is a common assembly method based on a standard machine The learning technique called "decision tree" has an accuracy of 99%. The DNN algorithm was able to detect attacks better than previous models and the ratios were as follows: DNN accuracy was 95% for DoS attacks and 99% for DDoS attacks. A standard specification computer was used with a data size of 1.1 million records. The model was trained with five hidden layers, each layer consisting of 50 cells [16].

Makuvaza [17] et all in 2021 in their research about Using DNN technology to detect DDoS attacks as they occur, The proposed model consists of four layers, each layer containing 128 neurons, and accuracy was The detection accuracy, of 97.25%.

Amaizu [18] et all in 2021 The researchers trained the models on a real database that was created over two days of network traffic monitoring and the features were obtained through the CICFlowMeter-V3 tool, the database is known as CICDDoS2019, which consists of 80 features and contains different types of malicious attacks such as TFTP and NetBIOS, they are used PCC was used as a method to select the best features. The training model consists of two DNN models, DNN 1 is composed of 7 hidden layers and two dropout layers, and DNN 2 consists of seven hidden layers, and two dropout layers, and the input layer of the second model receives its data from the output layer of the DNN model 1, and the accuracy score of 99.66 % [18].

2.2 HISTORY OF COMPUTER NETWORKS

The history of communication networks and their development to this day has not been easy. Although the invention technique and the marketing process are also not easy. However, it is necessary to identify the stages of development that this technology has gone through.

In the forties of the twentieth century, computers were of a large size, and thus all parts were vulnerable to damage, which led to the need to invent what is known today as semiconductors, and that was in 1947, through which computers were manufactured in small size and more reliable [19]. Work continued on the development of these semiconductors until the fifties of the same century, integrated electrical circuits were designed consisting of several transistors connected with semiconductors of small sizes. And in the sixties, computers were designed with a suitable small size known as minicomputers, which is considered a large size compared to the veil of current computers.

In 1977, the "13 Apple Computer" company produced a computer known as the Mac, and in 1981 the International Business Machines (IBM) announced the first personal computer (PC). In the latter half of the eighties, modems that were used to exchange files spread and was called bulletin board. One of the disadvantages of this device is the weakness of direct connection which requires each connection to have a modem, meaning if four people want to connect to the network at the same time, this means four modems must be connected to four phone lines different. From the sixties until the nineties, the US Department of Defense

worked on developing reliable Wide Area Networks (WANs) that were used for research for scientific and military purposes. This network was distinguished by the ability to connect many computers in various ways and they could send and receive data at the same time. The network decided how to send information between computers directly. This network is the same currently known as the World Wide Web [19].

2.3 PHYSICAL LAYER (PHYSICAL LAYER-LAYER 1)

It is the lowest, first and most complex layer of the seven OSI layers, which is responsible for the actual communication between network devices, in which communication requests that are in the form of bits are translated, converted to 0 and 1, and sent. It is a basic layer upon which the logical data structures that feed the higher-level functions in the network depend.

In it, means are defined to transfer raw bits instead of logical data packets. Bits can be assembled into encrypted symbols or words and converted into a physical signal via a ready transmission medium. In this layer, there are mechanical and electrical interfaces and media-specific operations used to transfer bits between network nodes [20][21].

2.4 WIRELESS PERSONAL AREA NETWORKS (WAPNS)

Networks consist of the connection of a number of computers and other small WPANS peripheral devices that connect the devices with each other over small distances wirelessly, which have a low data transfer rate and cover a short distance in terms of communication compared to fast WPAN networks, and the data transfer speed is estimated to 1 megabit per The second covers a distance of 10 meters.

The most famous application of this network is Bluetooth and HomeRF, and Bluetooth was developed as a technology to create a wireless connection between the electronic devices surrounding a person, while HomeRF was designed to build a non-wired network inside the home or in small business institutions, both technologies were aimed at increasing the volume of data transmitted And received between the devices of this network and cover a larger area.[22].

2.5 BLUETOOTH

Currently, Bluetooth has become a technology for information exchange. A general standard for a private network describes how digital phones and digital devices communicate with

each other. The Bluetooth network synchronizes devices wirelessly, to the Internet, accesses electronic mail through a local computer connection, accesses the application, and automatically logs in to the digital network.

Radio frequencies connect devices that support Bluetooth communication technology at a distance of up to 10 meters. Using these characteristics, you can connect your computer or PDA to other digital devices, digital phones, wireless cameras, speakers, and printers in the Bluetooth network.

The Bluetooth standard can transmit data and voice within a distance of 10 meters, and this distance can be increased to more than 10 meters, which may reach 100 meters using the wireless communication mechanism [23], it can provide asymmetric data transmission at a speed of 723.2 kilobits per second, and symmetric data transmission 433.9 kbps per channel with 3 simultaneous audio channels at 64 kbps, one voice channel and one asynchronous data channel.

2.6 WIRELESS LOCAL AREA NETWORK TECHNOLOGIES

The most important feature of wireless communication technology from other communication technologies is the use of air as a means of data transmission. Radiofrequency technology is used to transmit radio signals by air on channels of a specific frequency, and WLAN can be invested in developing and increasing the productivity and efficiency of any environment with flexibility in transmitting information and making backup copies difficult. WLAN networks in recent years, not only in institutions but also used in open places such as squares, exhibitions, parks, airports, and apartments. There are a number of technologies used to transmit information or wireless communication, including radio frequency technology, which relies on electromagnetic waves in its work, and another technology that uses frequencies of less value than visible light, which is infrared radiation that transmits data at frequencies lower than the frequency of visible light in the electromagnetic spectrum (3 x 1014 kHz / 850-950 nm). The technology of transmitting data using infrared rays is used in television sets, and communication using infrared rays as an alternative to cables at a low cost [25].

Mobile phones are designed with an infrared port, which can be used to connect to a wireless network, using IrDA (Infrared Data Association) protocols and standards.

These standards include low-cost, low-power components designed in such a way that

infrared devices are connected together.

Most of today's computers are designed with an infrared receiver and IR protocol, obtaining the address of the device, exchanging information, transfer rates, termination, and fixing errors by address range, and asynchronous and sequential data transfer speeds of 115.2 kbps (Kbps) or it may reach 4 Mbps (up to 16 Mbps) and the transfer takes place between devices that work with infrared technology. The mechanism of work is summarized: all information is sent over a communication line from the control device to the receiver device, and any device available in it can The required characteristics can be a control unit, the user can transmit information by infrared over short distances where there is a visible line between the transmitter and receiver [21], the receiving devices are detected automatically, at a rate of 9600 bits per second, the request panel sends a connection to other devices] 23].

After specifying the address, the bit rate is adjusted by the command and receiving stations, and the parameters are linked to the shared devices during the initial transmission. Finally, the control unit confirms the connection and sends the data to the device from which it is requested to receive the data. Here, the communication between the devices is opened and the data is transferred under the control of The control panel, it is not possible to open a connection between a group of infrared devices at the same time, and this is done if we install these devices on one computer, for example, a desktop computer, through the connection is made in the printer and the phone, and the connection is broadcast to another network all at the same time [23].

One of the disadvantages of wireless LANs that use infrared technology is its short range of work and its inability to penetrate hard objects, that Bluetooth technology is an application of WPAN that works with infrared technology.

WLAN has become used to transfer information as an alternative to (LANs) and covers areas or capacities of the LANS at a lower cost. It soon replaced current computer networks. WLAN systems were originally designed to complement existing wired networks (LANs). But recently, WLAN systems have begun to replace wired networks. In addition, WLAN systems are widely used to extend existing networks. WLAN systems are widely used in both business environments and individual users due to their low cost and performance. Through LANs, it is possible to build local networks when providing a means of communication with access points to a wireless transmitter (AP) instead of expensive cables, and the user gains flexibility in dealing with data transmission,

The WLAN system is not completely wireless, but it reduces the need for cables.
Wired networks can be used to provide an environment in which wireless networks work together. Access points are used to extend the scope of WLAN work using wireless adapters and to increase the quality of data transmission.
WLAN-to-WLAN communication is provided by this two-port communication device with WLAN and Ethernet interfaces and a transceiver antenna connected to this device. Through packet filtering and dynamic address learning, access points operating on OSI Layer 2 data links serve up to 250 users [22] and users can be mobile or fixed as long as they are located within the coverage of the service network.
The main components of a wireless network are access points (APS) and terminals.
There are two methods for wireless networks, namely the ad hoc (peer-to-peer) method and the infrastructure method, where users are connected directly to the infrastructure. In the case of access points, users interact with the access point, which acts as a bridge between users. These points connect to the Ethernet backbone network and connect to all wireless clients in the range (cells). The network traffic flow is controlled by the coverage area manager. Access points can be installed in corporate offices or homes in shops, bus stations, and public places, and from the wireless network components are the network interface card (NIC) of the wireless device that has the function Automatic Frequency Scan It can detect the arrival of a WLAN signal to the wireless device. Once the correct frequency channel is found on the network card, it will start the communication between the access point and the wireless device. If the cell does not provide enough coverage, you can add cells to extend the coverage. in this environment. This is called an "extended service area". This WLAN system consists of two cells. This wireless LAN is multicellular. A WLAN system is not completely wireless. Because. System Access Points (APS) must be connected to the network via a cable-like infrastructure that provides broadband or fiber-optic access services. Because. The WLAN system is not completely wireless, but it reduces the need for cables. When used from one device to another, no wires are required. Therefore, in this case, the complete configuration of the wireless network is performed.

2.7 WORKING PRINCIPLES OF WIRELESS LOCAL AREA NETWORKS SYSTEMS

The wireless phone is an easy and flexible communication system designed as an alternative to wired communication and the technology used in transmitting and exchanging information is a Radio Frequency (RF). Depending on the environment of use, this technology can provide services for areas with a radius of 25-100 meters indoors, and about 200 meters, this circle's radius includes the outdoor space. One of the advantages of radio frequencies is that it passes through solid objects if they are high (2.4 GHz and 5 GHz) used by WLAN systems. This feature is important in internal WLAN applications, because the line of sight is not possible due to concrete or furniture, because of solid objects, the signal is weakened, and this leads to a shortening of the time interval because the energy used to cover the distance is limited to (100 megawatts), so it is not possible to increase the distance due to the difficulty of increasing the power. When establishing this network, the location of the access points and the material environment of these points must be taken into account. Usually, an external antenna or an access point is installed connected with another access point in high locations, and this procedure is necessary to expand the area of the area to be covered, and then place other access points within this range Any digital device that has wireless access and a network connection can be used. As for digital devices that do not have wireless technology, they can connect to the network through an external wireless network adapter or an international memory card (PCMCIA)[26], or an ISA/PCI card if the digital devices are desktop computers. The access point is the one that connects customers with each other in the wireless network within the basic coverage area provided by the radio frequency, where there is no direct communication between customers, but rather through the access point [26].

2.8 DISADVANTAGES OF WIRELESS LAN SYSTEMS

It is not shown above that WLAN technology has many advantages, and also there are some disadvantages such as product selection, standardization of measurements, and cost-sharing, but these problems were solved in previous years, but not radical solutions and the most important negative aspect of WiFi technology is security Despite the use of intrusion prevention systems for unauthorized intrusions, they are more sensitive to attacks than wired systems.

Wireless systems use radio frequencies to provide communication service, but they cannot prevent the frequencies from interfering with each other. Usually, WLAN systems use a security mechanism known as WEP, which is an acronym for Wired Equivalent Privacy.

In these systems, data transmitted between users (sender and recipient) is secured by 64 or 128-bit fixed encryption [27] [28].

WPA uses a set of security keys that change over time and are proportional to the amount of data.

There is another defect in Wi-Fi technology, which is (interference). Usually, the ISM band is used by WLAN systems, so it is vulnerable to interference. WLAN systems interfere with each other due to stopping the radio frequency or bypassing lines in public places.

One of the shortcomings of this technology is the short communication distance, due to the narrow bandwidth used, which is about 100 meters [31] [32] in open areas with the possibility of increasing this distance in the case of using reinforced antennas. As for the presence of many physical obstacles such as walls and Furniture, the distance is less than that.

Despite the services provided by these systems, technical challenges are generated with them, and they are classified as defects, which require their development and solution to ensure a safe and long-life service, as the devices used to broadcast this service need maintenance, represented by the small life of the battery, for example.

2.9 WHERE WIRELESS LOCAL AREA NETWORK SYSTEMS ARE USED

The presence of wireless LANs in all aspects of life has become important, as it can be used in government offices, private sectors, education, and health [33], and in the field of education, it is increasingly being used, as the teaching staff and students can access wireless Internet inside the school and thus information can be exchanged anywhere A time that suits them, with WLAN technology, many channels can work on one line and thus it becomes easy to connect to the Internet from places such as home, workplace, malls, or airports And this is without the need to pull cables and get rid of their financial cost. I mean, if an organization is distributed to work on multiple buildings or offices, the WLAN provides a connection service between these offices at a cheap and fast cost [34] [35]. The WLAN also

served the field of health, where the medical committee or the specialist doctor can access the patient's information and identify his medical history through networked distributed systems that archive the patient's information and allow the authorized to browse it.

2.10 WIRELESS LOCAL AREA NETWORKS STANDARDS

At the beginning of the nineties, workplaces used WLAN systems for the first time, but their use was little due to a number of problems, including low data rate, incompatibility of devices, high material cost, and the lack of awareness of users of the mechanism of this technology and how to employ it in the service of their work environment. Since the 2000s, there has been A clear growth in the adoption of WLAN technology as a result of standardization, where different equipment of different brands can be used in building the network. This led to improved performance and reduced prices, thus increasing the demand by users to use them. Standards have been identified that match the industry standards that gain confidence among manufacturers and harmony in performance.

2.11 CYBER SECURITY

In fact, cyber security is not only concerned with information security, but it is more comprehensive. It is concerned with providing security for information and also provides security for the safety of users (the human factor) as well as assets such as electronic devices that are also a target for digital attacks.

(ITU) is an acronym for the International Telecommunication Union, which defines cybersecurity as: A set of security policies, tools, safeguards, security concepts, risk management methods, guidelines, procedures, best practices, training, and techniques that can be used to protect users, their data, and all assets that interact directly or indirectly with the Internet. environment [37].

All computers connected to a communications network, personnel, apps, infrastructure, communications systems, and data stored and transported in the cyber environment are all considered enterprise assets or user assets [37].

The technology revolution has affected practical and social life, storing information and organizing tasks in all aspects of life through designing applications or communication

systems that meet the user's needs and in a way that preserves data with the ability to process, retrieve and display data.

Through communication systems, the world has become a small village, and the need for security technologies has emerged to protect users and their data from attacks in cyberspace. These technologies are known as cybersecurity.

More than 50 countries have published a strategic document outlining the official position on cyberspace, cybercrime, and/or cybersecurity[38].

Mitnick and Simon describe information security as a process consisting of a set of actions, not a product or technology [38].

This security may require the provision of special products and devices through which security is applied to maintain the integrity and availability of the information.

Intrusion detection systems are considered part of cyber security because they counteract denial-of-service attacks (DoS)that target networks.

2.12 DISTRIBUTED DENIAL-0F-SERVICE ATTACK (DDOS)

DDoS is an acronym for distributed denial of service attack (DDoS attack), which targets web servers that use the Hypertext Transfer Protocol (HTTP) to perform DDOS attacks.

This attack is difficult to detect because it is able to simulate the authorized request to access servers and the HTTP protocol works to receive and respond to the request, this allows a connection between the client and the web server.

Attacks deplete network resources such as bandwidth, CPU, and memory.

The lower the network bandwidth, the greater the chance of an attack by disrupting the service.

There are different methods for carrying out DOS attacks, and in general, the attack may be an activity by one or more people to disrupt the services of a host connected to the network, which means either the inability of legitimate users to access the server or the expected bad service.

Network attacks fall into categories, including ICMP which is performed by sending a large number of fake IP packets, SYN which occurs by sending a large number of TCP packets, and DDOS which is more harm-effective, and is an acronym for distributed denial-of-service attack.

The DDOS attack occurs through a group of digital devices connected to the Distributed Internet in geographical locations far apart and controlled by the hacker, This procedure is known as a botnet.

The hacked party is called a Botmaster, and the digital devices that make up the botnet are called botnets.

The Botmaster has complete control over the botnet through network protocols such as HTTP (Hypertext Transfer Protocol) and IRC (InternetRelayChat). The attacker uses remote control programs (C&C) such as Zeus, Phatbot, and SpyEye.

The strength of a DDoS attack is directly proportional to The size of the botnet.

The attack begins with user requests to the network via an HTTP GET request to the servers, at this point, a TCP connection must be opened by the legitimate users to get a successful response from the servers to the request sent, in the next stage a list of requests is created to set a thread for the requested service Where this list keeps all the requests received without distinguishing whether these requests are legitimate or illegitimate, then the requests are processed from the request list and answered, where the server sends a response to HTTP.

In the event of an attack, this means that the queue is filled with requests. There is no space to store the requests of legitimate users. The server does not respond to them because it did not receive the request in the first place [39].

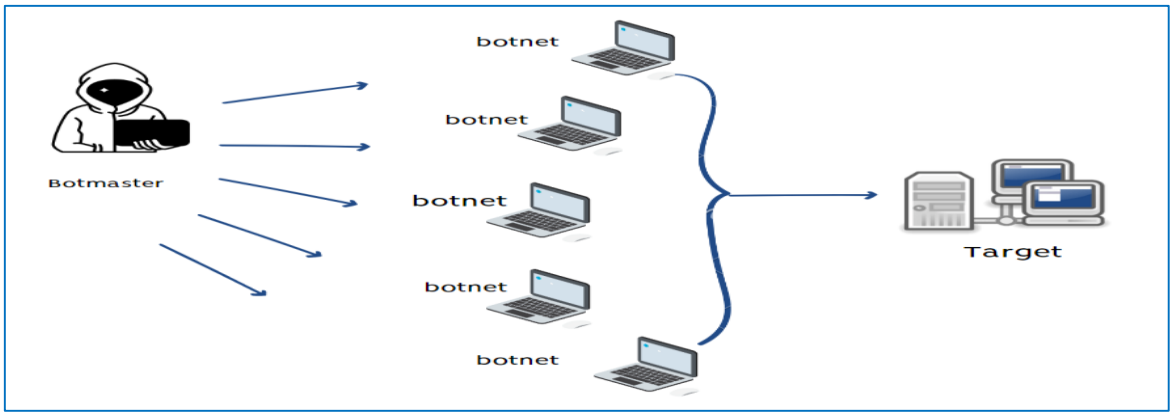

Figure 2.1: DDoS Attack.

2.12.1 Prevention

Security methods should be provided to prevent attacks by designing protection systems that only allow legitimate requests to pass through, block illegal requests, and increase network bandwidth as a precaution. It does not replace attack detection systems.

It is also recommended that network services be distributed over a number of independent servers so that if one of them is attacked, the rest of the servers will remain functional and meet the requests of users [39].

The systems that detect intrusion are known as intrusion detection systems, and these systems are designed using machine learning technology by creating a statistical model that can detect anomalies in requests sent to servers.

2.12.2 Intrusion Detection System

It is necessary to detect malicious attacks targeting a communication system or network, and this is done by special intrusion detection systems, which are systems that scan the network for the purpose of detecting malicious activity, and because the digital network is important in daily transactions, it has become necessary to secure the network from attacks.

Through machine learning technology, models were built to detect intrusion, and the efficiency of these systems is measured by the accuracy of their detection of attacks, and the efficiency of these systems increases as the false alarms decrease. Intrusion detection systems depend mainly on the analysis of network traffic data.

With the development of communications and the dependence of institutions and individuals on the Internet to organize their lives, and saboteurs' exploitation of network and host loopholes for the purpose of personal profit, whether for the purpose of financial gains, revenge or to change policies, which led to the expansion of attacks in recent years, and thus it became necessary to create the elements of security to deter attacks.

It was intrusion detection systems had the biggest role in confronting saboteurs, because well-known defense methods such as the firewall, which is the first line of defense for the network, and updating programs are not enough to win the war between attackers, wireless networks, and the host (computer, mobile, etc.).

The essence of the work of the intrusion detection system is to classify the network traffic and determine whether it is normal or abnormal while reducing false alarms, and one of the most prominent problems faced by these systems is their inefficiency in detecting the unknown network attack, and the high rate of false alarm [69]

Intrusion can be defined as an attempt to breach the confidentiality, data integrity, availability of information, and control over the host or network resources and services by unauthorized parties to access the host or network, or by legitimate users seeking additional privileges that are not within their authority, or Users who are authorized to access but abuse their privileges and harm the network or host.

intrusion detection systems are divided into

i. host-based systems (HIDS), monitor system logs, disk resources, and file systems, meaning they monitor only incoming and outgoing data packets from the device [69].

ii. network-based intrusion detection systems (NIDS) it is work on Monitoring data or requests that pass through the network to monitor attacks, distinguish legitimate requests, prevent illegal requests, and consider them an attack, and an alarm is issued to warn, which means tracking traffic to and from all devices on the network[69].

iii. Protocol Based Intrusion Detection System (PIDS): Protocol Dependent Intrusion Detection System (PIDS) consists of a system or proxy that is constantly present at the front end of the server and controls and interprets the protocol between the user/machine and the server as it tries to secure the web server by monitoring the HTTPS protocol flow.

iv. Application Protocol-Based Intrusion Detection System (APIDS): An (APIDS) is a system that is generally located within a group of servers and identifies intrusions by interpreting and monitoring communications on application-specific protocols. web server.

These systems are installed in a network in a strategic location to perform the tasks for which they are designed and with the best results, as they can be placed between the firewall and the network or placed behind the firewall [70].

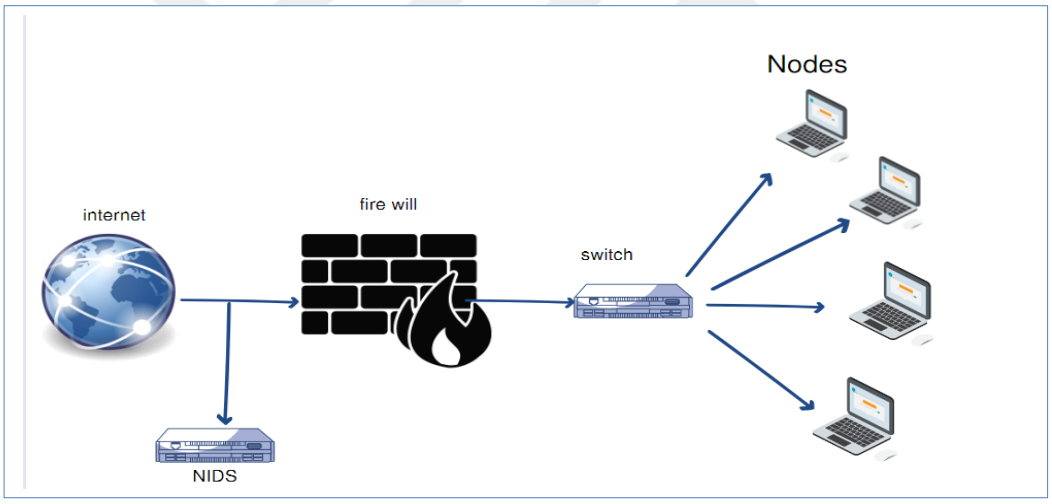

Figure 2.2: NIDS Between The Outside Network And The Firewall.

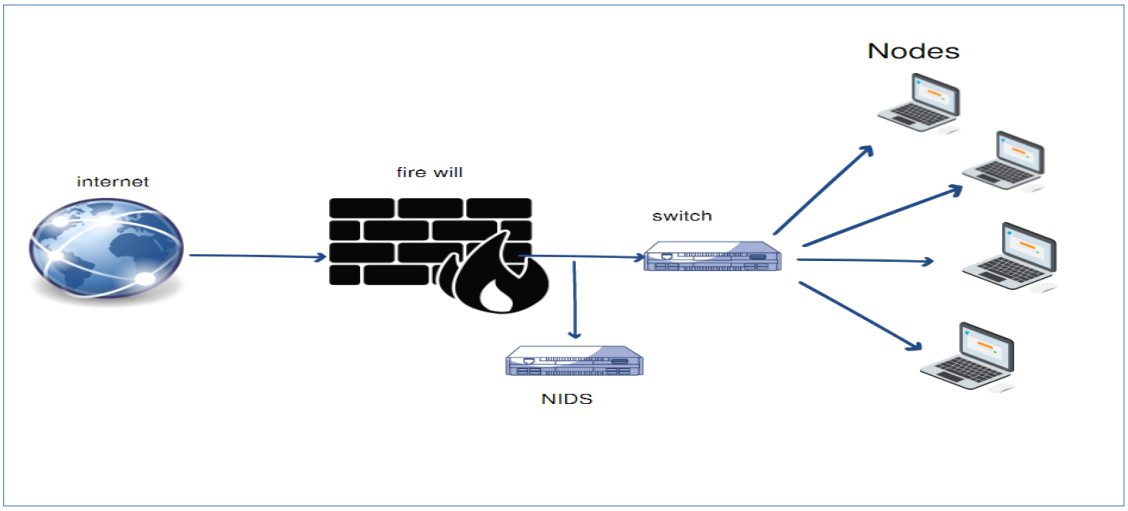

figure 2.3: Nids Between The Inside Network And The Firewall

the utilization of intrusion detection systems is in detecting intrusion at the moment it occurs and not after penetrating the network.

It is better not to specify the mission of IDS just the warn of attacks, but work to take any action to reduce the impact of the attack.

Also, the intrusion detection system must work automatically and continuously without human intervention, as is the case in the firewall where The network administrators manually specify the IP or the site allowed to pass, and this port may become a source of attack when it is controlled by hackers,

This danger can only be distinguished at the time of its occurrence [70].

There are two ways in which intrusion is detected:

i. SBS signature-based intrusion detection systems such as Snort are pattern recognition techniques by maintaining a database of previously known attack signatures for comparison with newly analyzed data.The alarm is issued when similarities have been identified This type of intrusion detection system is also called (an abuse-based system), this type is good for blocking only known attacks as it continues to be effective based on regular updates of signature databases, this means an increase in the size of the base And therefore the performance of the system engine decreases, so a group of intrusion detection engines is distributed on systems that contain a group of processors, and one of its most prominent negatives is that it

cannot detect a zero-day attack because it does not have the signature of this attack [70].

ii. ABS Abnormal Intrusion Detection Systems Also known as (Anomaly-Based Intrusion Detection Systems), These systems determine the network behavior and compare it with the previous behavior. In the event of a mismatch, the network is considered to have been attacked, and the network behavior is determined by network administrators this is what was addressed in the International Journal of Computer Applications (0975 - 8887) Volume 28 - 2011. It is considered the most important stage because it determines the efficiency of the IDS engine in penetrating protocols at all levels These systems are characterized by their ability to respond to unknown or instantaneous attacks, but there are disadvantages represented in the continuous change of network behavior that makes it difficult to determine the normal behavior of the network, for that the rate of false alarm for these systems is higher than others[71].

3. MATERIAL AND METHODS

3.1 INTRODUCTION

In this chapter, we review a model that is designed for intrusion detection using machine learning, and deep learning and which consists of several stages.

i. The stage of reading the UNSW- NB15 data set and preparing it for training.

ii. The stage of selecting the best features using Machine learning of the

iii. ExtraTreesClassifier algorithm and random forest algorithm.

iv. Data splitting for training and testing

v. Training and testing phase using the deep learning of the (Dennis) algorithm that works with DNN technology consisting of 3 hidden layers with an input layer and an output layer.

Chart (3.1) shows the model design algorithm that achieved 99.85% accuracy.

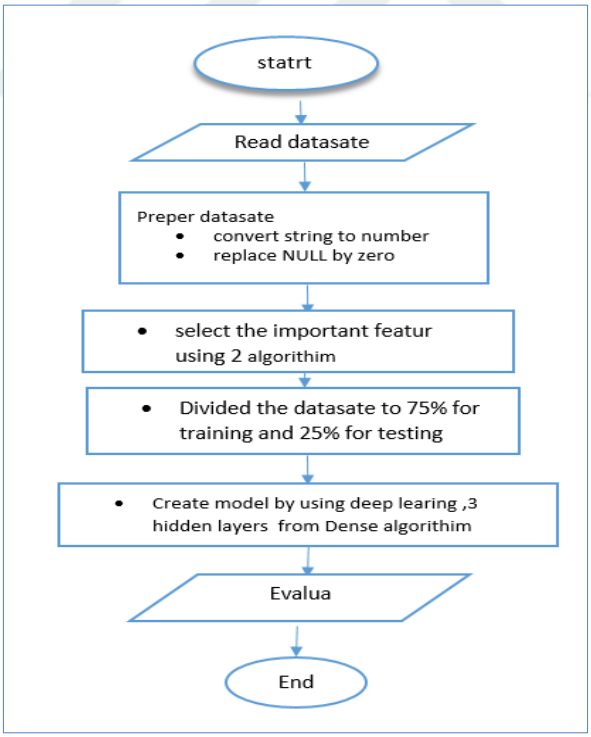

Chart 3.1: Model Algorithm

3.2 PREPARE THE DATASET

Unws_nb15 is the latest database used in intrusion detection systems and contains 9 types of attacks with real activity data arranged, including Survey, Shellcode, Fuzzers, Worms, Backdoors, Generic and DoS Exploits by 49 features (columns) Divided into six sections as follows (flow features, additional features, content features, basic features, time features) and 2.5 million records (rows) and has been used in many specialized studies in intrusion detection systems, the data set was created This is in 2015 at the University of New South Wales in the Cyber Security Labs, using the IXIA tool to collect network traffic and determine whether it is normal traffic or no [12].

It is better than previous databases KDDCUP99, KDD98, and NSLKDD It contains recent attacks[2] .and in In this research, these databases were prepared by placing zero values in the empty fields and replacing the text data with numeric data. These databases were divided into 70% for the purpose of training and 30% for the presentation of the test.

3.3 MACHINE LEARNING (ML)

It is one of the branches of artificial intelligence technology that develops the algorithms' mechanism of action and makes them predict and make decisions based on the available data. It differs from algorithm models that work steadily and conditionally with predefined inputs and outputs. It is able to improve its performance without human intervention. With its ability to deal with huge data and thus contribute to the production of highly efficient applications in terms of accuracy of results.

The main goal of machine learning is to design algorithms that can deal with huge data and process it in record time based on statistical analysis to obtain outputs within the prediction options and with high accuracy.

There are two approaches to designing these algorithms, which are supervised learning and unsupervised learning.

3.3.1 Supervised Learning

This type of machine learning requires that the training data contain correct and reliable inputs and outputs, as the machine trains and over time develops and learns to predict the outputs. The loss function acts as a criterion for the accuracy of the algorithm. Machine learning involves specifying the inputs and outputs of the algorithm for training, then applying what it learned during the training phase to new data. In this case, human assistance is required to prepare the data while describing the outputs, for as creating a database with a thousand photographs of a cat and a deer. It is a waste of time and effort since, after assembly, the image must be specified by default with the numbers 1 for cats and 0 for deer[72].

3.3.2 Unsupervised Learning

This type does not require specifying outputs for all inputs in the database, as it is used to analyze exploratory data because the algorithm works without human assistance in analyzing and diagnosing data. It is also used in consumer segmentation and image recognition because of its ability to identify similarities and differences.

Until it learns, the machine relies on repetition techniques. Deep learning is an approach that does not require training the algorithm on inputs and outputs, allowing it to identify if an image is of a cat or a deer without human interaction.

Because of this, it is used in processing the most complex data. It is most famously used in detecting intrusion in Internet networks and diagnosing e-mail as fraudulent (spam), as well as using social media applications in digital marketing by predicting user needs and selecting relevant advertisements for them[72].

3.4 CHOOSING IMPORTANT FEATURE

At this stage, the size of the data is reduced and the features that affect the performance of the model are identified to increase the efficiency of the model and reduce the training time by removing unimportant features and selecting the best.

Because the basis for the success of any model is a pure database rich in information related to the subject of the research [73].

3.4.1 Extra Trees Classifier Algorithm

It is an updated type of random forest algorithm with low computational cost because it does not depend on building decision trees on specific arithmetic equations and this requires processing time, but additional trees use the full data set to train decision trees[73], choose values randomly to split features and create sub-nodes and then depend on majority vote To improve classifier and arithmetic mean prediction in regression problems.[74].

3.4.2 Random Forest

A collective learning method, this forest builds decision trees by using arithmetic operations to divide features and create sub-nodes known as entropy Eq(3.1), and information gain Eq(3.2).

Where in the first step a database is created cloned from the main base known as bootstrapped to reduce the variance of the estimated predictive function and the decision tree is extracted from it and the process is repeated until the creation of the forest.Forest creation[74].

depend on majority vote To improve classifier and arithmetic mean prediction in regression problems.[74].

$$\mathrm{E}ntropy(s) = -p + log(p+) - p - \log(p-) \tag{3.1}$$

$$\mathrm{E}ntropy(s) = \sum_{i=1}^{c} - p_i \, log_c \, (p_i) \tag{3.2}$$

$$Gain\ (\mathrm{R,L}) = \mathrm{E(R)} - \sum_{i=1}^{k} p_i \ E(\mathrm{R,L}) \tag{3.3}$$

In the example below, we explain the importance of Entropy and how it is calculated

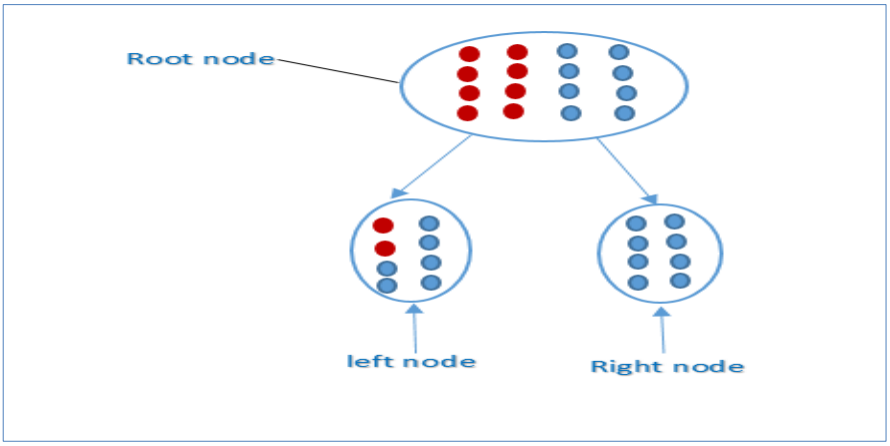

Figure 3.1: An Example Showing How To Calculate Entropy.

It is an important factor that controls the behavior of the decision tree and the distribution of data, and it is calculated for each node in the tree, and its value ranges between zero and one, and the information gain equation depends on it, where the nodes (the attribute) are divided, which has a higher information acquisition value and this leads to building a decision tree.

i. Entropy 0: means that the node is pure and contains only one type of data, for example, all in the node are males, only females, or only infected..etc, according to the categories of the data being processed[75].The number of blue balls is divided by the total number of balls, and it equals (8/16) there is also a possibility of red balls which equals (8/16).The result of the entropy of the tree nodes above is as follows

ii. Entropy 1: means the node is impure and contains an equal number of data characteristics, for example, the node contains 5 males and 5 females, and this means that the node needs to be divided[75].

iii. Entropy between zero and 1: and this also means that the node is not pure and contains data for more than one classification and in different proportions, for example (3 males and 6 females), and this leads to its division as well until reaching nodes with a low value of Entropy close to zero[75]. in the equation of Entropy Eq(3.1) the symbol (p) represents the probability of the data inside the node, for example, the root node in the above-numbered figure (3.1) has probabilities, which are

$$Entropy(root\ \ node) = -\frac{8}{16} + log_2\left(\frac{8}{16}\right) - \frac{8}{16} - log_2\left(\frac{8}{16}\right) = 1 \qquad (3.4)$$

$$Entropy(left\ node) = -\frac{6}{8} + log_2\left(\frac{6}{8}\right) - \frac{2}{8} - log_2\left(\frac{2}{8}\right) = 0.81 \quad (3.5)$$

$$Entropy(right\ node) = -\frac{8}{8} + log_2\left(\frac{8}{8}\right) - \frac{0}{8} - log_2\left(\frac{0}{8}\right) = 0 \quad (3.6)$$

If the node contains more than two classes, as shown in the numbered Figure (3.2), where it contains 4 colors, each color represents a class, then the entropy is calculated using the following equation Eq(3.4)

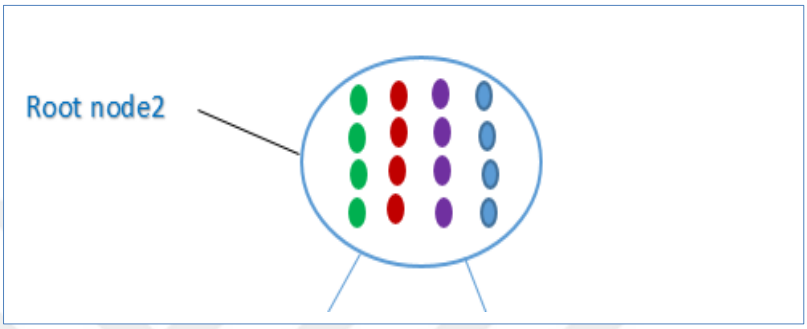

Figure 3.2: An Example Showing How A Node Contains More Than Two Classes.

$$Entropy(nod) = \sum_{i=1}^{k} p_i\ log_k\ (p_i) \quad (3.7)$$

The k represents the number of items and in our example shown in the numbered Figure (3.2)

$$Entropy(root\ node2) = \left(-\frac{5}{16} + log_4\left(\frac{5}{16}\right)\right) - \left(\frac{3}{16} - log_4\left(\frac{3}{16}\right)\right) - \left(\frac{2}{16} - log_4\left(\frac{2}{16}\right)\right) - \left(\frac{6}{16} - log_4\left(\frac{6}{16}\right)\right) \quad (3.8)$$

value of k = 4, and when this equation is applied to root nod 2 it is as follows

A tree whose entropy is close to zero or zero has the best advantage that can be used for prediction, but the entropy is not enough to evaluate, but we need another equation known as the information Gain Eq(3.3), which calculates the ratio of branch data compared to the tree root data and its formula is as follows, where R represents the root node, and L represents the terminal node (leaf tree) and P is the probability of the data example in the node (4 males and 5 females) The greater its value, the better, unlike entropy, the closer it is to zero, the

better.

The value of the Information gained in our example is calculated as follows.

$$G(R, L) = 1 - \left(\frac{8}{16}\right) * 0.81 - \frac{8}{16} * 0 = 0.595 \quad (3.9)$$

3.4.3 Bias

The phenomenon of bias skews an algorithm's output in favor of or against a certain idea.

The machine learning model itself experiences bias as a result of false assumptions made throughout the ML process.

Technically, bias is the discrepancy between the average model prediction and the actual data, or the difference between the model's predicted values and the actual or anticipated values. It also describes how well the model fits the training set of data .A more biased model wouldn't precisely fit the data set.The training data set will be closely matched by a low-bias model.

3.4.4 Variance

Bias' complete opposite is variation. Our model is given a limited number of opportunities to "view" the data during training in order to look for patterns. Insufficient time spent working with the data will result in bias because patterns won't be discovered. On the other hand, if our model is given access to the data too frequently, it will only be able to train very well for that data. The majority of patterns in the data will be captured, but it will also learn from extraneous data or noise that is there.

Variance can be thought of as the model's sensitivity to changes in the data. From the noise, our model might learn. This will lead our model to value unimportant features highly.

Alternatively, it might be the accuracy gap between machine learning during the testing phase and accuracy during the training phase, where new data could not have the exact same features and the model won't be able to predict it very well. Overfitting is the term for this.

3.4.5 Underfitting & Overfitting

The model's inability to match the data is referred to as underfitting or overfitting.

The accuracy of a model's fitting determines whether it will deliver precise forecasts from a given data collection.

When the model is unable to match the input data to the target data, underfitting occurs. This occurs when the model is underwhelming with the training dataset and is not complicated enough to match all the available data.When a model tries to match erroneous data, it is said to overfit.Dealing with extremely complex models so that the model matches all selected data points in an approximate manner. The result is an excellent performance with the training data.However, the model cannot generalize the data in the selection phase in order to predict accurately.

3.5 WHAT IS DEEP LEARNING

It is a method used in artificial intelligence and machine learning, where professionals attempt to mimic how people learn.

It varies from machine learning in that it attempts to find the most crucial aspects on its own, rather than relying on experts to do so. As a result, the deep learning algorithm takes a while to train and performs quickly during testing. In contrast to machine learning, where test time is inversely proportional to data amount.

Deep learning is an important technology in the data space, as it enters the fields of predictive modeling and statistics. It is useful for dealing with large data, as it makes the process of analyzing and interpreting this data easy and fast.

Deep learning can be described as a hierarchy of procedures in each layer that becomes increasingly complex.

The learning rate is one of the most important factors in this technique because it controls the number of changes that the model will go through depending on the errors that result from every time the weights of the model are changed. If the learning rates are small, the training process will be long and may fail to achieve the desired results.

Examples of the uses of deep learning are building applications for big data analytics, including NLP, stock market trading signals, language translation and text generation that matches the grammar of a specific language, network security, medical diagnosis, and image recognition as in the medical field Where deep learning models are developed to detect malignant tumors.

This is done through a set of algorithms that, thanks to its multi-layered structure, enable it to deal with huge data known as artificial neural networks (ANN). It is a sophisticated type of machine learning. the capabilities of neural networks in machine learning, making them

perform tasks such as regression, clustering, and classification[76]. In general, a neural network can do the work of machine learning algorithms, but machine learning algorithms cannot do the work of neural networks. This simple neuron consists of an input layer and output layer Complex neurons consist of an input and an output layer, and in the middle of them are hidden layers that may be one or more. Each layer contains a set of nodes, which is a graph of scalar values, each node is linked to the other node in the next layer by a "weight" which is a numerical value represented in the diagrams by the arrow connected between any two nodes [77].

3.5.1 Deep Neural Network

Deep learning, which is the latest technology in the field of machine learning, depends in its method on the hierarchy of concepts, and thus enables the machine to deduce information due to its ability to collect knowledge from experience, and the knowledge-gathering mechanism is devoid of human intervention because it is self-learning.

A DNN is a multi-layered artificial neural network and in order to understand the work of a DNN, you must understand the mechanism of the artificial neural network(ANN)[78].

3.5.2 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) belong to the technology of artificial intelligence.

ANN is a computer system aimed at realizing the capabilities of the human brain without help when programming these systems, traditional methods are often inadequate. In fact, he was inspired by human and animal biological neural networks when designing artificial neural networks. Electrical signals are passed from one neuron to another in order for neurons in the human brain to transmit information figure (3.3) explains the content of a neural cell. The transmitted signal can be stored in the brain (memory).

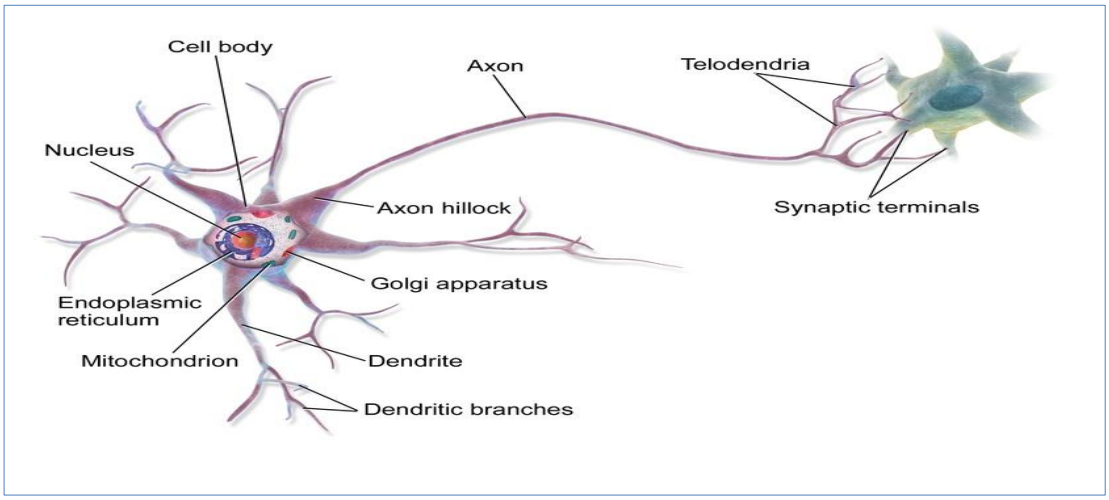

figure 3.3: Human Brain Neuron.

In other words, the memory that a computer uses to store it mimics the brain of a biological system. The purpose is the same, but the mechanism used is different. In short, ANN is considered a computer algorithm for data processing and is designed for one-off events.

This simple neuron consists of an input layer and an output layer Complex neurons consist of an input and an output layer, and in the middle of them are hidden layers that may be one or more Figure (3.4).

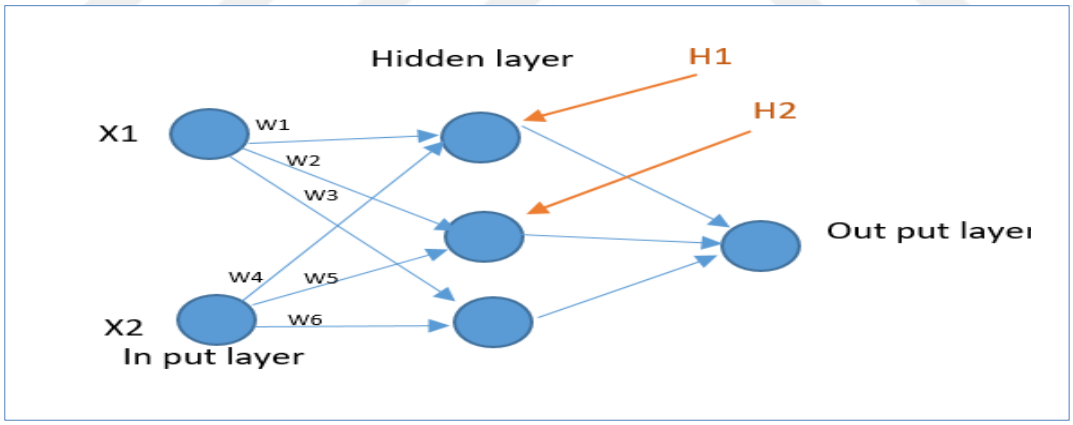

Figure 3.4: Feedforward ANN.

Each layer contains a set of nodes, which is a graph of scalar values, each node is linked to the other node in the next layer by a "weight" which is a numerical value represented in the diagrams by the arrow connected between any two nodes and is generated randomly, and each arrow has a different value than Another arrow and also generates a bias value (b) which is uniform for each hidden layer[79].

Data is sent from the input layer to perform a set of operations to determine its state and whether it is allowed to move to the next layer or not, this action is known as the activation function.

In short, each neuron receives a double copy of the input data (features) and then random weights are added to each layer with a fixed bias value. In the neural network, after reaching the output layer and obtaining the final results, the loss function is calculated, which represents the input in the input layer against the output, and then backpropagation is performed, at which point the weights are adjusted to reduce losses and retrain.

There are many types of these networks and their classification depends on many factors such as data flow, design structure, neuron density, and deep activation filters like

i. Perceptron: It is the best and simplest type of neuron composed of a single layer with four main parameters) value of the weight, input, bias, and activation function [80].

ii. convolutional neural network: It is the most famous type of artificial neural network, famous for its ability to process large data, and it was named convolutional because of the linear arithmetic equation between the data matrix. It consists of a number of layers, including the convolutional layer, the pooling layer, and the nonlinear layer[78].

iii. LSTM - long-term memory: It is a distinct type of artificial neural network that can learn itself through long-term consequences by each normal recurring node is replaced by a memory cell and was announced in 1997 by Hochreiter & Schmidhuber, and the basis of its work is to remember or retain information for long periods of time [81].

We should learn some vocabulary that we need in designing a neural network model for machine learning

i. epochs: It is training the neural network for one session and for all data.

ii. Batch: Epochs consist of one batch or more, where the data to be trained is divided into a number of groups to address the problem of the small size of

memory that is not commensurate with the size of the data to be trained. epochs from five batches.

iii. education rate: This parameter specifies the size of the weight updates to reduce the loss function in the network.

Each node contains an Activation function, which is necessary to determine the output value within a specified range. For example, if the Artificial Neural Network is for the purpose of a two-dimensional Classification process, the Activation function outputs the values of (0,1) or (1,-1) depending on the function used. There is a linear activation function and There is a non-linear activation function [81].

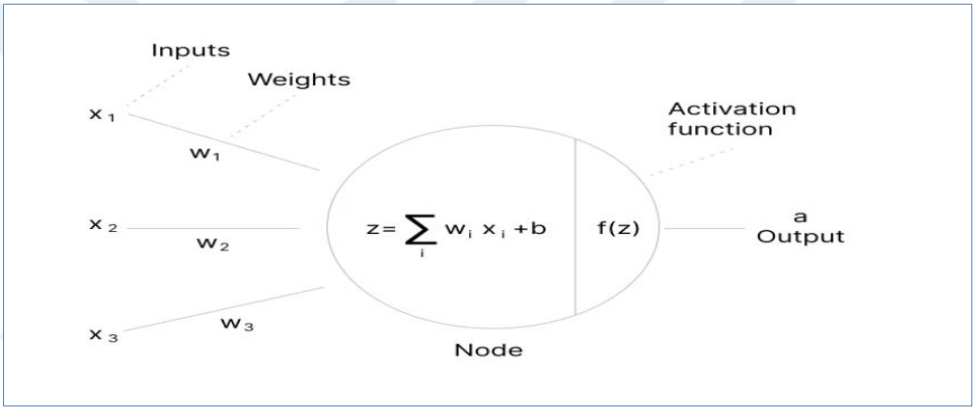

Figure 3.5: Diagram Showing The Components Of An Node In Artificial Neural.

i. Linear activation function

The value equation of the linear activation function based on our example shown in Figure 3.6 is Eq(3.9) and Eq(3.10), and Figure (3.3) shows the graph of a linear equation

$$\text{NODE= (x1w1+x2w2+x3w3)+b} \tag{3.10}$$

$$\sum_{i=1}^{m} wixi + bias = w1x1 + w2x2 + w3x3 + bias \tag{3.11}$$

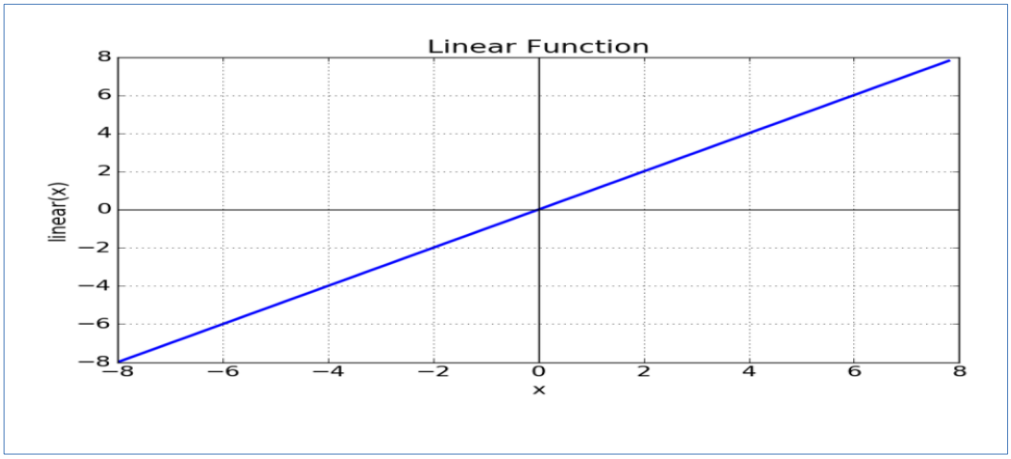

Figure 3.6: It Shows The Graph Of a linear Equation[15]

ii. Non-Linear activation function

Most modern neural networks use non-linear activation functions, which allow the mathematical model to discover complex relationships between network input and output, especially in the presence of multi-dimensional data, especially images, videos, and others[81].The most common nonlinear activation function used in neurons is.

Follow Sigmoid or Follow Logistic Activation It consists of the following equation Eq(3.12)

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.12}$$

The sigmoid function is characterized by the fact that it gives a flexible increment in the values, and the output values are always between 0 and 1. But one of the problems is that it cannot process a large value of the input, it is considered a sigmoid function costly in terms of an arithmetic point of view[9].

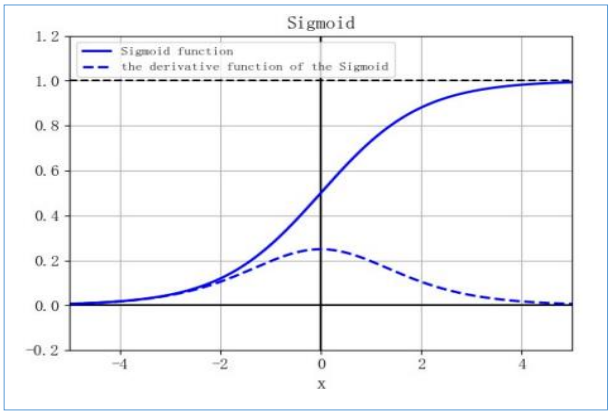

Figure 3.7: It Shows The Graph Of Non_Linear Equation(Sigmoid) [15].

The Softmax Activation Function It generates a set of probabilities in the output, it is widely used in neural networks that predict a set of Multiclass classification classes, where the function generates a probability value for each class, that is, if Z0 = 0.2 and Z1 = 0.5, that is, the first element in the array has the probability of being of the class The first is 0.2 while the probability of being of the second class is 0.5[81], From Eq(3.13) we extract the results of the activation function.

$$\delta(z)_{i=}\frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}}\ foe\ i = 1, \dots, k\ and\ z = (z_1, \dots, z_k) \in R^K \qquad (3.13)$$

Rectified Linear Units (ReLU) are the most used today in neural networks, especially in deep learning. It takes a value of zero if the value of the income is less than zero, or keeps the value of the income itself if it is greater than zero, so the function values are confined between 0 and Infinity. The function is differentiable and is a monotonic function. One of the negatives of the correction function is that it converts all negative values to zero, that is, in the event that there are different negative values in the training, the follower looks at it in the same way, which is zero, which may cause the loss of important information necessary for training the difficult network[82].

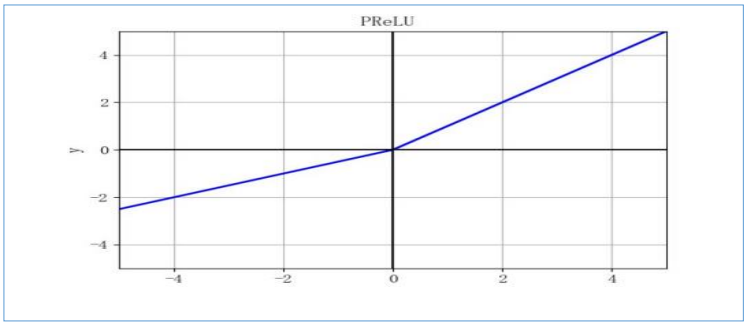

Figure 3.8: Shows The Diagram Of Rectified Linear Units (Relu).

3.5.3 Denes Algorithm

Also known as a fully connected neural network (FC) It is one of the types of artificial neural networks, in which the nodes of each layer are closely connected with all the nodes in the previous layer, and it is more used in building neural networks.

Suppose we have 5 input neurons (x1, x2, x3, x4, x5) and 3 output neurons (y1, y2, y3) in the FC layer Here we will generate a two-dimensional matrix of weights and its size consists of 5 rows and three columns. In a dense neural network, the input matrix is multiplied by the weight matrix In our example, the result is an array consisting of row 1 containing 3 columns[83]. This process can be summarized in the following equation Eq(3.14) and Fıger Explains the application of the equation.

$$y_j = \sum_{i=1}^{n} w_j \, x_i \qquad (3.14)$$

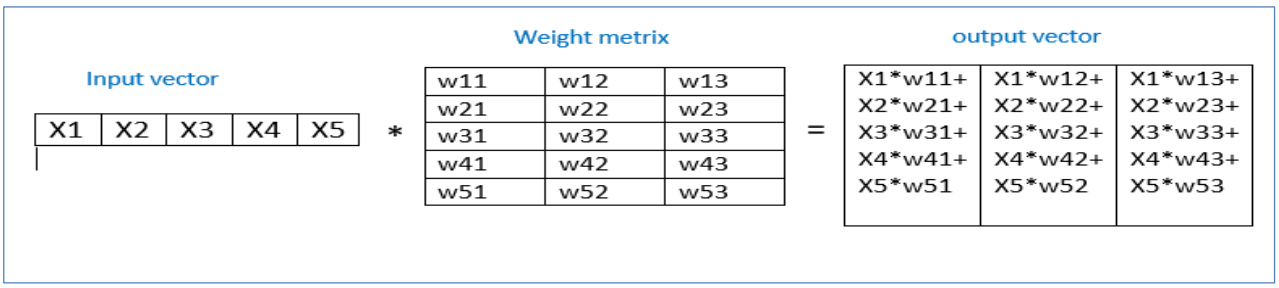

Figure 3.9: Application of The Equation

4. RESULTS

4.1 INTRODUCTION

In this chapter, we review the working environment of the proposed model, the type of devices used for implementation, the stages of work, and the results obtained The results for all stages are arranged based on their appearance in the third chapter.

4.2 SYSTEM REQUIREMENT

As it is known that the algorithms of deep learning and machine learning are suitable for processing large data, and this means that we need devices with high specifications to obtain satisfactory results.

In this research, devices and systems were used as shown in table (4.1).

Table 4.1: Environment specifications for the proposed system.

• Operating system(os)	Windows 11
• Hardware	
1. Central Processing Unit (CPU)	Intel(R) Core(TM) i7-1165G7 CPU.
2. RAM	8 GB.
3. Hard Disk	512 GB
• Programming Language	Python 3
• Work environment	Anaconda

4.3 ACTION STEPS ARE SUMMARIZED

i. get the database (UNSW_NB15) using the Pandas library.
ii. Prepare the database and make sure that it does not contain empty fields or text data using the Sklearn library.
iii. Filter the data and select the best features by Sklearn library by calling the Extra Trees Classifier algorithm.
iv. select the best data again using the random forest algorithm. Dividing the base into 70% training and 30% testing.
v. Training the model by calling the denes deep Sklearning algorithm by sk_learn library

which consists of 3 hidden layers with learning_rate=0.001, batch_size=5000, epochs = 50.

vi. Training the model by calling the denes deep learning algorithm by the Sklearn library which consists of 3 hidden layers with batch_size=5000, epochs = 100.

4.4 THE WORK ENVIRONMENT

They are the software systems and applications that contributed to the code book of the proposed model and its implementation to obtain the desired results.

4.4.1 Anaconda

It is a very popular work environment with more than 180 scientific packages including Conda and Python. The reason why anaconda is so widely used is that it has more than 1000 open-source libraries, and a simple installation process and Python is an object-oriented computer programming language with powerful cross-platform capabilities. It can be used in Linux, macOS, and Windows systems. and the strength of Python lies in its wide range of applications, covering artificial intelligence, scientific computing, web development, system operation and maintenance, big data and cloud computing, finance, and game development [84].

4.4.2 Scikit-Learn

The Scikit-Learn library is one of the libraries for machine learning in Python, and it contains many algorithms and methods used in designing machine learning models and using them in the stage of data processing and model evaluation. .

The Scikit-Learn library was launched in 2007 during the Google Summer of Code program by a group of volunteers. It is ranked second in the famous GitHub site in the list of machine learning libraries and is an easy entry point for anyone who wants to learn and apply machine learning and data science [85].

4.4.3 Keras

It is an open-source library used for neural network design and is built in Python language. Designed for quick testing on deep neural networks, it focuses on flexibility, ease of use, and scalability. It was developed by a Google engineer known as François Cholet, who is

the founder of the deep neural network model known as XCeption through research on the ONEIROS project.

Keras has ready-made offices for many of the main components used in a neural network design such as layers, activation functions, target functions, optimization methods, and a host of other tools to make it easier to manipulate image data and text data and write the code needed for deep neural network programming is clear, easy to understand and develop[86].

4.4.4 TensorFlow

It is an open-source and free machine-learning library. It has many uses but specializes in inferring and training deep neural networks.

It is a mathematics library based on data portability and differential programming developed by Google Brian Tencerflow for internal use at Google and announced in 2015 under the Apache 2.0 License.

It was first adopted in Python as a programming language to design and build many models, which were created mainly on machine learning algorithms and used in applications such as neural networks in the field of machine learning and depended on them as a research and production agent.

. It is currently used in many types of research and in various fields such as medicine, social sciences, economics, and engineering. It is also used in various types of cognitive and linguistic comprehension tasks such as text, facial, speech, and image recognition, as well as classification and discrimination processes.

Tensorflow makes it easy for developers to design, train, and test different models in deep learning. Although there are other open source libraries available to developers that serve the field of deep learning, they prefer Tensor Flow, it has been adopted by many giants such as IBM and Airbus, Twitter platform, and others - and that is thanks to the highly flexible in the system architecture[87].

4.5 DATA SETS

The database (UNSW_NB15) was obtained through Kaggle a documented website that includes a number of databases that are used free of charge in the fields of research[88].

4.6 SELECTION OF THE MOST IMPORTANT FEATURE

At this stage, the best features were selected from the database (UNSW_NB15) using the Extra tree classifier algorithm. We obtained a set of 35 features arranged in order of importance Figure (4.1)number chart, table (4.2), and 13 features were re-selected from the 35 features by the random forest algorithm Table (4.3), which led to the low variance.

The variance is the difference between the accuracy of the model's performance in the training phase and the testing phase.

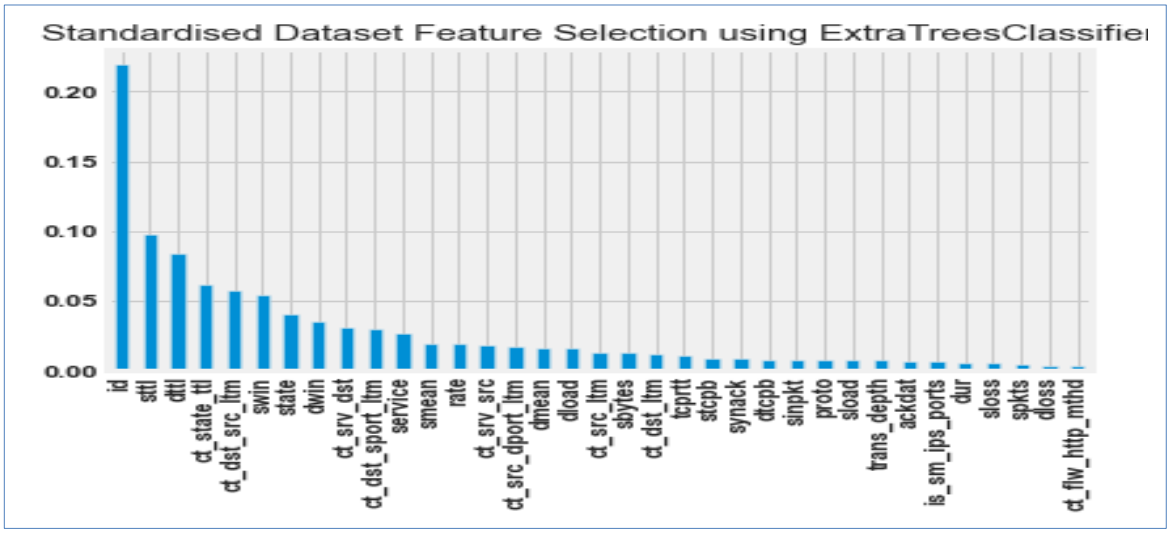

Figure 4.1: Top 20 features extracted from an algorithm Extra Trees Classifier algorithm.

Table 4.2: Top 35 features extracted from an algorithm Extra Trees

1	Id
2	Sttl
3	Dttl
4	Dttl
5	Ct_state_ttl
6	Ct_dst_src_ltm
7	Swin
8	State
9	Dwin
10	Ct_srv_dst
11	Ctdst_sport_ltm
12	Servies
13	Smean
14	Rate
15	Ct_srv_src
16	Ct_src_dport__ltm
17	Dmean
18	Dload
19	Ct_src_ltm
20	Sbytes
21	Ct_dst_ltm
22	Stcpb
23	Synack
24	Dtcpb
25	Sinpkt
26	Proto
27	Sload
28	Trans_depth
29	Ackat
30	Is_sm_ips_ports

Table 4.2: Top 35 features extracted from an algorithm Extra Trees (tables continued)

31	Dur
32	Sloss
33	Spkts
34	Dloss
35	Ct_flw_http_mthd

Table 4.3: random forest algorithm features

1	Id
2	dur
3	sbytes
4	dbytes
5	rate
6	sttl
7	sload
8	smean
9	Ct_state
10	Ct_dst_sport_ltm
11	Ct_dst_src_ltm
12	Ct_srv_dst
13	Dtype-object

4.7 MODEL FOR TRAINING AND TESTING

After reading and preparing the database (UNSW_NB15) to ensure that it does not contain empty fields and text data, the best features are extracted using the Additional Tree Classification Algorithm Figure (4.1) listed in Table (4.2) and using the Random Forest algorithm Table (4.3) to determine the best features obtained From the Extra Trees Classifier algorithm and divided into a training part and a test part, which is implemented using a deep learning model by an algorithm (denes) algorithm Figure (4.2). The model consists of 3

hidden layers with an input layer and an output layer with the use of two sets of parameters as follows

Group 1: batch_size=5000 epochs = 50.

Group 2: batch_size=5000 epochs = 100.

Activation function =relu,optimizer= adam ,Loss function = binary_crossentropy

This model combines the best features of the selection algorithms of low computational cost, low contrast, and low bias, and the high accuracy is 99.85%

```
Epoch 1/50
13/13 [==============================] - 3s 233ms/step - loss: 0.4751 - accuracy: 0.8154 - val_loss:
0.2767 - val_accuracy: 0.9082
Epoch 2/50
13/13 [==============================] - 3s 222ms/step - loss: 0.2193 - accuracy: 0.9216 - val_loss:
0.1562 - val_accuracy: 0.9451
Epoch 3/50
13/13 [==============================] - 3s 216ms/step - loss: 0.1368 - accuracy: 0.9544 - val_loss:
0.1101 - val_accuracy: 0.9650
Epoch 4/50
13/13 [==============================] - 3s 219ms/step - loss: 0.1054 - accuracy: 0.9631 - val_loss:
0.0883 - val_accuracy: 0.9697
Epoch 5/50
13/13 [==============================] - 3s 218ms/step - loss: 0.0823 - accuracy: 0.9701 - val_loss:
0.0666 - val_accuracy: 0.9769
Epoch 6/50
13/13 [==============================] - 3s 213ms/step - loss: 0.0621 - accuracy: 0.9811 - val_loss:
0.0510 - val_accuracy: 0.9843
Epoch 7/50
```

Figure 4.2: Training The Model.

4.8 RESULTS AND DISCUSSION

Through the accuracy chart and the confusion matrix can we determine the model's working efficieWhere the accuracy chart displays, the ratio of the accuracy of the model in the training phase with the ratio of the accuracy of the model in the testing phase, the accuracy value represents the ratio of the training phase, and val_acc is the accuracy of the model in the testing phase Figure (4.3).

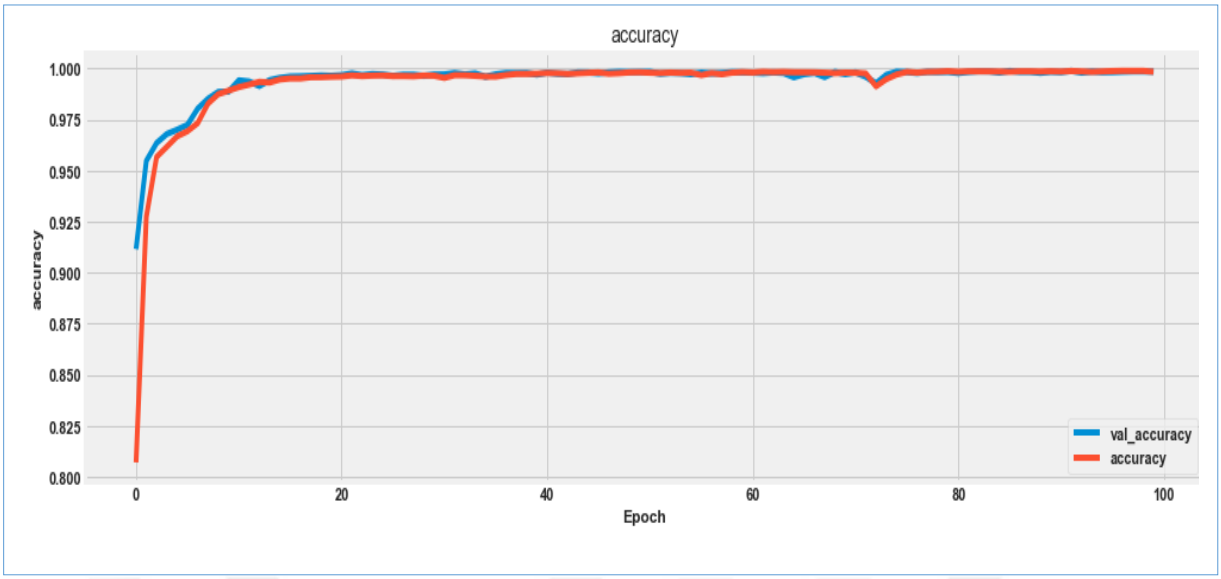

Figure 4.3: accuracy chart

and confusion matrix Figure (4.4) for Group1 and Figure (4.5) for Group2 Which is considered a standard for evaluating the performance of the model used in machine learning, Through it, we can get the value of Accuracy Eq (4.1), Precision Eq(4.2), Recall Eq(4.3), F1-Score Eq(4.4) which consists of four values (TP, FP, FN, TN) and they represent the following

True-Positives (TP) is When the real value is true and the expected value is true[17].

True-Negatives (TN) are When the actual value is negative and the forecast .is also negative [17].

False-Positives (FP) are When the actual is negative but the prediction.is positive Also known as type 1 error[17].

False-Negatives (FN) is When. the actual is positive but the. expectation is. negative. Also known.as type 2 error[17].

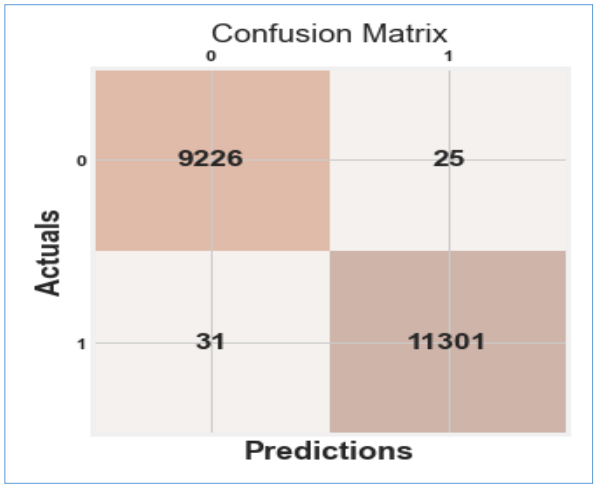

Figure 4.4: confusion matrix group 1.

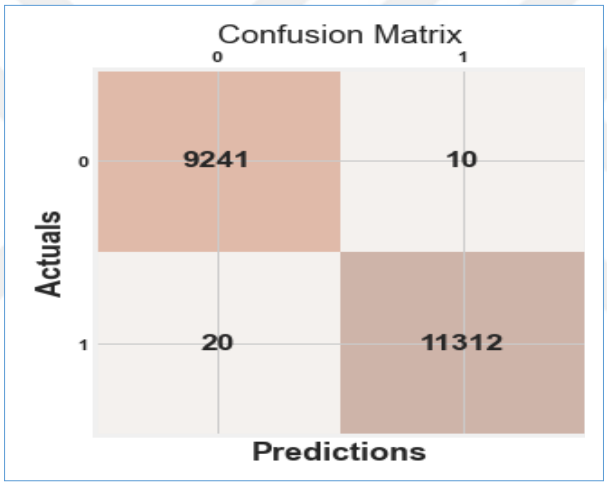

Figure 4.5: confusion matrix group 2.

i. Accuracy can be evaluated from the following equation

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{TP + FN + FP + TN} * 100 \qquad (4.1)$$

ii. Precision is the number of. accurately anticipated.records[17].

$$\text{Precision} = \frac{\text{TP}}{TP + FP} * 100 \qquad (4.2)$$

iii. Recall is the number of effectively, anticipated, records out of all records[17].

$$\text{Recall} = \frac{\text{TP}}{TP + FN} * 100 \tag{4.3}$$

iv. F1-Score is the average, between recall and, precision [17].

$$f1\ score = \frac{2 * \text{TP}}{2 * TP + FP + FN} * 100 \tag{4.4}$$

Table 4.4: Accuracy of Group Model No1

Group1	
Accuracy	$\text{Accuracy} = \frac{9226\ +\ 11301}{9226 + 26 + 31 + 11301} * 100 = 99.7\%$
Precision	$= \frac{11301}{11301 + 31} * 100 \quad = 99.7$
Recall	$= \frac{11301}{11301 + 25} * 100 \quad = 99.79\%$
fscore	$= \frac{2 * 11301}{2 * 11301 + 25 + 31} * 100 = 99.5\%$

Table 4.5: Accuracy of Group Model No.2

Group2	
Accuracy	$\frac{9241\ +\ 11312}{9241 + 10 + 20 + 11312} * 100 = 99.85\%$
Precision	$= \frac{11312}{11312 + 20} * 100 \quad = 99.8\%$
Recall	$= \frac{11312}{11312 + 10} * 100 \quad = 99.91\%$
fscore	$= \frac{2 * 11312}{2 * 11312 + 10 + 20} * 100 = 99.73\%$

The model was tested using only the extra trees classifier Algorithm to select the best 20 features Figer 4.6, so the accuracy was with a noticeable variation in the accuracy chart Figure (4.7).

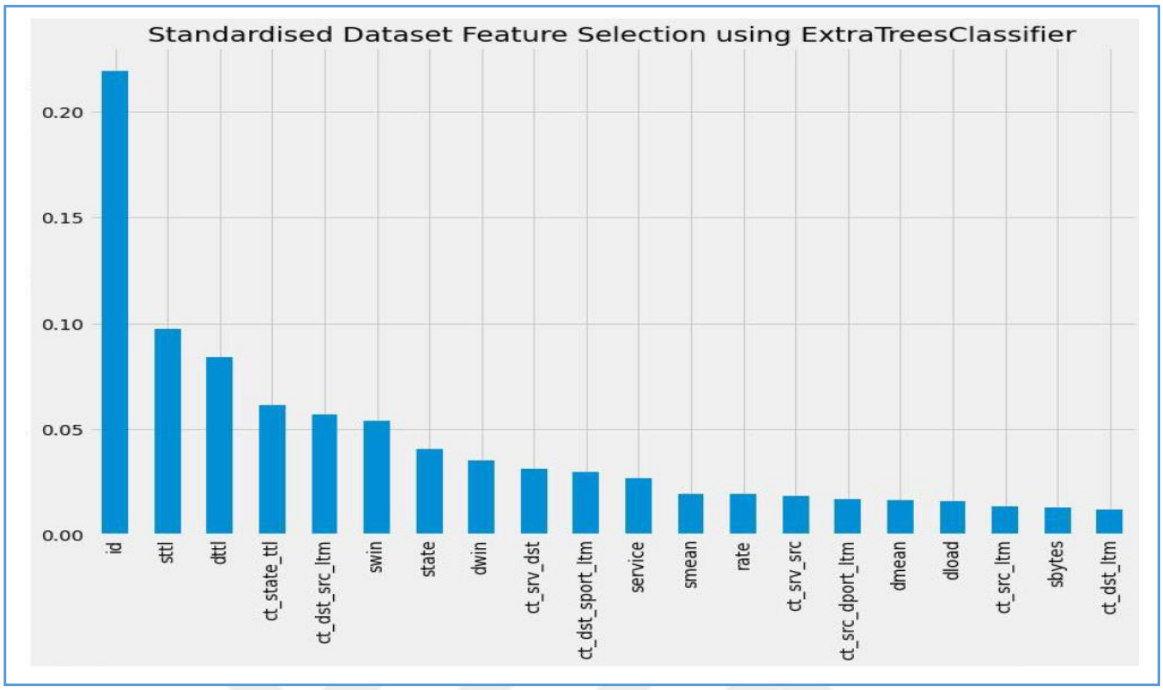

Figure 4.6: Best 20 Features

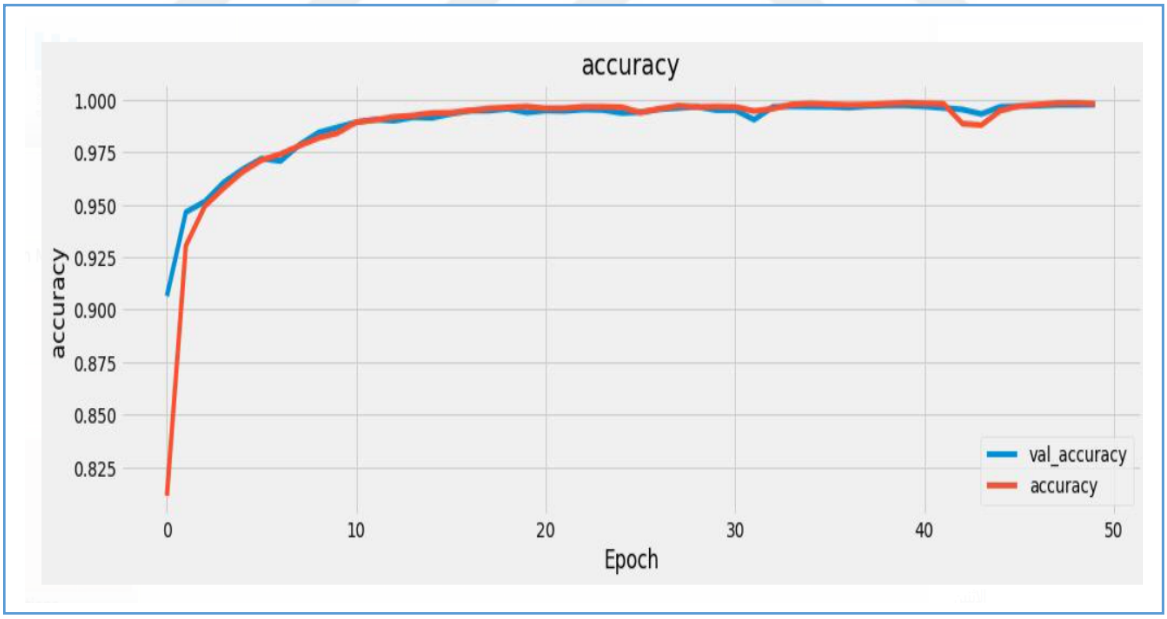

Figure 4.7: Accuracy Chart

Through the confusion matrix Figure (4.8) the results were as follows

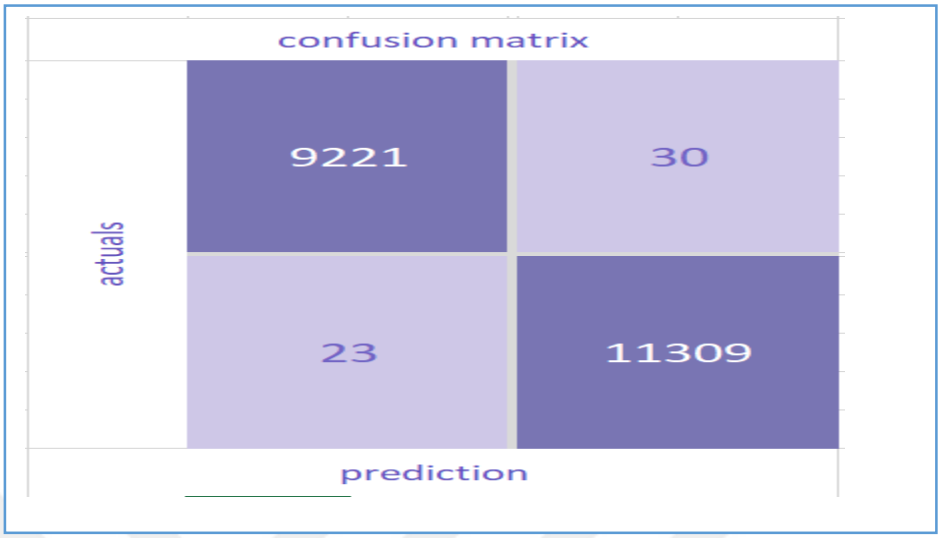

Figure 4.8: Confusion Matrix

Accuracy : 99.7%.

Precision: 99.79%

Recall: 99.73%

F1 score: 99.76%

Table 4.6: Summary of previous studies in terms of accuracy.

	researchers	Year	Data set	model	accuracy
1	V. Kanimozhi V. Kanimozhi, Prem Jacob	2019	UNSW-NB15	RandomForest algorithm for choosing the important feature and Decision tree classifier	%98
2	Jivitesh Sharma et all	2019	Unsw_nb	ExtraTrees selection + extreme learning machine(ELM)	98.24%
3	Sarika Choudhary, Nishtha Kesswani	2020	UNSW-NB15	used twenty hidden layers of the neural network (DNN)	99.2%.
4	A. M. ALEESA, et all	2021	UNSW-NB15	models (ANN and DNN and RNN)	accuracy of 99.26% and the result of the DNN technique was accuracy of 99.22%, with respect to the classification accuracy of RNN it was 85.42%.
5	Auther Makuvaza et all in 2021	2021	CICIDS 2017	MinMaxScaler for feature selection+Model (DNN)	97.25%.
6	G.C.Amaizu et all in 2021	2021	CICDDoS2019	PCC for feature selection + model (DNN)	99.66%
7	Alma D. Lopez et all in 2022	2022	CICIDS2017	Recursive Feature Elimination(RFE) for selection +models (MLR,KNN,RF,DNN)	88%,95%,99%,95%
	Accuracy of this research	2022	UNSW-NB15	the ExtraTreesClassifier +random forest + model denes	99.85%

5. CONCLUSION

Communication networks are very important today, and the Internet has become an indispensable element thanks to distributed applications that rely on the presence of the Internet. This wide use of the Internet brings with it many security holes and makes weak points of these networks and systems vulnerable to hackers and other attacks. Intrusion detection systems have been very important in addressing these attacks, especially systems that detect anomalies in network traffic. In this thesis, a deep learning model was designed to detect attacks using an extra tree classifier Which is characterized by a small computational cost as it builds decision trees randomly through it the data size was reduced and the best 35 features were selected from the database(UNWS_NB15) and the random Forest algorithm Which is characterized by low bias was used to choose the best of from the 35 features and 13 features were obtained, the model was trained and tested using a denes algorithm The epochs were obtained using two sets of parameters, and the best result was with an accuracy of 99.85% when the number of epochs = 100, which is greater than the first model, and this indicates that the model has learned to detect intrusion more than 50 times than the epochs model, which reached the size of the epochs = 50 This in turn also had a positive impact in reducing false alarms (FP), which numbered 10 out of 9,241 records, and FP, which numbered 20 out of 11,312 records.

REFERENCES

[1] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in *2017 IEEE 26th international symposium on industrial electronics (ISIE)*, 2017, pp. 1881–1886.

[2] S. Patel and J. Zaveri, "A risk-assessment model for cyber attacks on information systems.," *J. Comput.*, vol. 5, no. 3, pp. 352–359, 2010.

[3] B. Cashell, W. D. Jackson, M. Jickling, and B. Webel, "The economic impact of cyber-attacks," *Congr. Res. Serv. Doc. CRS RL32331 (washingt. DC)*, vol. 2, 2004.

[4] V. Srinivasan and C. Eswaran, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *J. Med. Syst.*, vol. 29, no. 6, pp. 647–660, 2005.

[5] A. Abraham and C. Grosan, "Evolving intrusion detection systems," in *Genetic systems programming*, Springer, 2006, pp. 57–79.

[6] I. Sofi, A. Mahajan, and V. Mansotra, "Machine learning techniques used for the detection and analysis of modern types of ddos attacks," *Int. Res. J. Eng. Technol*, 2017.

[7] P. Skrobanek, *Intrusion detection systems*. BoD–Books on Demand, 2011.

[8] R. G. Bace, *Intrusion detection*. Sams Publishing, 2000.

[9] R. Di Pietro and L. V Mancini, *Intrusion detection systems*, vol. 38. Springer Science & Business Media, 2008.

[10] H. B. Barlow, "Unsupervised learning," *Neural Comput.*, vol. 1, no. 3, pp. 295–311, 1989.

[11] V. Kanimozhi and P. Jacob, "UNSW-NB15 dataset feature selection and network intrusion detection using deep learning," *Int. J. Recent Technol. Eng*, vol. 7, pp. 443–446, 2019.

[12] S. Choudhary and N. Kesswani, "Analysis of KDD-Cup'99, NSL-KDD and UNSW-

NB15 datasets using deep learning in IoT," *Procedia Comput. Sci.*, vol. 167, pp. 1561–1573, 2020.

[13] A. M. Aleesa, M. Younis, A. A. Mohammed, and N. Sahar, "Deep-intrusion detection system with enhanced UNSW-NB15 dataset based on deep learning techniques," *J. Eng. Sci. Technol.*, vol. 16, no. 1, pp. 711–727, 2021.

[14] J. Sharma, C. Giri, O.-C. Granmo, and M. Goodwin, "Multi-layer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation," *EURASIP J. Inf. Secur.*, vol. 2019, no. 1, pp. 1–16, 2019.

[15] M. Rani, "Effective network intrusion detection by addressing class imbalance with deep neural networks multimedia tools and applications," *Multimed. Tools Appl.*, vol. 81, no. 6, pp. 8499–8518, 2022.

[16] A. D. Lopez, A. P. Mohan, and S. Nair, "Network traffic behavioral analytics for detection of DDoS attacks," *SMU data Sci. Rev.*, vol. 2, no. 1, p. 14, 2019.

[17] A. Makuvaza, D. S. Jat, and A. M. Gamundani, "Deep neural network (DNN) solution for real-time detection of distributed denial of service (DDoS) attacks in software defined networks (SDNs)," *SN Comput. Sci.*, vol. 2, no. 2, pp. 1–10, 2021.

[18] G. C. Amaizu, C. I. Nwakanma, S. Bhardwaj, J. M. Lee, and D.-S. Kim, "Composite and efficient DDoS attack detection framework for B5G networks," *Comput. Networks*, vol. 188, p. 107871, 2021.

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

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

[21] B. Škrbić and N. Đurišić-Mladenović, "Principal component analysis for soil contamination with organochlorine compounds," *Chemosphere*, vol. 68, no. 11, pp. 2144–2152, 2007.

[22] N. Patani and R. Patel, “A mechanism for prevention of flooding based ddos attack,” *Int. J. Comput. Intell. Res.*, vol. 13, no. 1, pp. 101–111, 2017.

[23] N. K. Gupta, *Inside Bluetooth low energy*. Artech House, 2016.

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

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

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

[27] S. Y. Yerima, M. K. Alzaylaee, and S. Sezer, “Machine learning-based dynamic analysis of Android apps with improved code coverage,” *EURASIP j. inf. Secur.*, vol. 2019, no. 1, 2019, doi: 10.1186/s13635-019-0087-1.

[28] M. Siddiqui, M. C. Wang, and J. Lee, “Data mining methods for malware detection using instruction sequences,” in *Artificial Intelligence and Applications*, 2008, pp. 358–363.

[29] S. Sinha and M. Sharma, “Simulation and Analysis of DDoS Attacks by Specialized Simulator using Virtualization,” *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 3, no. 2, 2015.

[30] B. Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, “Deep learning for classification of malware system call sequences,” in *AI 2016: Advances in Artificial Intelligence*, Cham: Springer International Publishing, 2016, pp. 137–149.

[31] B. Hang, R. Hu, and W. Shi, “An enhanced SYN cookie defence method for TCP DDoS attack,” *J. Netw.*, vol. 6, no. 8, 2011, doi: 10.4304/jnw.6.8.1206-1213.

[32] D. Wang, Zhu yufu, and J. Jie, “A multi-core based DDoS detection method,” *2010 3rd International Conference on Computer Science and Information Technology*.

IEEE, 2010, doi: 10.1109/iccsit.2010.5564969.

[33] L. Garg, C. Chakraborty, and S. Mahmoudi, *Healthcare Informatics for Fighting COVID-19 and Future Epidemics*. Springer International Publishing, 2022.

[34] S. M. Lee, D. S. Kim, J. H. Lee, and J. S. Park, "Detection of DDoS attacks using optimized traffic matrix," *Comput. Math. Appl.*, vol. 63, no. 2, pp. 501–510, 2012, doi: 10.1016/j.camwa.2011.08.020.

[35] H. Rahmani, N. Sahli, and F. Kamoun, "DDoS flooding attack detection scheme based on F-divergence," *Comput. Commun.*, vol. 35, no. 11, pp. 1380–1391, 2012, doi: 10.1016/j.comcom.2012.04.002.

[36] G. B. Senirkentli *et al.*, "Proton therapy for mandibula plate phantom," *Healthc.*, vol. 9, no. 2, p. 167, 2021, doi: 10.3390/healthcare9020167.

[37] V. Solms and R. Van Niekerk, "From information security to cyber security," *Comput. Secur.*, vol. 38, pp. 97–102, 2013.

[38] R. Reid and J. Van Niekerk, "From information security to cyber security cultures," in *2014 Information Security for South Africa*, 2014, pp. 1–7.

[39] R. R. Brooks, L. Yu, I. Ozcelik, J. Oakley, and N. Tusing, "Distributed denial of service (DDoS): A history," *IEEE ann. hist. Comput.*, vol. 44, no. 2, pp. 44–54, 2022, doi: 10.1109/mahc.2021.3072582.

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

[41] F. Jauro, H. Chiroma, A. Y. Gital, M. Almutairi, S. M. Abdulhamid, and J. H. Abawajy, "Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend," *Appl. Soft Comput.*, vol. 96, no. 106582, p. 106582, 2020, doi: 10.1016/j.asoc.2020.106582.

[42] A. M. Karim, F. V Çelebi, and A. S. Mohammed, "Software Development for Blood Disease Expert System," *Lect. Notes Softw. Eng.*, vol. 4, no. 3, 2016.

[43] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V Çelebi, "A new generalized deep learning framework combining sparse autoencoder and Taguchi Method for novel data classification and processing," *Math. Probl. Eng.*, vol. 2018, pp. 1–13, 2018, doi: 10.1155/2018/3145947.

[44] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953, doi: 10.1038/171737a0.

[45] A. Karim, "Development of secure Internet of Vehicle Things (IoVT) for smart transportation system," *Comput. Electr. Eng.*, vol. 102, no. 108101, p. 108101, 2022, doi: 10.1016/j.compeleceng.2022.108101.

[46] A. M. Karim, H. Kaya, V. Alcan, B. Sen, and I. A. Hadimlioglu, "New optimized deep learning application for COVID-19 detection in chest X-ray images," *Symmetry (Basel)*, vol. 14, no. 5, p. 1003, 2022, doi: 10.3390/sym14051003.

[47] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transp. Res. Part C Emerg. Technol.*, vol. 118, no. 102674, p. 102674, 2020, doi: 10.1016/j.trc.2020.102674.

[48] Y. Chen, "Voltages prediction algorithm based on LSTM recurrent neural network," *Opt.*, vol. 220, no. 164869, p. 164869, 2020, doi: 10.1016/j.ijleo.2020.164869.

[49] N. Andrei, "A Dai–Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization," *Appl. Math. Lett.*, vol. 21, no. 2, pp. 165–171, 2008, doi: 10.1016/j.aml.2007.05.002.

[50] M. M. Öztürk, İ. A. Cankaya, and D. İpekçi, "Optimizing echo state network through a novel fisher maximization based stochastic gradient descent," *Neurocomputing*, vol. 415, pp. 215–224, 2020, doi: 10.1016/j.neucom.2020.07.034.

[51] A. Koreanschi *et al.*, "Optimization and design of an aircraft's morphing wing-tip demonstrator for drag reduction at low speeds, Part II-Experimental validation using Infra-Red transition measurement from Wind Tunnel tests," *Chinese J. Aeronaut.*,

vol. 30, no. 1, pp. 164–174, 2017.

[52] A. Jentzen and P. von Wurstemberger, “Lower error bounds for the stochastic gradient descent optimization algorithm: Sharp convergence rates for slowly and fast decaying learning rates,” *J. Complex.*, vol. 57, no. 101438, p. 101438, 2020, doi: 10.1016/j.jco.2019.101438.

[53] S. Wang, L. Zhang, and J. Fu, “Adversarial transfer learning for cross-domain visual recognition,” *Knowl. Based Syst.*, vol. 204, no. 106258, p. 106258, 2020, doi: 10.1016/j.knosys.2020.106258.

[54] X. Yang, Y. Zhang, W. Lv, and D. Wang, “Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier,” *Renew. Energy*, vol. 163, pp. 386–397, 2021, doi: 10.1016/j.renene.2020.08.125.

[55] C. Li, S. Zhang, Y. Qin, and E. Estupinan, “A systematic review of deep transfer learning for machinery fault diagnosis,” *Neurocomputing*, vol. 407, pp. 121–135, 2020, doi: 10.1016/j.neucom.2020.04.045.

[56] Y. Wang, Y. Liu, W. Chen, Z.-M. Ma, and T.-Y. Liu, “Target transfer Q-learning and its convergence analysis,” *Neurocomputing*, vol. 392, pp. 11–22, 2020, doi: 10.1016/j.neucom.2020.02.117.

[57] Z. Qiu, S. Zhao, X. Feng, and Y. He, “Transfer learning method for plastic pollution evaluation in soil using NIR sensor,” *Sci. Total Environ.*, vol. 740, no. 140118, p. 140118, 2020, doi: 10.1016/j.scitotenv.2020.140118.

[58] S. K. Behera, A. K. Rath, and P. K. Sethy, “Maturity status classification of papaya fruits based on machine learning and transfer learning approach,” *Inf. Process. Agric.*, 2020, doi: 10.1016/j.inpa.2020.05.003.

[59] S. Kim, Y.-K. Noh, and F. C. Park, “Efficient neural network compression via transfer learning for machine vision inspection,” *Neurocomputing*, vol. 413, pp. 294–304, 2020, doi: 10.1016/j.neucom.2020.06.107.

[60] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, “Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks,” *Neural Netw.*, vol. 129, pp. 313–322, 2020, doi: 10.1016/j.neunet.2020.06.014.

[61] P. Kant, S. H. Laskar, J. Hazarika, and R. Mahamune, “CWT based transfer learning for motor imagery classification for brain computer interfaces,” *J. Neurosci. Methods*, vol. 345, no. 108886, p. 108886, 2020, doi: 10.1016/j.jneumeth.2020.108886.

[62] X. Li, Z. Hu, M. Xu, Y. Wang, and J. Ma, “Transfer learning based intrusion detection scheme for Internet of vehicles,” *Inf. Sci.*, vol. 547, pp. 119–135, 2021, doi: 10.1016/j.ins.2020.05.130.

[63] S. Saud, B. Jamil, Y. Upadhyay, and K. Irshad, “Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach,” *Sustain. Energy Technol. Assessments*, vol. 40, no. 100768, p. 100768, 2020, doi: 10.1016/j.seta.2020.100768.

[64] J. Wei and H. Chen, “Determining the number of factors in approximate factor models by twice K-fold cross validation,” *Econ. Lett.*, vol. 191, no. 109149, p. 109149, 2020, doi: 10.1016/j.econlet.2020.109149.

[65] T.-T. Wong, “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation,” *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, 2015, doi: 10.1016/j.patcog.2015.03.009.

[66] G. Jiang and W. Wang, “Error estimation based on variance analysis of k-fold cross-validation,” *Pattern Recognit.*, vol. 69, pp. 94–106, 2017.

[67] N. S. Harzevili and S. H. Alizadeh, “Analysis and modeling conditional mutual dependency of metrics in software defect prediction using latent variables,” *Neurocomputing*, vol. 460, pp. 309–330, 2021, doi: 10.1016/j.neucom.2021.05.043.

[68] C. Ni, X. Chen, F. Wu, Y. Shen, and Q. Gu, “An empirical study on pareto based multi-objective feature selection for software defect prediction,” *J. Syst. Softw.*, vol. 152, pp. 215–238, 2019, doi: 10.1016/j.jss.2019.03.012.

[69] M. Karoriya and G. Gogate, “A Review of Intrusion Detection Systems,” 2022.

[70] A. S. Ashoor and S. Gore, “Importance of intrusion detection system (IDS),” *Int. J. Sci. Eng. Res.*, vol. 2, no. 1, pp. 1–4, 2011.

[71] D. M. Best, S. Bohn, D. Love, A. Wynne, and W. A. Pike, “Real-time visualization of network behaviors for situational awareness,” *Proceedings of the Seventh International Symposium on Visualization for Cyber Security - VizSec ’10*. ACM Press, 2010, doi: 10.1145/1850795.1850805.

[72] Y. Baştanlar and M. Özuysal, “Introduction to machine learning. miRNomics: MicroRNA biology and computational analysis,” pp. 105–128, 2014.

[73] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely Randomized Trees,” *Mach. Learn.*, vol. 63, pp. 3–42, Apr. 2006, doi: 10.1007/s10994-006-6226-1.

[74] T. Hastie, R. Tibshirani, and J. Friedman, “Random forests,” in *The elements of statistical learning*, Springer, 2009, pp. 587–604.

[75] S. Tangirala, “Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, pp. 612–619, 2020.

[76] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[77] S. Skansi, *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Springer, 2018.

[78] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” *2017 International Conference on Engineering and Technology (ICET)*. IEEE, 2017, doi: 10.1109/icengtechnol.2017.8308186.

[79] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, USA, 2015.

[80] F. Rosenblatt, *Principal of NeurodynamY ics: Perceptrons and the theory of brain mechaY nisms*, vol. 616. Washington, DC: Spartan Boo@ s Editions, 1962.

[81] S. Sharma, S. Sharma, and A. Athaiya, “Activation functions in neural networks,” *Towar. data Sci.*, vol. 6, no. 12, pp. 310–316, 2017.

[82] S. Qiu, X. Xu, and B. Cai, “FReLU: flexible rectified linear units for improving convolutional neural networks,” in *2018 24th international conference on pattern recognition (icpr)*, 2018, pp. 1223–1228.

[83] W. Ma and J. Lu, “An equivalence of fully connected layer and convolutional layer,” *arXiv Prepr. arXiv1712.01252*, 2017.

[84] D. Rolon-Mérette, M. Ross, T. Rolon-Mérette, and K. Church, “Introduction to Anaconda and Python: Installation and setup,” *quant. methods psychol.*, vol. 16, no. 5, pp. S3–S11, 2020, doi: 10.20982/tqmp.16.5.s003.

[85] O. Kramer, “Scikit-Learn BT - Machine Learning for Evolution Strategies,” O. Kramer, Ed. Cham: Springer International Publishing, 2016, pp. 45–53.

[86] R. Atienza, *Advanced Deep Learning with Keras: Apply deep learning techniques, autoencoders, GANs, variational autoencoders, deep reinforcement learning, policy gradients, and more*. Packt Publishing Ltd, 2018.

[87] N. McClure, *TensorFlow Machine Learning Cookbook*, 2nd ed. Birmingham, England: Packt Publishing, 2018.

[88] A. Narayanan, E. Shi, and B. I. P. Rubinstein, “Link prediction by de-anonymization: How We Won the Kaggle Social Network Challenge,” *The 2011 International Joint Conference on Neural Networks*. IEEE, 2011, doi: 10.1109/ijcnn.2011.6033446.