REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technology

DETECTION OF DENIAL-OF SERVICE ATTACKS USING ARTIFICIAL NEURAL NETWORKS

Sarmad Jassim MOHAMMED

Master's Thesis

Supervisor

Prof. Dr. Osman Nuri UÇAN

Istanbul, 2023

DETECTION OF DENIAL-OF SERVICE ATTACKS USING ARTIFICIAL NEURAL NETWORKS

Sarmad Jassim MOHAMMED

Information Technology

Master’s Thesis

ALTINBAŞ UNIVERSITY

2023

The thesis titled DETECTION OF DENIAL-OF SERVICE ATTACKS USING ARTIFICIAL NEURAL NETWORKS prepared by SARMAD JASSIM MOHAMMED and submitted on 03/04/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technology.

Prof. Dr. Osman Nuri UÇAN

the Supervisor

Thesis Defense Committee Members:

Prof. Dr. Osman Nuri UÇAN	Department of Electrical and Electronic Engineering, Altınbaş University	_____
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Sciences and Engineering, Altınbaş University	_____
Assoc. Prof. Dr. Adil Deniz DURU	Department of Biomedical Engineering, 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.

Sarmad MOHAMMED

Signature

DEDICATION

First, I thank God for helping me and easing my difficulties. To my beloved father and great-mother are the secrets of my happiness and comfort. For the given spirit and symbol of achievement and encouragement my dear wife Parla Khlil. To my dear sister, the advice She didn’t diminish the help.

Everyone who gave me advice and support wished me success and encouraged me to study, especially them (Zaid Farhan, Mahdi Jaffeer).

ABSTRACT

DETECTION OF DENIAL-OF SERVICE ATTACKS USING ARTIFICIAL NEURAL NETWORKS

Mohammed, Sarmad

M.Sc., Information Technology, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri UÇAN

Date: April/2023

Pages: 51

In artificial neural networks, the layers between the input and the output are not visible to the naked eye (ANNs). Total neuronal synaptic weights. A number of different types of equations, predictions, and regressions may be created using ANNs. At the first stage of the learning process, ANN is in charge of fine-tuning synaptic parameters so that the system can acquire new information. At completion of training, the ANN's generalization capabilities should be evaluated using data that was not part of the training set. Machine learning refers to the ability of a system to acquire knowledge, identify patterns, and form opinions without any human intervention., which analyze data in a way that's analogous to how neurons do. The inputs to a synthetic neuron are weighted and then combined before the neuron is triggered by a sigmoid or threshold function. DOSS attacks have frequently been deadly since 1999. Radwar has discovered that denial-of-service (DOSS) assaults are among the most serious threats facing modern corporations. The security of online networks remains vulnerable to DOSS attacks. So, it clearly shows the most urgent measures for safety. We built a network monitoring system to keep an eye out for attacks. Confirming the identity of a user is essential for security purposes. Attacks of this kind include the Denial of Service (DoS) attack and others like it. Denial-of-service attacks make it impossible for legitimate users to get to the data or services the provider normally provides.

Keywords: Denial of Service, Information Security, ANN, GWO, DOSS Attack, Network Security.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

RNN	:	Recurrent Neural Network
DBN	:	Deep Belief Network
DL	:	Deep Learning

1. INTRODUCTION

Web and data services are increasingly used these days. They have become part of today's world. Whether it is education, art, business, or other human activity, they find their rightful place everywhere. However, due to their popularity, these services are often the target of attacks. A simple example for us can be online stores. In their case, the denial of web service means a direct financial loss. We could find many similar examples. From the point of view of preventing possible threats and limiting their consequences, it is good to know the performance limits of the devices on which we operate the given services. Spirent Communications' dedicated Avalanche 3100 can help us uncover these limits and other pitfalls. Thanks to this device, we are able to implement the entire spectrum of stress tests and targeted attacks, obtain the necessary results, and based on them, we can propose the necessary measures to increase the security and resistance of our devices. A DoS attack, or Denial of Service connection, means denial of service. There is also often talk about so-called DDoS attacks (Distributed Denial of Service), which are DoS attacks conducted from multiple sources. The consequence of these attacks is disruption or complete inaccessibility of the service for its regular users. The unavailability of services and systems causes significant financial losses to their operators, damage to the company's reputation, or the need to analyze and correct the problem. If we consider that most of today's companies and organizations are highly dependent on network services, we are dealing with a big problem. The popularity of these attacks among attackers is growing considerably. No in-depth knowledge is required for their implementation these days, as suitable tools for their implementation are commonly available. The fact that disrupting the operation of a network or system is often much easier than gaining access to it is just another reason for their expansion [12]. There are many motives for carrying out an attack. It can be, for example, attacks on competing companies with the aim of damaging its reputation or otherwise weakening its position on the market. In the media, we also often come across attacks that are conducted as a certain form of demonstration. In these cases, the target is most often government networks and systems, or groups and companies threatening freedom of speech. Last but not least, DDoS attacks are used as a terrorist threat. Nowadays, there are a huge number of different types of DDoS attacks. Deficiencies in the architecture and operation of TCP/IP protocols, which at the time were designed for use in an open and trusted

environment [12], are often exploited. Attacks are also aimed at occupying the transmission capacity of the line, exhausting system resources, errors in programs, packet routing systems or DNS. Attacks can also be specified and divided according to the TCP/IP layers through which they operate. In our case, we will focus on a brief analysis of those DDoS attacks provided by the Avalanche 3100 device. Biologically inspired computing and swarm intelligence have become really popular techniques in recent years [44][45]. Thanks to its simplicity and flexibility, swarm intelligence is used in many modern applications in various disciplines where traditional methods are not sufficiently accurate or do not yield satisfactory results. Swarm intelligence itself is part of a larger whole, which is represented by the already mentioned biologically inspired algorithms. In general, these are population-based algorithms. This means that the population of individuals (potential candidates for solutions) cooperates with each other and statistically improves over generations, eventually finding a solution. The concept of swarm intelligence includes multiplicity, stochasticity, randomness and also disorder.

Swarm intelligence is a set of nature-inspired algorithms that are based on the collective behavior of simple agents. They follow certain rules. The agent itself may be considered unintelligent, but the entire system of agents exhibits elements of self-organization and decentralization and thus behaves as a kind of collective intelligence. Algorithms in the group of swarm intelligence consist primarily of two phases that maintain a balance between exploration and exploitation. Thus, they force the whole swarm, i.e., a set of possible solutions, to adjust their positions. It is a variation and selection phase, while the variation phase examines different areas of the searched space, while the selection phase serves to use previous experiences.

2. ARTIFICIAL NEURAL NETWORKS

2.1 GENERAL PROPERTIES OF ARTIFICIAL NEURAL NETWORKS

In some ways, the maturation of neurons in the brain may be considered as analogous to the formation of an artificial neural network. Exposure to computers on a regular basis promotes cognitive flexibility and learning. The modeling done here might pave the way for the creation of more technological solutions that need fewer people to operate. When it comes to scientific progress, we believe neural networks are a major game changer. The job, which involves both abstract algebra and face recognition, is performed by computers. As a result, it might be difficult to design even simple patterns for computers. Computers lack the cognitive abilities necessary to assess, generalize, or implement learned patterns in the future. One way that people, for instance, might benefit from more complicated neural network analysis is by gaining a deeper understanding of the reasoning process. Our study's model [17], places special attention on the brain's capabilities. When people are involved, it might be difficult to investigate and understand many of these incidents. This method uses a new machine to store and analyze previously learned pattern data. The neural network has to be primed and trained before it can be put to use in resolving specific issues. In addition to thinking, interacting, forgetting, and reacting, our neural network is capable of a large variety of additional tasks. Even though the human brain's operation is undoubtedly a mystery, a specific form of cell that does not seem to proliferate as regularly as the cells in the rest of the body has been identified for any component of the brain's processor. In reality, this kind of cell can't replace the brain when it comes to our ability to recall past experiences and think critically about them. No one has ever been the first to tackle the difficulties the human brain presents in neural computing. The engineering firm that seeks answers has never been able to match the human brain's information-processing capacity. It's a creative solution to the problem of obstacles like machinery.

Let's check out a brief summary of each neuron. The "neuron" is the primary element of a neural network. Biological neurons take the input and do a number of intensive processes on it before delivering the final output. Dendrites, somas, and axons are the four structural components of nerve cells. It is well-known that dendrites have a role in the neurological system as inputs. Soma takes care of the data management. The axon connects neurons and

processes information received through synapses. In living things, the structure of neurons is complex and not simple.

Learning about synapses could be facilitated by studying biology. Network programmers might do well to study biological brains in order to improve their job. Short for "artificial neural networks," or ANNs are a class of computational techniques used widely in many disciplines to model complex phenomena in the actual world. The neurons that make up ANNs are highly linked and flexible parts of the computer architecture. While representing data, these neurons execute complex computations. Some of ANN's most distinguishing characteristics are its ability to "cope with faulty and insufficient information," "handle erroneous information," "resilience," "dropping tolerances," and "resistance to loss and mistakes."

Mathematical techniques are used in ANN-based computing to teach artificial neural networks to mimic human intelligence. Researchers have used this strategy because they believe it will help them better comprehend how the human brain works. Artificial neural networks (ANN) modeling has the potential to deliver almost precise answers to problems and processes that are now detectable only via laboratory and field observations." Modeling, ranking, pattern recognition, and multivariate analysis are just a few of the many uses for microbiological ANNs. It is possible to trace the origins of the concept of a dynamic image feature, or f (x, y), back to two crucial uses: the enhancement of humans via the acquisition of object knowledge and the encoding, transmission, and representation processing of data for autonomous sensing robots. Each of these uses revolve around some variation on the idea of a moving picture." Photographic photos may be converted to digital format with the use of digital technologies. There are many different parts that come together to form a digital photograph, and they all play important roles in the final product. The term "picture components" is used to describe all the individual elements that make up a picture, such as the pixels and textures. A versatile and extensive structure for making digital photographs.

There is a growing need to analyze high-dimensional data as both sensor resolution and computer memory capacities increase. There are several examples, such as the transmission of visual and auditory impulses and other forms of multisensory data. These signals, in whole or in part, are shown in a sample space as a result of the analysis. To illustrate this idea, consider a 16x16-pixel image representing a single data point. In 256-dimensional space, a point cloud represents a set of images. Quite a bit of interdependence often exists in multi-

sensor data sets, either between the sensors themselves or between their constituent parts. With regards to neighboring picture pixels, this is undeniably accurate. To assume that every physical experiment will generate dozens of new degrees of freedom is implausible. This suggests that in the not-too-distant future, we should expect to encounter reduced-dimensional representations of multi-dimensional data sets. It might be helpful to have a mix of methods for explaining topics available for these talks. Possible applications include determining the kind of data format or query being used, comparing two sets of data to find the most comparable entries in a database, or reconstructing lost data values.

Neural networks' primary benefits lie in their adaptability and their ability to generalize. Let me restate: a capable network is one that can identify data it has never seen before. The truth is that developers often only see a tiny fraction of the space filled by neural network patterns. The data collection must be split into three sections before any broad generalizations can be made: Currently Preparing to Conduct Experiments on Neural Networks This omission may be fixed at the planning stage.

How well your neural network performs on patterns it has not been trained to recognize may be determined with the use of a validation kit. Together, these tests determine the health of a neural network. If anyone exam is failed, all subsequent training will be placed on hold immediately. Once the net is at its widest, it can catch just about anything. If you don't stop studying, you'll eventually need more instruction, which will reduce your overall effectiveness. This is the case even if the percentage of erroneous examples in the training data keeps going down. Eventually, when the network has been trained, it will be put to the test using the research collection, the third and final dataset. Testing is performed at each and every stage of the ANNS training procedure.

Fault tolerance is a system's ability to keep operating normally despite the failure of some of its parts (or many internal failures). The amount of efficiency lost by a service may not necessarily increase in direct proportion to the scale of the loss, unlike with artificial systems where the amount of efficiency lost would be proportional to the size of the system. Failure tolerance is of the highest importance in highly efficient or life-critical settings. Failure of a machine part is said to have happened "gracefully" when it occurs in a way that is typical of the part's expected lifespan.

If the system is "defect-tolerant," it will continue to work, although at a lower level of efficiency, even if one or more of its parts are broken. A defect exists in software if it

suddenly ceases doing what it's supposed to. As used in this context, "incomplete" refers to computer systems whose performance is intermediate between "complete" and "partially failed," as seen by a lagged response time or partially failed output. So, the inaccessibility of the computer may occur regardless of whether the software or hardware is at fault. Wear and tear, corrosion, poor manufacturing, and accidents may all compromise a vehicle's structural integrity and functionality, but modern automobiles are meant to keep on rolling with just a little help from their friends. The vehicle is another example that demonstrates this idea.

To increase the system's tolerance to failure, it's important to do two things: (1) make an attempt to stabilize the system so that it may be included into an ideal system, and (2) plan for unexpected occurrences and provide a mechanism inside the system to fix them if they occur. These processes occur simultaneously. It is best to replicate anything you need if the consequences of employing an inefficient tool or the expense of improving it are significant. The program must be able to restore itself to a safe state in the case of a catastrophic failure in the underlying system. This regeneration is quite similar to an inversion; but, if there are people in the area, it may be the result of an intentional action on their part.

Analogy is a cognitive process that involves transforming a generalized notion, such as an analogy or context, into a more specific form, either in thought or speech. The inference or reasoning from one premise to another is less strictly defined than in the case of deduction, induction, or kidnapping when at least one hypothesis is universal. It is common practice in biology to use the term "analogy" to describe a connection between two entities that are neither chemically or physically similar. Rutherford's atomic theory establishes a correlation between molecular structure and planetary orbits. Some of the cognitive processes involved in addressing problems include deliberation, interpretation, generalization, memory, creativity, innovation, anticipation, empathy, characterization, conceptualization, and expression. Even if the language is not metonymic, it may help convey meaning if it is rich with instances, analogies, and metaphors, such as allegories and parables. At its very heart, the comparative semantic core consists of the act of comparing several elements of a subject. As there are no concrete examples being used to make a point, this cannot be a metonym. To determine the meaning of phrases like "as if," "like," "etc.," and "the same word," the reader must use his or her own inferences. In the sciences, the social sciences, the law, and the humanities, among other fields of study, the use of analogies is crucial. Analogy is linked

to many different ideas, including comparison, benchmarking, correspondence, mathematical and morphological unity, homomorphism, and iconicity. In certain cases, it may be instructive to compare the idea of conceptual metaphor to the concept of analogy, which is essential to the study of semantic linguistics. Each and every assertion and analysis based on empirical facts that is then applied to objects that were not a part of the first experiment requires the usage of analogy.

2.2 ARTIFICIAL NEURAL NETWORKS INPUTS ELEMENTS

This layer takes in all of the information that is available from the other parts of the system, including data, signals, processes, and measurements. These inputs are typically considered standardized so long as they continue to fall within the parameters defined by the activation function. The actual numerical values that are being followed by the sequence or pattern. The implementation of these rules has a positive impact on the capacity of the network to carry out mathematical operations with greater precision.

In artificial neural networks, each connection between neurons is given a "weight" or a "parameter," depending on the kind of network. Both weights are susceptible to change whenever new information is introduced into a neural network. In a different way of putting things, the neural network is able to comprehend.

The "weights" of a simulated neural network may be used to provide a rough approximation of a number of neuronal activities that occur in the actual world. On the other hand, myelination is a very important process. Myelination. Myelination. Not only is it conceivable for artificial neural networks (ANNs), but their use of both positive and negative weights is also strongly encouraged. An rise in the number of postsynaptic terminal neurotransmitter receptors is connected with the development and integration of vesicular and presynaptic terminal neurotransmitters, as well as an increasing combination of dendrites and dendritic synapses. Synaptic weights are responsible for the release of anticipatory neurotransmitters into the synaptic cleft. Positive synaptic weights function in the same manner as presynaptic terminals (i.e., glutamates). If a different cell is the one to receive the signal, then there is a greater possibility that the target cell will become activated.

Others have suggested that a connection might be made between the negative value and the synapses on neurotransmitters (i.e., GABA). One is required to carry a burden that is

unwelcome. In the event that there was a possibility that they might take part, the receiving unit would refrain from firing their weapons.

Myelination: As the process of myelination progresses, there is a reduction in the distance that separates axon movement and myelin thickness. Since the membrane does not have myelination, the stress potentials of the membrane are noticeably lower than those of the cell body. This is because the stress potentials of the cell body are myelinated. the equivalent of a pair of trousers that can be worn in a greenhouse. When the axon fails to myelinate, the garden trousers develop leaks, and since the water pressure is dropping, it is no longer feasible to use the pants because the pressure is too low (this is an essential factor for water management, the potential impact) (which is important for water control, the possible effect).

2.3 FUNCTION THAT IS ADDITIVE

An arithmetic f(n) function that is the sum of the functions of the integer when a and b are compressed into a smaller region is called an additive function in a positive integer. Even if the positive integral components of f(a)+f(b) are not co prime, the result is an additive function. The function represented by F / f has the additive feature (b). It is commonplace in this context to compare and contrast fully multiplicative functions with fully additive functions. An example sentence: F (1) = 0 if and only if f is a totally additive trait. A material utilized only for its adding characteristics is referred to as an additive. To what extent does the activation function The activation function is also known as the threshold function or the transition function. At this stage, output signals have been interpreted from the processes that initially activated neurons. Activation functions for neural networks may be obtained in a number of different ways. Identity functions, phases, component linear functions, and sigmoid functions are all instances of function forms.

2.4 LEARNING THAT IS SUPERVISED

Training sets are provided for guided teaching. This law covers a wide range of examples of a network's proper behavior. The expected results are derived once the inputs are supplied as a guided learning lesson. Parameters will be gently adjusted with the error signal during this investigation, parameters will be slowly updated with the error signal. The learning rule covers a wide range of scenarios (established trainings) as well as optimal network behavior. In this case, xn represents the network input, and dn represents the desired output. All inputs

are responsible for the final result. If you want more accurate results from your network, you may adjust its biases and weights with the help of the research rule. We use supervised learning to ensure that we give the computer the right response whenever it is given a question (d). The network parameter is externally adjusted based on the deviation between the actual response and the expected answer. In cases when the solution to the mistake is known, such as the analysis of input patterns, this error may be used to adjust the weight. In order to operate in the learning mode, a large number of input and output models are needed, in addition to the training collection.

2.5 LEARNING WITHOUT SUPERVISION

The process of unforeseen learning also makes use of self-organized analysis. In a learning environment without any boundaries, it is difficult to arrive at a reliable conclusion. In this case, the network data is used to calculate the weights and biases used by the model. When rearranging patterns, unattended research classification is utilized. When the network is applying unsupervised learning, specific error information cannot be used to modify the behavior of the network. This is due to the fact that knowledge gained in isolation cannot determine the appropriate response. Research must be conducted based on prior experiences with responses to withheld or novel results since there is no data that can be utilized to rectify incorrect replies. Unchecked learning algorithms rely on duplicate row data that lacks a mark for class or club membership. To define its parameters, the network must first identify any existing structures, objects, rules, and so forth. Because the teacher is not necessary, but must establish objectives, studying without attendance implies studying without the instructor. The importance of receiving input on neural networks cannot be overstated. Feedback is referred to be incremental learning, and it is critical for unrestricted learning.

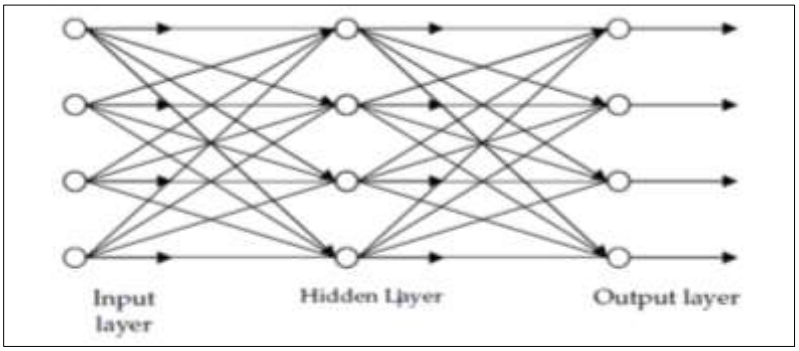

Figure 2.1: Multilayer Feed Forward Network [2].

2.6 SUPERVISED INSTRUCTION

Instructional inputs and targeted outcomes are conveyed to the student in a supervised learning environment. The data received is compared to the data provided via the network. There will be errors in the weight of the network as it grows because of this technique. As long as the person continues to make the effort to steadily change their weight, the cycle will continue. When students learn about networking, they repeatedly analyse the same data while the connection weights are increased. The simple process of collecting test data for analysis constitutes a "test session." In various situations, a network may have trouble comprehending. This is a possibility because the data used to produce the expected outcome lacked precise information. When there isn't enough information to do a thorough analysis, network convergence is impossible. It is possible that a backup copy of some data might be stored if all went according to plan. Many tiers of networks will be used, and many individual nodes will be responsible for storing data. Until the system was approved to track the network, the data would remain in the monitored testing phase. The system's decision to just store the information would be made at that moment. After the system was approved to track the network, the data would be stored in a supervised testing environment.

Inputs and outputs, the number of layers, the numerical components of each layer, layer-based networking, transitions, training features, and starting weights are all things to think about if the issue is too challenging for the network. The network can find a solution if the challenge is not too challenging. For the designer's work to be validated, it must comply with the requirements in a substantial way. As part of the training process, many separate suggestions (algorithms) are employed to feed in the right weight changes. In the present day, back-propagation is the method of choice. This isn't merely a precaution; it was designed to keep the network from being overloaded. Initially, the training data for the artificial neural network consists of the commonly observed statistical patterns. A discussion of common errors in the information then follows.

A client may request that the weights be frozen after the machine has been properly trained and no longer needs testing. We want to implement this network as physical hardware in the near future so that it can accommodate additional gadgets. The majority of electronics do not stop learning as they develop throughout time.

2.7 ADAPTIVE OR UNSUPERVISED TRAINING

The alternative educational approach is very flexible (learning). The network will benefit from the inputs, but the outputs are optional. The gadget has to provide an explanation of its operations before it can appropriately sort the incoming data. This idea may also be referred to as freedom of action or individual autonomy. External factors have no effect on the weighting of such networks. Instead, they are constantly monitored and assessed. The networks analyse the incoming data for repeating structures and then adjust the input accordingly. The network is also trying to figure out how to arrange itself without knowing whether its findings are correct. There are several ways in which the network's analytical algorithms and its underlying structure capitalize on these characteristics. An unsupervised learning technique will give greater weight to the inter-cluster cooperation that already exists. The clusters will work together using this strategy. Each node in a cluster has the potential to enhance the overall cluster's performance when prompted by an external input. Due to the increasing external input at the cluster nodes, the whole network may become unstable. Several processing components working in tandem may also aid comprehension. Competitive cluster preparation has the potential to enhance individuals' and teams' reactivity to novel situations. The relevant classes would then function together to provide a unique result. When a training competition goes well, the winner processing item's weight is sometimes changed. Due of the lack of knowledge around unsupervised education, further research is required.

2.8 LAWS OF LEARNING (ALGORITHMS)

Learning rules is also often used in many other contexts. When it comes to "Hebb's rules," the prevailing opinion is that it is the one that has been around the longest and is most often used. In accordance with Hebb's Law: Reliability Company’s Donald Hebb was the first to bring this up. As all nerves are relatively strong, removing the neuron from one will increase the pressure on the other nerves (the same sign). According to the Hopfield Law, whether the inputs are active or passive, the analysis rate should be raised. The Delta Rule is used to make constant changes to one's input connections in response to the gap between desired and achieved results (delta). Corrections for delta errors are made using linking weights and the Transfer derivative function, much as in the Delta Law. The eventual weight is determined by the shift factor, but there is another constant that is directly related to the pace

of learning. Kohonen's law states that this is beneficial for determining or adjusting the weight of an object. The victor should be the one who deserves the greatest acclaim from the players and the bystanders alike, and everyone engaged and watching will be carried up in the excitement. Only the victor and their near neighbors may shift the group's balance. Classifier performance is often measured in terms of sensitivity, specificity, and accuracy. Accurately predicting the susceptibility rate is crucial for a number of reasons, including the safety of otherwise healthy people. as a means of relieving the symptoms of epilepsy. Fair equality. In order to calibrate the statistics.

$$Sensitivity = \frac{TP}{TP+FN} \tag{2.1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{2.2}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{2.3}$$

2.9 NEURAL NETWORKS

Classification, in its broadest sense, is the process by which input items, usually those in the actual world, are separated into groupings, or classes, that have already been established. chapter will introduce you to the fundamentals of classifiers and classification, since Support Vector Machine is only one of several classifier machine learning approaches.

As a matter of course, man himself conducts the categorization on a daily basis, usually making use of preexisting norms or acquired knowledge and experience. For instance, he has the ability to divide, or "classify," apples into two groups—the "satisfactory" and "unsatisfactory"—based on his prior knowledge and experience. A fisherman who, after capturing a fish, measures and weighs it to determine whether or not he may retain it before releasing it back into the ocean is another example of a basic categorization. Financial firms utilize complex categorization strategies to evaluate loan applicants' creditworthiness. The bank determines whether or not a customer is at danger of defaulting on a loan by analyzing the client's age, occupation, marital status, monthly income, etc., and comparing that information with the bank's extensive experience with similar matters. It's obvious that a human can't make a judgment based on such a big quantity of information—for example, if we assume that the bank's expertise may represent hundreds of thousands of loans provided

in the past—and that an automated process is required. Hence, the bank's expertise may stand in for a database of information on previous customers, and the automated approach could be a computer program that analyzes information from a new customer and assigns it to the categories of dependable or unreliable.

Artificial neural networks (ANNs) are essentially a learning machine that take their cues from the brain's own neural connections. Created first by Rosenblatt [53], in 1958, ANNs are a simulation of biological neural networks that use artificial neurons (perceptron's). Several inputs are mixed in synthetic neurons before being transferred to an activation function like a sigmoid or threshold. The actions of a neuron determine its "voice".

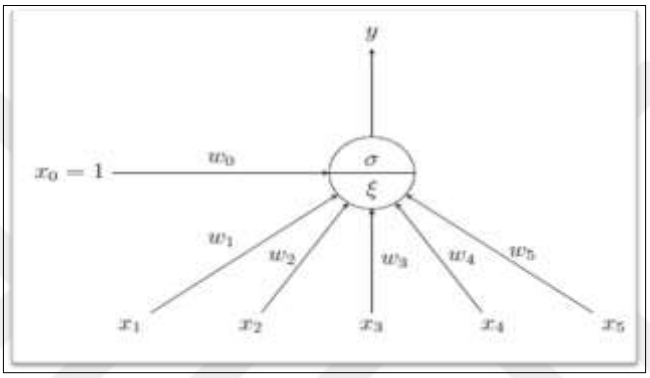

Figure 2.2: The Neuron Unit Connected with Five Inputs and Bias.

2.10 LONG SHORT-TERM MEMORY NETWORKS

One machine learning technique, support vector machine, may be used for both classification and regression. In 1963, vladimir vapnik introduced an algorithm for determining the best division plane for a linear binary classifier, the forerunner of today's widely used techniques. By employing kernel functions, Vapnik and his coworkers were able to implicitly transform a nonlinear classification issue into a linear problem in a different space (the feature space) in 1992 (Kernel trick). There may be a linear dividing plane between the training data samples of the two classes in this space, which tends to have a substantially greater dimension than the original one, and therefore an algorithm may be used to discover the ideal dividing plane of the linear issue. It was in the 1990s that the sum technique was created and put to use for a wide range of signal processing applications, including the recognition of handwritten zip code digits, the detection of faces in images, and the identification of voices. These instances of use are drawn from [1].

In this section, we will introduce the fundamental concept of SVM, which is to search for the ideal dividing plane between two classes while in training. is optimal because it minimizes the distance between the dividing plane and the closest training samples, resulting in a better trained classifier. Next, it will be shown how kernel functions may be used to transparently transform a nonlinear classification issue into a linear one.

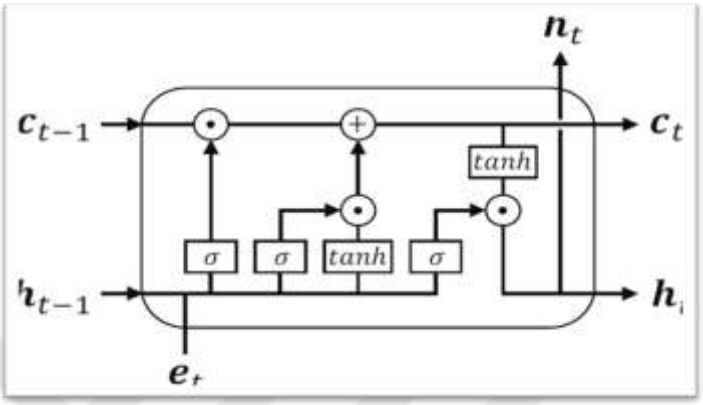

Figure 2.3: LSTM Cell Of the Recurrent Neural Network.

2.11 CONVOLUTIONAL NEURAL NETWORK

We showed how to set up a convolutional layer's settings. the moment has come to learn how these layers may be put to work for you. Making a convolutional classifier is one use for convolutional layers. A classifier is a kind of neural network that is trained on a data set and then predicts which category that set of data falls into. If we had a classifier for cats, for instance, we might tell which breed a given cat belonged to base on its weight, eye color, and tail length, among other characteristics. So how can we utilize convolutional layers to create this kind of classifier? The solution is to provide greater depth. The feature maps produced by the convolutional layers are then flattened in the first layer, which results in a vector. The feature vector is now ready to be fed into a standard fully linked layer. We need to pick how many classes to forecast before we can choose an activation function and the number of neurons. For instance, let's say we're trying to determine whether or not the animal in the image is a dog or a cat, and we need a way to classify it into one of two categories. In this scenario, we employ a single neuron with a sigmoid activation function to make predictions about the likelihood that a picture depicts a cat. Nevertheless, let's pretend we have a different scenario, where we want to figure out what item is in the snapshot from a list of n classes. Then the activation function SoftMax on n neurons is used. Figure 2.4 shows

how the infrastructure of such a CNN might look. In light of our current understanding of convolutional neural networks' fundamental components. Moving on to bigger networks is now possible. Figure 2.8 illustrates the main parts of this kind of CNN. The foundations of convolutional neural networks are now well understood. The transition to gan networks may now be made.

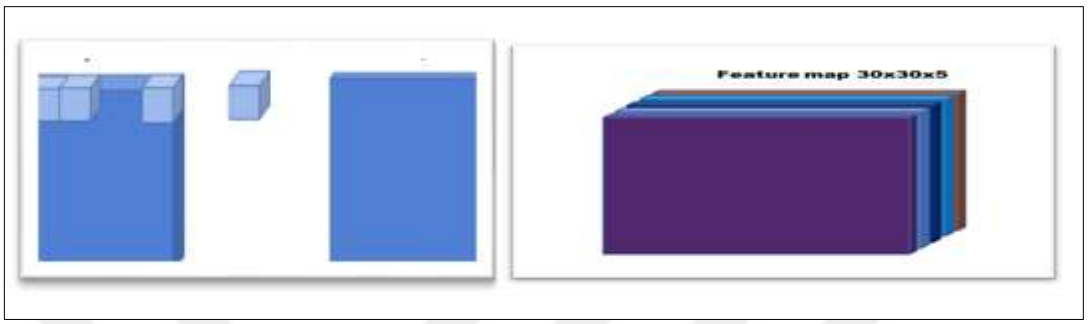

Figure 2.4: Identify the Effect of the Convolutional Layer on the Size of the Data.

a. When a filter of size 3x3x3 "single convolutional (middle)" is applied to the original image "left", the result is a feature map of size $30 \times 30 \times 50$
b. The image of size 30×30 is obtained from five convolutional filters applied to the original image with five feature channels.
c. Max pooling from an original image with size 2×2 resulting in a down-sampled output.
d. An image's effect after pooling.

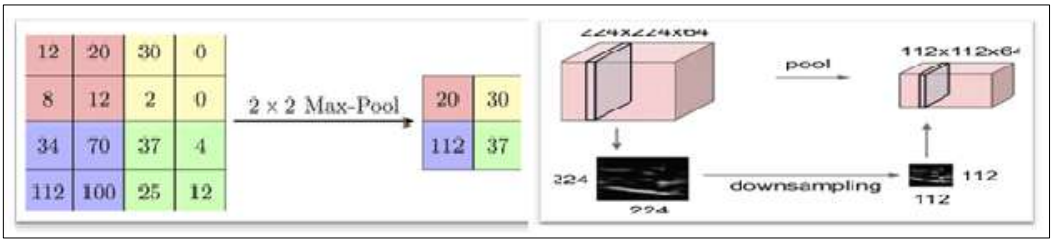

Figure 2.5: Max Pooling.

3. RELATED WORKS

As a result of the challenges involved in detecting distributed denial of service (DDoS) attacks, network and application security has deteriorated, prompting researchers to design and carry out an abundance of studies and tests. Both Mrinal Pandey and Vivek Kumar Sharmar presented findings from research that examined how standardized testing affects students' abilities to recall information while they are enrolled in higher education. After that, network and application administrators conceived of and carried out the plans for the results. As a consequence of this, they utilized four different procedures for decision trees, namely the NBtree technique, the J48 procedure, the Simple cart approach, and the REPtree procedure.

The following is how their plan was carried out:

They started by gathering data from the KDD99 dataset, then they performed analysis on a portion of that data, and then they moved on to the pre-processing step, where they used the J48 classifier. Use the J48 training method to conduct trials and tests for a total of 15 separate classes, which is organized in accordance with the steps of the J48 algorithm.

In order to present their findings, the researchers made use of a classifier known as the J48 (C4.5). to produce a result that is superior than the individual components if they were used individually. Numerous Applications of the Random Forest Model In the fifteen courses that came before this one, identical tasks were solved by using either the J48 decision tree technique or the random forest decision strategy. According to the information in the table that follows [7], the J48 decision tree approach was more successful than the random forest choosing strategy.

On this page, Nanak Chand and Preeti Mishra describe the findings of their earlier research endeavors. During their talk, they discussed the DOS attack discovery process, which included combining SVM theory with an ML methods notebook. IDS has relied heavily on SVM because to the high level of security and precision with which it can identify attacks as quickly as possible. This is especially important for IDS in mission-critical environments. In addition, they looked at the results that were created by combining the SVM with 9 alternative compilations and discovered that the combination of the SVM and Random Forest gave the greatest performance and accuracy of all of the combinations. Which may be determined by utilizing the equation that has been provided for you below [11]:

Accuracy is equal to (TN+TP)/(TN+TP+FN+FP) (3.1)

Training and testing are the two parts that make up the Naive Bayes Multinomial method, which was also created by these individuals. Using CAIDAS data, which frequently contained information about Doss attacks, he was able to track down the Doss attack that internet users had been subjected to, for example on anti-fraud sites that ERAY had researched in the past. This was possible because CAIDAS data often contained Doss attack information. This is helpful for controlling the security of open-source projects.

a. Brother as an illustration IDS has improved the speed of the network system by recording and displaying all network activity gleaned from files, including information on network vulnerabilities. As a direct consequence of this, open network traffic that includes two potentially dangerous detections has surfaced. In order to accomplish this goal, we used a file-based network activity tracking system. When used in conjunction with the scan, Bro v2.2 makes it much easier to identify outliers that are the result of premeditated attacks on uninvolved bystanders. The guys should read this script.

First Steps into the World of Supervised Machine Learning Classifiers In order to construct a classification model, as opposed to the decision tree and the Naive Bayes approaches, this method analyses training data in the form of layers. After that, the fundamental data is entered into the parent contract, and the decision tree then branches out into a contract to test the hypothesis. Naive Bayes is an alternative method to consider. Since it is a linear classifier, it sorts data sets into categories by using a straightforward linear ordering that is determined by the value of the top layer. Bro version 2.2 and Corsaro version 2.0.0 were the two free NIDS programs that the researchers used, as stated in the abstract of the study. Use a method such as the CART Decision Tree, which combines the positive aspects of two different machine learning classifiers. The Naive Bayes model illustrates how successful it is to train with 80% and 20% of the whole dataset, respectively, as opposed to the complete dataset. From 98 percent to 20 percent of the training data set exhibited positive detection findings. The Bro and the Corsaro are both equipped with the capability of leaping right in [12].

Tamara was able to arrive at the conclusion that the attack on Doss was the result of applying the decision tree. This is an algorithm that continually divides the input training set into new sets. The C5.0 method and organisms that are developing for the same class are two of the most well-known components of these subgroups. [C5.0 method] [organisms that are evolving for the same class]

Another method, known as the Random Forest algorithm, was developed by the researcher. This method, too, makes use of entropy as a measuring stick. This technique was created independently, as can be shown in the table that follows: Before generating its individual nodes, each tree must first choose a training set to use as a template. We generate a collection of qualities using a scale that was determined in advance, and we do so use a random method. As a consequence of this, we choose the feature that offers the most potential benefits from each stage, then we split, repeat, and divide once more until there is only one group left. According to what was reported in the press.

The author of the research stressed how vital it is to make use of frequent repetition in the ML. These repetition analyses are not only useful for delivering high-quality information in the form of an engineering framework, but they also serve as a tool for locating previously unknown connections. In my research, I looked at two different types of time series, and the results showed that regardless of whether or not an attack took place, a certain number of traffic milestones were always achieved. An assault denial of service (DoS) creates traffic as part of a larger series of training and testing that occurs over the course of time. The exponent, H [0.525, 0.575] [0.575, 0.625] [0.925, 0.975], was independently confirmed by the Hurst group via the use of experimentation, and it has remained the same over all 10 time periods.

The Unified Distribution Law uses a random selection process to choose a Hurst distribution to apply inside the confines of a single interval. According to the compiled experimental data that can be found in [13], the Random Forest technique generated the most successful results.

They suggested that the researchers utilize the DATA MINING ENGINE to spot the doss assault while it was still in its early stages. Cybersecurity is a complex and multifaceted field, therefore they made this recommendation. In addition, they stated that by developing it, it would be capable of analyzing data and requests that the server administers based on previous data, comparable to the CPU in a DATA MINING ENGIN that accumulates large quantities of information. This was stated in the context of the development of the technology. The following procedures are carried out by DATA MINING ENGIN in order to get the information that is required and provide Hlgom with a general response. The execution of these tasks is accomplished by applying various classification algorithms to large datasets as well as traffic patterns. The beginning of the Clustering Process, Phase 1

Putting the pieces of the puzzle together Making Use of Models Placement in the Order of Priorities

After gathering the information, they organized it according to the kind of data and the location from which it originated. After that, an analysis of the data was carried out, and the results of that were included into the clustering method. The parameters that were gathered through the clustering process were used. We are provided with a variety of different pathways that enable us to access the analysis as quickly as possible and on a wide range of time-structured data in order to facilitate the extraction of the primary components that make up the clustering process. This is done so that we can assist with the identification of the most important aspects of the clustering procedure. In order to construct an algorithm that can accurately extract genuine features and behavior in response to attacks, it is essential to do an analysis of the request's permissible activity. In conclusion, the objective of the distributed system is to identify both the original requests as well as the assaults once an attack has been identified and a response has been sent. The team conducted a literature assessment on studies that made use of DATA MINING ENGINe to achieve both of these aims [14].

Throughout the course of their investigation, Shital and Vaishali offered a suggestion for a method or algorithm that would be able to aid in improving the capture data in order to minimize data extraction processing and attack detection. Initially, raw capture packets would be stored in the field designated as D. The whole dataset was then subjected to the following approaches that they performed: It is reasonable for you to assume that the process that brought you to each line containing delimiters has been finished by this point. The primary goal is to develop a pre-processing that will increase the accuracy of algorithm identification. This will be accomplished by removing the delimiter, which is any symbol or combination of symbols that is used to split a string of texts, such as (,), (;), etc..., as well as the final word, which will be something like (so) delete.

After the completion of the data gathering program, the output was fed into the pre-processing algorithm that had been developed specifically for it. The Naive Bayes Multinomial classifier was then used to categorize assaults, As a consequence of its superior accuracy when compared to other approaches, this classifier was chosen for the classification.

In addition to the host Chuanhuang, Yan, Xiaoyong, Zhengjun, and Weiming all deployed deep learning models to detect DOSS intrusions. These models are often referred to as the "brain" of the SDN. The Convolutional Neural Network (CNN) was the first model to use deep learning; it consists of input layers, forward recursive levels, and output layers. It's possible that CNN will map each nuran of a local feature by using the idea of a local receptive field. Since it is customary, every nuran is responsible for carrying the same load. The functionality of pooling is determined by the scope of the feature. RNNs, also known as recurrent neural networks, make up the second building component of the system. It linked the hidden layers by using the information from the input layers, and after that, it carried out its calculations throughout the whole of the information contained in the hidden layers. They put four distinct deep learning strategies through their paces by using the ISCX2012 dataset as a test bed for their accuracy evaluations. They conducted research and found that 3LSTM deep learning provided the most accuracy; consequently, you need to make use of it.

Xiaoyong,Chuanhuang, They suggested the creation of a Deep Defense system that could identify doss attacks that were carried out via RNN, such as (LSTM,GRU).

In order to study the potential applications of RNN, the researchers picked 20 network traffic indicators from the ISCX2012 dataset. Recognition of spoken words and translation across languages are two examples of these applications. After that, the boolean fields were hashed using the hashing technique, and then transformed to binary format using the Bag of Words (BOW) approach. Think about the number assigned to the TCP/UDP port. They built a Bidirectional Recurrent Neural Network using this architecture, which had two sequence-to-sequence Recurrent Neural Layers in each direction.

Over the course of this endeavor, LSTM and GRU were both put to use.

LSTM was developed as a solution to the vanishing gradient issue that plagued RNNs. This allowed GRUs to be trained more quickly and with a less amount of data. the results of the Random Neural Network theory perform better than those of the Random Forest theory.

In the earlier work, Tran developed a hybrid approach that integrated SOM and k-NN. This approach was primarily geared for the SDN. A Self-Organizing Map is the name given to this kind of operation (SOM). In the area of artificial intelligence, the Self-Organizing Map, or SOM, is a method of unsupervised learning that functions in a manner similar to that of a neural network. The term "self-organizing map" is referred to by its abbreviation, "SOM." Which Kohonen was gracious enough to provide as a solution. The self-organizing map

(SOM) idea may be thought of as a map of neurons due to the fact that it is so efficient at gathering high-dimensional data. The ratio of the distance between two SOM vector inputs determines the number of weights that are included inside each cell.

The K-Nearest Neighbor theory performs well when applied to low-dimensional characteristics, in particular when it comes to the detection of a wide variety of DoS assaults. The researchers came up with five general criteria that may be used to data analysis and categorization. The following are some of these characteristics: • The randomness of the IP address used as the source (etpSrcIP). The entropy of the point of origin port (etpSrcP). Entropy of the port that is being targeted (etpDstP). The unpredictable behavior of the packet protocol's entropy (metaprotocol). The sum total of all of the individual packages (total Packet).

The neuron weights of the SOM include a representation of each of the aforementioned characteristics in the form of a dimensional vector. In order to standardize the data that was relevant to the aforementioned criteria, the use of z-scores was utilized. After the phase of preprocessing, the SOM + K-NN classification approach was used in order to identify a Doss attack in the manner that was specified.

After that, the SOM method need to be implemented in order to classify the distributed neurons.

It is included into each feed that is obtained from SOM. As a consequence of this, there must be cells that are able to prevail (BMU). It is possible to achieve this goal by determining the offset of the weight vector of each SOM cell with respect to its inputs. They recommended combining SOM with a center-distributed classification algorithm as a solution to this challenge in order to address it. After the conclusion of the SOM training and the selection of the median vector points, distances are computed between each input and a particular point denoted by the letter G. G= (g1, g2, gn). The following actions are taken inside SOM in the event that a center-distributed categorization is utilized: In conclusion, the results produced by applying the two innovative views to the same data performed much better than those obtained by using standard theories to the same information [16].

Chenggang came to the conclusion that the RDF-SVM theory, when coupled with the RDM theory, would be more useful than the SVM theory, which re-screens features, and the RDM theory, which calculates the significance of variables. By combining the two strategies, he was able to maximize his advantage while maintaining a high rate of detection throughout

the process. The notion of a random forest begins with randomly selecting samples, continuing with the random division of those samples in order to construct a decision tree, and then concluding with voting to choose which samples belong to which categories. Either utilize the Support Vector Machine (SVM) methodology as a second classification strategy that works well by addressing the imbalanced variable data set and has been picked to recheck the features; or use a different classification method. Since it is possible to remove a feature from the classification by accident while using the RDF-SVM methodology, the feature in question was reexamined by making use of the SVM theory. In order to determine which characteristics are most important for the procedure, the Random Forest hypothesis is used.

After the completion of my training and testing procedures on the KDD99 database, which includes 42 normal and attack dataset characteristics, I came to the following realization: As a direct result of this, there were a total of 39 different kinds of assaults found.

In order to have a better result from the finding of this theory, it has to be compared to a number of other theories employing the RDF-SVM theory to perform training and testing processes in order to get those procedures done. Not until then will we be in a position to produce a more favorable result.

The feature subset consists of 1, 0, 3, 10, 11, 8, 12, 13, 9, 6, 2. Despite this, both the RDF and SVM approaches failed to recognize the second feature. In addition, as can be seen in the figure that follows, it was the hypothesis of doss dataset that was the most effective in terms of testing. [18]. As a direct result of the rise in the frequency of distributed denial of service (DDoS) attacks, which attracted a great deal of attention from academics, those same experts have been very busy formulating a plan to defend cloud computing from attacks made using HTTP HTTP Cloud. Recently, precision was estimated by using a contemporary database of DOSS assaults that had been gathered at multiple network levels. These assaults included SIDDOS and HTTP flood. From this database, 27 characteristics were retrieved and compared using multi-layer machine learning and Random Forest. Doss has made public his attacks on the OWNCLOUD infrastructure, which offers users an alternative that is free and open source. It is possible for the SNORT intrusion detection system, which uses criteria to identify attacks, to recover any and all data taken from a server. The amount of time since a file was last modified and its permissions are two of the criteria for these rules. If the

appropriate rows are chosen, SNORT may provide direct output if necessary. In order to acquire all nine characteristics of the dataset, the following schedule was adhered to:

The classification of random, forest, naive bayes, and support vector machine data was accomplished with the use of three different machine learning theories. These theories were chosen for their relative value in determining whether or not an answer is right. Pattern recognition and security breach detection may be accomplished using a method known as the Support Vector Machine (SVM). Naive Bayes is the approach that ultimately chooses categorization and learns to predict, while RF is utilized for classification and regression because to its flexibility and simplicity of use. Naive Bayes is the strategy that ultimately learns to forecast. It is essential to the success of the Naive Bayes method that the probability you estimate for each class be as precise as possible. According to the findings, SVM was able to achieve the maximum accuracy in calculations, while both Naive Bayes and Random Forest both did quite well [18]. The gathering of data led to the discovery of this.

4. DENIAL OF SERVICE ATTACKS

A DOS attack, or Denial of Service connection, means denial of service. There is also often talk about so-called DDoS attacks (Distributed Denial of Service), which are DOS attacks conducted from multiple sources. The consequence of these attacks is disruption or complete inaccessibility of the service for its regular users. The unavailability of services and systems causes significant financial losses to their operators, damage to the company's reputation, or the need to analyze and correct the problem. If we consider that most of today's companies and organizations are highly dependent on network services, we are dealing with a big problem. The popularity of these attacks among attackers is growing considerably. No in-depth knowledge is required for their implementation these days, as suitable tools for their implementation are commonly available. The fact that disrupting the operation of a network or system is often much easier than gaining access to it is just another reason for their expansion [12]. There are many motives for carrying out an attack. It can be, for example, attacks on competing companies with the aim of damaging its reputation or otherwise weakening its position on the market. In the media, we also often come across attacks that are conducted as a certain form of demonstration. In these cases, the target is most often government networks and systems, or groups and companies threatening freedom of speech. Last but not least, DDoS attacks are used as a terrorist threat. Nowadays, there are a huge number of different types of DDoS attacks. Deficiencies in the architecture and operation of TCP/IP protocols, which at the time were designed for use in an open and trusted environment [12], are often exploited. Attacks are also aimed at occupying the transmission capacity of the line, exhausting system resources, errors in programs, packet routing systems, or DNS. Attacks can also be specified and divided according to the TCP/IP layers through which they operate. In our case, we will focus on a brief analysis of those DDoS attacks provided by the Avalanche 3100 device.

4.1 ANALYSIS OF AVAILABLE AVALANCHE ATTACK DEVICES

The Avalanche device from Spirent Communications offers, in addition to load testing (chapter 3), also the possibility of generating DDoS attacks. The characteristics of individual types of attacks are based on sources [13][12] and from the help of the Test Center Layer 4-7 Application program.

a. ARPFlood attack: It is an attack on computer networks and network devices, attempting to exploit their limitations in the area of managing ARP (Address Resolution Protocol) messages and their caching. With this attack, the Avalanche device generates ARP messages to the targeted device from a range of virtual source addresses. It also allows you to set the type of generated ARP messages (ARP request or ARP reply).
b. Evasive UDP attack: The attack generates a large stream of UDP datagrams with a variable size and a random source IP address. If datagrams are sent in large numbers, they can cause disruption to the victim's system.

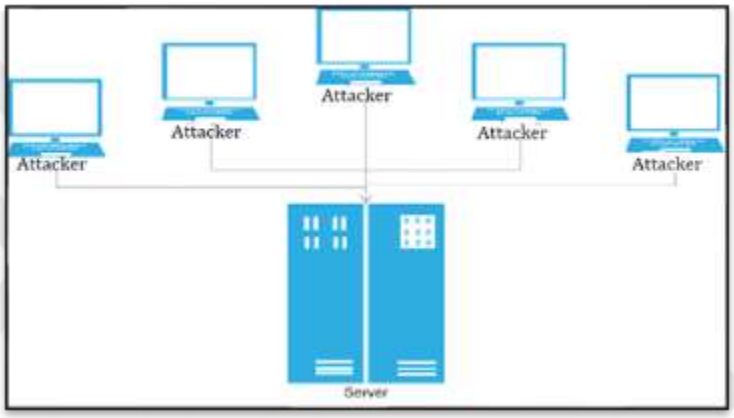

Figure 4.1: Dos Attack Scheme, Source**.**

c. Land attack: The core of the attack consists in forging a TCP SYN packet so that the source IP address and source port of the packet are the same as the IP address and port number of the victim's device. These spoofed packets are then sent to the victim's device, which starts trying to establish a connection with itself, which can cause it to crash.
d. Ping of death attack: This is the sending of an ICMP (Internet Control Message Protocol) packet with a length greater than 65536 bytes, which is the maximum allowed by the TCP/IP specification. For network transmission, the packet is fragmented and sent to the appropriate destination. The target system then tries to assemble the packet into the original, unauthorized size from the received fragments, which can lead to buffer overflow and system freezes or crashes.
e. Ping sweep attack: The ping sweep attack generates ICMP ping requests to a range of target network addresses. It is used to identify available network devices on the network. An attacker can thus create a map of active network devices that can be attacked.

f. Random unreachable host attack: The attack consists in sending ICMP error messages about the unavailability of the target to various network devices. The intent of the attack is to make the victim's system drop established connections based on error messages.

g. Reset flood attack: A sequence of TCP packets is generated on the victim's device, which have the task of resetting or terminate real active TCP connections.

h. Smurf attack: This attack generates an ICMP packet with a ping request (echo request) to the broadcast address of some network (e.g. broadcast address) with a forged address of the sender, which refers to the target of the attack. All computers on the network will respond to the ping, flooding the target. The network here acts as an amplifier of the original request.

i. Syn flood attack: This attack uses the basic principle of the TCP protocol when establishing a connection, the so-called "three-way handshake". During an attack, a SYN packet with a spoofed IP address is sent to the target device (e.g. server). The server reserves system resources for the potential connection and responds with a SYN/ACK packet to the sender to the spoofed IP address. However, there is no response from that, so the server sends the message again. After a certain amount of time has passed without a response from the other party, the server de-allocates system resources along with a record of the original connection initialization. If a large number of these packets are sent to the server, the number of "half-opened, unfinished" connections will increase, draining the server's system resources and stopping it from servicing new legitimate connection requests.

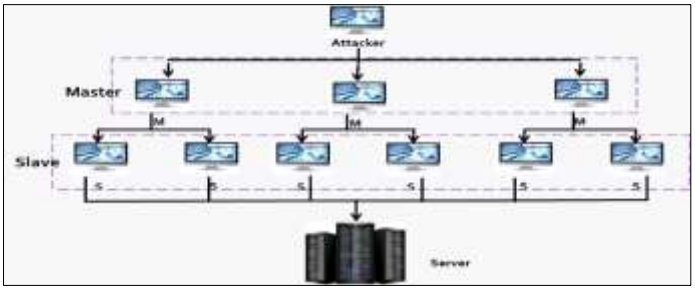

Figure 4.2: Distributed Denial of Service Attacks, Originating from Internal Processing.

j. TCPP or TScan attack: The TCPP or TScan attack generates a sequence of TCP SYN packets with different destination TCP ports on the victim's device. The intention is to find out on which ports TCP services are available. Port scanning is also used to exhaust

the device's system resources, which makes it impossible to establish new TCP connections.

k. Teardrop attack: Similar to the Ping of Death attack, the Teardrop Attack tries to cause the target system to collapse by sending multiple fragments that cannot be assembled correctly. Practically, it is a fragmented IP packet, the individual fragments of which overlap, thereby overwriting each other.
l. UDP flood attack: The attack generates a large number of UDP datagrams, which consumes network bandwidth and prevents network devices from establishing new UDP connections. The trick is also that UDP datagrams are sent to random transport ports of the target device. The device tries to find out what application is listening on these ports. When it detects that no application is listening on the given ports, it sends an ICMP reply about the destination being unreachable. The device must perform this overhead for each UDP datagram sent. The attacked device thus runs out of resources to establish new UDP connections.
m. UDPP or TScan attack: The UDPP or TScan attack generates a sequence of UDP data packets with different UDP destination ports on a given device. It is mostly used to identify UDP ports on which UDP services are available. Port scanning can also be used in a similar way to the UDPFlood attack to exhaust system resources and subsequently make it impossible to establish new UDP connections.
n. Unreachable host attack: This attack sends ICMP target unavailability error messages to the victim's device. The aim of the attack is to make the victim's system drop established connections with the target based on error messages.
o. Xmas tree attack: The Xmas tree attack generates a sequence of TCP packets with some set attributes (URG, PSH and FIN at the same time) used in the TCP header. Packets are aimed at the target of the attack and generated from a large number of IP addresses. This exploits a bug that exists in some network devices. The result is a system crash after receiving a Xmas tree packet.

4.2 ANALYSIS OF THE CURRENT STATE OF DDOS ATTACKS

Orientation in the area of the current state and development of DDoS attacks is quite crucial from the point of view of preventive protection. The final security reports of companies that analyze Internet traffic and develop security protection against DDoS attacks are very useful

in this regard. These security reports can provide us with valuable information about the trends, methods and tactics of DDoS attacks that currently pose the greatest threat. The following text is based on the current security reports for 2013 of three global companies1 dealing with this issue [14][15][16].

a. The most common types of attacks: From 2012 until mid-2013, the most common attacks were TCP flood (also known as SynFlood) and HTTP flood. The popularity of these attacks suddenly decreased in the second half of last year, and DNS flood became the predominant attack, accounting for more than 50% of the total number of attacks carried out in the second half of 2013. This change is attributed to increased protection capabilities against TCP flood and HTTP flood attacks and their increasingly frequent implementation in enterprises, hosting companies and data centers. The consequence is that cybercriminals target the DNS infrastructure, which remains one of the weakest links in the network and is therefore vulnerable to DDoS attacks. The percentage representation of the most frequently used DDoS attacks is shown in figures 4.1 and 4.2. The first of them is the result of the analysis for the first half of 2013, the second of them corresponds to the situation in the second half of 2013. The past year has also seen an increase in hybrid attacks, which consist of attacks based on different weaknesses in different protocols. A hybrid attack based on the TCP protocol has become the most popular. The second most common hybrid attack is based on ICMP, TCP and UDP protocols. The third place was then occupied by an attack using simultaneously ICMP, TCP, UDP and DNS protocol weaknesses. At the same time, the fact that DDoS attacks are used not only as a means of disrupting a certain service, but increasingly also as a kind of bait that aims to divert attention from another attack or data theft is revealed.
b. Duration of attacks and their size: The duration of most DDoS attacks in 2013 was 30 minutes or less. A similar trend was also noted in 2012. Attacks lasting less than 30 minutes accounted for 93.2% of all attacks recorded in the first half of 2013, and in the second half of the year it was 86.2%. In addition, the number of attacks that were shorter than 15 minutes also increased by almost 50% when comparing the first and second half of the year. This phenomenon is attributed to the shortening of the response time to an ongoing attack thanks to systems for diverting data traffic or protection against DDoS attacks. However, it should be noted that the strongest and most serious attacks recorded fall into the category with a duration of 8 hours or more. An increase of 4% was recorded

in this category during 2013. We can see the results of the duration of DDoS attacks in figures 4.3 (for the first half of 2013) and 4.4 (for the second half of 2013).

Regarding the data flow size of DDoS attacks, the most common are attacks with sizes up to 50 Mb/s. This is 80% of all attacks. These are mostly application layer attacks or hybrid attacks that have the potential to cause significant disruption for systems that are not prepared for similar types of attacks.

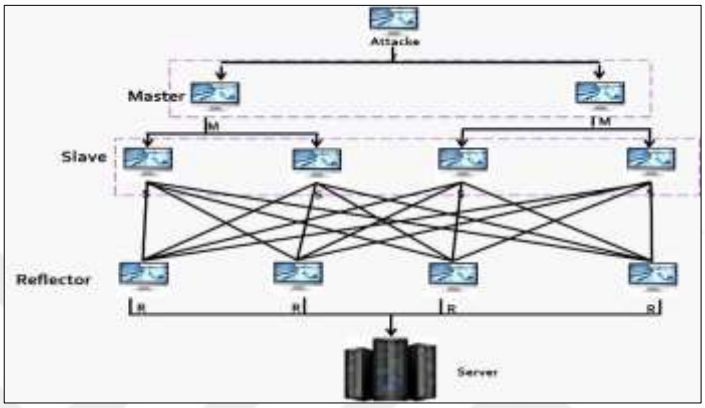

Figure 4.3: DRDOS Attack Distribution Scheme, Source: Own Processing.

c. The most common targets of attacks: The most attacked targets during 2013 can practically be divided into two main groups: government networks and online services. In the first of them, the motivation of the attacks is based on political or ideological motives, in the second group, the main motive is financial gain, competitive rivalry or even extortion. By monitoring DDoS attacks on a global scale, it was found that in the first half of 2013, banks and their services were the most frequent targets (43.3% of all attacks), while in the second half of the year, government and non-profit organizations took first place (49, 3% of all attacks). In comparison, attacks targeting banks fell to 6.5%. One of the possible explanations for this phenomenon is that the banking sector significantly strengthened its DDoS protection from the second half of the year after the first three stages of the massive DDoS attacks Operation Ababil (about this attack, which took place in the USA and is considered one of the most serious worldwide attacks is discussed in more detail in the Radware security report [14]). The most frequent attack targets for the second half of 2013 are clearly shown in Figure 4.5. In terms of attacks on devices, we most often encounter attacks on elements of backbone networks, firewalls and servers. Another interesting finding is the fact that more than 50% of DDoS attack

victims are attacked more than once. Usually there are two to ten attacks during the calendar year.

4.3 GENERAL DESIGN OF PROTECTIVE EQUIPMENT

We cannot prevent the attacks themselves, but nowadays there are a number of measures and tools to prevent attacks or to minimize the damage as much as possible.

At the beginning of the design of protected means, it is necessary to set individual goals, that is, to analyze what risk a DDoS attack represents for us and what financial or other damages may be its consequence [17]. Based on this, it is necessary to consider how many resources (finances, time, human resources) we are willing to invest in the measure. During the actual design and implementation of protective measures, we can proceed from individual systems, through the protection of our own network, to cooperation with the connection provider and the National Center for Cyber Security.

Against individual types of attacks, various measures are proposed at the level of devices or systems to defend against them to a large extent. An example can be the setting of SYN cookies on a given device or system as a measure against a SYN Flood attack (RFC 4987 [18]). We can apply similar settings for other types of attacks. These individual measures should not be underestimated, as the target of an attack can be any element of the network infrastructure. It goes without saying that the system should always be updated and patched, meeting the relevant RFC recommendations.

As part of the protection of the internal network, we can reach for specialized network elements whose task is to analyze network traffic and, in the event of an attack, detect it in time and stop or limit it. These devices mainly include various application firewalls, IPS (Intrusion Prevention Systems) or IDPS (Intrusion Detection and Prevention Systems) and others. Today, many companies focus on the development and sale of these devices. However, no solution provided is 100% perfect, and it is necessary to consider which product from which manufacturer is the most suitable for the network in terms of specifications (if at all). An important aspect in the fight against DDoS attacks is to distinguish them from legitimate traffic, which may not always be so simple. In many cases, even valid user traffic can show characteristics of DDoS attacks. An example can be a limited-time discount event of a popular international web portal, which suddenly becomes the target of a huge number of users (in principle, this is an HTTP flood attack).

A subsequent matter for protection against DDoS attacks is mutual cooperation with the Internet connection provider. Agreeing on certain crisis procedures and precautions helps to prevent a large part of the attack. Protection against massive attacks is also being fought at the national and transnational level. An example can be, for example, the Czech national internet node NIX.CZ, which is working on protection against large DDoS attacks and is planning more significant security of the domestic internet network [19]. Another such example is the NCKB (National Cyber Security Center), which recently 22 November 2013 published a document aptly titled "Recommendations in case of a DDoS attack - how to behave and how to proceed", which, among other things, deals with cooperation between the victim and the CERT/CSIRT security workplace [20].

Every network infrastructure is unique in its own way. When designing a specific measure against DDoS attacks, a deep knowledge of the issue and orientation in the given infrastructure is therefore needed. At the same time, it is always necessary to consider how much resources we are willing to invest in the measure and whether it will bring us the desired result. However, protection against DDoS attacks always needs to be paid extra attention these days.

5. IMPLEMENTATION AND RESULTS

5.1 INTRODUCTION

A machine learning experiment in this situation, as illustrated in figure 5.1, often follows an all-purpose template, which may be summarized in the following seven phases:

a. Dataset Collection.
b. Dataset Pre-Processing.
c. Feature Extraction.
d. Model Training.
e. Evaluation.
f. Classification.

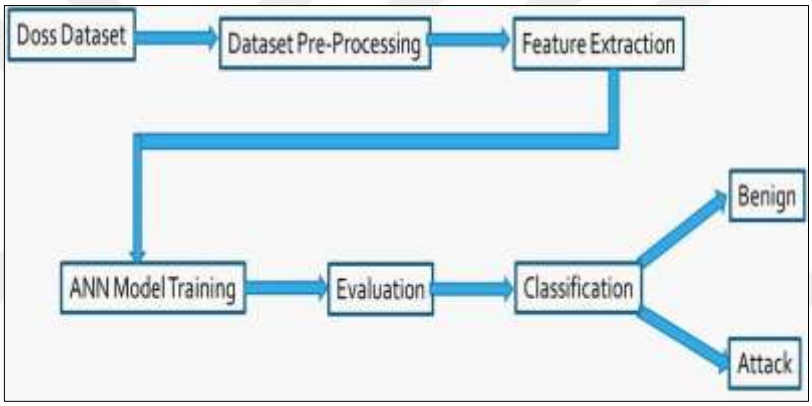

Figure 5.1: Research Methodology.

Prior chapters have covered the first two requirements. As a result, we'll be discussing those remaining five's methods in this section. Notably, even while these phases are often implemented in the sequence described above, it is normal for individual processes to be linked to some degree.

5.2 DATASET

So that machine learning and deep learning techniques may be applied to the task of categorizing traffic, this data set was compiled using the SDN Mininet simulator found in references [42]. Ten Mininet topologies are first built using switches linked to a Ryu controller. The TCP, UDP, and ICMP traffic seen in simulated networks is not always malicious (TCP Syn, UDP Flood, and ICMP attacks). Some of the 23 properties in the dataset were gleaned from the switches, while others were computed. The gleaned

characteristics are detailed below. In this case, we'll be using the following notation: "Source IP, Destination IP, Source/Destination IP, Source/Destination IP Port, Source/Destination IP, Source/Destination IP, Source/Destination IP, Source/Destination IP, Source/Destination IP, Source/Destination IP." The data bytes used for sending and receiving, or "tx bytes" and "rx bytes," are identical. A flow check is performed every 30 seconds, and the dt field is used to store the resulting numerical representation of the date and time. Port Bandwidth is the product of the transmit and receive data rates in kilobits per second (tx kbps and rx kbps, respectively).

The traffic class designation is shown in the last column (good or bad). A label of 0 indicates malicious traffic, whereas a label of 1 indicates benign traffic. A 250-minute network simulation produces 1,043,45 rows of data. The simulation runs for a predetermined period of time.

5.3 GREY WOLF OPTIMIZER (GWO)

This algorithm was proposed by Mr. S. Mirjalili and Mr. A. Lewis in 2014 [69,70]. GWO is based on the social hierarchy of the common wolf (Canis lupus). The mathematical model considers the best solution (optimum) as alpha (α). This means that the second and third best solutions will be beta (β) and delta (δ) according to the wolf hierarchy. Other solutions are considered omegas (ω). Optimization (hunting) in GWO is led by these three wolves, i.e. α, β and δ. The Omega Wolves follow them. The algorithm works on the following principles:

a. Encircling prey: When hunting, wolves first encircle the prey. Mathematically, this behavior is modeled by the equations: The first, second, and third responses are kept, and others examine the quest representations of the formulae There were alternatives for the first, second, and third places. Where a is a component linearly decreasing from 2 to 0 within iterations and $\vec{r}1$ and $\vec{r}2$ are random vectors (vectors of random numbers), in the ranges [0,1]. The position of the wolf can be updated depending on the positions of the prey. Various positions around the best agent. Different positions around the best agent in terms of the current position can be obtained by adjusting the values of the vectors A and C. Random vectors again they allow the wolf to reach any position in space.
b. Hunting: Alpha, beta and delta wolves have the best information about potential prey. This means that these three best solutions are saved and other agents (including omega

wolves) commit to update their positions depending on the positions of the best agents. In this sense, the following formulations are proposed:

c. This means that alpha, beta and delta estimate the position of the prey and the other wolves update their positions randomly around it.
d. Attacking prey: Mathematically, approaching prey is modeled by decreasing the value of a. When it decreases, the fluctuation range A also decreases. This means that A will be a random variable in the interval [-2a, 2a],
e. where a decreases from 2 to 0. For a value of A in the range [-1,1], the next updated position will be between the agent's current position and the prey's position. So as long as $|A| < 1$, wolves attack to meet prey.
f. Search for prey: The search process starts with the creation of a random population of wolves (candidate solutions). Each candidate solution updates its distance from the prey. The parameter a is reduced from 2 to 0 in order to emphasize attack and search. Candidate solutions diverge from the prey if $|A| > 1$ and as already mentioned in the previous principle, they converge to it as long as $|A| < 1$. The search is therefore realized by the convergence/divergence of the wolves towards each other to attack the prey. The parameter C adds a random weight to the prey. Its implementation is useful for searching and prevents stagnation in the local optimum, especially in the final iterations. Figure 14 shows the described principles in the form of pseudo code.

5.4 PARTICLE SWARM OPTIMIZATION (PSO)

Particle swarm optimization (PSO in short) is a swarm algorithm that is inspired by the swarming of birds or fish. This intelligence is based on a numerical optimization algorithm introduced in 1995 by Mr. James Kennedy and Russel Eberhart [46][47][48]. Social animals group together for a variety of reasons, and flocks or groups are always advantageous for the survival of individuals for several reasons:

a. Foraging: Individuals of the swarm can benefit from the discoveries and previous experiences of other individuals in the search for food.
b. defines against a predator: A group of animals (eg a flock of birds) can defend itself more effectively against a predator. More eyes mean a greater chance of spotting danger, the amount of potential prey in the herd reduces the danger to the individual, and a group of animals is able to confuse the enemy. Of course, this behaviour also brings disadvantages,

which are not simulated by PSO (e.g., weaker individuals do not die, as with the genetic algorithm). In PSO, all individuals survive and try to:

a. Proximity principle: The population should be able to perform simple spatial and temporal calculations.
b. Quality principle: The population should be able to respond to quality factors in the environment.
c. Diverse response principle: The population should not engage in the same response.
d. Stability principle: The population should not change its behaviour style every time the environment changes.
e. Adaptability principle (Adaptability principle): The population should be able to change its behaviour as long as it is advantageous in terms of computational cost.

With PSO, the solution is obtained through random search (Random search), which is equipped with swarm intelligence. This means that PSO is a swarm intelligent search algorithm. The search is carried out by randomly generated potential solutions. This collection of potential solutions is called a swarm, and each potential solution is called a particle.

Browsing is influenced by two ways of learning. Each particle learns from another particle, which represents the so-called social learning and at the same time from her own experiences, the so-called cognitive learning. Thus, thanks to social learning, the particle obtains the best solutions visited by any particle in the swarm, called gbest (global best).

Through cognitive learning, it reaches the best solution that each particle has found and visited so far, called pbest (personal or particle best).

Each particle has its own trajectory, called position xi and velocity vi, moving in the searched space by gradually updating its trajectory. Particle populations adjust their trajectories based on pbest and gbest positions. All particles have a fitness value, which is evaluated by an objective function designed for optimization. The particles move in the solution space by following the optimal particles. The algorithm initializes a group of particles at random positions and then searches for optima by updating generations. In each iteration, the particle is updated based on the two best positions pbest and gbest. In D-dimensional space, the position and velocity of the i-th particle in the time step t are represented by a D-dimensional vector, $xt = (xt, xt, \ldots, xt)\ T$ and $vt = (vt, vt, \ldots, vt)\ T$.

5.5 RESULTS

Following the use of standardization method as a form of preprocessing approach on the dataset, we will explicitly use the mental imbalance dataset in this inquiry to test our hypotheses using ANN, ANN-PSO, and ANN-GWO on handled datasets.

The arrangement exactness of ANN, ANN-PSO and ANN-GWO were gathered a few times in the wake of making the concealed layer 5 with 1000 emphasess. Besides, looking at the MSE of all chose shrouded layer esteems shows that the MSE of the concealed layer with the worth - is considerably less than the others.

Table 5.1: Performance Comparison Among Different Algorithms with 5 Layers.

Algorithm	Training Error	Training Accuracy	Testing Error	Testing Accuracy
ANN	0.013	91.68	0.010	91.97
PSO	0.011	93.35	0.010	93.45
GWO	0.009	95.74	0.008	95.83

Table 5.2: Performance Comparison Among Different Algorithms with 10 Layers.

Algorithm	Training Error	Training Accuracy	Testing Error	Testing Accuracy
ANN	0.012	91.70	0.013	91.63
PSO	0.011	93.35	0.010	93.45
GWO	0.009	95.79	0.007	95.90

Table 5.3: Performance Comparison Among Different Algorithms with 15 Layers.

Algorithm	Training Error	Training Accuracy	Testing Error	Testing Accuracy
ANN	0.023	91.69	0.017	91.21
PSO	0.011	93.35	0.010	93.45
GWO	0.009	95.73	0.008	95.87

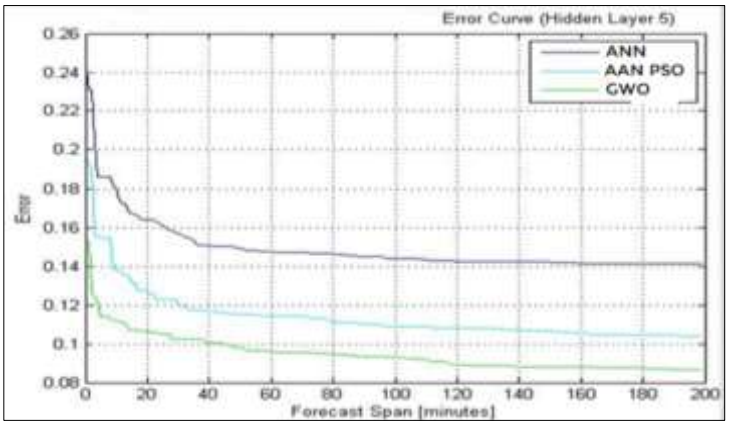

Figure 5.2: Error Curve for Algorithms on 5 Layers.

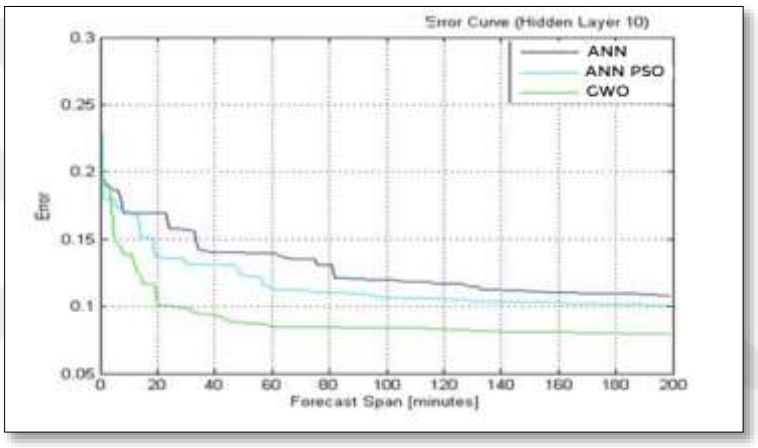

Figure 5.3: Error Curve for Algorithms on 10 Layers.

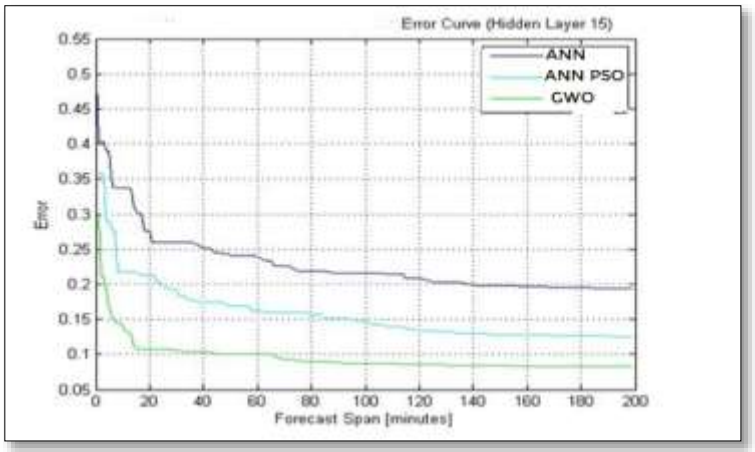

Figure 5.4: Error Curve for Algorithms on 15 Layers.

6. CONCLUSION

For the purpose of this research, a neural network framework for the diagnosis of autism spectrum disease was created using three different optimization strategies. According to the results, which can be seen in Table 4.3, the ANN-GWO surpassed the other categorization approaches in terms of the quality of the output it produced.

In this research, we propose three conditions for neural connection that need to be met in order to advance the area of feed ANN. These criteria make it possible for neurons to communicate with one another across layers. Moreover, these criteria might be used to build horizontal connections between neurons that are located in the same layer. Not only did we find out that the suggested method did not link those Input Neurons to various ANNs, but we also found out that this was not the only issue that needed to be addressed. This indicates that the neural technique was not used in the evaluation of the ANN output. This process is known as dimension reduction, and it is used in input models. The challenges posed by silicone and the identification of objects generally result in a weighted detection rate of one hundred percent; however, the challenges posed by glass and spiral have proved to be less effective.

REFERENCES

[1] G. Beni and J. Wang, "Swarm intelligence in cellular robotic systems," in Robots and Biological Systems: Towards a New Bionics, vol. 102 of NATO ASI Series, pp. 703–712, Springer, Berlin, Germany, 1993.

[2] X. Yao, "Evolving artificial neural networks," Proceedings of the IEEE, vol. 87, no. 9, pp. 1423–1447, 1999.

[3] E. Alba and R. Martí, Metaheuristic Procedures for Training Neural Networks, Operations Research/Computer Science Interfaces Series, Springer, New York, NY, USA, 2006.

[4] J. Yu, L. Xi, and S. Wang, "An improved particle swarm optimization for evolving feedforward artificial neural networks," Neural Processing Letters, vol. 26, no.3, pp. 217–231, 2007.

[5] M. Conforth and Y. Meng, "Toward evolving neural networks using bio-inspired algorithms," in IC-AI, H. R. Arabnia and Y. Mun, Eds., pp. 413–419, CSREA Press, 2008.

[6] Y. Da and G. Xiurun, "An improved PSO-based ANN with simulated annealing technique," Neurocomputing, vol. 63, pp. 527–533, 2005.

[7] X. Yao and Y. Liu, "A new evolutionary system for evolving artificial neural networks," IEEE Transactions on Neural Networks, vol. 8, no. 3, pp. 694–713, 1997.

[8] D. Rivero and D. Periscal, "Evolving graphs for ann development and simplification," in Encyclopedia of Artificial Intelligence, J. R. Rabuñal, J. Dorado, and A. Pazos, Eds., pp. 618–624,2009.

[9] H. M. Abdul-Kader, "Neural networks training based on differential evolution algorithm compared with other architectures for weather forecasting," International Journal of Computer Science and Network Security, vol. 9, no. 3, pp. 92–99, 2009.

[10] K. K. Kuok, S. Harun, and S. M. Shamsuddin, "Particle swarm optimization feedforward neural network for modelling runoff," International Journal of Environmental Science and Technology, vol. 7, no. 1, pp. 67–78, 2010.

[11] B. A. Garro, H. Sossa, and R. A. Vázquez, "Back-propagation vs particle swarm optimization algorithm: which algorithm is better to adjust the synaptic weights of a feed-forward ANN," International Journal of Artificial Intelligence, vol.7, no.11, pp. 208–218, 2011.

[12] B. Garro, H. Sossa, and R. Vazquez, “Evolving neural networks: a comparison between differential evolution and particle swarm optimization,” in Advances in Swarm Intelligence, Y. Tan, Y. Shi, Y. Chai, and G. Wang, Eds., Lecture Notes in Computer Science, pp. 447–454,2011.

[13] B. A. Garro, H. Sossa, and R. A. Vazquez, “Design of artificial neural networks using a modified particle swarm optimization algorithm,” in Proceedings of the International Joint Conference on Neural Networks (IJCNN '09), IEEE, Atlanta, Ga, USA, pp. 938–945, 2009.

[14] B. Garro, H. Sossa, and R. Vazquez, “Design of artificial neural networks using differential evolution algorithm,” in Neural Information Processing. Models and Applications, K. Wong, B. Mendis, and A. Bouzerdoum, Eds., vol. 6444 of Lecture Notes in Computer Science, Springer, Berlin, Germany, pp. 201–208, 2010.

[15] B. A. Garro, H. Sossa, and R. A. Vazquez, “Artificial neural network synthesis by means of artificial bee colony (abc) algorithm,” in Proceedings of the IEEE Congress on Evolutionary Computation (CEC '11), IEEE, New Orleans, La, USA, pp. 331–338, 2011.

[16] R. C. Eberhart, Y. Shi, and J. Kennedy, Swarm Intelligence, “The Morgan Kaufmann Series in Evolutionary Computation,” Morgan Kaufmann, Boston, Mass, USA, 1st edition, 2001.

[17] M. Chen, “Second generation particle swarm optimization,” in Proceedings of the IEEE Congress on Evolutionary Computation, Hong Kong, pp. 90–96, 2008.

[18] Y. Shi and R. C. Eberhart, “Empirical study of particle swarm optimization,” in Proceedings of the Congress on Evolutionary Computation, IEEE, Washington, DC, USA, 1999.

[19] M. Løvberg, T. K. Rasmussen, and T. Krink, “Hybrid particle swarm optimizer with breeding and subpopulations,” in Proceedings of the Genetic and Evolutionary Computation Conference, Morgan Kaufmann, San Francisco, Calif, USA, pp. 469–476, 2001.

[20] N. Higashi and H. Iba, “Particle swarm optimization with Gaussian mutation,” in Proceedings of the IEEE Swarm Intelligence Symposium, IEEE, pp. 72–79, 2003.

[21] S. Mohais, R. Mohais, C. Ward, and C. Posthoff, “Earthquake classifying neural networks trained with random dynamic neighborhood PSOs,” in Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference, ACM, New York, NY, USA, pp.110–117, 2007.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, D. E. Rumelhart, J. L. McClelland, and PDP Research Group, Eds., MIT Press, Cambridge, Mass, USA, pp. 318–362,1986.

[23] J. A. Anderson, An Introduction to Neural Networks, The MIT Press, 1995.

[24] P. J. Werbos, “Backpropagation through time: what it does and how to do it,” Proceedings of the IEEE, vol. 78, no. 10, pp. 1550–1560, 1990.

[25] D. N. A. Asuncion, UCI machine learning repository, 2007.

[26] R. A. V. E. De Los Monteros and J. H. Sossa Azuela, “A new associative model with dynamical synapses,” Neural Processing Letters, vol. 28, no. 3, pp. 189–207, 2008.

[27] B. A. Garro and R. A. Vázquez, “Designing Artificial Neural Networks Using Particle Swarm Optimization Algorithms,” Computational Intelligence and Neuroscience, 2015.

[28] V. G. Gudise and G. K. Venayagamoorthy, “Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks,” in Proceedings of the 2003 IEEE Swarm Intelligence Symposium, pp. 110–117,2003.

[29] “Pima Indians Diabetes Database.” Available: https://www.kaggle.com/uciml/pima-indians-diabetes-database.

[30] Bagley, J. D. “The Behaviour of Adaptive Systems Which Employ Genetic and Correlative Algorithms.” PhD thesis, University of Michigan, Ann Arbor, 1967.

[31] Bauer, P., Bodenhofer, U., Klement, E. P. "A fuzzy algorithm for pixel classification based on the discrepancy norm.” In Proc. 5th IEEE Int. Conf. on Fuzzy Systems, pp. 2007–2012.

[32] Bodenhofer, U. “Tuning of fuzzy systems using genetic algorithms.” Master’s thesis, Johannes Kepler University¨ at Linz, 1996.

[33] Bodenhofer, U., Bauer, P. “A formal model of interpretability of linguistic variables.” In Interpretability Issues in Fuzzy Modelling, J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, Eds.,Studies in Fuzziness and Soft Computing. Springer, Berlin, pp. 524–545,2003.

[34] Bodenhofer, U., And Herrera, F. “Ten lectures on genetic fuzzy systems.” In Preprints of the International Summer School: Advanced Control—Fuzzy, Neural, Genetic, R. Mesiar, Ed. Slovak Technical University, Bratislava, pp. 1–69,1997.

[35] Bodenhofer, U., Klement, E. P. “Genetic optimization of fuzzy classification systems — a case study.” In Computational Intelligence in Theory and Practice, B. Reusch and K.-H. Temme, Eds., Advances in Soft Computing. Physical-Verlag, Heidelberg, pp. 183– 200,2001.

[36] Bonarini, A. ELF, “Learning incomplete fuzzy rule sets for an autonomous robot.” In Proc. EUFIT’93, vol. 1, pp. 69–75,1993.

[37] Bonarini, A. “Evolutionary learning of fuzzy rules: Competition and cooperation.” In Fuzzy Modelling: Paradigms and Practice, W. Pedrycz, Ed. Kluwer Academic Publishers, Dordrecht, pp. 265–283,1996.

[38] Bonarini, A. “Anytime learning and adaptation of hierarchical fuzzy logic behaviours.” Adaptive Behaviour, pp.281–315,1997.

[39] Bulirsch, R., Stoer, J. “Introduction to Numerical Analysis.” Springer, Berlin, 1980.

[40] E. Augustin, M. Carre, E. Grosicki, J.-M. Brodin, E. Geoffrois, and F. Preteux. “Rimes evaluation campaign for handwritten mail processing.” In IWFHR, pp.231–235, 2006.

[41] A. B. Bernard, F. Menasri, R. H. Mohamad, C. Mokbel, C. Kermorvant, and L. Sulem. “Dynamic and contextual information in HMM modeling for handwritten word recognition.” IEEE, pp.2066–2080, 2011.

[42] Ahuja, Nisha; Signal, Gaurav; Mukhopadhyay, Debajyoti, “DDOS attack SDN Dataset”, Mendeley Data,2020.

[43] A. Britto, R. Sabourin, F. Bortolozzi, and C. Suen. “A two-stage HMM-based system for recognizing handwritten numeral strings.” In ICDAR, pp.396–400, 2001.

[44] Y. Chherawala, P. P. Roy, and M. Cheriet. “Feature design for offline Arabic handwriting recognition,” handcrafted vs automated, In ICDAR, pp.678-682,2013.

[45] P. Doetsch, M. Hamdani, H. Ney, A. Gimenez, J. Andres-Ferrer, and A. Juan. "Comparison of Bernoulli and Gaussian HMMs using a vertical repositioning technique for off-line handwriting recognition." In ICFHR, pp.37, 2012.

[46] A. El-Yacoubi, R. Sabourin, C. Y. Suen, and M. Gilloux. "An HMM-based approach for off-line unconstrained handwritten word modelling and recognition." IEEE T-PAMI, pp.752–760, 1999.

[47] H. Fujisawa. "Forty years of research in character and document recognition an industrial perspective." Pattern Recognition, pp.2435–2446, 2008.

[48] A. Graves. Rnnlib, "A recurrent neural net-work library for sequence learning problems." http://sourceforge.net/projects/rnnl/.

[49] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber. "A novel connectionist system for unconstrained handwriting recognition." IEEE T-PAMI, pp.855–868, 2009.

[50] E. Grosicki and H. E. Abed. "Hand-writing recognition competition. In ICDAR,"2009.

[51] S. Haykin. "Neural Networks: A Comprehensive Foundation." Prentice Hall, 1998.

[52] G. E. Hinton, S. Osindero, and Y. Teh. "A fast-learning algorithm for deep belief nets." Neural Compute., pp.1527–1554, 2006.

[53] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. "On combining classifiers." IEEE Trans. PAMI, pp.226–239, 1998.

[54] A. L. Koerich, R. Sabourin, and C. Y. Suen. "Recognition and verification of unconstrained handwritten words." IEEE Trans. PAMI, pp.1509– 1522, 2005.

[55] U. V. Marti and H. Bunke. "Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system." IJPRAI, pp.65–90, 2001.

[56] F. Menasri, J. Louradour, A. L. Bianne-Bernard, and C. Kermorvant. "The A2iA French handwriting recognition system at the RIMES-ICDAR2011 competition." In DRR XIX, 2012.

[57] R. Mohamad, L. Likforman-Sulem, and C. Mokbel. "Combining slanted-frame classifiers for im-proved HMM-based Arabic handwriting recognition." IEEE Trans. PAMI, pp.1165–1177, 2009.

[58] M. P. Schambach and S. F. Rashid. “Stabilize sequence learning with recurrent neural networks by forced alignment.” In ICDAR, pp. 1270–1274, 2013.

[59] A. Vinciarelli and S. Bengio. “Offline recognition of unconstrained handwritten texts using HMMs and statistical language models.” IEEE, pp.709–720, 2004.

[60] S. Young. “The HTK Book,” Version 3.4. Cambridge Univ. Eng. 2006.

[61] N. R. Stepanek and B. Dorn, “Automated Analysis of Lecture Video Engagement Using Student Posts,” in Artificial Intelligence in Education, pp. 565–569,2017.

[62] R. L. U. Cazarez and C. L. Martin, “Neural Networks for Predicting Student Performance in Online Education,” IEEE Latin America Transactions, vol. 16, no. 7, pp. 2053–2060, 2018.

[63] G. Zhao *et al.*, “Deep convolutional neural network for drowsy student state detection,” Concurrency and Computation: Practice and Experience, vol. 30, no. 23,2018.

[64] C. Lee *et al.*, “PSO-Based Fuzzy Markup Language for Student Learning Performance Evaluation and Educational Application,” IEEE Transactions on Fuzzy Systems, vol. 26, no. 5, pp. 2618–2633, 2018.

[65] Y. Ma, C. Cui, X. Nie, G. Yang, K. Shaheed, and Y. Yin, “Pre-course student performance prediction with multi-instance multi-label learning,” Sci. China Inf. Sci., vol. 62, no. 2, 2019.

[66] P. Cortez and A. Silva. “Using Data Mining to Predict Secondary School Student Performance.” In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology, pp. 5-12, 2008.

[67] Heikki, Mannila, “Data mining: machine learning, statistics, and databases,” IEEE, 1996.

[68] U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” AAAI Press / The MIT Press, Massachusetts Institute Of Technology, 1996.

[69] J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufmann, 2000.

[70] Alaa el-Halees, “Mining students’ data to analyze e-Learning behavior: A Case Study”, 2009.

[71] U. K. Pandey, and S. Pal, “Data Mining: A prediction of performer or underperformer using classification”, (IJCSIT) International Journal of Computer Science and Information Technology, 2011.

[72] S. T. Hijazi, and R. S. M. M. Naqvi, “Factors affecting student’s performance: A Case of Private Colleges”, Bangladesh e-Journal of Sociology, vol. 3, no. 1, 2006.

[73] Z. N. Khan, “Scholastic achievement of higher secondary students in science stream”, Journal of Social Sciences, vol. 1, no. 2, pp. 84-87, 2005.

[74] Galit.et.al, “Examining online learning processes based on log files analysis: a case study”. Research, Reflection and Innovations in Integrating ICT in Education 2007.

[75] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, “Mining student data using decision trees”, International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[76] U. K. Pandey, and S. Pal, “A Data mining view on class room teaching language”, (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, pp. 277-282, 2011.

[77] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, “Data mining model for higher education system”, Europen Journal of Scientific Research, vol.43, no.1, pp.24-29, 2010.

[78] M. Bray, “The shadow education system: private tutoring and its implications for planners, “(2nd ed.), UNESCO, PARIS, France, 2007.

[79] B.K. Bharadwaj and S. Pal. “Data Mining: A prediction for performance improvement using classification”, International Journal of Computer Science and Information Security (IJCSIS), vol. 9, no. 4, pp. 136-140, 2011.

[80] J. R. Quinlan, “Introduction of decision tree: Machine learn”, pp. 86-106, 1986.

[81] Vashishta, S. “Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm.” IJACSA - International Journal of Advanced Computer Science and Applications, pp. 77-80,2011.

[82] Kumar, V. “An Empirical Study of the Applications of Data Mining Techniques in Higher Education.” IJACSA - International Journal of Advanced Computer Science and Applications, pp. 80-84,2011.

[83] B. Kumar and S. Pal, “Mining Educational Data to Analyze Students Performance,” International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, 2011.

[84] Alzaidi, K.M.S. Bayat, O. Uçan, O.N. “Multiple DGs for Reducing Total Power Losses in Radial Distribution Systems Using Hybrid WOA-SSA Algorithm,” International Journal of Photoenergy,2019.

[85] Mohammed, T.A. Bayat, O. Uçan, O.N. Alhayali, S. “Hybrid Efficient Genetic Algorithm for Big Data Feature Selection Problems,” Foundations of Science, pp.1-17,2019.

[86] Al-hayali, S. Ucan, O. Bayat O. “Genetic Algorithm for finding shortest paths Problem,” Proceedings of the Fourth International Conference on Engineering & MIS, 27,2018.

[87] Alzaidi, K.M.S. Bayat, O. Uçan O.N. “A heuristic approach for optimal planning and operation of distribution systems,” Journal of Optimization, Hindawi,2018.

[88] Hayali, S.A. Ucan, O. Rahebi, J. Bayat, O. “Detection of Attacks on Wireless Sensor Network Using Genetic Algorithms Based on Fuzzy,” International Journal of Renewable Energy Development ,20019.

[89] Muetı Hadı, Wasan. “Predicting DOSS attacks using neural networks.” (Yayınlanmamış yüksek lisans tezi). Altınbaş Üniversitesi, Lisansüstü Eğitim Enstitüsü, İstanbul,2020.