WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Malware detection approaches using deep learning and ml

[1]Gajendra Kumar, [2]Kushagra Jain

[1]Research Scholar, Department of Computer Science and Engineering, Arya institute of engineering and technology Jaipur, Rajasthan. India. E mail: - gajendrajat2021@gmail.com

[2]Research Scholar, Department of Computer Science and Engineering, Arya institute of engineering and technology Jaipur, Rajasthan. India. E mail: - kjkushagra1234@gmai.com

ABSTRACT:

Due to technological advancements, the internet has grown significantly, necessitating the necessity for data storage. When data is altered when uploading or downloading, a variety of malware infections are heavily present in the data. With the advent of diverse technologies, technology use will grow more complicated. Different machine learning processes exist based on how the technologies are used effectively. We offer a survey on the use of deep learning algorithms for virus detection in this one. The fundamental analysis of the relevant material is covered first. After that, a summary of deep learning techniques and infection types is given. A brief report on malware detection and deep learning techniques is included in this study.

Keywords: CNN, RNN, Hybrid, Dynamic, and Static analysis.

INTRODUCTION: -

One method within the machine learning family is called deep learning. Artificial intelligence's neural network architecture serves as its foundation. Deep Learning is a technique that teaches computers to perform tasks that humans precisely anticipate. It is categorized as part of the artificial intelligence family that uses hierarchical learning technique, which is further divided into supervised and unsupervised learning. One term for deep learning in artificial intelligence is "Black box". A harmful software is one that changes other computer programs and copies itself. The purpose of the anti-virus software is to detect the presence of malicious software by comparing its definitions with regularly updated data. Malware detection based on signatures is the term for this Programs such as malware, viruses, spam, etc. are the ones that infect systems. The biggest issue with cyber security is the detection and identification of malware. Machine learning provides parabolic accuracy in malware detection and categorization together with the flexibility to reduce a great deal of the human labor required with traditional techniques to malware analysis. This research presents an efficient method for demonstrating the effectiveness of the Deep Learning algorithm over several cyber security methods... Deep learning algorithms are regarded as a reliable method of problem solving since they take into account several traditional and machine learning techniques. The conclusion of this article is that malware detection must be done quickly and automatically.
identify fresh malware classifications emerging in daily data storage.

RELATED WORK: -

Over the years, deep learning applications for security issues have gained attention. The uses of deep learning techniques in the field of security threat detection have been closely described by researchers. These days, deep learning techniques are quite important in the field of security. There are several methods for data to get compromised. In order to classify diseases, it is necessary to identify the significant noise present in data storage. A small number of serious varieties of infections have been identified. Specifically, malware, worms, and viruses. These illnesses often fall under one of two categories: work related to infection detection or work related to infection prevention. Malware is the main infection that affects the of the aforementioned illnesses. The final issue in terms of security concerns is malware analysis. The two main phases of malware detection with a machine learning technique are automated detection and feature extraction. Malware may be detected using two different methods: static analysis and dynamic analysis. Numerous studies have examined the existence of different infection classifications and the detection of infections using a variety of Deep Learning functions and approaches.

NEED OF STUDY: -

This survey article could inspire a novel deep learning strategy. Numerous papers have been published outlining different machine learning methodologies. Furthermore, in the advanced survey, a variety of machine learning method applications were covered. Methods for cyber security intrusion detection using machine learning (ML) and data mining (DM) have been studied. Klaine has reviewed machine learning algorithms and their

self-organizing cellular network solutions, providing insightful categorization and comparison. [29] Although machine learning techniques have been used in many different fields, there are still a few more where research and development should be directed. In his study, Jun Feng Xie conducted a survey of the use of machine learning algorithms in SDN. Hodo et alML-based intrusion detection systems (IDS) are also the topic of. The deep learning-based IDS, which is also briefly discussed in, is the primary distinction between and the integrated choices are the static and dynamic ones, which are combined together.

Feature vectors are used in machine learning approaches for coaching and classification. DL, also referred to as "hidden boxing" AI, has several more benefits that machine learning is unable to match. These are a variety of polls that illustrate the necessity for DL in light of different security threats.

DEEP LEARNING: -

Introduction: -

AI has a significant influence on developing technologies. A few subsets of AI go by the term of "machine learning," which serves as a primer for deep learning. The neurons that make up DL are linked together. The most significant AI technique is called deep learning, and it is made up of a network of neurons. Connected layers are used to process the structure of neural networks. A neuronal network is composed of several layers, including input, output, and hidden layers. The term "hidden layer" refers to every layer that is between the input and output layers. A deep network is one that connects more than two layers. There are connections between neurons. The weight, bias, and activation function of the input layer determine how strong the signal is sent to the following layer. The brain system's capacity for complicated learning increases with layer depth.

Deep Learning Approaches: -

Three types of deep learning algorithms exist: unsupervised, semi-supervised, and supervised learning. Additionally, it has been divided into two categories: deep reinforcement learning and reinforcement learning. There are two main stages to the learning process. To produce a statistical model of the data as an output, nonlinear transformation of the input data is applied in the first phase. The second stage involves refining the data using a derivative mathematical model. To get the desired result, these two steps are done several times. Deep Supervised Learning, Deep Semi-Supervised Learning, Deep Unsupervised Learning, Reinforcement Learning, and Deep Reinforcement Learning are the different deep learning methodologies.

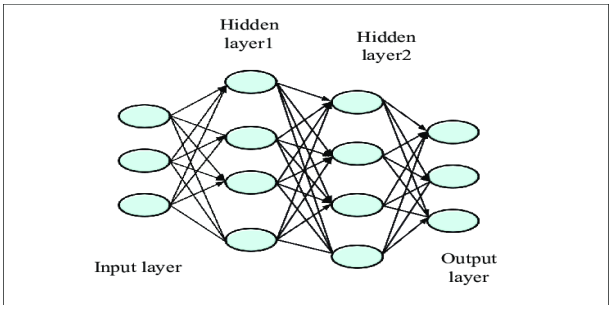


Fig 1: Simple Structure of Deep Neural Network

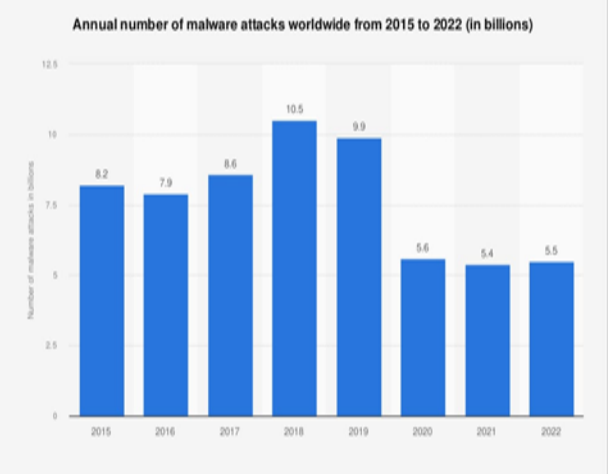
Architecture of Deep Learning Network: -

Based on neural networks, the architecture of deep learning can be broadly categorized as reinforcement learning, feed forward neural networks (FFN), recurrent neural networks (RNN), and convolutional neural networks (CNN). Multiple layers of neurons combine in different ways to form different combinations. The DL neural network architecture consists of multiple layers of neurons arranged in combination. where the input and output layers are separated by a number of hidden layers

MALWARE: -

Classification: -

Malicious software, also known as malware, is any program intended to harm individuals, networks, or computer systems. Malware comes in several forms. It's critical that people and organizations understand the various forms of malware and take precautions to safeguard their systems. Some of these precautions include using antivirus software, updating software and systems, and exercising caution when opening email attachments or downloading files from the internet



3. 2. 1. Viruses

It is a legal file with a malicious executable code attached. When an infected file is transferred from one machine to another, it may propagate. They can alter, remove, or spy on data using it. A virus may activate when a file is opened. A software virus will infect other programs on the computer once it becomes active. The first virus was created in 1986 and went by the moniker BRAIN. It was infected via a floppy disk by two hacker brothers, Basit and Amjad Farooq Alv.

3. 2. 2. Worms

Like viruses, worms attach themselves to a file and search for paths between machines, such as a computer network that shares common file storage regions, in order to multiply themselves on the system and cause destruction. Networks are often slowed down by worms. Worms are self-contained programs, while viruses require a host software to function. Once a worm has infected a host, it may swiftly propagate throughout the network. The Morris or Internet worm (2 NOV 1988) was the first known worm assault, and it was disseminated over the Internet.

3. 2. 3. Trojan horse

Hackers utilize it the most frequently of all. The tale of the Trojan Horse inspired its name. It installs itself on a computer in the appearance of a trustworthy application. Typically, the distribution technique involves an attacker hiding harmful code within legitimate software using social engineering in an attempt to obtain users' system access through the software. The Rakhni Trojan and Tiny Banker are two instances of trojan horses.

3. 2. 4. Ransomware

After seizing control of a computer system or the data on it, ransomware locks it until the victim pays. The user-secret key is used to encrypt data on the computer. To get the data back, the user needs to give the perpetrators a ransom. The victim can use the system again when the money has been paid. Joseph Popp created the AIDS trojan in 1989, which was the first ransomware assault. The trojan's design flaw was so bad that it was not required to pay the extortionist at all, even though Popp intended to develop a trojan.

3. 2. 5. Adware

On the PC, it shows intrusive pop-ups and advertisements. Packages and software downloads are included with it. By displaying advertisements, it brings in money for the software distributor. Industry insiders believed that the first ad-supported software, which debuted in 1995, belonged to the wider category of spyware.

3. 2. 6. Spyware

Its goal is to steal confidential data for a third party from a system. Information is gathered by spyware and sent to the hacker. On October 16, 1995, the term "spider" was first used in writing.

3. 2. 7. Logic Bombs

A malicious software that employs a trigger to activate its harmful code is known as a logic bomb. The logic bomb doesn't start working until that trigger event takes place. When a logic bomb goes off, it activates harmful code that damages a computer. Known as the Trans-Siberian Pipeline event, it happened for the first time in 1982. This incident's tale featured all the elements of a spy film: secret documents, international intrigue, the KGB and CIA.

3. 2. 8. Rootkits

A rootkit creates a backdoor by altering the OS. Then, via the backdoor, attackers may remotely access the machine. Originally identified as NT Rootkit, it was originally detected by Greg Houland in 1999. Hacker Defender followed in 2003, and the first Mac arrived in 2009.

3. 2. 9. Backdoors

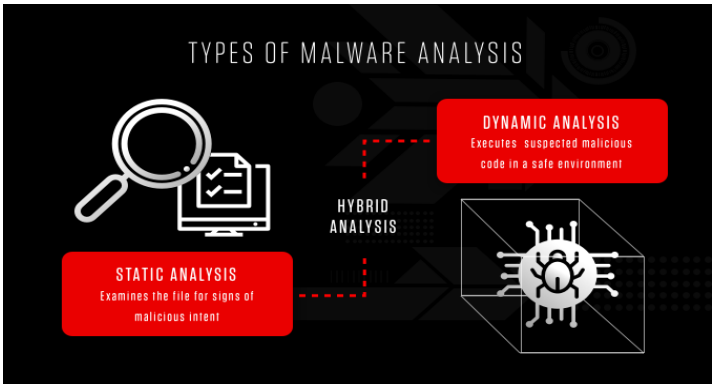
A backdoor allows access to a system without using the standard authentication. Even if the company resolves the initial security flaw that allowed the hacker to access the system, the backdoor's goal is to provide the hacker access to the system in the future. When there is no other means to obtain system access, the hacker makes sure it is hidden correctly and uses it.

3. 2. 10. Keyloggers

In order to gather passwords and other sensitive information, Keylogger logs everything a user enters on their computer and sends it to the program's originator. Keylogger monitoring of the victim's activities is its primary function. Wikipedia describes the many usages of keyloggers for government covert operations during the 1970s and early 1980s.

Malware Analysis Technique :

Malware analysis techniques protect users from security risks associated with different malware assaults [19]. Malware analysis is classified into three main categories. They might be hybrid, dynamic, or static. Static code is another term for this. used for troubleshooting software without running the application.



Analysis of Static Data

It is not necessary to run the code in order to do basic static analysis. Static analysis, on the other hand, looks for indications of malicious intent in the file. Finding malicious libraries, compressed files, or infrastructure might be helpful.
To ascertain whether a file is malicious, one can employ technical indications such as file names, hashes, strings like IP addresses and domains, and file header data. Furthermore, in order to learn more about how malware functions, one may monitor the virus without really executing it using tools like disassemblers and network analyzers.

Dynamic Analysis: -

Its behavior is tested and its usefulness is discovered via the use of dynamic analysis. This may also include the domain name, IP address, and other information.

Hybrid Analysis: -

Its behavior is tested and its usefulness is discovered via the use of dynamic analysis [. This mother Hybrid analysis gives security teams the best of both worlds by fusing static and previously unknown code with dynamic analysis methodologies. This allows it to identify malicious code that is attempting to conceal and extract a large number of additional indications of compromise (IOCs). It also includes the domain name, IP address, and other information.

Malware detection Approach: -

Malware detection is classified into two broad category improved analysis of signatures. Deep learning for the identification of malware. recursive disassembling. examining data on various operating systems. Innovation in sandboxing. There are two main categories for malware detection. both heuristic and signature based.

The two types of signature-based detection techniques are byte signature and hash signature. There are two types of heuristic-based detection approaches: static and dynamic.

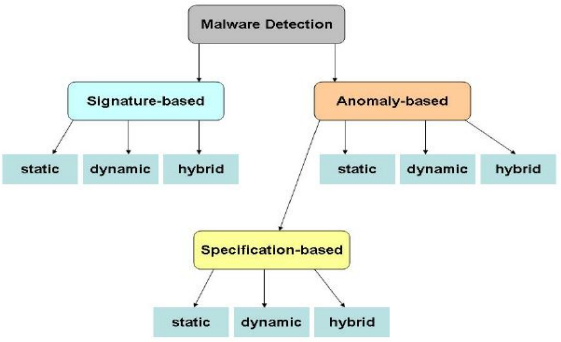


Fig:3.1 Detection Technique

Deep Learning Algorithms in Malware detection: -

Deep learning algorithms, which can pick up complex patterns and traits from raw data, have shown promise in the realm of malware identification. Here are a few popular deep learning algorithms:

1. CNNs, or convolutional neural networks:

In addition to being often utilized for image classification tasks, CNNs may be applied to malware detection by taking into account binary data, such as pictures.

Through convolutional layers, they are able to learn hierarchical representations of characteristics in the data, which makes them efficient malware detectors.

2. RNNs, or recurrent neural networks:

Because RNNs excel at processing sequential data, they are useful for examining system call or API call sequences in malware investigation.

Because they can capture long-term dependencies, RNN versions with Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are often utilized.

3. GANs, or Generative Adversarial Networks:

The generator and discriminator neural networks, which make up GANs, are trained concurrently.

They produce malware that isn't real.

Conclusion: -

First, the book discusses the necessity of data storage due to advances in technology, malware infections, and the rise in technical complexity. The findings underline the critical need for quick and automated malware detection and emphasize the significance of deep learning techniques in virus identification and malware analysis in the current cyber security environment.

REFERENCE:

1. Malwarebytes by Marcin Kleczynski
2. Wikipedia by Lawrence Mark Sanger
3. A STUDY OF CYBER SECURITY CHALLENGES AND ITS EMERGNING TRENDS ON LATEST TECHNOLOGIES by NIKHITA REDDY and G.J. UGANDER REDDY.