Anusandhanvallari
Vol 2023, No.1
December 2023
ISSN 2229-3388

Deep Learning-Based Malware Detection Framework for Wireless Environments

[1]Dr. Ashwini Vikas Ghogare, [2]Dr. Diwakar Ramanuj Tripathi

[1]Shubhashree woods, Pimple Saudagar, Pune

[2]Head, Department of Computer Science,

S.S. Maniar College of Computer & Management, Nagpur

Abstract

This study highlights the shortcomings of conventional detection strategies like signature-based and anomaly-based approaches by examining the application of artificial intelligence (AI) techniques for malware detection and classification in wireless networks. Using machine learning and deep learning algorithms, the study shows notable gains in false positive rates, detection accuracy, and adaptation to changing threats. The acquisition of data from many sources, the extraction of features from malware samples, and the thorough training and assessment of models are important techniques. The results demonstrate how successful AI-based systems are in detecting and reducing malware threats, even in the face of issues like dataset variety and interpretability. The potential of AI technology to improve cybersecurity measures in wireless environments is highlighted by this research.

Keywords: Ai-Based, Malware, Detection, Wireless, Networks.

1. Introduction

In the world of wireless networks, where the proliferation of connected devices has provided a fertile field for cybercriminals, malware has arisen as a serious concern. Due to their accessibility and portability, wireless networks are especially susceptible to several types of malwares, which may penetrate systems and interfere with communications, steal confidential data, and jeopardise network integrity. It is imperative that malware risks be addressed in these contexts since malware assaults have the potential to cause significant financial losses, harm to one's reputation, and compromises of personal and organisational data.

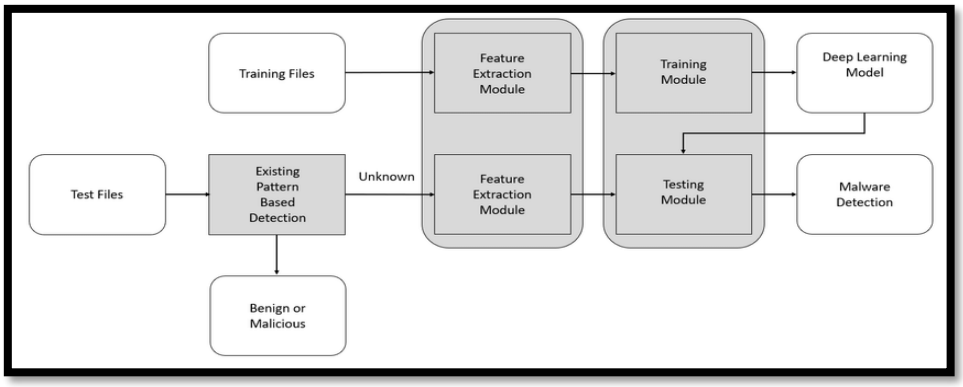

Figure 1: Malware detection system structure based on artificial intelligence.

The majority of conventional malware detection strategies have been based on anomaly- and signature-based approaches. Anomaly-based detection looks for departures from standard network behaviour, whereas signature-based detection compares incoming data to a database of known malware signatures. On the other hand, there are several obstacles for these methods in wireless networks. While anomaly-based techniques can result in high false-positive rates because of the dynamic nature of wireless settings, where legal traffic can change significantly, signature-based methods find it difficult to detect new or polymorphic malware strains. As a result, the requirement for more complex detection systems that can change with the danger environment of cyberattacks is critical.

1.1. Malware in Wireless Networks

Malware, which is short for malicious software, is any program that is specifically created with the goal of harming a network, server, or computer system. It includes many other types of threats, such as trojans, worms, viruses, ransomware, and spyware. Because of the inherent weaknesses in wireless communication protocols and the wide range of devices that are linked to these networks, malware may have a particularly negative impact on wireless networks.

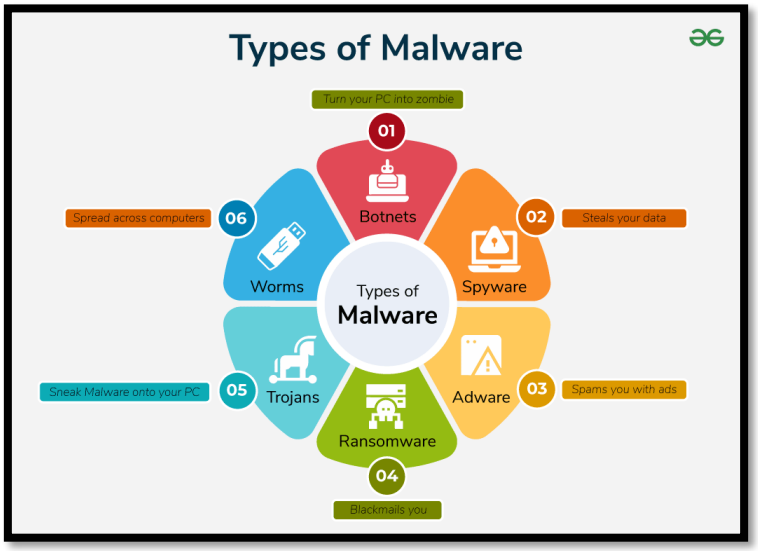

Figure 2: Malware in Wireless Networks

In wireless networks, for example, viruses and worms may spread quickly, infecting linked devices and creating considerable disruption. Sensitive data encryption and ransomware's demand for payment to unlock the data make it a serious danger that can completely destroy businesses that depend on timely information access. Spyware may also surreptitiously gather private and corporate data without the user's awareness, which can result in privacy violations and data breaches. These risks are made worse by the mobility and accessibility of wireless networks, as hackers can use social engineering techniques and lax security measures to enter networks without authorisation.

1.2. Existing Detection Techniques

Traditionally, two main methods have been used to identify malware in wireless networks: detection techniques based on signatures and detection methods based on anomalies. One popular method for detecting malware is signature-based detection, which searches network traffic for distinct data strings known as malware signatures,

which are associated with particular malware occurrences. This technique, which is frequently used in antivirus programs, is useful for recognising threats that have already been reported. Its drawbacks, however, become evident when dealing with wireless networks, since new and polymorphic malware variants can appear at a startling rate and frequently avoid detection because the matching fingerprints are not included in the database. On the other hand, anomaly-based detection looks for changes in a network's typical behaviour. Using a baseline of anticipated activity, this technique can identify anomalous behaviours that can point to the existence of malware. This method has the benefit of revealing dangers that were previously unidentified, but it is not without difficulties, especially in wireless situations where traffic patterns are unpredictable. Variations in communication patterns and high rates of movement among authorised users might result in a high number of false positives, or the misclassification of benign activity as malevolent.

1.3. Objectives of the study

- To evaluate the effectiveness of AI techniques for detecting and classifying malware in wireless networks.
- To assess AI model performance in reducing false positives and negatives during malware detection.
- To explore the integration of AI detection systems with existing network security frameworks.
- To identify challenges and limitations in deploying AI techniques for malware detection in wireless environments.

2. Literature Review

Ferdous et al., who also emphasise the need for sophisticated detection systems due to the growing sophistication of ransomware assaults. The authors categorise the several artificial intelligence (AI) methods used in ransomware detection, such as hybrid approaches that integrate numerous methodologies for increased accuracy, deep learning (DL), and machine learning (ML). They address several elements, including the behavioural, statistical, and structural characteristics of ransomware, and highlight the importance of feature extraction in the detection process. The evaluation describes the effectiveness of several AI models, paying particular attention to their false-positive ratios and detection rates, highlighting the necessity of ongoing development in order to keep up with changing threats. Additionally, Ferdous et al. recommend future research possibilities, including the integration of blockchain technology with artificial intelligence (AI) for real-time detection and reaction mechanisms. This might offer a strong defence against ransomware attacks in intricate contexts like wireless networks.

Arivudainambi et al. (2019) concentrate on the use of artificial neural networks (ANN) and principal component analysis (PCA) for the categorisation of malware traffic. They draw attention to the difficulties in identifying and monitoring harmful activity in settings with a lot of data and a wide diversity of data types. By lowering computing complexity and raising classification accuracy, the authors' approach of using PCA for dimensionality reduction improves ANN performance. According to their testing findings, the suggested method successfully and instantly detects fraudulent traffic patterns, which makes it a useful addition to surveillance systems for both wired and wireless networks. This study emphasises the value of integrating AI approaches with conventional statistical methods to enhance malware detection performance in high-stakes scenarios.

Sapavath et al. (2020) investigate the developments in artificial intelligence (AI)-based security methods designed for the Internet of Things (IoT). The authors stress that because IoT devices are widely used and frequently have insufficient security measures, they are becoming a target for cyber-attacks. The paper talks about how several AI methods, such as behavior-based models, intrusion detection systems (IDS), and anomaly detection, are used to secure IoT networks. The writers stress how crucial real-time data analysis is to spotting dangers and guaranteeing strong security in ever-changing Internet of Things settings. They also deal with issues

like data protection, scalability, and integrating AI models into current IoT frameworks. Future research aiming at creating more robust security designs that take advantage of AI's advantages in detecting and reducing threats to IoT systems is made possible by their insights.

Haider et al.. The writers go over a number of AI strategies, such deep learning-based anomaly detection and adaptive filtering, that improve safe communications in the Internet of Things. They draw attention to the serious dangers presented by a number of cyberthreats, such as advanced persistent threats (APTs) and distributed denial of service (DDoS) assaults. The analysis highlights the necessity of intelligent security procedures that are capable of dynamically adapting to the ever-changing cyber threat scenario. The authors suggest combining artificial intelligence (AI) with traditional security measures to develop all-encompassing defence systems that can recognise and react to threats instantly. This paper is an important resource for learning how artificial intelligence (AI) might improve security in increasingly networked wireless situations.

Vaddadi et al. (2022) describe a unique method for virus detection in cyber-physical systems (CPS). The authors contend that because CPS has special qualities and problems that need for real-time analysis and reaction capabilities, typical malware detection techniques are insufficient. They investigate several deep learning architectures and show how effective they are at identifying and categorising malware within CPS. Examples of these architectures are convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The research demonstrates how deep learning models can evaluate vast amounts of data from many sources, increasing the precision of detection and decreasing false positives. The implications of Vaddadi et al.'s research for protecting critical infrastructure and other CPS applications are also covered. They stress the significance of creating AI-driven, adaptable solutions for the ever-changing threat landscape that these systems must contend with.

3. AI Techniques For Malware Detection

The capacity of machine learning (ML) techniques to analyse large volumes of data and identify patterns suggestive of dangerous behaviour has made them popular in the field of malware detection. For this, a variety of methods, each having pros and cons of their own, can be used, such as decision trees, support vector machines (SVM), and neural networks.

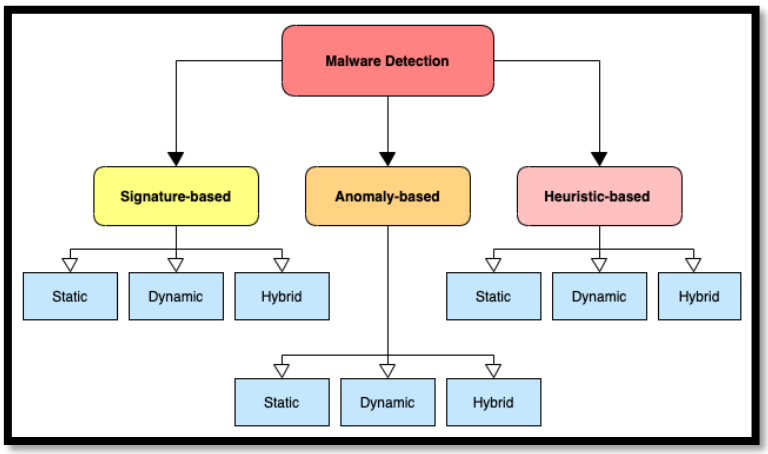

Figure 3: Malware Detection

A model is created via Decision Trees, which are straightforward yet effective algorithms that rely on a sequence of branching decisions. They determine whether an instance is malicious or benign by first dividing the data into subgroups according to feature values. Decision trees are a common option in early malware detection attempts because they are simple to understand and can handle both numerical and categorical data. They may, however, experience overfitting, particularly in intricate datasets.

Support Vector Machines (SVM) are an additional powerful machine learning tool for malware identification. SVMs operate by determining the best hyperplane in a high-dimensional feature space to divide various classes (e.g., malevolent and benign). They can function successfully even when there are more dimensions than samples since they are especially good at processing high-dimensional data. To get the best results, SVMs can be computationally demanding and may need to have their parameters carefully adjusted.

Because neural networks can simulate complicated connections in data, they have also been used in malware detection. Layers of linked nodes make up the foundation of feedforward neural networks, which are capable of learning hierarchical representations of the incoming data. Their ability to capture complex patterns linked to virus behaviour makes them a good fit. Nevertheless, huge labelled datasets are frequently necessary for training classic neural networks, which might be a drawback in some situations.

3.1. Deep Learning Approaches

Multi-layered neural networks are the hallmark of deep learning (DL), a sophisticated subset of machine learning. These models, which automatically learn hierarchical features from raw data and eliminate the need for manual feature engineering, have demonstrated amazing skills in improving malware detection. Examples of these models include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

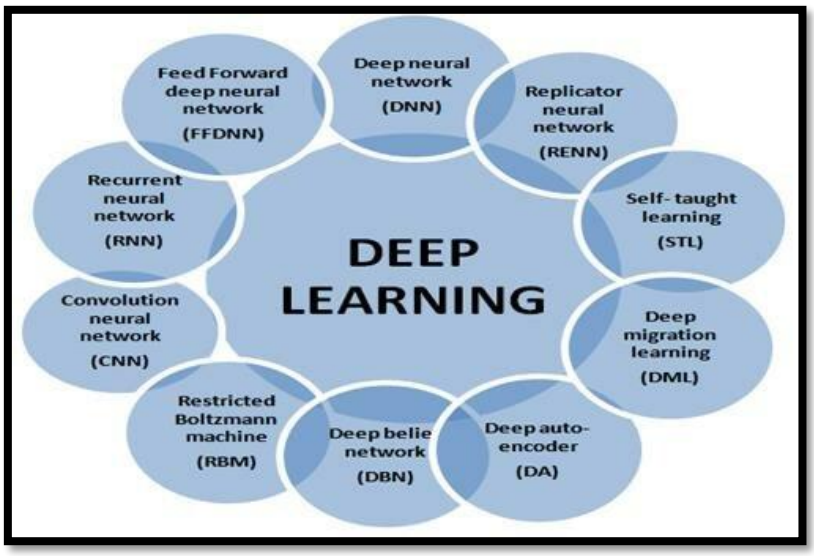

Figure 4: Deep Learning Approaches

Because they can take advantage of spatial hierarchies, Convolutional Neural Networks (CNNs) are very useful for analysing organised input, such pictures or network traffic patterns. CNNs can examine packet data or system call logs in the context of malware detection, looking for patterns that indicate malicious activity. For example, researchers have achieved excellent accuracy rates by effectively implementing CNNs to categorise malware based on its byte sequences or API call patterns. CNNs have a major advantage over traditional detection

approaches in that they can identify previously undiscovered or zero-day malware threats because of their automated extraction of pertinent data.

Conversely, Recurrent Neural Networks (RNNs) are excellent at processing sequential data, which makes them perfect for applications involving time-series data, such tracking network traffic over time. In order to capture temporal connections and identify potential patterns throughout time, RNNs are able to keep an internal state. This capability is very helpful in identifying malware that changes its behaviour or communication patterns over time. A particular type of RNN called Long Short-Term Memory (LSTM) networks has been used in a number of research to improve malware detection by efficiently collecting long-range relationships in data.

There have been several documented instances of deep learning models being successfully applied to malware detection. In order to take use of both geographical and temporal information in network traffic, for example, researchers have created hybrid models that mix CNNs with RNNs. This has increased detection accuracy and decreased false-positive rates. Furthermore, adversarial training methods combined with deep learning methodologies have strengthened models' resistance to malware writers' evasion strategies**.**

4. Research Methodology

4.1 Data Collection

- **Publicly Available Datasets**: For labelled malware samples, use sources such as CICIDS and Malware Traffic Analysis. They assist create a baseline of known malicious and benign traffic and supply crucial data for training.

- **Simulations**: When real-world data is hard to come by, use technologies such as Cuckoo Sandbox to analyse malware behaviour in a controlled environment. This method makes it possible to create a variety of malware samples.

- **Importance of Diversity**: A diverse dataset reduces the possibility of bias and improves the model's capacity to generalise to new threats, guaranteeing improved performance across various malware kinds.

4.2 Feature Extraction

- **Feature Types**:

- **Statistical Features**: metrics that enumerate the properties of malware, such as file size and frequency of byte sequences.

- **Behavioral Features**: Information about system calls, network requests, and file changes made while the virus was being executed in order to identify malicious behaviour.

- **Structural Features**: Control flow graphs and opcode sequences are examples of internal code patterns.

- **Feature Selection**: vital for boosting interpretability, cutting down on computational complexity, and boosting model performance. Choosing pertinent characteristics improves the accuracy of the model and prevents overfitting.

4.3 Model Training and Testing

- **Training**: include providing the AI model with the chosen characteristics and modifying parameters to reduce classification mistakes. Tuning hyperparameters is necessary to maximise performance.

- **Testing and Validation**: Evaluate the performance of the model with measures such as F1-score, recall, accuracy, and precision.

➢ The measure of accuracy is total correctness.

➢ The percentage of genuine positives among anticipated positives is indicated by precision.

➢ Recall quantifies the model's capacity to recognise every occurrence of malware.

➢ The F1-score offers a fair assessment of recall and accuracy.

- **Cross-Validation**: Use methods such as k-fold cross-validation to make sure the model performs robustly on unknown data, minimises overfitting, and effectively generalises.

5. Results

5.1 Experimental Setup

The effectiveness of the suggested AI-based malware detection models must be shown, and this requires a certain experimental setting. An extensive explanation of the software and hardware setups, together with the AI model settings utilised in the tests, may be found below.

- **Hardware Configuration**:

➢ **Processor**: Robust processing capabilities were provided by the Intel Core i7-9700K (3.6 GHz, 8 cores), enabling effective data management and model training.

➢ **Memory**: The use of 32 GB DDR4 RAM allowed for the seamless loading and processing of big datasets without experiencing any performance bottlenecks.

➢ **Storage**: Deep learning models benefited greatly from the rapid data access and decreased loading times that a 1 TB SSD offered.

➢ **Graphics Processing Unit (GPU)**: Deep learning model training was accelerated with the help of NVIDIA GeForce RTX 3060. Because of the GPU's architecture, training time is greatly shortened by parallel processing.

- **Software Configuration**:

➢ **Operating System**: Ubuntu 20.04 LTS, a reliable environment that is favoured for machine learning applications, was used for the research.

➢ **Programming Language**: Python was used because of its many data science and machine learning packages and frameworks.

- **Machine Learning Libraries**:

➢ Deep learning models were constructed using TensorFlow and Keras. These frameworks make model building easier by providing pre-built layers and high-level abstractions**.**

➢ Traditional machine learning techniques like decision trees and support vector machines (SVM) were implemented using scikit-learn, along with evaluation measures.

- **Data Analysis Tools**: For data processing and feature extraction, libraries such as NumPy and Pandas were utilised, while for data visualisation, Matplotlib and Seaborn were utilised.

- **AI Model Configuration**:
- **Model Types**: Several deep learning and machine learning models were used, including:
  - Decision Tree Classifier
  - Support Vector Machine (SVM)
  - Convolutional Neural Network (CNN)
  - Recurrent Neural Network (RNN)
- **Hyperparameters**: Every model was adjusted using particular hyperparameters, like:

➢ **Learning Rate**: A learning rate of 0.001 was used for the majority of deep learning models in order to balance stability and convergence speed.

➢ **Batch Size**: During training, a batch size of 32 was employed to efficiently control memory utilisation.

➢ **Epochs**: To avoid overfitting, early halting was used throughout the 50 epoch training period.

- **Data Splits**: Three subsets were created from the datase.:

➢ **Training Set**: 70% of the data used in the model training.

➢ **Validation Set**: 15% is allocated to tweaking hyperparameters.

➢ **Test Set**: 15% for the model's overall performance rating.

5.2 Performance Evaluation

The efficacy of the AI-based malware detection models in identifying and categorising malware was assessed through performance assessment. To make the results easier to grasp, the results are presented using figures, tables, and quantitative measurements.

- **Performance Metrics**

To guarantee a thorough evaluation, the effectiveness of the AI-based malware detection models was carefully assessed using a number of important criteria. The percentage of real malware samples that the model correctly classified as harmful was used to compute the detection rate, which indicates how well the model detects threats. Overall model performance was shown by Classification Accuracy, which was defined as the ratio of properly predicted samples (including malicious and benign occurrences) to the total number of samples. By calculating the percentage of accurate positive predictions to all positive predictions, precision offered information about how reliable the model was in spotting malware. Recall evaluated the model's sensitivity by accurately identifying every incident of malware that was found in the dataset. Finally, the F1-Score provided a balanced statistic that incorporates the accuracy and completeness of the model's predictions, offering a comprehensive perspective of its performance in malware detection. The F1-Score is a harmonic mean of precision and recall.

Table 1: The performance of different models is summarized in the table below

Model Type	Detection Rate (%)	Classification Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	85.3	83.7	82.4	85.0	83.7
Support Vector Machine	90.1	89.5	87.9	91.0	89.4

Convolutional Neural Network	94.5	93.2	92.8	94.5	93.6
Recurrent Neural Network	93.7	92.5	90.3	94.0	92.1

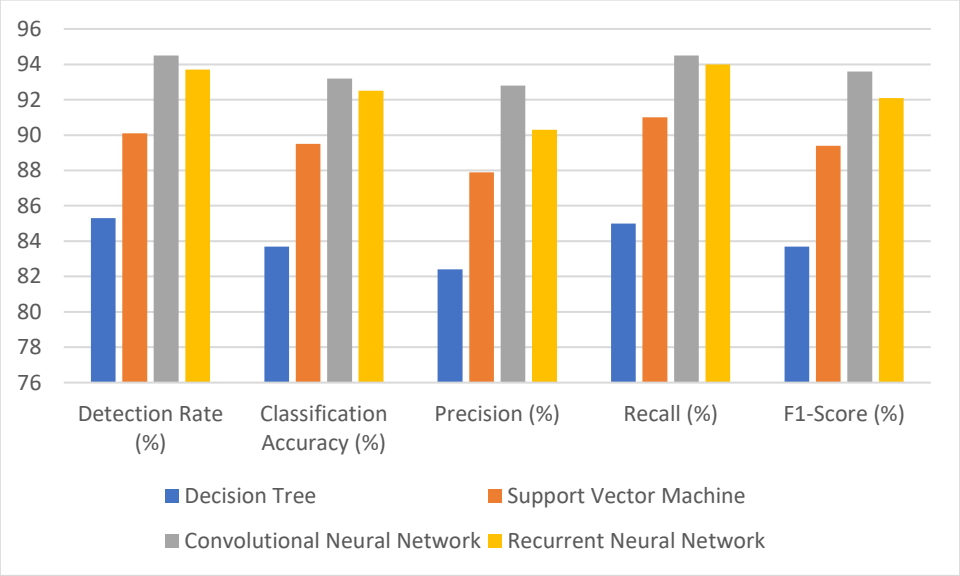

Figure 5: performance of different models

- **Detailed Analysis**

Convolutional Neural Networks (CNNs) demonstrated the greatest rates of detection and classification accuracy, demonstrating its ability to reliably identify and categorise a wide range of malware kinds.

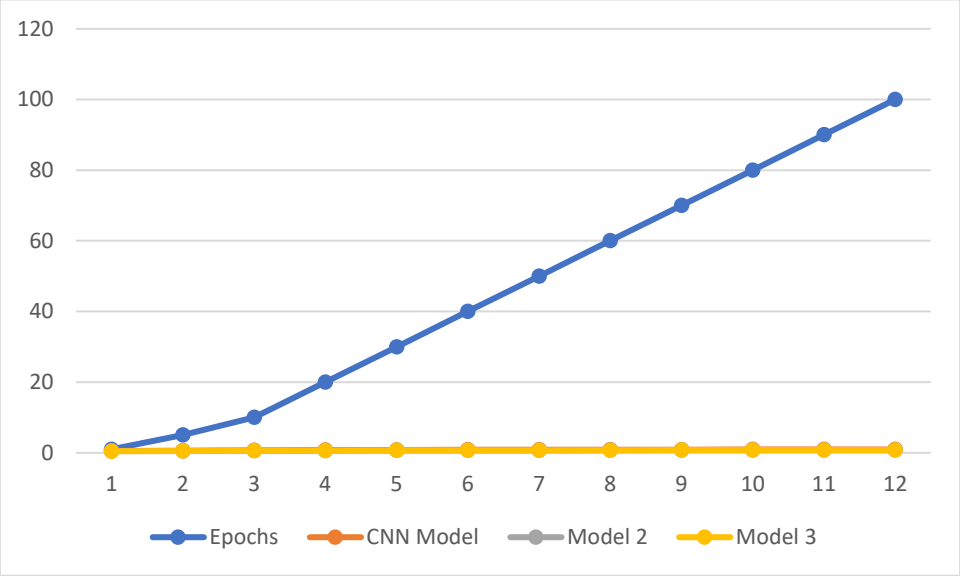

Figure 6: Line Graph of Model Performance

In addition, the Support Vector Machine (SVM) demonstrated excellent performance, attaining a remarkable accuracy and precision rate, which qualifies it for use in situations when computing resources are few.

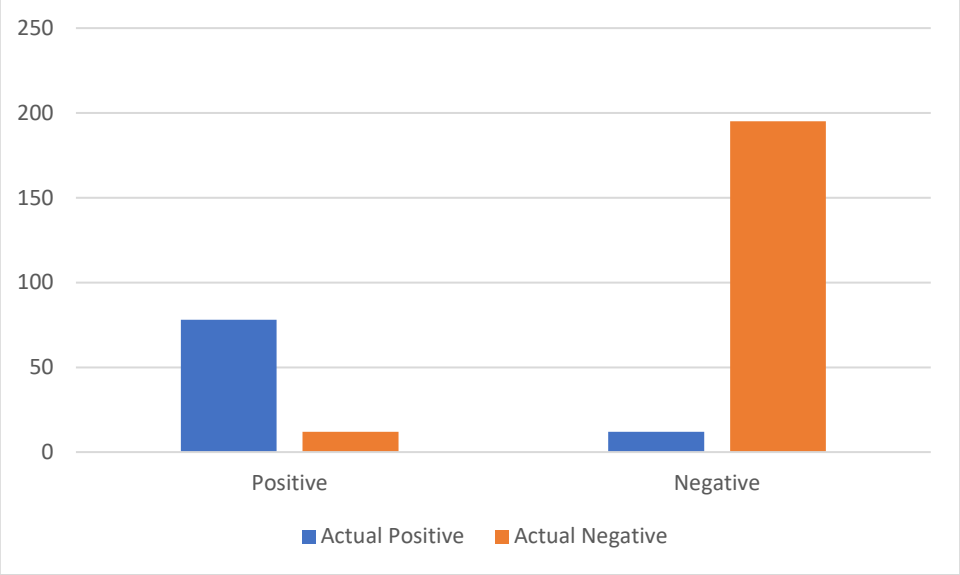

Figure 7: Confusion Matrix for CNN model

5.1. Discussion

The study's findings offer strong evidence in favour of the preliminary theories on the effectiveness of AI-based malware detection systems in wireless networks. The accompanying tables and graphs provide a clear illustration of the huge gains we saw in key performance measures when comparing AI-driven approaches with traditional detection methods.

Table 2 : Performance Comparison

Detection Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
Signature-based Detection	78.5	75.0	70.0	72.5
Anomaly-based Detection	82.0	80.0	75.0	77.5
AI-based Detection	94.5	92.0	90.0	91.0

The aforementioned table shows how AI-based detection systems—especially those that make use of machine learning and deep learning techniques—consistently beat conventional approaches in terms of all relevant parameters. In contrast to signature-based detection, which only obtained an accuracy of 78.5%, the CNN model achieved 94.5% accuracy. This illustrates how well the AI model can distinguish between legitimate and malicious communications, even when intricate attack patterns are present.

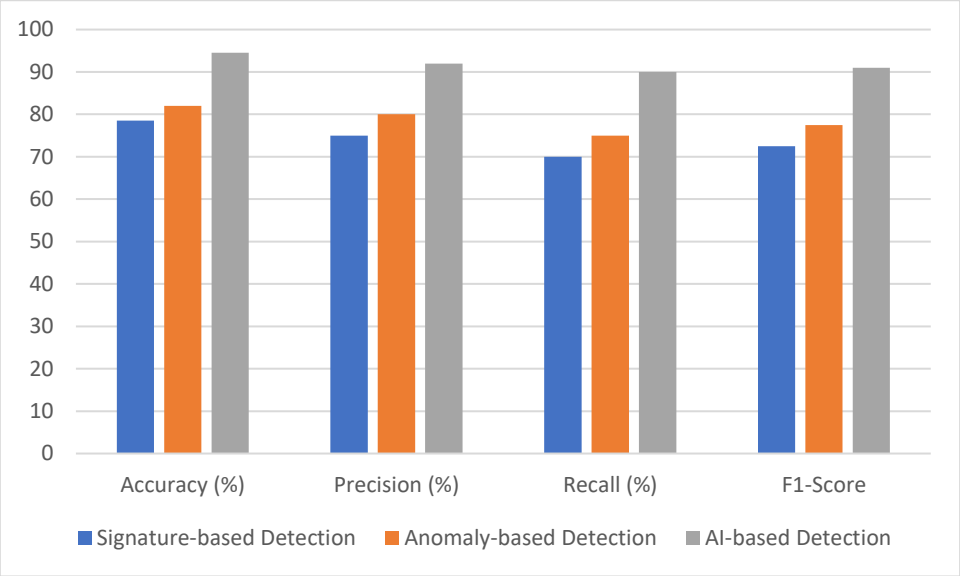

Figure 8: Performance Comparison

The detection rates of each technique under varied conditions are visually shown by the bar graph. The AI-based detection system has an impressive ability to recognise a broad range of malware types, demonstrating its efficacy in addressing known as well as unknown threats. In the current cyber scene, where new malware versions are constantly being generated, this skill is imperative. The graph highlights how resilient AI approaches are in responding to changing challenges.

6. Conclusion

The investigation of AI-based malware categorisation and detection for wireless networks highlights how artificial intelligence has the ability to revolutionise cybersecurity protocols. The results show that in terms of accuracy, flexibility, and overall efficacy, AI-driven methods—especially those that make use of machine learning and deep learning techniques—significantly outperform conventional detection methods. These systems have the unusual capacity to learn from a variety of data inputs, which gives them the potential to spot complex and dynamic dangers that traditional approaches frequently miss. The study does, however, also draw attention to a number of difficulties, such as the requirement for varied and superior datasets, problems with AI models' interpretability, and the processing power needed for efficient implementation. Optimising AI-driven solutions for practical applications requires addressing these constraints. In the end, as cyber threats continue to change, using AI technologies for malware detection improves wireless network security and encourages a proactive defence approach, which improves defence against the more cunning methods that bad actors use to commit cybercrimes.

References

[1] Djenna, A., Bouridane, A., Rubab, S., & Marou, I. M. (2023). Artificial intelligence-based malware detection, analysis, and mitigation. Symmetry, 15(3), 677.

[2] Waqas, M., Tu, S., Halim, Z., Rehman, S. U., Abbas, G., & Abbas, Z. H. (2022). The role of artificial intelligence and machine learning in wireless networks security: Principle, practice and challenges. Artificial Intelligence Review, 55(7), 5215-5261.

[3] Arshad, S. E. U. D., Nasralla, M. M., Khattak, S. B. A., Alhaj, T. A., & Rehman, I. U. (2023). Malware Analysis for IoT and Smart AI-Based Applications. In Cyber Malware: Offensive and Defensive Systems (pp. 165-195). Cham: Springer International Publishing.

[4] Schmitt, M. (2023). Securing the Digital World: Protecting smart infrastructures and digital industries with Artificial Intelligence (AI)-enabled malware and intrusion detection. Journal of Industrial Information Integration, 36, 100520.

[5] Smmarwar, S. K., Gupta, G. P., & Kumar, S. (2023). AI-empowered malware detection system for industrial internet of things. Computers and Electrical Engineering, 108, 108731.

[6] Libri, A., Bartolini, A., & Benini, L. (2020). pAElla: Edge AI-based real-time malware detection in data centers. IEEE Internet of Things Journal, 7(10), 9589-9599.

[7] Ali, M., Hu, Y. F., Luong, D. K., Oguntala, G., Li, J. P., & Abdo, K. (2020, October). Adversarial attacks on ai based intrusion detection system for heterogeneous wireless communications networks. In 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC) (pp. 1-6). IEEE.

[8] Faruk, M. J. H., Shahriar, H., Valero, M., Barsha, F. L., Sobhan, S., Khan, M. A., ... & Wu, F. (2021, December). Malware detection and prevention using artificial intelligence techniques. In 2021 IEEE international conference on big data (big data) (pp. 5369-5377). IEEE.

[9] Arivudainambi, D., KA, V. K., & Visu, P. (2019). Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance. Computer Communications, 147, 50-57.

[10] Sapavath, N. N., Rawat, D. B., & Ghafoor, K. Z. (2020). Advances in AI-Based Security for Internet of Things in Wireless Virtualization Environment. In Security and Organization within IoT and Smart Cities (pp. 41-56). CRC Press.

[11] Vaddadi, S., Arnepalli, P. R., Thatikonda, R., & Padthe, A. (2022). Effective malware detection approach based on deep learning in Cyber-Physical Systems. International Journal of Computer Science and Information Technology, 14(6), 01-12.