International Journal on Emerging Research Areas (IJERA) *ISSN:2230-9993*

DOI:10.5281/zenodo.14670928

Survey on AI Malware Detection Methods and Cybersecurity Education

Syam Gopi
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
syamgopi@amaljyothi.ac.in

Evelyn Susan Jacob
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
evelynsusanj@gmail.com

Joel John
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
joeljohn5112@gmail.com

Raynell Rajeev
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
raynellrajeev007@gmail.com

Steve Alex
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
stevealex365@gmail.com

Abstract— **This paper provides an extensive overview of recent advancements in deep learning-based methods for detecting malware and in programs for educating people about cybersecurity. The overview includes hybrid models, detection based on images, and advanced techniques for extracting features such as texts and images. The main techniques assessed include Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and hybrid models that combine CNNs with Recurrent Neural Networks (RNNs). Furthermore, this article assesses strategies for cybersecurity education, with a focus on engaging employees, providing targeted education for at-risk groups, and integrating digital learning tools. While deep learning models have greatly enhanced the accuracy of malware detection, challenges like the quality of datasets, computational expenses, and adversarial attacks continue to exist. In the field of cybersecurity education, promoting awareness through interactive and gamified techniques has been proven to be effective in creating a more resilient workforce. This overview examines these challenges and suggests future directions, including hybrid models for improved malware detection and scalable digital tools for cybersecurity education.**

Keywords— Cybersecurity, Malware detection, Artificial Intelligence, Convolutional Neural Networks, Cybersecurity Education

I. Introduction

Cyber threats have grown increasingly with criminals constantly developing new tactics to evade security measures. These evolving threats include the use of advanced persistent threats, ransomware, polymorphic malware to exploit human vulnerabilities. Thus cybersecurity has shifted toward using artificial intelligence and machine learning to detect these sophisticated threats. These newly developed systems can automatically detect patterns in malware behavior and adapt to new forms of malware through continuous learning of large data. Deep learning models such as CNNs, LSTMs, and hybrid models, are now widely used to detect hidden and complex patterns in malware datasets.

However, the increasing reliance on AI-powered detection must also be supported by cybersecurity education to create a more resilient defense framework. As most cyber-attacks exploit human errors, it is very important to educate users about cyber threats and practices. By combining advanced malware detection systems with cybersecurity education, organizations can safeguard their network and assets while fostering a security conscious culture.

DOI: **10.5281/zenodo.14670928**

II. Literature Review

A. Malware Detection Approaches

Traditional methods of malware detection have limitations against the modern evolving threats. Methods like signature based detection works on a database consisting of known malware signatures, which fail in identifying threats like zero-day threat. These traditional approaches struggle with the sheer volume and complexity of modern malware. To overcome these limitations we incorporate technologies like Machine Learning and Deep Learning, which can analyse new patterns. Which in turn increases the efficiency of malware detection systems.

- *Convolutional Neural Networks*

Convolutional Neural Networks (CNN) is an artificial neural network which has the ability to recognize patterns in images. CNN provides a dense network which helps in performing efficient identification.

CNN has shown remarkable results in image based malware detection[1]. In image based malware detection, malware binaries are converted into grayscale and RGB images. Grayscale images are effective against obfuscation techniques as they help in understanding the underlying code structure. RGB images can provide more accurate information by using three channels, which highlights the difference between malware families.

CNNs use a convolutional layer to process the input and a pooling layer to reduce the dimensionality of the output from three dimensions to one. CNN architectures like EfficientNet[3] help in achieving high performance with few parameters, which makes it very efficient in speed and memory usage when it comes to large datasets like those involved in malware detection.

- *Long Short-Term Memory*

Long short-term memory is an artificial recurrent neural network (RNN) which processes and analyses sequential data, such as time series, speech and text. It is highly efficient in capturing long-term dependencies, thus making it ideal for sequence prediction tasks [2]. They use a memory cell and gates to control the flow of information, allowing them to selectively retain or discard information as needed and thus avoid the vanishing gradient problem that plagues traditional RNNs [4].

LSTM networks can analyse patterns in time series data or sequences, making them useful for detecting malicious based sequences of system calls, API invocations, or network traffic logs. LSTMs are efficient in identifying behavioural sequences, which is useful in detecting zero-day or previously unseen patterns.

- *Hybrid Models*

By combining different model architectures, hybrid models improve the ability to detect sophisticated malware with higher precision. Hybrid models can process multiple features like images, graphs and sequences thus making them more adaptable to various malware detection scenarios.

In the CNN-BiLSTM model[5], CNNs are adept at capturing spatial patterns from malware data and BiLSTM processes sequential data in both forward and backward directions. This bidirectional processing allows the model to understand relationships in the data that may be missed by unidirectional models, improving detection accuracy for malware that exhibits complex, non-linear behaviours over time.

Another variant, CNN-GRU models [5], pairs CNNs with Gated Recurrent Units (GRUs). CNN is responsible for extracting spatial features from malware binaries and the GRU layers process the extracted features to analyse temporal dependencies, such as sequences of system calls or network traffic data.

Hybrid models with transfer learning utilises existing pre-trained models like ResNet or VGG16. They are fine-tuned on malware datasets to reduce the training time and have increased accuracy. Even though hybrid models have advantages over traditional models, their computational costs remain high.

B. Cybersecurity Education Models

Gamified learning and educational tools have become very prominent in the realm of cybersecurity education. Gamification helps in the integration of game-like elements which creates a dynamic learning environment that promotes active participation and retention of information. These tools aim to foster a culture of vigilance, where users are encouraged to actively participate in cybersecurity defence. For example, simulated phishing attacks, interactive quizzes, and security awareness games are widely used to teach users to identify malicious emails or avoid clicking on harmful links. One other major advantage of gamification is its ability to provide instant feedback.

A study on the impact of gamification based on personality traits revealed that individuals tend to engage more with gamified environments compared to traditional learning models.

DOI: **10.5281/zenodo.14670928**

Cybersecurity games simulate real-world cyber threats in a safe, controlled environment. These games allow users to practise identifying, responding to, and mitigating cyberattacks. For example, many organisations use capture-the-flag (CTF) competitions where participants solve a series of cybersecurity challenges to earn points.

Behavioural tracking is another key component of many gamified cybersecurity platforms. These systems monitor user behaviour, rewarding those who follow security policies consistently and offering additional training or reminders to those who do not.

C. Challenges in Existing Solutions

A critical challenge in the development and training of malware detection systems is the availability of quality of datasets. Most publicly available datasets are either outdated or lack the necessary variety of malware types, especially zero-day or sophisticated variants. Without diverse and representative datasets, models struggle to detect unseen malware.

One other major problem faced is class imbalance in malware datasets, where some malware families have significantly more samples than the others. This can make the model biased to the detection of the more prevalent types, reducing the ability to detect rare or other dangerous malware. Data augmentation or synthetic data generation can reduce the problem of class imbalance to some extent. Advanced evasion techniques like code obfuscation, encryption and polymorphism allow the malware to disguise itself and avoid static analysis tools. Thus, many existing malware detection systems struggle to identify these disguised threats.

Evasive tactics such as packing, where the malware is compressed and encrypted, are designed to defeat static signature-based detection. Adversarial attacks are subtle alterations made to the malware code to trick machine learning models. Thus, models must be robust enough to handle adversarial modifications without being misled by minor changes, which requires ongoing research into more resilient systems.

III. Proposed Methodology

A. Data Collection

Introduction to Malware Datasets

The efficiency of malware detection models significantly relies on the quality and diversity of the dataset. A well-rounded dataset comprises both benign and malicious samples, which is critical for training models to distinguish between normal and malicious behaviour. Balanced datasets can reduce bias and increase the generalisation ability of machine learning models, making them more dependable in real-world settings.

Dataset Sources

Malware datasets are sourced from both public repositories and commercial sources. Common public sources include:

- VirusTotal: A commonly used service that aggregates virus samples.
- Kaggle: Contains many cybersecurity-related datasets, including malware samples.
- VirusShare: Provides access to an extensive collection of malware binaries.

Using real-world malware samples is essential since synthetic data lacks the complexity of actual infections. However, obtaining real malware samples involves legal, ethical, and security challenges, which necessitate careful handling and adherence to cybersecurity norms.

Dataset Composition

A comprehensive malware dataset must contain various malware families (e.g., ransomware, trojans, worms) to ensure that the trained model can generalise across diverse threat types. By integrating multiple malware classes, the model is better suited to detect new and unseen infections. Additionally, a balanced mix of benign data is critical for the model to learn the distinction between normal and harmful behaviour.

Preprocessing and Data Labelling

Preprocessing ensures that the dataset is clean and suitable for training the model. Key preprocessing tasks include:

- Normalisation: Standardising data formats for consistency.
- Deduplication: Removing duplicate samples to avoid bias toward repeated patterns.
- Cleaning Noisy Samples: Identifying and removing corrupted or incomplete data points.

DOI:10.5281/zenodo.14670928

- Accurate Labelling: Ensuring correct labelling of benign or malicious samples is crucial for supervised learning models. Mislabeling can lead to inaccurate predictions.

Class Imbalance and Solutions

In most malware datasets, there is a considerable class imbalance between benign and malicious samples, or between different types of malware. This imbalance can skew the model's predictions. Strategies to address this include:

- Oversampling: Increasing the minority class samples through duplication or synthetic data generation (e.g., SMOTE).
- Undersampling: Reducing the majority class samples to balance the dataset.
- Data Augmentation: Modifying existing samples to provide additional training data (e.g., rotating or flipping malware images).

Balancing the dataset ensures that the model performs well across all classes, reducing bias and improving overall accuracy.

B. Feature Extraction

Text-based Features

- N-grams: N-grams represent contiguous sequences of N objects, such as bytes or words, found in a file or script. In malware analysis, N-grams help capture the sequential patterns of code or operations within malware. Examples:
  - Bigrams: Capture two consecutive API calls or commands.
  - Trigrams: Capture sequences of three objects, which provide deeper context. Larger N values capture more information but increase computational complexity. N-grams enable models to detect common sequences or behaviors associated with malware.
- Word2Vec: Word2Vec translates textual data (e.g., code snippets, API calls) into vector representations that capture semantic relationships. This model learns connections between malware components and helps discover patterns in API calls or malware commands. Word2Vec is particularly effective for detecting malware that injects code or manipulates scripts, as it identifies relationships between commands in context.

Image-based Features

- Grayscale and RGB Images: In image-based malware detection, malware binaries are transformed into visual formats:
  - Grayscale Images: Each byte in a binary file is represented as a pixel, creating a grayscale image.
  - RGB Images: The binary file is represented as an image with three colour channels (Red, Green, Blue) for more detailed representation. Both types of images allow models to detect structural patterns in malware code, often undetectable in raw data.
- Visual Patterns in Malware: Malware exhibits identifiable patterns in visual form, often due to repetitive structures in the binary code. CNNs (Convolutional Neural Networks) can identify these patterns and have proven effective in detecting malware by recognizing visual irregularities in binary structures.

Importance of Combining Text and Image Features

- Text-based Features: Focus on semantic and logical flow, ideal for detecting script-based malware.
- Image-based Features: Focus on structural patterns, effective for detecting binary-based malware.
  Combining both approaches enables a comprehensive analysis of malware, improving detection accuracy by capturing both behavioral and structural elements.

C. Model Architecture

Convolutional Neural Networks (CNNs)

- CNN Basics: CNNs are designed to recognize spatial patterns, making them ideal for image-based malware detection. They consist of:
  - Convolutional Layers: Detect features like edges and textures.
  - Pooling Layers: Reduce data dimensions while retaining essential information. CNNs can identify spatial irregularities in binary structure, differentiating benign and malicious samples.

DOI:10.5281/zenodo.14670928

- Specific Application to Malware Detection: CNN layers capture low-level patterns (e.g., file headers) in early stages and more abstract patterns (e.g., suspicious byte sequences) in later layers.

EfficientNet-B4

- Architecture Overview: EfficientNet-B4 balances network depth, width, and resolution, optimising performance. This balance allows it to achieve high accuracy with fewer parameters compared to other models.
- Why EfficientNet-B4?: It offers superior performance with fewer computational resources. EfficientNet-B4 can also be fine-tuned using transfer learning, adapting pre-trained weights for malware-specific tasks.

LSTM Integration

- LSTM (Long Short-Term Memory): LSTMs capture temporal dependencies in sequential data, such as API call sequences or event logs, and are more robust against the vanishing gradient problem. LSTMs help detect malware that relies on time-dependent behaviour (e.g., delayed actions).
- CNN-LSTM Hybrid Model: Combining CNNs and LSTMs creates a powerful architecture where CNNs extract spatial features from images and LSTMs capture temporal patterns in event sequences.

Training Process

- Data Splitting: The dataset is split into training, validation, and test sets (e.g., 70%-15%-15%).
- Hyperparameter Tuning: Learning rate, batch size, and number of epochs are optimised to achieve the best model performance. Techniques such as grid search are used for this purpose.
- Regularisation: Techniques such as dropout, data augmentation, and early stopping are applied to prevent overfitting and improve generalisation.

D. Cybersecurity Education Module

Importance of Cybersecurity Education

As cyber threats become more frequent and sophisticated, cybersecurity education is essential for reducing risks. By increasing awareness, users can recognize potential threats and take proactive measures, particularly against social engineering and malware.

Gamification Techniques

Gamification integrates game-like elements (e.g., points, badges, levels) to increase learner engagement and motivation. In the education module, learners earn points and badges for completing tasks, progressing through levels, and competing on leaderboards.

Interactive Learning

Quizzes, assessments, and simulated malware analysis environments will provide hands-on experience. Real-life scenarios will help learners apply their knowledge in practical contexts, strengthening their ability to detect and respond to malware.

Target Audience and Learning Outcomes

The education module targets a broad audience, including students, professionals, and general users. Upon completing the module, learners will:

- Understand different types of malware and their operation.
- Recognize social engineering tactics.
- Respond effectively to cybersecurity incidents.

User Feedback and Iteration

Feedback from users will be collected to refine the module, ensuring it remains relevant and effective as cybersecurity challenges evolve.

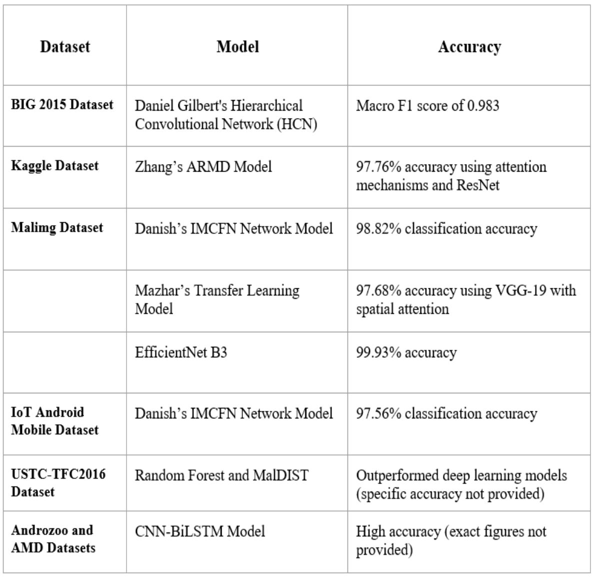
IV. Experiments and Results

The high accuracy of various deep learning models in malware detection across multiple datasets is highlighted in the table.

EfficientNet B3 and Danish's IMCFN Network Model consistently show impressive performance, achieving up to 99.93% and 98.82% accuracy on the Malimg dataset, respectively. Mazhar's VGG-19 model, using transfer learning approaches, also demonstrates strong performance, indicating the effectiveness of pre-trained models in this field.

In encrypted traffic classification with the USTC-TFC2016 dataset, classical machine learning models like Random Forest (RF) and MalDIST exhibit strong results, although their specific accuracies are not provided.

The results across different datasets emphasize that both deep learning and classical methods are capable of achieving high detection accuracy, with some models achieving close to perfect accuracy in malware classification.



Dataset	Model	Accuracy
BIG 2015 Dataset	Daniel Gilbert's Hierarchical Convolutional Network (HCN)	Macro F1 score of 0.983
Kaggle Dataset	Zhang's ARMD Model	97.76% accuracy using attention mechanisms and ResNet
Malimg Dataset	Danish's IMCFN Network Model	98.82% classification accuracy
	Mazhar's Transfer Learning Model	97.68% accuracy using VGG-19 with spatial attention
	EfficientNet B3	99.93% accuracy
IoT Android Mobile Dataset	Danish's IMCFN Network Model	97.56% classification accuracy
USTC-TFC2016 Dataset	Random Forest and MalDIST	Outperformed deep learning models (specific accuracy not provided)
Androzoo and AMD Datasets	CNN-BiLSTM Model	High accuracy (exact figures not provided)

TABLE I

V. DISCUSSION

A. Key Findings

Enhanced Accuracy in Detection: DL models such as CNNs, LSTMs, and hybrid structures show a much higher performance when compared to traditional methods achieving accuracies of over 95%. For example - EfficientNet B3 got a high accuracy of 99.93% on the Malimg dataset.

Advancements in Feature Extraction Techniques: Techniques like N-gram analysis, Control Flow Graphs (CFGs), and representing malware as images have increased the capability of deep learning models to identify even heavily disguised malware. Thus to improve such a feature extraction methods has become a crucial step in malware detection.[1]

Success of Hybrid Models: The combination of various deep learning architectures (e.g., CNNs with LSTMs or GRUs) has resulted in better detection performance. The CNN-BiLSTM model surpassed the CNN-GRU in detecting Android malware.

B. Challenges and Limitations

Dataset Size and Quality: For malware detection deep learning models, the effectiveness depends highly on the diversity and availability of datasets.. The availability of these datasets is directly dictated by private organizations which leads to the problem of access to these datasets. models require large amounts of data that improve training to improve their overall accuracy. An imbalance in the malware family datasets can lead to poor generalization of malware families which in turn affects detection of older and newer families of malware.

Computational Demand: Deep learning models, especially hybrid and ensemble architectures, can have significant computational expenses during training which will be difficult to meet and will be limited to availability of hardware. When there is high energy consumption for the hardware it will eventually lead to environmental sustainability issue as well. For example training a complex model like ResNet-50 for malware detection will lead to high usage of computational resources

Adversarial Attacks: Malware detection models are always prone to adversarial attacks such as the modification of malware to evade detection. Attackers use white box and black box attacks to bypass detection. This will be an active challenge throughout future research as attackers now have access to generative AI to help with such attacks.

C. Implications for future Research

Optimization of Hybrid Model : The importance of hybrid models to improve malware detection should direct future research towards improving them. There should be a balance between accuracy and efficiency in the usage of computational resources. This will include using transfer learning to improve the process of training these models, increasing model efficiency and integrating various improved models to build future hybrid models.

Adversarial Attack Resilience: In order to overcome the problem of adversarial attacks we must focus on adversarial training which involves introducing new kinds of malware into the training dataset. Model hardening which involves gradient masking and defensive distillation will also contribute to solving the problem of adversarial attacks.

D. Cybersecurity Education and Awareness

The review of cybersecurity education frameworks show that raising awareness is crucial for preventing malware attacks. The various methods to promote cybersecurity education is as follows:

Engaging Employees: To encourage a proactive mindset for employees in organisations the most effective ways to train them is by conducting interactive workshops and simulated phishing exercises. These methods use psychological and

DOI: **10.5281/zenodo.14670928**

behaviour based frameworks to engage employees and ensure compliance with cybersecurity protocols.

Tailored Education for At-Risk Groups: Specific demographics such as the elderly are always at risk of cybersecurity attacks so it is important to develop a framework that is customised for them. The Synergistic Cybersecurity Awareness Model for the Elderly (SCASAM- Elderly) is one of the most effective models to enhance knowledge among the elderly by including theory and practical exercises.

Closing the Cybersecurity Skills Gap: In order to improve cybersecurity skills we can have project based, problem based and hands-on models. We can include remote labs, virtual classrooms and Massive Open Online Courses (MOOCs) to improve the overall learning which will help in expanding cybersecurity education.

VI. Conclusion

This literature survey covers various deep learning techniques for detecting malware such as CNNs, LSTMS, hybrid models and the significance of transfer learning for enhancing accuracy. There is a necessity for developing robust models to combat adversarial attacks and obfuscated malware while also improving scalability and computational efficiency. Importance should be given to expanding diverse datasets which will lead to enhanced real-world performance. Cybersecurity awareness also plays a huge role in preventing various malware attacks and the paper focuses on various education frameworks for employees in organisations and vulnerable groups.

References

[1] https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/#:~:text=LSTM%20(Long%20Short%2DTerm%20Memory,ideal%20for%20sequence%20prediction%20tasks.

[2] https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/lstm#:~:text=LSTMs%20are%20able%20to%20process,problem%20that%20plagues%20traditional%20RNNs

[3] Huijuan Wang, Boyan Cui, Quanbo Yuan, Ruonan Shi, Mengying Huang, A review of deep learning based malware detection techniques, Neurocomputing, Volume 598, 2024, 128010, ISSN 0925-2312

[4] Agarap, Abien Fred. (2017). Towards Building an Intelligent Anti-Malware System: A Deep Learning Approach using Support Vector Machine (SVM) for Malware Classification. 10.48550/arXiv.1801.00318.

[5] Halit Bakır, Rezan Bakır,DroidEncoder: Malware detection using auto-encoder based feature extractor and machine learning algorithms,Computers and Electrical Engineering,Volume 110,2023,108804,ISSN 0045-7906.

[6] Adi Lichy, Ofek Bader, Ran Dubin, Amit Dvir, Chen Hajaj,When a RF beats a CNN and GRU, together—A comparison of deep learning and classical machine learning approaches for encrypted malware traffic classification, Computers & Security,Volume 124,2023,103000,ISSN 0167-4048.

[7] Haq, Ikram & Khan, Tamim & Akhunzada, Adnan. (2021). A Dynamic Robust DL-based Model for Android Malware Detection. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3079370.

[8] Rajasekhar Chaganti, Vinayakumar Ravi, Tuan D. Pham,Deep learning based cross architecture internet of things malware detection and classification,Computers & Security,Volume 120,2022,102779,ISSN 0167-4048.

[9] Aslan, Omer & Yılmaz, Abdullah. (2021). A New Malware Classification Framework Based on Deep Learning Algorithms. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3089586.

[10] Yadav, Pooja & Menon, Neeraj & Ravi, Vinayakumar & Vishvanathan, Sowmya & Pham, Tuan. (2022). EfficientNet Convolutional Neural Networks-based Android Malware Detection. Computers & Security. 115. 10.1016/j.cose.2022.102622.

[11] Rajasekhar Chaganti, Vinayakumar Ravi, Tuan D. Pham,Deep learning based cross architecture internet of things malware detection and classification,Computers & Security,Volume 120,2022,102779,ISSN 0167-4048.

[12] Akhtar, Muhammad & Feng, Tao. (2022). Detection of Malware by Deep Learning as CNN-LSTM Machine Learning Techniques in Real Time. Symmetry. 14. 2308. 10.3390/sym14112308.

[13] Algarni, Musaad & Alroobaea, Roobaea & Almotiri, Jasem & Ullah, Syed Sajid & Hussain, Saddam & Umar, Fazlullah. (2022). An Efficient Convolutional Neural Network with Transfer Learning for Malware Classification. Wireless Communications and Mobile Computing. 2022. 1-8. 10.1155/2022/4841741.

[14] Prima, B. & Bouhorma, Mohammed. (2020). USING TRANSFER LEARNING FOR MALWARE CLASSIFICATION. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLIV-4/W3-2020. 343-349. 10.5194/isprs-archives-XLIV-4-W3-2020-343-2020.