CyberSecurity dan Forensik Digital e-ISSN: 2615-8442
Vol. 8, No. 1, Mei 2025, hlm. 72-79

Multiclass Klasifikasi Malware Berbasis Gambar Menggunakan *Vision Transformer Architecture*

Diash Firdaus[1], Idi Sumardi[2], Chalifa Chazar[3] , Muhamad Zufar Dafy[4]

[1,3,4] Informatics, Institut Teknologi Nasional, Bandung, Indonesia
[2] Informatics Engineering, STMIK JABAR, Bandung, Indonesia

[*]Email: [1]Diash@itenas.ac.id, [2]Idis@stmikjabar.ac.id, [3]chalifa@itenas.ac.id, [4]dafyluck@gmail.com

Abstrak

Perkembangan malware yang semakin canggih telah menjadi ancaman serius bagi keamanan siber global, mengakibatkan kerugian finansial yang signifikan. Metode deteksi tradisional seperti deteksi berbasis tanda tangan dan analisis dinamis memiliki keterbatasan dalam mendeteksi varian malware baru. Sebagai solusi inovatif, analisis malware berbasis gambar mengubah file biner malware menjadi representasi gambar, memanfaatkan pemrosesan citra digital dan pembelajaran mesin untuk identifikasi yang lebih efisien. Penelitian ini menggunakan arsitektur Vision Transformer (ViT) untuk klasifikasi malware multikelas berbasis gambar, menawarkan pendekatan yang lebih efektif dibandingkan CNN tradisional seperti EfficientNet dan VGG16. ViT muncul sebagai pendekatan baru yang menarik karena fleksibilitasnya dalam memahami hubungan objek dalam gambar dan mendeteksi pola penting. Dengan kemampuannya mempelajari hubungan jangka panjang antar data, ViT dapat mendeteksi perbedaan halus antara berbagai jenis malware dan mencapai akurasi lebih tinggi. Dataset yang digunakan adalah Malimg, yang merupakan hasil konversi malware biner menjadi format gambar. Hasil penelitian menunjukkan Vision Transformers mencapai akurasi pelatihan 99.96%, validasi 98.05%, dan pengujian 97.49%, meningkatkan akurasi dibandingkan CNN. Keberhasilan ini menunjukkan kemajuan signifikan dalam akurasi deteksi, mengindikasikan arah menjanjikan untuk penelitian dan aplikasi keamanan siber di masa depan. Studi ini menekankan pentingnya teknik pembelajaran mesin yang canggih untuk meningkatkan deteksi malware.

Kata kunci: Vision Transformers, Klasifikasi Malware, Deep learning

Image-Based Malware Multiclass Classification Using Vision Transformer Architecture

Abstract

The advancement of increasingly sophisticated malware has become a serious threat to global cybersecurity, resulting in significant financial losses. Traditional detection methods such as signature-based detection and dynamic analysis have limitations in detecting new malware variants. As an innovative solution, image-based malware analysis transforms malware binary files into image representations, leveraging digital image processing and machine learning for more efficient identification. This research utilizes the Vision Transformer (ViT) architecture for image-based multi-class malware classification, offering a more effective approach compared to traditional CNNs like EfficientNet and VGG16. ViT emerges as an intriguing new approach due to its flexibility in understanding object relationships in images and detecting important patterns. With its ability to learn long-term relationships between data, ViT can detect subtle differences between various types of malware and achieve higher accuracy. The dataset used is Malimg, which is the result of converting binary malware into image format. The study's results show that Vision Transformers achieve a training accuracy of 99.96%, validation accuracy of 98.05%, and testing accuracy of 97.49%, improving accuracy compared to CNNs. This success indicates significant progress in detection accuracy, suggesting a promising direction for future research and cybersecurity applications. This study emphasizes the importance of advanced machine learning techniques to enhance malware detection.

***Keywords**: Vision Transformers, Malware Classification, Deep learning*

1. INTRODUCTION

In recent years, the evolution of malware has become a serious threat to global cybersecurity (Aboaoja et al. 2022). Malware continues to evolve with increasingly with sophisticated techniques, make traditional detection and classification methods less effective (Aslan and Yilmaz 2021). Ransomware, trojans, worms, and spyware attacks are causing significant financial losses and disruptions for both individuals and large organizations (Beaman, Barkworth, and David 2020) (Singh Bhadouria 2022). The proliferation of malware across the

internet and its infection of communication devices has become a severe issue in cybersecurity. In 2020, about 360,000 new malware files were detected each day, with the number growing by 5.2% daily. This highlights the need for better ways to identify and classify malware(Awan et al. 2021).

Traditional methods in malware analysis, such as signature-based detection and dynamic analysis, have several limitations. Signature-based detection relies on a database of known patterns, making it ineffective against new or modified malware variants and depend on Database (Odii, Hampo, and Onwuama 2019) (Basak et al. 2024). Dynamic analysis, which involves running the malware in a controlled environment, is time-consuming and risky if not done properly. These methods struggle to keep up with the increasing amount of malware (Sihwail, Omar, and Ariffin 2018).

One innovative solution to address these limitations is image-based malware analysis. This technique converts malware binary files into image representations, where byte patterns are extracted as pixels (Nataraj et al. 2011). This approach leverages the advantages of digital image processing and machine learning to identify malware characteristics without executing potentially harmful code, making it safer and more efficient (Le et al. 2018).

Deep learning, a part of machine learning, has been successful in areas like computer vision and image classification. Convolutional Neural Networks (CNNs) are often used for this because they can automatically find important features in images. However, CNNs struggle to see relationships between distant parts of an image, which can be important for identifying complex malware patterns (Alam et al. 2021) (He and Kim 2019).

Conventional approaches that rely on signature and heuristic methods to detect malicious software are not effective enough in identifying new or unknown malware. This indicates that Machine Learning (ML) methods could be a solution to this problem. More advanced Deep Learning methods, combined with transfer learning techniques, have been used to enhance resilience and accuracy in detecting malware without requiring deep security knowledge (Alshomrani et al. 2024).

In related research, (Yadav et al. 2022) proposed a two-stage Deep Learning framework for detecting and classifying DEX file images, utilizing the EfficientNetB0 model to extract relevant features from malware images. This method achieved impressive results, with 100% accuracy in binary classification and 92.9% accuracy in five-class classification. Additionally, (Khan et al. 2023) introduced a malware detection framework called Deep Squeezed-Boosted and Ensemble Learning (DSBEL). This framework combines CNN with multi-path dilated convolution operations to capture malicious patterns globally, achieving an accuracy of 98.50% on the image IOT_Malware dataset. Researcher (Asam et al. 2022) introduced a CNN-based architecture called the IoT Malware Detection Architecture (iMDA). This architecture is designed for effective detection using various feature learning schemes, achieving an accuracy of 97.93% on the IoT dataset.

In this study by Rezende et al. [29], they developed a neural network architecture utilizing transfer learning with ResNet-50. They used RGB images sized 224 × 224 and applied the Glorot uniform approach for weight initialization, along with the Adam optimizer, training the model for 750 epochs and achieving a final accuracy of 98.62%. Additionally, they implemented GIST features with kNN, which resulted in an accuracy of 97.48%.

Group researchers, (Khan, Zhang, and Kumar 2019) conducted a comprehensive analysis on the use of transfer learning for malware classification using ResNet and GoogleNet. They set up their data pipeline and identified the best model. The model accuracies achieved by ResNet 18, 34, 50, 101, and 152 were 83%, 86.51%, 86.62%, 85.94%, and 87.98%, respectively.

ViT is a good algorithm for performing multiclass classification for irregular images and has similarities, for example in the following study which performed Multi-Class Classification on X-Ray images (Hadhoud et al. 2024). This suggests that ViT can effectively perform multiclass classification on image-based malware. Additionally, ViT has emerged as an exciting new approach due to its attention mechanism, which provides flexibility in understanding relationships between objects within images and detecting important patterns. With its ability to handle irregular images and learn long-range relationships in the data, ViT can identify subtle differences between various types and subtypes of mage-based malware, leading to higher classification accuracy (Katar and Yıldırım 2024).

ViT can significantly enhance the model's accuracy compared to traditional CNNs. This improvement is primarily driven by the ability of the self-attention mechanism in ViT to capture long-range spatial dependencies and global context across an entire image, a capability often limited in CNN architectures that focus on local features. This capability directly challenges the reliance of CNNs on a strong inductive bias for locality. While this bias makes CNNs highly efficient on smaller datasets, it can also be a constraint. In contrast, ViT, with its weaker bias, has the flexibility to learn unexpected patterns directly from the data, provided that a massive amount of training data is available to guide it

2. MATERIAL AND METHODS

The flowchart illustrates a structured process for research or project development, delineating a sequence of steps from initiation to conclusion. Figure 1 is a flow for the research method.

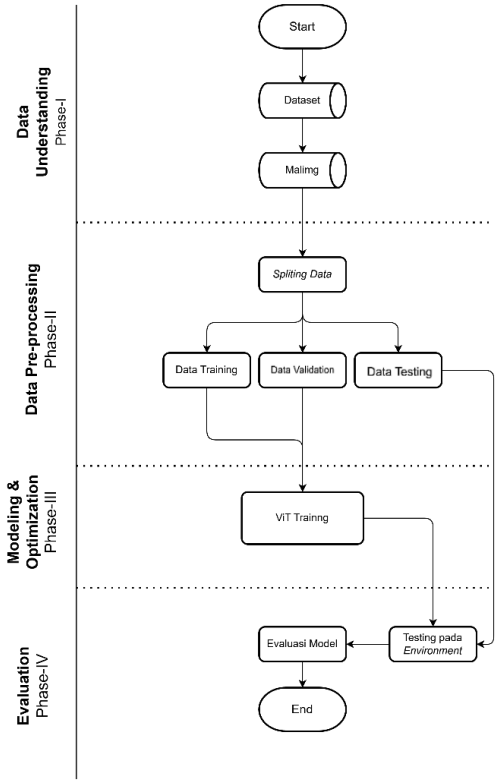

Figure 1. Research Method

In our study, we use the "Malimg" dataset, which includes a wide variety of malware images. We divide this dataset into three parts: 80% training, 10% testing, and 10% validation, to support thorough model development and assessment. We train the model using the Vision Transformer (ViT), which is excellent at identifying long-range patterns, making it ideal for classifying malware. After training, we carefully evaluate the model with the testing dataset to check its accuracy and how well it can handle different types of malware. Finally, we perform environmental testing to confirm the model's effectiveness with validation data, ensuring it works well in real-world situations.

2.1. Schematic diagram of the proposed method

This paper introduces a ViT model for classifying malware images. Rather than using the full images directly, the model divides them into patches and converts them into vectors. This method enhances speed and efficiency by allowing the model to process smaller image segments. A schematic diagram of this proposed method for malware image classification is shown in Figure 2.

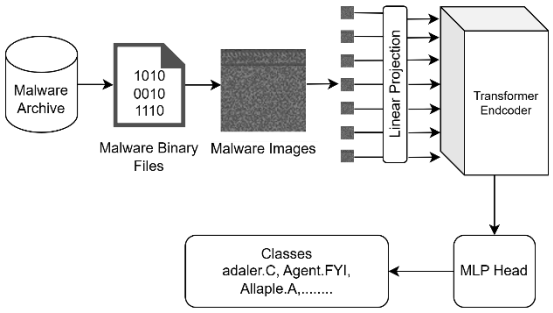

Figure 2. Schematic diagram of proposed method

The success of deep learning models largely depends on the quality, diversity, and size of the dataset used. A well-curated and comprehensive dataset of malware images is essential for the ViT model to learn meaningful patterns and features that differentiate various types of malware. By leveraging a large and varied dataset, the ViT model can effectively extract important patterns and structures from input images. The model's capability to split images into smaller patches and convert them into vectors allows it to utilize spatial relationships in the dataset. This approach captures detailed features within each patch, enabling accurate and efficient classification. Additionally, a diverse dataset enhances the ViT model's ability to generalize. Exposure to a wide range of malware samples with different characteristics helps the model adapt to new and real-world scenarios, thereby increasing the robustness and reliability of the proposed malware image classification method.

2.2. Dataset

The Malimg dataset is a prominent collection used in cybersecurity research, consisting of grayscale images derived from malware binaries. Each image represents a specific malware type, providing a diverse range of categories for comprehensive analysis. Malimg Have 9458 images with 25 class, the dataset offers a substantial amount of data for training and testing machine learning models. By converting malware binaries into visual formats, it allows researchers to apply image processing techniques to identify patterns and enhance malware classification. The dataset is instrumental in developing and evaluating algorithms, helping address cybersecurity challenges by exploring visual patterns for effective malware detection (Nataraj et al. 2011). Figure 3 shows 4 examples from the total of 25 classes in the Malimg dataset.

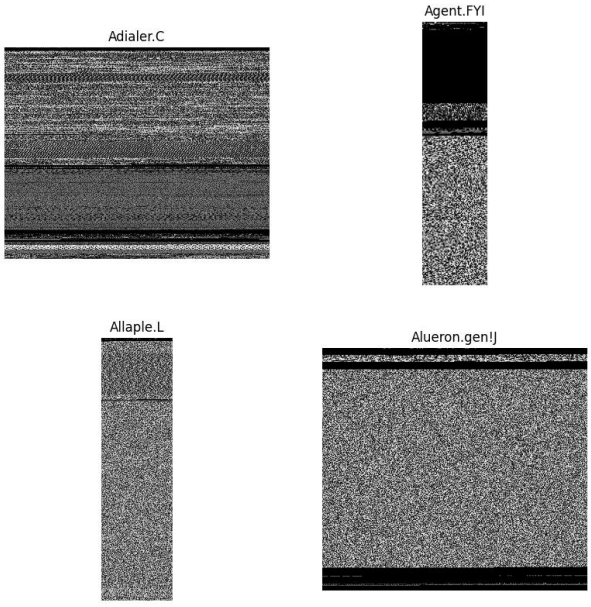

Figure 3. Malimg Dataset (Nataraj et al. 2011)

2.3. Vit Model

The significant impact of transformer networks on natural language processing is well-recognized. Building on the success of the original transformer architecture, (Dosovitskiy et al. 2021) introduced the ViT model, specifically designed for image processing. The ViT model features self-attention blocks and MLP networks, using linear projection and positional embedding to handle input images effectively. The ViT architecture divides input images into fixed-size, non-overlapping patches. These patches are then flattened, and a spatial embedding is applied using linear projection to retain the spatial information of the original image. The resulting vector is processed through a series of N transformer encoder blocks. The structure of these encoder blocks, used for feature extraction in the ViT model, is detailed in Figure 4.

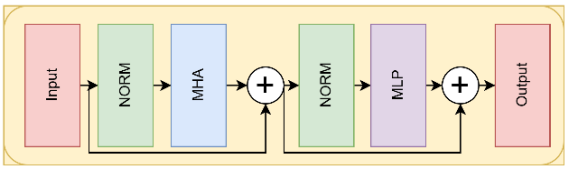

Figure 4. ViT Detail Architecture (Katar and Yıldırım 2024)

To effectively classify malware images, the ViT model is designed with a set of carefully chosen parameters. These parameters ensure the model's ability to process and analyze the data efficiently while maintaining high accuracy. Table 1 outlines the parameters used for the ViT model designed for malware image classification

Table 1. Environment Setup

No	Parameter	Value
1	image_size	256
2	patch_size	32
3	num_classes	25
4	dim	1024
5	depth	6
6	heads	16
7	mlp_dim	2048
8	dropout	0.1
9	emb_dropout	0.1
10	epoch	10

. The input image size is set to 256x256 pixels, enabling uniform processing. Each image is divided into 32x32 pixel patches, creating 64 patches per image for detailed analysis. The model categorizes images into 25 output classes, representing various malware types. Each patch is transformed into a 1024-dimensional embedding after linear projection, capturing rich information for further processing. With 6 Transformer Encoder layers, each containing Multi-Head Attention and MLP, the model captures complex patterns effectively using 16 attention heads. The MLP block's hidden layer size is 2048, supporting robust computation. To enhance generalization, a dropout rate of 0.1 is applied within the Transformer layers, along with an embedding dropout rate of 0.1 before the Transformer Encoder. These parameters collectively contribute to the model's efficiency and reliability in classifying malware images.

2.4. Environment Setup

In this stage, we focus on setting up the development environment necessary for the research. This includes the hardware, software, and tools required. Ensuring that all technical components are ready is crucial for supporting model development and testing. By having everything well-integrated, the process of developing and evaluating the model can be carried out efficiently and effectively.

Table 2. Environment Setup

No	Name	Version
1	Operating System	Windows 11 Ubuntu 22
2	Language	Python 3
3	Tools	Google Colab
4	Library	Os
		Numpy
		matplotlib.pyplot
		Seaborn
		PIL
		Torch
		ViT_pytorch
		Tqdm

These libraries collectively support data processing, visualizatio

2.5. Testing dan Performance Evaluation

The model evaluation is conducted to determine how effectively the model can perform detection. This evaluation process involves measuring detection errors using various methods, including the Classification Report and Confusion Matrix, to test the classification performance of the developed algorithm (Firdaus and Rianti 2023). Calculating accuracy, precision, recall, and F1-Score is a crucial step in assessing the algorithm's performance to determine the model's accuracy level. The results of these calculations are presented in Figure 5.

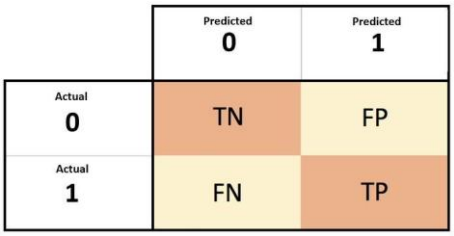

Figure 5. Confussion Matrix (Firdaus and Rianti 2023)

True Positive (TP) refers to predictions that are correctly identified as positive, where both the predicted and actual values are positive. True Negative (TN) refers to predictions that are correctly identified as negative, where both the predicted and actual values are negative. False Positive (FP) occurs when the prediction is positive, but the actual value is negative. False Negative (FN) is when the prediction is negative, but the actual value is positive (Firdaus, Munadi, and Purwanto 2020). The formulas for classification evaluation can be found in Equations (1), (2), (3), and (4) below.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1\text{-Score} = \frac{2\times(\text{Presisi}\times\text{Recall})}{\text{Presisi}+\text{Recall}} \quad (4)$$

3. RESULT AND DISCUSSION

The ViT model was trained in the Google Colab environment utilizing samples from the Malimg dataset. By employing optimized ViT weights as the initial parameters instead of random weights, the model was able to achieve high accuracy rates in a relatively short duration. The early stopping function determined that the validation accuracy of 98.04%, reached during the 7 th epoch, would serve as the stopping criterion, as there was no further improvement in the following twenty epochs. The weights obtained during this process were preserved for application in the test phase. Additionally, the performance curves of the ViT model throughout the training and validation phases are illustrated in Figure 6

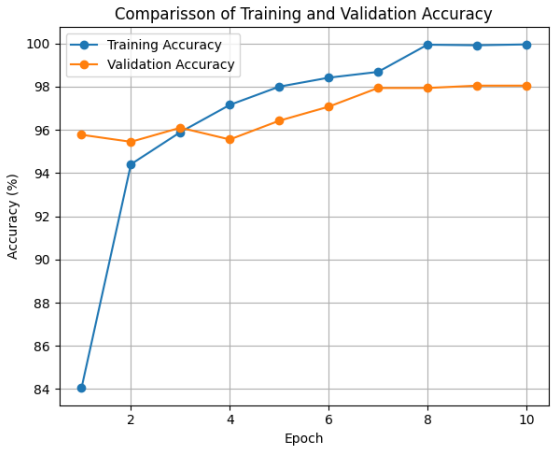

Figure 6. Comparison of Training and Validation Accuracy

Figure 7 shows the comparison of training and validation loss over 10 epochs. The training loss decreases sharply from the first to the second epoch, indicating rapid initial learning, and continues to decline, suggesting good fitting to the training data. The validation loss initially increases slightly around the second epoch, indicating potential early overfitting, but then stabilizes and gradually decreases, showing improved generalization to unseen data. Both losses decrease and stabilize, indicating effective learning with minimal overfitting, as evidenced by the small gap between the two losses.

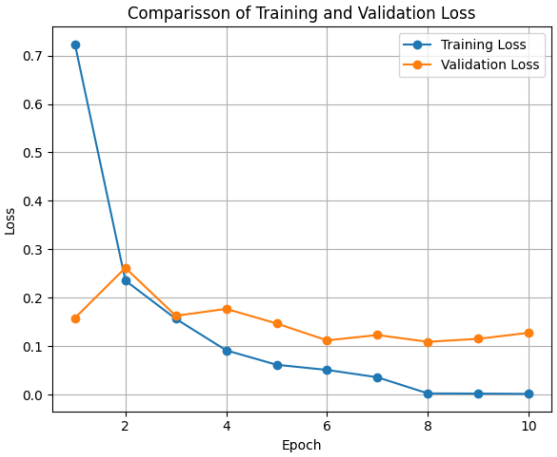

Figure 7. Comparison of Training and Validation Loss

Hasil dari Training dan testing model dievaluasi menggunakan confusin matrix yang bisa dilihat pada Figure. 8

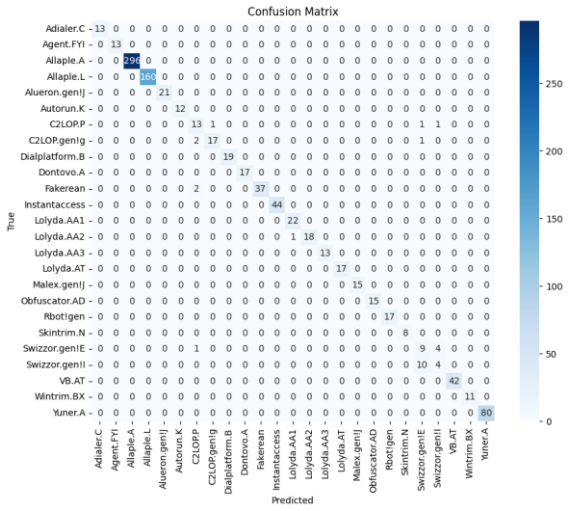

Figure 8. Confusion Matrix Malware Classification

The confusion matrix provides a detailed overview of the model's performance across different classes. Each row represents the actual class, while each column represents the predicted class. The diagonal elements indicate correct predictions for each class. For instance, the model accurately predicted 296 instances of the "Allaple.A" class and 160 instances of the "Allaple.L" class. Off-diagonal elements represent misclassifications. For example, the model predicted "Allaple.L" instead of "Allaple.A" in 0 instances. The matrix highlights strong performance for most classes, with high true positive rates and minimal confusion between different classes. This indicates effective classification with few errors, as shown by the concentrated values along the diagonal.

The confusion matrix indicates that the model performs well in classifying most classes with high accuracy. Most predictions fall on the diagonal elements, reflecting a low error rate. There are few misclassifications, meaning the model can effectively distinguish between classes. However, a few classes might need additional attention to further reduce errors. Overall, the model demonstrates strong and reliable detection capabilities for the task.

Based on the accuracy results from training, validation, and testing using the ViT architecture, the outcomes can be seen in Table 3. The highest training accuracy achieved is 99.96%, the highest validation accuracy is 98.05%, and the testing accuracy is 97.49%. This indicates that the model has a good accuracy level in classifying malware.

Table 3. Comparisson Train and Validation Accuracy

Epoch	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss
1	84.05%	95.77%	0.7233	0.1584
2	94.41%	95.45%	0.2356	0.2618
3	95.90%	96.10%	0.1567	0.1629
4	97.16%	95.56%	0.0916	0.1772
5	98.00%	96.42%	0.0612	0.1470
6	98.42%	97.07%	0.0512	0.1121
7	98.69%	97.94%	0.0360	0.1232
8	99.95%	97.94%	0.0028	0.1092
9	99.92%	98.05%	0.0025	0.1115
10	99.96%	98.05%	0.0018	0.1278

In cybersecurity, accurately classifying malware images is crucial for identifying and mitigating a wide range of digital threats. Traditionally, this task was labor-intensive, but integrating machine learning to automate key processes is now essential, especially in information security. With advancements in deep learning, particularly through CNNs, new architectures continue to emerge, enhancing threat detection capabilities. This study explores the use of ViTs, which have recently gained popularity, for high-accuracy malware classification. Table 4 highlights similar research about Malware Detetcion. Researcher (Awan et al. 2021) Using spatial attention CNN and VGG architecture achieved 97.42% accuracy. For instance, (Patil et al. 2021) achieved 93.00% accuracy using RF, 93.70% for EfficientNet-B0, and 92.00% for VGG-16 models. Additionally, researcher (Naeem et al. 2020) using Kernel-based ELM with statistical texture features achieved 94.25% accuracy.

Table 4. Comparison Accuracy with another architecture

Researcher	Methodology	Accuracy
(Awan et al. 2021)	Spatial Attention CNN and VGG Architecture	97.42%
(Patil et al. 2021)	RF	93.00%
	EfficientNet-B0	93.70%
	VGG-16 Models	92.00%
(Naeem et al. 2020)	Kernel-based ELM with Statistical Texture Features	94.25%
Our Proposed	**Vision Transformers**	**98.05%**

4. CONCLUSION

The research highlights various methodologies for malware detection, demonstrating the effectiveness of different architectural approaches. achieved notable success with spatial attention CNN and VGG architecture, reaching an accuracy of 97.42%. another research explored multiple models, with EfficientNet-B0 achieving the highest accuracy at 93.70%. then another researcher (2020) effectively utilized Kernel-based ELM with statistical texture features, achieving 94.25% accuracy. The proposed methodology using Vision Transformers outperformed the others, achieving an impressive accuracy of 98.05%. This indicates that Vision Transformers offer significant advancements in accurately for image-based multiclass classification malware, suggesting a promising direction for future research and application in cybersecurity. Overall, this study underscores the importance of leveraging

advanced machine learning techniques to enhance malware detection capabilities

To improve the accuracy of our research, consider implementing a hybrid approach by combining Vision Transformers (ViT) with other Deep Learning architecture like efficientnet, mobilenet, and so on. Additionally, expand the dataset by incorporating more image-based malware datasets to create a comprehensive training set, enhancing the model's ability to generalize across various types of malware. Apply data augmentation techniques to increase variability and robustness, and use advanced cross-validation methods to ensure consistent performance. Finally, experiment with different configurations and hyperparameters in your hybrid model to find the optimal settings for maximum accuracy.

REFERENCES

Aboaoja, Faitouri A., Anazida Zainal, Fuad A. Ghaleb, Bander Ali Saleh Al-rimy, Taiseer Abdalla Elfadil Eisa, and Asma Abbas Hassan Elnour. 2022. "Malware Detection Issues, Challenges, and Future Directions: A Survey." *Applied Sciences (Switzerland)* 12(17). doi: 10.3390/app12178482.

Alam, Mehmood, Adeel Akram, Talha Saeed, and Sobia Arshad. 2021. "DeepMalware: A Deep Learning Based Malware Images Classification." *2021 International Conference on Cyber Warfare and Security, ICCWS 2021 - Proceedings* (February):93–99. doi: 10.1109/ICCWS53234.2021.9703021.

Alshomrani, Mohammed, Aiiad Albeshri, Badraddin Alturki, Fouad Shoie Alallah, and Abdulaziz A. Alsulami. 2024. "Survey of Transformer-Based Malicious Software Detection Systems." *Electronics (Switzerland)* 13(23):1–34. doi: 10.3390/electronics13234677.

Asam, Muhammad, Saddam Hussain Khan, Altaf Akbar, Sameena Bibi, Tauseef Jamal, Asifullah Khan, Usman Ghafoor, and Muhammad Raheel Bhutta. 2022. "IoT Malware Detection Architecture Using a Novel Channel Boosted and Squeezed CNN." *Scientific Reports* 12(1):1–13. doi: 10.1038/s41598-022-18936-9.

Aslan, Omer, and Abdullah Asim Yilmaz. 2021. "A New Malware Classification Framework Based on Deep Learning Algorithms." *IEEE Access* 9:87936–51. doi: 10.1109/ACCESS.2021.3089586.

Awan, Mazhar Javed, Osama Ahmed Masood, Mazin Abed Mohammed, Awais Yasin, Azlan Mohd Zain, Robertas Damaševičius, and Karrar Hameed Abdulkareem. 2021. "Image-based Malware Classification Using Vgg19 Network and Spatial Convolutional Attention." *Electronics (Switzerland)* 10(19). doi: 10.3390/electronics10192444.

Basak, Mainak, Dong Wook Kim, Myung Mook Han, and Gun Yoon Shin. 2024. "Attention-Based Malware Detection Model by Visualizing Latent Features Through Dynamic Residual Kernel Network." *Sensors* 24(24). doi: 10.3390/s24247953.

Beaman, Craig, Ashley Barkworth, and Toluwalope David. 2020. "Ransomware: Recent Advances, Analysis, Challenges and Future Research Directions." (January).

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. "An Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale." *ICLR 2021 - 9th International Conference on Learning Representations*.

Firdaus, Diash, Rendy Munadi, and Yudha Purwanto. 2020. "DDoS Attack Detection in Software Defined Network Using Ensemble K-Means++ and Random Forest." *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020* 164–69. doi: 10.1109/ISRITI51436.2020.9315521.

Firdaus, Diash, and Resa Rianti. 2023. "DETEKSI ANOMALI DAN SERANGAN LOW RATE DDOS DALAM LALU LINTAS JARINGAN MENGGUNAKAN NAIVE BAYES." 05(02):140–48.

Hadhoud, Yousra, Tahar Mekhaznia, Akram Bennour, Mohamed Amroune, Neesrin Ali Kurdi, Abdulaziz Hadi Aborujilah, and Mohammed Al-Sarem. 2024. "From Binary to Multi-Class Classification: A Two-Step Hybrid CNN-ViT Model for Chest Disease Classification Based on X-Ray Images." *Diagnostics* 14(23):1–16. doi: 10.3390/diagnostics14232754.

He, Ke, and Dong Seong Kim. 2019. "Malware Detection with Malware Images Using Deep Learning Techniques." *Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019* 95–102. doi: 10.1109/TrustCom/BigDataSE.2019.00022.

Katar, Oğuzhan, and Özal Yıldırım. 2024. "Classification of Malware Images Using Fine-Tunned ViT." *Sakarya University Journal of Computer and Information Sciences* 7(1):22–35. doi: 10.35377/saucis...1341082.

Khan, Riaz Ullah, Xiaosong Zhang, and Rajesh Kumar. 2019. "Analysis of ResNet and

GoogleNet Models for Malware Detection." *Journal of Computer Virology and Hacking Techniques* 15(1):29–37. doi: 10.1007/s11416-018-0324-z.

Khan, Saddam Hussain, Tahani Jaser Alahmadi, Wasi Ullah, Javed Iqbal, Azizur Rahim, Hend Khalid Alkahtani, Wajdi Alghamdi, and Alaa Omran Almagrabi. 2023. "A New Deep Boosted CNN and Ensemble Learning Based IoT Malware Detection." *Computers and Security* 133(January):103385. doi: 10.1016/j.cose.2023.103385.

Le, Quan, Oisín Boydell, Brian Mac Namee, and Mark Scanlon. 2018. "Deep Learning at the Shallow End: Malware Classification for Non-Domain Experts." *Proceedings of the Digital Forensic Research Conference, DFRWS 2018 USA* 26:S118–26. doi: 10.1016/j.diin.2018.04.024.

Naeem, Hamad, Farhan Ullah, Muhammad Rashid Naeem, Shehzad Khalid, Danish Vasan, Sohail Jabbar, and Saqib Saeed. 2020. "Malware Detection in Industrial Internet of Things Based on Hybrid Image Visualization and Deep Learning Model." *Ad Hoc Networks* 105:102154. doi: 10.1016/J.ADHOC.2020.102154.

Nataraj, L., S. Karthikeyan, G. Jacob, and B. S. Manjunath. 2011. "Malware Images: Visualization and Automatic Classification." *ACM International Conference Proceeding Series* (July). doi: 10.1145/2016904.2016908.

Odii, Juliet, JohnPaul Hampo, and Nwokoma Onwuama. 2019. "Comparative Analysis of Malware Detection Techniques Using Signature, Behaviour and Heuristics." *International Journal of Computer Science and Information Security*, Vol. 17(March):33–50.

Patil, Shruti, Vijayakumar Varadarajan, Devika Walimbe, Siddharth Gulechha, Sushant Shenoy, Aditya Raina, and Ketan Kotecha. 2021. "Improving the Robustness of Ai-Based Malware Detection Using Adversarial Machine Learning." *Algorithms* 14(10). doi: 10.3390/a14100297.

Sihwail, Rami, Khairuddin Omar, and K. A. Z. Ariffin. 2018. "A Survey on Malware Analysis Techniques: Static, Dynamic, Hybrid and Memory Analysis." *International Journal on Advanced Science, Engineering and Information Technology* 8(4–2):1662–71. doi: 10.18517/ijaseit.8.4-2.6827.

Singh Bhadouria, Aashi. 2022. "Study of: Impact of Malicious Attacks and Data Breach on the Growth and Performance of the Company and Few of the World's Biggest Data Breaches." *International Journal of Scientific and Research Publications* X(September):1–11. doi: 10.29322/IJSRP.X.2022.p091095.

Yadav, Pooja, Neeraj Menon, Vinayakumar Ravi, Sowmya Vishvanathan, and Tuan D. Pham. 2022. "A Two-Stage Deep Learning Framework for Image-Based Android Malware Detection and Variant Classification." *Computational Intelligence* 38(5):1748–71. doi: 10.1111/coin.12532.