НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2024 Том 24 № 5 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
September–October 2024 Vol. 24 No 5 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2024-24-5-849-857

Obfuscated malware detection using deep neural network with ANOVA feature selection on CIC-MalMem-2022 dataset

Mourad Hadjila[1]✉, Mohammed Merzoug[2], Wafaa Ferhi[3], Djillali Moussaoui[4], Al Baraa Bouidaine[5], Mohammed Hicham Hachemi[6]

[1,2,3,4,5] University of Tlemcen, Tlemcen, 13000, Algeria
[6] University of Oran, Oran, 31000, Algeria

[1] mhadjila.2009@gmail.com✉, https://orcid.org/0000-0002-6554-3925
[2] mohammed.merzoug@univ-tlemcen.dz, https://orcid.org/0009-0002-9117-047X
[3] wafaa.ferhi@univ-tlemcen.dz, https://orcid.org/0009-0005-7574-8368
[4] djilali.moussaoui@univ-tlemcen.dz, https://orcid.org/0000-0003-3478-263X
[5] albaraa.bouidaine@univ-tlemcen.dz, https://orcid.org/0009-0005-2204-9117
[6] hicham.hachemi@univ-usto.dz, https://orcid.org/0000-0003-3967-6609

Abstract

Malware analysis is the process of dissecting malicious software to understand its functionality, behavior, and potential risks. Artificial Intelligence (AI) and deep learning are ushering in a new era of automated, intelligent, and adaptive malware analysis. This convergence of AI and deep learning promises to revolutionize the way cybersecurity professionals detect, analyze and respond to malware threats. This paper proposed a Deep Neural Network (DNN) model built from features selected by ANalysis Of Variance (ANOVA) F-test (DNN-ANOVA) to increase accuracy by identifying informative features. ANOVA is a feature selection method used for numerical input data when the target variable is categorical. The top k most relevant features are those whose score values are greater than a certain threshold equal to the ratio between the sum of all features scores and the total number of features. Experiments are conducted on CIC-MalMem-2022 dataset. Malware Analysis is performed using binary classification to detect the presence or absence of malware and multiclass classification to detect not only the malware but also its type. According to the test results, DNN-ANOVA model achieves best values of 100 %, 99.99 %, 99.99 %, and 99.98 % in terms of precision, accuracy, F1-score and recall respectively for binary classification. In addition, DNN-ANOVA outperforms the current works with an overall accuracy rate of 85.83 %, and 73.98 % for family attacks and individual attacks respectively in the case of multiclass classification.

Keywords

malware detection, deep learning, ANOVA feature selection, binary classification

For citation: Hadjila M., Merzoug M., Ferhi W., Moussaoui D., Bouidaine A.B., Hachemi M.H. Obfuscated malware detection using deep neural network with ANOVA feature selection on CIC-MalMem-2022 dataset. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 5, pp. 849–857. doi: 10.17586/2226-1494-2024-24-5-849-857

УДК 004.491

Обнаружение скрытого вредоносного программного обеспечения с использованием глубокой нейронной сети с выбором признаков ANOVA на наборе данных CIC-MalMem-2022

Мурад Хаджила[1]✉, Мохаммед Мерзуг[2], Вафа Ферхи[3], Джилали Муссауи[4], Аль Бараа Буйден[5], Мохаммед Хишам Хашеми[6]

[1,2,3,4,5] Университет Тлемсена, Тлемсен, 13000, Алжир
[6] Университет Орана, Оран, 31000, Алжир

[1] mhadjila.2009@gmail.com✉, https://orcid.org/0000-0002-6554-3925
[2] mohammed.merzoug@univ-tlemcen.dz, https://orcid.org/0009-0002-9117-047X
[3] wafaa.ferhi@univ-tlemcen.dz, https://orcid.org/0009-0005-7574-8368
[4] djilali.moussaoui@univ-tlemcen.dz, https://orcid.org/0000-0003-3478-263X
[5] albaraa.bouidaine@univ-tlemcen.dz, https://orcid.org/0009-0005-2204-9117
[6] hicham.hachemi@univ-usto.dz, https://orcid.org/0000-0003-3967-6609

Аннотация

Анализ вредоносного программного обеспечения включает исследование функциональности, поведения и потенциальных рисков. Искусственный интеллект и глубокое обучение открывают возможности автоматизированного, интеллектуального и адаптивного анализа вредоносного программного обеспечения. В работе предлагается модель глубокой нейронной сети (Deep Neural Network, DNN), созданная на основе признаков, выбранных с помощью F-теста дисперсионного анализа (ANalysis Of VAriance, ANOVA), для повышения точности распознавания путем выявления информативных признаков. DNN-ANOVA представляет собой метод выбора признаков, используемый для анализа числовых входных данных, когда целевая переменная является категориальной. К наиболее релевантным признакам относятся те, значения оценки которых превышают определенный порог, равный отношению суммы оценок всех признаков к общему числу признаков. Эксперименты выполнены на наборе данных CIC-MalMem-2022. Проведен анализ обнаружения или отсутствия вредоносного программного обеспечения с использованием бинарной классификации, а также полиномиальной классификации для определения его типа. Согласно результатам F-теста, модель DNN-ANOVA достигает наилучших значений: 100 % — precision, 99,99 % — accuracy, 99, 99 % — F1-score и 99,98 % — recall для бинарной классификации. Кроме того, DNN-ANOVA превосходит текущие работы с общим показателем точности (accuracy) 85,83 % для групповых атак и 73,98 % для индивидуальных атак в случае полиномиальной классификации.

Ключевые слова

обнаружение вредоносного программного обеспечения, глубокое обучение, выбор признаков ANOVA, бинарная классификация, полиномиальная классификация, набор данных

Ссылка для цитирования: Хаджила М., Мерзуг М., Ферхи В., Муссауи Д., Буйден А.Б., Хашеми М.Х. Обнаружение скрытого вредоносного программного обеспечения с использованием глубокой нейронной сети с выбором признаков ANOVA на наборе данных CIC-MalMem-2022 // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 5. С. 849–857 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-5-849-857

Introduction

Malware, or malicious software, refers to various programs designed to harm devices, steal data, or monitor user activities without their knowledge [1]. It includes worms, viruses, trojans, ransomware, and spyware, which can infect networks and devices [2]. Malware spreads through methods like USB drives, collaboration tools, and drive-by downloads, and is used for purposes, such as stealing passwords [3], accessing confidential data [4], and deceiving governments.

The rapid increase in internet-connected devices has posed significant challenges for malware analysts, researchers, and antivirus companies [5]. Cybercriminals increasingly use the internet for illegal activities, such as financial fraud, data theft, and unauthorized access to systems. In response, researchers have developed new security measures to combat evolving malware. The term “malware” now broadly encompasses any malicious program that compromises the confidentiality, integrity, and availability of systems, networks, or services, whereas previously all such programs were referred to as computer viruses.

The cyber world is essential to daily life, enabling services like networking, banking, and shopping. However, it also brings significant threats, with malware being one of the most dangerous. The growing volume and complexity of malware make it increasingly difficult to identify and classify [6]. Traditional methods like signature-based detection, along with static and dynamic analysis, are becoming less effective against new malware variants. As a result, there is growing interest and ongoing research in developing more effective methods for malware classification and detection [7].

Artificial Intelligence (AI) has experienced various cycles of interest and decline since its beginnings in the late 1940s, but recent advancements in Machine Learning (ML) and Deep Learning (DL) have solidified its role as a transformative technology. AI now impacts nearly every aspect of life, including information security [8], where it is particularly valuable in combating the vast and costly threat of malware. AI-powered algorithms can analyze

file behavior to detect and classify malware, reducing the workload on security analysts and speeding up detection. Despite its effectiveness, AI has limitations such as false positives, but its role in malware detection is expected to grow as the technology evolves [9].

This paper proposes developing a deep neural network to detect and classify malware using the CIC-MalMem-2022 dataset. To improve accuracy, the study involves a pre-processing step that includes feature selection using the ANalysis Of Variance (ANOVA) F-test to identify the most relevant features. The analysis covers both binary and multiclass classification, targeting family and individual malware attacks. The Python scikit-learn library is used for implementing the feature selection method.

Related works

Several studies have explored malware analysis using ML. One such study [10] evaluates various ML classifiers for anomaly detection, malware detection, and Intrusion Detection Systems (IDS). The authors also propose a methodology to integrate these ML models into a real-world network security framework, discussing the challenges faced during implementation and the solutions developed to address them. The authors in [11] focus on using ML to detect malware by analyzing memory dumps. They created a new dataset that includes various ransomware types (like BlackCat and REvil) and benign samples. Different ML models were tested, with XGBoost performing the best. The study in [12] introduces a Multi-Attack Detection Framework (MMAD) combining DL methods — Convolutional Neural Network (CNN), Deep Neural Network (DNN), E2E architectures, Recurrent Neural Network (RNN), and Multi Level Platform (MLP) — to detect 11 malware types effectively. Review [13] summarizes and evaluates recent ML and DL contributions to malware detection and suggests future research directions. Sharjeel et al. in [14] propose a DL-based classification method for malware detection in IoT environments. Their methodology includes preprocessing (data cleaning, scaling, etc.), applying an ensemble of CNN and Long Short-Term Memory (LSTM) techniques, and comparing model performance. Xiaofei et al. [15] introduce a new malware detection model using dimensionality reduction and auto-encoder techniques tested on an Android dataset with promising results. Akhtar and Feng [16] developed a malware detection system using Decision Tree (DT), CNN, and Support Vector Machine (SVM). DT achieved the highest accuracy at 99 %, followed by CNN at 98.76 %, and SVM at 96.41 %. Shafin et al. [17] proposed a system combining CNNs and bidirectional LSTMs to detect obfuscated memory malware on resource-constrained devices, using models called CompactCBL and RobustCBL. Mezina and Burget [18] introduced an extended convolutional neural network for classifying obfuscated malware in memory dumps, outperforming traditional classifiers like Decision Tree, SVM, and Random Forest in both binary and multi-class classification.

ANOVA F-Test Feature Selection

Feature selection is the process of reducing the number of input variables to those that are believed to be most useful for predicting the target variable. The goal is to decrease both the computational cost of modeling and, in many cases, to enhance the model performance [19]. Feature selection is typically straightforward when dealing with real-valued input and output data, as it can be done using methods like Pearson's correlation coefficient. However, it can be more challenging when working with numerical input data and a categorical target variable [20]. When dealing with numerical input data and a categorical target variable, two of the most commonly used feature selection methods are the ANOVA F-test statistic and the mutual information statistic [21]. ANOVA is a parametric statistical hypothesis test for determining whether the means from two or more samples of data come from the same distribution or not. The results of this test can be used for feature selection where those features that are independent of the target variable can be removed from the dataset [19].

Methodology

Dataset description

The obfuscated malware dataset is designed to evaluate memory-based detection techniques for concealed malicious software. It includes common malware types like Spyware, Ransomware, and Trojan Horses, making it suitable for testing detection systems [22]. To closely mimic real-world conditions, the dataset uses a debug mode during memory dumps, ensuring the process is not visible, similar to an average user's experience during a malware attack. The dataset is balanced with 50 % benign and 50 % malicious samples, totaling 58,596 records, evenly split between 29,298 benign and 29,298 malicious memory dumps.

The CIC-MalMem-2022 dataset includes three main malware families: Spyware, Trojans, and Ransomware.

— **Spyware.** Infiltrates systems without user consent, monitors activities, and collects data, posing serious risks like identity theft and data breaches [23–25].

— **Trojan.** Disguises itself as legitimate software to deceive users and gain unauthorized access, often used to steal financial information [26, 27].

— **Ransomware.** Encrypts or locks systems, demanding payment to restore access. Typically financially motivated, but can also have political or ideological motives. Prevention and increasing attack difficulty are key protective measures [28–31].

Data pre-processing

Before starting pre-processing task, it is typically to use a separation of data into training and test set. The training set is into validation set and training set. Test size and validation_split arguments are set to 0.2 and 0.2 respectively.

In this work, CIC-MalMem-2022 dataset contains 55 features and two targets called "Class" and "Category". "Class" target is used for binary classification while "Category" target is used for multiclass classification.

A bar chart of the feature importance scores for each input feature is created (Fig. 1). This clearly shows that some features might be the most relevant (according to test statistic) when comparing to others and that are the most relevant. Since the used dataset has numerical inputs and a categorical output, we opted for the ANOVA method where a threshold value for selecting the most relevant features is given according to the following formula:

$$Threshold = \sum_{i=0}^{N-1} \text{score}(i)/N,$$

where score(i) is the test statistic value for feature i (value indicated on the y-axis of Fig. 1), and N is the number of features where the value is different to "nan".

For each dataset feature, score parameter represents the F-statistic calculated as [32]:

$$F = MSB/MSE,$$

where MSB is the Mean Sum of Squares between the groups, and MSE is the Error Mean Sum of Squares.

MSB is calculated by dividing the Sum of Squares (SS) between the groups by the between group degrees of freedom. MSB is given by the following:

$$MSB = SS(Between)/(m - 1),$$

where SS(*Between*) is the Sum of Squares between the group means and the grand mean.

There are $m - 1$ degrees of freedom associated with the factor of interest when m groups are compared.

MSE is calculated by dividing the Sum of Squares within the groups by the error degrees of freedom. MSE is given by the following:

$$MSE = SS(Error)/(n - m),$$

where SS(*Error*) is the sum of squares between the data and the group means. There are $n - m$ degrees of freedom of error when n total data points are collected and m groups are compared.

The k most relevant features are calculated with the following equation:

$$k = \{Features | score > Threshold\}.$$

In the dataset, the 16 most relevant features were selected using the ANOVA F-test. These features were retained based on their high relevance scores, as shown in Fig. 1, which also includes a threshold value. Since all features are numerical, feature scaling was essential to normalize the data, making the values comparable. The categorical target variable, "Class" was encoded for binary classification (Benign = 0, Malware = 1). For multiclass classification, the "Category" target values were encoded as 0, 1, 2, and 3, representing benign, trojan, ransomware, and spyware, respectively. For more detailed classification, the "Category" values were encoded from 0 to 15.

Proposed architecture

The model is a sequential neural network with four hidden layers. The first hidden layer has 512 neurons and expects 16 input variables, while the subsequent layers have 256, 128, and 64 neurons. The output layer varies depending on the classification task: 1 neuron for binary classification, 4 neurons for detecting malware families, and 16 neurons for detecting individual malware types. The ReLU activation function is used for hidden layers, with sigmoid for binary output and softmax for multiclass output. L2 regularization is applied to prevent overfitting. The model is trained for 500 epochs with a batch size of 256, using accuracy as the evaluation metric and Adam optimizer with a 0.001 learning rate. The loss functions used are binary cross-entropy and categorical cross-entropy, depending on the task.

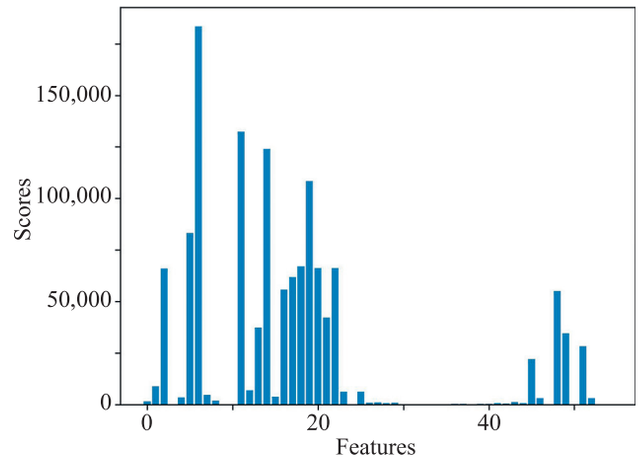

Fig. 1. Bar Chart of the Input Features vs. the ANOVA F-test Feature Importance

Results and discussion

Binary classification

The CIC-MalMem-2022 dataset is a balanced dataset with 58,596 records, equally split between benign and malicious samples. It is designed for binary classification, distinguishing between benign and malware instances using the "Class" target variable. The model built for this task achieved near-perfect performance, with 100 % precision, 99.99 % accuracy and F1-score, and 99.98 % recall on the test set. The confusion matrix (Fig. 2) confirms the model effectiveness, showing very few errors in classifying the samples, making it highly reliable for malware detection in cybersecurity.

Multi-class classification

The CIC-MalMem-2022 dataset, designed for obfuscated malware testing, includes malware families like Trojan Horse, Spyware, and Ransomware, making it suitable for multiclass classification. Two types of multiclass classification are discussed: one for identifying the malware families (Trojan, Spyware, Ransomware) and another for classifying individual attacks within each family.

Detection of malware families. The CIC-MalMem-2022 dataset includes a target variable called "Category" which contains values for benign samples and three malware families: trojan, ransomware, and spyware. The malware names were standardized by renaming all entries starting with "ransomware-" "spyware-", and "Trojan-" to

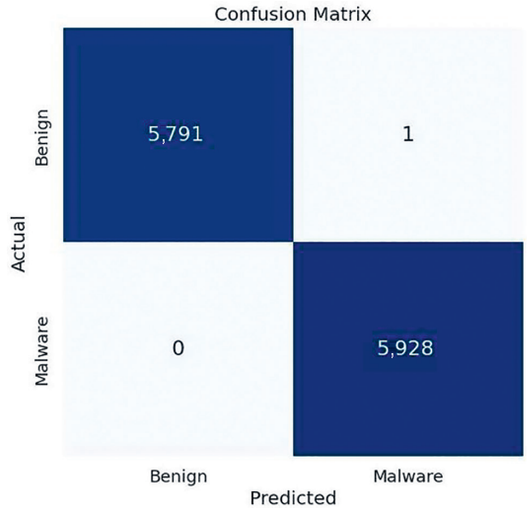

Fig. 2. Confusion matrix for binary classification

Table 1. Number of Instances in case of malware families

Type	Number of instances
Benign	29,298
Spyware	10,020
Ransomware	9,791
Trojan	9,487

"ransomware", "spyware", and "trojan", respectively, resulting in four classes: one for benign samples and three for the malware families.

Table 1 summarizes the occurrence of different types of instances in the "Category" target. The dataset contains 29,298 benign instances, followed by 10,020 spyware instances, 9,791 ransomware instances, and 9,487 trojan instances.

The DNN-ANOVA model achieved an accuracy of 85.83 %, with precision, recall, and F1-score all at 86 %, indicating strong performance and a good balance between precision and recall. The confusion matrix shown in Fig. 3 provides a visual representation of the model accuracy for multiclass classification, specifically in detecting malware families, such as trojan, ransomware, and spyware.

Detection of individual malwares. The "Category" target in the CIC-MalMem-2022 dataset includes 58,596 records, evenly split between benign and malicious samples. The malicious records are categorized into three main families — trojan, ransomware, and spyware – each further divided into subfamilies with five types of attacks. Overall, the "Category" target can take 16 values: one for benign samples and 15 for individual malware types. Fig. 4 illustrates the distribution of these malware families.

Table 2 details the number of instances for various types of malware and their subtypes in the "Category" target. The most common type is benign with 29,298 instances. Among the malicious types, "spyware-transponder" is the most frequent with 2,410 instances, followed by "spyware-gator" with 2,200 instances. For ransomware, the most common subtype is "shade" with 2,128 instances, and for trojans, "refroso" is the most prevalent with 2,000 instances.

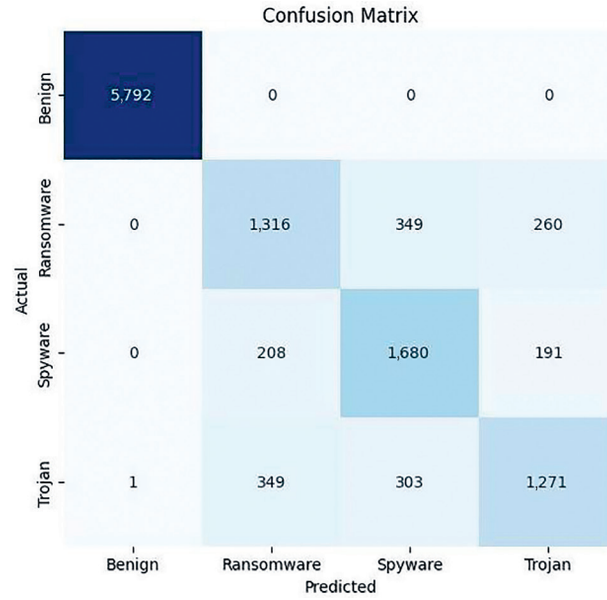

Fig. 3. Confusion matrix for malware families detection

Simulation results show that DNN-ANOVA model achieves accuracy of 73.98 %, and precision, recall, and F1-score, all of 74 %. Fig. 5 shows the confusion matrix for individual attacks detection of the defined dataset.

Comparative study

This section provides a comparative study of our work with previous research. Table 3 compares Mezina and Burget [18], RobustCBL [17], CompactCBL [17] and DNN-ANOVA (proposed work) with respect to accuracy, precision, recall and F1-score. The DNN-ANOVA model has the highest accuracy, precision, recall, and F1-score

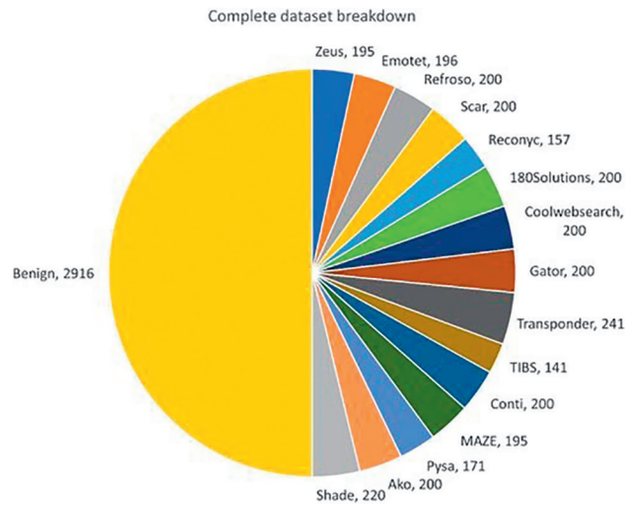

Fig. 4. Complete dataset breakdown[1]

[1] T. Carrier, P. Victor, A. Tekeoglu, and A. H. Lashkari, "CIC-MalMem-2022: Malware Memory Analysis Dataset," Canadian Institute for Cybersecurity, University of New Brunswick, 2022. [Online]. Available: https://www.unb.ca/cic/datasets/malmem-2022.html (accessed: 02.08.2024).

Table 2. Number of Instances in case of individual malwares

Type	Number of instances
Benign	29,298
Spyware-Transponder	2,410
Spyware-Gator	2,200
Spyware-180solutions	2,000
Spyware-CWS	2,000
Spyware-TIBS	1,410
Ransomware-Shade	2,128
Ransomware-Ako	2,000
Ransomware-Conti	1,988
Ransomware-Maze	1,958
Ransomware-Pysa	1,717
Trojan-Refroso	2,000
Trojan-Scar	2,000
Trojan-Emotet	1,967
Trojan-Zeus	1,950
Trojan-Reconyc	1,570

with 85.83 %, 86 %, 86 %, and 86 %, respectively. RobustCBL [17] and CompactCBL [17] have almost the same performance measures, whereas Mezina and Burget has the lowest measures.

The performance metrics of the different models for different classes of malware (malware families) are shown in Table 4. All models have perfect accuracy, precision and recall for the benign class. For the ransomware class, DNN-ANOVA has the highest accuracy, precision, and recall, whereas Mezina and Burget [18] has the lowest accuracy, precision, and recall. For the spyware class, DNN-ANOVA has the highest accuracy, precision, and recall, while RobustCBL [17] has the lowest accuracy, precision, and recall. Regarding the trojan category, DNN-ANOVA has the highest precision and shares with RobustCBL [17] the best F1-score, though it falls slightly behind in recall. Conversely, Mezina and Burget [18] record the minimal scores in accuracy, precision, and recall for this class.

Table 5 shows the performance metrics for different models in the case of individual malwares detection. The comparison is made in terms of accuracy, precision, recall, and F1-score. DNN-ANOVA has the highest accuracy, precision, recall, and F1-score at 73.98 %, 74 %, 74 %, and 74 %, respectively. Robust-CBL [17] and CompactCBL [17] have lower performance metrics, with RobustCBL [17] having slightly higher precision, recall, and F1-score than CompactCBL [17].

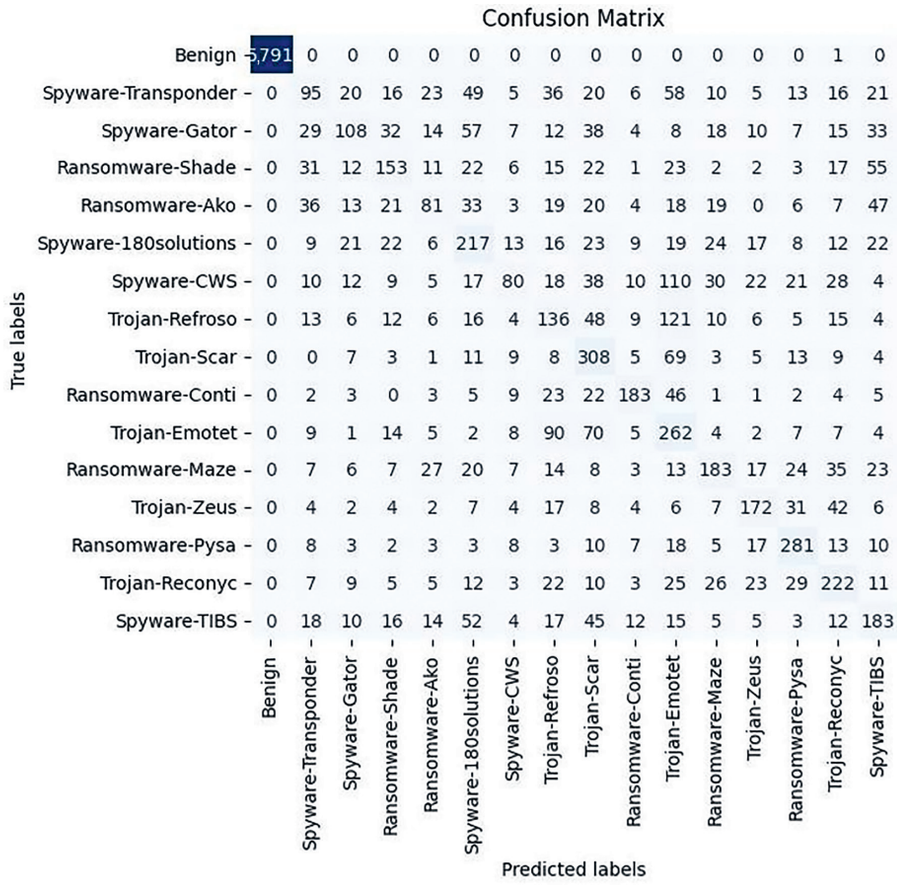

Fig. 5. Confusion matrix for individual malware detection

Table 3. Attack family detection performance comparison, %

	Accuracy	Precision	Recall	F1-score
Mezina & Burget [18]	83.53	75.79	75.18	75.13
RobustCBL [17]	84.56	85.00	85.00	84.00
CompactCBL [17]	84.22	84.00	84.00	84.00
DNN-ANOVA	85.83	86.00	86.00	86.00

Table 4. Attack families detection performance for each class, %

Class	Mezina & Burget [18]			RobustCBL [17]			DNN-ANOVA		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Benign	100	100	100	100	100	100	100	100	100
Ransom-ware	62	66	64	67	62	64	70	68	69
Spyware	67	76	71	69	77	73	72	81	76
Trojan	73	57	64	71	67	70	74	66	70

Table 5. Attack family detection performance comparison, %

	Accuracy	Precision	Recall	F1-score
RobustCBL [17]	72.60	73.00	73.00	72.00
CompactCBL [17]	71.42	72.00	71.00	71.00
DNN-ANOVA	73.98	74.00	74.00	74.00

Conclusion

In this paper, a Deep Neural Network was proposed to detect and classify malwares. A pre-processing task was necessary to first select the most relevant features using ANalysis Of Variance (ANOVA) F-test feature selection strategy where the *k* most relevant features are those whose score values exceeds a certain threshold evaluated as the ratio between the sum of all features scores and the total number of features. Once features are selected, target variable was encoded. Malware analysis is done either by binary classification to determine if there is malware, or by multiclass classification to detect not only the malware but also its type. The proposed model is trained and assessed using CIC-MalMem-2022 dataset. To demonstrate the efficiency of the suggested model, its performance has been studied and compared with other previous research. Results reveal that the suggested model outperforms previous works with 85.83% accuracy, 86 % precision, 86 % recall, and 86 % F1-score when dealing with malware families, and outperforms other works with 73.98 % accuracy, 74 % precision, 74 % recall, and 74 % F1-score in the case of individual attacks detection. Our future research will focus on refining the model by applying graph neural network and deep reinforcement learning.

References

1. Kramer S., Bradfield J.C. A general definition of malware. *Journal in Computer Virology*, 2010, vol. 6, no. 2, pp. 105–114. https://doi.org/10.1007/s11416-009-0137-1
2. Li C., Gaudiot J.L. Detecting malicious attacks exploiting hardware vulnerabilities using performance counters. *Proc. of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). V. 1*, 2019, pp. 588–597. https://doi.org/10.1109/compsac.2019.00090
3. Sinanovic H., Mrdovic S. Analysis of Mirai malicious software. *Proc. of the 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2017, pp. 1–5. https://doi.org/10.23919/softcom.2017.8115504
4. Singh R., Kumar H., Singla R.K., Ketti R.R. Internet attacks and intrusion detection system: A review of the literature. *Online Information Review*, 2017, vol. 41, no. 2, pp. 171–184. https://doi.org/10.1108/oir-12-2015-0394
5. Yadav B., Tokekar S. Deep learning in malware identification and classification. *Malware Analysis Using Artificial Intelligence and Deep Learning*. Springer, Cham, 2021, pp. 163–205. https://doi.org/10.1007/978-3-030-62582-5_6

Литература

1. Kramer S., Bradfield J.C. A general definition of malware // Journal in Computer Virology. 2010. V. 6. N 2. P. 105–114. https://doi.org/10.1007/s11416-009-0137-1
2. Li C., Gaudiot J.L. Detecting malicious attacks exploiting hardware vulnerabilities using performance counters // Proc. of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). V. 1. 2019. P. 588–597. https://doi.org/10.1109/compsac.2019.00090
3. Sinanovic H., Mrdovic S. Analysis of Mirai malicious software // Proc. of the 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM). 2017. P. 1–5. https://doi.org/10.23919/softcom.2017.8115504
4. Singh R., Kumar H., Singla R.K., Ketti R.R. Internet attacks and intrusion detection system: A review of the literature // Online Information Review. 2017. V. 41. N 2. P. 171–184. https://doi.org/10.1108/oir-12-2015-0394
5. Yadav B., Tokekar S. Deep learning in malware identification and classification // Malware Analysis Using Artificial Intelligence and Deep Learning. Springer, Cham, 2021. P. 163–205. https://doi.org/10.1007/978-3-030-62582-5_6

6. Kertysova K., Frinking E., van den Dool K., Maricic A., Bhattacharyya K. *Cybersecurity: Ensuring awareness and resilience of the private sector across Europe in face of mounting cyber risks-Study*. Bruxelles, Belgium, European Economic and Social Committee, 2018.
7. Gopinath M., Sethuraman S.C. A comprehensive survey on deep learning based malware detection techniques. *Computer Science Review*, 2023, vol. 47, pp. 100529. https://doi.org/10.1016/j.cosrev.2022.100529
8. Faruk M.J.H., Shahriar H., Valero M., Barsha F.L., Sobhan S., Khan M.A., Whitman M., Cuzzocrea A., Lo D., Rahman A., Wu F. Malware detection and prevention using artificial intelligence techniques. *Proc. of the IEEE International Conference on Big Data (Big Data)*, 2021, pp. 5369–5377. https://doi.org/10.1109/bigdata52589.2021.9671434
9. Vigna G. How AI will help in the fight against malware. *Retrieved from TechBeacon*, 2020.
10. Schmitt M. Securing the Digital World: Protecting smart infrastructures and digital industries with Artificial Intelligence (AI)-enabled malware and intrusion detection. *Journal of Industrial Information Integration*, 2023, vol. 36, pp. 100520. https://doi.org/10.1016/j.jii.2023.100520
11. Aljabri M., Alhaidari F., Albuainain A., Alrashidi S., Alansari J., Alqahtani W., Alshaya J. Ransomware detection based on machine learning using memory features. *Egyptian Informatics Journal*, 2024, vol. 25, pp. 100445. https://doi.org/10.1016/j.eij.2024.100445
12. Ababneh M., Aljarrah A. Cybersecurity: Malware multi-attack detector on android-based devices using deep learning methods. *Journal of Theoretical and Applied Information Technology*, 2024, vol. 102, no. 1, pp. 144–166.
13. Majid A.A.M., Alshaibi A.J., Kostyuchenko E., Shelupanov A. A review of artificial intelligence based malware detection using deep learning. *Materials Today: Proceedings*, 2023, vol. 80, part 3, pp. 2678–2683. https://doi.org/10.1016/j.matpr.2021.07.012
14. Riaz S., Latif S., Usman S.M., Ullah S.S., Algarni A.D., Yasin A., Anwar A., Elmannai H., Hussain S. Malware Detection in Internet of Things (IoT) devices using deep learning. *Sensors*, 2022, vol. 22, no. 23, pp. 9305. https://doi.org/10.3390/s22239305
15. Xing X., Jin X., Elahi H., Jiang H., Wang G. A malware detection approach using autoencoder in deep learning. *IEEE Access*, 2022, vol. 10, pp. 25696–25706. https://doi.org/10.1109/access.2022.3155695
16. Ucci D., Aniello L., Baldoni R. Survey of machine learning techniques for malware analysis. *Computers & Security*, 2019, vol. 81, pp. 123–147. https://doi.org/10.1016/j.cose.2018.11.001
17. Shafin S.S., Karmakar G., Mareels I. Obfuscated memory malware detection in resource-constrained IoT devices for smart city applications. *Sensors*, 2023, vol. 23, no. 11, pp. 5348. https://doi.org/10.3390/s23115348
18. Mezina A., Burget R. Obfuscated malware detection using dilated convolutional network. *Proc. of the 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2022, pp. 110–115. https://doi.org/10.1109/icumt57764.2022.9943443
19. Brownlee J. How to perform feature selection with numerical input data. *Machine Learning Mastery*, 2020.
20. Brownlee J. How to choose a feature selection method for machine learning. *Machine Learning Mastery*, 2019.
21. Cai J., Luo J., Wang S., Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing*, 2018, vol. 300, pp. 70–79. https://doi.org/10.1016/j.neucom.2017.11.077
22. Payton A.M. A review of spyware campaigns and strategies to combat them. *Proc. of the 3rd Annual Conference on Information Security Curriculum Development*, 2006, pp. 136–141. https://doi.org/10.1145/1231047.1231077
23. Carrier T., Victor P., Tekeoglu A., Lashkari A. Detecting obfuscated malware using memory feature engineering. *Proc. of the 8th International Conference on Information Systems Security and Privacy ICISSP. V. 1*, 2022, pp. 177–188. https://doi.org/10.5220/0010908200003120
24. Mallikarajunan K.N., Preethi S.R., Selvalakshmi S., Nithish N. Detection of spyware in software using virtual environment. *Proc. of the 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 1138–1142. https://doi.org/10.1109/icoei.2019.8862547

6. Kertysova K., Frinking E., van den Dool K., Maricic A., Bhattacharyya K. Cybersecurity: Ensuring awareness and resilience of the private sector across Europe in face of mounting cyber risks-Study. Bruxelles, Belgium: European Economic and Social Committee, 2018.
7. Gopinath M., Sethuraman S.C. A comprehensive survey on deep learning based malware detection techniques // Computer Science Review. 2023. V. 47. P. 100529. https://doi.org/10.1016/j.cosrev.2022.100529
8. Faruk M.J.H., Shahriar H., Valero M., Barsha F.L., Sobhan S., Khan M.A., Whitman M., Cuzzocrea A., Lo D., Rahman A., Wu F. Malware detection and prevention using artificial intelligence techniques // Proc. of the IEEE International Conference on Big Data (Big Data). 2021. P. 5369–5377. https://doi.org/10.1109/bigdata52589.2021.9671434
9. Vigna G. How AI will help in the fight against malware // Retrieved from TechBeacon. 2020.
10. Schmitt M. Securing the Digital World: Protecting smart infrastructures and digital industries with Artificial Intelligence (AI)-enabled malware and intrusion detection // Journal of Industrial Information Integration. 2023. V. 36. P. 100520. https://doi.org/10.1016/j.jii.2023.100520
11. Aljabri M., Alhaidari F., Albuainain A., Alrashidi S., Alansari J., Alqahtani W., Alshaya J. Ransomware detection based on machine learning using memory features // Egyptian Informatics Journal. 2024. V. 25. P. 100445. https://doi.org/10.1016/j.eij.2024.100445
12. Ababneh M., Aljarrah A. Cybersecurity: Malware multi-attack detector on android-based devices using deep learning methods // Journal of Theoretical and Applied Information Technology. 2024. V. 102. N 1. P. 144–166.
13. Majid A.A.M., Alshaibi A.J., Kostyuchenko E., Shelupanov A. A review of artificial intelligence based malware detection using deep learning // Materials Today: Proceedings. 2023. V. 80. Part 3. P. 2678–2683. https://doi.org/10.1016/j.matpr.2021.07.012
14. Riaz S., Latif S., Usman S.M., Ullah S.S., Algarni A.D., Yasin A., Anwar A., Elmannai H., Hussain S. Malware Detection in Internet of Things (IoT) devices using deep learning // Sensors. 2022. V. 22. N 23. P. 9305. https://doi.org/10.3390/s22239305
15. Xing X., Jin X., Elahi H., Jiang H., Wang G. A malware detection approach using autoencoder in deep learning // IEEE Access. 2022. V. 10. P. 25696–25706. https://doi.org/10.1109/access.2022.3155695
16. Ucci D., Aniello L., Baldoni R. Survey of machine learning techniques for malware analysis // Computers & Security. 2019. V. 81. P. 123–147. https://doi.org/10.1016/j.cose.2018.11.001
17. Shafin S.S., Karmakar G., Mareels I. Obfuscated memory malware detection in resource-constrained IoT devices for smart city applications // Sensors. 2023. V. 23. N 11. P. 5348. https://doi.org/10.3390/s23115348
18. Mezina A., Burget R. Obfuscated malware detection using dilated convolutional network // Proc. of the 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). 2022. P. 110–115. https://doi.org/10.1109/icumt57764.2022.9943443
19. Brownlee J. How to perform feature selection with numerical input data // Machine Learning Mastery. 2020.
20. Brownlee J. How to choose a feature selection method for machine learning // Machine Learning Mastery. 2019.
21. Cai J., Luo J., Wang S., Yang S. Feature selection in machine learning: A new perspective // Neurocomputing. 2018. V. 300. P. 70–79. https://doi.org/10.1016/j.neucom.2017.11.077
22. Payton A.M. A review of spyware campaigns and strategies to combat them // Proc. of the 3rd Annual Conference on Information Security Curriculum Development. 2006. P. 136–141. https://doi.org/10.1145/1231047.1231077
23. Carrier T., Victor P., Tekeoglu A., Lashkari A. Detecting obfuscated malware using memory feature engineering // Proc. of the 8th International Conference on Information Systems Security and Privacy ICISSP. V. 1. 2022. P. 177–188. https://doi.org/10.5220/0010908200003120
24. Mallikarajunan K.N., Preethi S.R., Selvalakshmi S., Nithish N. Detection of spyware in software using virtual environment // Proc. of the 3rd International Conference on Trends in Electronics and Informatics (ICOEI). 2019. P. 1138–1142. https://doi.org/10.1109/icoei.2019.8862547
25. Jonasson D., Sigholm J. What is Spyware?. TDDC03 Projects, Department of Computer and Information Science. Sewden: Linkopings University, 2005.

25. Jonasson D., Sigholm J. *What is Spyware?. TDDC03 Projects, Department of Computer and Information Science*. Sewden, Linkopings University, 2005.
26. Pelchen-Matthews A., Raposo G., Marsh M. Endosomes, exosomes and Trojan viruses. *Trends in Microbiology*, 2004, vol. 12, no. 7, pp. 310–316. https://doi.org/10.1016/j.tim.2004.05.004
27. Liu Y., Mondal A., Chakraborty A., Zuzak M., Jacobsen N., Xing D., Srivastava A. A survey on neural trojans. *Proc. of the 21st International Symposium on Quality Electronic Design (ISQED)*, 2020, pp. 33–39. https://doi.org/10.1109/isqed48828.2020.9137011
28. Brewer R. Ransomware attacks: detection, prevention and cure. *Network Security*, 2016, vol. 2016, no. 9, pp. 5–9. https://doi.org/10.1016/s1353-4858(16)30086-1
29. Tuttle H. Ransomware attackers turn to double extortion. *Risk Management*, 2021, vol. 68, no. 2, pp. 8–9.
30. Nershi K., Grossman S. Assessing the Political Motivations Behind Ransomware Attacks. *SSRN Electronic Journal*, 2023. https://doi.org/10.2139/ssrn.4507111
31. Casas P., Blancas J., Villanueva A. Ransomware Report 2023: targets, motives, and trends. *Outpost24*. 07 Feb. 2023. Available: https://outpost24.com/blog/ransomware-report-2023- targets-motives-and-trends/ (accessed: 01.08.2024).
32. Sawyer S.F. Analysis of variance: the fundamental concepts. *Journal of Manual & Manipulative Therapy*, 2009, vol. 17, no. 2, pp. 27E–38E. https://doi.org/10.1179/jmt.2009.17.2.27e

26. Pelchen-Matthews A., Raposo G., Marsh M. Endosomes, exosomes and Trojan viruses // Trends in Microbiology. 2004. V. 12. N 7. P. 310–316. https://doi.org/10.1016/j.tim.2004.05.004
27. Liu Y., Mondal A., Chakraborty A., Zuzak M., Jacobsen N., Xing D., Srivastava A. A survey on neural trojans // Proc. of the 21st International Symposium on Quality Electronic Design (ISQED). 2020. P. 33–39. https://doi.org/10.1109/isqed48828.2020.9137011
28. Brewer R. Ransomware attacks: detection, prevention and cure // Network Security. 2016. V. 2016. N 9. P. 5–9. https://doi.org/10.1016/s1353-4858(16)30086-1
29. Tuttle H. Ransomware attackers turn to double extortion // Risk Management. 2021. V. 68. N 2. P. 8–9.
30. Nershi K., Grossman S. Assessing the Political Motivations Behind Ransomware Attacks // SSRN Electronic Journal. 2023. https://doi.org/10.2139/ssrn.4507111
31. Casas P., Blancas J., Villanueva A. Ransomware Report 2023: targets, motives, and trends // Outpost24. 07 Feb. 2023 [Электронный ресурс]. URL: https://outpost24.com/blog/ransomware-report-2023-targets-motives-and-trends/ (дата обращения: 01.08.2024).
32. Sawyer S.F. Analysis of variance: the fundamental concepts // Journal of Manual & Manipulative Therapy. 2009. V. 17. N 2. P. 27E–38E. https://doi.org/10.1179/jmt.2009.17.2.27e

Authors

Mourad Hadjila — D.Sc., Lecturer-Researcher, University of Tlemcen, Tlemcen, 13000, Algeria, SC 56440246000, https://orcid.org/0000-0002-6554-3925, mhadjila.2009@gmail.com
Mohammed Merzoug — D.Sc., Lecturer-Researcher, University of Tlemcen, University of Tlemcen, Tlemcen, 13000, Algeria, SC 55309175500, https://orcid.org/0009-0002-9117-047X, mohammed.merzoug@univ-tlemcen.dz
Wafaa Ferhi — PhD Student, University of Tlemcen, Tlemcen, 13000, Algeria, https://orcid.org/0009-0005-7574-8368, wafaa.ferhi@univ-tlemcen.dz
Djillali Moussaoui — D.Sc., Lecturer-Researcher, University of Tlemcen, Tlemcen, 13000, Algeria, SC 56360232600, https://orcid.org/0000-0003-3478-263X, djilali.moussaoui@univ-tlemcen.dz
Al Baraa Bouidaine — PhD Student, University of Tlemcen, Tlemcen, 13000, Algeria, https://orcid.org/0009-0005-2204-9117, albaraa.bouidaine@univ-tlemcen.dz
Mohammed Hicham Hachemi — D.Sc., Lecturer-Researcher, University of Oran, Oran, 31000, Algeria, SC 57196009731, https://orcid.org/0000-0003-3967-6609, hicham.hachemi@univ-usto.dz

Авторы

Хаджила Мурад — доктор наук, преподаватель-исследователь, Университет Тлемсена, Тлемсен, 13000, Алжир, SC 56440246000, https://orcid.org/0000-0002-6554-3925, mhadjila.2009@gmail.com
Мерзуг Мохаммед — доктор наук, преподаватель-исследователь, Университет Тлемсена, Тлемсен, 13000, Алжир, SC 55309175500, https://orcid.org/0009-0002-9117-047X, mohammed.merzoug@univ-tlemcen.dz
Ферхи Вафа — аспирант, Университет Тлемсена, Тлемсен, 13000, Алжир, https://orcid.org/0009-0005-7574-8368, wafaa.ferhi@univ-tlemcen.dz
Муссауи Джилали — доктор наук, преподаватель-исследователь, Университет Тлемсена, Тлемсен, 13000, Алжир, SC 56360232600, https://orcid.org/0000-0003-3478-263X, djilali.moussaoui@univ-tlemcen.dz
Буйден Аль Бараа — аспирант, Университет Тлемсена, Тлемсен, 13000, Алжир, https://orcid.org/0009-0005-2204-9117, albaraa.bouidaine@univ-tlemcen.dz
Хашеми Мохаммед Хишам — доктор наук, преподаватель-исследователь, Университет Орана, Оран, 31000, Алжир, SC 57196009731, https://orcid.org/0000-0003-3967-6609, hicham.hachemi@univ-usto.dz

Received 27.10.2023
Approved after reviewing 13.08.2024
Accepted 25.09.2024

Статья поступила в редакцию 27.10.2023
Одобрена после рецензирования 13.08.2024
Принята к печати 25.09.2024