International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EVALUATING AI-POWERED CYBER DEFENSE MECHANISMS AGAINST ZERO-DAY ATTACKS

[1*]V.S.Balaji, [2] Radhakrishnan Arikrishna Perumal, [3]Dr Dinesha H A, [4]Dr.Vijayakumar Sangamesvarappa, [5]Mrs. Anupa Amol Pawar, [6]Swathi Nelavalli

[1*]SASTRA Deemed University, Thanjavur, Tamil Nadu, India, Orcid Id: 0000-0002-4599-2854, Email Id: biobala_mtech@eie.sastra.edu
[2]Principal, Software Architect Anchor, General Insurance Agency, Orcid Id: https://orcid.org/0009-0007-4919-911X , Email Id: krishtna.ar@gmail.com
[3]Professor, Visvesvaraya Technological University, Email Id: sridini@gmail.com
[4]Computer Applications, Aditya University, Surampalem ,533437, Orcid Id: 0009-0009-3021-445X, Email Id: vijayakumars@adityauniversity.in
[5]Assiatant Professor, Sri. Balaji University, Pune, Email Id: haridas.annu009@gmail.com
[6] Assistant Professor, RVR&JC College of Engineering, Email Id: swathi.nelavalli@gmail.com

Abstract

The ability to identify known or unknown cyber threats is an urgent need that can only be addressed with adaptive cyber defense strategies. This study evaluates the use of AI-powered defense mechanisms based on the use of supervised machine learning and anomaly detection models within the UGRansome dataset, a large dataset of behavioral, transactional and network-level indicators associated with ransomware activity. The methodology combines Random Forest, Support Vector Machine and Naive Bayes classifiers with a zero-day simulation where ransomware families are withheld during the training process, and a model of Isolation Forest with data on only known ransomware families to detect deviations in the behavior of unknown ransomware attacks. Results show that Random Forest has a near perfect accuracy on known families (99.31%) and has high generalization under the zero-day environment (96.22%), and SVM has also a competitive performance, but Naive Bayes has serious limitations. The Isolation Forest detector further puts a separation between zero-day families with a much higher anomaly rate (+23.37%), which makes it a useful complementary behavioral layer. These findings underscore the power of hybrid architectures of artificial intelligence defence approaches that make use of both supervised classification and anomaly detection in order to build more resilience to evolving threats. The study ends with stressing the importance of conducting real-time evaluation, deep learning models, and adversarial robustness studies to further enhance AI-driven cyber security systems.

Keywords: AI-powered cyber defense, Zero-day ransomware detection, Ensemble machine learning, Anomaly detection, UGRansome dataset.

1. Introduction

The speed with which ransomware attacks have increased has prompted them to top the list of the most disruptive and financially damaging cyber threats in the world. These attacks are increasingly automated, obfuscated and fast variant generation, which has led to traditional signature defence not being adequate for modern threat landscapes. As organizations across the healthcare industry, finance, manufacturing and critical infrastructure have been digitizing their operations, adversaries have been building out their arsenals with sophisticated malware families, which can evade static detection systems. Recent researches highlight that ransomware is now evolving with shorter cycle time and tend to use polymorphism and behaviour masking techniques to circumvent the standard intrusion

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

detection techniques, which generate an urgent need for dynamic, adaptable and intelligent defence systems[1]. In response, artificial intelligence (AI) and machine learning (ML) have become promising paradigms able to analyze high dimensional behavioral, transactional and network telemetry to identify attack patterns not captured by static methods.

AI-powered cyber defense architectures increasingly utilize models that have learned the behavioral signatures of cyber or malware through supervision and have adapted themselves to new attack modalities or attack methods. Supervised ML models are being trained on labeled malware datasets which have been used to train models with exceptional performance to detect well known ransomware families based on features such as process behavior, network communication pattern and transaction anomalies. At the same time, unsupervised anomaly detection approaches, such as autoencoders and clustering algorithms have shown their potential in detecting novel or unseen attack vectors based on the deviation of the normal system activity. However, more recent evaluations identify that the quality of the dataset, variability of the features, and timeliness are still ongoing challenges in achieving strong and powerful detection capabilities that are not limited to known attacks [2]. Consequently, creating AI systems with a balance between their predictive power, generalization of behavior, and their ability to respond in real time is an important research challenge.

Despite the growing power of AI-driven cybersecurity solutions, the performance of these solutions against zero-day ransomware attacks that are completely novel behaviors not included in the training data is still inconsistent. A major drawback of many supervised ML models is that they are reliant on historical labeled data which limits how accurately they can detect behavioral patterns that belong to ransomware families that are new to them. Research has shown that when whole families of ransomware are kept out of training, things can get much dumber - models will not be able to generalize across unseen behavioural clusters [3]. Furthermore, while the aforementioned techniques for detecting anomalies can point out suspicious deviations, they often have a high false-positive rate, low interpretability or unstable performance across a heterogeneous threat environment. These limitations suggest that there is a methodological gap in the current state of cyber defense research: There are no integrated frameworks that combine supervised classification and anomaly-based detection to offer layering resilience policies against novel, rapidly evolving ransomware threats.

This study examines the efficacy of AI-powered cyber defense mechanisms with the use of UGRansome dataset, including known and zero-day ransomware detection. The scope of the research includes three supervised models (Random Forest, Support Vector Machine SVM and Naive Bayes) that are used to classify behavioral and transactional indicators linked to ransomware activities. To approximate realistic zero-day conditions, a family hold-out experiment is used where the 5 ransomware families with the greatest frequency are taken as training data and the remaining ransomware families are taken as unseen data. In parallel, an Isolation Forest anomaly detector is trained on only the data of the known family in order to test the detector's capability to detect behavioral deviations in samples that are not yet seen. Performance is measured in terms of a combination of confusion matrices, accuracy factors and anomaly score distributions. However, the synthetic components in UGRansome dataset may not accurately reflect the variability of real-world ransomware attacks, and the fact that the study did not use more computationally intensive deep learning models such as LSTMs, CNNs, and reinforcement learning agents that have shown strong performance in the related fields [4].

The contribution of this study to the cybersecurity literature is the development of a comprehensive evaluation framework that incorporates both supervised and unsupervised ransomware attack detection mechanisms for known and zero-day ransomware attacks. First of all, the work shows the robustness of ensemble learning methodologies, especially Random Forest, in achieving better classification accuracy for different ransomware families, which provides evidence that ensemble-based models are still very powerful in high dimensional malware analysis [5]. Second, the work confirms the value of detecting anomalies in order to identify new ransomware behaviors and the

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

emerging consensus that hybrid threat detection architectures (which combine prediction and anomaly-based awareness) offer the best defense capabilities. Third, by using a controlled family hold-out design, the research is consistent with suggestions for more realistic zero-day evaluation frameworks, something that is a longstanding methodological shortcoming in cybersecurity experimentation [6]. Finally, the findings have some practical implications for intrusion detection system (IDS) designers and security practitioners who are interested in developing robust, scalable and adaptive defensive models that can respond to fast-evolving ransomware ecosystems.

By comparing supervised classifiers with a specialized anomaly detector in known and zero day scenarios, this paper extends the recent research on the necessity of multi-layered cyber defense systems. Prior evaluations of the frameworks for ransomware detection have consistently demonstrated the failure of static or single-layer ransomware detection frameworks to capture the complexity of modern ransomware execution patterns, and thus the importance of having diverse analytical layers to capture financial, temporal, and behavioral indicators [7]. Moreover, the research makes use of structured behavioral and transactional features that complement results in recent AI based ransomware detection reviews, which emphasize on granular feature engineering as well as dynamic behavioral analysis to distinguish subtle differences between ransomware families [8]. The methodological integration as used in this work also reflects the calls that have been made within the literature for the adoption of hybrid models combining supervised and anomaly-based detection processes to increase robustness and resilience in high-uncertainty cyber environments [9]. Additionally, the zero-day simulation implemented in this study offers empirical evidence to the feasibility of ensemble learning and the anomaly detection technique in dealing with previously unseen threats to contribute to the discourse on adaptive and scalable cyber defense strategies [10].

Based on the gap identified in the existing literature on cybersecurity and the evolution of the threat posed by ransomware, the present study aims to do the following:

1. To assess the predictive performance of supervised machine learning algorithms, namely Random Forest, SVM, and Naive Bayes in detecting ransomware activities by means of structured network and transactional data.
2. To examine the generalization power of supervised models under simulated zero-day conditions by not using full families of ransomware models when training the models.
3. To investigate the effectiveness of anomaly-based methods (Isolation Forest) for identifying behavioral deviations of ransomware families that have never been seen before.

2. Literature Review

The explosive growth of ransomware and zero-day attacks has been a catalyst for accelerated development of machine learning (ML) and artificial intelligence (AI) cyber defense systems, where the recent literature shows both methodology progress and remaining limitations. Research in this space tends to focus more on the significance of feature engineering, hybrid learning models, dataset quality and generalization capabilities - factors that jointly determine the power of modern threat detection tools. This review is a synthesis of important contributions from recent works, in order to place the current investigation into perspective.

Early research highlights the importance of strong feature engineering and hybrid architectures to deal with the complexity of the behavior of ransomware. Zhang et al. [11] showed that by using feature selection techniques like Recursive Feature Elimination (RFE) and SHAP-based interpretablity, we can significantly increase the accuracy of the models and also make them more interpretable. Their results show that economic and behavioral indicators, represented by transaction patterns and attributes at the system level, offer good predictive value, which supports the necessity of multi-dimensional feature representation. Complementing this view, Por et al. [12] listed some persistent challenges concerning data availability, algorithmic complexity, and real-time applicability

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

in the case of zero-day detection, and identified the need for more flexible and context-aware AI systems.

Ransomware research has also been aided by innovations in dynamic analysis and behavioural modelling. Ayub et al. [13] proposed a static-informed dynamic analysis method (RWArmor) that achieved a high detection accuracy at the early detection stage by utilizing behavior logs in the first 30-120 seconds of execution. Their results indicate the importance of time-sensitive behavioral signals, which can permit fast response before irreversible file encryption can take place. Extending the conversation to zero day detection, Alkasassbeh et al. [14] showed off the power of deep reinforcement learning (specifically Deep Q-Networks) in effectively classifying attack patterns, both known and unseen, and exhibiting strong generalization capabilities in zero-day evaluation setups.

In parallel, domain-specific approaches to detection have emerged in order to protect critical infrastructures such as IoT-based healthcare systems. Iqbal et al. [15] proposed a multimodal ransomware detector (RThreatDroid) using image, text and code analysis (high accuracy using a hybrid static-dynamic architecture). This work is the illustration of how threat detection pipelines that are tailored for specific domains can be effective in bettering threat detection in environments with unique operation constraints. Meanwhile, the ransomware and zero-day research themes of dataset quality are currently booming. Nkongolo et al. [16] proposed UGRansome, which is a dataset that seeks to represent realistic behavioral and cyclostationary behaviors of ransomware. Their study raised problems such as class imbalance, and feature variability and inherent difficulties in creating datasets that represent ransomware diversity in the real world.

As cyber threats became more sophisticated new learning frameworks have been explored to comply with distributed data sources and privacy requirements. Verma et al. [17] proposed a federated learning-based architecture for zero day attacks detection in 5G-enabled IIoT system by leveraging dual autoencoders and one-class SVM in order to preserve the adaptability of the model and sovereignty of data. Their results point to the promise of decentralized learning to offer high performance without compromising data privacy. Ensemble learning has also been demonstrated to bring good performance benefits. Yan et al. [18] proposed a two-stage ensemble model for the detection of ransomware and classification of ransomware families, which produced an almost 100% accuracy in the detection stage and competitive results in attribution analysis - further underlining the benefits of using multiple classifiers to capture complex signatures of the behavior.

Recent evaluations involving the detection using ML has also brought to light the significance of choice of algorithms and training strategies. Nhlapo and Nkongolo [19], showed ensemble techniques and tree classifiers outperforming SVM and Naive Bayes classifiers for ransomware and zero day attacks detection issues, especially the application of UGRansome. Their results are in line with the general trend that nonlinear, ensemble-based models better model higher order feature interactions that are central to cyber threat identification. Finally, zero-shot learning has taken shape as a promising direction to detect previously unseen attacks in an environment which has a scarce labeled data. Gao et al. [20] demonstrated that generative zero-shot learning with LightGBM-based feature extraction can be used for high accuracy IoT zero-day detection with significant improvement than traditional supervised methods by using latent semantic feature spaces.

Collectively, the literature hints that the best cyber defense strategies incorporate a diversity of learning paradigms, i.e. supervised, unsupervised, reinforcement and zero-shot, to meet the multifaceted nature of evolving cyber threat. Across studies, there are some common themes: a need for better datasets of behavior, a need for better generalization to unseen attacks, the use of hybrid detection pipelines and the development of mechanisms to make pipelines less dependent on handcrafted signatures. These insights form a strong foundation for the study of advanced AI-powered frameworks that can identify known and new ransomware behaviors in dynamic and high-risk environments.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Methodology

3.1 Research Design

This study uses quantitative research, a quasi-experimental research design, and its main objective is to assess the AI-powered cyber defense mechanisms along with detecting Zero-day attacks by using the structured cybersecurity dataset. The design has been motivated by the need to simulate realistic threat environments in which new or unknown malware behaviours may not be represented as part of training the model. The research is based on a particular process that involves understanding the data, preprocessing, and transforming features; building and training machine learning supervised models and unsupervised algorithms for detecting anomalies; and using a controlled zero-day simulation framework, with previously unseen ransomware families or clusters of behavioural patterns held back from training, to simulate new attack patterns. This way, the study can examine both the predictive ability on known threats and the generalization ability of behaviors that have not yet been observed or exhibit anomalies. Through this design, the research provides a replicable and rigorous evaluation of AI-based cyber defense mechanisms in the context of cybersecurity challenges that closely approximate real-world cybersecurity challenges involving zero-day attacks.

3.2 Data Collection Methods

The study uses the UGRansome dataset, a publicly available cybersecurity dataset developed for the analysis of ransomware, network anomalies and zero-day cyber threats [16]. The data set comprises more than 200,000 observations, each of which describes an event in the network using attributes such as timestamps, protocol identifiers, attack flags, behavioural clusters, ransomware families, cryptocurrency payment values, port numbers, threat types and machine-generated attack signatures. Data acquisition was done by retrieving the fully anonymized data set from a well-known public research repository with the assurance that no personal or sensitive data was contained. Preprocessing steps were needed to prepare the dataset to be analyzed, and these included: removing invalid or inconsistent values, removing duplicate values, and cleaning malformed address identifier data. All categorical attributes such as family, protocol type, threat category, and attack flag were encoded to a number as suitable for machine learning algorithms. Numerical features like network byte counts, BTC and USD values, and the port number were normalized to ensure that the feature space is on the same scale. The cleaned and encoded dataset was then split into training and testing datasets in order to create a structured and reliable basis for the modelling and evaluation procedures to follow.

3.3 Population and Sampling

The target population for this study is network level cyberattack behaviours, specifically ransomware events, malicious communication patterns and threat signatures that are commonly encountered in cybersecurity monitoring environments. Although there are synthetic elements in the dataset, it is still representative of life cycles, behavioural indicators of compromise, and anomalous patterns that are characteristic of modern cyber threats. A stratified sampling method was used to ensure that the various ransomware families and threat categories in the dataset are represented in proportion, especially in the face of natural imbalances in the frequency of certain types of attack. This technique maintains the integrity of the dataset statistically when the data are split into training and test data. To simulate conditions of a zero day attack, deliberate holdout sampling was introduced: entire ransomware families, behavioural clusters or machine-generated synthetical signatures were excluded from the training data and introduced only when testing the model. This exclusion strategy mimics situations where a cyber defense mechanism based on AI comes into contact with a novel attack pattern to which it was never exposed previously, and thus allows an empirical judge of zero-day detection capability. By a combination of stratified sampling for assessed known threats with holdout sampling for zero-day emulation the study covers both traditional predictive performance, as well as resilience to emerging attack variants.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.4 Data Analysis Techniques

The analytical phase uses a mix of supervised learning machine models and unsupervised, anomaly anomaly detection methods to test AI-powered opponent defensive mechanisms to ransom ware and zero-day attacks. Three supervised classifiers were applied: Random forest, Support Vector Machine (SVM) and Naive Bayes which were trained with the task of classifying threat categories with behavioural, transactional and network level attributes extracted from the dataset. The dataset was divided with the help of a stratified 70/30 split, in order to have a proportional distribution between the classes and avoid bias caused by the class imbalance.

In order to measure the zero-day attack detection capability, the family hold-out experimental design was adopted. The five most common families of ransomware were considered "known" and used only for model training, and all others were not made available and were considered unseen "zero-day" threats. The Random Forest model was then tested on known and zero day subsets, to allow generalization performance to be measured against previously unobserved malicious behaviours.

For anomaly-based detection, the research used Isolation Forest and was only trained on samples from known ransomware families. This unsupervised model finds the deviation in behavioural patterns by adopting isolation of the anomalies using recursive partitioning. Its goal is to determine if there are behavioural differences between unseen ransomware families, when compared to known ransomware families. Evaluation enrolled analysing anomaly scores and computing anomaly rates on known versus zero day subsets.

Model performance was evaluated using various metrics such as accuracy, precision, recall, F1-score and confusion matrices for supervised models while using anomaly score distributions and anomaly rates for the unsupervised detector. This combination of supervised classification and anomaly detection gives a multi-layered evaluation of defensive performance against well known cyber threat scenarios, as well as zero day cyber threats.

3.5 Ethical Considerations

The dataset used in this study is fully anonymized and publicly available with no personal identifiable information, thus ensuring the compliance of data protection and ethical research standards. All analyses have been performed in a simulated computational environment without deploying, executing, or reverse engineering any malicious code, and no cybersecurity risks were introduced in conducting the analyses or developing the models and in reporting the results.The models developed and the findings reported are intended for defensive and academic purposes only, with an express avoidance of any application that would enable the conduct of offensive cybersecurity activities..

4. Results

The results of evaluation of supervised machine learning models, zero-day attack simulations, and an anomaly-based approach for the ransomware-focused dataset are presented in this section. The results demonstrate both the capability to accurately predict and generalise to behaviours not included in this study, across areas of known and yet-unseen ransomware families.

4.1 Data Presentation

The curated dataset had 148,831 observations representing ransomware - associated behavioral patterns and network activity characteristics, and financial attributes. After preprocessing and encoding, important predictors such as BTC, USD, Flag, Threats and ExpAddress were found to be highly informative in classification. These attributes match both transactional and operational signatures that vary from ransomware family to family.

4.2 Supervised Learning Performance

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Three supervised classifiers were tested: the Naive Bayes, Support Vector Machine (SVM), and Random Forest classifiers. Their overall accuracies are presented in Table 1.

Table 1. Supervised Model Accuracies

Model	Accuracy
Naive Bayes	**0.7196**
SVM	**0.9649**
Random Forest	**0.9931**

Random Forest performed the best (99.31%), followed by SVM (96.49%) with Naive Bayes being way down because its assumptions are not compatible with the complex interactions among features of the dataset.

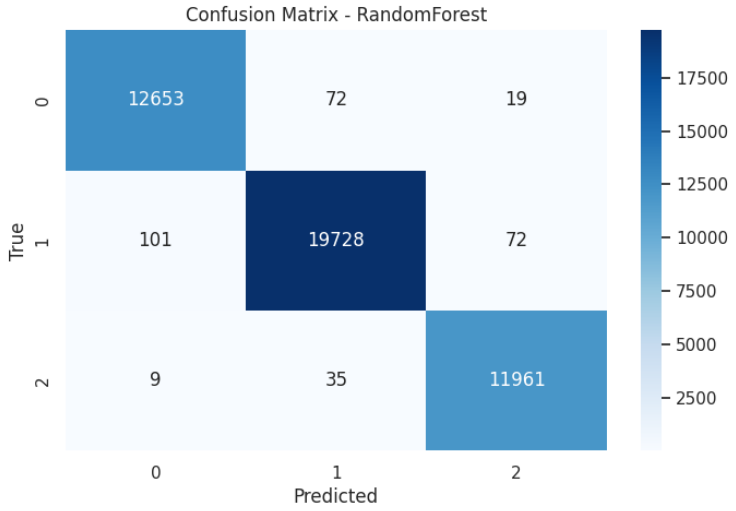
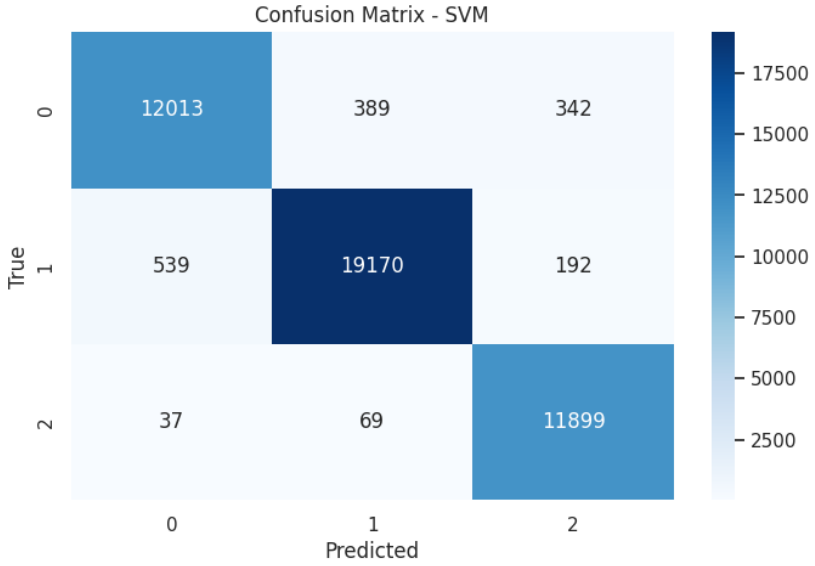


Figure 1. Confusion Matrix — Random Forest (All Families)

The confusion matrix for the Random Forest classifier for all the families is given in Figure 1. The strongly concentrated diagonal values correspond to exceptionally low misclassification, showing the potential of the model to describe multivariate ransomware behavior. The apparent segregation between the three classes in figure 1 establishes Random Forest as the most sound supervised model in this study.



Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 2. Confusion Matrix — SVM

Confusion matrix of SVM classifier is given in Figure 2. While SVM also performs well in terms of accuracy, the moderate rise of off-diagonal entries, especially between classes 0 and 1, would indicate that SVM is slightly more sensitive to the overlap of classes. As shown in figure 2, the model sets strong boundaries that separate behaviors into different parts, but at the same time is not as adaptable in terms of more subtle behavioral differences within families of ransomware.

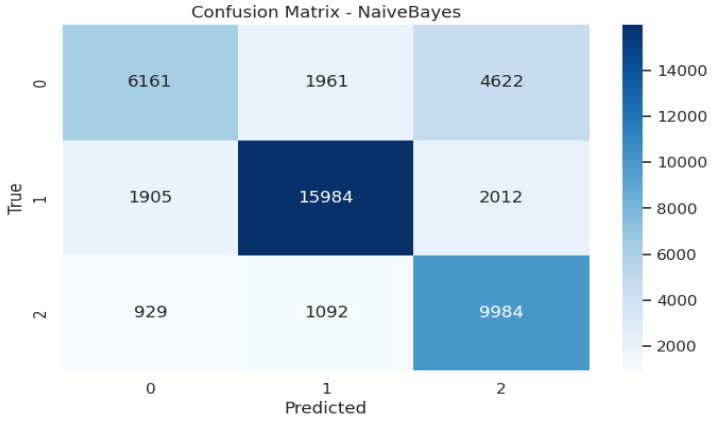


Figure 3. Confusion Matrix — Naive Bayes

Figure 3 shows the confusion matrix for Naive Bayes which shows strong misclassification in all the classes. The widespread distribution of erroneous assumptions that can be seen on figure 3 reinforces the conclusion that Naive Bayes is not appropriate for this dataset based on the existence of strong correlations among features which violate the independence assumptions of this model.

4.3 Zero-Day Attack Simulation (Family Hold-Out Evaluation)

In order to assess the resilience of the system to unseen attack families, five out of the dominant families of ransomware were used for training while all the remaining categories of ransomware were considered zero-day attacks. The Random Forest model was re-trained using the known families only and was evaluated on the known and unknown subsets.

Table 2. Zero-Day Evaluation Accuracy

Evaluation Group	Accuracy
Known Families	**0.9909**
Zero-Day Families	**0.9622**

Although the model was never exposed to the zero day families when training, it had a 96.22% accuracy which provides evidence of strong generalization.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

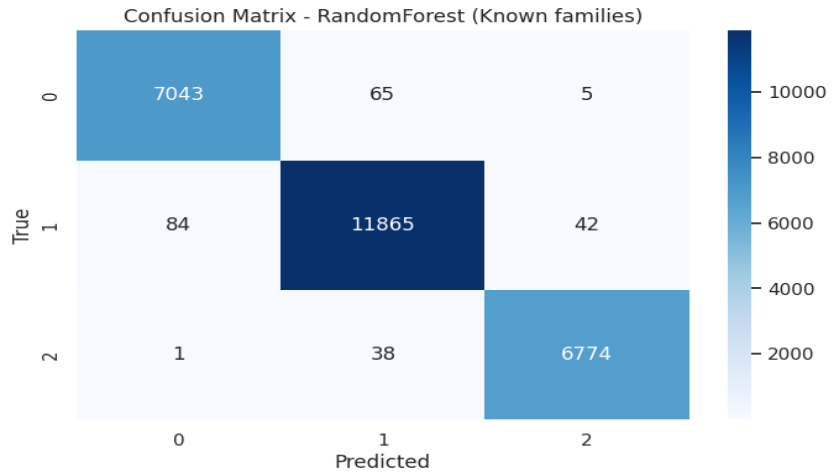


Figure 4. Confusion Matrix — Random Forest (Known Families)

The confusion matrix for predictions on known families is shown in Figure 4. The tight concentration of diagonal entries and the small off-diagonal values reflect the fact that the model perfectly internalized the behavioral structure of the families that were known to it. As shown in figure 4, the model is able to keep high recall and precision across all classes in familiar conditions.

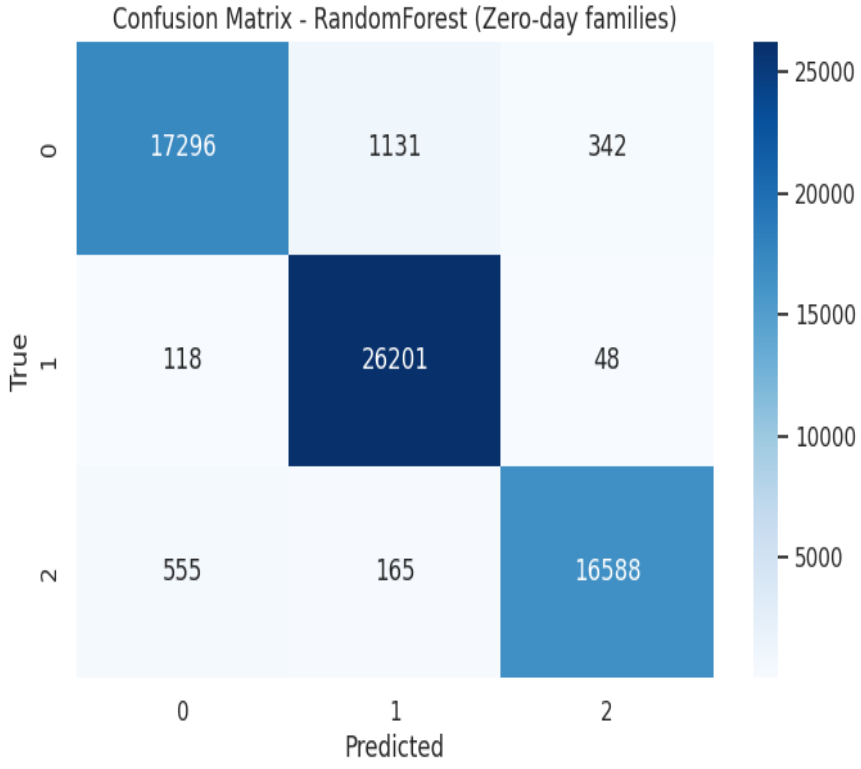


Figure 5. Confusion Matrix — Random Forest (Zero-Day Families)

The confusion matrix of the zero-day families is given in figure 5. Despite the fact that all these families are completely left out of the training, the model still has a high predictive performance, as most samples lie along the diagonal. The moderate level of confusion between the classes 0 and 1 associated with the figure 5 is not surprising given the fact that there has been no prior exposure, but the overall structure shows Random Forest can be an effective way to capture transferable behavioral signatures between unfamiliar ransomware strains.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.4 Anomaly Detection Evaluation (Isolation Forest)

Isolation Forest was employed as an unsupervised based anomaly detector to determine whether ransomware families that were not personally witnessed have behavioral deviations from known ransomware families. The model was only trained with the help of known-family data.

Anomaly Detection Results

- Anomaly rate for known families: 0.4560
- Anomaly rate for zero-day families: 0.6797

The much higher rate of anomaly for zero-day families (+23.37%) suggests that the model was able to classify them as behaviorally distinct.

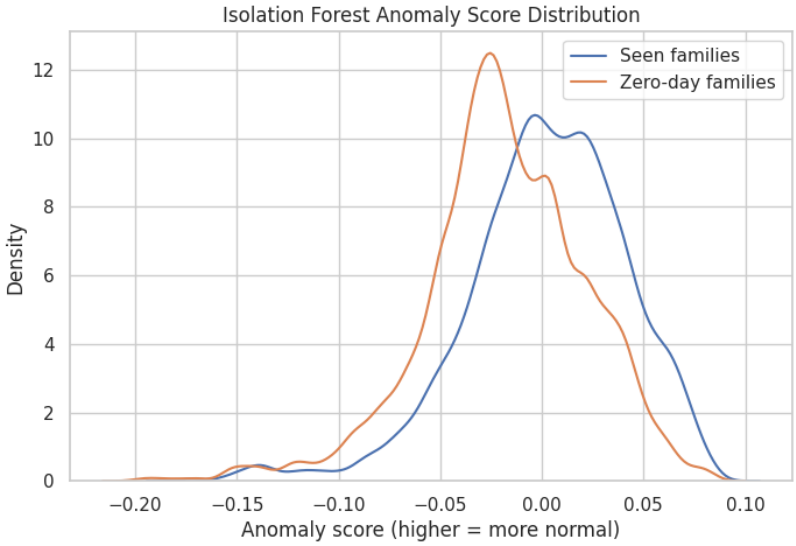


Figure 6. Isolation Forest Anomaly Score Distribution

The curves of the distribution of anomaly scores for known and zero-day families are presented in Figure 6. The zero-day curve is strongly shifted towards lower anomaly scores, proving that these samples are more often isolated by the model. This separation in Figure 6 validates the power of Isolation Forest in detecting behaviors that have yet to be seen by the model supporting its role as a robust zero-day detection mechanism based on anomalies.

4.5 Key Findings and Trends

The results show that the Random Forest classifier can always perform better among all models and the accuracy rate is almost 100% and the shape of the confusion matrix indicates the misclassification is highly concentrated. Although Support Vector Machine(SVM) also provides high accuracy, the moderate misclassification rate classifies it as a good but secondary classifier. In contrast, the Naive Bayes model is unsuitable for this dataset because it has a very simple statistical assumptions and does not pick up the complicated feature dependencies that exist in ransomware behavior. The evaluation of zero-day further gives us that Random Forest keeps its performance strong even when it faces the ransomware families that it hasn't seen before, which means that the Random Forest has strong generalization capability. Complementing this, the Isolation Forest algorithm is effective at detecting behavioral deviations between known and zero day families as is evident from its clear separation of scores of anomalies, highlighting its suitability as an anomaly-based zero day detector. Collectively, combining supervised classifiers into the anomaly detection mechanism shows the importance of a multi-layered approach using AI to combat known and emerging cyber threats.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. Discussion

The objective of this research was to assess the performance of cyber defense mechanisms that are powered by artificial intelligence in the detection of known and zero-day ransomware attacks. By combining supervised classification algorithms and an anomaly approach to detection, the research looked at the successes and failures of various machine learning approaches when presented with known ransomware activity, as well as attack families that had not been observed before. The empirical evidence shows that ensemble learning methods perform much better than classical statistical models, and hybrid designs with the addition of anomaly detection provide a much better resiliency to evolving cyber threats. This section is about interpreting the findings, comparing them to previous research, and drawing practical implications, limitations and future directions.

The supervised learning results showed that there was a hierarchy in model performance. Random Forest achieved the highest accuracy of 99.31% significantly higher than SVM that achieved the accuracy of 96.49% and Naive Bayes that achieved the accuracy of 71.96%. This disparity can be explained by the complexity of interdependencies between features present in the ransomware dataset, which do not meet the independence assumptions on which the Naive Bayes algorithm is based. In contrast, Random Forest is a good way to capture non-linear interactions, hierarchical relationships of features, and the non-linear decision boundaries, making it well suited for high-dimensional cybersecurity data. The zero-day simulation also produced useful insights; even if the model were trained only on the five most common types of ransomware families, the Random Forest model still had a high accuracy, of 96.22%, when tested among zero-day families. The resulting confusion matrix patterns confirmed that even though there was moderate misclassification between classes 0 and 1, the classifier did a good job of generalizing to ransomware classes that had not been seen before. This indicates that latent behavioral similarities - like similar network flow signatures, payment characteristics or propagation characteristics - remain across ransomware families, which allows the model to be extrapolated outside of its training distribution.

The anomalous detection component supported these findings. The result of the Isolation Forest model was an anomaly rate of 0.4560 for known family and 0.6797 for zero-day family, a deviation of +23.37%, and thus showed its ability to isolate unknown behavioral patterns. The distribution of the anomaly scores revealed clear clustering with known families producing higher (more normal) values with zero-day families having a lower valued distribution associated with anomalous activity. This behavior is the crucial role of anomaly-based models in the detection of previously unknown attack variants from known deviations from well-known baselines. Collectively, these results justify the value of combined hybrid defense strategies for security where known threats are classified with the help of supervised models, and novelty or unpredictability patterns are detected by anomaly detectors, providing further resilience against changing cyber threats.

The findings are in good alignment with the existing literature about intrusion detection through artificial intelligence. Numerous studies have shown the superiority of ensemble learning techniques for modeling complex ransomware behavior and especially when the dataset is rich for behavioral and transactional information as in UGRansome. Prior work on feature selection and sensitivity analysis has similarly suggested predictive value of the attributes, such as BTC, USD, Flags, and Threats, corresponding with the present study upon which these variables are found to be key predictors. Anomaly detection research also gives further support to the idea that models like Isolation Forest and One-Class SVM are often better than purely supervised methods in identifying zero-day behaviors, particularly where there are large deviations from the learned behaviors. The zero-day generalization is observed to correlate with the results of reinforcement learning-based intrusion detection frameworks, which can show excellent performance even if whole families of attacks are removed from the training data. Modern surveys also call for hybrid architectures combining the approaches of supervised classification and anomaly detection, which is operationalized and empirically validated in this research.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A number of important implications for practical cyber defense are suggested by these findings. The outstanding performance of ensemble models, especially Random Forest, suggests that these types of algorithms should be used to anchor ransomware detection components in an intrusion detection system. Their capability in handling complex, high-dimensional sets of features makes them robust irrespective of the ransomware variant that is in play (known and unknown). The shown ability of Isolation Forest to identify the behavior deviations confirms the need for the inclusion of the layers on the detection of anomalies in AI-powered security systems, even more so as cybercriminals are actively developing new ransomware classes to bypass signature-based filters. The prominence of behavioral, temporal, and transactional indicators, including BTC, USD, Threats, and ExpAddress doubles down a general industry move toward dynamic behavioral analytics. The successful zero day simulation, additionally, suggests that family hold-out evaluation should become a standard component of model validation, so that researchers and practitioners can better evaluate generalization in the real world. Finally, the good results obtained with AI-based models compared to signature-driven approaches shows that machine learning models can help to reduce dependency on traditional update cycles, and that more adaptive, predictive and forward-looking cyber defense is possible.

Despite these encouraging results, there are a number of limitations that should be acknowledged. Although the UGRansome dataset has structurally rich information, it includes synthetic components that may not accurately represent the behavior of real-world ransomware and earlier studies have focused on the limitations of the dataset including feature variability and temporal relevance. The research further used a limited number of machine learning algorithms, excluding more complex architectures such as autoencoders, LSTM, graph neural networks, deep reinforcement learning models, etc. which have shown good results in similar areas. While the standard preprocessing steps have been applied, more complex feature engineering such as SHAP-based feature interpretability or the recursive feature elimination were not performed, despite their demonstrated ability to enhance the transparency and the predictive power of the model. In addition, all the experiments were offline, whereas real-world intrusion detection systems must be deployed in environments governed by their own set of requirements: high throughput, low latency, and adversarially adaptive systems. The study also did not measure adversarial robustness, although more and more ransomware authors are employing evasion techniques to evade machine learning detection. Consequently, the model's defensive ability against adversarial perturbation has not been tested.

Looking ahead, there are a number of directions in research that are promising based on this work. Future research should investigate the real-time deployment of online learning and learning models that are aware of concept drift, which will allow systems to cope with changing traffic patterns. Integrating deep learning and reinforcement learning architectures might further increase this generalization capacity, which is the basis of the recent success with deep Q-networks and LSTM-based architecture in zero-day detection. Incorporating federated and distributed learning mechanisms may also prove to be of great benefit, especially in the privacy-preserving intrusion detection domain across decentralized environments. Given the emergence of adversarial techniques, the future systems ought to focus more on adversarial training, model hardening, and perturbation awareness detection strategies. Broadening feature sets based on dynamic analysis, such as execution traces, sequence of API calls, and memory forensics, could also be used for better representation of behavior. Overall, this study underlines that AI-powered cyber defense systems based on ensemble learning and anomaly detection provide good capabilities in detecting known and zero-day ransomware threats. While there are still limitations, the findings lay ample paths for building more adaptive, resilient and comprehensive cybersecurity frameworks that are capable of responding to the complexities of contemporary cyber threat landscapes.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

6. Conclusion

This study examined the efficacy of artificial intelligence (AI)-powered cyber defense mechanisms for detection of ransomware and zero-day attacks with the help of supervised machine learning models and an anomaly-based detection approach. The results give strong empirical evidence that ensemble-based supervised learning and unsupervised anomaly detection are a solid foundation for modern ransomware defense systems. Random Forest was able to show near-perfect accuracy for known ransomware families and was able to perform well even in realistic zero-day scenarios, with 96.22% accuracy even though it was trained with only a subset of families. This shows that it has a great generalization ability and can model complex behavior and transactional indicators such as BTC, USD, Flags, and Threats. Complementing this, we found the Isolation forest anomaly detector detected a much higher number of anomalies for the zero-day samples than for known families confirming their ability to flag on behavioural deviations that are associated with unfamiliar ransomware strains. Together, these find the validation of hybrid architectures to combine both predictive modeling with anomaly-based awareness.

The implications of these results are broadened to the practical problem of cybersecurity system design. The demonstrated performance of ensemble models implies that they should be considered for prioritization in intrusion detection systems for ransomware classification. Additionally, anomaly detection has to be incorporated as a complementary layer, to detect previously unseen variants of an attack, to address the main limitations of purely signature-based or completely supervised detection pipelines. The study also demonstrates the importance of adding behavioral, temporal and financial attributes to significantly improve detection accuracy of threats, which highlights the importance of multi-faceted feature engineering in operational systems.

Despite these promising outcomes, there are still a number of limitations. The UGRansome dataset is full of synthetic elements which may not be able to represent the complexity of real-world ransomware ecosystems. The research also tested a small number of algorithms and did not use advanced deep learning and reinforcement learning models. In addition, all the experiments were carried out in offline settings without real-time constraints and adversarial attack scenarios.

Future research should consider real-time deployments or adversarial robustness fundraising, dynamic feature extractions and deep learning/federated detection frameworks integration. Such advancements will lead to further strengthening of the AI giveaways against cyber threats and cyber defence capabilities, against increasingly sophisticated ransomware and zero day attacks.

REFERENCES

[1] A. Khraisat, A. Alazab, S. Singh, T. Jan, and A. Gomez Jr., “Survey on federated learning for intrusion detection system: Concept, architectures, aggregation strategies, challenges, and future directions,” *ACM Computing Surveys*, vol. 57, no. 1, pp. 1–38, 2024.

[2] E. Rios-Ochoa, J. A. Pérez-Díaz, E. Garcia-Ceja, and G. Rodriguez-Hernandez, “Ransomware family attribution with ML: A comprehensive evaluation of datasets quality, models comparison and a simulated deployment,” *IEEE Access*, 2025.

[3] F. Deldar and M. Abadi, “Deep learning for zero-day malware detection and classification: A survey,” *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–37, 2023.

[4] F. El Husseini, H. Noura, O. Salman, and A. Chehab, “Advanced machine learning approaches for zero-day attack detection: A review,” in *Proc. 2024 8th Cyber Security in Networking Conf. (CSNet)*, 2024, pp. 297–304.

[5] G. Gowthami and S. S. Priscila, “Zero-day threat detection: A machine learning paradigm for intrusion prevention,” in *Proc. 2024 7th Int. Conf. Circuit Power Comput. Technol. (ICCPCT)*, vol. 1, 2024, pp. 852–857.

Received: August 05, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[6] I. Mbona and J. H. Eloff, "Detecting zero-day intrusion attacks using semi-supervised machine learning approaches," *IEEE Access*, vol. 10, pp. 69822–69838, 2022.
[7] J. A. Herrera-Silva and M. Hernández-Álvarez, "Dynamic feature dataset for ransomware detection using machine learning algorithms," *Sensors*, vol. 23, no. 3, p. 1053, 2023.
[8] J. Ferdous, R. Islam, A. Mahboubi, and M. Z. Islam, "AI-based ransomware detection: A comprehensive review," *IEEE Access*, 2024.
[9] J. Ispahany, M. R. Islam, M. Z. Islam, and M. A. Khan, "Ransomware detection using machine learning: A review, research limitations and future directions," *IEEE Access*, vol. 12, pp. 68785–68813, 2024.
[10] K. Alam, M. F. Monir, M. J. Hossain, M. S. Uddin, and M. T. Habib, "Adaptive defense: Zero-day attack detection in NIDS with deep reinforcement learning," *IEEE Access*, 2025.
[11] K. Zhang, Y. Wang, U. A. Bhatti, Y. Zhou, and M. Jin, "Enhanced ransomware attacks detection using feature selection, sensitivity analysis, and optimized hybrid model," *Journal of Big Data*, vol. 12, no. 1, p. 245, 2025.
[12] L. Y. Por *et al.*, "A systematic literature review on the methods and challenges in detecting zero-day attacks: Insights from the recent CrowdStrike incident," *IEEE Access*, 2024.
[13] M. A. Ayub, A. Siraj, B. Filar, and M. Gupta, "RWArmor: A static-informed dynamic analysis approach for early detection of cryptographic Windows ransomware," *Int. J. Inf. Secur.*, vol. 23, no. 1, pp. 533–556, 2024.
[14] M. Alkasassbeh, E. H. Omoush, M. Almseidin, and A. Aldweesh, "A self-adaptive intrusion detection system for zero-day attacks using deep Q-networks," *IEEE Access*, 2025.
[15] M. J. Iqbal, S. Aurangzeb, M. Aleem, G. Srivastava, and J. C. W. Lin, "RThreatDroid: A ransomware detection approach to secure IoT-based healthcare systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 2574–2583, 2022.
[16] M. Nkongolo, J. P. Van Deventer, and S. M. Kasongo, "UGRansome1819: A novel dataset for anomaly detection and zero-day threats," *Information*, vol. 12, no. 10, p. 405, 2021.
[17] P. Verma, N. Bharot, J. G. Breslin, D. O'Shea, A. Vidyarthi, and D. Gupta, "Zero-day guardian: A dual model enabled federated learning framework for handling zero-day attacks in 5G enabled IIoT," *IEEE Trans. Consumer Electron.*, vol. 70, no. 1, pp. 3856–3866, 2023.
[18] P. Yan, T. T. Khoei, R. S. Hyder, and R. S. Hyder, "A dual-stage ensemble approach to detect and classify ransomware attacks," in *Proc. 2024 IEEE 15th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, 2024, pp. 781–786.
[19] S. J. Nhlapo and M. N. W. Nkongolo, "Zero-day attack and ransomware detection," *arXiv preprint* arXiv:2408.05244, 2024.
[20] X. Gao, K. Chen, Y. Zhao, P. Zhang, L. Han, and D. Zhang, "A zero-shot learning-based detection model against zero-day attacks in IoT," in *Proc. 2024 9th Int. Conf. Electron. Technol. Inf. Sci. (ICETIS)*, 2024, pp. 309–314.

Received: August 05, 2025