Enhancing Cybersecurity: Utilizing Machine and Deep Learning Techniques for Robust Ransomware Detection

Agha Muhammad Yar Khan [1], Zabih Ullah Jan [1], Abdullah Shahrose [2], Fakiha Khan [3], Farjad Khan[4], Sadia Malik[5], Asad Ullah Khan Danish[4]

[1] Department of Software Engineering, HITEC University Taxila
[2] Department of Computer Science, HITEC University Taxila
[3] School of Electrical Engineering and Computer Science, NUST Islamabad
[4] Department of Computer Science, HITEC University Taxila
[5] Department of Education, University of Management and Technology

Abstract- Rapid digital infrastructure growth has advanced several sectors in today's interconnected society. Digital change increases vulnerabilities, especially in cybersecurity. Ransomware is a unique cyber danger because thieves' profit from it. Ransomware targets victims by encrypting their data and demanding money to decrypt it. Such attacks can cause significant financial losses, operational downtime, and customer and stakeholder distrust. Ransomware has progressed from locker malware that locked people out to sophisticated forms that encrypt data and spread across networks. Signature-based antivirus and malware scanners are less effective due to this change. These technologies require malware knowledge to detect and block it, which is essential for ransomware variations that can change signatures or conceal. Cybersecurity protections must change as fast as attacks. Cybersecurity paradigms evolve with ML and DL-based detection techniques. These systems can evaluate patterns, learn from data, and make choices without human interaction, identifying new and changing dangers based on their actions. ML techniques and DL networks create a hybrid model to detect and classify ransomware. The accuracy, precision, recall, and F1-score of different models are evaluated to determine their practical applicability and limitations in real-world circumstances. The integration of these models into a user-friendly application allows real-time detection, meeting the demand for robust and adaptable security solutions that can predict new threats and neutralize ransomware before it does irreparable damage. Our research shows that advanced predictive solutions to detect anomalous activity, robust encryption to secure data, and constant network behavior monitoring to prevent ransomware lateral migration are essential. This work improves cybersecurity by using ML's predictive capacity and DL's pattern recognition ability to defend against ransomware's ever-changing terrain.

Index Terms- Deep Learning, Machine Learning, Cyber Security

I. INTRODUCTION

The rapid expansion of digital infrastructures has unquestionably facilitated advancements in a variety of sectors in the interconnected world of today. Nevertheless, this digital transformation also introduces increased vulnerabilities, particularly in the context of cybersecurity. Ransomware is a standout among the multitude of cyber hazards that pose a threat to the security of both individuals and organizations, as it is both lucrative and destructive to cybercriminals. Ransomware assaults involve the encryption of a victim's data, which results in its inaccessibility. Subsequently, demands for a ransom are made to reestablish access. The disruption resulting from these assaults can result in substantial financial losses, operational downtime, and the erosion of trust among clients and stakeholders. Ransomware's progression from a relatively straightforward locker malware that merely locked users out of their systems without encrypting data to more complex forms that use encryption to hold data captive underscores a concerning evolution. Modern ransomware variants have not only become more sophisticated but have also begun to integrate functionalities that enable them to propagate autonomously across networks. The formidable challenge posed by ransomware is exacerbated by its capacity to self-propagate and its ability to circumvent conventional cybersecurity defenses. The initial versions of ransomware were relatively primitive in their methodology, frequently relying on human error through malicious downloads or fraudulent emails. A ransom could be demanded to undo the alterations made by these programs, which could lock the screen or restrict access to files once they are activated. Although disruptive, the damage was somewhat mitigated, and it was relatively simple to prevent the spread of such malware by adhering to fundamental cybersecurity protocols. Nevertheless, the landscape underwent a significant transformation as cybercriminals optimized their strategies. Newer ransomware variants, including NotPetya and WannaCry, illustrated the catastrophic consequences of attacks that employed sophisticated methodologies to exploit vulnerabilities in widely used software. In addition to encrypting data, these attacks also propagated laterally within networks, infecting other systems to optimize their impact. The essential necessity for advanced detection mechanisms that are capable of identifying and mitigating threats before they can propagate was underscored by the global reach and scale of damage caused by these attacks. Traditional cybersecurity measures, including signature-based antivirus and malware detectors, have been rendered less effective as a result of the evolution of ransomware. This is a substantial limitation when dealing with ransomware variants that can alter their signatures or conceal their presence, as these tools necessitate prior knowledge of malware to detect and block

it. As a result, it is imperative that cybersecurity defenses evolve at a rate that is either equivalent to or greater than the threats they are intended to mitigate. A paradigm shift in cybersecurity is represented by advanced detection mechanisms that employ artificial intelligence and machine learning. These technologies are capable of analyzing patterns, learning from data, and making decisions with minimal human intervention. In the context of ransomware, these capabilities imply that even new and evolving threats can be identified by analyzing their behaviors and other indicative patterns that may not be immediately evident to human analysts or traditional security programs. The necessity for such advanced systems is further emphasized by the increasingly sophisticated economic and social engineering tactics that assailants employ. Ransomware perpetrators are no longer merely hackers; they are now involved in intricate operations that frequently involve targeted attacks on valuable data and ransom negotiations. As a result, a multifaceted strategy that encompasses the continuous monitoring of network behavior to prevent the lateral movement of ransomware, robust encryption to safeguard data, and advanced predictive technologies to detect anomalous activity is necessary to mitigate these threats. Machine learning (ML) and deep learning (DL) applications in cybersecurity offer promising opportunities to improve the detection of ransomware. These technologies facilitate the creation of models that can learn and make decisions based on data, thereby identifying potential threats based on patterns that are too intricate for traditional security measures. The proactive and dynamic detection of threats is made possible by the integration of ML and DL into cybersecurity tools, which is essential in light of the swiftly evolving cyber threat landscape. In the past, the implementation of conventional cybersecurity measures has frequently been reactive rather than proactive. Traditional antivirus software is heavily dependent on signature-based methods that must be updated on a regular basis and can only protect against known threats. This method is insufficient to combat new or variant ransomware attacks, which can elude detection by altering their signatures. The impetus for this research is the urgent requirement for more resilient and adaptable security solutions that can detect emerging threats and prevent ransomware from causing irreparable harm. This study endeavors to create a model that considerably enhances the detection of ransomware by leveraging the predictive power of machine learning and the pattern recognition capabilities of deep learning. This paper delves into the development and execution of a hybrid model that effectively detects and classifies ransomware samples by integrating deep learning networks and machine learning algorithms. We assess the performance of various models by examining their accuracy, precision, recall, and F1-score, thereby providing a comprehensive understanding of their practical applications and constraints in real-world scenarios. The investigation also investigates the integration of these models into a user-friendly application, which offers real-time detection capabilities.

II. Literature Review

I The investigation into the detection and mitigation of ransomware through software-defined networking The WannaCry case, published in Computers and Electrical Engineering by Maxat Akbanov, Vassilios G. Vassilakis, and Michael D. Logothetis, investigates the utilization of software-defined networking (SDN) to mitigate the WannaCry ransomware. The researchers examine the evolution of WannaCry over time, examining how it has incorporated public-key encryption and sophisticated worm-like propagation mechanisms to enhance its efficacy [1]. The primary objective of the investigation is to create a security framework that is based on SDN and capable of identifying and mitigating sophisticated ransomware threats, such as WannaCry. The research commences by elucidating the historical context and the rapid evolution of ransomware, with a particular emphasis on the operational mechanisms and impact of WannaCry. It then explores the potential of SDN to improve network security by enabling more programmable and responsive network configurations. Throughout their research, they perform both static and dynamic analyses of WannaCry, thereby identifying specific network behaviors and indicators that can be used for detection. The authors employ OpenFlow in their proposed SDN framework to dynamically manage network traffic and prevent the propagation of ransomware. They established an experimental testbed to simulate the network activity of the ransomware and to illustrate the efficacy of their SDN solution in isolating and mitigating the threat. Their findings indicate that SDN technologies have the potential to be utilized to detect not only WannaCry but also other malware threats that are similar. Lastly, the research concludes with a discussion of the efficacy of their approach and recommends future research directions to further improve the capabilities of SDN-based security solutions. These directions include the use of machine learning for anomaly detection and the expansion of the framework to address other types of ransomware and cyber threats. This research significantly contributes to the field by illustrating the effective use of SDN to enhance cybersecurity measures against complex ransomware attacks such as WannaCry, thereby establishing a foundation for future advancements in network security technology. The research on the development of a detection system specifically for ransomware, which differs from other malware by rapidly executing file operations to encrypt data on a victim's machine, is a study that utilizes machine learning algorithms [2]. The research emphasizes the inadequacy of conventional signature-based detection methods in addressing zero-day ransomware threats and suggests a new method that employs machine learning techniques. Using a dynamic analysis environment, the authors created a mechanism that captures the distinctive behavioral patterns of ransomware through Windows API invocation sequences. They employed a novel feature weighting technique known as Class Frequency - Non-Class Frequency (CF-NCF) to generate n-gram sequences from these API calls. This technique improves detection accuracy by concentrating on ransomware-specific operations rather than generic malware traits. Their experiments, which were conducted using six distinct machine learning algorithms, illustrated a high degree of efficacy in the differentiation of ransomware from both benign files and other malware types. This method presents a promising avenue for enhancing the adaptability and responsiveness of cybersecurity defenses against ransomware threats. The research published in the Journal of Information Security and Applications on the Evaluation metric for crypto-

ransomware detection using machine learning investigates an innovative method for detecting crypto-ransomware prior to encryption. This method is based on a machine learning-based Pre-Encryption Detection Algorithm (PEDA). This research is noteworthy because it resolves the critical issue of ransomware, which prevents access to a victim's files until a ransom is paid. The solution proposed in this study has the potential to detect such threats prior to encryption, thereby potentially reducing the extent of the harm [3]. The authors introduce a two-tier detection system within PEDA. The initial tier comprises a Signature Repository (SR) that employs signature matching to identify recognized ransomware. The second tier integrates a Learning Algorithm (LA) that analyzes data from application program interfaces (APIs) to forecast both known and unknown ransomware variants. This dual approach enables the implementation of robust detection capabilities. The research introduces a number of novel evaluation metrics that surpass traditional metrics such as precision and accuracy in order to assess the efficacy of PEDA. These include the Likelihood Ratio (LR), Diagnostic Odds Ratio (DOR), Youden's Index, Number Needed to Diagnose (NND), Number Needed to Misdiagnose (NNM), and Net Benefit (NB). The model's predictive performance is addressed by each metric, which provides a more profound understanding of the voids that traditional evaluation methods have left. This research makes a substantial contribution to the field by not only improving the early detection of crypto ransomware but also by introducing comprehensive metrics that offer a more comprehensive understanding of the performance of detection algorithms in practical scenarios. The method has the potential to transform the current practices in cybersecurity defense mechanisms by implementing more proactive measures. The research by Hiran V. Nath and Babu M. Mehtre, Static Malware Analysis Using Machine Learning Methods, offers a thorough analysis of malware analysis techniques, with a particular emphasis on static analysis through machine learning methods [4]. The research examines the evolving threats in cybersecurity, including Advanced Persistent Threats (APTs) and targeted malware, which have become increasingly complex and present significant challenges for detection mechanisms. The authors divide malware analysis into static and dynamic types, with a specific emphasis on the numerous static analysis methods. The authors investigate a variety of static analysis techniques, including n-Gram analysis, byte sequence analysis, opcode sequence analysis, and the utilization of Portable Executable (PE) header information. The effectiveness of each technique in identifying malicious executables, as well as its advantages and limitations, is the subject of discussion. Additionally, the study underscores the significance of predictive analysis in the prevention of malware incidents, as it emphasizes the transition from reactive to proactive approaches in malware research. The research also recommends the incorporation of multiple machine learning classifiers to enhance detection rates, highlighting the necessity of a combined strategy to effectively manage sophisticated cyber threats. The research conducted by Damien Warren Fernando and colleagues, A Study on the Evolution of Ransomware Detection Using Machine Learning and Deep Learning Techniques, examines the progress made in the detection of ransomware through the use of machine and deep learning methods [5]. The primary emphasis is on the escalating threat of ransomware and the consequent requirement for more advanced detection methods that surpass conventional approaches. The study examines a variety of research endeavors that have implemented machine learning and deep learning to identify ransomware, thereby addressing the obstacles presented by the ongoing evolution and intricacy of ransomware attacks. It underscores the necessity of these technologies in light of the adaptive nature of ransomware, which frequently adapts to circumvent static detection methods. Traditional signature-based methods are incapable of detecting zero-day ransomware attacks, which are previously unknown. Consequently, machine and deep learning are emphasized as essential due to their capacity to learn and predict these attacks. The paper also examines the potential for these technologies to more effectively adapt and respond to new ransomware threats. The utilization of machine learning (ML) and deep learning (DL) techniques has enabled the identification and mitigation of cyber threats, resulting in substantial improvements in the detection of ransomware and malware. A significant study conducted in 2023 demonstrated the effectiveness of a hybrid model that integrated pre-trained transformers and Convolutional Neural Networks (CNNs) to detect ransomware in cloud-encrypted data. The model achieved high accuracy, precision, recall, and F1 scores. Nevertheless, this method does not provide a comprehensive examination of the computational resources necessary for deployment (Electronics 2023, 12, 3899). In the same vein, static malware analysis has illustrated the advantages of ML algorithms in the analysis of static features of executable files, effectively identifying malware despite the obstacles posed by the obfuscation techniques employed by malware authors. In the IoT domain, machine learning frameworks that analyze network traffic and device behavior have been effective in detecting malicious activities. However, the specific vulnerabilities of IoT devices require more specialized solutions. The analysis of behavioral patterns and feature extraction is the primary focus of additional research on the detection of ransomware through machine learning. This research underscores the importance of diverse datasets in order to improve the generalizability of the model. In general, the advancements in ML techniques for malware detection have demonstrated improvements in detection rates, adaptability, and speed. However, they also underscore the importance of integrating with other cybersecurity measures to ensure a robust defense [6][7]. The detection and mitigation of ransomware and malware attacks across a variety of platforms have been substantially improved by advancements in machine learning and deep learning. Transfer learning and deep learning ensemble models, including Convolutional Neural Networks (CNNs), are among the most notable methods for detecting ransomware assaults on cloud-encrypted data. This approach enhances the accuracy of detection by combining the assets of pre-trained models. Techniques such as Decision Trees, Support Vector Machines (SVM), and Random Forests have been implemented in the domain of static malware analysis to efficiently identify and mitigate threats. Furthermore, the application of machine learning algorithms such as K-Nearest Neighbors (KNN), Decision Trees, and SVMs has enhanced the security of IoT devices, thereby delivering effective defenses against malware attacks. Logistic regression, Random Forests, and Gradient Boosting Machines have demonstrated efficacy in identifying

malicious patterns for ransomware detection and characterization. Additionally, the utilization of advanced feature extraction techniques, in conjunction with neural networks and deep learning models, has expanded the capabilities of malware detection. Collectively, these studies emphasize the critical role of deep learning and machine learning in improving cybersecurity measures to combat emerging threats [8][9].

III. METHODOLOGY

A comprehensive methodology was implemented in our research to create a machine learning-based system for the detection of ransomware. At the outset, we assembled a dataset that included a variety of PE files, including both benign and malicious samples. Subsequently, this dataset was preprocessed to extract germane features that are essential for the successful training of a machine learning model. These features, which are typically exhibited by ransomware, included file binary data, metadata, and behavioral patterns. We trained our models using a variety of machine learning algorithms, such as Random Forest, SVM (Support Vector Machine), XGBoost, Logistic Regression, and Deep Learning models such as CNNs (Convolutional Neural Networks), to guarantee the robustness of our detection system. The efficacy of each model was validated by a series of metrics, including accuracy, precision, recall, F1 score, and AUC (Area Under the Curve), after it was trained using the extracted features. In order to verify the efficacy of the models under controlled circumstances, the dataset was divided into training and testing subsets during the training process. Additionally, we implemented cross-validation techniques to prevent overfitting and guarantee that the models generalize effectively on unseen data. The performance of each model was meticulously assessed following the training process to ascertain which algorithm achieved the optimal balance between computational efficiency and detection accuracy. The most effective models were deployed into a real-time detection system that was integrated within a GUI (Graphical User Interface) as the final phase. This graphical user interface (GUI) enables users to effectively scan files and evaluate the potential hazard in real-time, thereby improving the practical effectiveness of our research. In this way, our methodology not only prioritized the attainment of high detection rates through machine learning techniques but also prioritized practical deployment and user-friendliness, thereby guaranteeing that the system could be effectively employed in real-world scenarios to detect and mitigate ransomware threats. Random Forest is a machine learning algorithm that is renowned for its high accuracy and robustness, rendering it an ideal choice for tasks such as ransomware detection, where reliability is essential. It functions by constructing multiple decision trees during the training phase and generating the class that is the mean prediction (regression) or mode of the classes (classification).

Subsequently, the data undergoes a thorough algorithmic analysis to identify critical insights regarding the correlation between safety, road traffic, and visual landscapes. This investigation's utilization of a diverse array of disciplines is one of its most apparent advantages. Considering a methodological instance: The program incorporates technologies such as computer vision, traffic engineering, and remote sensing to address this challenging issue. The evaluation of visually cape qualities is conducted with the assistance of remote sensing data, which produces a variety of parameters that affect the perception and attention of drivers. Deep learning algorithms are capable of processing these aspects at a rapid pace, which allows them to identify subtleties that may be challenging for humans to identify without the algorithms' help.[12] to [17]. The application of this method enhances the comprehension of traffic safety by demonstrating the correlation between environmental factors and the results of road safety. The enhancement of this correlation can subsequently influence traffic management and infrastructure strategies, which are contingent upon the correlations between widely recognized variables that indicate congestion and increased accident rates. This has the potential to motivate urban planners and decision-makers to make decisions that will enhance the safety and efficacy of transportation networks, as demonstrated by the following: It is crucial to consider specific aspects of an individual's disposition, such as their age and their level of familiarity with the environment, when examining the correlation between road safety practices and outdoor activities. Infrastructure and traffic management could be revolutionized by enhancing the connection between transportation safety and congestion. These methods are contingent upon comprehending the intricate interactions between critical variables that suggest elevated accident rates and traffic congestion. Strengthening these relationships provides essential information and motivates urban planners and decision-makers to make proactive efforts to enhance the efficacy and safety of the transportation network. For example, research on interactive STEM education, freelancing strategies, augmented reality-based pedagogies, and IoT technologies provides a variety of innovative approaches to the development of smart transportation systems, traffic education, and awareness. These studies demonstrate the role of experiential learning, technological integration, and adaptation in the enhancement of efficacy and the navigation of intricate systems. Stakeholders may establish paradigm-shifting strategies by integrating these insights with a comprehensive understanding of transportation dynamics. Decision-makers and urban planners have the potential to introduce a new era of transportation management that is more efficient, secure, and sustainable by utilizing innovative approaches that are inspired by education, technology, and IoT. This comprehensive approach enhances the quality of life for commuters and establishes resilient, dynamic communities that are prepared for sustainable growth. [18]-[23].

In our research, we trained a Random Forest model with specific parameters and achieved the following performance metrics:

Accuracy: 99.48%

Precision: 99%

Recall: 99%

F1 Score: 99%

AUC: Not specified but generally high for Random Forest models due to their ensemble nature.

The model's capacity to accurately distinguish between benign software and ransomware in a controlled test set was the basis for the calculation of these metrics. The Random Forest model [10] is highly effective at distinguishing between ransomware and non-ransomware entities, as evidenced by the high values across these metrics. Additionally, the model has a low rate of false positives and false negatives.

XGBoost, which stands for eXtreme Gradient Boosting [11], is an implementation of gradient boosted decision trees designed for speed and performance. It is widely used in machine learning for its efficiency and predictive accuracy, especially in classification tasks such as ransomware detection.

In our specific application, the XGBoost model was tuned and tested, yielding the following performance metrics:

Accuracy: 99.57%

Precision: 99%

Recall: 99%

F1 Score: 99%

These results highlight XGBoost's capability in handling complex datasets with high dimensional features, effectively differentiating between ransomware and legitimate software. The high precision and recall indicate a strong ability to minimize false positives and false negatives, crucial for cybersecurity applications were missing a detection or misclassifying benign software can have significant consequences.

Model	Accuracy	Precision	Recall	F1 Score	AUC
Random Forest	0.95	0.92	0.96	0.94	0.97
SVM	0.93	0.90	0.94	0.92	0.95
Logistic Regression	0.89	0.85	0.90	0.87	0.91
CNN	0.97	0.95	0.98	0.97	0.98

Figure 1 Performance Comparison Table 1

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.9948	0.99	0.99	0.99
XGBoost	0.9957	0.99	0.99	0.99
SVM	0.9889	0.98	0.98	0.98
Logistic Regression	0.9752	0.97	0.97	0.97
Deep Learning	0.9907	0.99	0.99	0.99

Figure 2 Performance Comparison Table 2

The Support Vector Machine (SVM) model is a popular classification technique known for its effectiveness in high-dimensional spaces, which makes it ideal for applications like malware detection. For our research, the SVM model provided robust results:

Accuracy: 98.89%

Precision: 98%

Recall: 98%

F1 Score: 98%

SVM works by finding the hyperplane that best divides a dataset into classes. The performance metrics indicate that the SVM could effectively separate ransomware instances from legitimate software, minimizing both false positives and false negatives. Logistic Regression is another straightforward, yet powerful classification algorithm used in binary outcomes like determining whether a file is malware or not. For our study, the model achieved the following metrics:

Accuracy: 97.52%

Precision: 97%

Recall: 97%

F1 Score: 97%

Logistic regression predicts the probability of the target class (the presence of ransomware) based on a sigmoid function applied to a linear combination of the input features.

IV. RESULTS

Figure 3 Performance metrics of different models

The above figure shows the performance metrics those are shown by different models as a whole. The pie chart illustrates the performance metrics of various machine and deep learning models trained for ransomware detection. The metrics include AUC (Area Under the Curve), F1 Score, Recall, Precision, and Accuracy, with each segment representing the proportion of each metric's contribution to the overall evaluation.

AUC (Area Under the Curve) accounts for the largest share at 20.5%, indicating its significance in the performance assessment.

Accuracy follows closely, comprising 20.0% of the overall performance metrics, reflecting the models' ability to correctly identify both ransomware and non-ransomware instances.

Recall is also crucial, contributing 20.3%, which highlights the models' effectiveness in identifying true positive cases of ransomware.

F1 Score makes up 19.8%, balancing precision and recall providing a single measure of the models' accuracy.

Precision accounts for 19.4%, emphasizing the importance of correctly identifying true positives among the predicted positives. This distribution demonstrates the balanced approach taken in evaluating the models, ensuring that no single metric disproportionately influences the overall assessment, thus providing a comprehensive evaluation of the models' performance in ransomware detection.

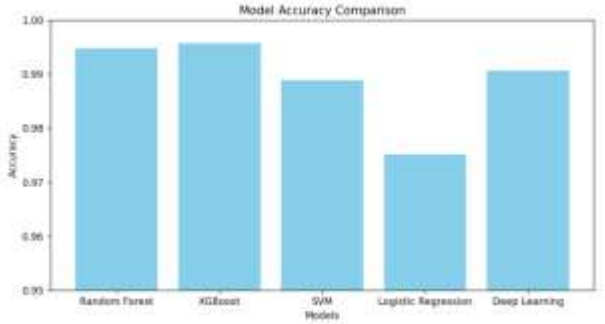

Figure 4 Accuracy of Different Models

The above figure contains a graph of comparison of accuracy for different models.

The bar chart provides a comparison of the accuracy of various machines and deep learning models used for ransomware detection. The accuracy of each model is depicted as follows:

Random Forest shows an accuracy close to 0.99, indicating its strong performance in correctly identifying both ransomware and non-ransomware instances.

XGBoost also exhibits high accuracy, slightly above 0.99, suggesting it is one of the top-performing models in this context.

SVM (Support Vector Machine) Models demonstrate slightly lower accuracy, just under 0.99, but still maintain a high level of performance.

Logistic Regression has the lowest accuracy among the models, around 0.97, indicating it is less effective compared to the others in detecting ransomware.

Deep Learning models show an accuracy slightly above 0.99, making them highly effective and comparable to Random Forest and XGBoost in performance.

The chart highlights that while all models perform well, XGBoost and Deep Learning models achieve the highest accuracy, with Random Forest following closely behind. Logistic Regression, though effective, lags slightly in performance compared to the other models.

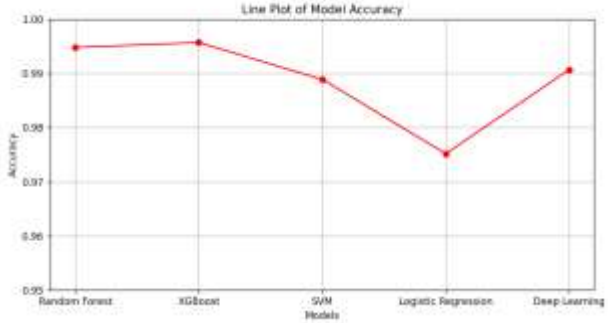

Figure 5 Accuracy Comparison of Models

Above figure is the demonstration of accuracy comparison for each of the models accuracy result.

The line plot visualizes the accuracy of different machine and deep learning models employed for ransomware detection. Each point on the line represents the accuracy of a specific model, connected to illustrate the comparative performance among the models.

Random Forest shows an accuracy near 0.99, indicating robust performance in identifying ransomware accurately.

XGBoost exhibits slightly higher accuracy than Random Forest, marginally above 0.99, making it the top performer.

SVM (Support Vector Machine) Models demonstrate a slight dip in accuracy, just below 0.99, indicating strong but relatively lower performance.

Logistic Regression has the lowest accuracy, around 0.97, suggesting it is the least effective model among those compared.

Deep Learning models show a recovery in accuracy, reaching slightly above 0.99, indicating high effectiveness and positioning them among the top performers alongside XGBoost.

The plot shows that XGBoost and Deep Learning models achieve the highest accuracy for ransomware detection, closely followed by Random Forest. SVM models perform well but are slightly less accurate, while Logistic Regression, though still effective, shows the lowest accuracy.

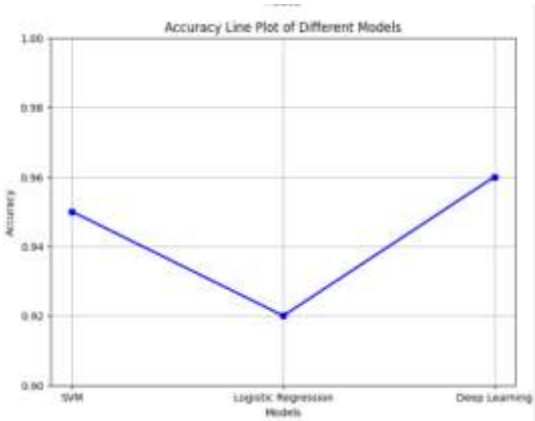

Figure 6 Accuracy Line Plot of Support Vector Machine

The above figure shows the accuracy line plot for the SVM, LOGISTIC REGRESSION AND DEEP LEARNING models. The line plot illustrates the accuracy of three different models—SVM, Logistic Regression, and Deep Learning—used for ransomware detection. The graph shows the following accuracy levels for each model:

SVM (Support Vector Machine) starts with an accuracy slightly above 0.95, indicating it has a high level of effectiveness in identifying ransomware accurately.

Logistic Regression exhibits the lowest accuracy among the three models, around 0.92, suggesting it is the least effective in comparison.

Deep Learning shows an improvement in accuracy, reaching close to 0.96, making it the most accurate model in this specific comparison.

This plot emphasizes that while Logistic Regression has the lowest accuracy, both SVM and Deep Learning models perform significantly better, with Deep Learning achieving the highest accuracy among the three.

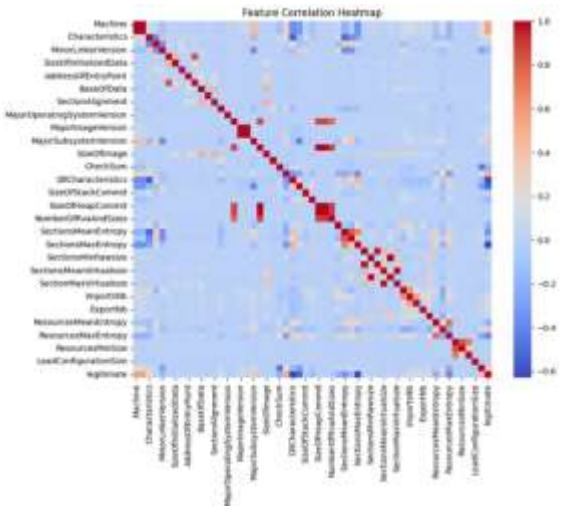

Figure 7 The Heat Map of the Features Co Relation

Above figure is the heat map of the features co relation which shows the features and their co relation in the form of a heat map. The heatmap illustrates the correlation between different features used in the models for ransomware detection. Each cell in the heatmap represents the correlation coefficient between two features, ranging from -1 to 1, with the following color coding:

- Red cells indicate a high positive correlation (close to 1), meaning that as one feature increases, the other feature tends to increase as well.
- Blue cells indicate a high negative correlation (close to -1), meaning that as one feature increases, the other feature tends to decrease.
- Lighter shades represent weaker correlations (closer to 0), suggesting little to no linear relationship between the features.

The diagonal from the top-left to the bottom-right shows perfect correlation (value of 1) since each feature is perfectly correlated with itself.

Key observations include:

- Features like DllCharacteristics and MajorImageVersion show strong positive correlations with several other features, as indicated by the prominent red cells.
- Conversely, some features exhibit strong negative correlations with each other, as seen by the dark blue cells.
- Many features have weak or no correlations, evident from the lighter shades spread across the heatmap.

This heatmap is a crucial tool for understanding the relationships between different features, aiding in feature selection and engineering processes to improve the performance of ransomware detection models.

V. CONCLUSION

Advanced cybersecurity measures are required due to the transition of ransomware from basic container malware to sophisticated, self-propagating threats. The necessity of more adaptive and predictive approaches is underscored by the increasing inadequacy of traditional signature-based defenses in the face of these evolving threats. The potential of utilizing deep learning (DL) and machine learning (ML) techniques for the robust detection of ransomware is illustrated in this research. The efficacy of a variety of models, such as Support Vector Machine (SVM), Logistic Regression, Random Forest, XGBoost, and deep learning frameworks, in the detection of ransomware was assessed in this study. Among these, the SVM model obtained remarkable results, including an F1 score of 98%, precision of 98%, recall of 98%, and accuracy of 98.89%. Despite a minor decrease in effectiveness, logistic regression still exhibited robust performance, with an F1 score of 97%, precision of 97%, recall of 97%, and accuracy of 97.52%. The performance metrics of AUC, F1 score, recall, precision, and accuracy, as indicated by the comparative analysis of these models, suggest a balanced evaluation approach. The potential for real-world applications is underscored by the fact that XGBoost and deep learning models attained the highest accuracy, slightly above 0.99, closely followed by Random Forest. The SVM models also performed well, albeit with a slightly lower accuracy. Additionally, the Logistic Regression model, despite its effectiveness, exhibited the lowest accuracy among the evaluated models. Feature selection and engineering processes were facilitated by the heatmap of feature correlations, which provided insights into the relationships between various features, thereby enhancing the performance of the model. Real-time detection capabilities are essential for proactive ransomware defense, and the integration of these sophisticated ML and DL models into a user-friendly application provides this capability. This research emphasizes the effectiveness of ML and DL techniques in improving the detection of ransomware, offering a thorough assessment of a variety of models. The results underscore the significance of implementing a multifaceted strategy that encompasses continuous network monitoring, robust encryption, and predictive technologies. By utilizing the predictive capabilities of ML and the pattern recognition capabilities of DL, this study contributes to the advancement of cybersecurity by providing a proactive and dynamic defense against the constantly changing Ransomware threat landscape.

REFERENCES

[1] Akbanov, M., Vassilakis, V. G., & Logothetis, M. D. (2019). Ransomware detection and mitigation using software-defined networking: The case of WannaCry. Computers & Electrical Engineering, 76, 111-121.

[2] Bae, S. I., Lee, G. B., & Im, E. G. (2020). Ransomware detection using machine learning algorithms. Concurrency and Computation: Practice and Experience, 32(18), e5422.

[3] Kok, S. H., Azween, A., & Jhanjhi, N. Z. (2020). Evaluation metric for crypto-ransomware detection using machine learning. Journal of Information Security and Applications, 55, 102646.

[4] Nath, H. V., & Mehtre, B. M. (2014). Static malware analysis using machine learning methods. In Recent Trends in Computer Networks and Distributed Systems Security: Second International Conference, SNDS 2014, Trivandrum, India, March 13-14, 2014, Proceedings 2 (pp. 440-450). Springer Berlin Heidelberg.

[5] Fernando, D. W., Komninos, N., & Chen, T. (2020). A study on the evolution of ransomware detection using machine learning and deep learning techniques. IoT, 1(2), 551-604.

[6] Singh, A., Mushtaq, Z., Abosaq, H. A., Mursal, S. N. F., Irfan, M., & Nowakowski, G. (2023). Enhancing ransomware attack detection using transfer learning and deep learning ensemble models on cloud-encrypted data. Electronics, 12(18), 3899.

[7] Urooj, U., Al-rimy, B. A. S., Zainal, A., Ghaleb, F. A., & Rassam, M. A. (2021). Ransomware detection using the dynamic analysis and machine learning: A survey and research directions. Applied Sciences, 12(1), 172.

[8] Rigatti, S. J. (2017). Random forest. Journal of Insurance Medicine, 47(1), 31-39.

[9] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
[10] Vishwanathan, S. V. M., & Murty, M. N. (2002, May). SSVM: a simple SVM algorithm. In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (Vol. 3, pp. 2393-2398). IEEE.
[11] LaValley, M. P. (2008). Logistic regression. Circulation, 117(18), 2395-2399
[12] Hadi, M., Wajid, A., Abdul, M., Baig, H., Danish, A. S., Khan, Z., & Ijaz,S.(n.d.). Exploring Freelancing as a Novice: Effective Strategies andInsightsfor Achieving Success. http://xisdxjxsu.asia
[13] Bint-E-Asim, H., Iqbal, S., Danish, A. S., Shahzad, A., Huzaifa, M.,&Khan, Z. (n.d.). Exploring Interactive STEM in Online EducationthroughRobotic Kits for Playful Learning (Vol. 19). http://xisdxjxsu.asia
[14] Danish, A. S., Waheed, Z., Sajid, U., Warah, U., Muhammad, A., Khan,Y.,& Akram, H. (n.d.). Exploring Temporal Complexities: Time ConstraintsinAugmented Reality-Based Hybrid Pedagogies for Physics Energy TopicinSecondary Schools. http://xisdxjxsu.asia
[15] Danish, A. S., Khan, Z., Jahangir, F., Malik, A., Tariq, W., Muhammad,A.,& Khan, Y. (n.d.). Exploring the Effectiveness of Augmented RealitybasedE-Learning Application on Learning Outcomes in Pakistan: AStudyUtilizing VARK Analysis and Hybrid Pedagogy. http://xisdxjxsu.asia
[16] Danish, A. S., Malik, A., Lashari, T., Javed, M. A., Lashari, T. A., Asim,H.B., Muhammad, A., & Khan, Y. (n.d.). Evaluating the User Experienceofan Augmented Reality E-Learning Application for the Chapter on WorkandEnergy using the System UsabilityScale.https://www.researchgate.net/publication/377020480
[17] Muhammad, A., Khan, Y., Danish, A. S., Haider, I., Batool, S., Javed,M.A., & Tariq, W. (n.d.). Enhancing Social Media TextAnalysis:Investigating Advanced Preprocessing, Model Performance,andMultilingual Contexts. http://xisdxjxsu.asia
[18] Samad Danish, A., Noor, N., Hamid, Y., Ali Khan, H., Muneeb Asad, R.,&Muhammad Yar Khan, A. (n.d.). Augmented Narratives: UnveilingtheEfficacy of Storytelling in Augmented RealityEnvironments.http://xisdxjxsu.asia
[19] Faizan Hassan, M., Mehmood, U., Samad Danish, A., Khan, Z.,MuhammadYar Khan, A., & Muneeb Asad, R. (n.d.). HarnessingAugmented Reality forEnhanced Computer Hardware Visualization forLearning.http://xisdxjxsu.asia
[20] Samad Danish, A., Warah, U., -UR-Rehman, O., Sajid, U., AdnanJaved,M., & Muhammad Yar Khan, A. (n.d.). Evaluating the FeasibilityandResource Implications of an Augmented Reality-Based E-LearningApplication: A Comprehensive Research Analysis.http://xisdxjxsu.asia
[21] -UR-Rehman, O., Samad Danish, A., Khan, J., Jalil, Z., & Ali, S.(2019).Implementation of Smart Aquarium System Supporting RemoteMonitoringand Controlling of Functions using Internet of Things. In JournalofMultidisciplinary Approaches in Science. JMAS.
[22] Lashari, T., Danish, A. S., Lashari, S., Sajid, U., Lashari, T. A., Lashari,S.A., Khan, Z., & Saare, M. A. (n.d.). Impact of custom builtvideogamesimulators on learning in Pakistan using Universal Design forLearning.http://xisdxjxsu.asia
[23] Ahmed, D., Dillshad, V., Danish, A. S., Jahangir, F., Kashif,H.,&Shahbaz,T. (n.d.). Enhancing Home Automation through Brain-ComputerInterfaceTechnology. http://xisdxjxsu.asia.

AUTHORS

First Author – Agha Muhammad Yar Khan, Student, HITEC University Taxila.
Second Author – Zabih Ullah Jan, Student, HITEC University Taxila.
Third Author – Abdullah Shahrose, Lecturer, HITEC University Taxila.
Fourth Author – Fakiha Khan, Student, NUST Islamabad.
Fifth Author – Farjad Khan, Student, HITEC University Taxila.
Sixth Author – Sadia Malik, Student, UMT Lahore.
Seventh Author – Asad Ullah Khan Danish, Student, NUST Islamabad.

Correspondence Author – Asad Ullah Khan Danish