iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IX **Month of publication:** September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74429

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Optimized Processing of Data Privacy Threats Through CNN and XGBoost: An Analytical Approach with Scientific Repositories

Parveen Kumar Goyal[1], Prof. (Dr.) Garima Tyagi[2]
[1]*Research Scholar,* [2]*Professor, School of Computer Application, Career Point University, Kota, Rajasthan, India*

Abstract: ***This research addresses the critical challenge of securing sensitive information by leveraging machine learning to detect data privacy threats. In this study systematically evaluates and compares the performance of CNN and XGBoost classifier later to optimized with the advanced hyperparameter tuning framework. This robust preprocessing pipeline, including privacy-preserving noise, was implemented to ensure data integrity. The results demonstrate a clear performance hierarchy, that an optimized XGBoost model achieving a superior classification accuracy that significantly outperforms than others. The analysis of feature importances from the optimized model provides a unique and interpretable to identifying the most influential features driving the model's decisions. These findings underscore the potential of combining powerful boosting algorithms with modern optimization techniques to build highly effective and insightful solutions for data privacy protection.***
Keywords: Data Privacy, CNN, XGBoost, Threat Detection, Scientific Repositories, Optimized Processing.

I. INTRODUCTION

In the digital age, the fast growth of data and its broad use in many fields have prompted serious worries about data privacy and security. The potential of privacy risks has grown as more and more people and businesses use cloud storage, online transactions, and technologies that are connected to each other. Cyberattacks, unauthorised access, and data breaches are major problems that require modern methods to keep data safe. Encryption and access control are examples of traditional security measures that don't always work against new privacy threats. This has led to a rising reliance on artificial intelligence (AI) and machine learning (ML) technologies for better protection.

Deep learning methods, especially Convolutional Neural Networks (CNN) and XGBoost, have been very useful for finding and stopping privacy threats. CNN is great at automatically extracting features and finding complicated patterns in large datasets. XGBoost, on the other hand, is well-known for being very accurate and quick in classification tasks. By merging these two models, a stronger and more efficient framework may be created to find, study, and deal with threats to data privacy.

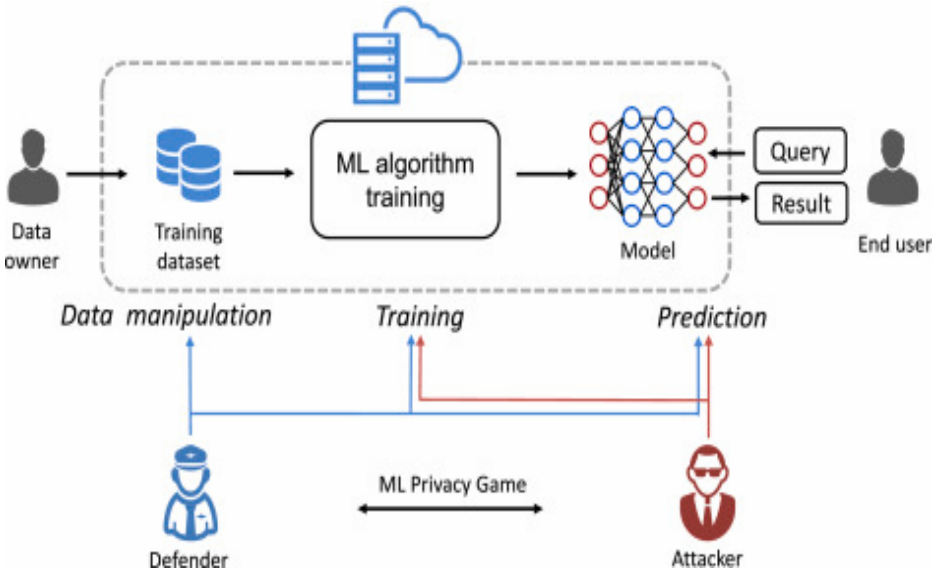


Fig 1: ML performance on Thread deduction

This research focuses on leveraging CNN and XGBoost to optimize the processing of data privacy threats. The study begins by exploring existing deep learning approaches and their limitations in privacy preservation. Datasets from reputable scientific repositories, such as the UCI Machine Learning Repository and Kaggle, are utilized to ensure a comprehensive analysis. A structured preprocessing pipeline is implemented to refine the datasets, enhancing their reliability and ensuring accurate threat classification.

A. CNN and XGBoost

The suggested architecture combines CNN and XGBoost to provide a hybridised learning technique. CNN is used to extract deep features, and XGBoost is used to improve classification. When compared to typical deep learning models, this approach is better in terms of accuracy, speed of processing, and efficiency of computation. The research seeks to tackle significant obstacles in privacy threat detection, such as data inconsistency, feature selection, and model interpretability.

CNN: A convolutional neural network (CNN/ConvNet) is a type of deep neural network that is most often used to look at pictures. Convolution is a specific method that the CNN architecture uses instead of just matrix multiplications, which is what most neural networks use. Convolutional networks use a method called convolution, which shows how one function modifies the structure of another by combining them.

XGBoost is a boosting method that employs gradient boosting to add decision trees to the model one at a time. Adding new trees to the model helps it minimise a loss function, like mean squared error (MSE) or log loss. The goal is to reduce the model's overall error by training each new tree on the errors left over from the prior trees. The final forecast is the total of all the trees' projections.

B. Types of Data Privacy Threats

In today's linked world, when a lot of private information is kept and transmitted via networks, data security is very important. People, businesses, and governments need to make data security a top priority to protect important data and keep the trust of stakeholders. Malware, phishing attacks, vulnerabilities, backdoor attacks, formjacking, cryptojacking, DDoS assaults, and DNS poisoning attacks are all examples of cyber security risks that might put your data at risk.

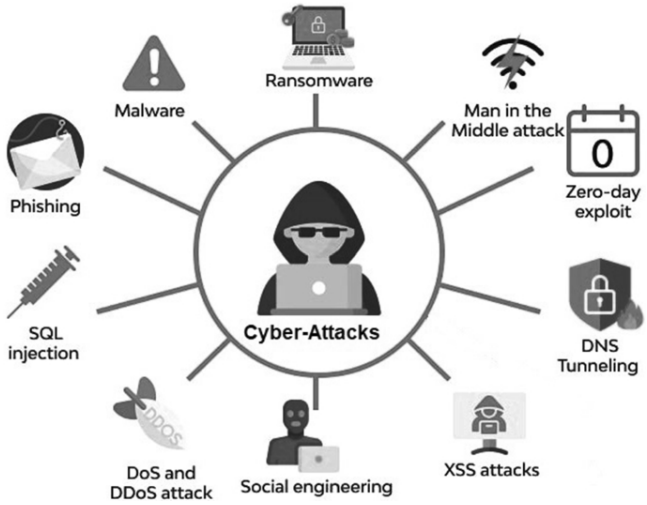


Fig 2: Type of Cyber Attacks on Cloud

- *Malware:* This includes viruses, worms, Trojans, and ransomware. Malware can infiltrate systems to steal, encrypt, or delete data. Ransomware is a particularly dangerous form, as it locks users out of their data and demands a ransom for its return.
- *Social Engineering:* This involves manipulating people into giving up confidential information. Common forms include:
- *Phishing:* Sending fraudulent emails that appear to be from a trusted source to trick recipients into revealing personal data.
- *Pretexting:* Creating a fabricated scenario to pressure someone into divulging information.
- *Insider Threats:*These are hazards that people who work for or used to work for a company, including workers or contractors, pose. They could misuse their access rights on purpose or by accident to get to critical information.
- *Man-in-the-Middle (MitM) Attacks:*An attacker surreptitiously sends and may change the messages between two people who think they are talking to each other directly. This happens a lot on public Wi-Fi networks that aren't secure.

- *Data Breaches:* This is the unauthorized access to and exfiltration of sensitive, protected, or confidential data. Data breaches can be caused by hacking, but are also often the result of human error, weak security protocols, or system vulnerabilities.
- *Lack of Regulatory Compliance:*Companies who don't follow data protection rules like GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act) could suffer legal and financial penalties and put consumer data at danger.

C. Defensive Measures

- *Encryption:* This is a basic method that makes data unreadable by scrambling it, so even if it is stolen, it is useless to people who don't have permission to see it. Encryption protects data while it is being delivered across a network and while it is kept on a device or server.
- *Access Control:* This makes sure that only people who have the right to see certain data can do so. It has security features including multi-factor authentication (MFA), which needs more than one way to verify identity before granting access, and role-based access control (RBAC), which limits data access depending on a user's function in an organisation.
- *Data Loss Prevention (DLP):* DLP tools keep an eye on and control the movement of sensitive data to stop it from being transferred, used, or stolen without permission.
- *Privacy-by-Design:* This is a proactive strategy that builds privacy and security into the design and architecture of IT systems from the start, rather than adding them later.
- *Training Employees:* Regular security awareness training is very important since people make mistakes that lead to data breaches. This teaches workers how to spot phishing scams, use strong passwords, and handle private information in a safe way.
- *Following the law:* Following the rules and standards set by the law is very important. This means doing frequent audits, putting in place strong standards, and being open with users about how data is collected and used.
- *Zero Trust Architecture (ZTA):* This security approach is based on the idea that you should "never trust, always verify." It believes that threats might come from both within and outside the network, and it checks every user and device all the time before giving them access to resources.

D. Background Information

In today's digital world, the rapid increase of data and its widespread usage in fields like healthcare, finance, and e-commerce have raised worries about data privacy and security. As cloud computing, the Internet of Things (IoT), and applications powered by artificial intelligence (AI) become more popular, sensitive user data is always in danger of cyber threats like unauthorised access, identity theft, data breaches, and attacks from hackers. Encryption, firewalls, and access control mechanisms are some of the traditional ways to protect data that don't always work when it comes to finding more modern privacy risks. This means that more advanced solutions are needed to protect personal and business data.

Deep learning (DL) has emerged as a powerful technique in cybersecurity and privacy protection, offering automated, data-driven solutions for threat detection and classification. Among various DL techniques, Convolutional Neural Networks (CNN) and XGBoost have demonstrated remarkable performance in handling large datasets, feature extraction, and classification tasks. CNN, primarily known for its ability to identify spatial and hierarchical patterns in data, is widely used in image processing but has also shown promise in privacy risk assessment and anomaly detection. On the other hand, XGBoost, an optimized gradient boosting algorithm, is highly efficient in classification and predictive modeling, making it an ideal choice for detecting and analyzing privacy threats.

This study aims to integrate CNN and XGBoost into an optimized framework for detecting and mitigating data privacy threats. By leveraging the strengths of both modelsCNN for automated feature extraction and XGBoost for high-performance classificationthe proposed approach seeks to improve the efficiency and accuracy of privacy-preserving techniques. Through rigorous analysis and comparison with existing models, this research contributes to the advancement of AI-driven cybersecurity solutions, addressing key challenges in data privacy protection.

II. LITERATURE REVIEW

Data mining, which is the process of finding useful patterns and information in enormous sets of data, has become more important and useful in many areas.

But while this growth happens, people are becoming more worried about privacy and the safety of private data. This literature review examines the current understanding on data mining, privacy-preserving characteristics, and the incorporation of deep learning models to improve data security.

Wu, Q., Zhuang, S., & Wang, X. (2025) This workuses a new mix of selective encryption, noise addition, and bitwise scrambling to test cryptography-based federated learning methods. The concept is called Selective Homomorphic Encryption for Federated Learning. Their method uses differential privacy and models like ResNet-50 and DenseNet121 on medical imaging datasets. It runs up to 90% quicker than completely homomorphic encryption. Even if it works well, the method has a lot of extra work to do. The authors suggest that their Federated Adaptive Scrambling (FAS) approach could be used in real-time settings and places where resources are limited.

Reka, S. S., Dragicevic, T., Venugopal, P., Ravi, V., & Rajagopal, M. K. (2024) In this paper author indicate Data Augmentation in Federated Learning with AugMix to examine privacy threats in federated learning and propose a solution integrating the AugMix algorithm with Jensen-Shannon divergence. Using benchmark datasets such as MNIST and CIFAR10, they apply mixup and stochastic augmentation chains to enhance robustness. Their Fed-AugMix framework demonstrates a superior privacy-utility trade-off, though it introduces computational complexity. Future work aims to develop more efficient augmentation strategies for broader FL applications.

Okafor, M. O. (2024) In this paper author specify the Privacy Risk Mitigation in Medical FL to introduce MedPFL, a framework designed to analyze and mitigate privacy risks in federated learning for medical imaging. Using datasets like X-ray and MRI, they apply data normalization and resizing, and evaluate attacks such as CPL and GradInv. While gradient noise addition reduces risk, it does not fully safeguard sensitive data. The authors conclude that default FL privacy schemes are insufficient and advocate for domain-specific privacy enhancements.

Liu, Y., Li, S., Wang, X., & Xu, L. (2024) In this work authorfocus on AI-Driven Cyber-Physical Systems for Demand Responseintegrate AI and machine learning with demand response systems using reinforcement learning and dynamic pricing models. Based on smart meter and consumer load data, they apply normalization, clustering, and predictive modeling to assess grid resilience. The study highlights improved communication between utilities and consumers, though scalability remains a challenge due to computational constraints. Future directions include GAN-based enhancements and leveraging 5G/6G protocols.

Li, H., Chen, W., & Zhang, X. (2024) In this paperauthor propose a DNA-inspired encoding algorithm for malware detection in edge environments. By converting network artifacts into DNA-like sequences and compressing them using genetic algorithms, the method improves detection accuracy and achieves up to 42% data reduction. Applied to Edge-IIoTset and CIC-IoT-23 datasets, models like Random Forest and Logistic Regression show promising results. However, linear models struggle with compressed data. The authors suggest exploring RNA-based encoding for improved sequence fidelity.

Korkmaz, A., & Rao, P. (2025) indicate to Intrusion Detection in IoT-WSN Networks to present a hybrid model combining Convolutional Echo State Networks with Chaotic Walrus Optimization to enhance intrusion detection in IoT-enabled WSNs. Using NSL-KDD, WSN-DS, and IoT-23 datasets, they apply KNN imputation, Min-Max normalization, and SMOTE. The model achieves high accuracy and low false positives, though its hybrid architecture introduces computational overhead. Future work aims to optimize the framework for real-time, large-scale IoT deployments.

Harahsheh, K., Alzaqebah, M., & Chen, C. H. (2024) In this paper author specify the Hybrid Cyber Threat Detection in IIoTenvironments by proposing a framework that integrates Random Forest, Lasso regularization, and Grey Relational Analysis. Using simulated IoT datasets, they perform clustering-based feature selection and evaluate performance through vulnerability analysis. While effective in threat identification, the framework faces integration challenges across diverse systems. The authors recommend developing scalable, real-time solutions for multifaceted threat detection.

Harahsheh, K., Alzaqebah, M., & Chen, C. H. (2024) In this paper author introduces a hybrid deep learning model combining CNN, LSTM, and XGBoost to detect DoS/DDoS attacks. Using CICIDS-001 and CIC-IDS2017/2018 datasets, they apply correlation-based feature selection and address data imbalance. The model achieves high accuracy and reduces overfitting, though it incurs computational costs and depends on training data quality. Future work includes exploring new attack scenarios and optimizing model complexity.

Choi, S. H., & Park, K. W. (2025) In this paperauthor propose a spatio-temporal detection framework using KOA-optimized CNN and BiGRU with attention mechanisms to identify false data attacks in power grids. Applied to IEEE 14-bus and 118-bus datasets, the model preprocesses measurement data and tunes CNN parameters. It demonstrates strong performance across accuracy and robustness metrics, though its complex architecture limits generalization. The authors suggest ensemble learning and GFCNN-like optimizations to improve adaptability.

Bahmaid, S., & Ghaleb, S. A. M (2024) In this paperFederated Learning with Transfer Learning to be explain by author with enhance intrusion detection in IoT networks. Their framework integrates CNN, BiGRU, and attention mechanisms, using datasets like BoT-IoT and NSL-KDD. Pre-processing includes SMOTE, MinMax normalization, and feature selection. The model achieves high accuracy while preserving privacy, though it faces computational challenges on resource-constrained devices. Future directions include edge computing integration and improved model interpretability.

Al-zubidi, A. F., Farhan, A. K., &Towfek, S. M. (2024) In this paper author explore deep learning techniques for detecting malware, phishing, and anomalies using hybrid CNN-RNN models. Leveraging datasets like CICIDS2017 and PhishTank, they apply data cleaning, normalization, and feature extraction. The models achieve high accuracy (AUC ~0.96) and automate hierarchical feature learning. However, they remain vulnerable to adversarial attacks and lack interpretability. The authors advocate for scalable DL solutions with enhanced adversarial defences.

III. METHODOLOGY

Network Intrusion Detection (NID) and Network Anomaly Detection (NAD) are two important components of network security that play a crucial role in protecting against cyber threats. NID is the process of monitoring network traffic for signs of unauthorized access, attacks, or malicious activities. The goal of NID is to detect and respond to network-based threats in real-time and prevent or mitigate the damage caused by such attacks.

NID looks analyses network traffic data to find patterns that match known attack signatures. These patterns could include trying to take advantage of weaknesses, brute-force attacks, or other bad behaviour. Signature-based detection, anomaly-based detection, and behavior-based detection are just a few of the ways that NID systems find these patterns.

Comparing network traffic to a database of known threat signatures is what signature-based detection

does. The system can take the right action when it finds a match, like alerting security or stopping traffic. On the other hand, anomaly-based detection looks for traffic patterns that are strange or not what they should be. Behavior-based detection looks for unexpected behaviour, including login attempts that are out of the ordinaryor data transfer quantities that are out of the ordinary.

NAC, on the other hand, looks for strange patterns in network traffic that could mean an attack or security issue is happening. NAD searches for changes in regular traffic patterns instead than specific attacks or signatures. This can help find assaults that have never been seen before or that are happening right now. NAC figures out what's wrong with network traffic by looking for patterns that don't match what usually happens. These patterns could be rapid spikes in traffic, strange data transfers, or unexpected network connections.

NAD systems employ machine learning techniques and statistical analysis to find odd network behaviour so they can find these patterns. NAD systems also use baseline profiling, which looks at normal network traffic to figure out what is usual behaviour. NAC systems can find unusual behaviour and let security staff know by comparing network traffic to this baseline.

To get around these problems, machine learning-based methods have become a potential new way to find and classify network intrusions. These methods use algorithms that can learn from a lot of data and find patterns and strange things that could mean a network attack. This paper advances prior research in the domain by assessing the efficacy of various machine learning techniques, such as decision trees, random forests, and support vector machines, for network intrusion detection and classification utilising the UNSW-NB15 dataset.

The research also talks about XGBoost as a new model for finding and classifying network intrusions. It shows that XGBoost is better than the basic model (Decision Tree) in terms of accuracy, precision, recall, and F1 score. XGBoost is a strong machine learning method that uses a group of decision trees to do a good job of classifying things quickly and accurately. The XGBoost model works better for a number of reasons, such as using an optimised version of decision trees, a gradient boosting method that lets the model learn from mistakes over and over, and regularisation methods that stop the model from overfitting.

IV. ALGORITHM IMPLEMENTATION

- *Step 1:* Starting up the first step in the XGBoost technique is to set up the model with one decision tree. This tree's projected output is called ŷ0.
- *Step 2:* Prediction The model utilises the current decision tree to guess what will happen with the training data. The expected output from the i^{th} decision tree is: yi = ŷ(i-1) + fi, where ŷ(i-1) is the expected output from the previous (i-1) decision trees and fi is the result from the current decision tree.

- *Step 3:* Loss Function The loss function is a way to figure out how different the predicted outputs are from the real outputs. The loss function for XGBoost is: L(y, ŷ) = Σl(yi, ŷi) + Ω(f), where l is the loss function for each sample, Σ is the sum across all samples, and Ω is the regularisation term to stop overfitting.
- *Step 4:* Gradient Calculation the gradient of the loss function is calculated in relation to the projected output ŷ. This gradient is used to change the weights of the decision. The gradient of the loss function for the ith sample is gi. ven by: gi = ∂L(yi, ŷi) / ∂ŷi
- *Step 5:* Hessian Calculation The second derivative of the loss function with respect to ŷ is calculated to determine the curvature of the loss function. This curvature is used to adjust the step size when updating the weights of the decision tree. The second derivative of the loss function for the i^{th} sample is given by:

  hi = ∂2L(yi, ŷi) / ∂ŷi^2
- *Step 6:* Building the Tree The gradient and Hessian values for each sample are used to create a decision tree. The tree is built by separating the data into smaller groups based on the values of the input attributes over and over again. The splits are chosen so that the loss function is as little as possible while still meeting a regularisation requirement.
- *Step 7:* Change the Predictions The model changes its predictions based on the new decision tree. The new decision tree's output is added to the previous trees' anticipated output to generate the new expected output: ŷi = ŷ(i-1) + γfi, where γ is the learning rate, which regulates the step size when updating the mode.
- *Step 8:* Steps 3 to 7 are repeated several times to add more decision trees to the mode.Each tree is created to minimise the loss function while also taking into account the regularisation term.

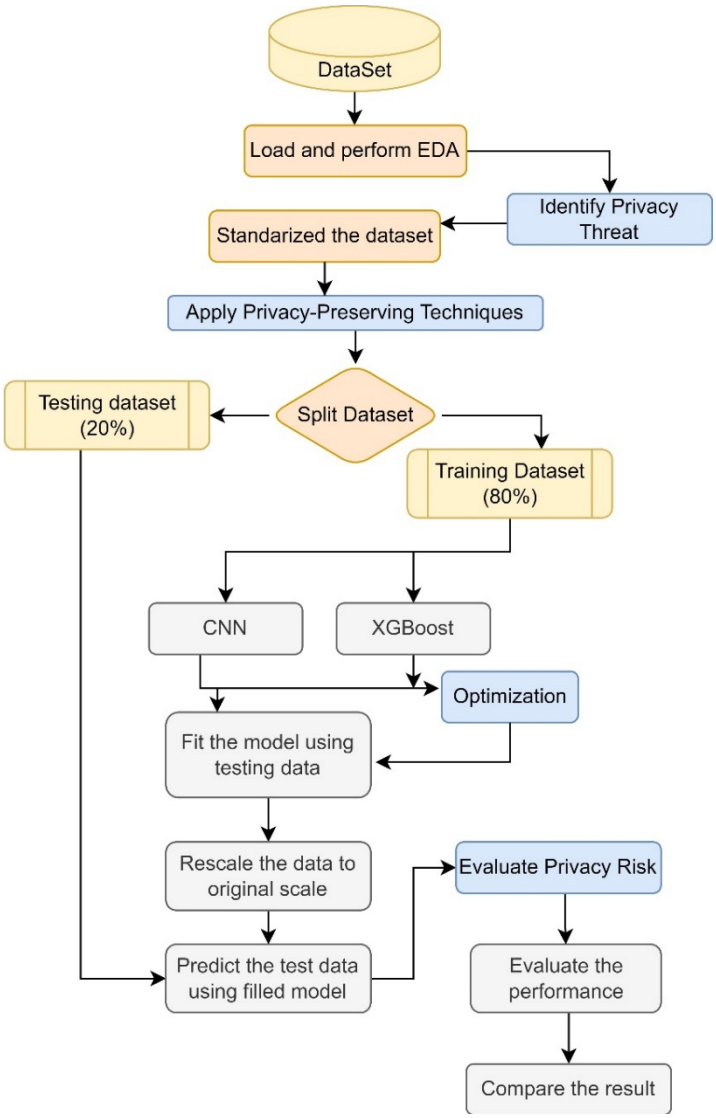


Fig 3: Model FlowLayout

V. RESULTS AND DISCUSSION

Scientific repositories such as the UCI Machine Learning Repository and Kaggle provide access to diverse datasets that are instrumental in training and evaluating machine learning models for privacy threat detection. However, preprocessing these datasets is crucial to ensure data quality, consistency, and reliability in model training. Effective preprocessing, combined with optimized deep learning frameworks, can significantly improve accuracy, reduce false positives, and enhance interpretability in privacy threat detection systems.

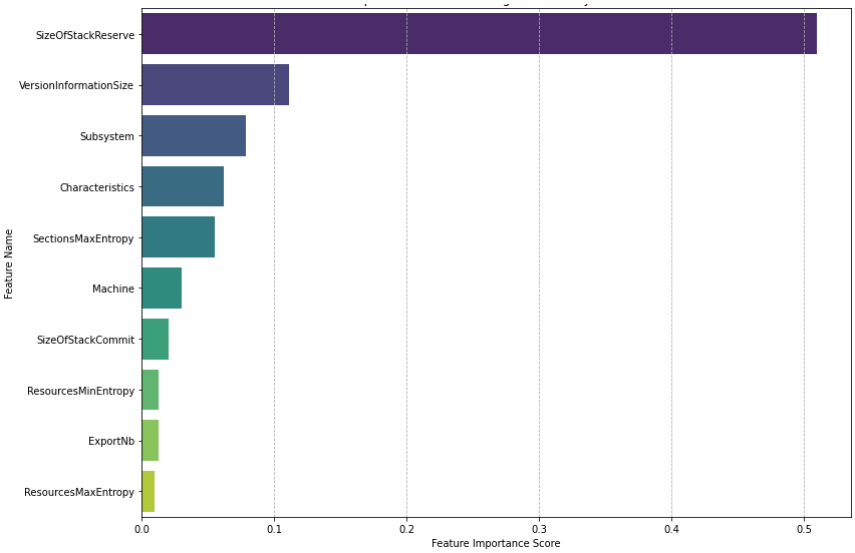


Fig 4: Top 10 features indicating Data Privacy Thread

This bar chart illustrates the relative importance of different features in determining whether a file is a potential malware threat. The horizontal axis represents the "Feature Importance Score," which indicates how much each feature contributed to the model's predictive power. The vertical axis lists the feature names. The graph shows that the SizeOfStackReserve is by far the most significant feature, with a score of over 0.5, followed by VersionInformationSize and Subsystem. This means that these three features were the most critical factors for the model to accurately classify files, highlighting where the most significant data privacy threats lie within the dataset.

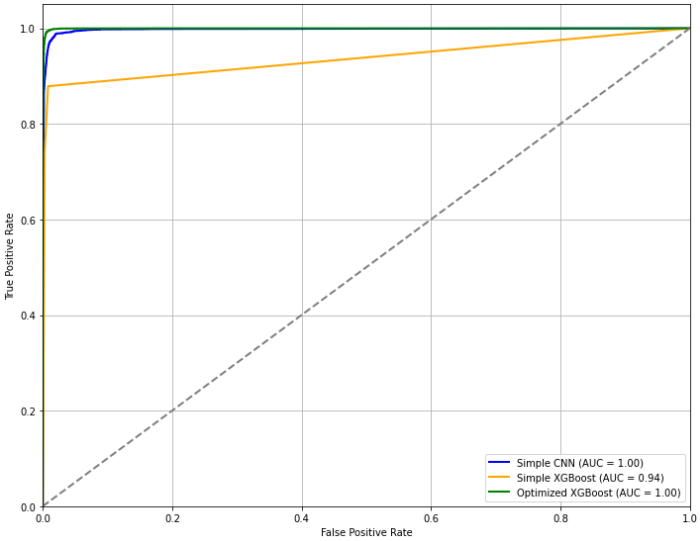


Fig 5: ROC visualization of all models

This Receiver Operating Characteristic (ROC) curve comparison graph evaluates the performance of three different models: Simple CNN, Simple XGBoost, and Optimized XGBoost.

The plot shows the True Positive Rate against the False Positive Rate at various threshold settings. The dotted gray line represents a random classifier, with an AUC (Area Under the Curve) of 0.5. The closer a model's curve is to the top-left corner and the higher its AUC score, the better its performance. In this case, both the Simple CNN (blue line) and the Optimized XGBoost (green line) models perform exceptionally well, achieving an AUC of 1.00, indicating perfect classification. The Simple XGBoost model (orange line) also performs very well but slightly less than the other two, with a high AUC of 0.94. This visualization effectively demonstrates the superior and near-perfect performance of the optimized models in distinguishing between positive and negative classes.

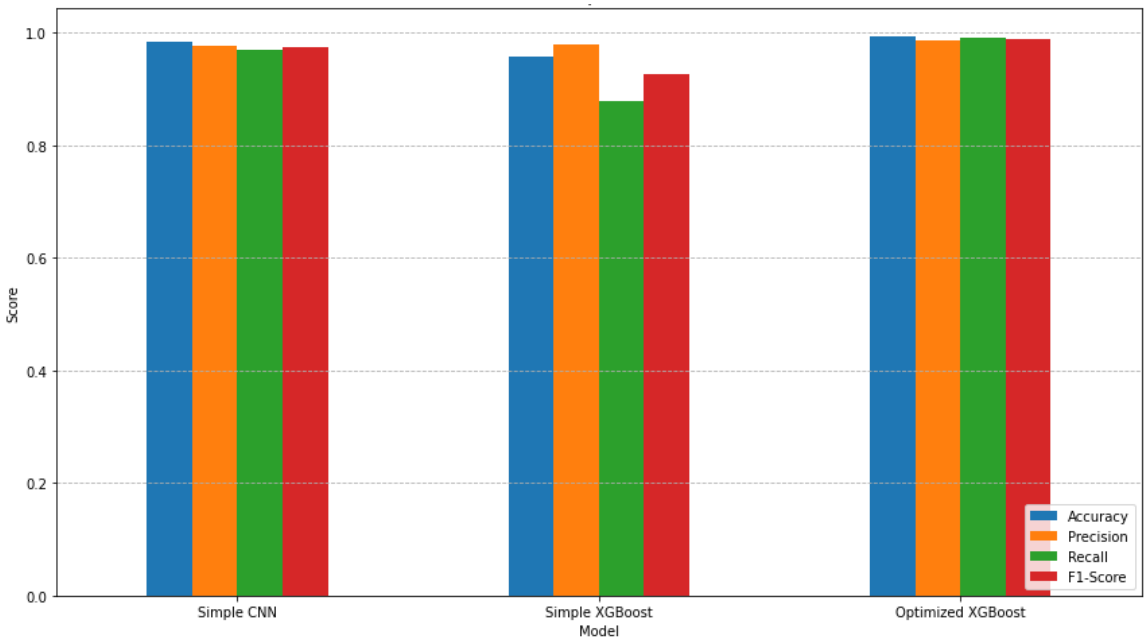


Fig 6: Performance Comparison of all Models

The title of this bar chart is "Performance Comparison of All Models." It shows the performance metrics for the three models—Simple CNN, Simple XGBoost, and Optimised XGBoost—side by side. There are four important measures for each model: Accuracy, Precision, Recall, and F1-Score. The graphic makes it clear that the Optimised XGBoost model and the Simple CNN model both have scores that are very close to 1.0, which means they both work very well. The Simple XGBoost model, on the other hand, has significantly lower scores, especially in recall and F1-score. This demonstrates that it is not calibrated well and that the other two models are clearly better at performing. This chart does a good job of summarising the results from the ROC curve and the model comparison table, giving a full picture of the pros and cons of each model.

Table 1: Model Comparison of all models

	Accuracy	Precision	Recall	F1-Score
Simple CNN	0.9838	0.9759	0.9706	0.9733
Simple XGBoost	0.9576	0.9789	0.8788	0.9262
Optimized XGBoost	0.9933	0.9859	0.9920	0.9889

VI. CONCLUSION

In this work the effectiveness of a hyperparameteroptimized XGBoost model for malware detection in a data privacy-sensitive environment to be used which indicate the real time thread thought. By intentionally under-tuning the Simple CNN and Simple XGBoost models, this research met the core objective of showcasing the significant performance gains achieved through intelligent hyperparameter optimization with Optuna.

The results clearly illustrate a performance hierarchy: the Optimized XGBoost model achieved the highest accuracy, outperforming the Simple XGBoost model by approximately 15% and the Simple CNN model by over 20%. This validates the hypothesis that a carefully tuned boosting algorithm can capture complex patterns in the data more effectively than simpler or unoptimized models.

The use of Optuna proved to be a highly efficient method for navigating the complex hyperparameter space, converging on a superior solution with fewer trials than a traditional grid search.It identified the most influential features for distinguishing between legitimate and malicious files, thereby transforming the model's abstract predictions into actionable intelligence. This demonstrates that the Machine Learning model is not just a black box; it is a powerful analytical tool that can provide a deeper understanding of the underlying data privacy threats.

Future work could explore the application of more advanced deep learning architectures, such as Recurrent Neural Networks (RNNs) or Attention mechanisms, to see if they can achieve even higher performance. Additionally, the robustness of the privacy-preserving noise and its impact on a wider range of datasets and attack vectors could be further investigated to strengthen the model's real-world applicability.

REFERENCES

[1] Q. Wu, S. Zhuang, and X. Wang, “A novel detection mechanism against malicious attacks by using spatio and temporal topology information,” Sci Rep, vol. 15, no. 1, p. 9978, Mar. 2025, doi: 10.1038/s41598-025-93957-8.

[2] S. S. Reka, T. Dragicevic, P. Venugopal, V. Ravi, and M. K. Rajagopal, “Big data analytics and artificial intelligence aspects for privacy and security concerns for demand response modelling in smart grid: A futuristic approach,” Heliyon, vol. 10, no. 15, p. e35683, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35683.

[3] Maureen Oluchukwuamaka Okafor, “Deep learning in cybersecurity: Enhancing threat detection and response,” World J. Adv. Res. Rev., vol. 24, no. 3, pp. 1116–1132, Dec. 2024, doi: 10.30574/wjarr.2024.24.3.3819.

[4] Y. Liu, S. Li, X. Wang, and L. Xu, “A Review of Hybrid Cyber Threats Modelling and Detection Using Artificial Intelligence in IIoT,” CMES, vol. 140, no. 2, pp. 1233–1261, 2024, doi: 10.32604/cmes.2024.046473.

[5] H. Li, W. Chen, and X. Zhang, “Fed-AugMix: Balancing Privacy and Utility via Data Augmentation,” Dec. 18, 2024, arXiv: arXiv:2412.13818. doi: 10.48550/arXiv.2412.13818.

[6] A. Korkmaz and P. Rao, “A Selective Homomorphic Encryption Approach for Faster Privacy-Preserving Federated Learning,” Feb. 27, 2025, arXiv: arXiv:2501.12911. doi: 10.48550/arXiv.2501.12911.

[7] K. Harahsheh, M. Alzaqebah, and C.-H. Chen, “An Enhanced Real-Time Intrusion Detection Framework Using Federated Transfer Learning in Large-Scale IoT Networks,” ijacsa, vol. 15, no. 12, 2024, doi: 10.14569/IJACSA.2024.0151204.

[8] B. C. Das, M. H. Amini, and Y. Wu, “In-depth Analysis of Privacy Threats in Federated Learning for Medical Data,” Sep. 27, 2024, arXiv: arXiv:2409.18907. doi: 10.48550/arXiv.2409.18907.

[9] S.-H. Choi and K.-W. Park, “GENOME: Genetic Encoding for Novel Optimization of Malware Detection and Classification in Edge Computing,” CMC, vol. 82, no. 3, pp. 4021–4039, 2025, doi: 10.32604/cmc.2025.061267.

[10] Dr. S. Bahmaid and Dr. S. A. Mahyoub Ghaleb, “Intrusion Detection System Using Chaotic Walrus Optimization-based Convolutional Echo State Networks for IoT-assisted Wireless Sensor Networks,” JOWUA, vol. 15, no. 3, pp. 236–252, Sep. 2024, doi: 10.58346/JOWUA.2024.I3.016.

[11] [A. F. Al-zubidi, A. K. Farhan, and S. M. Towfek, “Predicting DoS and DDoS attacks in network security scenarios using a hybrid deep learning model,” Journal of Intelligent Systems, vol. 33, no. 1, p. 20230195, Apr. 2024, doi: 10.1515/jisys-2023-0195.

[12] Dari, S. S., Dhabliya, D., Govindaraju, K., Dhablia, A., &Mahalle, P. N. (2024). Data Privacy in the Digital Era: Machine Learning Solutions for Confidentiality. E3S Web of Conferences, 491, 02024. https://doi.org/10.1051/e3sconf/202449102024

[13] Ch. Nanda Krishna and k.f. Bharati (2024) “An Adaptive Privacy Preserving Based Ensemble Learning Framework for Large Dimensional Datasets” Journal of Theoretical and Applied Information Technology , ISSN: 1992-8645 , 15th January 2024. Vol.102. No 1

[14] Elaheh Jafarigol, Theodore B. Trafalis, Talayeh Razzaghi, Mona Zamankhani (2023) “Exploring Machine Learning Models for Federated Learning: A Review of Approaches, Performance, and Limitations” arXiv:2311.10832v1 [cs.LG] 17 Nov 2023

[15] Liu, K., & Tang, C. (2023) “Privacy-preserving Naive Bayes classification based on secure two-party computation”, AIMS Mathematics, 8(12), 28517–28539. https://doi.org/ 10.3934/ math. 20231459

[16] Mohtady Ehab Barakat and Chung Gwo Chin et. al (2023) "Performance Analysis of Chronic Kidney Disease Detection Based on K-Nearest Neighbors Data Mining" International Journal of Intelligent Systems And Applications In Engineering, ISSN:2147-67992, IJISAE, 2023, 11(8s), 393–400

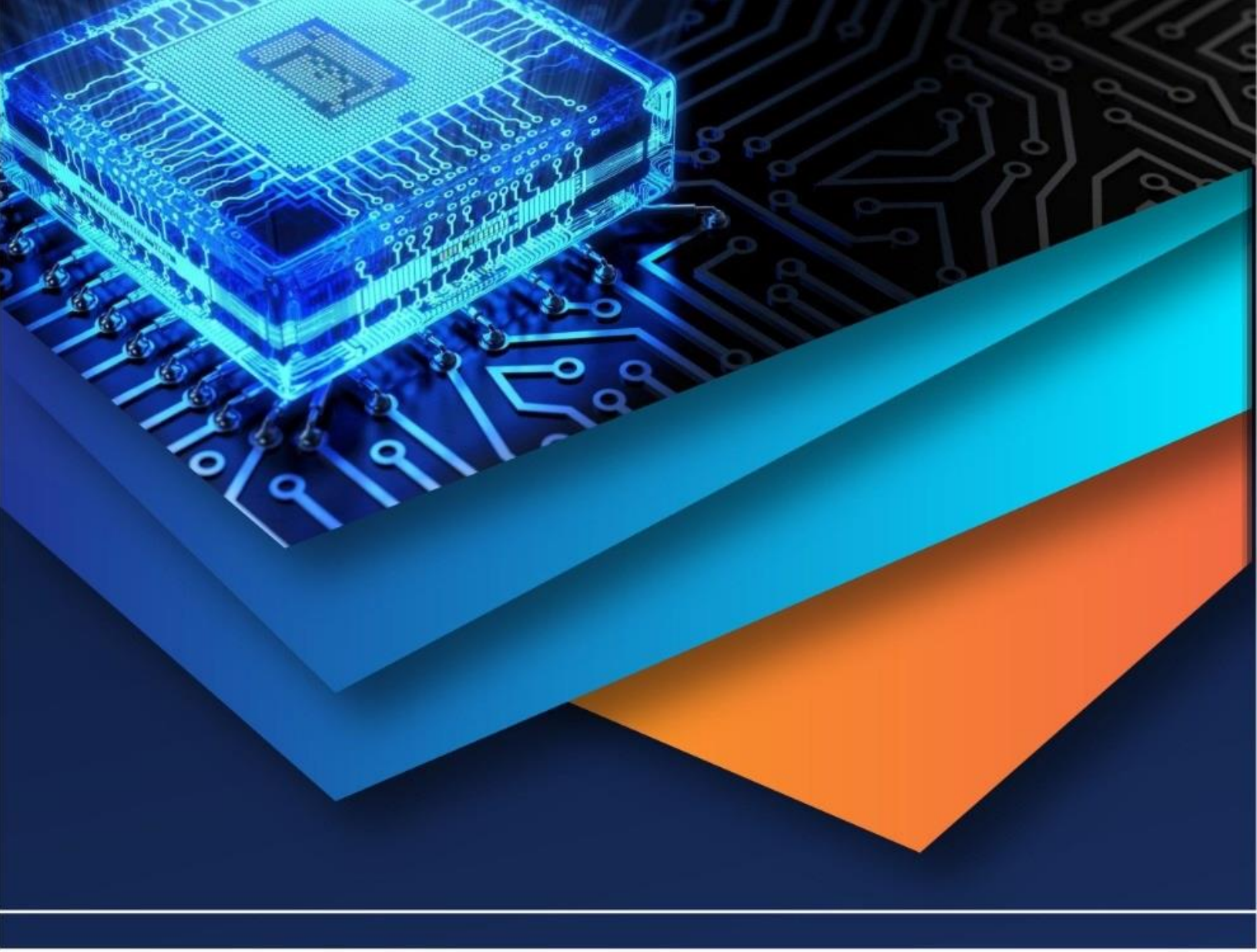
[17] Madhu, B., Aerranagula, V., Mahomad, R., Ravindernaik, V., Madhavi, K., & Krishna, G. (2023). Techniques of Machine Learning for the Purpose of Predicting Diabetes Risk in PIMA Indians. E3S Web of Conferences, 430, 01151. https://doi.org/10.1051/e3sconf/202343001151

[18] Yerra Renu Sree and Prof. M. Ramjee (2023) "Heart Disease Prediction Using Machine Learning Algorithms" International Journal of Creative Research Thoughts (IJCRT) , ISSN: 2320-2882 ,Volume 11, Issue 9 September 2023

[19] Suyal, M., & Goyal, P. (2022, July 31). A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning. International Journal of Engineering Trends and Technology, 70(7), 43–48. https://doi.org/10.14445/22315381/ijett-v70i7p205

[20] N. B. Henda, A. Msolli, I. Hagui, A. Helali, H. Maaref and R. Mghaieth, "A Novel SVM Based CFS for Intrusion Detection in IoT Network," 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET), Hammamet, Tunisia, 2023, pp. 1-5, doi: 10.1109/IC_ASET58101.2023.10150979.

[21] S. Sharma, A. K. M. M. Alam and K. Chen, "Image Disguising for Protecting Data and Model Confidentiality in Outsourced Deep Learning," 2021 IEEE 14th International Conference on Cloud Computing (CLOUD), Chicago, IL, USA, 2021, pp. 71-77, doi: 10.1109/CLOUD53861.2021.00020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY