Development of Cybercrime Prediction Model using Hybridized Algorithms

Chukumeka Gift Iroanwusi, Friday Eleonu Onuodu, *Davies Isobo Nelson, Bassey Aniefiok Tom

[1&4](Department of Computer Science Iqnatius Ajuru University of Education, Port Harcourt Nigeria)
[2](Department of Computer Science, University of Port Harcourt, Nigeria)
[3](Department of Computer Science, Rivers State University, Port Harcourt Nigeria)

Corresponding Email: isobo.davies@ust.edu.ng

--------------------------------------***************************-------------------------------

Abstract:

The growing sophistication of cyber threats demands intelligent, adaptive security solutions capable of detecting malicious network activity in real time. This study presents a hybrid cybercrime prediction model that integrates Artificial Neural Networks (ANN), Autoencoders, Support Vector Machines (SVM), and XGBoost for the classification of network traffic as either benign or malicious. Structured network traffic features extracted from the CIC-DDoS2019 dataset including IP addresses, ports, timestamps, protocol types, and payload sizes were used to train and evaluate the model using a 70:30 train-test split. To address privacy concerns in network data analysis, the system incorporates the CKKS homomorphic encryption scheme, enabling secure computation on encrypted data without exposing sensitive information. Additionally, the system employs heuristic URL analysis integrated with Google Safe Browsing and VirusTotal APIs for phishing site detection. Experimental results demonstrate that the system accurately classifies both DDoS and benign traffic patterns, and correctly identifies phishing URLs with a 35% detection rate among tested URLs. Furthermore, the encryption and decryption operations performed within milliseconds confirm the practical efficiency of the system's privacy mechanism. The proposed framework offers a robust, privacy-preserving approach to cybercrime prediction, combining predictive accuracy with data confidentiality for real-world deployment.

Keywords-**Cybercrime Detection, Machine Learning, Artificial Neural Network, Support Vector Machine, XGBoost, Homomorphic Encryption, DDoS Attack, Phishing Detection, Network Traffic Classification, Cybersecurity**

I. INTRODUCTION

In recent years, the rapid digitisation of modern life has made businesses, governments, and individuals increasingly reliant on internet-based platforms for communication, financial transactions, healthcare, and critical infrastructure [1]. While this transformation has driven efficiency and connectivity, it has equally expanded the attack surface for cybercriminal activity. Global cybercrime costs are projected to reach trillions of dollars annually within the next decade, underscoring the urgent need for more intelligent, proactive security solutions [2].

Cyber threats have grown substantially in scale and sophistication since the widespread adoption of internet connectivity in the 1990s. However, today's threat landscape encompasses malware, phishing, data breaches, advanced persistent threats, and distributed denial-of-service (DDoS) attacks [3]. Technically, DDoS attacks are disruptive due to their capacity to overwhelm network resources and cripple online services. Conventional defences such as signature-based antivirus tools, rule-based detection systems, and static blacklists, while effective against known threats, they often struggle against novel or evolving attack patterns such as zero-day attacks [4]. This limitation has driven growing interest in Machine Learning (ML) as an adaptive alternative.

Machine Learning models enables systems to learn from historical network traffic data by analysing features such as IP addresses, ports, protocols, timestamps, and payload characteristics to distinguish benign activity from malicious behaviour in real time [5]. Further, supervised learning algorithms, including Artificial Neural Networks (ANN) and Support Vector Machines (SVM), have demonstrated strong performance in anomaly detection and threat classification [6]. However, models relying on a single algorithm can be limited in handling complex, dynamic attack patterns.

This study addresses that gap by proposing a hybrid cybercrime prediction model that integrates ANN and SVM with a cryptographic privacy mechanism based on homomorphic encryption. Using structured network traffic features from the CIC-DDoS2019 dataset, the model was able to classify network traffic activities as benign or malicious while preserving the confidentiality of sensitive data during analysis. This hybrid setup presents the system with predictive robustness and privacy-aware design.

II. RELATED WORKS

Cybercrime poses an increasing threat to organizations and individuals globally, with criminals employing advanced techniques to breach security systems and access sensitive data. The paper of [7] provides a comprehensive survey of recent advancements in cybercrime prediction.

Class imbalance in Intrusion Detection was addressed by [8]. Their study aimed to evaluate various machine learning algorithms under different levels of class imbalances, using resampling.

The study of [9], analyses some machine learning techniques proposed in recent years. In their study, several classifications were made to detect anomalous behaviour in network traffic. The models was built and evaluated based on the CIC-IDS-2017 dataset. The authors conducted their experiment using different ML models and their ensemble approach with the model outperforming the rest.

[10] proposed an optimized stacking ensemble method for phishing website detection. They conducted their optimisation using genetic algorithm (GA) to tune the parameters of several ensemble machine learning methods, including random forests, AdaBoost, XGBoost, Bagging, GradientBoost, and LightGBM. The study employed three different datasets (Phishing Websites Data from UCI, Phishing Dataset for Machine Learning from Mendeley, and Datasets for Phishing Websites Detection from Mendeley) Their experimental results showed an improvement using the optimized stacking ensemble method, thereby achieving accuracy in detection of 97.16%, 98.58%, and 97.39% for the three datasets respectively.

[11] developed a machine learning approach to forecast Phishing and DDoS attacks. The system employed several techniques such as Logistic Regression, KNN, SVC, Random Forests, Decision Trees, and Naive Bayes to analyse these attack patterns in real-time. Their experimental findings revealed that the Random Forest algorithm outperformed all other models, achieving an accuracy of 97.3%.

[12] developed an effective Network Intrusion Detection System (NIDS) based on ML and feature selection techniques. Their system employed four different ML models and achieved an impressive detection accuracy rate of 99.72%.

A novel security model was proposed by [13]. The authors employed Fully Homomorphic Encryption (FHE) to safeguard data privacy in pervasive computing environments. The study outlined a four-layered architecture for data collection, encryption, encrypted computation, and decryption. By employing the Brakerski/Fan-Vercauteren (BFV) scheme through Microsoft SEAL, the model achieved 95.8% accuracy in security and minimal privacy loss of 0.6%, with a processing overhead of 720 ms.

Researchers have developed a multi-objective optimization-based hybrid method that employs lightweight deep learning models to enhance cybercrime detection accuracy. The study utilized QR code images embedded with diverse datasets, while the MobileNetV2 Convolutional Neural Network (CNN) offers efficient image processing and feature extraction. After training, a genetic algorithm (GA) refines the classification by identifying top features, achieving an overall success rate of around 96%. The SVM classifier demonstrated strong performance with an F1-score of 0.86 and accuracy of 96.2%, while the KNN classifier achieved approximately 85.78% accuracy [14].

Researchers presented an Intrusion Detection and Prevention System (IDPS) using a novel and Adaptive Hybrid Case-Based Neuro-Fuzzy System (HCBNFS). The Case-Based Reasoning (CBR) component serves as the primary detection engine to identify network traffic patterns, while the Neuro-Fuzzy Inference System (NFIS) enhances the analysis of unknown traffic and refines CBR's reverse phase inquiry. The model was trained and tested using the CIC-IoT2022 dataset, achieving 99% accuracy in intrusion detection, along with precision, recall, and F1-Score metrics of 99.5% [15].

III. METHODOLOGY

The proposed system employed a structured network traffic features such as IP addresses, ports, timestamp, protocol types, and payloads which was extracted from the CIC-DDoS2019 dataset to predict whether a traffic is benign or malicious. Note, for this study, the collected dataset was slitted using 70:30 ratio with 70 for training and 30 for testing.

The collected dataset was pre-processed and cleaned before feeding them into the proposed prediction model. The system was implemented using Visual Studio Code for interface design and Python programming language with plugged-in ML libraries for the development of the ML models. Further the study employed MySQL database as the system's database server.

In this study, Artificial Neural Network (ANN) and Autoencoder were employed for the prediction, while Support Vector Machine (SVM) and XGBoost were utilized for classification of network traffic patterns. Additionally, the study employed an homomorphic cryptography method for data security and privacy. Presented in Fig. 1 is the architecture of the cybercrime prediction.

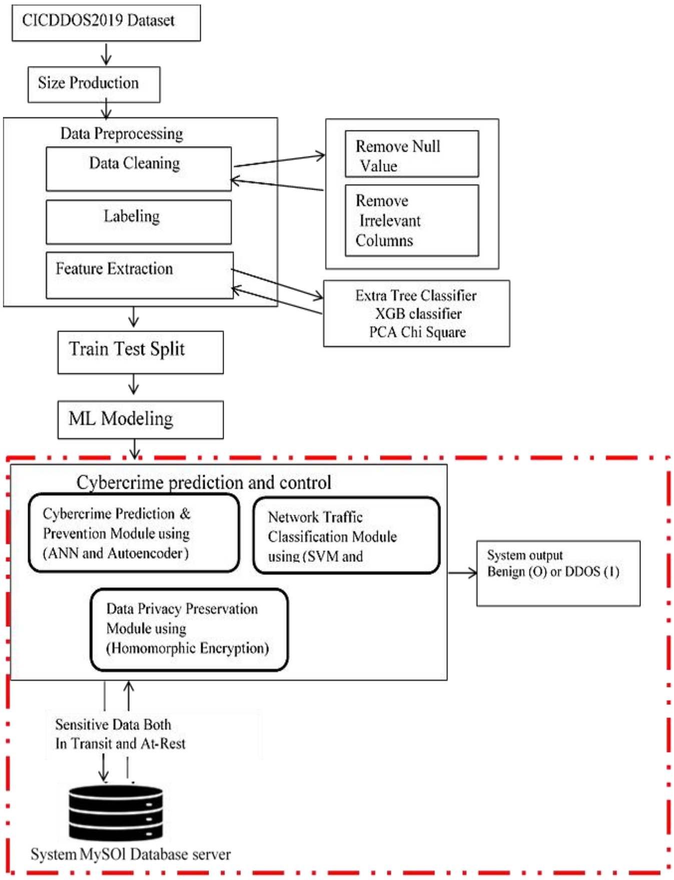

CICDDOS2019 Dataset
Size Production
Data Preprocessing
Data Cleaning
Labeling
Feature Extraction
Remove Null Value
Remove Irrelevant Columns
Extra Tree Classifier
XGB classifier
PCA Chi Square
Train Test Split
ML Modeling
Cybercrime prediction and control
Cybercrime Prediction & Prevention Module using (ANN and Autoencoder)
Network Traffic Classification Module using (SVM and
System output Benign (O) or DDOS (1)
Data Privacy Preservation Module using (Homomorphic Encryption)
Sensitive Data Both In Transit and At-Rest
System MySOl Database server

Fig 1: Architecture of the Proposed System

This study applied Accuracy, Precision, Recall, and F1-Score performance metric to evaluate the proposed system. The formula for these metric are as follows:

Accuracy:

$$CybercrimeModel_{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

(1)

Precision:

$$CybercrimeModel_{Precision} = \frac{TP}{TP+FP}$$

(2)

Recall:

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

F1-Score:

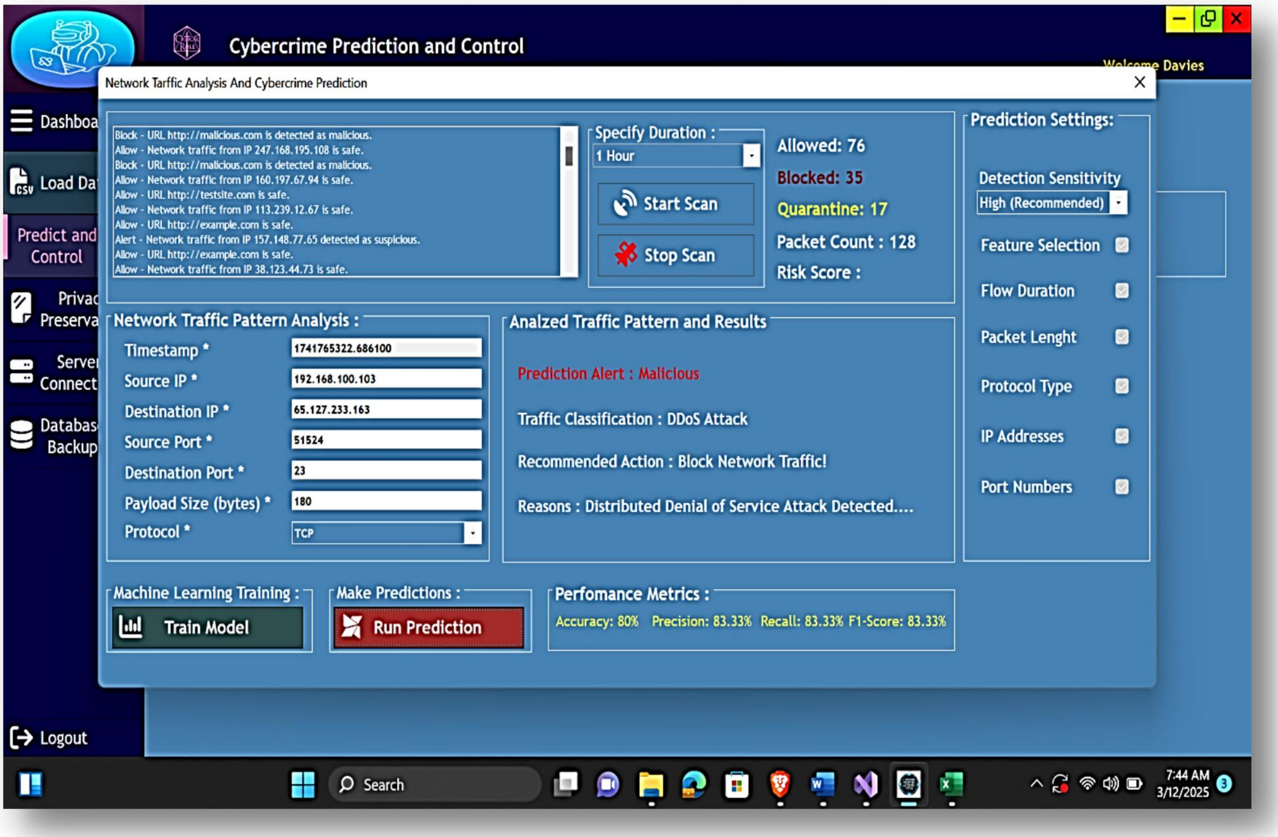
$$CybercrimeModel_{F1-Score} = \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

IV. RESULTS

Fig. 2 captures the screenshot results of the developed system for successful prediction and classification of DDoS attack. The system analysed network traffic to predict potential cybercrime or threats using Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The classification results for different network-based traffic patterns are presented in Table I.



Network Tarffic Analysis And Cybercrime Prediction

```
Block - URL http://malicious.com is detected as malicious.
Allow - Network traffic from IP 247.168.195.108 is safe.
Block - URL http://malicious.com is detected as malicious.
Allow - Network traffic from IP 160.197.67.94 is safe.
Allow - URL http://testsite.com is safe.
Allow - Network traffic from IP 113.239.12.67 is safe.
Allow - URL http://example.com is safe.
Alert - Network traffic from IP 157.148.77.65 detected as suspicious.
Allow - URL http://example.com is safe.
Allow - Network traffic from IP 38.123.44.73 is safe.
```

Specify Duration : 1 Hour

Start Scan

Stop Scan

Allowed: 76
Blocked: 35
Quarantine: 17
Packet Count : 128
Risk Score :

Prediction Settings:

Detection Sensitivity: High (Recommended)

Feature Selection ☑
Flow Duration ☑
Packet Lenght ☑
Protocol Type ☑
IP Addresses ☑
Port Numbers ☑

Network Traffic Pattern Analysis :

Timestamp *	1741765322.686100
Source IP *	192.168.100.103
Destination IP *	65.127.233.163
Source Port *	51524
Destination Port *	23
Payload Size (bytes) *	180
Protocol *	TCP

Analzed Traffic Pattern and Results

Prediction Alert : Malicious

Traffic Classification : DDoS Attack

Recommended Action : Block Network Traffic!

Reasons : Distributed Denial of Service Attack Detected....

Machine Learning Training : Train Model

Make Predictions : Run Prediction

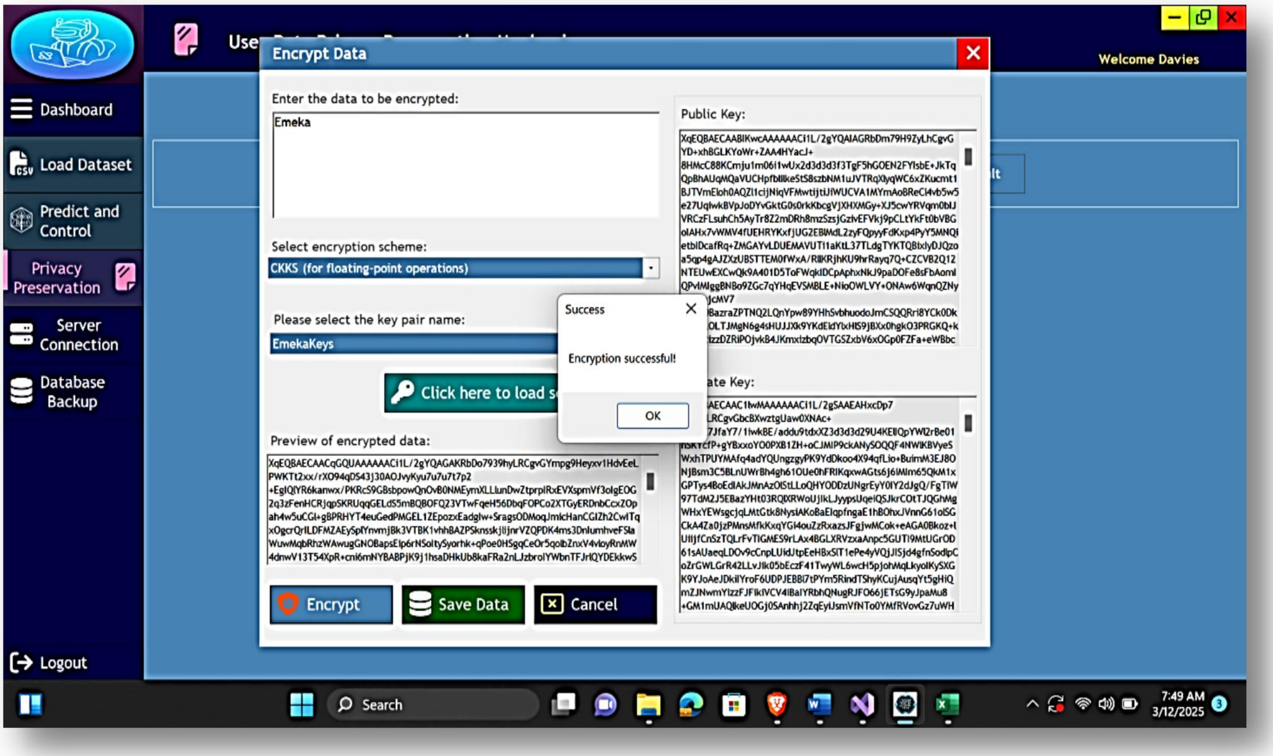
Perfomance Metrics :
Accuracy: 80% Precision: 83.33% Recall: 83.33% F1-Score: 83.33%

Fig 2: Prediction and Classification Screenshot Interface

TABLE I
TESTED NETWORK TRAFFIC PATTERNS

Timestamp	Source IP	Source Port	Destination IP	Destination Port	Protocol	Payload	Output
1525880074.015853	192.168.100.103	51524	65.127.233.163	23	TCP	180	DDoS
1525880022.015990	192.168.100.103	41101	111.40.23.49	23	TCP	60	DDoS
1525880078.015841	192.168.100.103	60905	131.174.215.147	23	TCP	180	DDoS
1525880066.731484	192.168.100.103	44301	91.42.47.63	23	TCP	60	DDoS
1525880070.045846	192.168.100.103	50359	139.124.68.11	2323	TCP	60	DDoS
1525880017.016316	192.168.100.103	43763	117.35.253.72	28970	UDP	40	Benign
1525880073.006919	192.168.100.103	33756	141.241.245.73	9527	TCP	180	DDoS
1526282631.03078	192.168.100.103	37003	96.102.99.11	23	TCP	60	DDoS
1526282628.03075	192.168.100.103	43763	179.85.180.245	21619	UDP	40	Benign
1526282681.06039	192.168.100.103	43763	220.209.211.228	18712	UDP	40	Benign
1525880147.04637	192.168.100.103	45218	8.162.132.20	57933	TCP	180	Benign
1525880093.01614	192.168.100.103	43763	255.82.48.235	21837	UDP	40	Benign
1525880087.16667	154.205.133.202	3	192.168.100.103	3	ICMP	68	Benign
1525880088.00669	192.168.100.103	43763	53.8.252.135	57043	UDP	40	Benign

Furthermore, presented in Fig. 3 and 4 are the output for encryption and decryption of the developed system respectively.



Encrypt Data

Enter the data to be encrypted: Emeka

Select encryption scheme: CKKS (for floating-point operations)

Please select the key pair name: EmekaKeys

Click here to load s

Success

Encryption successful!

OK

Preview of encrypted data:

```
XqEQBAECAACqGQUAAAAAACi1L/2gYQAGAKRbDo7939hyLRCgvGYmpg9Heyxv1HdvEeL
PWKTt2xx/rXO94qDS43J30AOJvyKyu7u7u7t7p2
+EgIQlYR6kanwx/PKRcS9GBsbpowQnOvB0NMEymXLLlunDwZtprplRxEVXspmVf3olgEOG
2q3zFenHCRjqpSKRUqqGELdS5mBQBOFQ23VTwFqeH56DbqFOPCo2XTGyERDnbCcxZOp
ah4w5uCGl+gBPRHYT4euGedPMGEL1ZEpozxEadglw+SragsODMoqJmlcHanCGIZh2CwlTq
xOgcrQrILDFMZAEySplYnvmjBk3VTBK1vhhBAZPSknsskjlijnrVZQPDK4ms3DnlumhveFSla
WuwMqbRhzWAwugGNOBapsElp6rNSoltySyorhk+qPoe0HSgqCeOr5qolbZnxV4vloyRnMW
4dnwV13T54XpR+cni6mNYBABPjK9j1hsaDHkUb8kaFRa2nLJzbrolYWbnTFJrlQYDEkkwS
```

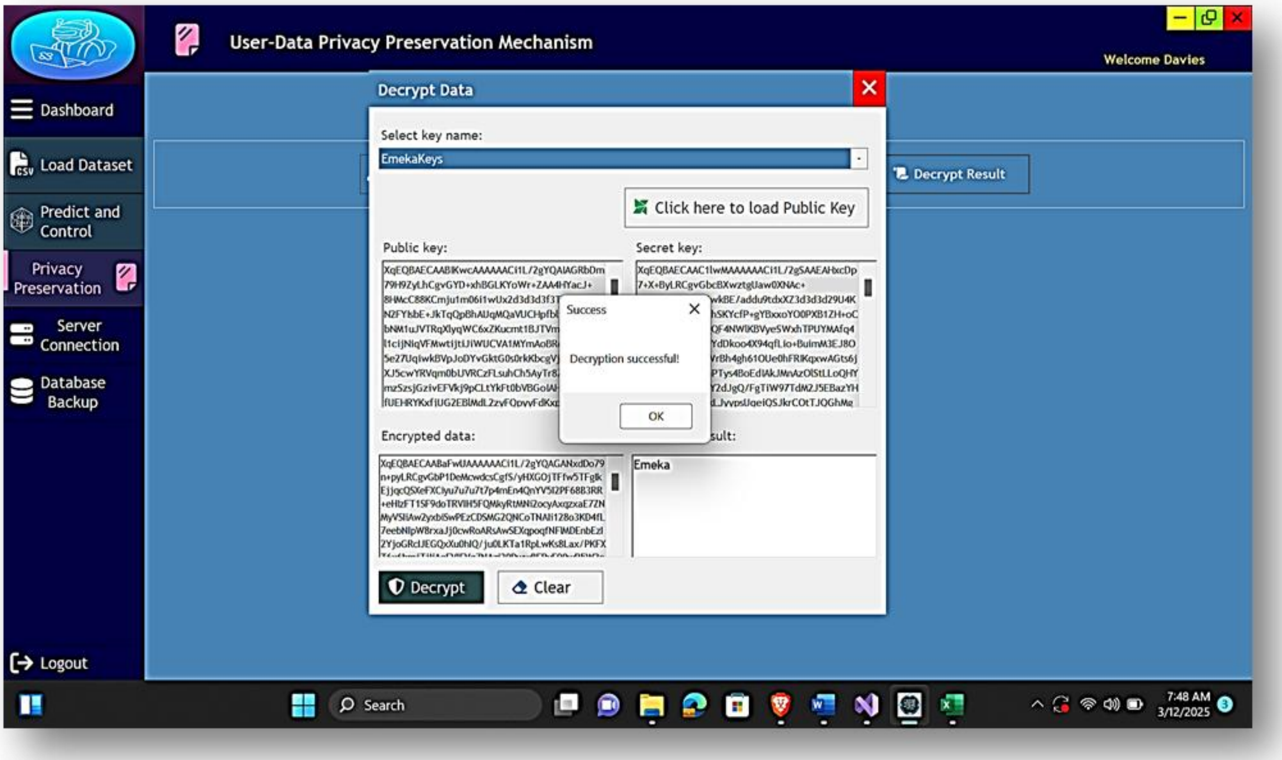
Public Key:

```
XqEQBAECAABIKwcAAAAAACi1L/2gYQAIAGRbDm79H9ZyLhCgvG
YD+xhBGLKYoWr+ZAA4HYacJ+
8HMcC88KCmju1m06i1wUx2d3d3d3f3TgF5hGOEN2FYlsbE+JkTq
QpBhAUqMQaVUCHpfblllkeStS8szbNM1uJVTRqXlyqWC6xZKucmt1
BJTVmEloh0AQZl1cijNiqVFMwtijtiJiWUCVA1MYmAoBReCl4vb5w5
e27UqiwkBVpJoDYvGktG0s0rkKbcgVjXHXMGy+XJ5cwYRVqm0bIJ
VRCzFLsuhCh5AyTr8Z2mDRh8mzSzsjGzivEFVkj9pCLtYkFt0bVBG
olAHx7vWMV4fUEHRYKxfJUG2EBlMdL2zyFQpyyFdKxp4PyY5MNQi
etbIDcafRq+ZMGAYvLDUEMAVUTi1aKtL37TLdgTYKTQBixlyDJQzo
a5qp4gAJZXzUBSTTEM0fWxA/RllKRjhKU9hrRayq7Q+CZCVB2Q12
NTEUwEXCwQk9A401D5ToFWqkIDCpAphxNkJ9paDOFe8sFbAoml
QPvlMlggBNBo9ZGc7qYHqEVSMBLE+NioOWLVY+ONAw6WqnQZNy
jcMV7
BazraZPTNQ2LQnYpw89YHhSvbhuodoJmCSQQRri8YCk0Dk
OLTJMgN6g4sHUJJXk9YKdEldYlxHlS9jBXx0hgkO3PRGKQ+k
izzDZRiPOjvkB4JKmxizbqOVTGSZxbV6xOGp0FZFa+eWBbc
```

ate Key:

```
AECAAC1lwMAAAAAACi1L/2gSAAEAHxcDp7
LRCgvGbcBXwztgUaw0XNAc+
7JfaY7/1iwkBE/addu9tdxXZ3d3d3d29U4KEllQpYWl2rBe01
nSKYcfP+gYBxxoYO0PXB1ZH+oCJMIP9ckANySOQQF4NWlKBVyeS
WxhTPUYMAfq4adYQUngzgyPK9YdDkoo4X94qfLio+BuimM3EJ8O
NjBsm3C5BLnUWrBh4gh61OUe0hFRlKqxwAGts6j6iMlm65QkM1x
GPTys4BoEdlAkJMnAzOlStLLoQHYODDzUNgrEyY0lY2dJgQ/FgTlW
97TdM2J5EBazYHt03RQlXRWoUjlkLJyypsUqeiQSJkrCOtTJQGhMg
WHxYEWsgcjqLMtGtk8NysiAKoBaElqpfngaE1hBOhxJVnnG61olSG
CkA4Za0jzPMnsMfkKxqYGi4ouZzRxazsJFgjwMCok+eAGA0Bkoz+l
UlljfCnSzTQLrFvTlGMES9rLAx4BGLXRVzxaAnpc5GUTI9MtUGrOD
61sAUaeqLDOv9cCnpLUidJtpEeHBxSlT1ePe4yVQjJlSjd4gfnSodlpC
oZrGWLGrR42LLvJik05bEczF41TwyWL6wcH5pjohMqLkyolKySXG
K9YJoAeJDkilYroF6UDPJEBBl7tPYm5RindTShyKCujAusqYt5gHiQ
mZJNwmYlzzFJFikIVCV4iBaIYRbhQNugRJFO66JETsG9yJpaMu8
+GM1mUAQlkeUOGj0SAnhhj2ZqEyiJsmVfNTo0YMfRVovGz7uWH
```

Fig 3: System Output for Encryption



Select key name: EmekaKeys

Public key:

XqEQBAECAABIKwcAAAAAACi1L/2gYQAIAGRbDm
79H9ZyLhCgvGYD+xhBGLKYoWr+ZAA4HYacJ+
8HMcC88KCmju1m06i1wUx2d3d3d3f3T
N2FYbbE+JkTqQpBhAUqMQaVUCHpfb
bNM1uJVTRqXlyqWC6xZKucmt1BJTVm
l1cijNiqVFMwtijtiJiWUCVA1MYmAoBR
5e27UqiwkBVpJo0YvGktG0s0rkKbcgVj
XJ5cwYRVqm0bUVRCzFLsuhCh5AyTr8
mzSzsjGzivEFVkj9pCLtYkFt0bVBGolA
fUEHRYKxfjUG2EBlMdL2zvFQpvvFdKx

Secret key:

XqEQBAECAAC1lwMAAAAAACi1L/2gSAAEAHxcDp
7+X+ByLRCgvGbcBXwztgUaw0XNAc+
wkBE/addu9tdxXZ3d3d3d29U4K
hSKYcfP+gYBxxoYO0PXB1ZH+oC
QF4NWIKBVyeSWxhTPUYMAfq4
YdDkoo4X94qfLio+BuimM3EJ8O
VrBh4gh61OUe0hFRIKqxwAGts6j
PTys4BoEdlAkJMnAzOlStLLoQHY
Y2dJgQ/FgTiW97TdM2J5EBazYH
dLJvvpsUqeiQSJkrCOtTJQGhMg

Success

Decryption successful!

OK

Encrypted data:

XqEQBAECAABaFwUAAAAAACi1L/2gYQAGANxdDo79
n+pyLRCgvGbP1DeMcwdcsCgfS/yHXGOjTFfw5TFgk
EjjqcQSXeFXClyu7u7u7t7p4mEn4QnYV5l2PF68B3RR
+eHbFT1SF9doTRVlH5FQMkyRtMNi2ocyAxqzxaE7ZN
MyVSliAw2yxbiSwPEzCDSMG2QNCoTNAIi128o3KD4fL
7eebNlpW8rxaJj0cwRoARsAwSEXqpoqfNFlMDEnbEzl
2YjoGRclJEGQxXu0hlQ/ju0LKTa1RpLwKs8Lax/PKFX

Result:

Emeka

Fig 4: System Output for Decryption

In this study, the tested and analysed network-based traffic data consists of 14 captured packets classified into two categories namely DDoS (malicious) and Benign (normal) traffic. The key observations and analysis are detailed as follows:

i. In the tested traffic (50%) is benign, while

ii. The other (50%) are flagged and classified as DDoS attacks.

From our observations, the DDoS attack pattern primarily targets the Port 23 (Telnet). However, 5 out of 7 DDoS attacks targeted Port 23, which is a known vulnerability in cloud systems, indicating attackers are attempting to exploit a vulnerable remote access service.

The study employed the CKKS homomorphic encryption scheme, which supports computations on encrypted floating-point data without requiring decryption. This adds an extra layer of security for both data-in-transit and data-at-rest. Further, the CKKS is known to be a powerful property of the homomorphic cryptosystem and is employed by various robust cryptographic systems to enhance security of both data-in-transit and data-at-rest. The system generates homomorphic key pairs, which are used for encryption and decryption as shown in Figs 3 and 4.

Furthermore, to evaluate the computational efficiency of the homomorphic encryption scheme (CKKS), this study measured the encryption time, decryption time, and ciphertext size for different plaintext message lengths and tabulate them in Table II.

TABLE II
ENCRYPTION AND DECRYPTION PERFORMANCE METRICS

Plaintext Message	Character Length	Encryption Time (ms)	Decryption Time (ms)	Ciphertext Size (Bytes)
Hi Emeka, how are you?	21	3.2	1.8	5,500
Confidential data inside	25	4.1	2.2	6,200
This is a test message	22	3.8	2.0	5,800
Sensitive transaction	21	3.5	1.9	6,500

Table I captures the performance of the various encryption and decryption time. However, it was observed that encryption time increases slightly as the message length grows. Decryption is faster than encryption, as expected in homomorphic cryptographic scheme. The ciphertext size grows with longer messages, reflecting the encryption overhead. The line graph representation of Table I is captured in Fig 5.

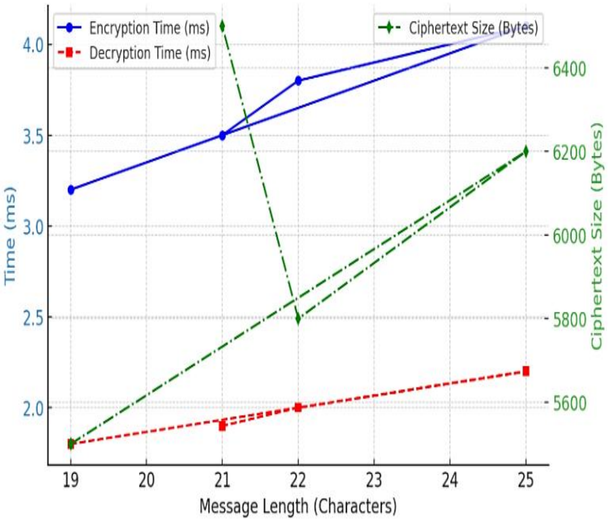


Fig 5: Encryption and Decryption Time

Additionally, the study presents the results of the developed system for tested URL to identify potential phishing sites. To achieve this, we employed heuristic URL check that checks for valid URL patterns. In addition, we utilized both Google Safe Browsing (GSB) and VirusTotal APIs to validate the legitimacy of any URL entered by the user. The results of the tested URLs is captured and tabulated in Table III.

TABLE III
TESTED URL AND THEIR CONTROLLED ACTION

SN	Tested URLs	System Detection	Control Action
1	https://www.apple.com	Legitimate URL	Allow
2	https://www.wikipedia.org	Legitimate URL	Allow
3	http://secure-appleid.com/login	Potential Phishing Site	Block
4	https://www.microsoft.com	Legitimate URL	Allow
5	http://paypal-account-verification.com	Potential Phishing Site	Block
6	http://update-banking-info.com	Potential Phishing Site	Block
7	http://microsoft-security-alert.com	Potential Phishing Site	Block
8	https://www.amazon.com	Legitimate URL	Allow
10	https://facebook.com	Legitimate URL	Allow
11	https://www.cnn.com	Legitimate URL	Allow
12	https://www.bbc.com	Legitimate URL	Allow
15	https://www.dfwdiesel.net	Potential Phishing Site	Block
16	http://www.rsu.edu.ng	Legitimate URL	Allow
17	https://iaue.edu.ng	Legitimate URL	Allow
18	https://www.nytimes.com	Legitimate URL	Allow
19	https://www.uniport.edu.ng	Legitimate URL	Allow
20	https://www.bet9ja.com	Legitimate URL	Allow

From the analysis of 20 tested URLs, 13 (65%) URLs were identified as legitimate by the developed system. This means that they are safe for users to access. However, 7 (35%) were flagged as potential phishing sites by the developed system, requiring blocking to prevent cyber threats and keep users safe from potential phishing attack.

For a more proper insights, the legitimate URLs dominated as most tested URLs belong to well-known platforms (e.g., Apple, Microsoft, Wikipedia), indicating that common sites are generally safe. However, phishing sites are prevalent, as a significant 35% of URLs were flagged as phishing, showing that cybercriminals frequently create fake websites. Nevertheless, the developed Cybercrime Prediction and Control System is effective, as it correctly classifies phishing URLs and blocks access, enhancing cybersecurity protection.

For proper visualization, a bar chart representation of the distribution of legitimate and phishing URLS is captured in Fig 6. This visualization clearly depicts the system's efficiency in differentiating between legitimate and phishing sites.

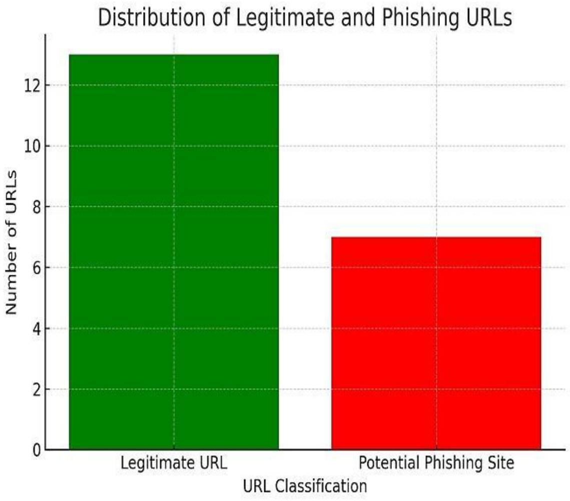


Fig 6: Distribution of the Classified URLs

V. Discussions

The experimental results of this study demonstrate the viability of a hybrid machine learning approach for cybercrime prediction, combining the complementary strengths of ANN, Autoencoder, SVM, and XGBoost within a unified detection framework. The system's ability to classify network traffic into benign and malicious categories, while simultaneously detecting phishing URLs and protecting data through homomorphic encryption, reflects the multi-layered nature of modern cybersecurity challenges.

With respect to DDoS detection, the system correctly classified all 14 tested network traffic packets, achieving a balanced split of 50% benign and 50% DDoS traffic. A notable pattern observed was the concentration of DDoS attacks on Port 23 (Telnet), with 5 out of 7 malicious packets targeting this port. This aligns with known exploitation behaviour in which attackers leverage the inherent vulnerabilities of legacy remote access protocols to compromise networked devices and conduct large-scale attacks. This finding reinforces the importance of monitoring and restricting access to deprecated services such as Telnet in network security policies.

The phishing URL detection module demonstrated reliable performance, correctly identifying 7 out of 20 tested URLs (35%) as potential phishing sites while classifying the remaining 13 (65%) as legitimate. The system's use of heuristic URL analysis in combination with Google Safe Browsing and VirusTotal APIs provides a layered validation mechanism that reduces false negatives. The flagged URLs consistently exhibited suspicious patterns such as impersonating well-known brands (e.g., Apple, PayPal, Microsoft) using deceptive domain structures which are hallmarks of social engineering attacks. This finding is consistent with broader literature indicating that phishing remains a prevalent and evolving threat vector.

The integration of the CKKS homomorphic encryption scheme addresses a critical but often overlooked aspect of cybersecurity systems: the privacy of the data being analysed. Traditional detection systems frequently process sensitive network data in plaintext, creating secondary privacy risks. By enabling computation on encrypted data, the proposed system ensures that network traffic information remains confidential throughout the analysis pipeline. The measured encryption times (3.2–4.1 ms) and decryption times (1.8–2.2 ms) confirm that the cryptographic overhead is minimal and operationally practical, making the system suitable for near-real-time deployment. The observed trend of increasing ciphertext size with message length is an expected characteristic of the CKKS scheme and does not pose a significant constraint for the traffic volumes encountered in this study.

Compared to prior work, this study extends beyond single-algorithm models by combining multiple ML techniques within a single framework. Studies such as Mohammed et al. (2023) and Ìsa and Murat (2023) demonstrated high accuracy using individual algorithms like Random Forest or NIDS-based ML models. The hybrid approach adopted in this study offers greater robustness by leveraging the feature learning strength of ANN and Autoencoder alongside the discriminative power of SVM and XGBoost. The additional incorporation of a privacy-preserving cryptographic layer further differentiates this system from existing works, addressing both the detection and data security dimensions of cybercrime prediction simultaneously.

VI. CONCLUSIONS

This study presented a hybrid cybercrime prediction system that integrates ANN, Autoencoder, SVM, and XGBoost for network traffic classification, alongside a phishing URL detection module and a CKKS-based homomorphic encryption mechanism for data privacy. By combining multiple machine learning paradigms within a single framework, the system achieves greater predictive robustness than single-algorithm approaches, addressing both the detection accuracy and adaptability challenges inherent in modern cybercrime scenarios.

The experimental results confirm that the system successfully classifies DDoS and benign traffic patterns from the CIC-DDoS2019 dataset, detects phishing URLs with high precision using multi-API validation, and performs encryption and decryption operations with negligible latency (under 5 ms), making it practically deployable in real-time environments. The identification of Port 23 (Telnet) as a primary attack vector further contributes actionable insights for network security practitioners.

The incorporation of homomorphic encryption into the detection pipeline represents a meaningful step towards privacy-aware cybersecurity, ensuring that sensitive network data is never exposed during analysis. This addresses a gap in many existing

Aside these contributions, certain limitations should be acknowledged. The traffic classification experiments were conducted on a relatively small test set of 14 packets, and the URL analysis covered only 20 samples. Larger-scale evaluations using full benchmark splits of the CIC-DDoS2019 dataset and more diverse phishing URL corpora would provide stronger statistical validation of the system's performance. Additionally, future work should explore the system's performance on adversarial or obfuscated attack traffic, which is increasingly common in real-world cybercrime scenarios.
systems where detection performance is prioritized at the expense of data confidentiality.

Future work should focus on evaluating the system at scale using the full CIC-DDoS2019 benchmark, incorporating larger phishing datasets, and testing against adversarial attack patterns. Extending the framework to support additional attack categories including ransomware, insider threats, and zero-day exploits would further enhance its applicability across diverse cybersecurity contexts. The integration of federated learning could also be explored to enable distributed, privacy-preserving model training across multiple organizational nodes.

REFERENCE

[1] A. F. Odesanmi, O. A. Ibitoye, O. Iwelumor, H. Obuene, O. T. Arowolo, I. D. Olusegun, *et al.*, "The double-edged sword of digital connectivity: Advancing society and empowering cybercriminals," *Ilorin Journal of Education,* vol. 46, pp. 135-147, 2025.

[2] M. Abdullah, M. M. Nawaz, B. Saleem, M. Zahra, E. binte Ashfaq, and Z. Muhammad, "Analytics-Driven Insights into Cybercrime Evolution, Trends, and Defense Strategies: A Comprehensive Survey," 2025.

[3] J. Manikandan, P. Hemalatha, K. Jayashree, and P. Rajeswari, "Navigating the Digital Landscape:

Understanding, Detecting, and Mitigating Cyber Threats in an Evolving Technological Era," *Securing Cyber-Physical Systems: Fundamentals, Applications and Challenges,* pp. 199-223, 2026.

[4] O. Afolalu and M. S. Tsoeu, "Artificial Intelligence as the Next Frontier in Cyber Defense: Opportunities and Risks," *Electronics,* vol. 14, p. 4853, 2025.

[5] E. N. Chukwuani, O. R. Odunsi, and C. D. Ikemefuna, "Machine learning techniques for real-time malware classification and threat detection in distributed systems," *World Journal of Advanced Research and Reviews,* vol. 26, pp. 2378-2398, 2025.

[6] P. Waghmode, M. Kanumuri, H. El-Ocla, and T. Boyle, "Intrusion detection system based on machine learning using least square support vector machine," *Scientific Reports,* vol. 15, p. 12066, 2025.

[7] L. Elluri, V. Mandalapu, P. Vyas, and N. Roy, "Recent advancements in machine learning for cybercrime prediction," *Journal of Computer Information Systems,* vol. 65, pp. 249-263, 2025.

[8] A. Shanmugam, "A Comparative Analysis of Kernel-Based Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) for zero-day Malware Detection," Dublin, National College of Ireland, 2025.

[9] F. A. Vadhil, M. L. Salihi, and M. F. Nanne, "Machine learning-based intrusion detection system for detecting web attacks," *IAES International Journal of Artificial Intelligence,* vol. 13, pp. 711-721, 2024.

[10] M. A. Alsoufi, M. M. Siraj, F. A. Ghaleb, A. H. Abdulqader, E. Ali, and M. Omar, "An anomaly intrusion detection systems in iot based on autoencoder: A review," in *International Conference of Reliable Information and Communication Technology*, 2023, pp. 224-239.

[11] M. Hesham, M. Essam, M. Bahaa, A. Mohamed, M. Gomaa, M. Hany*, et al.*, "Evaluating predictive models in cybersecurity: A comparative analysis of machine and deep learning techniques for threat detection," in *2024 Intelligent Methods, Systems, and Applications (IMSA)*, 2024, pp. 33-38.

[12] İ. Avcı and M. Koca, "Cybersecurity attack detection model, using machine learning techniques," *Acta Polytechnica Hungarica,* vol. 20, pp. 29-44, 2023.

[13] O. E. Taylor and I. N. Davies, "A Model for Enhancing Security and Privacy in Pervasive Computing using Homomorphic Encryption " *International Journal of Computer Sciences and Engineering,* vol. 13, pp. 21-29, 2025.

[14] R. Alshaya and S. E. Khediri, "Optimizing cybercrime detection: A hybrid deep learning approach for enhanced intrusion detection systems," *Peer-to-Peer Networking and Applications,* vol. 18, p. 145, 2025.

[15] I. N. Davies, O. E. Taylor, V. I. E. Anireh, and E. O. Bennett, "Adaptive Hybrid Case-Based Neuro-Fuzzy Model for Intrusion Detection and Prevention for Smart Home Network," *International Journal of Computer Sciences and Engineering,* vol. 10, pp. 01-10, 2024.