Journal of Theoretical and Applied Information Technology
15th April 2025. Vol.103. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STRENGTHENING SECURITY PROTOCOL TO COMBAT FINANCIAL FRAUD ADVANCES IN AUTHENTICATION AND ACCESS CONTROL

JANJHYAM VENKATA NAGA RAMESH [1 A,B], DR.M.V.RAJESH [2], B.VEERAJYOTHI [3], VEMULA JASMINE SOWMYA [4], AMIT VERMA [5],REFKA GHODHBANI [6*]

[1a] Adjunct Professor, Department of CSE, Graphic Era Hill University, Dehradun,248002, India.
[1b] Adjunct Professor , Department of CSE, Graphic Era Deemed To Be University , Dehradun,248002, Uttarakhand, India.
[2] Professor, Department of IT, Aditya University, Surampalem, India,
[3] Associate Professor, Department of Information Technology, Chaitanya Bharathi institute of Technology, Hyderabad, India.
[4] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh,India.
[5] University Centre for Research and Development, Chandigarh University, Gharuan Mohali, Punjab, India.
[6] Center for Scientific Research and Entrepreneurship, Northern Border University, Arar, Saudi Arabia.

E-mail: [1] jvnramesh@gmail.com,[2] Rajesh.masina@adityauniversity.in,[3] veerajyothi_it@cbit.ac.in, [4] vemulajasmine@gmail.com, [5] amit.e9679@cumail.in, [6] Refka.Ghodhbani@nbu.edu.sa

ABSTRACT

This research delves into the potential of combining machine learning with multi-factor authentication (MFA) to strengthen financial system security protocols. The main objective is to fight fraud using improved authentication and access control methods. The study looks at practical implementation techniques to fix current security flaws, such as issues with centralised authentication systems that are vulnerable to hacks and system malfunctions. The study highlights the value of energy-efficient and ecologically friendly security measures in addition to traditional encryption and biometric techniques. Different architectures, consensus protocols, applications, services, and implementation goals are all covered in the analysis of various security frameworks. Six machine learning models are thoroughly examined to see how well they identify fraudulent activity and improve overall money security. The results demonstrate how ML and MFA together may greatly enhance network intrusion detection and fraud prevention, eventually improving the security of financial transactions.

Keywords: *Machine Learning, Network Intrusion Detection, Multi-Factor Authentication Security Frameworks*

1. INTRODUCTION

New cybersecurity dangers and attack tactics have challenged traditional security systems [1]. In this research, researchers combine MFA with machine learning (ML) to overcome these crucial challenges. Analyse expert research to determine how these technologies may improve network security [2].

Modern cybersecurity is continually changing, and hackers are fast to adapt [3-5]. They use sophisticated hacking and cyber threats to obtain illegal access. Cybercriminals continue to evade security measures, even if password-based authentication has proven less successful [6-8]. Researchers must innovate to enhance system protections [9].

This research introduces ML-driven methodologies, MFA, and network intrusion detection systems to provide a more dynamic and responsive security architecture [10-15]. This study addresses a major social issue, not only technology. Cyberattacks may damage sensitive data, business assets, and key infrastructure, threatening organisations and nations [16-17]. These occurrences demonstrate the essential need for cyber threat-adaptive security systems.

Cybersecurity's biggest issue is creating a real-time, adaptive security solution that can combat varied cyber-attacks. Traditional security measures are meant to resist emerging threats, but adversaries typically develop more complex attack methods quicker than security systems can react, creating exploitable flaws [19]. ML, MFA, and network intrusion detection systems reduce cyber threats by using data-driven security advancements [20-23]. This study examines how these integrated security components work together to strengthen system resilience against unwanted access, data breaches, and cyberattacks.

Password-based authentication systems have been vulnerable to security breaches and crypto-attacks [24]. Modern digital environments, including IoT devices and smart networks, need a comprehensive, adaptable, and intelligent network security strategy. Despite cybersecurity advances, loopholes enable attackers to exploit flaws quicker than conventional security methods [25]. Few studies have examined how ML and MFA might improve network security together. This research tackles this gap by integrating ML and MFA into a cybersecurity architecture. Researchers study how various technologies can collaborate to defend against cyberattacks. The research also assesses this approach's real-world efficacy, emphasizing its potential advantages.

Structure of the rest of the paper: Section 2 covers methodology, Section 3 model execution and assessment, Section 4 experimental results, Section 5 findings, Section 6 real-world implications, and Section 7 research contributions and practical applications

A. Problem statement

Despite cybersecurity advancements, gaps remain in tackling the growing threat scenario. Traditional security solutions typically lag. Due to the fast growth of cyber threats, attackers may swiftly exploit vulnerabilities. Even with a great deal of study on multi-factor authentication and machine learning, there is still a clear knowledge gap. Comprehensive studies incorporate technologies to increase network security via synergy. This research addresses problems by integrating machine learning and multi-factor authentication into a unified network security architecture. By investigating how various technologies could cooperate to offer efficient cyber threat protection, the research seeks to close the gap. The study evaluates this integrated approach's performance in actual scenarios, demonstrating its applicability and possible advantages.

2. METHODOLOGY:

Recent advancements in AI, including ML as well as DL, have significantly enhanced automated intrusion detection. Network security is becoming more challenging due to the complexity of attackers. In addition, developing security components has always been difficult due to data volume. To get around these issues, this study makes use of novel technologies such as confusion matrices for model evaluation and Synthetic Minority Over-sampling Technique [SMOTE] for imbalanced datasets. To secure sensitive digital assets, researchers employ XGBoost to achieve successful dimension reduction and stringent performance evaluation criteria to evaluate the dependability of AI-driven automation model.

2.1 Networks Privacy via Integration Driven by ML:

To get the desired result, the hybrid technique takes into account all aspects of the study problem. Among the many important data treatment techniques used by this blend plan are label encoding, SMOTE data balancing, extensive feature scaling, and suitable data pre-processing processes for handling missing values. The next step is to use the feature selection approach's output vector as input to various ML and DL algorithms that rely on the tree-boosting technique as their primary decision-support methodology. Researcher looks over ML/DL techniques, like RF, DT, and KNN, along with the way they relate to CNN, ANN, and MLP approaches. Researchers have developed a method that can reliably detect network breaches.

Based on the distinctive idea of securing networks through Machine Learning-Driven Integration, this part gives a short outline of the main ideas and methods that make it work.

2.1.1 Sources for the dataset:

Several datasets are used in the study for

the network intrusion detection, like the TON_IoT, CIDDS-001, and 5G-NIDD Dataset. These datasets were chosen due to their comprehensiveness and variety in representing various network traffic situations and attack types.

2.1.2 Methods for preprocessing:

Data cleaning, feature scaling, and categorical variable encoding are all part of the preprocessing methods that deal with the data that is lacking and outliers. To fix the problem of unbalanced data and make sure all classes are fairly represented, SMOTE is used to equalize each class in the dataset.

The hybrid approach relies on feature selection to restrict the feature's bandwidth and maintain the maximum degree of effectiveness possible. Feature selection reduces computational cost and improves model performance by identifying a better subset of attributes. Before this, the Extreme Gradient Boosting (XGBoost) technique was used, known for its accuracy in selecting important traits. In XGBoost, a method for structured data, complex methods including Newton's 2nd order gradient curve and lasso (L1) and ridge regression (L2) regularization are used. These factors aid in locating key parameters, perhaps improving generalization and reducing overfitting.

The XGBoost software primarily consists of two parts: creating decision trees and using these trees to make estimations. It has a compounded objective function which comprises a loss function (l), which is a way to measure the difference between expected and actual outcomes, and a regularization function (Ω), which controls how complex the model is and whether it fits properly. For the method to work, iterations must be done so that the goal function's increase sums out to its entire value. The XGBoost usually uses 2nd order approaches to get even better results.

To examine the features and choose the most important ones in the datasets, for effective feature processing, the researcher can use the XGBoost algorithm. The ML-driven integrated security network solution is based on the following table, which will summarize the key points about the fundamental methods and features of network security that are combined with ML technology. Use a horizontal bar graph to display the attributes (Figure 1).

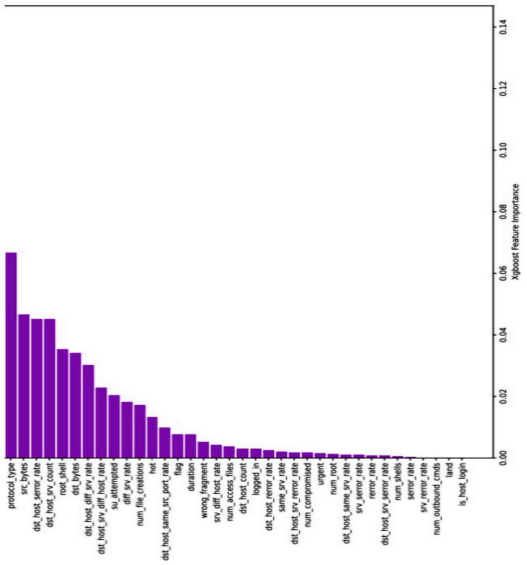


Figure 1: Graph for Feature Significance of XGBoost Algorithm

The network intrusion detection model cannot be improved in terms of efficiency and accuracy without this feature selection technique.

In this work, researchers conducted the machine learning experiments using characteristics that were carefully selected using a rigorous method. The approach made use of the following ML methods such as Multi-Layer Perceptron (MLP), Random Forest (RF), Decision Trees (DT), and k-Nearest Neighbours (KNN). Feature selection in general may be summarized as follows:

First, the qualities were ranked from most important to least important. The first round of choosing attributes was determined on how the XGBoost program ranked the features. A sample of the scored traits was used to test how accurate all machine learning methods were. The goal was to figure out how well the algorithms worked by breaking them down into different sets of features.

Researchers offered a method for selecting algorithms for the candidate feature set that achieve accuracy levels higher than a certain threshold (T_s). At first, k was equal to the whole set of traits (N). Then, the researcher gradually decreased k by two until it was higher than zero. As the feature dataset got smaller, this led to carefully testing how well the methods worked. It could be difficult to use k=1 for a data set with 41 characteristics since

it might result in a time complexity of T(C). A temporal complexity of T(C/10) would be the outcome of setting k=10, however, it may not be all that's needed to find the key performance aspects. Researcher pick for k=2 because it allows for a more comprehensive examination of feature subsets and reduces the temporal complexity into T(C/2).

A collection of attributes may only be considered a candidate set if their accuracy is at least 99.96%, which is higher than the previously defined threshold (T_s). The lowest set of features that satisfied the accuracy criterion for each method was used to choose the final feature collection. The researcher made sure to find the smallest set of features that would still give the required accuracy.

The full feature selection procedure is shown graphically in Figure 1. The feature dataset having a value of k equivalent to N, where N is the total number of features, was then subjected to the XGBoost algorithm. Researchers then used ML classifiers, keeping a careful eye on each one's accuracy levels. The relevant characteristics were found and recorded as an ideal feature set, and just those classifiers that satisfied the accuracy requirement were included. In the absence of such circumstance, researchers calculated a smaller set of attributes that nevertheless met the accuracy requirements by dividing the value for k by two.

2.2. Proposed Experimental Architecture:

By using ML technologies & conducting a thorough multi-factor authentication test, the integrated architecture described in the research study provides a method to improve network security. A methodical block diagram outlining five consecutive processes in this innovative approach is shown in Figure 2.

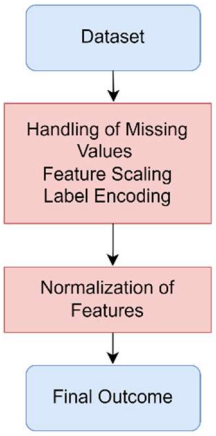


Figure 2 Framework for Preprocessing Data

1) Preprocessing of Data:

Topics like intensive data preparation are part of this critical primary phase. Methods include completing blanks, standardizing scales for attributes, and converting categorical data to a form suitable for modelling. All subsequent follow-up research is based on these particular data changes, which provide an initial platform for further exploitation.

2) Equilibrating Data Utilizing SMOTE:

The primary goal of SMOTE is to balance the various training data classes used in ML. Recognizing the importance of the data balance. Getting the datasets for equitable representation is the focus of this phase. SMOTE is intelligently carried out to assist in restoring the devices to the dataset if it indicates where there is an inconsistency among the data. Therefore, this approach may handle the previously described problem of data inconsistency and seeks to increase the reliability of subsequent studies.

3) Choice of Features with XGBoost:

In this step, an important part of the design is introduced, and the algorithm known as XGBoost is used smartly. The aim is to identify and keep the most important traits in the data set (Table 1). This step improves the model's ability to determine the difference between things while also lowering the number of dimensions. It does this by getting rid of features that don't have strong relationships with the

class names.

Table 1 Feature Selection

Features	F1	F2	F3	Fn
R1	0	0	1	Sn
R2	0	0	....	
R3				

4) Separating Data with K-Fold Cross-Validation:

At this stage, the design begins with the model assessment. In this step, researchers use the tried-and-true K-fold cross-validation method to carefully split the pre-processed datasets in two parts: the training set and the testing set. This method encourages generalization while also guaranteeing accurate model evaluation.

5) Assessment of Performance:

In this stage, ML techniques are used to bring the suggested architecture to a close. In this case, algorithms undergo extensive testing and intensive training. Several important measures, including recall, F1-score, efficiency and precision, are used to thoroughly evaluate performance. For network intrusion detection, the model with the best performance is suggested (Figure 3). The effectiveness of this model is then determined by a thorough comparison with other models.

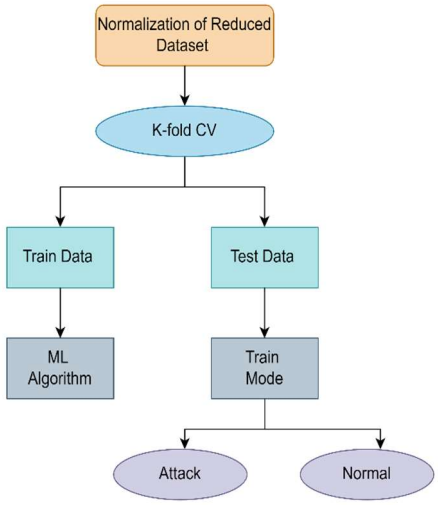


Figure 3 Block Diagram for Feature Selection and Intrusion Detection.

3. MODEL EXECUTION AND ASSESSMENT

A new combined method to make computer networks safer is presented in this study. This approach combines the SMOTE with the XGBoost algorithm for successful feature selection to fix problems with uneven data. A lot of different ML & DL methods are used to find the most robust model. Through many tests on different datasets, this technique has been carefully verified and indicated to be of adequate quality. Finally, the researcher gives a full explanation for the dataset descriptions and subsequently explains the way to prepare the data and train the model.

3.1. ML Model Training and Testing Data Sets:

Comparing the success of ML models and figuring out how well they work in real life is made easier by the collected data. Therefore, it provides the foundations for developing, trying, and confirming these methods. The database contains many types of incidents that are put on, like mimic real-life situations. Every IP protocol message in this location is collected for research. This dataset includes statistics on both normal and attack connections, which can represent everyday connections or specific forms of assault. Labeling enables learners to receive guided learning. Machine learning models will learn causes and nodule type/abnormality discrimination. Lack of consistency in conduct reflects invasion accounts. The dataset contains sufficient information for model training and testing, with 42 numerical and qualitative metrics for each unique link.

As the foundation for training and evaluating the ML models for detection of network intrusions, researchers carefully gathered and prepared the dataset for the study.

Cleaning the Dataset: After repairing errors and missing values, researchers cleaned the dataset thoroughly and used it for trustworthy data analysis.

Balance techniques: Researchers used ingenious filtering techniques that were carefully calculated to ensure that the different classes were shown equally to avoid the over-representation of certain intrusions and not "squeeze" a system in a gap.

Advancements: The method included improvements like categorical variables coding, administration, and disappearing exceeding values, and dataset separation into testing and training subsets, in addition to necessary preprocessing of concepts.

3.2. Collecting Data:

Researchers supplemented the data to eliminate noise as a part of their data cleaning duties. Following this, the researcher made greater use of it while simultaneously selecting a dataset to create a DL-DeMo model. The whole process of figuring out the right amount, along with the feature size for data, is part of data preparation. And finally, one important part of preparing data is working with numbers that are missing. Along with the usual preparation steps, the advanced techniques are used to feature engineering, like principal component analysis and other methods that decrease the number of dimensions, to make the model work better and make it easier to compute.

3.2.1. Managing Missing Values:

Missing data often results in data sets that do not have numbers for occurrences. Since neural networks and ML may be taught with incomplete data and potentially provide incorrect new data, it would be beneficial to provide a data preparation service. Using straightforward and easy-to-understand processes, researchers fill in the dataset's missing values. This can only be accomplished by removing duplicate entries, null rows (RE), negative rows (RN) and identical (I).

3.2.2. Stabilization-Based Feature Scaling:

To standardize the values of features to a consistent range, scaling of features is the initial step. Standardized dataset features provide uniform scaling across measurements. The model accuracy is improved by standardizing feature values. Variations in measurement units may greatly affect model dependability. Standardizing attribute values helps avoid this issue by limiting them to an acceptable range. In standardization, each feature is normalized by subtracting the mean & dividing by its SD. The formula is:

$$X_s = sd(y)y - mean(y)$$

The standardized value is denoted by Xs. y is the attribute's true value. The real value's mean is represented by mean(y). The standard deviation of the real value is represented by sd(y).

3.2.3. Encode of Labels:

Transforming categorical data into numerical values is the process of label encoding. To be utilized as input to the training part of machine learning models, categorical traits must be stored as numeric data. To show each group as a number about zero to n-1, this mapping is used. The numbers between 0-4 are utilized to show the categorical information for all of the five groups. Utilizing the encoded labels for both binary as well as multilabel classification is part of the study of the 5G-NIDD Dataset. Table 2 as well as Table 3 demonstrate the way this process works.

Table 2 Binary Classification.

Types of Attack	Encoding Label
Normal	1
Attack	0

Table 3 Multi-Label Classification.

Types of Attack	Encoding Label
Denial-of-Service	0
Normal	1
Root to Local attacks	3
User to Root attack	4
Probe	2

3.2.4. Process of Training:

The research's important training step involves applying machine learning methods to the pre-processed dataset. This section contains the software and hardware configurations of the system being used for training, as well as the tools & libraries needed for creating the model and preparing the data. Model training times were accelerated by using GPU acceleration where appropriate in a high-performance computer environment. Grid search was used to enhance model performance through hyperparameter adjustment.

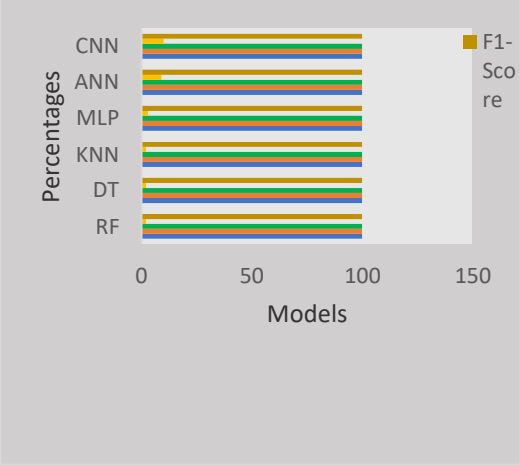
Windows 10 Pro on an HP laptop serves as the foundation for the learning environment. The researcher used Google Colab, which is also the Python 3.8.6 programming language, as our working tool. The Dask & SymPy libraries are used

when working specifically with data processing and manipulation. The people in charge of data visualization, however, rely on Seaborn and Matplotlib. To complete machine learning tasks and basic data analysis, researchers rely on the Pylearn2 package.

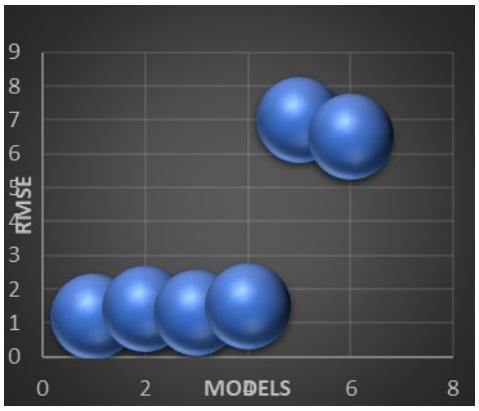
The proposed method of action is evaluated using several metrics, including recall F1-score, receiver operating characteristic curve, RMSE, accuracy and precision. Researchers evaluated the model's performance using a variety of metrics, such as confusion matrices and precision-recall curves, to find out the extent to which it could identify uncommon attack types.

4. EXPERIMENTAL OUTCOMES:

This article is a complete overview of the outcomes obtained by various ML-based models for finding network intrusions. Evaluation criteria include thoroughness, correctness, and usefulness. Figures 4 and 5 show a graph that summarizes the F1-score along with RMSE. The recall, precision, and accuracy levels demonstrated F1-score performance, indicating that the ML models were very reliable. The RF's performance achieved the highest accuracy of 99.98%, while DT closely behind it secured 99.97%. The k-Nearest Neighbors classifier has achieved corresponding accuracy rates of 99.96%. Although ANN performed well in this regard as well, the RMSE is superior to those of other tree-based models.



Figures 4 ML Models' Performance



Figures 5 RMSE for ML Models

The data has been supplemented with bar graphs to demonstrate the progress. The line plots show that computers with different learning models do better than computers without those models. Despite providing much visual guidance to the reader in interpreting the results, the comparison plots make it simple to compare and contrast many models. The study's results show how many machine learning techniques might be useful when creating intrusion detection systems and add to the growing body of information in the area of intrusion detection. These key performance indicators show potential for an autonomous learning model that can safely manage the networks. Next steps. In addition, studies might be conducted to use these methods or make greater use of the features of these models when used together. The technique uses many detection algorithms to improve detection accuracy while keeping costs low.

4.1. Analysis of Result:

The results of a thorough analysis of several approaches for spotting illegal access activities in computer networks are the main topic of this presentation. This research set out to determine the best effective model for network intrusion detection by assessing several performance indicators. All features, chosen features, and the recommended features were included in this study. The results of the analysis show that the recommended feature set performed better than the remaining two sets combined.

4.1.1 Experimental Procedure:

For both static and the multiple classes' identity jobs, the researcher used the intrusion detection

methods that worked well in the tests. A k-fold cross-validation using a k-value of 11 and additional hyperparameters that didn't change the quality of the model were used to evaluate the models' accuracy (Figure 6). The data is divided into 10 smaller sets, with 81% being used for training and 19% for testing.

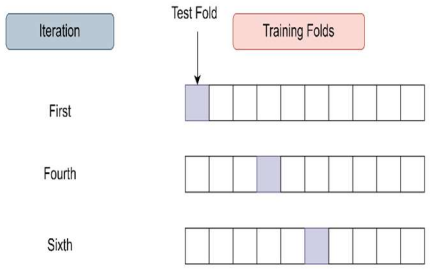


Figure 6 K-Fold Cross Validation.

- **Outcome of Binary Classification:**

This article provides a thorough review of each model's performance across several measures, including precision, recall, F1-score, F1-score, along with RMSE, which is assessed using several machine learning models for network intrusion detection. An impressive 99.98% accuracy rate was shown by RF, indicating that it is very successful in identifying both correct & incorrect instances of network infiltration. With a recall rating of 99.99% and an accuracy rate of 99.97%, the model shows that it can identify a significant number of true invasions with a very small incorrect instances rate. Its excellent performance was further confirmed by the fact that it achieved an F1-Score of 99.99%, which is a measure of accuracy and recall. With a RMSE of only 1.22, the model seems to be making quite accurate predictions.

The DT system achieved 99.97% accuracy. Although somewhat less effective than the RF, this result is still impressive. Its accuracy rate of 99.96% indicates few incorrect positives, while its recall score of 99.98% indicates exceptional accuracy in identifying positive instances. F1-Score of 99.98% shows an effective mix between accuracy and recall, yet it's lower than RF. While the RMSE of 1.35 is within acceptable limits, it indicates a substantially higher prediction error than RF.

An accuracy rate of 99.96% was reached by the KNN model, which means it did quite a job, generally at finding network intrusions. The KNN model has a low incorrect positive rate that keeps it accurate 99.95% of the time. This means that most attacks that are found are real. There is an ideal mix between accuracy and recall, as shown by the memory number of 99.98% and the F1-number of 99.98%. There is a little more forecast error with the RMSE of 1.28 than with the RF, but it is still fairly less.

The MLP models achieved a very effective 99.93% accuracy, even though they performed considerably worse than the other models. Although its incorrect positive rate is significantly higher than the other models, its accuracy rate of 99.92% indicates that positive predictions are very accurate. Although it is not as high as the top models, its recall rate of 99.95% shows that it does an excellent job of detecting positive events. With an F1-Score of 99.94%, the MLP model achieves a satisfactory mix between recall and accuracy, however it falls just short of the best performance. While still rather low, the analyzed models' RMSEs of 1.44 indicate the largest prediction error.

4.2. Precision of Neural Network Models:

The success of two neural network models is also looked at: CNN & ANN. Data in Figure 7 shows that CNN is better than ANN at binary classification, with a 0.03% improvement. Additionally, we recommended the effective utilization of both neural network models, CNN & ANN, during the decision-making processes for multilabel classification.

Journal of Theoretical and Applied Information Technology
15th April 2025. Vol.103. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

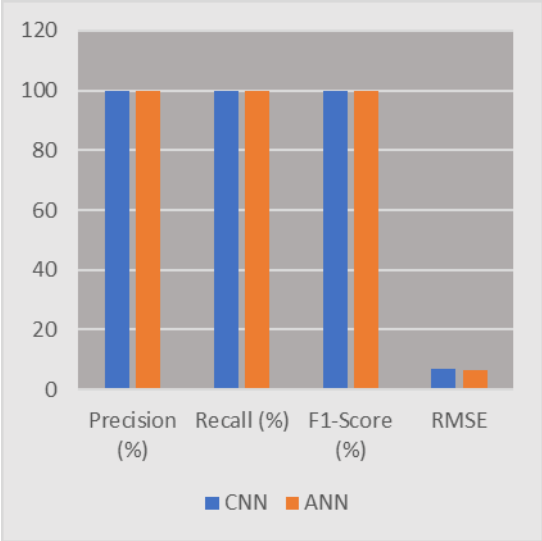


Figure 7 Neural Network Model Performance Analysis.

- **Confusion Matrix:**

Figure 8 shows the confusion matrices, which show the extent to which the model worked. If it comes to intrusion detection, RF & DT are more successful since they generate more Correct Positives (CP) & Negatives (CN) with less Incorrect Positives (IP) & Incorrect Negatives (IN).

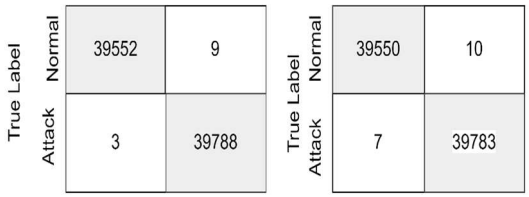


Figure 8 Binary Classification Confusion Matrices.

As the bar graph in Figure 9 illustrates, the suggested approach achieves far higher accuracy in multilabel categorization. The RMSE rates are considerably lower than when employing specific characteristics, and the accuracy rates increase significantly in all circumstances.

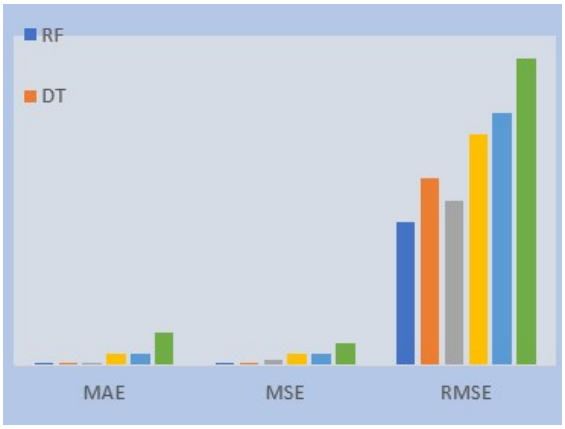


Figure 9 Analysis of Binary Classification Performance

4.3. Comparison of Model Efficiency:

Figures 10 through 12, the researcher provided the full picture of the four models work for finding network intrusions.

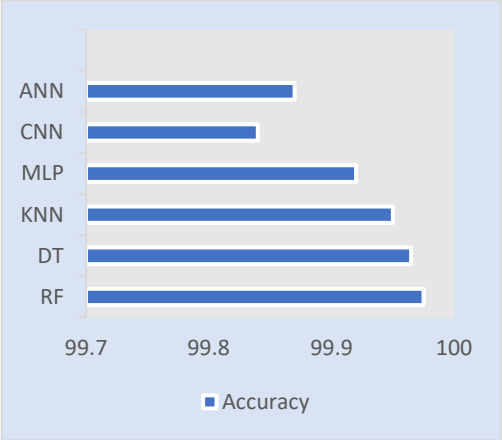


Figure 10 ML Models Accuracy Comparison

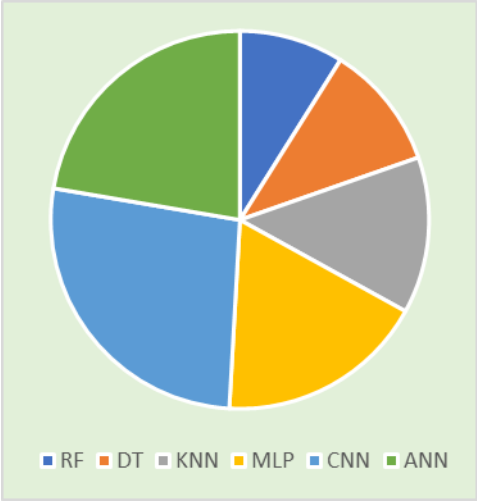


Figure 11 ML Models Training Speed Comparison

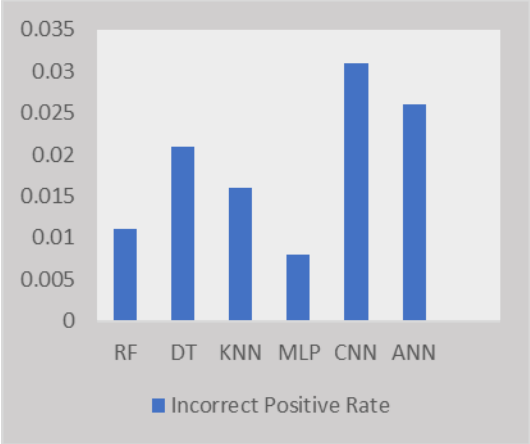


Figure 12 Incorrect Positive Rates Comparison for ML Models

The F1-score, accuracy, precision, recall & RMSE were used to assess these models. In addition to the outcomes of other ML models, the CNN and ANN findings provide further details on how well these neural network designs perform for network intrusion detection.

4.4. Real Time Application:

The suggested ways to use ML to find network intrusions. Adoption of multi-factor identification has been positive. Including testing in several real-life situations. Higher protection is ready to be added to web sites, apps, and browser extensions using these technologies.

Several well-known web servers have strengthened their security architecture by using the intrusion detection systems powered by ML in conjunction with MFA. For example, the technology was used by a major e-commerce platform to protect customer data and stop unauthorized people from getting into their systems. The platform increased its protection against cyber-attacks and saw a significant decrease in security breaches after integrating strong ML algorithms with MFA mechanisms.

Many banks have started using the breach monitoring system with multi-factor login to keep internet operations safe. Assistance in keeping fake behavior out of client accounts. As a result, these financial institutions saw an enormous decrease in transactions that were not legitimate along with a greater trust from customers in their online banking services.

Users now have an extra layer of security while interacting with websites due to the proposed technologies that have been integrated into web browsers. To detect and prevent malicious websites from being accessible in real-time, a prominent online browser integrated the intrusion detection technology that is based on ML. Through the implementation of methods that use several forms of authentication. The browser ensures that users have safe and secure browsing experiences by protecting against malware-infested phishing attacks.

This research stands out from the others because it uses multi-factor authentication (MFA) to improve network intrusion detection and uses sophisticated machine learning algorithms, such as XGBoost for feature selection and SMOTE for data balancing. While many studies examine ML or MFA alone, this study fills a significant gap in adaptive cybersecurity solutions by combining the two in a novel way. By minimizing data imbalance and improving feature selection, this study exhibits higher accuracy and efficiency than studies that just use conventional ML methods. For example. Data cleaning responsibilities included researchers supplementing the data to remove noise. After that, the researcher started using it more often while also choosing a dataset to build a DL-DeMo model. Additionally, a lot of research focuses on certain attack types or datasets; this study evaluates the models using a variety of datasets, which increases applicability. A more comprehensive knowledge of

the efficacy of several machine learning models, such as CNN and ANN, in network intrusion detection is made possible by the inclusion of a thorough performance comparison across these models. Beyond merely algorithmic performance, the discussion of real-world applications and the focus on the complementary impact of ML and MFA in thwarting contemporary cyberthreats constitute a noteworthy contribution, providing useful insights for putting improved security measures into place in a variety of online environments. Lastly, the research incorporates real-world application, which is sometimes lacking in previous studies, in addition to detection.

5. DISCUSSION:

To test the efficacy of the suggested model, researchers compare it to others developed utilizing the 5G-NIDD Dataset & the CIC SGG Dataset. Comparing the proposed model to competing models in binary & multi-label classification tasks, the results show that the suggested model performs better. Table 4 summarizes the comparative findings regarding the Network Intrusion Detection dataset and presents the assessment of the 2000 5G-NIDD Dataset. The classification skills of XGBoost and the better data-balancing capabilities of SMOTE are responsible for the proposed model's superiority over state-of-the-art methods.

Table 4 Analysis of the Network Intrusion Detection Datasets.

Method for Feature Selection	Algorithm Classification	Chosen Features	Accuracy [Percentage]
FGCC+CFA	DT	11	95.04
IGR+CR+ S+CS	PART	13	99.33
EFS	RF	16	93.91
PSO	ANN	21	98.01
CR	DNN	31	99.41
BAT	SVM	26	94.12

The results make it clear that the suggested model is more accurate than other methods for binary classification. To test the suggested model, researchers also examine other models that were built using the 5G-NIDD Dataset and CIC SGG Dataset. When compared to other models for binary and multi-label classification, the findings show that the proposed model is the most accurate. Table 4 shows the tabular comparison results, which mostly concentrate in the molecular level performance of several models on the network intrusion detection dataset. Based on the findings, it may be concluded that using SMOTE to balance the data and using the XGBoost as the model were significant factors.

6. REAL-TIME EFFECTS:

The information collected has real-world effects for network intruder detecting tools. With the model's improved accuracy and speed, researchers can use the method to better analyze network attacks in several different situations. This model is useful by finding threats faster and reducing fake alarms. The design could not pose a problem in theory, but it might be hard to use in real life because it needs a lot of computing power and integrating with other systems. Finding solutions to these problems will require improving the steps that are taken while the framework needs to be run in real time, while still following current security measures.

7. CONCLUSION:

This paper shows a unique and very efficient hybrid machine learning method for Network Intrusion Detection Systems (NIDS), hence advancing detection accuracy and dependability. Using XGBoost for strong feature selection and SMote for data balancing helps us to efficiently handle unbalanced datasets and challenging attack patterns. Showcasing up to 99% accuracy across many datasets (CIC SGG, AU, MSS, 5G-NIDD), the empirical findings highlight the excellence of our methodology above many modern cybersecurity research and conventional techniques. This study is new in the strategic integration of SMote and XGBoost along with the thorough assessment of several machine learning and deep learning classifiers (DT, KNN, MLP, RF, CNN, ANN). Consistent excellent performance of the Random Forest (RF) classifier across all datasets confirms the effectiveness of our feature selection and data balancing techniques. Further adding to the generalization and robustness of the model is XGBoost's capacity to improve feature accuracy using gradient boosting and regularizing methods.

This study has great influence. By providing a scalable and sensible solution for real-time intrusion detection, the suggested hybrid pipeline

helps companies to proactively reduce developing cyber risks. This method is easily implemented in many network contexts because of the shown adaptability and great detection rates. Moreover, the model's natural elasticity helps it to be well-positioned for future developments, which will enable ongoing responsiveness to changing data environments and new security concerns. Investigating sophisticated feature selection methods—especially in concert with neural networks—will be part of future studies to help further improve the NIDS. Using neural network-based feature extraction and investigating group selection of features can help to improve detection accuracy and provide better understanding of feature interactions. Maintaining and enhancing the efficacy of the model is the long-term goal, thus guaranteeing its indispensable use in the always changing terrain of cybersecurity concerns. In the end, our study offers a highly accurate and flexible NIDS architecture, therefore greatly enhancing the state-of-the-art in network security and offering a strong protection against contemporary assaults.

8. ACKNOWLEDGEMENT:

The authors extend their appreciation to Northern Border University, Saudi Arabia, for supporting this work through project number (NBU-CRP-2025-2461)

REFERENCES:

[1] Mahmood, Rafah Kareem, et al. "Optimizing Network Security with Machine Learning and Multi-Factor Authentication for Enhanced Intrusion Detection." *Journal of Robotics and Control (JRC)* 5.5 (2024): 1502-1524.

[2] Williamson, Gregory D., and G. E. Money–America's. *Enhanced authentication in online banking*. Diss. Utica College, 2006.

[3] Saxena, Neetesh, and Bong Jun Choi. "State of the art authentication, access control, and secure integration in smart grid." *Energies* 8.10 (2015): 11883-11915.

[4] Lynch, Jennifer. "Identity theft in cyberspace: Crime control methods and their effectiveness in combating phishing attacks." *Berkeley Tech. LJ* 20 (2005): 259.

[5] Jimmy, Fnu. "Enhancing data security in financial institutions with blockchain technology." *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023* 5.1 (2024): 424-437.

[6] Sanyaolu, Temitope Oluwafunmike, et al. "Harnessing blockchain technology in banking to enhance financial inclusion, security, and transaction efficiency." *International Journal of Scholarly Research in Science and Technology, August* 5.01 (2024): 035-053.

[7] Rajulu GG, Rani MJ, Deepa D, Mamodiya U, Deshmukh RG, Kumar TR. Cloud-Computed Solar Tracking System. Computer Communication, Networking and IoT: Proceedings of 5th ICICC 2021, Volume 2. 2022 Oct 4;459:75.

[8] Ali, Guma, Mussa Ally Dida, and Anael Elikana Sam. "Two-factor authentication scheme for mobile money: A review of threat models and countermeasures." *Future Internet* 12.10 (2020): 160.

[9] Mewada, Shivlal, Anil Saroliya, N. Chandramouli, T. Rajasanthosh Kumar, M. Lakshmi, S. Suma Christal Mary, and Mani Jayakumar. "Smart Diagnostic Expert System for Defect in Forging Process by Using Machine Learning Process." *Journal of Nanomaterials* 2022 (2022).

[10] Onyesolu, Moses O., and Amara C. Okpala. "Improving security using a three-tier authentication for automated teller machine (ATM)." *International Journal of Computer Network and Information Security* 9.10 (2017): 50.

[11] Hossain, Mohammad Ikbal, et al. "Enhancing Data Integrity and Traceability in Industry Cyber Physical Systems (ICPS) through Blockchain Technology: A Comprehensive Approach." *arXiv preprint arXiv:2405.04837* (2024).

[12] Khalifa, Nada, Wael Elmedany, and Saeed Sharif. "Leveraging Digital Identity and Open Banking Data for Fraud Prevention in the Financial Industry." *2024 11th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, 2024.

[13] Gai, Keke, et al. "Security and privacy issues: A survey on FinTech." Smart Computing and Communication: First International Conference, SmartCom 2016, Shenzhen, China, December 17-19, 2016, Proceedings 1. Springer International Publishing, 2017.

[14] Hasan, Mohammad Kamrul, et al. "A review on security threats, vulnerabilities, and counter measures of 5G enabled Internet-of-Medical-Things." *IET communications* 16.5 (2022): 421-432.

[15] Jasper, K. Daniel, et al. "FMDB Transactions on Sustainable Computing Systems." (2023).

[16] Alsayed, A., and Anwar Bilgrami. "E-banking security: Internet hacking, phishing attacks, analysis and prevention of fraudulent activities." *International Journal of Emerging Technology and advanced engineering* 7.1 (2017): 109-115.

[17] Usman, Ahmad Kabir, and Mahmood Hussain Shah. "Critical success factors for preventing e-banking fraud." *Journal of Internet Banking and Commerce* 18.2 (2013): 1-15.

[18] Shoniregun, Charles Adetokunbo. "Are existing Internet security measures guaranteed to protect user identity in the financial services industry?." *International Journal of Services Technology and Management* 4.2 (2003): 194-216.

[19] Maheshwaran, T., et al. "Securing E-Commerce Strategies With Cloud, Blockchain, AI, and ML." *Strategies for E-Commerce Data Security: Cloud, Blockchain, AI, and Machine Learning*. IGI Global, 2024. 470-500.

[20] Asmar, Muath, and Alia Tuqan. "Integrating machine learning for sustaining cybersecurity in digital banks." *Heliyon* 10.17 (2024).

[21] Nadeem, Muhammad, et al. "Phishing attack, its detections and prevention techniques." *Int. J. Wirel. Secur. Netw* 1 (2023): 13-25.

[22] Mishra, Aditya Dev, and Youddha Beer Singh. "Big data analytics for security and privacy challenges." *2016 international conference on computing, communication and automation (ICCCA)*. IEEE, 2016.

[23] Obaidat, Mohammad S., Issa Traore, and Isaac Woungang, eds. *Biometric-based physical and cybersecurity systems*. Cham: Springer International Publishing, 2019.

[24] Ahmad, Ahmad Yahiya Ahmad Bani. "Fraud Prevention in Insurance: Biometric Identity Verification and AI-Based Risk Assessment." *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*. Vol. 1. IEEE, 2024.

[25] Centobelli, Piera, et al. "Blockchain technology design in accounting: Game changer to tackle fraud or technological fairy tale?." *Accounting, Auditing & Accountability Journal* 35.7 (2022): 1566-1597.