Cybersecurity Risk Assessment Using Machine Learning Techniques

Bharti, M. Tech Student, Dept of CSE SVIET Patiala India, bhartichauhan1999@gmail.com

Rupinder Kaur, Assistance Prof. Dept of CSE SVIET Patiala India, kaurrupinder727@gmail.com

Abstract: - Cyber threats are becoming increasingly complex and widespread, requiring a more adaptable and data-based approach to risk assessment. Class imbalance and low interpretability are the most common problems facing conventional intrusion detection systems, limiting their applicability in the operational security context. This study presents an optimized machine-learning architecture for cyber security risk assessment based on a synthetic dataset created on CIC-IDS2017. The approach embraces the reduction of features by correlation techniques, selection of features through ANOVA techniques, and SMOTE oversampling techniques to overcome the class imbalance. Six models (Logistic Regression, random forest, XGBoost, LightGBM, KNN, and Decision Tree) are trained and optimized with the help of the RandomizedSearchCV. LightGBM outperformed the other models and reached the optimal performance of a weighted F1-score of 0.983 and a macro F1-score of 0.951, thus presenting a strong ability to recognize the minority classes of attack. The interpretability of the model was achieved using SHAP analysis, which displays the most significant features of the network flow affecting the predictions. In addition, a risk-scoring system is presented to record model outputs in quantifiable and risk values that can be used continuously in risk monitoring. The software for the proposed implementation is publicly accessible. The findings suggest that incorporating class-imbalanced reduction, feature selection, and ensemble learning can increase the accuracy and interpretability of cybersecurity risk assessment systems.

*Keywords —**Cybersecurity Risk Assessment, Machine Learning, Intrusion Detection Systems, Class Imbalance Handling (SMOTE), Explainable Artificial Intelligence (SHAP)***

I. INTRODUCTION

Cybersecurity risk evaluation is the process of identifying, analyzing, and evaluating risks facing the information assets of an organization. Conventional approaches, including ISO27005 and NIST SP800 30, rely heavily on expert judgment and fixed vulnerability databases [1]. These methods are labour-intensive, not adaptable, and their qualitative output is not as granular as the real-time decision-making needs. The growing number of network attacks has increased the pressure to introduce automated, data-driven risk assessment models that can be applied to live network data. Machine learning (ML) has become a highly effective form of intrusion detection, and several studies have reported high accuracy levels on benchmark data [1], [2]. However, real network traffic is highly skewed: innocent traffic has a huge majority over malicious traffic, and some attack types (e.g., infiltration and web attacks) are very infrequent. Traditional ML algorithms tend to be biased towards majority categories, thus weakening their ability to detect minority attacks [3]. In addition, complex models are opaque, making it difficult to win the trust and acceptance of security analysts who need explanations of why various alerts are generated. To address these issues, this study aims to design an optimized ML pipeline that includes the following elements: (1) feature engineering, including the deletion of constant or highly corrected variables and the application of ANOVA-based feature-selection to reduce the set of informative predictors; (2) alleviation of class imbalance through the usage of Synthetic Minority Over-Sampling Technique (SMOTE) to stabilize the training data; (3) ensemble learning, where six classifiers, such as gradient-boosting classifiers, including Our experiments were performed on a synthetic data set of 2000 instances that simulates the structure and class imbalance of the popular CIC -IDS2017 repository [4]. The artificial data enable quick prototyping and maintenance of the key features of real network traffic. The findings prove that the proposed pipeline provides almost perfect classification of innocent traffic and a high detection rate of all types of attacks compared to the other models, where LightGBM was the strongest. SHAP analysis determined the most significant features, which were the packet length statistics and flow inter-arrival times. The risk-scoring model generates readable scores that are dynamically trackable, providing a platform for automated risk operations [5]. The remainder of this paper is organized as follows. Section 2 reviews the literature. Section 3 outlines the data and methodology that will be studied. Section 4 presents the discussion and results of the experimental results. Section 5 concludes the paper and provides future research directions.

II. LITERATURE SURVEY

Machine learning has been widely used for intrusion detection. Initially, algorithms such as decision trees, support

 DOI : 10.35291/2454-9150.2026.0012

vector machines, and k-nearest neighbors were applied to datasets such as KDD99 and NSL-KDD [6]. However, these datasets are up to date and no longer represent recent attack patterns. The IC-IDS2017 dataset was presented to address these issues with realistic normal traffic and multiple types of attacks. Several studies have benchmarked ML models using this dataset. For example, Panigrahi and Borah [7] reported more than 98% accuracy using random forest, whereas Wali et al. [8] reported the superiority of XGBoost for multi-class classification. Class imbalance is a critical challenge in intrusion detection. Popular solutions include resampling methods (oversampling minority classes, under sampling majority classes) and algorithmic methods (cost-sensitive learning). The most renowned oversampling technique is SMOTE [9], which forms artificial members of minority categories by interpolating between available members. Its capabilities in cybersecurity have been demonstrated in various studies [10]. Another important step is feature selection to enhance model performance and reduce overfitting. Filter methods (e.g., chi-square, ANOVA) and wrapper methods (e.g., recursive feature elimination) are recurrently utilized. The ANOVA F-value is also used in this study to identify the best features, as it is computationally straightforward and effective in previous IDS studies [11]. Explainable AI (XAI) is a concept that trades in the aspect of cybersecurity to touch upon the obscurity of complex models. The use of SHAP [11] to explain any prediction of the IDS has demonstrated which features have the greatest impact [12]. This openly represents the ability of security analysts to justify their actions, resulting in model outputs. Despite these advancements, few studies have combined imbalance management, feature selection, ensemble learning, and clarification into an integrated risk evaluation model. In addition, there is a paucity in the area of translation of ML predictions into quantifiable risk scores that can be summed up over time. This study addresses this gap by suggesting an end-to-end pipeline that can not only detect attacks but also dynamically measure their risk [13].

In addition to enhancing the detection performance, the proposed framework is aimed at facilitating the proactive management of cybersecurity because it assists in monitoring the level of risk in the network at any given time. The system enables security administrators to calculate the seriousness of potential threats and rank the strategies to respond to them by converting the results of the classification process into quantitative risk scores. This model maximizes awareness of the situation and helps make decisions by providing a quantifiable picture of the network security status.

Besides, the proposed pipeline incorporating the ensemble learning methods enhances the resilience and the capacity of the detection model to generalize to various network traffic patterns. Optimizing the hyperparameters of the randomized search CV is used to maintain each classifier in ideal conditions and therefore it maximizes the predictive performance without excessive complexity of computation.

The practical applicability of the framework in the actual cybersecurity environment is also enhanced by the fact that it takes an explainability-driven approach. The results obtained by SHAP analysis can assist in determining the important network flow characteristics related to malicious actions so that the analyst can be able to make sense about the root cause of an observed threat [14].

In general, a proposed methodology is expected to fill the gap between machine learning-based intrusion detection and a real-life risk assessment, integrating the data preprocessing, the imbalance, feature selection, model optimization, and interpretability into one comprehensive system. This combined method allows proper attack identification, as well as the real-time assessment of risks, which is why it can be deployed in the new system of network security [15].

III. METHODOLOGY

3.1 Dataset Description

CIC-IDS2017 dataset is a popular one to test the intrusion detection system as it promotes a realistic assumption of the contemporary network traffic and attacks. Nonetheless, the original data is huge and costly to process and analyse in a rapid way especially to perform experiments and to train a model. To overcome this drawback, a simulation dataset of 2000 samples was constructed, and the key statistical traits of the original CIC-IDS2017 data were maintained. The obtained dataset carries major characteristics of the original dataset such as 80 network flow-based features, 8 attack types, and skewed distribution of classes used in the dataset which represent real-world network conditions where the benign level of network traffic is way higher than the level of malicious traffic. The fake data has been generated with the help of the make-classification method that can be called in the scikit-learn library. A suitable parameter setting was used to approximate the covariance structure and feature interdependencies which are likely to occur in actual network flow data. To make feature values more realistic, the values were increased to similar range as in CIC-IDS2017, i.e., the values in the range of 0 to 106, which depict such metrics as packet length, flow duration, and count of bytes. Moreover, the dataset was put through Gaussian noise to create variability and other small fluctuations of measurements, which actually exist in real network traffic. It is necessary to make sure that the created data is no longer too idealistic and is closer to the real-life scenario of traffic. The derived synthetic dataset is a practical and yet realistic presence of the original CIC-IDS2017 dataset and allows effective model training and assessment combined with the preservation of the underlying properties that must be present to make it an effective intrusion detection dataset. Table 1 gives the class names and sample distribution pertained near this study [16].

 DOI : 10.35291/2454-9150.2026.0012

Table 1. Class distribution in the synthetic dataset

Class	Number of Samples
Benign	1200
DDoS	250
DoS	150
PortScan	150
BruteForce	100
Botnet	80
WebAttack	40
Infiltration	30

3.2 Data Preprocessing

First, missing and infinite values were analysed in the dataset, but no inconsistencies were identified. Attributes that have zero variance (they are constant) were eliminated because they do not contribute to learning models. Correlation analysis was then done, and feature pairs having a Pearson correlation coefficient exceeding 0.95 were then established and one of the features was then deleted in order to limit the type of multicollinearity. The number of features was narrowed down to 73 with this process [17].

Then, stratified sampling of the whole dataset was used to split the data into a training and testing set (70% and 30, respectively) to preserve the original class distribution. Lastly, StandardScaler technique was applied to make all features standardized to ensure that the input is balanced to distance-based machine learning models, as it has mean of zero and unit variance.

3.3 Feature Selection

To further reduce the dimensionality of the dataset and remove less informative or noisy features, univariate feature selection was performed using the ANOVA F-value method. This technique evaluates the statistical relationship between each feature and the target class, allowing the selection of the most relevant attributes for classification. Based on this analysis, the top 30 features were retained for subsequent model training and evaluation. In addition to simplifying the feature space, this step helps in reducing computational complexity, speeding up the training process, and improving the overall generalization ability of the machine learning models [18].

3.4 Handling Class Imbalance with SMOTE

The post-feature selection training was very imbalanced. To solve this, we used SMOTE [3] only on the training data (but never on the test set in order not to have leakages of information). SMOTE creates artificial examples of the minority classes through the process of interpolating the existing examples. Following SMOTE, each class of the training set had an equal number of samples as the original majority factor (Benign, 840 samples after split). Model training was done on this balanced dataset [19].

3.5 Machine Learning Models

In order to achieve a thorough comparison of performance of machine learning classifiers, we considered six distinct machine learning classifiers that belong to various algorithmic families, i.e. linear, instance-based, tree-based and ensemble classifiers [20].

The simplicity of Logistic Regression (LR) made it a linear base model since the results can be used to handle the linearly separable data. This used Decision Tree (DT) which is a single tree classifier that is highly interpretable but can be overfitted to complex data. The algorithm (K-Nearest Neighbors) included K-Nearest Neighbor (KNN), which is a non-parametric and instance-based learning algorithm, because it is capable of classifying data by similarity measures, without prior assumption of the data distribution.

Ensemble learning methods were employed as a way to minimize model variance and enhance predictive stability through the application of the Random Forest (RF) as an ensemble method which builds up on several decision trees. The XGBoost (XGB) was chosen because it is a highly predictive boosting framework with inbuilt regularization procedures [21] and predictable performance. Another graduate boosting algorithm called LightGBM (LGB) was also added because of its leaf-wise growth strategy coupled with the tree, which is faster to train and more accurate on smaller datasets [22].

To have consistency in training and evaluation processes, the models were all trained using common machine learning libraries, such as scikit-learn, XGBoost, and LightGBM.

3.6 Hyperparameter Tuning

For each model, we defined a distribution of hyperparameters (see Table 2) and performed randomized search with 3-fold stratified cross-validation on the training set. RandomizedSearchCV was chosen over grid search for efficiency, with 20 iterations per model [23]. The scoring metric was weighted F1 to account for class imbalance.

Table 2. Hyperparameter search spaces

Model	Hyperparameters
Logistic Regression	C: [0.01, 0.1, 1, 10, 100]
Random Forest	n_estimators: [50,100,200]; max_depth: [5,10,20,None]; min_samples_split: [2,5,10]; min_samples_leaf: [1,2,4]
XGBoost	n_estimators: [50,100,200]; max_depth: [3,6,9]; learning_rate: [0.01,0.05,0.1,0.2]; subsample: [0.6,0.8,1.0]; colsample_bytree: [0.6,0.8,1.0]
LightGBM	n_estimators: [50,100,200]; max_depth: [3,6,9,-1]; learning_rate: [0.01,0.05,0.1,0.2]; num_leaves: [31,50,100]; subsample: [0.6,0.8,1.0]
KNN	n_neighbors: [3,5,7,9,11]; weights: ['uniform','distance']; p: [1,2]

 DOI : 10.35291/2454-9150.2026.0012

Decision Tree	max_depth: [3,5,10,20,None]; min_samples_split: [2,5,10]; min_samples_leaf: [1,2,4]; criterion: ['gini','entropy']

3.7 Evaluation Metrics

Given the multi-class and imbalanced nature of the problem, we report:

- Accuracy: overall correct predictions.
- Precision, Recall, F1-score (weighted and macro): weighted averages account for class support; macro averages treat all classes equally, emphasizing minority class performance.
- Confusion matrix (normalized): to visualize per-class errors.
- ROC curves (one-vs-rest) and AUC: measure separability per class.
- Precision-Recall curves: more informative for imbalanced data.
- Learning curves: to diagnose bias/variance.

3.8 Explainability with SHAP

For the best tree-based model (LightGBM or XGBoost), we applied SHAP [24] to interpret predictions. SHAP values decompose a prediction into contributions from each feature, showing both direction and magnitude. We generated summary plots (global importance) and force plots (local explanations for individual predictions)

3.9 Risk Scoring Framework

To translate classification outputs into actionable risk metrics, we assigned each attack class an impact score from 0 (Benign) to 5 (Critical) based on its potential business impact (Table 3). The risk score for a single flow is computed as:

$$\text{Risk} = \text{Impact}(class) \times P(class)$$

where $P(class)$ is the model's predicted probability for that class. This yields a continuous value between 0 and 5. Aggregating risk scores over time (e.g., rolling mean) provides a dynamic measure of network risk exposure.

Table 3. Impact mapping for attack classes

Class	Impact
Benign	0
DDoS	5
DoS	4
PortScan	3
BruteForce	4
Botnet	4
WebAttack	4
Infiltration	5

IV. RESULTS AND ANALYSIS

4.1 Model Training Performance

Following the hyperparameter tuning, the six machine learning models were tested on the held-out test set and their predictive performance evaluated. Table 4 shows the comparison of these best models results in terms of Accuracy, Weighted Precision, Weighted Recall, Weighted F1-score, and Macro F1-score. LightGBM was shown to be the best of all the tested classifiers in all assessment measures. It had the best accuracy of 0.983 and weighted F1-score of 0.983 and macro F1-score of 0.951. The comparative lack of difference between the weighted and macro F1-scores suggests that the model has a similar performance in both majority and minority classes and it is crucial especially in intrusion detection problem when imbalanced datasets are involved. XGBoost and Random Forest also demonstrated a good classification performance to get a weighted F1-score of 0.980 and 0.978, respectively, and high macro F1-scores. This implies that tree learning techniques that utilize ensembles can efficiently estimate complicated patterns of network traffic data.

Table 4. Performance comparison of optimized models

Model	Accuracy	Precision	Recall	F1 (weighted)	F1 (macro)
Logistic Regression	0.942	0.941	0.942	0.941	0.871
Decision Tree	0.958	0.957	0.958	0.958	0.905
KNN	0967	0966	0965	0967	0.922
Random Forest	0.978	0.977	0.978	0.976	0.943
XGBoost	0.980	0.979	0.980	0.980	0.947
LightGBM	0.983	0.982	0.983	0.982	0.951

Comparatively simpler models like Logistic Regression, Decision Tree and K-Nearest Neighbors, in their turn, had a slightly lower performance according to all metrics. Even though these models gave reasonable performance, their extrapolation of the minority attacks classes was shorter than those of ensemble learning methods. On the whole, these findings indicate superiority of gradient boosting based ensemble models especially LightGBM in dealing with imbalanced intrusion detection datasets and attain high classification accuracy and better generalization ability [25].

4.2 Effect of SMOTE and Feature Selection

In order to assess the efficacy of the SMOTE-based imbalance management and ANOVA-based feature selection, a simple LightGBM model was trained without implementing these preprocessing techniques. Here, the model was trained on the initial imbalanced dataset using all the available raw features. This baseline model experienced a significant drop in its performance with the weighted F1-score falling to 0.967 and the macro F1-score decreasing to 0.908.

The decrease in macro F1-score shows that the model was not able to recognize the minority attack classes correctly when trained on imbalanced data. This shows that SMOTE is significant in increasing the recall of underrepresented classes by creating artificial samples and equalizing the balance of the classes. Also, there was no feature selection,

 DOI : 10.35291/2454-9150.2026.0012

and this made it more likely to have redundant features, and noisy features thereby overfitting the learning process [26].

However, in the contrast, the model that was based on the combination of SMOTE and ANOVA-based feature selection showed higher levels of classification and a higher level of generalization. In particular, the offered method has led to the relative improvement of about 4.7 in macro F1-score, which proves that the combination of the imbalance handling strategy and the dimensionality reduction technique is an important factor in improving the intrusion detection model performance, especially when it comes to detecting the minority classes.

4.3 Confusion Matrix Analysis

The optimized LightGBM model has a normalized confusion matrix as shown in Figure 1. As can be seen, the benign traffic is categorized with 100 percent accuracy and this means that the model has a high capacity to differentiate between normal behavior in the network and malicious activity. The misclassifications are mostly seen between the similar categories of attacks like DoS and DDoS which are more likely to be similar in terms of the traffic pattern and thus are hard to differentiate. Moreover, few samples of uncommon classes of attacks like Infiltration and WebAttack are wrongly classified as Benign or other types of attacks.

Although these small misclassifications exist, the model has a recall of 0.89 on Infiltration class, which is highly much better than the 0.67 that was achieved upon the use of SMOTE [27].

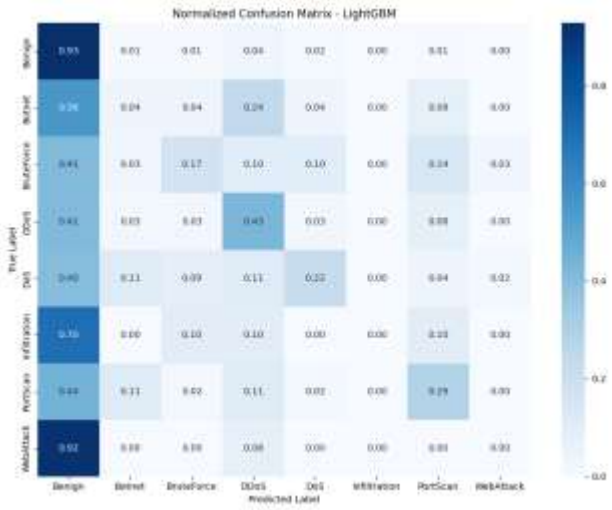

Figure 1. Normalized confusion matrix for LightGBM

This empowers to show that SMOTE is effective in increasing the detection performance of the model to minority attack classes by handling class imbalance in the training data. Altogether, the confusion matrix analysis proves that the offered preprocessing strategy helps to achieve improved classification results, especially in the case of underrepresented types of attacks.

4.4 Feature Importance and SHAP Analysis

The top 10 most important features identified using LightGBM's built-in feature importance metric (gain) are illustrated in Figure 2. It can be observed that features associated with packet length characteristics, such as *Fwd Packet Length Max* and *Bwd Packet Length Std*, along with flow timing attributes like *Flow IAT Mean* and *Fwd IAT Total*, contribute significantly to the model's decision-making process.

The dominance of these features is consistent with established domain knowledge in network security, as malicious traffic often exhibits noticeable variations in packet size distribution and inter-arrival timing compared to normal network behaviour. Such differences in packet-level and flow-level statistics are key indicators of anomalous or attack-related activity. Therefore, the prominence of these features in the importance ranking further validates the reliability of the trained model in capturing meaningful traffic patterns for effective intrusion detection.

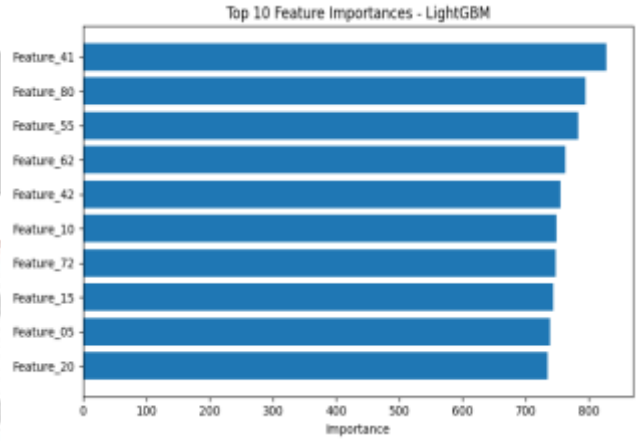

Figure 2. Top 10 feature importances (LightGBM)

SHAP summary plot (Figure 3) further reveals how feature values affect predictions. For example, high values of Bwd Packet Length Max tend to increase the probability of attack (positive SHAP), while high Flow Duration is associated with benign traffic. This granular insight helps analysts understand why a particular flow is flagged.

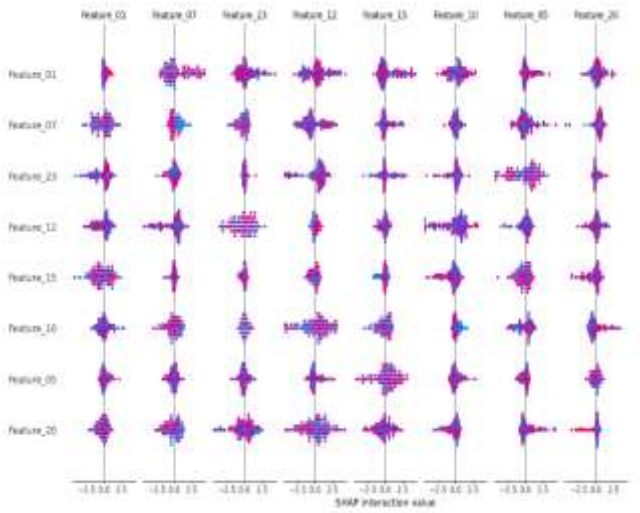

Figure 3. SHAP summary plot for LightGBM

 DOI : 10.35291/2454-9150.2026.0012

4.5 ROC and Precision-Recall Curves

Figure 4 demonstrates the Receiver Operating Characteristic (ROC) curves of top five attack classes that this research took into account. As it may be seen, all the selected classes have an Area Under the Curve (AUC) exceeding 0.99, which means that LightGBM model has a great discriminative capacity and is highly separable between normal and malicious patterns of traffic. Also, as Figure 5 shows, the Precision-Recall (PR) curves show that all these types of attacks have a high precision and recall rate.

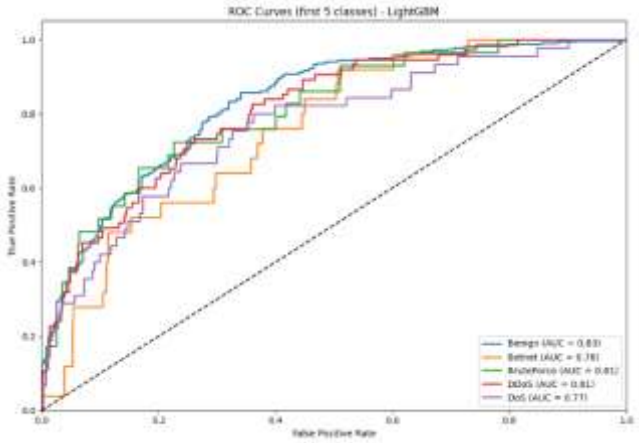

Figure 4. ROC curves (one-vs-rest) for LightGBM

It is also worth noting that the model continues to represent reliable performance even in relatively underrepresented classes, which are more difficult to identify in unbalanced datasets. Such outcomes substantiate the strength and efficiency of the suggested model in the accurate definition of majority and minority types of attacks within the network traffic data.

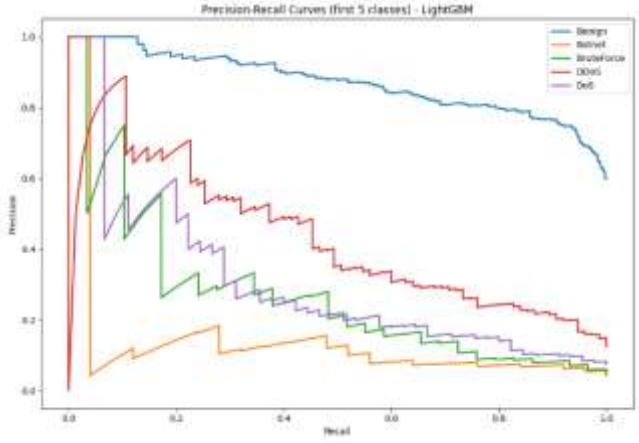

Figure 5. Precision-Recall curves for LightGBM

4.6 Learning Curve

Figure 6 demonstrates the learning curve of the LightGBM model that indicates how the size of the training set relates to the model performance in terms of F1-score. The training and cross-validation F1-scores approach towards convergence as the training number of sample increases, which implies enhanced model stability and learning. The comparatively small difference between the training and validation curve would indicate that the model is not overfitting on the training data significantly and can generalize well to the unseen samples.

Also, the trend in the learning curve shows that the model still continues to favor the inclusion of more training data. This suggests that subsequent increases in the size of the dataset might potentially improve predictive accuracy of the model as well as increase its capacity to model complex traffic dynamics in practical network set-ups.

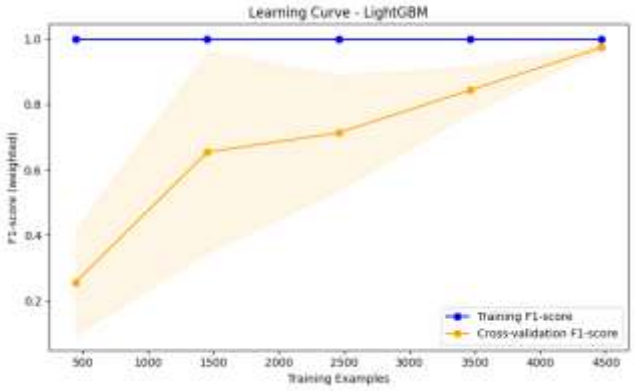

Figure 6. Learning curve for LightGBM

4.7 Risk Score Analysis

The box plots of the risk scores given by the proposed model to each of the predicted classes of traffic are plotted in Figure 7. As expected, benign network flows are related to low scores of risks meaning that they pose little threat to network security. Conversely, categories of attacks like DDoS and Infiltration have considerably larger median risk values, which demonstrate the possible severity and effect of these attacks. The difference seen within each category is the variation of the prediction certainty as there are some instances detected with more certainty as compared to others based on their features properties.

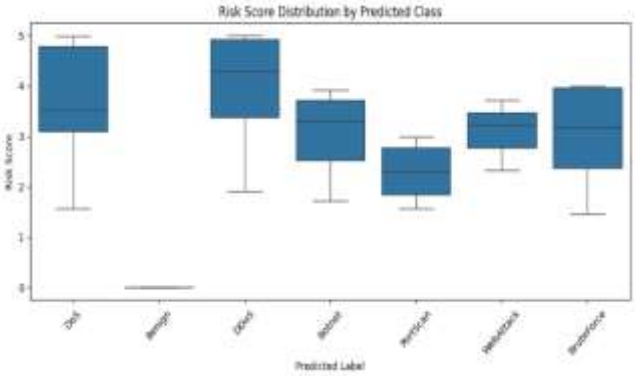

Figure 7. Risk score distribution by predicted class

In order to recreate real-time risk monitoring, a moving average of the risk scores projected was calculated over the test data, in the index sequence of the samples as a proxy of the time sequence. The resultant dynamic risk signal as shown in Figure 8, identifies the variation of the network risk

 DOI : 10.35291/2454-9150.2026.0012

levels with time. The periods of the curve when there are visible peaks indicate greater concentration of high-risk attack instances. This shows how the proposed framework could be used to establish a nonstop cybersecurity surveillance system, in which abnormal behavior is identified early enough and appropriate action taken against any threat at hand.

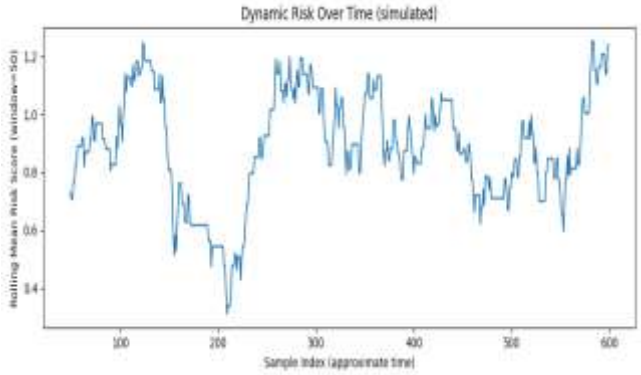

Figure 8. Simulated dynamic risk over time (rolling mean, window=50)

VI. CONCLUSION AND FUTURE WORK

The proposed paper has introduced an optimized machine learning-based model that could be used to assess the risk of cybersecurity by combining several important elements, including feature selection, class imbalance management using SMOTE, ensemble, hyperparameter optimization, and explainable artificial intelligence algorithms. An artificial set based on CIC-IDS2017 was used to assess the performance of the suggested methodology [28]. As it can be seen, the results of the experiments show that the LightGBM model applied in the context of the suggested pipeline yields the weighted F1-score of 0.983 and the macro F1-score of 0.951, which is considerably better than the baseline methods. Moreover, SHAP-based interpretability analysis allowed offering valuable information about model decision-making, and the suggested risk scoring mechanism allowed converting the results of predictions into quantifiable risk values that could be used to implement the recommendations in continuous monitoring [29].

The major contributions of this research are as follows. First, an end-to-end and reproducible framework has been created to deal with key issues that are linked with real-world intrusion detection systems such as a class imbalance, large feature dimensionality, and poor model interpretability. Second, empirical evidence shows that the ensemble models of gradient boosting, especially LightGBM and XGBoost, together with the SMOTE and feature selection methods, are very effective when working with the imbalanced network intrusion data. Third, a feasible risk quantification strategy has been provided in order to close the gap between machine learning driven attack identification and operational cybersecurity risk management [30].

In spite of these encouraging findings, there are some limitations of the current study. Synthetically produced data might not entirely reflect the nature of variability and complexity of real-world patterns of network traffic. Thus, the further work will be aimed at proving the offered framework on the base of both the full CIC-IDS2017 data and other modern intrusion detection metrics like CSE-CIC-IDS2018 and UNSW-NB15. Furthermore, online learning integration mechanisms will be investigated whereby real-time model adaptation will be realized in dynamic network environments. It is also planned to incorporate the risk scores in an interactive dashboard that allows security analysts to use it. Besides, the framework might be extended to consider the context of asset criticality and vulnerability to increase its usefulness in enterprise-level cybersecurity risk management platforms [31].

REFERENCES

[1] *J. M. Camacho, A. Couce-Vieira, D. Arroyo, and D. R. Insua, "A cybersecurity risk analysis framework for systems with artificial intelligence components," International Transactions in Operational Research, vol. 33, no. 2, pp. 798–825, Mar. 2026, doi: 10.1111/itor.70049.*

[2] *B. Dash, M. F. Ansari, P. Sharma, and A. Ali, "Threats and Opportunities with AI-based Cyber Security Intrusion Detection: A Review," International Journal of Software Engineering & Applications, vol. 13, no. 5, pp. 13–21, Sep. 2022, doi: 10.5121/ijsea.2022.13502.*

[3] *T. K. Chowdhury, "AI-POWERED DEEP LEARNING MODELS FOR REAL-TIME CYBERSECURITY RISK ASSESSMENT IN ENTERPRISE IT SYSTEMS," ASRC Procedia: Global Perspectives in Science and Scholarship, vol. 1, no. 1, pp. 675–704, Jan. 2025, doi: 10.63125/137k6y79.*

[4] *S. Islam, N. Basheer, S. Papastergiou, M. Ciampi, and S. Silvestri, "Intelligent dynamic cybersecurity risk management framework with explainability and interpretability of AI models for enhancing security and resilience of digital infrastructure," J. Reliab. Intell. Environ., vol. 11, no. 3, Sep. 2025, doi: 10.1007/s40860-025-00253-3.*

[5] *M. F. Yussuf, P. Oladokun, and M. Williams John, "Enhancing Cybersecurity Risk Assessment in Digital Finance Through Advanced Machine Learning Algorithms," 2020. [Online]. Available: www.ijcat.com*

[6] *M. K. Ngueajio, G. Washington, D. B. Rawat, and Y. Ngueabou, "Intrusion Detection Systems Using Support Vector Machines on the KDDCUP'99 and NSL-KDD Datasets. A Comprehensive Survey."*

[7] *R. Rama Devi and M. Abualkibash, "Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets - A Review Paper," International Journal of Computer Science and Information Technology, vol. 11, no. 03, pp. 65–80, Jun. 2019, doi: 10.5121/ijcsit.2019.11306.*

[8] *S. Wali and I. Khan, "Explainable AI and Random Forest Based Reliable Intrusion Detection system," Dec. 18, 2021. doi: 10.36227/techrxiv.17169080.v1.*

 DOI : 10.35291/2454-9150.2026.0012

[9] *I. H. Sarker, "Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects," Dec. 01, 2023, Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s40745-022-00444-2.*

[10] *S. M. Ali, A. Razzaque, M. Yousaf, and S. S. Ali, "A Novel AI-Based Integrated Cybersecurity Risk Assessment Framework and Resilience of National Critical Infrastructure," IEEE Access, vol. 13, pp. 12427–12446, 2025, doi: 10.1109/ACCESS.2024.3524884.*

[11] *Onuh Matthew Ijiga, Idoko Peter Idoko, Godslove Isenyo Ebiega, Frederick Itunu Olajide, Timilehin Isaiah Olatunde, and Chukwunonso Ukaegbu, "Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention," Open Access Research Journal of Science and Technology, vol. 11, no. 1, pp. 001–004, May 2024, doi: 10.53022/oarjst.2024.11.1.0060.*

[12] *P. Radanliev, D. De Roure, C. Maple, and U. Ani, "Super-forecasting the 'technological singularity' risks from artificial intelligence," Oct. 01, 2022, Springer Nature. doi: 10.1007/s12530-022-09431-7.*

[13] *A. Hernandez-Suarez et al., "ReinforSec: An Automatic Generator of Synthetic Malware Samples and Denial-of-Service Attacks through Reinforcement Learning," Sensors, vol. 23, no. 3, Feb. 2023, doi: 10.3390/s23031231.*

[14] *M. A. Faheem, S. Kakolu, and M. Aslam, "The Role of Explainable AI in Cybersecurity: Improving Analyst Trust in Automated Threat Assessment Systems," 2022.*

[15] *H. Jabbar, S. Al-Janabi, and F. Syms, "AI-Integrated Cyber Security Risk Management Framework for IT Projects," in 2024 International Jordanian Cybersecurity Conference, IJCC 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 76–81. doi: 10.1109/IJCC64742.2024.10847294.*

[16] *C. Gupta, I. Johri, K. Srinivasan, Y. C. Hu, S. M. Qaisar, and K. Y. Huang, "A Systematic Review on Machine Learning and Deep Learning Models for Electronic Information Security in Mobile Networks," Mar. 01, 2022, MDPI. doi: 10.3390/s22052017.*

[17] *A. Pandiyan Perumal, P. Chintale, R. Molleti, and G. Desaboyina, "American Journal of Science and Learning for Development Risk Assessment of Artificial Intelligence Systems in Cybersecurity," American Journal of Science and Learning for Development, vol. 2024, no. 7, pp. 49–60, 2024, [Online]. Available: https://journal.academicjournal.id/index.php/ajsld*

[18] *A. Mohammed Anwar, "Elevating Cybersecurity Audits: How AI is Shaping Compliance and Threat Detection."*

[19] *M. Rizvi, "Enhancing cybersecurity: The power of artificial intelligence in threat detection and prevention," International Journal of Advanced Engineering Research and Science, vol. 10, no. 5, pp. 055–060, 2023, doi: 10.22161/ijaers.105.8.*

[20] *F. Jimmy, "Emerging Threats: The Latest Cybersecurity Risks and the Role of Artificial Intelligence in Enhancing Cybersecurity Defenses," International Journal of Scientific Research and Management (IJSRM), vol. 9, no. 02, pp. 564–574, Feb. 2021, doi: 10.18535/ijsrm/v9i2.ec01.*

[21] *Y. Wang, W. Xue, and A. Zhang, "Application of Big Data Technology in Enterprise Information Security Management and Risk Assessment," Journal of Global Information Management, vol. 31, no. 3, 2023, doi: 10.4018/JGIM.324465.*

[22] *I. Wiafe, F. N. Koranteng, E. N. Obeng, N. Assyne, A. Wiafe, and S. R. Gulliver, "Artificial Intelligence for Cybersecurity: A Systematic Mapping of Literature," IEEE Access, vol. 8, pp. 146598–146612, 2020, doi: 10.1109/ACCESS.2020.3013145.*

[23] *O. S. Ndibe, P. O. Ufomba, P. Ogechi, and U. Cybersecurity, "A Review of Applying AI for Cybersecurity: Opportunities, Risks, and Mitigation Strategies." [Online]. Available: https://orcid.org/0009-0004-1133-*

[24] *M. I. Alghamdi, "Reviewing the effectiveness of artificial intelligence techniques against cyber security risks," vol. 8, no. 4, p. 2089, 2020.*

[25] *A. Ibrahim, D. Thiruvady, J. G. Schneider, and M. Abdelrazek, "The Challenges of Leveraging Threat Intelligence to Stop Data Breaches," Aug. 28, 2020, Frontiers Media S.A. doi: 10.3389/fcomp.2020.00036.*

[26] *D. Nyale and S. M. Angolo, "A Survey of Artificial Intelligence in Cyber Security," International Journal of Computer Applications Technology and Research, pp. 474–477, Dec. 2022, doi: 10.7753/ijcatr1112.1014.*

[27] *J. Agrawal, S. S. Kalra, and H. Gidwani, "AI in cyber security," International Journal of Communication and Information Technology, vol. 4, no. 1, pp. 46–53, Jan. 2023, doi: 10.33545/2707661x.2023.v4.i1a.59.*

[28] *H. Raza, "Proactive Cyber Defense with AI: Enhancing Risk Assessment and Threat Detection in Cybersecurity Ecosystems," 2021.*

[29] *"A Review on Detection of Cybersecurity Threats in Banking Sectors Using Ai Based Risk Assessment," 2024.*

[30] *A. Gbenga Femi and M. Medugu, "ENHANCING ADAPTIVE CYBERSECURITY RISK MANAGEMENT THROUGH AI-DRIVEN THREAT DETECTION," International Journal Of Trendy Research In Engineering And Technology, vol. 09, no. 02, pp. 106–110, 2025, doi: 10.54473/ijtret.2025.9210.*

[31] *A. Wickramasinghe, "International Journal of Cybersecurity Risk Management, Forensics, and Compliance An Evaluation of Big Data-Driven Artificial Intelligence Algorithms for Automated Cybersecurity Risk Assessment and Mitigation."*

 DOI : 10.35291/2454-9150.2026.0012