SINERGI Vol. 29, No. 3, October 2025: 615-624
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.2025.3.005

An intelligent approach for detection and classification of security attacks in a Passive Optical Network using Light Gradient Boosting Machine

Sumayya Bibi[1], Nadiatulhuda Zulkifli[1]*, Farabi Iqbal[1], Sajid Iqbal[2], Arnidza Ramli[1], Adam Wong Yoon Khang[3]
[1]Department of Communication Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia
[2]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Saudi Arabia
[3]Department of Engineering Technology, Faculty of Electronics and Computer Technology and Engineering, Universiti Teknikal Malaysia Melaka, Malaysia

Abstract
Over the past decade, Passive Optical Networks (PONs) have emerged as a leading solution for next-generation broadband access, providing high-speed and cost-effective communication. However, PONs face significant security challenges, including data interception, denial-of-service (DoS) attacks, and resource exhaustion caused by malicious Optical Network Units (ONUs). Machine learning (ML), particularly advanced models like Light Gradient Boosting Machine (LightGBM), has proven to be a promising solution for managing complex security issues in PONs. Leveraging its ability to handle imbalanced, high-dimensional datasets, LightGBM was employed in this study to detect and classify malicious ONUs based on bandwidth usage patterns. The model achieved an impressive accuracy of 95.27%, a Matthews Correlation Coefficient (MCC) of 90%, and a precision rate of 93%. While traditional classifiers, such as Naïve Bayes (NB), achieved an accuracy of 88.53%, LightGBM demonstrated superior robustness in addressing class imbalance and enhancing detection accuracy. This work highlights the potential of LightGBM in enhancing PON security and enabling intelligent, resilient broadband networks.

Keywords:
Attack detection system;
Classification;
LightGBM;
Machine Learning;
Naïve Bayes;

Article History:
Received: September 18, 2024
Revised: January 13, 2025
Accepted: February 12, 2025
Published: September 1, 2025

Corresponding Author:
Nadiatulhuda Zulkifli
Universiti Teknologi Malaysia,
81310 UTM, Johor, Malaysia
Email: nadiatuhuda@utm.my

INTRODUCTION

Passive Optical Networks (PONs) have emerged as a premier approach for alleviating access congestion challenges in recent years. Their capability to deliver higher transmission speeds, guaranteed consistent quality of service (QoS), and cost effectiveness has solidified their position as the leading fiber-access network option [1, 2, 3, 4, 5, 6]. It functions through tree topology, which connects one point to multiple endpoints, providing user access. In a standard time-division multiplexing (TDM) PON configuration, an optical fiber is passively branched by an optical power splitter, allowing a single fiber to route traffic exchange in the connection linking the optical line terminal (OLT) and the optical network units (ONUs) [7]. Typically, the communication channels connecting these two elements utilize distinct wavelengths: 1490 nm is used for downstream transmissions, while 1310 nm is used for upstream transmissions. Due to the inherently passive design of the PON network, it offers a high level of security, creating substantial difficulties for any potential attackers attempting to intercept the optical signal [5]. For example, Gigabit PON (GPON) incorporates

security measures like data encryption, identity authentication, and key management, along with other functionalities. However, recent studies have shown that attackers have devised multiple techniques, including splitting and bending attacks, to illicitly access a PON network [8]. This setup can potentially be exploited by malicious entities. aiming to disrupt the standard operations related to the ONU within the medium access control (MAC) layer. In these scenarios, rogue ONUs might intercept sensitive information intended for other ONUs, which could result in stealing information.

Every ONU is required to comply and function in accordance with the dynamic bandwidth algorithm (DBA) agreement, which might lead to network vulnerabilities that potentially undermine the security of the DBA mechanism. Avoiding this is crucial for optimal GPON functionality. During the DBA process, a degradation attack attempts to acquire additional bandwidth at the expense of other ONUs rather than causing a complete disruption of GPON operations. Nevertheless, countering degradation attacks remains a difficult challenge. Numerous strategies have been suggested in scholarly literature to counter such network threats. For instance, one method to mitigate IP spoofing and DOS attacks involves labelling network packets and tracking their origin at the perimeter routers. Another method is to block out these spoofed packets at the perimeter routers using hop limit or time to live filter as a criteria [9].

Given that PON operates within access networks, whereas its DBA operates primarily within the medium access control layer, a DoS attack directed within the network and transport layers would notably increase traffic frames in both the downstream and upstream links of a PON [10]. Several other potential attacks include IP spoofing, routing attacks, selective forwarding attacks, session hijacking attacks, port scanning attacks, and distributed denial-of-service attacks. Specifically, an ONU under attack will experience a heightened demand for bandwidth in the upstream shared link. The increased bandwidth demand will decrease the bandwidth available to other normal ONUs. A typical (DBA) scheme for managing upstream bandwidth often falls short in addressing this situation.

The high accuracy of predictions when machine learning (ML) methods with real traces are employed in Next Generation Ethernet Passive Optical Network (NG-EPON)for detecting network traffic has been illustrated in [11]. The proposed approach utilizes a single Long Short-Term Memory (LSTM) model at the location of the Optical Line Terminal (OLT) for forecasting the bandwidth requirements of all ONUs under various network loads. It was demonstrated that applying ML algorithms for traffic prediction enhances performance in the context of NG-EPON. This success was mainly attributable to the ability to gather and utilize knowledge effectively. This method demonstrates that high-performing, intelligent communication strategies can significantly enhance or potentially replace traditional network control in the near future.

Additionally, [12] proposed an intelligent approach for classification and prediction within PON. The authors introduced an advanced classification technique that autonomously and incrementally predicts and categorizes future traffic into various types using LSTM and Gated Recurrent Unit (GRU) models. Similarly, [13] sought to illustrate the detection and classification of events within the PON applications for network traffic monitoring by incorporating Long Short-Term Memory (LSTM) with (a) an ensemble classifier and (b) a neural network, respectively. Also, [14] focused on demonstrating fault detection in PON by applying a Support Vector Machine (SVM) classifier.

Nevertheless, most of the existing DBA algorithms, with only a few exceptions, lack security awareness and tend to overlook potential network attacks, hence security has become an emerging topic in optical access networks. Notable studies on secure bandwidth allocation algorithms include Drakulic et al. [15] and Fadila et al. [16]. However, they do not incorporate ML techniques as security measures, including threat detection and mitigation techniques rely on collision monitoring per ONU. This approach identifies only the ONU with the fewest collisions as a measure of potential threat and imposes penalties accordingly.

Previous studies primarily utilized algorithms such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and mostly LSTMs for classification tasks. While these models demonstrated satisfactory performance in terms of classification accuracy, region of convergence (ROC), and precision, their performance was inferior to modern ensemble techniques like Light Gradient Boosting Machine (LightGBM). For instance, a previous study in [12] employed LSTM and GRU in industrial passive optical networks for a dynamic bandwidth allocation algorithm based on traffic classification. Similarly, the study in [22] explored the synergistic use of XGBoost, TABPFN, and LightGBM for enhancing classification performance. LightGBM's strengths, such as handling class imbalance and

efficiently processing large datasets, enabled superior classification of security attacks in PONs. It outperformed older methods in accuracy, precision, and ROC metrics, reliably distinguishing between attack types and advancing real-world PON security applications.

In summary, the results from studies relating to machine learning methods applied in PON models have potential that are outlined as follows [12, 13, 14, 15]:

1. Supervised learning techniques are commonly used. While K-Nearest Neighbors (KNN), SVM, and Bayesian algorithms have witnessed increased research attention given their prevalence in many studies, limited studies are available on PON models.
2. Most of the supervised learning methods consistently achieve high mean accuracies, exceeding 90% in detection effectiveness across different assessment criteria.
3. In PON implementations, the majority of studies have utilized SVM and Decision Tree (DT) methods.
4. Different kinds of datasets have been employed. Certain studies have utilized data from online sources like Kaggle, whereas some have generated datasets through flow generation methods for use with machine learning algorithms.
5. No studies have identified the main attributes of flow records in PON (including priority and action attributes) for all types of datasets employed in existing machine learning approaches.
6. Precision, recall, and F1-score are the primary frequently employed evaluation metrics for assessing the performance of ML algorithms in most research. Conversely, accuracy and execution time are rarely utilized as performance measures.

In view of the above trends, this paper aims to address the gap in detecting and mitigating various security threats, such as eavesdropping, DoS attacks, masquerading, and Theft of Service (ToS) in PON, by classifying malicious vs normal ONU using ML algorithms. Additionally, we propose a novel approach using Borderline-SMOTE post data processing that can further refine a model's performance on imbalanced datasets, especially after an initial model has been trained. This method focuses on adjusting and enhancing the model's predictions by generating synthetic samples specifically in regions where the model misclassifies minority class instances.

It is an effective strategy for handling class imbalance, particularly when the initial model has difficulty with minority class instances near the decision boundary. By generating synthetic samples in these critical regions and re-training the model, better classification performance through improved recall and more balanced precision can be achieved, ultimately leading to a more robust and reliable model.

The rest of this paper is structured as follows: Section II reviews related work on machine learning algorithms applied in the PON domain. The proposed methods are outlined in Section III. Section IV provides a detailed account of the experimental findings and discussion. In the end, Section V concludes the paper and highlights future directions.

METHOD

The model proposed throughout the study consists of two primary stages: detection and classification. Figure 1 illustrates the proposed model for identifying and classifying malicious ONUs. The initial phase involves distinguishing between malicious and normal ONUs. In this phase, the algorithms assess the impact of a DDoS attack to monitor the behavior of ONUs. Features of a DDoS attack, such as bandwidth usage, are crucial for distinguishing between normal and malicious ONUs. Consequently, the results from these feature-checking methods determine whether the ONUs are normal or malicious. Normal ONUs are passed directly, while malicious ONUs are sent to the next phase for further classification. The algorithms proposed for detecting malicious ONUs, the SVM algorithms, were developed and implemented to boost their effectiveness regarding accuracy and execution time.

The SVM algorithm was selected because it has demonstrated strong performance in prior research across various PON applications [19]. Figure 1 shows the algorithmic steps employed in detecting malicious ONUs. The processes included in the detection stage are as follows.

1. Execute and operate the method.
2. The method examines the characteristics of ONUs.
3. The method examines the priority and bandwidth usage of each ONU.
4. The malicious ONUs are forwarded to the classification algorithm.

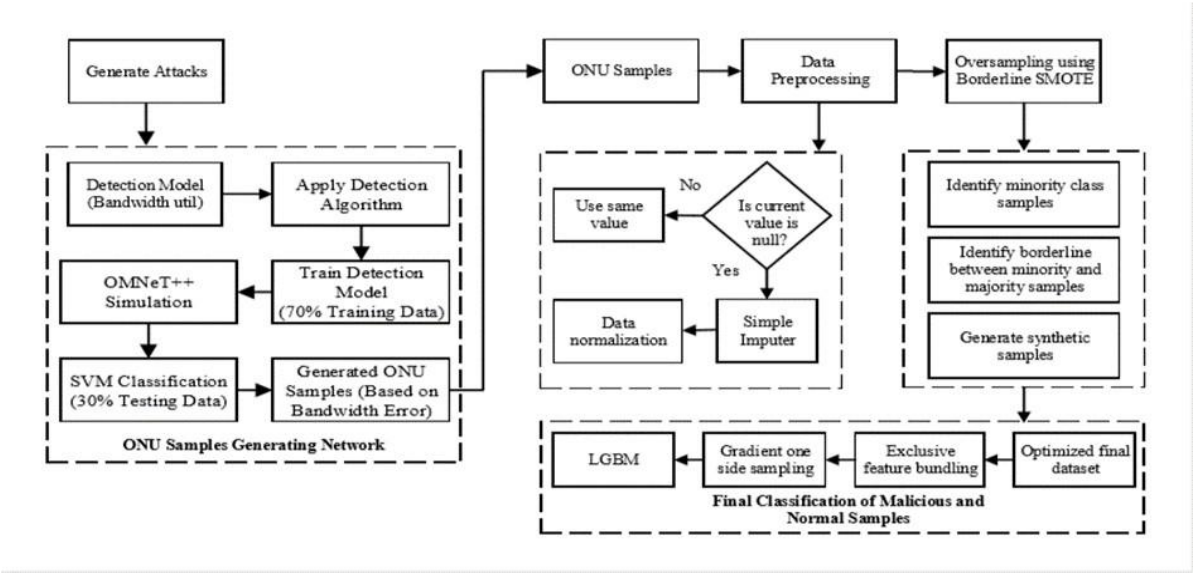

Figure 1. The proposed model for identifying and classifying ONUs

The second stage of the proposed framework involves the classification of ONUs. In this stage, the DDOS attacks identified during the detection stage are analyzed by an algorithm to assess the behavior of the ONUs. The three characteristics of ONUs are priority, bandwidth, and time. Once the checking process is complete, the ONUs is classified. Figure 1 also shows how the LightGBM algorithm is used to classify malicious ONUs. The procedure involved in the classification stage can be outlined as follows:

1. Execute and operate the LightGBM classifier method.
2. The method starts by detecting malicious ONUs.
3. The method examines the priority and bandwidth usage of each ONU.

It then classifies the types of ONUs based on the features assessed in step 3.

Gradient Boosting Algorithm

Gradient boosting is a type of ensemble learning method. Unlike the Naïve Bayes method, where models are created independently, ensemble boosting builds models sequentially, iteratively reducing the errors of previously learned models [18]. It develops a predictive model by combining *M* additive tree models (f_0, f_1, f_2, . . ., f_M) to forecast the outcomes (1).

$$f(x) = \sum_{m=0}^{M} (f_m(x)) \quad (1)$$

The tree ensemble model is optimized by minimizing the expected generalization error L, as described in (2).

$$L = \sum_{i}^{n} (\mathrm{y} - \hat{\mathrm{y}})^2 \quad (2)$$

L represents a loss function that quantifies the difference between the target value yi and the predicted value $\hat{y}$ for a given data point. There are three main motivations to use an ensemble-based approach.

Statistical

Combining and averaging multiple learners enhances data learning and reduces the risk of selecting inappropriate classifiers.

Computational

During learning, finding a local optimum to accurately represent data, such as decision boundaries, is computationally challenging. For instance, neural networks use gradient descent to minimize the loss function, starting from a single point. Ensemble methods, however, leverage multiple starting points for local searches, enabling more accurate estimations of functions like decision boundaries compared to individual classifiers.

Representational

In some cases, a single classifier may struggle to capture complex decision boundaries. Ensemble-based learning addresses this by combining diverse decision boundaries from multiple classifiers [20]. Gradient boosting enhances classifier robustness by reducing variance and bias while mitigating individual shortcomings. This study utilizes LightGBM, a

novel and highly efficient gradient boosting algorithm, to build a more robust model.

Light Gradient Boosting Machines

LightGBM [21] is a gradient boosting method that employs a vertical, leaf-level tree-building approach. LightGBM selects the leaf with the greatest loss reduction for splitting and uses histogram-based methods to identify optimal splits. To improve training, it employs Gradient-based One-Side Sampling (GOSS), which prioritizes data samples with larger gradients while ignoring those with smaller gradients, assuming they have fewer errors and are well-trained [20, 21, 22, 23, 24, 25, 26].

Thus, GOSS recommends ignoring less-informative data points and using the remaining ones to compute information gain for optimal splits. However, this can introduce bias toward samples with larger gradients and distort the original data distribution.

To address this issue, GOSS uses random sampling for low-gradient data while retaining high-gradient points. It compensates by increasing the weights of low-gradient points during information gain calculation. LightGBM uses a unique feature grouping algorithm to address data sparsity. LightGBM effectively handles data sparsity and imbalance by merging mutually exclusive features in a nearly lossless manner, reducing feature count while retaining key information. Using Gradient-based One-Side Sampling (GOSS), it prioritizes samples with higher errors, as those with lower errors are considered adequately trained. Additionally, its "Exclusive Feature Bundling" method enables efficient processing of high-dimensional data, a common challenge in sentiment analysis.

Algorithm LighgtGBM Training Process

Input: Training data $D = \{(x_i, y_i)\}$, $(x_i \epsilon R^n, y_i \epsilon Y)$

Output: Multi-Class Classification

1. Target column contains multiple k classes.
2. For $i = 1$ *to* T
3. Fit the classifier y_i, calculate $P(Y_m)$ using Bayes theorem,
   $P(Y_m|X) = \frac{P(Y_m)P(X|Y_m)}{P(X)}$, 1≤m≤k
4. **Maximum Likelihood Estimation**
   Given $(x_i, y_i)\epsilon Y_i \leftrightarrow P(Y_i|X) > P(Y_m|X), i \neq m$
5. Compute $P(Y_m) = \frac{frequency(Y_m \text{ in } x_{train})}{size(x_{train})}$
6. Using conditional independence assumption, Find;
   $P(X|Y_m) = \prod_{i=1}^{m} P(x_i|Y_m)$
7. Compute $P(x_i|Y_m) = g(x_i, \mu_{y_m}, \sigma_{y_m})$, where μ_{y_m}, σ_{y_m} represent mean And standard deviation.
8. **end for.**

Figure 2. LightGBM Training Procedure

The operation of LightGBM is illustrated by the algorithm depicted in Figure 2. Additionally, it has been shown that LightGBM achieves quicker convergence than other algorithms within the gradient boosting framework. As part of this study, we adopt the identical hybrid method, expected to be detailed in the section on related work.

Naïve Bayes

The algorithm applies Bayes' theorem, assuming variable independence relative to the class variable, a simplification rarely accurate in practice, hence the term "Naive." Nonetheless, it performs efficiently in controlled classification tasks [27][28], as shown in (3) and (4) for probability calculations under known conditions.

$$P(A|B) = P(A)P(B|A) \quad (3)$$

$$\frac{P(A)P(\frac{B}{A})}{P(B)} \quad (4)$$

We develop a LightGBM classifier to differentiate between ONUs (i.e., normal and malicious). The performance of classifiers is widely recognized as being highly dependent on the features used for training. Throughout this research, our goal is to accomplish the following.

1. Develop a classification model using LGBM to analyze ONUs affected by DDOS attacks in terms of bandwidth utilization.
2. Examine how the proposed features perform on our dataset.
3. Explore the correlation between normal and malicious ONUs derived from our dataset.
4. Compare LightGBM with another classifier, specifically Naïve Bayes.
5. Assess the effectiveness of the analysis.

We outline the additions of this study as follows.

1. Create a cutting-edge LightGBM-based model for analyzing DDOS attacks on ONUs.
2. Conduct comprehensive evaluations of classification algorithms using different feature subsets through experiments on our dataset.

Where P(A|B) represents the probability of event A occurring given that event B has occurred. P(A) is the probability of event A occurring. P(B|A) is the probability of the occurrence of event B when event A occurs, P(B) is the probability of event B occurring. The concept behind the Naïve Bayes algorithm is to determine the posterior probability of a data instance ti in a class cj in belonging to a class within the data model.

Algorithm Naïve Bayes Training Process

Input: Training data $D = \{(x_i, y_i)\}$, $(x_i \epsilon R^n, y_i \epsilon Y)$

Output: Multi-Class Classification

1. Target column contains multiple k classes.
2. For $i = 1$ *to* T
3. Fit the classifier y_i, calculate $P(Y_m)$ using Bayes theorem,
   $P(Y_m|X) = \frac{P(Y_m)P(X|Y_m)}{P(X)}$, $1 \leq m \leq k$
4. **Maximum Likelihood Estimation**
   Given $(x_i, y_i) \epsilon Y_i \leftrightarrow P(Y_i|X) > P(Y_m|X), i \neq m$
5. Compute $P(Y_m) = \frac{frequency(Y_m \text{ in } x_{train})}{size(x_{train})}$
6. Using conditional independence assumption, Find;
   $P(X|Y_m) = \prod_{i=1}^{m} P(x_i|Y_m)$
7. Compute $P(x_i|Y_m) = g(x_i, \mu_{y_m}, \sigma_{y_m})$, where μ_{y_m}, σ_{y_m} represent mean And standard deviation.
8. **end for.**

Figure 3. The Naïve Bayes algorithm Training Process

The posterior probability P $(t_i|c_j)$ represents the likelihood that t_i can be assigned the label c_j. P $(t_i|c_j)$ can be determined by multiplying the probabilities of all attributes of the data instance within the data model.

$$\mathrm{P}(t_i \ |c_j)) = \prod_{k=1}^{p} \mathrm{P}\ (t_i \ |c_j) \quad (5)$$

Where P represents the number of attributes in each data instance. The posterior probability is computed for all classes, and the class with the maximum probability is assigned as the label for the instance. The flowchart for this algorithm is shown in Figure 3.

RESULTS AND DISCUSSION

This section discusses the implementation of Naive Bayes and LightGBM and the analysis of the performance of these models based on evaluation metrics. This work uses OMNET++ simulated data network comprising 64 ONUs and one OLT with a fiber distance of 40 km in the ODN. All the models were trained on a synthetic dataset that was created to train a classifier to detect malicious ONUs within a PON based on their bandwidth usage patterns. This dataset includes bandwidth demand profiles from ONUs recorded under normal conditions and during simulated attacks. The response variable for the binary classification is labeled as 0 for normal ONUs and 1 for malicious ONUs. Predictive features include each ONU's average and peak bandwidth demands. Data cleaning involved removing outliers from the bandwidth data. During the simulation, ONUs were labeled as either malicious or normal based on their behavior in attack scenarios.

The Accuracy, Precision and MCC of the proposed methods are computed using the formulas given by (6), (7), and (8), respectively.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Accuracy = \frac{TP + TN}{(TP + TN - FP + FP} \quad (7)$$

$$MCC = TP\ x\ TN - FP\ x\ FN \quad (8)$$

The initial model used for analysis is Naïve Bayes. The model produces predictions based on the validation set. Three distinct metrics were computed for the predictions generated by the model: Precision, Matthews Correlation Coefficient (MCC), and Accuracy. The Naive Bayes classifier achieved a precision of 81.185%, an MCC of 80.503%, and an accuracy of 88.359% using the validation data. Subsequently, the same trained model was used to estimate the labels for the test data. The values for True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) can be determined by plotting the confusion matrix comparing the actual predictions with the values predicted by our model.

The detailed results of the confusion matrix include True Positives (TP=31). The confusion matrix offers a detailed breakdown of the classifier's performance by displaying actual versus predicted classifications, with True Positives (TP=31). The model accurately identified 31 instances as positive. True Negatives (TN=16). The model accurately predicted 16 instances as negative. False Positives (FP=5) were incorrectly predicted as positive, while False Negatives (FN=0) were incorrectly predicted as negative, as illustrated in Figure 4.

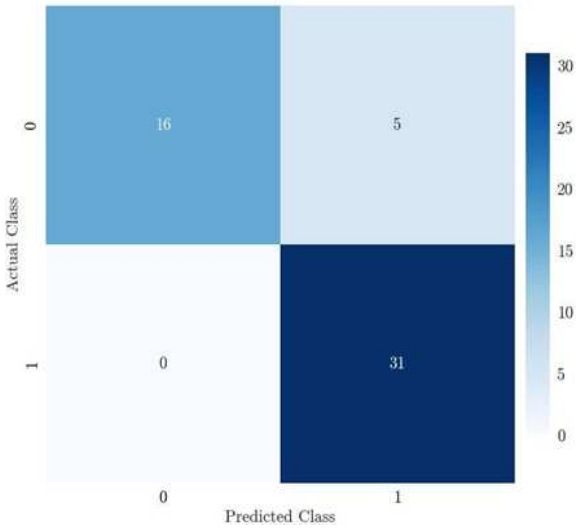

Figure 4. Confusion matrix for actual class and predicted class detection

Roc Curve

The ROC curve demonstrates the balance between sensitivity and specificity by plotting the true positive rate against the false positive rate at various threshold settings, illustrating the trade-off between sensitivity and specificity.

AUC (Area under Curve=0.88)

The AUC value of 0.88 indicates that the model has good discriminative ability, as shown in Figure 5. A model with an AUC closer to 1 is considered excellent, while an AUC closer to 0.5 suggests no discriminative power.

These Performance Metrics are calculated. The bar charts provide a summary of key performance metrics: Accuracy, Precision, and MCC are shown in Figure 6. The NB classifier achieved a precision of 81.185%, an MCC of 80%, and an accuracy of 88.359% using the validation data.

Accuracy

The high accuracy demonstrates that the model is reliable in its predictions.

Precision

The precision value indicates the model's effectiveness in minimizing false positives, which is particularly important in scenarios where false positives are costly.

MCC (Matthews Correlation Coefficient)

A high MCC score indicates a strong overall performance, accounting for true and false positives and negatives. This is useful for a comprehensive perception of the model's performance.

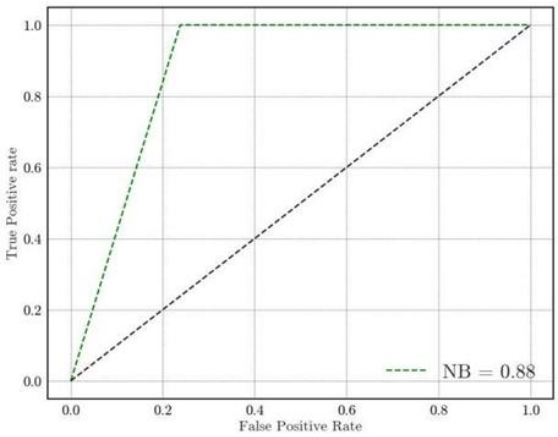

Figure 5. ROC for Naïve Bayes (NB)Classifier with AUC 0.88 for ONU fault

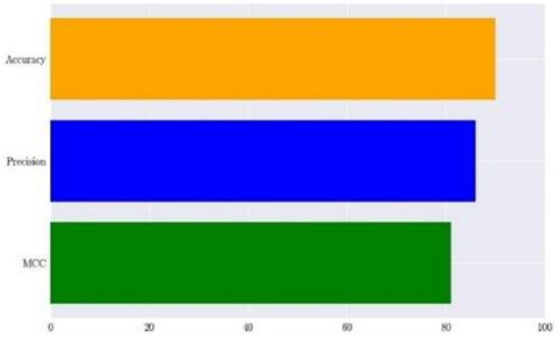

Figure 6. Performance Metrics of Naïve Bayes (NB) Classifier for ONU Fault Detection: Accuracy, Precision, and MCC

From the above discussion, it is observed that the LightGBM model demonstrates excellent performance across all evaluation metrics. Figure 7 shows the confusion matrix for actual and predicted classes using the LGBM model, with true negatives (19), false positives (2), false negatives (0), and true positives (31). The model demonstrates high accuracy, with only two misclassifications, indicating strong predictive performance.

The model's exceptional discriminative ability is highlighted by the ROC curve in Figure 8, where the LightGBM (LGBM) classifier achieves an AUC of 0.95. This high AUC highlights the model's strong ability to distinguish between classes, outperforming the Naive Bayes classifier, which typically has lower AUC scores due to its simplifying assumptions and limitations.

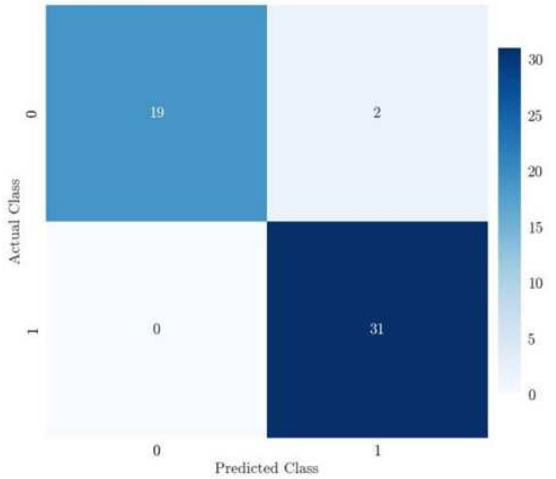

Figure 7. Confusion matrix for the actual class and predicted class.

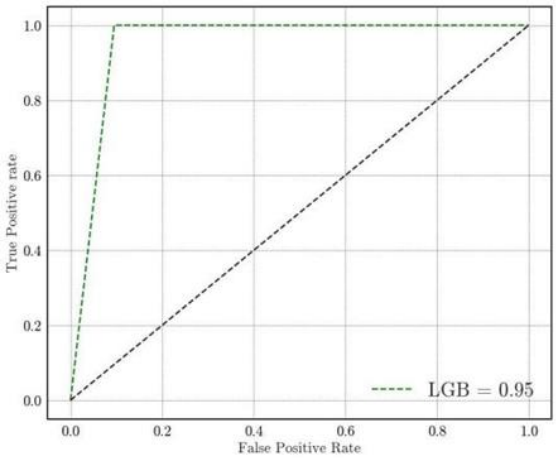

Figure 8. ROC for LightGBM Classifier with AUC 0.95 for ONU fault detection

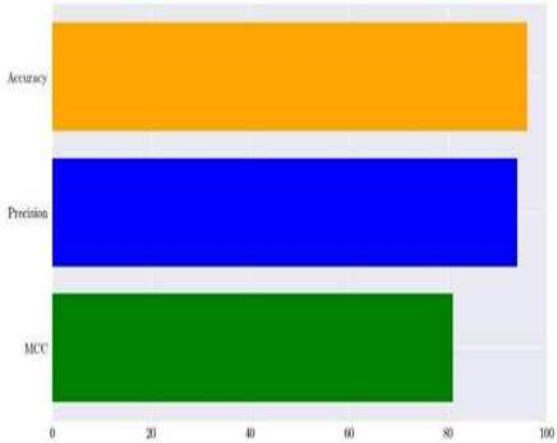

Figure 9. Performance Metrics of LightGBM Classifier for ONU Fault Detection: Accuracy, Precision, and MCC

Figure 9 shows the performance metrics of the LightGBM classifier for ONU fault detection, focusing on Accuracy, Precision, and MCC. The LGBM classifier achieved a precision of 93%, an MCC of 90%, and an accuracy of 95.27% using the validation data. The high accuracy reflects the model's overall correctness in predictions, while the high precision indicates its ability to minimize false positives, making it reliable for fault detection. Additionally, the high MCC score demonstrates the classifier's balanced and robust performance, accounting for all aspects of the confusion matrix, even in scenarios with potential class imbalance. Overall, the results confirm the effectiveness of the LightGBM classifier in accurately detecting ONU faults.

Based on Table 1, the LGBM model outperforms NB in all metrics, making it the superior choice for detecting security threats and classifying ONUs. While NB is simpler and more efficient, its performance is lower.

Table 1. Performance of our models at detecting security threats and classifying ONUs

Model	Accuracy (%)	MCC (%)	Precision (%)
LGBM	95.27	90	93
NB	88.36	80.50	81.19

For optimal threat detection with minimal false positives, LGBM is preferred.

CONCLUSION

This study evaluates the classification of normal and faulty ONUs using LightGBM and Naive Bayes (NB). It also compares the performance of LGBM, an advanced classification method, against NB, one of the earliest classification algorithms, demonstrating that LightGBM achieved superior scores. The findings indicate that the LGBM classifier outperforms others in terms of accuracy, precision, and Matthews correlation coefficient (MCC). LightGBM excels in addressing class imbalance issues, delivering the best results with a detection accuracy of 98.49%, outperforming NB. LGBM shows robustness in accuracy, precision, and MCC, achieving the highest scores among the evaluated techniques. Future work could explore the impact of varying dosage levels on classification performance. To the best of our knowledge, this is the first effort to apply machine learning algorithms for detecting and classifying the nature of ONUs.

ACKNOWLEDGMENT

This work was supported/funded by the Ministry of Higher Education under the Fundamental Research Grant Scheme with reference number of (FRGS/1/2023/TK07/UTM/02/8).

REFERENCES

[1] H. S. Abbas and M. A. Gregory, “The next generation of passive optical networks: A review,” *Journal of Network and Computer Applications*, vol. 67. Academic Press, pp. 53-74, May 01, 2016. doi: 10.1016/j.jnca.2016.02.015

[2] J. F. Balarezo, S. Wang, K. G. Chavez, A. Al-Hourani, and S. Kandeepan, “A survey on DoS/DDoS attacks mathematical modelling for traditional, SDN and virtual networks,” *Engineering Science and Technology*, vol. 31, pp. 101065, 2022, doi: 10.1016/j.jestch.2021.09.011.

[3] S. Sinha and G. Kruthi, “Network layer DoS Attack on IoT System and location identification of the attacker,” *2021 Third International Conference on Inventive*

Research in Computing Applications (ICIRCA), 2021, pp. 22–27, doi: 10.1109/ICIRCA51532.2021.9545071.

[4] A. Sarkunavathi and V. Srinivasan, "A Scrutinized study on DoS attacks in Wireless Sensor Networks and need of SDN in Mitigating DoS attacks," *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, doi: 10.1109/ICCCI50826.2021.9402459

[5] R. Singh and M. Kumar, "A comprehensive analysis for the Performance of Next Generation Passive Optical Network," in *2021 International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON*, Institute of Electrical and Electronics Engineers Inc., 2021, doi: 10.1109/SMARTGENCON51891.2021.9645886.

[6] R. Bonk, "The Future of Passive Optical Networks," *25th International Conference on Optical Network Design and Modelling (ONDM 2021),* 2021, pp. 1–3, doi: 10.23919/ONDM51796.2021.9492398.

[7] F. Obite, E. T. Jaja, G. Ijeomah, and K. I. Jahun, "The evolution of Ethernet Passive Optical Network (EPON) and future trends," *Optik (Stuttgart)*, vol. 167, pp. 103–120, 2018, doi: 10.1016/j.ijleo.2018.03.119.

[8] P. Pinho and D. Camacho, "Analysis tool for passive optical access network," *Journal of Microwaves, Optoelectronics and Electromagnetic Applications,* vol. 20, no. 2, pp. 395–406, Jun. 2021, doi:10.1590/2179-10742021V20I21185.

[9] Y. Li, Y. Zhao, J. Li, X. Yu, Y. Zhao, and J. Zhang, "DDoS Attack Mitigation Based on Traffic Scheduling in Edge Computing-Enabled TWDM-PON," *IEEE Access*, vol. 9, pp. 166566–166578, 2021, doi: 10.1109/ACCESS.2021.3134671.

[10] A. N. Kadhim and S. B. Sadkhan, "Security Threats in Wireless Network Communication-Status, Challenges, and Future Trends," in *2021 International Conference on Advanced Computer Applications, ACA 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 176–181. doi: 10.1109/ACA52198.2021.9626810.

[11] J. A. Hatem, A. R. Dhaini, and S. Elbassuoni, "Deep learning-based dynamic bandwidth allocation for future optical access networks," *IEEE Access*, vol. 7, pp. 9730797318, 2019, doi: 10.1109/ACCESS.2019.2929480.

[12] Y. Wang et al., "Dynamic Bandwidth allocation algorithm based on traffic classification with the aid of LSTM and GRU for industrial passive optical networks," *2023 21st Int. Conf. Opt. Commun. Networks, ICOCN*, 2023, pp. 1–3, doi: 10.1109/ICOCN59242.2023.10236158.

[13] T. Horvath, A. Tomasov, P. Munster, P. Dejdar, and V. Oujezsky, "Unsupervised Anomaly Detection Using Bidirectional GRU Autoencoder Neural Network for PLOAM Message Sequence Analysis in GPON," *Int. Conf. Electr. Comput. Commun. Mechatronics Eng. ICECCME 2022*, November 2022, pp. 1–5, doi: 10.1109/ICECCME55909.2022.9988508.

[14] A. Usman, N. Zulkifli, M. R. Salim, and K. Khairi, "Fault monitoring in passive optical network through the integration of machine learning and fiber sensors," *International Journal of Communication Systems*, vol. 35, no. 9, Jun. 2022, doi: 10.1002/dac.5134.

[15] M. Luqman and A. R. Faridi, "An overview on security issues in internet of things," *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 2018, pp. 6–8, doi: 10.1109/CCAA.2018.8777560.

[16] F. M. Atan et al., "Security enhanced dynamic bandwidth allocation algorithm against degradation attacks in next generation passive optical networks," *Journal of Optical Communications and Networking*, vol. 13, no. 12, pp. 301-311, Dec. 2021, doi: 10.1364/JOCN.434739.

[17] M. Mohinur Rahaman and M. Azharuddin, "Wireless sensor networks in agriculture through machine learning: A survey," *Computers and Electronics in Agriculture*, vol. 197, pp. 106928, 2022, doi: 10.1016/j.compag.2022.106928.

[18] A. Tomasov, M. Holik, V. Oujezsky, T. Horvath, and P. Munster, "GPON PLOAMd message analysis using supervised neural networks," *Applied Sciences*, vol. 10, no. 22, pp. 1–12, 2020, doi: 10.3390/app10228139.

[19] R. Barona and E. Baburaj, "An efficient DDoS attack detection and categorization using adolescent identity search-based weighted SVM model," *Peer-to-Peer Networking and Applications*, vol. 16, no. 2, pp. 1227–1241, Mar. 2023, doi: 10.1007/s12083- 023-01460-6.

[20] F. Alzamzami, M. Hoda, and A. El Saddik, "Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A

Comparative Evaluation," *IEEE Access*, vol. 8, pp. 101840–101858, 2020, doi: 10.1109/ACCESS.2020.2997330.

[21] D. Agrawal, S. Minocha, and A. K. Goel, "Gradient boosting based classification of ion channels," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS 2021)*, 2021, pp. 102–107, doi: 10.1109/ICCCIS51004.2021.9397161.

[22] Sarwo and Y. D. Prabowo, "Enhancing Classification Performance Through the Synergistic Use of XGBoost, TABPFN, and LGBM Models," *Proceedings of 2023 15th International Congress onAdvanced Applied Informatics WinterIIAI-AAI-Winter 2023*, Bali, Indonesia, December 2023, pp. 255–259, 2023, doi: 10.1109/IIAI-AAI-inter61682.2023.00054.

[23] N. Zaeri and R. R. Qasim, "Intelligent Wireless Sensor Network for Gas Classification Using Machine Learning," *IEEE Systems Journal*, vol. 17, no. 2, pp. 1765–1776, Jun. 2023, doi: 10.1109/JSYST.2023.3238357.

[24] S. Gore, Y. Nagalakshmi, P. Knowles, K. G. Gupta, N. S. Jagtap, and R. P. Sali, "Improvised Ensemble Model for Fast Prediction of DoS/DDoS Attacks in Various Networks," *2023 1st International Conference on Cognitive Computing and Engineering Education (ICCCEE)*, 2023, pp. 1–5, doi: 10.1109/ICCCEE55951.2023.10424447.

[25] M. Osman, J. He, F. M. M. Mokbal, N. Zhu, and S. Qureshi, "ML-LGBM: A Machine Learning Model Based on Light Gradient Boosting Machine for the Detection of Version Number Attacks in RPL-Based Networks," *IEEE Access*, vol. 9, pp. 83654–83665, 2021, doi: 10.1109/ACCESS.2021.3087175.

[26] M. R. Youcefi, F. S. Boukredera, K. Ghalem, A. Hadjadj, and C. P. Ezenkwu, "Development of an expert-informed rig state classifier using naive bayes algorithm for invisible loss time measurement," *Applied Intelligence*, vol. 54, no. 17–18, pp. 7659–7673, 2024, doi: 10.1007/s10489-024-05560-5.

[27] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Systems*, vol. 192, p. 105361, 2020, doi: 10.1016/j.knosys.2019.105361.

[28] A. Irwanto and L. Goeirmanto, "Sentiment analysis from twitter about Covid-19 vaccination in indonesia using Naive Bayes and Xgboost classifier algorithm," *SINERGI*, vol. 27, no. 2, June 2023, pp. 145-152, doi: 10.22441/sinergi.2023.2.001