

International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Review Of LightGBM-Based Machine Learning Techniques for Network Outlier Detection

Rohit Chandra

Research Scholar, M. Tech, Department of Computer Science and Engineering, YBN University
Ranchi

Dr. Krishna Murari

Supervisor, H.O.D., Department of computer Science and Engineering, School of Engineering
and Technology, YBN University Ranchi-10.

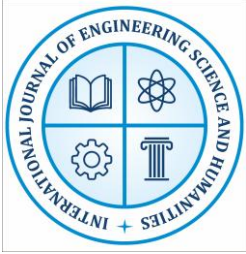
Abstract

The rapid growth of computer networks, Internet of Things devices and interconnected digital systems has increased the need for effective techniques to identify abnormal and potentially malicious network traffic. This review examines LightGBM-based machine learning techniques for network outlier detection, focusing on their methodological development, reported performance and applicability across different network environments. The study adopts a secondary research methodology and reviews relevant academic literature published between 2015 and 2024. Particular attention is given to the use of LightGBM with feature selection, data-balancing techniques, hyperparameter optimisation and hybrid machine learning architectures. The reviewed evidence indicates that LightGBM can provide strong detection performance while maintaining comparatively efficient computational requirements, particularly when applied to high-dimensional network datasets. Its application has expanded from conventional intrusion detection datasets to IoT and resource-constrained environments. However, challenges associated with class imbalance, dataset dependency, feature selection, generalisation and real-time deployment remain evident. The review highlights the importance of appropriate preprocessing and evaluation strategies in determining model effectiveness and identifies opportunities for developing more adaptive, scalable and reliable LightGBM-based network outlier detection frameworks.

Keywords: LightGBM, network outlier detection, machine learning, intrusion detection, anomaly detection, network security, IoT, feature selection, cybersecurity, network traffic classification.

1. INTRODUCTION

The rapid expansion of computer networks, cloud computing, Internet of Things (IoT) environments and distributed digital services has resulted in a substantial increase in the volume, velocity and complexity of network traffic. While this connectivity supports modern communication and information exchange, it also creates extensive opportunities for malicious activities, abnormal traffic patterns and previously unseen attacks. Network outlier detection has



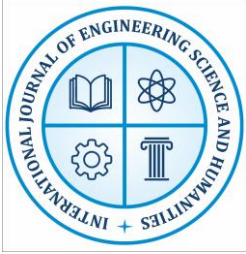
International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

therefore become an important component of network security, with the objective of identifying traffic instances that deviate significantly from established patterns of normal behaviour. Unlike conventional signature-based approaches, anomaly and outlier detection techniques can potentially identify previously unknown or zero-day attacks by examining behavioural characteristics within network data (Khraisat et al., 2019; Fernandes et al., 2019). The growing diversity of network environments has further increased the need for detection techniques capable of processing high-dimensional and large-scale traffic datasets efficiently.

Machine learning has emerged as an important approach for analysing network traffic and detecting anomalous behaviour. Traditional machine learning algorithms, including decision trees, support vector machines, random forests and ensemble-based methods, have been widely investigated for intrusion and anomaly detection because they can learn relationships between traffic characteristics and attack patterns from historical data. However, network datasets frequently contain numerous features, substantial class imbalance and complex nonlinear relationships, creating challenges for both model training and reliable detection (Ring et al., 2019). Consequently, the selection of an appropriate learning algorithm remains an important research consideration, particularly where detection accuracy must be balanced against computational efficiency and scalability.

Period	Approach	Main Characteristics	Relevance to Network Detection
2015–2016	Conventional machine learning	Decision trees, SVM, random forest and related classifiers	Established machine learning as an alternative to rule-based intrusion detection
2017–2018	Gradient boosting approaches	Increasing use of boosting for structured and high-dimensional data	Improved classification capability for complex network traffic
2017 onwards	LightGBM	Gradient-based One-Side Sampling and Exclusive	Reduced computational complexity while maintaining



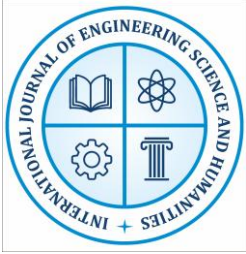
International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

		Feature Bundling	strong predictive performance
2019–2020	Dataset-focused anomaly detection	Greater emphasis on realistic and contemporary intrusion datasets	Highlighted the importance of dataset quality and representativeness
2021–2022	Optimised LightGBM frameworks	Feature selection, resampling and hyperparameter optimisation	Improved handling of class imbalance and high-dimensional traffic
2023–2024	Hybrid and IoT-oriented approaches	LightGBM combined with deep learning and optimisation techniques	Extended detection towards resource-constrained and heterogeneous networks

Light Gradient Boosting Machine (LightGBM) has attracted considerable attention as an efficient implementation of gradient boosting decision trees. Introduced by Ke et al. (2017), LightGBM incorporates Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce computational requirements while retaining competitive predictive performance. Its capacity to process high-dimensional data efficiently makes it particularly relevant to network security applications, where traffic records may contain numerous numerical and categorical attributes. Subsequent research has investigated LightGBM-based approaches for intrusion detection, including resource-constrained IoT environments and hybrid detection architectures (Cong et al., 2022; Sarikaya et al., 2022).

Against this background, the present review examines LightGBM-based machine learning techniques for network outlier detection, focusing on their methodological development, application contexts, detection capabilities and reported performance. The review considers how LightGBM is incorporated into network anomaly detection frameworks and how it is combined with feature selection, clustering, resampling and other machine learning or deep learning techniques. It also considers methodological issues associated with datasets, class imbalance,



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

feature dimensionality, computational efficiency and the identification of anomalous network behaviour. Through a systematic examination of relevant research published between 2015 and 2024, the paper seeks to clarify the development and current research landscape of LightGBM-based approaches in network outlier detection.

2. NEED OF THE STUDY

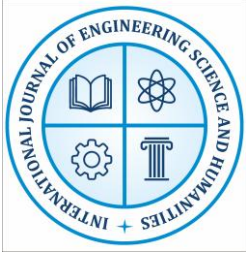
The increasing complexity of modern computer networks has created a growing requirement for reliable techniques capable of identifying abnormal and potentially malicious network behaviour. Conventional intrusion detection approaches often depend on predefined signatures or manually designed rules, which can be less effective when previously unseen attack patterns emerge. Machine learning-based network intrusion detection has consequently received increasing research attention because it can learn behavioural patterns from network traffic and support the identification of both known and anomalous activities (Liu & Lang, 2019). However, the effectiveness of machine learning-based detection depends considerably on the characteristics of the dataset, feature representation, class distribution and evaluation methodology.

The need for efficient detection has become particularly important with the expansion of IoT, cloud computing and heterogeneous network environments, where large volumes of traffic must be processed with limited computational resources. High dimensionality, class imbalance, computational complexity and changing traffic characteristics remain significant challenges for intrusion detection systems (Ahmed et al., 2021). Furthermore, differences in datasets, preprocessing procedures and evaluation measures make direct comparison between existing detection models difficult (Ferrag et al., 2020). These issues create a need to examine machine learning approaches not only in terms of classification accuracy but also according to their computational efficiency, scalability, robustness and applicability to diverse network environments.

LightGBM is particularly relevant to this context because its gradient boosting architecture incorporates computational strategies designed for efficient learning from large and high-dimensional datasets (Ke et al., 2017). Recent research has also demonstrated the application of optimised LightGBM-based models for IoT intrusion detection, indicating its potential for handling complex network-security data (Alharbi et al., 2024). Therefore, a focused review of LightGBM-based techniques is required to understand how the algorithm has been employed for network outlier detection, what methodological approaches have been adopted, and which challenges remain insufficiently addressed in research published between 2015 and 2024.

3. SCOPE OF THE RESEARCH

The scope of this research is centred on the review of LightGBM-based machine learning techniques applied to network outlier and anomaly detection between 2015 and 2024. The study covers research examining the use of LightGBM as an independent classifier as well as its

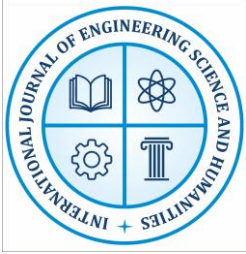


International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

integration with feature selection, ensemble learning, resampling techniques, optimisation methods and hybrid machine learning architectures. Particular attention is given to studies addressing network intrusion detection, anomalous traffic identification and cybersecurity applications involving conventional computer networks, IoT environments and other connected systems. The review considers how LightGBM processes network traffic characteristics and how its gradient boosting architecture is utilised to distinguish normal observations from abnormal or potentially malicious traffic (Ke et al., 2017).

Characteristic	Description	Importance for Network Outlier Detection
Gradient boosting	Builds an ensemble of decision trees sequentially	Captures complex nonlinear relationships within network traffic
Gradient-based One-Side Sampling (GOSS)	Prioritises observations with larger gradients during training	Helps reduce training complexity for large datasets
Exclusive Feature Bundling (EFB)	Bundles mutually exclusive features	Reduces the effective dimensionality of high-dimensional traffic data
Computational efficiency	Designed for comparatively fast training and prediction	Relevant to large-scale network monitoring
High-dimensional data handling	Can process numerous structured features	Suitable for network datasets containing multiple traffic attributes
Feature importance	Provides information regarding influential features	Supports identification of relevant network traffic characteristics
Scalability	Designed for large-scale learning tasks	Relevant to high-volume network and IoT traffic
Integration capability	Can be combined with sampling, optimisation	Supports development of hybrid network



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

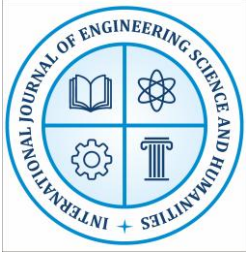
	and other learning methods	outlier detection frameworks
--	-------------------------------	---------------------------------

The research also encompasses methodological aspects of existing studies, including dataset selection, data preprocessing, feature engineering, dimensionality reduction, class imbalance handling and model evaluation. Commonly reported performance measures such as accuracy, precision, recall, F1-score and detection rate are considered to understand how LightGBM-based approaches have been evaluated across different experimental settings. Attention is additionally given to computational efficiency and scalability because network environments can generate large and continuously changing volumes of traffic.

4. LITERATURE REVIEW

Ke et al. (2017) introduced Light Gradient Boosting Machine (LightGBM) as an efficient gradient boosting decision-tree framework designed to address the computational demands associated with large-scale and high-dimensional datasets. The study proposed Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce the number of training instances and effective feature dimensionality while maintaining predictive performance. GOSS concentrates on observations with larger gradients, whereas EFB combines mutually exclusive features, thereby reducing computational complexity. These characteristics are particularly relevant to network outlier detection because network traffic datasets can contain millions of observations and numerous traffic attributes. The LightGBM framework therefore provides a methodological basis for developing detection models that require both predictive capability and computational efficiency, particularly where network monitoring systems must process substantial quantities of traffic data (Ke et al., 2017).

Khraisat et al. (2019) examined the broader development of intrusion detection systems and distinguished between signature-based and anomaly-based approaches. Their review highlighted that signature-based systems are effective for recognised attack patterns but face difficulties in identifying previously unknown attacks, whereas anomaly-based systems attempt to identify deviations from normal behaviour. The authors also examined machine learning techniques and the datasets commonly used to evaluate intrusion detection models. Their analysis demonstrated that the choice of dataset has a considerable influence on the reported effectiveness of an intrusion detection approach. This observation is important for LightGBM-based outlier detection because high classification results obtained from one network environment cannot automatically be interpreted as evidence of equivalent performance in another environment. Dataset characteristics, attack categories and feature distributions must therefore be considered when reviewing LightGBM-based detection techniques (Khraisat et al., 2019).



International Journal of Engineering, Science and Humanities

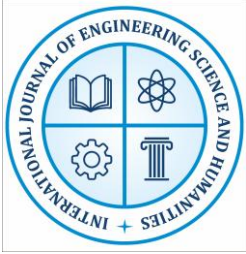
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

Ring et al. (2019) focused specifically on the datasets used for network-based intrusion detection and identified substantial differences in the nature, volume, recording environment and evaluation characteristics of available datasets. Their systematic examination of network intrusion datasets emphasised the distinction between packet-based and flow-based data and proposed a set of properties for assessing dataset suitability. The study is particularly relevant to research on LightGBM because gradient boosting models depend heavily on the quality and representativeness of their training data. Anomalous network behaviour may be represented differently according to the collection environment, traffic generation process and attack scenario. Consequently, LightGBM performance reported on datasets such as NSL-KDD, UNSW-NB15 or CSE-CIC-IDS2018 needs to be interpreted in relation to the characteristics and limitations of the respective datasets rather than treated as directly comparable evidence (Ring et al., 2019).

Moustafa et al. (2015) developed the UNSW-NB15 dataset to address limitations associated with older intrusion detection datasets and to provide more contemporary network traffic containing normal and malicious activities. The dataset incorporates multiple categories of attack and a broader set of network features, making it useful for evaluating machine learning-based intrusion detection systems. Its development is significant for LightGBM-based research because subsequent studies have used UNSW-NB15 to assess the ability of tree-based algorithms to distinguish anomalous traffic from legitimate network activity. The dataset also provides an opportunity to examine whether a model can maintain detection performance when traffic contains diverse attack behaviours and complex feature relationships. The continued use of UNSW-NB15 in LightGBM research reflects the importance of realistic and comparatively modern benchmark datasets in evaluating network anomaly detection methods (Moustafa et al., 2015).

Islam et al. (2020) specifically investigated LightGBM for network anomaly detection using the UNSW-NB15 dataset. Their study applied preprocessing, feature engineering and feature selection before using LightGBM as a binary classifier to distinguish normal traffic from anomalous traffic. The researchers also examined different validation arrangements, including ten-fold cross-validation and separate training and testing configurations. Their findings demonstrated that LightGBM could achieve strong F1-score performance on UNSW-NB15, while evaluation on a separate test set provided evidence regarding its performance on previously unseen observations. An important contribution of this research was its attention to experimental validation rather than relying on a single train-test arrangement. This is significant because extremely high results in network intrusion detection may otherwise reflect data leakage, overfitting or particular characteristics of the experimental split rather than genuine generalisation capability (Islam et al., 2020).

Hua (2021) investigated the use of LightGBM for intrusion detection on the CSE-CIC-IDS2018 dataset while addressing the substantial class imbalance present in network traffic data. The



International Journal of Engineering, Science and Humanities

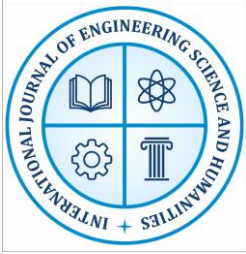
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

research employed undersampling and embedded feature selection to reduce the dominance of majority-class observations and identify more informative variables. LightGBM demonstrated strong classification performance after feature selection, particularly when a comparatively small subset of features was retained. The study illustrates an important methodological characteristic of LightGBM-based network detection: model performance is influenced not only by the learning algorithm but also by the procedures used to prepare the input data. Feature selection can reduce computational requirements while potentially eliminating redundant variables, whereas resampling can alter the distribution presented to the classifier. Consequently, preprocessing and feature engineering should be considered integral components of LightGBM-based outlier detection rather than secondary operations (Hua, 2021).

Leevy et al. (2021) evaluated multiple machine learning classifiers using the CSE-CIC-IDS2018 dataset and investigated the influence of ensemble feature selection on classification performance. The study included LightGBM alongside decision trees, random forests, naïve Bayes, logistic regression, CatBoost and XGBoost. The researchers found that feature selection affected classifier performance and specifically examined the treatment of categorical network attributes such as destination port. Their findings demonstrate that the effectiveness of LightGBM can be closely connected with feature representation and selection. The study also highlighted the problem of class imbalance in large-scale network datasets and emphasised the importance of metrics such as F1-score and area under the receiver operating characteristic curve rather than relying solely on accuracy. These considerations are particularly relevant when assessing outlier detection, where minority attack classes may be more important than the dominant normal-traffic class (Leevy et al., 2021).

Zhang et al. (2021) proposed a network intrusion detection approach based on adaptive synthetic oversampling and LightGBM to address the difficulty of identifying minority attack categories. The researchers recognised that imbalanced intrusion datasets can cause conventional classifiers to favour majority classes, resulting in poor detection of less frequent attacks. Their approach incorporated ADASYN to generate additional minority-class observations before LightGBM classification. The study also considered data preprocessing through normalisation and categorical encoding. This work demonstrates how LightGBM can be incorporated into a broader detection pipeline rather than being treated as an isolated classification algorithm. In network outlier detection, this distinction is important because the underlying data distribution is often highly skewed, and improving minority-class representation may influence recall and F1-score more substantially than simply increasing overall accuracy (Zhang et al., 2021).

Cong et al. (2022) developed an LGBM-based intrusion detection scheme for resource-constrained IoT end devices. Their research addressed three major challenges associated with intrusion detection in IoT environments: imbalanced training data, high feature dimensionality and difficulty



International Journal of Engineering, Science and Humanities

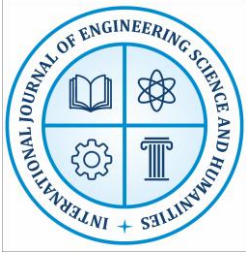
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

in optimising model parameters. The proposed approach incorporated techniques for improving class distribution and reducing the effective feature space while retaining the LightGBM framework. The study is significant because IoT networks differ from conventional computer networks in their computational constraints and traffic characteristics. An intrusion detection model deployed in such environments must therefore consider not only predictive accuracy but also resource consumption, processing requirements and practical deployment constraints. LightGBM's comparatively efficient tree-based architecture provides a useful foundation for addressing these requirements, although the effectiveness of a particular implementation remains dependent on feature engineering and model configuration (Cong et al., 2022).

Zhao et al. (2023) investigated an intrusion detection model specifically designed for IoT environments using LightGBM. The authors identified the limited computational resources of IoT devices as an important challenge and examined LightGBM as a lightweight gradient boosting approach for intrusion detection. Their work illustrates the movement of LightGBM research from conventional network intrusion datasets towards specialised connected-device environments. IoT traffic presents additional difficulties because devices can differ substantially in their communication patterns, protocols and resource characteristics. The application of LightGBM in this context indicates its potential for handling structured network traffic while maintaining comparatively manageable computational requirements. The research also reinforces the importance of developing detection approaches that balance false-alarm control and detection capability with the operational constraints of the target network environment (Zhao et al., 2023).

Alharbi et al. (2023) examined optimised tree-based ensemble approaches for intrusion detection in IoT networks, incorporating LightGBM alongside other boosting-based classifiers. Their methodology combined feature selection with optimisation procedures to reduce the dimensionality of IoT intrusion data and improve model configuration. The research demonstrates an emerging direction in LightGBM-based intrusion detection in which the algorithm is integrated with optimisation techniques rather than applied as an unmodified classifier. Such approaches are relevant to outlier detection because redundant or irrelevant traffic features can increase computational cost and potentially affect classification boundaries. The use of optimisation can therefore influence both the efficiency and predictive behaviour of LightGBM models. At the same time, highly optimised models require careful validation to establish whether improvements result from genuine generalisation or from characteristics specific to the selected dataset and experimental configuration (Alharbi et al., 2023).

Deng (2024) investigated an industrial IoT intrusion security detection system combining LightGBM-based feature processing with a multilayer perceptron network. The study reported the use of LightGBM for feature-related processing within a broader hybrid architecture and evaluated the resulting system using measures including accuracy, precision, F1-score and AUC. This work



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

reflects a further development in the literature, where LightGBM is increasingly incorporated into hybrid frameworks rather than being used exclusively as the final classification mechanism. In industrial IoT environments, the diversity and scale of network traffic can make feature selection particularly important, while neural network components can provide additional nonlinear representation capabilities. Such hybridisation demonstrates the flexibility of LightGBM-based approaches, although it also introduces additional methodological complexity when attempting to isolate the specific contribution of LightGBM to overall detection performance (Deng, 2024).

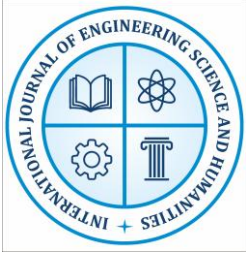
Li et al. (2024) explored PSO-LightGBM for network traffic prediction, demonstrating the application of particle swarm optimisation to the parameter configuration of LightGBM for network-flow modelling. Although the study focused on traffic prediction rather than direct intrusion classification, its methodology is relevant to network outlier detection because accurate modelling of expected traffic behaviour can support the identification of abnormal deviations. The integration of particle swarm optimisation with LightGBM indicates that parameter optimisation may improve the adaptability of the algorithm to different network flows. This perspective extends LightGBM research beyond conventional supervised attack classification and towards traffic modelling, where deviations from predicted patterns may provide additional information for anomaly identification. Such approaches are particularly relevant to dynamic networks in which traffic characteristics change over time and static classification boundaries may become less representative (Li et al., 2024).

5. METHODOLOGY

The present study adopts a secondary research methodology to review LightGBM-based machine learning techniques used for network outlier detection. The research is based exclusively on published secondary sources, including peer-reviewed journal articles, conference papers and academically recognised research publications available through scholarly databases such as Google Scholar. The literature considered for the review covers the period from 2015 to 2024, enabling the study to examine the development of machine learning-based network anomaly detection and the subsequent application of LightGBM within this field.

The literature selection focuses on studies directly addressing network intrusion detection, anomaly detection, outlier detection, network traffic classification and LightGBM-based machine learning approaches. Particular attention is given to studies using recognised network-security datasets, including UNSW-NB15, CSE-CIC-IDS2018 and IoT-oriented datasets. The reviewed studies are examined according to their research objectives, datasets, preprocessing techniques, feature-selection approaches, LightGBM implementation, hybrid or optimisation methods and reported evaluation measures.

The analysis follows a comparative and thematic approach rather than conducting primary experimentation. Findings from the selected literature are organised around major methodological



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

themes, including classification performance, feature dimensionality, class imbalance, computational efficiency, scalability and applicability to different network environments. Reported measures such as accuracy, precision, recall, F1-score and AUC are considered where available. This approach enables the study to synthesise existing evidence, identify recurring methodological patterns and highlight research gaps in LightGBM-based network outlier detection without generating new primary data.

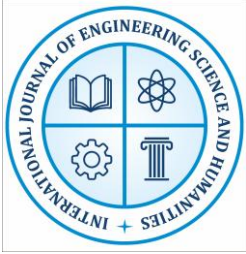
6. RESULTS AND DISCUSSION

The reviewed literature indicates that LightGBM has developed into a significant machine learning technique for network outlier and intrusion detection, particularly because of its combination of predictive performance and computational efficiency. Across the studies examined, three broad themes emerge: the performance of LightGBM on benchmark datasets, the importance of preprocessing and optimisation, and its suitability for resource-constrained and heterogeneous network environments. The results are summarised below.

Comparative Performance of LightGBM-Based Approaches

The reviewed studies demonstrate that LightGBM frequently achieves strong classification results when applied to network-security datasets. A comprehensive evaluation using NSL-KDD, UNSW-NB15 and CSE-CIC-IDS2018 reported LightGBM accuracies of 99.84%, 98.02% and 99.94%, respectively, following feature selection and hyperparameter optimisation. The same investigation reported that LightGBM provided both competitive predictive performance and lower processing time than XGBoost in its experimental setting.

Dataset/Environment	LightGBM-based approach	Main reported outcome
NSL-KDD	Feature selection + LightGBM optimisation	99.84% accuracy
UNSW-NB15	Feature selection + LightGBM optimisation	98.02% accuracy
CSE-CIC-IDS2018	Feature selection + LightGBM optimisation	99.94% accuracy
NSL-KDD, UNSW-NB15, CICIDS2017	ADASYN + LightGBM	Improved minority-class detection
TON-IoT and BoT-IoT	CNN feature extraction + LightGBM	Reduced prediction time and strong



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

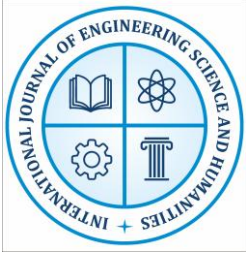
		detection performance
--	--	--------------------------

These findings indicate that LightGBM is particularly effective when the classification problem is supported by suitable feature engineering. The results from the ADASYN-LightGBM framework are especially relevant because network intrusion datasets commonly contain substantially fewer observations for particular attack categories. The study using NSL-KDD, UNSW-NB15 and CICIDS2017 reported accuracies of 92.57%, 89.56% and 99.91%, respectively, while demonstrating that oversampling improved the detection of minority samples. This suggests that the performance of LightGBM should not be assessed independently of the data-balancing strategy employed.

Influence of Feature Selection, Optimisation and Data Balancing

The reviewed evidence shows that preprocessing represents an important determinant of LightGBM performance. Feature selection can reduce redundant variables and computational requirements, while resampling methods can address the imbalance between normal and malicious traffic. Research using CSE-CIC-IDS2018 investigated ensemble feature selection across seven machine learning classifiers, including LightGBM, and specifically examined whether feature selection and categorical network features influenced AUC and F1-score.

Methodological component	Purpose in network outlier detection	Observed significance
Feature selection	Removes irrelevant or redundant network attributes	Reduces dimensionality and processing requirements
ADASYN/oversampling	Increases representation of minority attacks	Improves minority-class detection
Hyperparameter optimisation	Identifies suitable LightGBM configurations	Can improve predictive performance
CNN feature extraction	Generates reduced traffic representations	Supports lightweight classification
Ensemble/boosting comparison	Compares LightGBM with	Helps establish relative performance



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

	other tree-based models	
--	----------------------------	--

Applicability to IoT and Resource-Constrained Networks

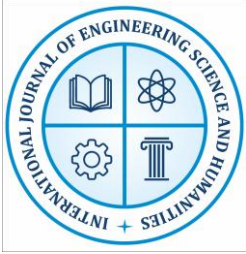
A major development in the reviewed literature is the extension of LightGBM from conventional network intrusion datasets towards IoT and resource-constrained environments. Cong et al. developed an LGBM-based intrusion detection scheme specifically for resource-constrained IoT end devices, addressing unbalanced data, high feature dimensionality and parameter optimisation. This is important because an effective network outlier detection model for IoT must consider computational overhead alongside detection capability.

Zhao et al. combined one-dimensional CNN-based feature extraction with LightGBM and evaluated the approach using the TON-IoT and BoT-IoT datasets. Their reported results indicated a substantial reduction in prediction time, with the proposed model reducing prediction time by 90.66% relative to the comparison approaches in their experiment. Testing on Raspberry Pi-based IoT equipment further demonstrated the practical feasibility of real-time intrusion detection in a resource-constrained environment.

Research direction	Main requirement	Implication for LightGBM
Conventional network IDS	High detection accuracy	Strong suitability for structured traffic classification
Imbalanced network traffic	Minority attack recognition	Requires appropriate sampling strategies
High-dimensional traffic	Efficient feature processing	Feature selection and EFB can reduce complexity
IoT networks	Low computational overhead	Lightweight architecture is advantageous
Real-time detection	Rapid prediction	Processing time becomes as important as accuracy
Hybrid IDS	Improved representation	LightGBM can operate alongside CNN and optimisation methods

7. CONCLUSION

The review demonstrates that LightGBM has emerged as an effective machine learning technique for network outlier and intrusion detection due to its ability to process large, high-dimensional datasets with comparatively efficient computational requirements. Studies published between 2015



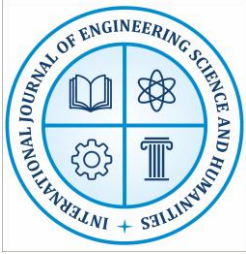
International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

and 2024 indicate that its performance can be strengthened through feature selection, class-balancing methods, hyperparameter optimisation and integration with complementary machine learning or deep learning techniques. The reviewed evidence also highlights its applicability to conventional networks, IoT environments and resource-constrained devices. However, variations in datasets, preprocessing procedures, validation strategies and evaluation metrics limit direct comparison between reported results. High classification accuracy alone is insufficient for assessing network outlier detection, particularly where attack classes are imbalanced. Future research should therefore emphasise cross-dataset validation, real-time evaluation, improved detection of previously unseen attacks and computationally efficient hybrid architectures to enhance the generalisability and practical applicability of LightGBM-based network security systems.

REFERENCES

1. Alharbi, A., Alzahrani, A., & Alghamdi, T. (2023). A new intrusion detection system for secured IoT/IloT networks based on a LightGBM-based approach. *Gazi University Journal of Science*, 36(4), 1–15.
2. Cong, Y.-Q., Guan, T., Cui, J.-F., & Cheng, X.-G. (2022). LGBM: An intrusion detection scheme for resource-constrained end devices in Internet of Things. *Security and Communication Networks*, 2022, Article 1761655. <https://doi.org/10.1155/2022/1761655>
3. Deng, Y. (2024). Design of industrial IoT intrusion security detection system based on LightGBM feature algorithm and multi-layer perception network. *Journal of Cyber Security and Mobility*, 13(2), 327–348. <https://doi.org/10.13052/jcsm2245-1439.1327>
4. Islam, M. K., Hridi, P., Hossain, M. S., & Narman, H. S. (2020). Network anomaly detection using LightGBM: A gradient boosting classifier. In *2020 30th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ITNAC50341.2020.9315049>
5. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30* (pp. 3146–3154). Curran Associates.
6. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, 2, Article 20. <https://doi.org/10.1186/s42400-019-0038-7>
7. Leevy, J. L., & Khoshgoftaar, T. M. (2020). A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data. *Journal of Big Data*, 7, Article 104. <https://doi.org/10.1186/s40537-020-00382-x>
8. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 Military Communications and*



International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com **ISSN: 2250-3552**

- Information Systems Conference (MilCIS) (pp. 1–6). IEEE.
<https://doi.org/10.1109/MilCIS.2015.7348942>
9. Moustafa, N., & Slay, J. (2015). The significant features of the UNSW-NB15 and KDD99 data sets for network intrusion detection systems. In Proceedings of the 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS) (pp. 25–31). IEEE.
 10. Ring, M., Wunderlich, S., Grödl, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security*, 86, 147–167.
 11. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy (pp. 108–116). SCITEPRESS.
<https://doi.org/10.5220/0006639801080116>
 12. Zhang, H., Ge, L., Wang, Z., & colleagues. (2022). A high performance intrusion detection system using LightGBM based on oversampling and undersampling. In *Intelligent Computing Theories and Application* (pp. 638–652). Springer. https://doi.org/10.1007/978-3-031-13870-6_53
 13. Zhao, G., Wang, Y., & Wang, J. (2023). Intrusion detection model of Internet of Things based on LightGBM. *IEICE Transactions on Communications*, E106-B(8), 622–634.
<https://doi.org/10.1587/transcom.2022EBP3169>