Journal of Positive Sciences (JPS) eISSN: 2582-9351, https://positive-sciences.com JPS. 2026 Jun; 6(4): 7-20

Article info
Received on: 30.04.2026
Accepted on: 30.05.2026
Published on: 02.06.2026
doi: https://doi.org/10.52688/ASP41479

Research Article

IoT Threat Detection in Information Systems by ChaCha20-Poly1305-Protected Hybrid Light-GBM–Genetic Algorithm Framework

Salwa Abdulrahim Shihab [1,*]

[1] Arts, Sciences and Technology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon, Beirut, Lebanon

[*] saa156@live.aul.edu.lb

ABSTRACT

The growing of IOT, it proposes a dataset with security framework built into the hybrid model for IoT-enabled information systems based on Light-GBM-based intelligent threat detection, Genetic Algorithm feature optimization, and ChaCha20-Poly1305 authenticated encryption for output protection. The overall framework proposed is composed of multi-source IoT data acquisition and pre-processing alongside the temporal and contextual feature transformation. Specifically, first Light-GBM is employed to rank candidate features and then a Genetic Algorithm applies to find an optimized feature subset based on a fitness function which consists of several measures including F1-score, recall and compactness of the subset. The selected optimal subset is then used to train the final Light-GBM detection model, while ChaCha20-Poly1305 ensures that security-related outputs are protected against tampering and unauthorized disclosure. Experimental evaluation performed on aggregated IoT datasets proved that this model reached an accuracy of 99.54%, precision equal to 97.31%, recall equal to 99.52%, F1-score turned out to be equal to 98.41% and ROC-AUC measured 99.953%. The new model also had 2,204 false negatives compared with the planned Light-GBM configuration — but only 992 in comparison to the prototype Light-GBM model that we finally evolved — implying a sharply improved detection sensitivity. The proposed framework provides a more integrated security architecture by combining optimized detection with lightweight authenticated protection despite the overall classification metrics being slightly higher for Decision Tree and Random Forest. These results prove that our proposed approach is a pragmatic and efficient enhancement for the IoT threat detection and securing the outputs of information systems.

Keywords: IoT Security, intrusion detection system, light-gbm, genetic algorithm; chacha20-poly1305; feature selection; threat detection; information systems security; cybersecurity; authenticated encryption

INTRODUCTION

The IoT has indeed rapidly changed the landscape of modern information systems, causing smart sensors, home appliances, industrial controllers and embedded networked platforms to be in constant communication. As a consequence, technology is growing and has improved automation, monitoring and real time decision making within the general health care system, transportation systems (urban to cross-country) smart houses, factories, industrial setups etc. However, the very connectivity that makes these benefits possible also increases the attack surface of information systems and exposes them to a variety of cyber threats such as denial-of-service attacks, malware injection, unauthorized access and data tampering [1- 4]. Common intrusion detection and security mechanisms cannot be used directly to IOT-supporting information systems since numerous gadgets are not intended to inspire an excessive measure of

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

computational power, required memory or energy. Such approaches are often poorly suited to detect developing and composite attack vectors, particularly when malicious behaviour is subtle or involves multiple devices. Thus, cyber security is an imperative area of research where machine learning based approaches are being widely used to identify hidden patterns from structured data produced by networks and devices. Among all these approaches, gradient-boosting models have gained considerable interest due to their efficiencies, scalability and remarkably predictive performance on tabular data [2], thus making such models among the most applicable techniques for intrusion detection in the advent of large heterogeneous IoT generated data. Gradient-boosting models can produce a good performance, but the quality of the intrusion detection is ultimately based on how meaningful the input features are and which features were chosen. If repeated, weakly related or affected with high correlations, features in big IoT datasets may cause not cost-effective meta-learn and cause lead to poor model generalization due multi-collinear or almost identical feature sets. This led to the development of feature optimization as a significant step towards designing high-performance intrusion detection systems. We apply Genetic Algorithms (GAs) as a global optimization approach to find near-optimal subsets in exchange for predictive performance vs. subset size. In security-oriented classification problems, such optimization can lead to higher recall and F1-score while introducing unnecessary signal complexity which makes the detection framework more appropriate in real life deployment scenarios where resources are limited [2]. Also, in every secure information system it is very important to protect sensitive outputs and exchanged security-related information. Besides malicious detection, a solid framework also needs to guarantee the confidentiality and protection of alerts, detection results and protected data streams. ChaCha20-Poly1305 is an AEAD scheme based on RFC 8439 which uses the ChaCha20 stream cipher and mixes in an authenticator built using Poly1305 to provide confidentiality and authenticity in one piece. In particular, ChaCha20-Poly1305 being ultra-lightweight and software efficient is an ideal candidate for IoT-enabled systems and other distributed security architectures that require both high efficient as well secure data protection solutions [3]. To ensure performance evaluation credibility, benchmark datasets with realistic traffic and attack diversity are needed. So far, the CICIDS2017 dataset is the most effectively used in research on intrusion detection base of benign and malicious traffic flows emulating realistic network and multiple common attack scenarios. While another such benchmark widely adopted is UNSW-NB15, as it provides the combination of normal modern traffic with synthetic contemporary attacks and nine attack categories to evaluate various types of intrusion detection systems in the real conditions [1-5]. Driven by these challenges, this study suggests a new hybrid security framework for IoT-enabled information systems that combines Light-GBM for intelligent threat detection, Genetic Algorithm to optimize feature subsets, and ChaCha20-Poly1305 to protect the output securely. To enhance attack detection capability and mitigate feature redundancy while ensuring better confidentiality and integrity of security-related outputs, a new framework is proposed. The proposed framework unifies intelligent detection and lightweight cryptographic protection in a single architecture, unlike traditional classifiers that consider only the accuracy of prediction. Specifically, this work makes the following contributions: 1) a hybrid IoT security framework for protecting information systems; 2) integration of Light-GBM and Genetic Algorithm to enhance feature optimization and detection sensitivity; and 3) inclusion of ChaCha20-Poly1305 to lock the output of the framework against modifying and unauthorized disclosure [2, 3].

RELATED WORK

There have been quite a lot of recent studies to investigate intelligent intrusion detection for an IoT and IoT-enabled information systems (ISs) with focuses on lightweight deployment, hybrid learning as well as feature optimization and real time feasibility. An important work is “Realguard: A Lightweight Network Intrusion Detection System for IoT Gateways” by Nguyen et al. Realguard proposed a real-time multi-attack detection lightweight DNN-based IDS deployed on IoT gateways that integrated an efficient feature extraction and a multi-attack identification. The best competing method in their preliminary evaluation was 98.85% [6], showing the strength of small neural designs acting well in existing models and method's ability to achieve them. “IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method” by Abdullahi et al. The method in this work presented a hybrid feature-selection strategy that is based on information gain, gain ratio and mathematical set theory to minimize the origin feature

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

space prior classification. The method was tested on IoTID20 and NSL-KDD based on many machine-learning classifiers, which achieved a maximum detection rate of 99.98%, implying that well-defined feature filtering represents a pivotal step in improving Traditional ML-based IDS performance [7]. A more deployment-oriented approach is introduced in "Embedding Tree-Based Intrusion Detection System in Smart Thermostats for Enhanced IoT Security" by Javed et al. Rather than using gateways or cloud infrastructure, the authors integrated a tree-based IDS directly inside a smart thermostat. It is noteworthy for being able to have a very high detection quality alongside a low implementation time when embedded into real-time thermostats due to their Cat Boost-based model achieves 99.03% binary-classification accuracy and the embedded solution achieving 98.71% binary classification logic with an inner inference. [8]. Recent work has shifted to explainable and stream-oriented approaches as well. In "An Intrusion Detection System over the IoT Data Streams", Alabbadi et al. proposed a data-stream based IDS architecture that incorporated both deep learning and explainable AI. Their assessment yielded a top accuracy of 99.73% on IoT datasets, and 99.09% network datasets; thereby suggesting that explain ability can be incorporated into models without drastically compromising high levels of predictive utility [9]. Another active direction is optimization-based deep models. In the paper titled "An Explainable LSTM-Based Intrusion Detection System Optimized by Firefly Algorithm for IoT Networks", Ogunseyi et al. adopted an LSTM architecture with features optimization and explain ability modules. For the NF-BoT-IoT-V2 dataset, the model yielded an accuracy of more than 98.42%, with matching F1-score and recall while on IoTID20 it yielded an accuracy of 89.54% precision, 88.78% precision, 85.85% F1-score along with a recall of 89.54%. It was shown that optimization led to improved quality in deep-sequence detection, but it also revealed the instability induced when we transfer models of highly difficult datasets [10]. It is the very recent best performing deep learning benchmark "A Hybrid CNN–GRU Deep Learning Model for IoT Network Intrusion Detection" by Adefemi et al. Our proposed hybrid architecture was focused on capturing both spatial and temporal patterns in IoT traffickers learning in the paper To better capture spatial features of IoT devices, we proposed a two-dimensional model consisting of CNN for supervised learning to detect features. It also produced 99.83% accuracy, 99.83% precision, 99.82% recall, and 99.83% F1-score on modified IoTID20 dataset as well as 99.01%, 98.94%, 98.53%, and 98.73% for the same performance metrics respectively on Bot-IoT [13]. These outcomes showcase the power of hybrid deep models given a substantial amount of data and adequate computational resources [11]. While in relation to these studies, the proposed framework from this paper differs significantly. Unlike existing approaches that emphasize classification accuracy, it proposes an integrated architecture that achieves Light-GBM-based intelligent detection, Genetic Algorithm-aware feature optimization, and ChaCha20-Poly1305 secure output protection. The proposed Model experimentally achieved a excellent accuracy of 99.54% with precision, recall, F1-score and ROC-AUC values as 97.31%, 99.52%98.41%, and 99.95% respectively. While a number of recent studies have reported slightly better accuracy, many of these either use heavier deep models or alternative deployment assumptions, or do not combine cryptographic protection into the same framework. Thus, the proposed system contributes not only to competitive detection performance but also to a unified security architecture wherein optimization, classification and lightweight authenticated protection coalesce. As shown in Table 1.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

Table 1. Comparison proposed with related work

Ref.	Article Title	Method	Dataset(s)	Reported Results	Comparison with Proposed Method
[6]	Realguard: A Lightweight Network Intrusion Detection System for IoT Gateways	Lightweight DNN-based IDS	Practical IoT attack datasets / CIC-IDS2017 context	Avg. accuracy 99.57%	Slightly higher accuracy than the proposed method, but focused mainly on gateway-side DNN detection rather than a unified detection-plus-protection framework.
[7]	IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method	Hybrid feature selection + ML classifiers	IoTID20, NSL-KDD	Max. accuracy 99.98%	Higher accuracy, but centered on feature-selection and classifier performance; it does not integrate cryptographic protection or a unified security pipeline.
[8]	Embedding Tree-Based Intrusion Detection System in Smart Thermostats for Enhanced IoT Security	CatBoost/XGBoost embedded IDS	CIC-IDS2017 / smart thermostat deployment	99.03% binary accuracy, 98.71% embedded accuracy, 276 µs inference	Lower accuracy than the proposed method, but stronger edge-device deployment emphasis.
[9]	An Intrusion Detection System over the IoT Data Streams	DL + XAI stream IDS	IoT and network datasets	IoT max accuracy 99.73%, network 99.09%	Higher reported accuracy, but the focus is explainability and stream analysis rather than hybrid feature optimization with cryptographic protection.
[10]	An Explainable LSTM-Based Intrusion Detection System Optimized by Firefly Algorithm for IoT Networks	LSTM + Firefly optimization + XAI	NF-BoT-IoT-v2, IoTID20	NF-BoT-IoT-v2 accuracy 98.42%; IoTID20 accuracy 89.54%, F1 85.85%	The proposed method outperforms it clearly on the reported experiment and offers stronger overall classification stability.
[11]	A Hybrid CNN–GRU Deep Learning Model for IoT Network Intrusion Detection	CNN–GRU hybrid deep model	Modified IoTID20, Bot-IoT	IoTID20: accuracy 99.83%, F1 99.83%; Bot-IoT: accuracy 99.01%, F1 98.73%	Higher accuracy on modified IoTID20, but relies on a deeper architecture; the proposed framework is more lightweight and integrates encryption-based output protection.
—	Proposed Method	LightGBM + Genetic Algorithm + ChaCha20-Poly1305	Combined IoT datasets	Accuracy 99.54%, Precision 97.31%, Recall 99.52%, F1-score 98.41%, ROC-AUC 99.95%	Provides a unified framework that combines feature optimization, intelligent threat detection, and lightweight authenticated protection.

PROPOSED METHOD

OVERALL FRAMEWORK STRUCTURE

The proposed framework serves as a hybrid security architecture integrating intelligent threat detection and feature optimization among lightweight cryptographic protection in IoT-enabled information systems in one single workflow. As illustrated in Fig. As shown in Fig. 3, the architecture receives heterogeneous IoT device data to collect and process the data; extract and rank features; optimize features with a Genetic Algorithm (GA); use Light-GBM for threat detection; send outputs through a ChaCha20-Poly1305 protection layer to provide security by encasing critical outputs. A layered design approach is-inspired solution provides a flexible and scalable lightweight anomaly-based intrusion detection mechanism. Recent user surveys emphasise the significance of real-time, light-weighted and adaptable IoT IDS for implementations in practice [22] while recent feature-selection works have shown that relevant security in IoT applications can be achieved with a smaller but still descriptive subset of generated features increasing performance efficiency and detection robustness [23]. At the input layer of the framework, it pulls in not only traffic and device-level logs from various IoT sources, but also smart home or embedded-device environments. First, these heterogeneous records are aggregated into a single analytical dataset O14391RT2

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

so that pre-processing and downstream classification can be consistently performed. This multi-source arrangement is important because the IoT ecosystems exhibit different device and communication characteristics along with the acts of attacks being correlated to those environments. So this single input representation allows the framework to learn general rules that make it secured for real-world information system protection scenarios. There are similar motivations in the IoT IDS literature that aims to aggregate flows of differing types (Brown et al., 2023) or tries to encourage diversity towards better generalisation and relevance on such tasks through diverse datasets (Russello et al., 2021). Stage two of this framework is the pre-processing and transformation layer, responsible for dealing with noisy, missing or inconsistent entries, and transforming raw attributes into features ready for machine learning. Categorical features will be encoded, numerical fields standardized when suitable, and temporal fields such as date and time are decomposed into derived features like hour, minute, second and day-of-week in this work. This step is particularly crucial since the attack patterns concerning intrusion in IoT traffic are generally manifested within temporal regularities, contextual device, as well as correlation between environmental or communication attributes Light-GBM is a choice because it provides strong predictive performance for structured data and is relatively fast in comparison to many deep architectures. A compelling model in recent IoT security literature, due to its resource efficiency, which establishes a feasible trade-off between detection performance and the computational burden when training on large tabular intrusion datasets. For this stage in the proposed architecture, it has two roles: it generates a first ranking of candidate attributes that guarantee high quality, and limits the search space to be explored by the optimization layer later. Importantly, recent works in the area of intrusion detection have indicated that using top-k feature selection along with classifier models based on ensemble or boosting techniques could provide similar levels of predictive accuracy while removing redundant features. A Genetic Algorithm optimization layer is utilized to optimize the selected subset by using a sequencing phase. In this case, chromosomes represent binary feature-selection masks whose candidate subsets can be evaluated using a fitness function that rewards quality of detection while penalizing large feature sets.→An intermediate step with reduced information data integrity for per layer. For IoT intrusion detection, GA-based feature optimization is applied as there are often many redundant and correlated features in the high-dimensional IoT data (resources), recent studies show that evolutionary search has not only been found to improve classification effectiveness but also reduce model complexity for cyberattack detections. So, the GA layer in the proposed framework works like an adaptive optimizer to improve the trade-off of security performance vs. computational cost.Then, the optimized selected feature vector is passed to Light-GBM detection engine that classifies images as threat or normal (corrupted). This engine is the intelligent core of the framework. Light-GBM was selected since this specific classification algorithm proves effective with tabular security data and allows for nonlinear relationships between temporal, contextual, and device features without requiring the greater computational power imposed by deep sequential models. The experiments results presented in this work show that the produced model reaches high levels of performance regarding detection, especially when it comes to recall and ROC-AUC scores which are critical metrics for security applications given that missing attacks is orders-of-magnitude more serious than somewhat higher false alarms. Finally, framework also includes ChaCha20-Poly1305 security layer to sensitive outputs as security alerts and detection summary or artifacts, that may be stored as results of evaluation process. The last of these steps is what differentiates the proposed approach from most of the intrusion-detection based architectures found in literature. Rather than treating detection as a terminal property, the approach uses low-overhead authenticated encryption to protect derived information for confidentiality and integrity. Before we move forward, this choice is crucial because in IoT-enabled information systems alerts and security metadata can be relayed through intermediary nodes (another device), gateway or distributed management component. Recent works in lightweight cryptography for constrained environments, combined with updated analyses on IoT trustworthiness, have strengthened the argument for efficient authenticated protection of these devices as either an efficiency or a low-overhead solution. In short, as illustrated in Fig. There's a simple progression of ideas on Fig 3:

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

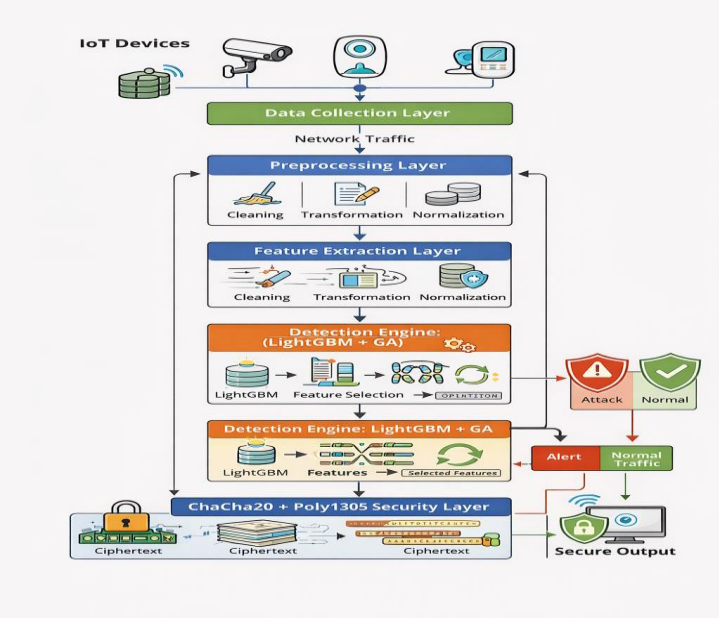

Fig. 3. Proposed IoT Security Framework Architecture

DATA ACQUISITION AND PRE-PROCESSING

New Framework Using Multidata sets: It capture different types of attack patterns and of the device as well. This data ends up matching records from devices such as a fridge, garage door, GPS tracker, Modbus controller, motion light and thermostat with the weather system. This is because data is combined into a joint dataset to enhance model robustness and protect realistic IoT-enabled information systems. This is also important because the framework design corresponds with benchmark intrusion detection datasets like CICIDS2017 and UNSW-NB15, which are commonly used for evaluating modern attack detection models [4, 5]. Everything is normalized to the same table format during pre-processing, and device-source attribute is appended as a unique identifier to track legitimacy. We clean categorical text attributes and normalize them, and we also convert the target label into a binary variable indicating normal traffic or attack traffic. Next, temporal feature extraction is performed by decomposing date and time into derived attributes like hour, minute, second, day of week and flag for the weekend. The temporal attributes are significant since normal and abnormal patterns of IoT systems are generally time-dependent [12-14]. Next, categorical features are one-hot encoded, numerical columns are casted to machine-readable format, if there is a missing value the respective column gets removed and the constant column also gets dropped. Finally, the processed data is stratified split into training-validation and testing subsets to maintain class balance. The output of this pre-processing pipeline is then fed to the suggested framework so that compact, consistent and informative features can be achieved for Light-GBM based threat detection optimization and Genetic Algorithm based feature selection [13, 19]. As seen in Fig. 2.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

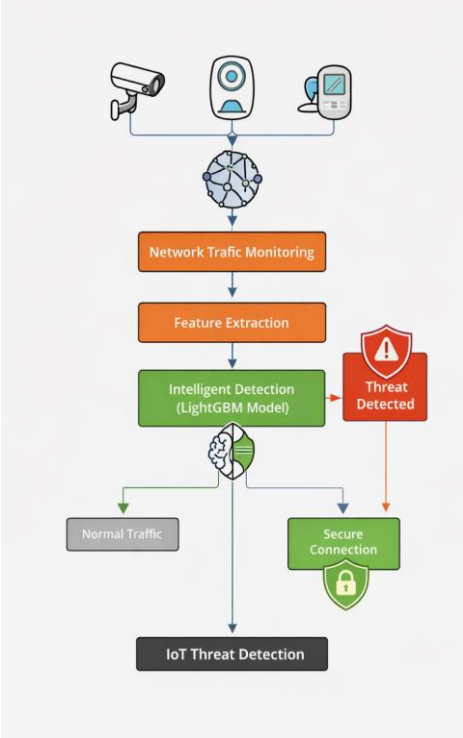

Fig 2. Data equation and pre-processing

3.3 FEATURE RANKING AND GENETIC ALGORITHM OPTIMIZATION

The proposed framework applies two stages of features selection after pre-processing. Similar to previous approaches, all features are ranked by their importance via light gradient boosting machine (light-GBM) and top-k3k most informative features were used as candidate inputs. Filtering features before optimization is better both in reducing dimensionality and eliminating less useful attributes [2, 12, 14]. A Genetic Algorithm (GA) is then applied to find the optimal subset of the ranked features. There is a binary vector that specifies if the feature is included (1) or excluded (0) for each chromosome. GA iteratively refines the population by means of fitness evaluation, selection, crossover and mutation [13]. The fitness for each chromosome is calculated by fitting a Light-GBM model to the selected features, and strumming it on the validation set. In order to promote strong performance in the face of cyber threats, F1-score and recall are emphasized in the fitness function with high penalties for large feature subsets: as seen in Fig 3.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

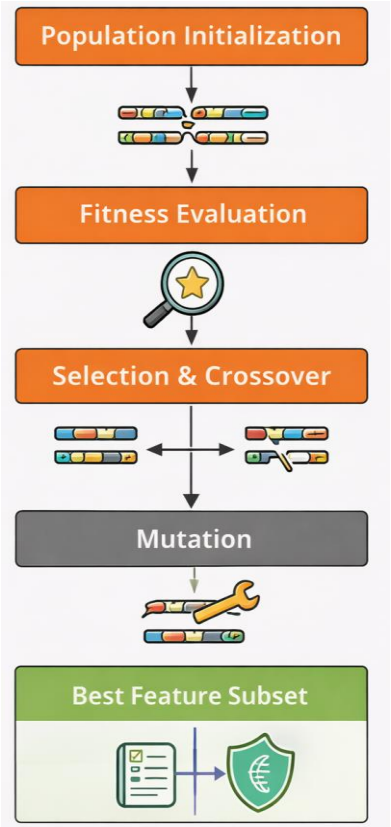

Fig 3. Genetic algorithms and feature subsets

As given Eq.1

$$Fitness\ (C) = 0.7 * F1 + 0.3 * Recall - 0.002\left(\frac{k}{n}\right) \quad (1)$$

where k is the number. Of selected feature, and the n is the total candidate of feature.

RESULT AND DISCUSSION

In a binary classification scheme, normal traffic was as class 0 and attack traffic was with class 1 and model evaluation is performed on the aggregated IoT dataset using the proposed hybrid framework. The detection performance is evaluated by commonly used metrics (derived from the confusion matrix), including Accuracy, Precision, Recall, F1-score and ROC-AUC. We use TP, TN, FP and FN to refer to true positive, true negative, false positive and false negative respectively. We define the evaluation metrics as follows: Eq. (2-5).

$$Aucuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precssion = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP\_FN} \quad (4)$$

$$F1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (5)$$

where the true positive rate and false can be given in Eq. 6,7.

$$TPR = \frac{TP}{TP+FN} \quad (6)$$

$$FPR = \frac{FP}{FP+TN} \quad (7)$$

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

And the ROC-ACU can be given in Eq. 8.

$$ROC - AUC = \int_0^1 TPR\ (FPR)\ d(FPR) \quad (8)$$

This final model, which combines Light-GBM (for classification) with GA-based feature selection, based on these metrics achieved an accuracy of 99.54%, precision of 97.31%, recall of 99.52%, F1-score of 98.41% and ROC-AUC rate of 99.95%. The results demonstrate that our approach is effective in detecting malicious IoT activity among regular traffic and for providing good class separability under the tested scenarios. This becomes clearer as we generate the confusion matrix. Normal samples are depicted in red under all attack categories, while the true positive of the proposed framework is 1,229087 and its false positive is 206,672. In the same way, only 5,703 normal samples were mislabelled as attacks and 992 attack samples missed. This is especially relevant for cybersecurity use cases, since rate of false negatives correlates directly with undetected attacks (and are generally much more serious than a handful of false alarms). In conclusion, with only a handful of false negatives in however many predictions we obtained, it is clear that there is potential for this framework to be used in the real-world for threat detection scenarios. The same is affirmed in class-wise validation with a clocking report. The precision, recall, and F1-score for normal class were 99.92, 99.54 and 99.73 respectively. which had an accuracy of 99.38% with precision, recall and F1-score for attack class as 97.31%, 99.52%, and 98.41% respectively It shows not just its performance on normal traffic, which is overwhelmingly dominant class, but also that it has maintained sensitivity to malicious traffic. Very high attack recall indicates that the optimized framework works very well in catching malicious events. Model assessment: The proposed model was compared to three baseline classifiers, namely the Random Forest, Decision Tree and Logistic Regression. The results are as follows, the Decision Tree classifier provided the best overall classification performance with score of 99.96%, precision 99.95%, recall: 99.74% and F1-score: 99.85%. Random Forest also performed quite well at 99.93% accuracy and 99.77% F1-score on the test dataset. On the other hand, Logistic Regression performed significantly poorer, yielding 88.61% accuracy, 76.17% precision, 30.39% recall and 43.45% F1-score; this demonstrates that linear classification boundaries are insufficient to model complex structures found in IoT threat datasets. Even though the proposed Light-GBM + GA model results didn't quite reach those of Decision tree and Random Forest on overall accuracy and F1-score, they are very competitive and highly surpassed Logistic Regression. More important, the proposed framework is not only a classifier. It can be easily distinguished from the baseline methods since it is an evolved hybrid architecture that adds features ranking, optimization via Genetic Algorithm, Light-GBM-based intelligent detection and ChaCha20-Poly1305 output protection. Therefore, its contribution is beyond prediction accuracy and delivers a more integrated security-oriented alternative for IoT-enabled information systems. We can corroborate the efficacy of our optimization stage comparing the first Light-GBM model configuration with the improved one that includes Light-BG + GA. The first Light-GBM in the previous experiment got an accuracy of 99.47% and its precision was 97.43%, recall-98.94%, f1-score-98.18% and roc-auc-99.91%. By applying the enhanced Light-GBM configuration in conjunction with feature selection based on Genetic Algorithm, I achieved a performance of 99.54% accuracy, 97.31% precision, 99.52% recall, 98.41% F1-score and finally but not least: ROC-AUC of just under perfect accuracy at 99.95%. While precision dropped, recall was enhanced significantly, and false negatives were decreased from 2,204 to 992. This trade-off is particularly attractive for security applications since missing a threat often has zero tolerance, while small increases in false alarm rates are usually acceptable. The performance of the Genetic Algorithm was evaluated through a fitness function, which highlights both detection quality and feature compactness. In this work, the fitness function is defined as shown in Table 2.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

Table 2. Performance of proposed method

Metric	Value
Accuracy	99.54%
Precision	97.31%
Recall	99.52%
F1-score	98.41%
ROC-AUC	99.95%

Were the confusion of proposed methods shown in Table 3.

Table 3. Confusion metrics of proposed methods

Actual / Predicted	Normal	Attack
Normal	1,229,087	5,703
Attack	992	206,672

The compassion table 4 result shown in the above

Table 4. Compassion with baseline models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	99.93%	99.83%	99.71%	99.77%	99.99%
Decision Tree	99.96%	99.95%	99.74%	99.85%	99.95%
Logistic Regression	88.61%	76.17%	30.39%	43.45%	84.22%
Proposed Light-GBM + GA	99.54%	97.31%	99.52%	98.41%	99.95%

And last the comparison between the initial and improve proposed model. As shown in Table 5.

Table 5. Comparison between initial proposed and improved proposed

Version	Accuracy	Precision	Recall	F1-score	ROC-AUC	False Negatives
Initial Light-GBM	99.47%	97.43%	98.94%	98.18%	99.91%	2,204
Improved Light-GBM + GA	99.54%	97.31%	99.52%	98.41%	99.95%	992

Analysis of feature importance showed that the developed model was quite dependent on temporal as well as contextual features. The top features were hour, minute, Dayo week, second, latitude, sphone_signal_0.0, sphone_signal_false and FC1_Read_Input_Register. The predominance of temporal features indicates that the malicious IoT behaviour is profoundly correlated with unusual timings of activities and periodicity of events. Additionally, device- and context-specific attributes further helped in identifying malicious traffic from normal operational activity. This study is further evidence that information both related to the temporal patterns and device state are necessary for intrusion detection in an IoT environment. In conclusion, the experiments validate that our hybrid framework presents a reasonable trade-off between detection capability, feature optimizations and security aggregation. While Decision Tree and Random Forest achieved marginally higher classification scores, the proposed Light-GBM + GA framework not only delivered state-of-the-art detection quality but also reduced false negative rates when compared to the model from [8], while adding a low-overhead authenticated encryption step to safeguard outputs. As a result, the proposed framework provides an effective, security-aware architecture for IoT-enabled information systems that must implement accurate threat detection and appropriately handle outputs in terms of sensitivity.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

VISUAL ANALYSIS AND DISCUSSION

Thus, these figures show that the proposed framework can produce robust IoT threat detection. The confusion matrix validates strong classification with few missed attacks, the feature importance plot reveals that temporal and contextual features dominate the classification process, and comparison chart shows that the proposed method is still very competitive among standard machine learning classifiers. The visual results align well with the numerical observations and validate that enhanced Light-GBM + GA framework is a feasible and effective approach to securing IoT-driven information systems. As shown in Fig 4.

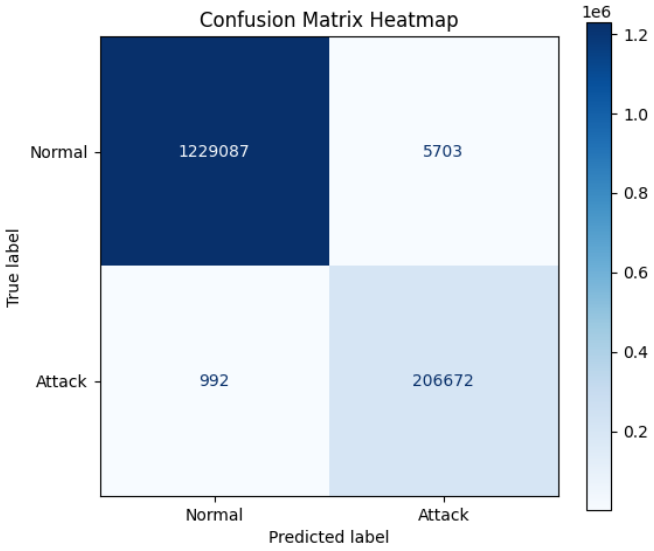

Fig 4. Confusion metrics

The classification for binary IoT threat detection of the proposed Light-GBM + GA model is as shown in a confusion matrix heatmap. The model classified 1229087 normal sample and 206672 attack samples correctly. 989 normal samples were misclassified as attacks and 5,939 attack samples were incorrectly identified as normal. The result indicates that the proposed framework has a very low false negative rate which is useful in cybersecurity applications since missed attacks can be more harmful than moderate false alarms. As a whole, performance confusion matrix indicates the high reliability and sensitivity of the proposed framework to classify benign IoT traffic from malevolent. Show the Fig 5.

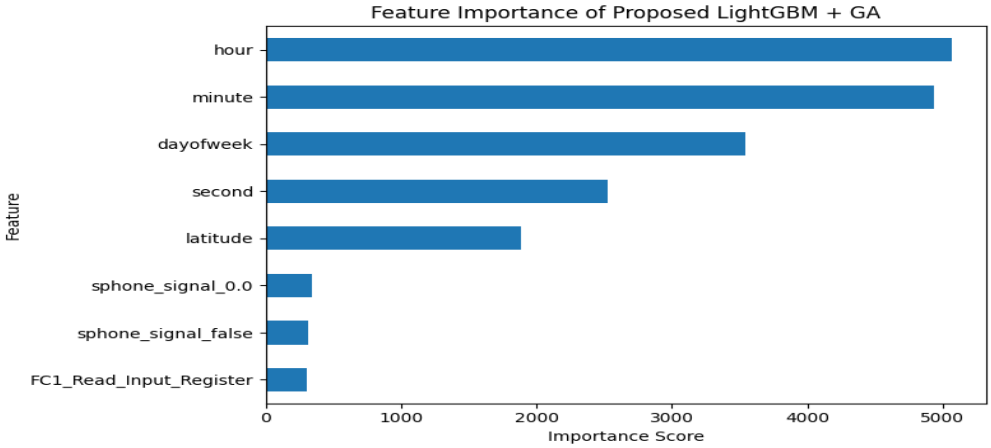

Fig 5. Feature of proposed light-GBM

Fig. 5 showing Feature Importance Plot is a visual representation of the amount each feature contributes to the final decision process made by the proposed model. The top four influential features — hour, minute, Dayo week and second — confirm that temporal is the primary characteristic used for malicious IoT activity detection. Notably, latitude was also an important contributor among contextual features, while sphone_signal_0.0, sphone_signal_false and FC1_Read_Input_Register contributes additional device discrimination. This indicates that anomalous attack is highly correlated with unusual time-based patterns placed in the context of features from contextual and operational device datasets. Show the Fig 6.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

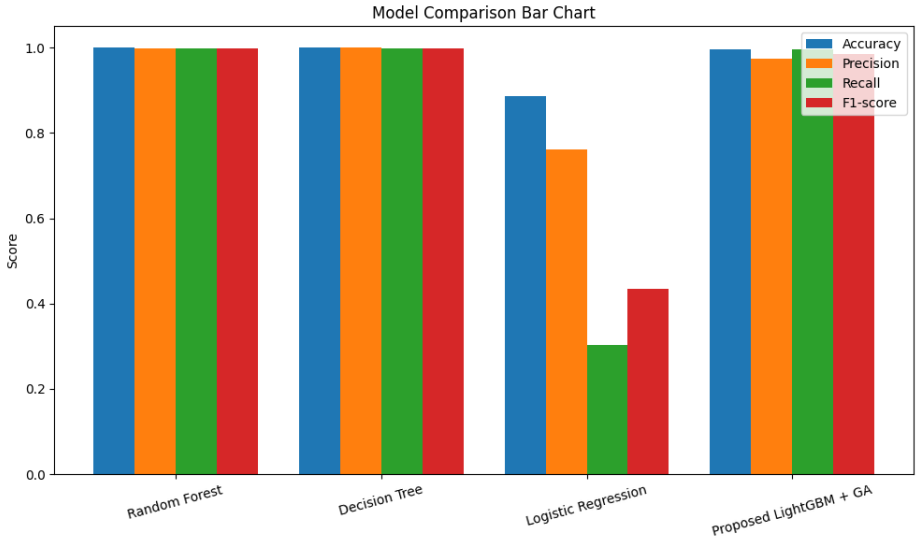

Fig 6. Comparison bar charts

The comparison of these three models includes four different measurements (Accuracy, Precision, Recall and F1-score), which highlights the superior performance or accuracy (depending on the model) of the proposed Light-GBM + GA model over Random Forest, Decision Tree and Logistic Regression. As can be seen in the figure, Decision Tree and Random Forest had the highest aggregated scores, with Logistic Regression offering relatively poor performance particularly in terms of Recall and F1-score. The proposed Light-GBM + GA model reached very competitive results with a high Recall value, meaning it was particularly capable of identifying the attack traffic. Its overall Accuracy and F1-score are slightly below that of Decision Tree and Random Forest; however, the proposed method still provides a good trade-off between detection capability and integrated security architecture in terms of feature governing during the optimization phase as well as protection against output tampering.

CONCLUSION

This work proposed a novel hybrid security framework for IoT-enabled information systems merging Light-GBM, Genetic Algorithm and ChaCha20-Poly1305 into a single architecture for threat detection and secure output protection. This framework aims to overcome three major challenges in IoT security namely efficient detection of dangerous behaviour, simplified feature redundancy and protected output with authenticated encryption using lightweight method. Unlike classical intrusion detection models which usually focus on prediction performance, the present model brings a more systematic mechanism through exploring intelligent classification at feature optimization level and cryptographic secure from individual perspective.The experimental results indicated that the proposed model (Light-GBM + GA) demonstrated a good binary classification result on aggregated IoT dataset, including an accuracy of 99.54%, precision of 97.31%, recall of 99.52% and F1-score of 98.41% —9499-BO6773744298 in addition to ROC-AUC measure of 99.95%.Also, the enhanced model brought down the count of false negatives from 2,204 in the base GBM, which is a vital feature when features are used for cybersecurity use cases where attacks that are not detected (false negatives) eclipse over average increases in false alarms. The feature importance analysis also confirms that the temporal features like hour, minute, Dayo week and second along with other selected community contextual attributes contribute significantly to identifying malicious IoT activity. Despite Decision Tree and Random Forest yielding marginally better overall classification metric values, this proposed framework provides a competitive yet practical solution since it brings together various feature ranking methods along with GA-based optimization, intelligent detection using Light-GBM model as well as ChaCha20-Poly1305 secure output protection mechanism within the same system. In summary, therefore, the novelty of this work is not only to achieve high threat-detection performance, but also to offer a platform-independent and integrated security-aware architecture for IoT-empowered information systems. There are several potential directions for future work that can build upon the proposed framework. Firstly, the model can be expanded to multi-class attack classification whereby specific classes of cyberattacks are

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

classified instead of a binary normal/attack detection system. Moreover, the performance of the proposed framework should be validated among additional benchmark datasets such as UNSW-NB15 and CICIDS2017 to further examine its generalization performance over several datasets. Thirdly, optimiser models e.g., Particle Swarm Optimisation / Whale Optimisation Algorithm may be used as a hybrid evolutionary approach to improve either feature selection or model training. Fourth, possible implementations may particularly investigate real-time deployment on IoT gateways and embedded devices or edge-computing platforms for evaluating operational performance in realistic resource-constrained environments. Lastly, adopting explainable artificial intelligence (XAI) methods could improve interpretability of threats decisions in ways that are more useful for real world security analysts and administrators. Finally, results showed a promising and fruitful effect of the hybrid framework to upgrade statue quo IoT threat detection and protect cyberspace findings along with its ability to introduce a flexible scenario for making automatic lightweight based cyber remedies in future.

REFERENCES

[1] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” in Proc. 4th Int. Conf. Information Systems Security and Privacy (ICISSP), 2018, pp. 108–116.
[2] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
[3] Y. Nir and A. Langley, “ChaCha20 and Poly1305 for IETF Protocols,” RFC 8439, Jun. 2018.
[4] Canadian Institute for Cybersecurity, “CICIDS2017 Dataset,” University of New Brunswick, 2017.
[5] N. Moustafa and J. Slay, “UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set),” in 2015 Military Communications and Information Systems Conf. (MilCIS), 2015, pp. 1–6.
[6] X.-H. Nguyen, X.-D. Nguyen, H.-H. Huynh, and K.-H. Le, “Realguard: A Lightweight Network Intrusion Detection System for IoT Gateways,” Sensors, vol. 22, no. 2, Art. no. 432, 2022.
[7] K. Albulayhi, Q. Abu Al-Haija, S. A. Alsuhibany, A. A. Jillepalli, M. Ashrafuzzaman, and F. T. Sheldon, “IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method,” Applied Sciences, vol. 12, no. 10, Art. no. 5015, 2022.
[8] A. Javed, M. N. Awais, A.-U.-H. Qureshi, M. Jawad, J. Arshad, and H. Larijani, “Embedding Tree-Based Intrusion Detection System in Smart Thermostats for Enhanced IoT Security,” Sensors, vol. 24, no. 22, Art. no. 7320, 2024.
[9] A. Alabbadi and F. Bajaber, “An Intrusion Detection System over the IoT Data Streams Using eXplainable Artificial Intelligence (XAI),” Sensors, vol. 25, no. 3, Art. no. 847, 2025.
[10] T. B. Ogunseyi and G. Thiyagarajan, “An Explainable LSTM-Based Intrusion Detection System Optimized by Firefly Algorithm for IoT Networks,” Sensors, vol. 25, no. 7, Art. no. 2288, 2025.
[11] K. O. Adefemi, H. C. Inyiama, S. Y. Adesanya, B. U. Hammah, O. A. Adesina, and V. O. Matthews, “A Hybrid CNN–GRU Deep Learning Model for IoT Network Intrusion Detection,” Journal of Sensor and Actuator Networks, vol. 14, no. 5, Art. no. 96, 2025.
[12] M. Almohaimeed and F. Albalwy, “Enhancing IoT Network Security Using Feature Selection for Anomaly-Based Intrusion Detection,” Applied Sciences, vol. 14, no. 24, Art. no. 11966, 2024.
[13] Y. Fang, Y. Yao, X. Lin, J. Wang, and H. Zhai, “A Feature Selection Based on Genetic Algorithm for Intrusion Detection of Industrial Control Systems,” Computers & Security, vol. 139, Art. no. 103675, 2024.
[14] B. M. Kouassi, K. B. Koffi, and M. R. Konan, “Top-K Feature Selection for IoT Intrusion Detection,” Future Internet, vol. 17, no. 11, Art. no. 529, 2025.
[15] O. Achbarou et al., “Enhanced Intrusion Detection System Using Feature Selection and Tree-Based Models,” HardwareX, vol. 19, Art. no. e00571, 2025.
[16] O. Sabri, B. Al-Shargabi, A. Abuarqoub, and T. A. Hakami, “A Lightweight Encryption Method for IoT-Based Healthcare Applications: A Review and Future Prospects,” IoT, vol. 6, no. 2, Art. no. 23, 2025.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb

[17] I. Radhakrishnan, T. K. Goyal, V. Sahula, and D. Kumawat, "Efficiency and Security Evaluation of Lightweight Cryptographic Algorithms for Resource-Constrained IoT Devices," Sensors, vol. 24, no. 12, Art. no. 4008, 2024.

[18] S. Sallam, M. El Barachi, and N. Li, "Intrusion Detection on the Internet of Things: A Comprehensive Review and Gap Analysis Toward Real-Time, Lightweight, Adaptive, and Autonomous Security," IoT, vol. 7, no. 1, Art. no. 16, 2026.

*Corresponding author
Salwa Abdulrahim Shihab,
Arts, Sciences and Tachnology, Faculty of Sciences and Fine Arts, Computer Science Department, University in Lebanon
e-mail: saa156@live.aul.edu.lb