The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Development of a Homogeneous Hybrid Machine Learning for Intrusion Detection in the IOT Network

Kosar Ibrahim mirza[1], Dalshad Jaafar Hussein[2], Haval Muhammed Sidqi [3], Nyaz Aziz Ali[4]

[1,2] Department of Database, Computer Science Institute, Sulaimani Polytechnic University, Sulaimani, Iraq

[3] Technical Department of IT, College of Informatics, Sulaimani Polytechnic University, Sulaimani, Iraq

[4] Department of IT, College of Commerce, Sulaimani University, Sulaimani, Iraq

Email: kosar.ibrahim@spu.edu.iq[1], dalsha.jaafar@spu.edu.iq[2], haval.sidqi@spu.edu.iq[3], nyaz.aziz@univsul.edu.iq[4]

Abstract:

In the rapidly evolving landscape of the Internet of Things (IoT), which underpins advancements in e-health, smart homes, e-commerce, and various other digital domains, fog computing emerges as a pivotal IoT platform, bringing to the forefront significant cybersecurity challenges. Traditional intrusion detection mechanisms falter in the IoT context, hindered by the unique constraints of IoT environments, such as limited-resource devices, data imbalance, and specialized protocol stacks and standards. Particularly, the prevalence of unbalanced data in IoT-related network attack datasets compromises the efficacy of conventional intrusion detection systems. Addressing these challenges, this study introduces a novel Group Intrusion Detection Mechanism (Group-based Machine Learning Mechanism for Enhanced Security) GMMES tailored for IoT networks, specifically designed to mitigate malicious activities, with a focus on botnet attacks targeting DNS, HTTP, and MQTT protocols. The GMMES model innovatively integrates correlation-based feature selection, Gaussian mixture model clustering, and ensemble stacking techniques. When benchmarked against contemporary IoT intrusion detection models using the UNSW-NB 15 dataset, based on Attack Detection Precision (ADP) and Early Warning Precision (EWP) metrics, the GMMES model demonstrates superior performance in identifying Dos, Exploits, and Generic attacks compared to other models, including deep neural networks and Adaboost learning algorithms. However, its efficacy in detecting Worms remains consistent with previous models. Furthermore, the incorporation of correlation-based feature selection and parallel processing in the(Group Intrusion Detection Mechanism GMMES model significantly enhances training efficiency, presenting a promising avenue for efficient and effective IoT cybersecurity measures, and the study of the training time of the proposed model also showed that it could reduce the training time by using correlation-based feature selection and parallel processing.

The limitations are incorporating, heterogeneity, optimization, dynamic learning and adversarial Training.

Keywords: Feature Selection via Correlation Analysis, Stacked Ensemble Techniques, Models for Detecting Internet of Things Network, GMMES Model, Gaussian Mixture Model.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

الملخص:

في ظل التطور السريع لإنترنت الأشياء (IoT)، الذي يدعم التقدم في مجالات الصحة الإلكترونية، والمنازل الذكية، والتجارة الإلكترونية، وغيرها من المجالات الرقمية، تبرز الحوسبة الضبابية كمنصة محورية لإنترنت الأشياء، مما يُبرز تحديات الأمن السيبراني الكبيرة. تتعثر آليات كشف التسلل التقليدية في سياق إنترنت الأشياء، بسبب القيود الفريدة لبيئات إنترنت الأشياء، مثل الأجهزة محدودة الموارد، واختلال توازن البيانات، وحزم البروتوكولات والمعايير المتخصصة. وعلى وجه الخصوص، يُضعف انتشار البيانات غير المتوازنة في مجموعات بيانات هجمات الشبكات المتعلقة بإنترنت الأشياء فعالية أنظمة كشف التسلل التقليدية. ولمعالجة هذه التحديات، تُقدم هذه الدراسة آلية جديدة لكشف التسلل الجماعي (GMMES) مُصممة خصيصًا لشبكات إنترنت الأشياء، ومُصممة خصيصًا للحد من الأنشطة الخبيثة، مع التركيز على هجمات شبكات الروبوتات التي تستهدف بروتوكولات DNS وHTTP وMQTT. يدمج نموذج GMMES بشكل مبتكر تقنيات اختيار الميزات القائمة على الارتباط، وتجميع نماذج خليط غاوس، وتكديس المجموعات. عند مقارنته بنماذج كشف تسلل إنترنت الأشياء المعاصرة باستخدام مجموعة بيانات UNSW-NB 15، استنادًا إلى مقاييس دقة اكتشاف الهجوم (ADP) ودقة الإنذار المبكر (EWP)، يُظهر نموذج GMMES أداءً متفوقًا في تحديد هجمات Dos و Exploits و Generic مقارنةً بنماذج أخرى، بما في ذلك الشبكات العصبية العميقة وخوارزميات التعلم Adaboost. ومع ذلك، لا تزال فعاليته في اكتشاف الديدان متوافقة مع النماذج السابقة. علاوة على ذلك، فإن دمج اختيار الميزات القائم على الارتباط والمعالجة المتوازية في نموذج GMMES يُعزز كفاءة التدريب بشكل كبير، مما يُمثل سبيلًا واعدًا لتدابير أمن سيبراني فعّالة في إنترنت الأشياء. كما أظهرت دراسة وقت تدريب النموذج المقترح أنه يُمكنه تقليل وقت التدريب باستخدام اختيار الميزات القائم على الارتباط والمعالجة المتوازية.

الكلمات المفتاحية: اختيار الميزات عبر تحليل الارتباط، تقنيات المجموعات المكدسة، نماذج الكشف عن شبكات إنترنت الأشياء، نموذج GMMES، نموذج الخليط الغاوسي.

پوختە:

لە دیمەنی پەرەسەندنی خێرای ئینتەرنێتی شتەکان (IoT) کە بنەمای پێشکەوتنەکانی تەندروستی ئەلیکترۆنی، ماڵە زیرەکەکان، بازرگانی ئەلیکترۆنی و دۆمەینە دیجیتاڵییە جیاوازەکانی تر، کۆمپیوتەری تەم وەک پلاتفۆرمێکی سەرەکی IoT سەرهەڵدەدات، کە تەحەدای بەرچاوی ئاسایشی ئەلیکترۆنی دەخاتە پێشەوە. میکانیزمە تەقلیدییەکانی دیاریکردنی دەستدرێژی لە چوارچێوەی IoTدا دەلەرزن، بەهۆی سنووردارکردنی ناوازەی ژینگەکانی IoT، وەک ئامێرە سەرچاوە سنووردارەکان، ناهاوسەنگی داتا و ستاک و ستاندارده تایبەتمەندەکانی پرۆتۆکۆڵ. بە تایبەتی، بڵاوبوونەوەی داتا ناهاوسەنگەکان لە کۆمەڵە داتاکانی هێرشی تۆری پەیوەست بە IoT کاریگەری سیستەمی دیاریکردنی دەستدرێژی ئاسایی دەخاتە مەترسییەوە. ئەم توێژینەوەیە بۆ چارەسەرکردنی ئەم تەحەدایانە، میکانیزمێکی نوێی دۆزینەوەی دەستدرێژی گروپی (GMMES) دەناسێنێت کە بۆ تۆڕەکانی IoT دارێژراوە، کە بە تایبەتی بۆ کەمکردنەوەی چالاکییە زیانبەخشەکان دارێژراوە، لەگەڵ گرنگیدان بە هێرشەکانی بۆتنێت کە پرۆتۆکۆڵەکانی DNS، HTTP و MQTT دەکەنە ئامانج. مۆدێلی GMMES بە شێوەیەکی داهێنەرانە هەڵبژاردنی تایبەتمەندی لەسەر بنەمای پەیوەندی، کۆکردنەوەی مۆدێلی تێکەڵەی گاوسی و تەکنیکەکانی کۆکردنەوەی ئەنسەمبڵ یەکدەخات. کاتێک کە بەراورد دەکرێت لە بەرامبەر مۆدێلەکانی دیاریکردنی دەستدرێژی IoT هاوچەرخ بە بەکارهێنانی کۆمەڵە داتاکانی UNSW-NB 15، لەسەر بنەمای پێوەرەکانی وردبینی دیاریکردنی هێرش (ADP) و وردبینی ئاگادارکردنەوەی پێشوەختە (EWP)، مۆدێلی GMMES ئەدای بەرزتر لە ناسینەوەی هێرشەکانی Dos، Exploits و Generic بەراورد بە مۆدێلەکانی تر نیشان دەدات، لەوانەش تۆڕە دەماریيە قوولەکان و فێربوونی Adaboost ئەلگۆریتمەکان. بەڵام کاریگەرییەکەی لە دیاریکردنی کرمەکاندا لەگەڵ مۆدێلەکانی پێشوودا یەکدەگرێتەوە. سەرەڕای ئەوە، جێگیرکردنی هەڵبژاردنی تایبەتمەندی لەسەر بنەمای پەیوەندی و پرۆسێسی هاوتەریب لە مۆدێلی GMMES بە شێوەیەکی بەرچاو کارایی ڕاهێنان بەرز دەکاتەوە، ڕێگایەکی بەڵێندەر بۆ ڕێوشوێنی ئاسایشی ئەلیکترۆنی IoT کارا و کاریگەر دەخاتە ڕوو، و لێکۆڵینەوە لە کاتی ڕاهێنانی مۆدێلی پێشنیارکراوی هەروەها نیشانیدا کە دەتوانێت کاتەکانی ڕاهێنان کەم بکاتەوە بە بەکارهێنانی هەڵبژاردنی تایبەتمەندی لەسەر بنەمای پەیوەندی و پرۆسێسی هاوتەریب. **وشەی سەرەکی:** هەڵبژاردنی تایبەتمەندی لە ڕێگەی شیکاری پەیوەندی، تەکنیکەکانی کۆمەڵەی کۆکراوە، مۆدێلەکان بۆ دیاریکردنی تۆڕی ئینتەرنێتی شتەکان، مۆدێلی GMMES، مۆدێلی تێکەڵەی گاوسی

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

کلیلە وشە: هەڵبژاردنی تایبەتمەندی لە ڕێگەی شیکاری پەیوەندی، تەکنیکەکانی کۆکردنەوەی کۆکراوە، مۆدێلی دیاریکردنی IoT، مۆدێلی GMMES، مۆدێلی تێکەڵەی گاوسی.

I. INTRODUCTION

The rapid growth of information and communication technology has led to the emergence of the Internet of Things (IoT). This technology enables interaction between people and objects in the real world and data and virtual environments(Alghanam, Almobaideen et al. 2023). Smart homes, wise medical care, and smart cities have formed a thriving digital society.

During the design of Internet of Things products, the security of these products is not considered. For example, improving the safety of CCTV cameras causes a significant increase in production costs, which is not cost-effective for manufacturing companies. However, connecting these products to the Internet has left no choice but to pay attention to their security. Although many IoT products do not have enough memory or processing power to perform extensive hacking operations compared to computers or mobile phones(Roets and Tait 2023). But by infecting them with malware and turning them into a part of an attacking botnet network, they can use their capabilities to attack websites and web services and steal information. On the other hand, DDOS attacks create a high volume of traffic on the attacked sites; such attacks force the companies that provide services to the attacked websites to stop providing services to them because maintaining such sites will cost a lot. Based on the mentioned materials, it will be significant to provide an intelligent intrusion detection system to protect network traffic in the Internet of Things(Mostafa, Khalaf et al. 2023). Since it is technically impossible to create computer systems without weak points and security failures, detecting penetration in the Internet of Things network is critical. Intrusion detection systems help system security administrators detect intrusions and attacks.

A. Problem description

The problem in building an expected behavior model is the selection of features that are used as input to build the model. In current models, the security expert determines the input features, and there is no guarantee that all the practical components in intrusion detection are correctly selected. Also, not removing parts unrelated to intrusion can reduce the efficiency of intrusion detection. On the other hand, intrusion detection systems based on machine learning each have advantages and disadvantages(Mhawi, Aldallal et al. 2022).

In machine learning models, the selection of algorithms plays a vital role in achieving favorable outcomes. The model selection is influenced by various factors related to the problem, including the quantity, dimensions, and nature of the data distribution. A low bias and variance model is suitable for learning the typical pattern. In hybrid machine learning methods, basic models are combined as building blocks to create more complex models. These models do not perform well independently and have high bias or variance. We must first select the base models to create a hybrid machine-learning method. A single basic learning algorithm is used in many cases, especially bagging and boosting methods. Therefore, we have several identical basic models trained in different ways, called homogeneous hybrid models. Different types of basic learning algorithms are used in other methods, which are called heterogeneous hybrid models(Talukder, Hasan et al. 2023). This article aims to

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

present a new method using feature selection and collective learning models to improve intrusion detection systems in the Internet of Things network. To investigate the effect of training different basic models and their combination through teaching a metamodel in increasing the accuracy of intrusion detection systems.

B. Contributions

This work introduces a novel stacked ensemble machine-learning framework designed to enhance the prediction of anomalous request patterns. The core contributions of this study are summarized as follows:

1. Integrated Stacking Approach: We developed a sophisticated ensemble method that synergizes the predictive capabilities of diverse machine learning algorithms, including KNN classifier, Logistic Regression, and Support Vector Machine (SVM), by stacking them with a neural network meta-model. This integrated approach leverages the foundational models' individual predictions to drive the meta-model's final prediction, displaying a significant advancement in ensemble learning techniques.
2. Strategic Inclusion of SVM: The SVM plays a pivotal role in our ensemble due to its exceptional classification performance, especially in high-dimensional spaces. Its inclusion is justified by its maximal margin classification, versatility in handling linear and non-linear data through kernel functions, strong generalization properties to prevent overfitting, and its complementary role in the ensemble, enhancing the overall predictive accuracy and robustness.
3. Methodological Advancements: Our methodology involves a systematic training process that includes partitioning the training data, selectively training the foundational models, and utilizing their predictions to train a neural network meta-model. This structured approach ensures a comprehensive utilization of data and model predictions, contributing to the robustness and reliability of the final output.
4. Application to Anomalous Pattern Identification: The ensemble model's effectiveness is demonstrated through its application to identifying patterns in anomalous user request data. By integrating multiple models, our approach achieves superior accuracy and robustness in detecting complex patterns, offering significant implications for security and data analysis fields.

These contributions underscore the effectiveness of our integrated stacking approach and the strategic inclusion of SVM in enhancing the predictive performance of machine learning ensembles, particularly in complex pattern recognition tasks such as the identification of anomalous request patterns.

II. LITERATURE REVIEW

Data confidentiality, authentication, privacy protection, and access control within the Internet of Things network have been investigated in previous research, for example:

The signature-based intrusion detection method uses known attack patterns to identify and detect intrusion(Shaikh and Gupta 2022). This method can detect attacks whose attack patterns are stored in the database. Still, this method cannot detect new attacks whose attack patterns are outside the database. Can use anomaly-based intrusion detection methods. This statistical method tries to find

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

activities that do not match the typical behavior pattern and seem abnormal. In (Asgharzadeh, Ghaffari et al. 2023),the convolutional neural network is chosen to determine the typical behavior pattern. The selection of features using a convolutional neural network leads to computational complexity and delay in recognizing the penetration pattern. There is no guarantee Will select all practical features in intrusion detection correctly. The deep neural network has been used for intrusion detection in reference(Thakkar and Lohiya 2023). Automatic learning of features, accuracy, and generalization power of results in the deep neural network is high, which helps to identify new and hidden patterns in training data. But the high number of features and the correlation between the inputs lead to increased computing time and high cost. The decision tree has been used for intrusion detection in reference(Louk and Tama 2023). One of the advantages of the decision tree algorithm is the elimination of unnecessary comparisons, which increases the speed of intrusion detection. The decision tree learning model is not suitable in cases where the goal is to predict a function with continuous values. The performance of this model is low in issues where we are faced with a large number of categories and a small training sample. Also, producing a decision tree has a high computational cost.

The random forest model has used for intrusion detection in reference)Bhavani, Rao et al. 2019(. High prediction accuracy and the ability to learn non-linear relationships are among the advantages of the random forest model, which helps identify new patterns in training data. Increasing their characteristics and correlation leads to increased computing time and high cost in creating this model.

The artificial immune system (AIS) model has been used to detect intrusion in reference(Sabitha, Gopikrishnan et al. 2022). This model selects appropriate and optimal input parameters for model training. Still, slow convergence to the global optimum and the instability of the results are among the problems of this model.

Hybrid Fuzzy C-Means algorithm (HFCM) and Neural Network algorithm (NN) have been used in the Internet of Things network(Ananthi and Parthipan 2022). One of the advantages of the approach used in this research is the use of fuzzy clustering, which reduces computing time. The clustering of intrusion events is determined only from raw data; therefore, the effort required to set up IDS is diminished. The drawback is the potential to generate false alarms. Like non-linear optimization methods, the fuzzy clustering method depends on the initial values (number of clusters, initial centers, and fixed weights), so the application of this algorithm is minimal.

The use of deep learning as a suitable tool for intrusion detection systems (IDS) and the Internet of Things (IoT) is presented in(Elghamrawy, Lotfy et al. 2022). The fundamental advantage of deep learning over other techniques is that it eliminates most of the feature extraction process while maintaining the system's accuracy, efficiency, and reliability. The disadvantages of the used model are that correlation of input features has yet to be checked so that the results may converge. Increasing the input features leads to an increase in computing time and high computing cost.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

III. PROPOSED MODELS

The proposed GMMES model, an advanced ensemble learning framework, is designed for the detection of botnet attacks within Internet of Things (IoT) networks, leveraging TCP/IP protocols with a focus on MQTT, DNS, and HTTP. This model integrates correlation-based feature selection, Gaussian mixture model clustering, and ensemble stacking to form a comprehensive approach to intrusion detection. The framework unfolds in three pivotal steps: establishing a feature set, selecting features based on correlation coefficients to identify those with the lowest correlations yet significant for discerning legitimate from malicious patterns, and applying an ensemble method for classification.

The ensemble component employs three machine learning techniques: K-nearest neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF). These techniques are orchestrated to distribute data for analysis based on a defined error function, enhancing the detection accuracy by leveraging the strengths of each. The ensemble is further optimized through the AdaBoost algorithm, contributing to an adaptive Network Intrusion Detection System (NIDS) capable of efficiently classifying network records as normal or attack-related.

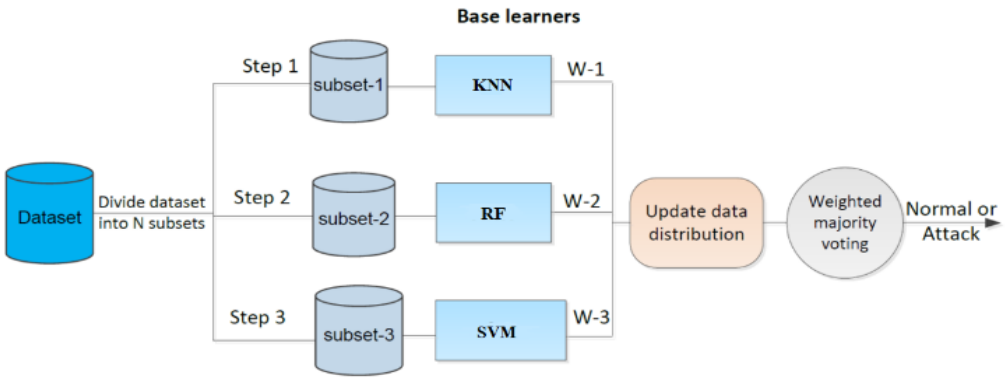

Fig. 1 Structure of the proposed method

The utility of the GMMES model is demonstrated through its application to the UNSW-NB 15 dataset, which comprises 49 features extracted from network traffic. These features, generated through tools like Argus and Bro-IDS, encompass a range of network traffic attributes including flow information, basic packet characteristics, content specifics, temporal behaviors, connection states, and statistical summaries. This rich feature set enables the GMMES model to accurately identify a variety of attack types such as Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms, offering a robust solution for IoT security.

For an in-depth understanding of the 49 features and their relevance to intrusion detection, the "UNSW-NB15_features.csv" file provides detailed descriptions, underscoring the comprehensive nature of this dataset in evaluating NIDS capabilities. This feature set is crucial for the GMMES model's ability to discern complex attack patterns in IoT networks, showcasing its potential as a cutting-edge solution in the domain of cybersecurity (Moustafa and Slay 2015).

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

This framework enables the deployment of elementary learners through a systematic review of observations during the training phase. Observations that are misclassified by the preceding initial learner are assigned greater weight in subsequent iterations of the training process. The fundamental principle of the Boosting technique involves the iterative application of an initial learner to modify the training phase's design, resulting in a series of initial learners over a predetermined number of iterations. Initially, all observations are assigned equal weights, and each iteration utilizes these weighted samples as the basis for the initial learner. In the context of a data distribution, the weight of misclassified observations increases while the weight of correctly classified observations diminishes. The final model produced by the Boosting algorithm is a linear amalgamation of multiple initial learners, each weighted according to their respective performance. In this study, the AdaBoost technique emerges as the predominant method within ensemble learning frameworks for distributing input data across various machine learning approaches. The flowchart and procedural steps of the AdaBoost technique are illustrated in Figure 5, which elucidates its application to the proposed stream features.

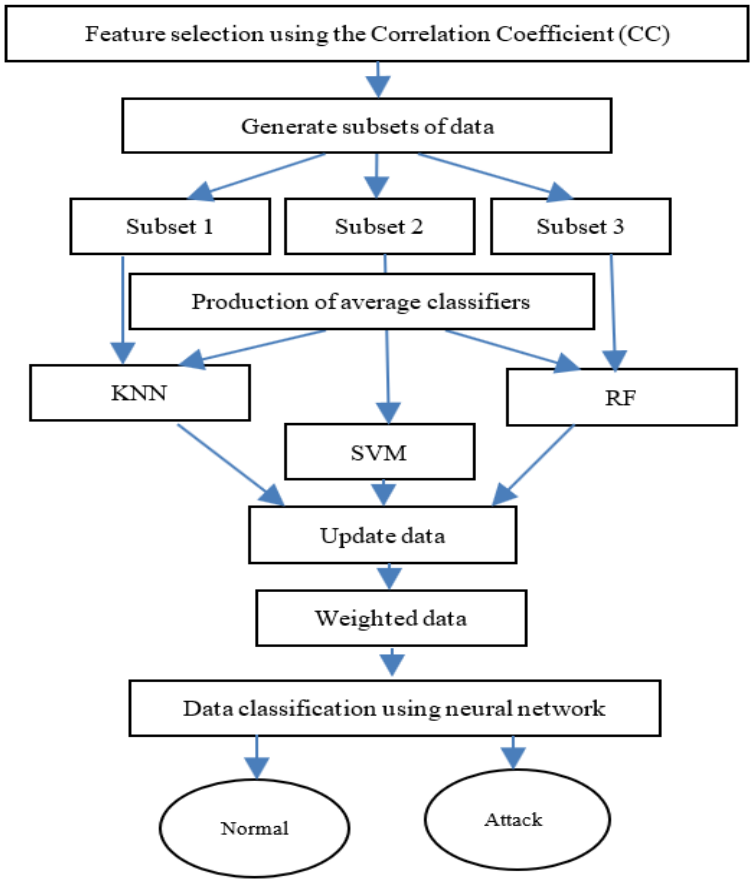

Fig. 2 Flowchart of the proposed method

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

A. Feature Selection Method

Feature selection plays a crucial role in network intrusion detection systems to select essential features and remove unnecessary features, which can help distinguish legitimate and suspicious samples and improve the overall performance of any NIDS. The purpose of feature selection is to reduce the computational cost of NIDS, remove redundant information, enhance the accuracy of NIDS, and help analyze the normality of network data. In this research, the simplest feature selection method was used, especially the correlation coefficient (CC), which calculates the degree of stability (ability) between several features. The features with the lowest ranking N are selected as the essential features, which are transferred to the AdaBoost method to identify the abnormal behavior of DNS and HTTP instances. The CC of features f1 and f2 is calculated as follows.

$$CC(f_1, f_2) = \frac{cov(f_1, f_2)}{\delta_{f_1} . \delta_{f_2}} \quad (1)$$

$$CC(f_1, f_2) = \sum_{i=1}^{N} \frac{(a_i - M_{f_1})(b_i - M_{f_2})}{\sqrt{\sum_{i=1}^{N}(a_i - M_{f_1})^2} . \sqrt{\sum_{i=1}^{N}(b_i - M_{f_2})^2}} \quad (2)$$

In the above equation, δ is the standard deviation of the feature, $cov(f_1, f_2)$ is the covariance of the features, and a_i and b_i represent the values of f_1 and f_2, respectively, and f_1 and f_2 through $M_{f_1} = \sum_i^N a_i / N$ and $M_{f_2} = \sum_i^N b_i / N$ are calculated. In equation (2), the results obtained from the correlation coefficient are in a constant range of [+1,1]. If this value is close to +1 or -1, it indicates a strong correlation between two features f_1 and f_2.

B. Generate subsets of data

In the proposed method, we use Model-Based Clustering to create subsets. In this method, Model-Based Clustering has been assumed for the data; the purpose of clustering based on the model is to estimate the statistical distribution parameters along with the hidden variable introduced as the label of the clusters in the model. According to the number of clusters, for example, k, we present the likelihood function to find the sets as follows

$$L_M(\Theta_1, \Theta_2, \Theta_3, \dots, \Theta_k; t_1, t_2, \dots, t_k) = \prod_{i=1}^{n} \sum_{j=1}^{k} t_j f_j(x_i, \Theta_j) \quad (3)$$

In most cases, the mixed normal distribution is considered for the data. In this case, the mixed distribution is presented in a form where the percentage of mixing (percentage of data belonging to each distribution) is present as a parameter in the model.

$$f(x) = \sum_{j=1}^{k} p_j \Phi(X; \mu_j, \sum j) \quad (4)$$

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

In the above relationship, Φ is the normal distribution with parameters $\boldsymbol{\mu_j}$ and $\sum \boldsymbol{j}$ for the jth distribution, and $\boldsymbol{p_j}$ is the mixing percentage for the jth distribution. Figure 6 shows an example of a bivariate normal mixed distribution.

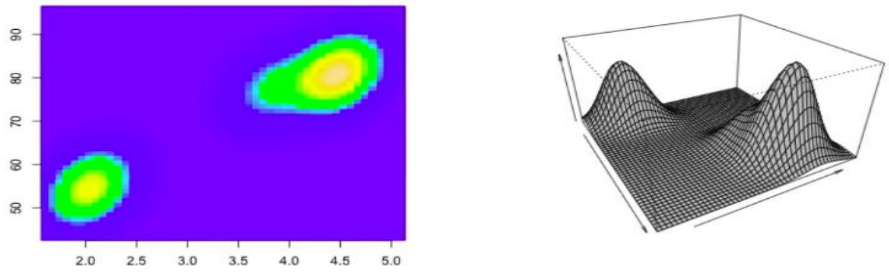

Fig. 3 A bivariate normal mixed distribution

There are several choices for the variance-covariance matrix or Σ. One of these modes is to consider independence and the same variance for all distributions. It is also possible to view the variance of each distribution differently from other distributions and operate without the condition of independence. In the first case, the complexity of the model and the number of estimated parameters will be low. In the second case, the complexity of the model and the number of estimated parameters will be high.

C. Training Model

Input: dataset $\boldsymbol{D} = (\boldsymbol{x_1}, \boldsymbol{y_1}), (\boldsymbol{x_2}, \boldsymbol{y_2}), \ldots, (\boldsymbol{x_m}, \boldsymbol{y_m})$, L is the initial learning techniques, I is the number of iterations and $D_i(j)$ is the weight distribution parameter, Each training sample is given a weight in each stage, and the set of weights of the samples in each stage of repetition t affects the determination of the weak learner ht. This set of weights D(t) is updated in each iteration step. The weight is reduced for samples that are correctly classified and increased for samples that are incorrectly classified. The amount of training error in each step is equal to the total weight of misclassified data, so this method focuses on the misclassified samples in the previous steps. For this reason, this algorithm is adaptive and is named Adaptive Boosting.

In the first stage, all the data have the same weight, and the total weight in each set is equal to one. The data's weight is considered a discrete probability distribution, which is initially a uniform distribution. As a result, all data have the same effect on learning in the first stage.

In step 2, the weight of the data correctly classified by the previous step is reduced. The weight of the data misclassified by the previous step is increased. The data that were wrongly classified in the previous steps have a more significant effect on determining the learner. This procedure is repeated to identify the more difficult samples to classify. The similarity between Adaboost and SVM is in the same focus on complex data, with the difference that Adaboost affects all data in each step. While in SVM, only Support vector samples are effective in the result of classification; however, all data play a role in determining support vector samples. Weight update stage using the following formula:

$$D_n^{(t+1)} = D_n^t \cdot \frac{exp\,[-\alpha_t y_n h_t(x_n)]}{z_t} \qquad (5)$$

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Because we are facing a binary problem, the result of the learner $\boldsymbol{h_t}$ is either +1 or -1. Therefore, if a sample is correctly classified, the product of $\boldsymbol{h_t(x_n)}$ in the data label $\boldsymbol{x_n}$ will equal positive yn and otherwise negative yn. The negative sign inside the exponent causes the sample weight $\boldsymbol{x_n}$ to decrease for the next step in correct classification and increase in the case of wrong classification. The coefficient $\boldsymbol{z_t}$ is a constant value for all samples in each step, which makes the sum of all weights equal to one in each step. The error function is equal to the sum of the weights of misclassified data. The sum of all weights must become one so that the error becomes one if all the data are misclassified. The at value also called the confidence value and depended only on the at error, is obtained by minimizing the training error. The formula of the final hypothesis shows that the larger this value is, the weaker the learner $\boldsymbol{h_t}$ will have a greater effect on the final result. This value of at is obtained using the following loss function:

$$\boldsymbol{G^{AB}(\alpha) = \sum_{n=1}^{N} \exp\{-y_n(\alpha h_t(x_n) + f_{t-1}(x_n))} \qquad (6)$$

where in

$$\boldsymbol{f_{t-1}(x_n) = \sum_{r=1}^{t-1} \alpha_r h_r(x_n)} \qquad (7)$$

In the theory of boosting, it is proven that minimizing the training error leads to reducing the test error, which can be confirmed from practical examples.

IV. EXPERIMENTS

This section discusses the evaluation of the proposed method on the UNSW-NB 15 dataset. Therefore, first, the dataset used for the experiments is introduced. Then, the criteria used to evaluate the methods are introduced. In the following, the methods selected for comparisons are explained, and the proposed method is compared with other methods regarding accuracy and error warning percentage.

A. Dataset

The UNSW-NB 15 dataset's raw network traffic was synthesized using the IXIA PerfectStorm tool at the Cyber Range Lab of UNSW Canberra, blending authentic normal activities with simulated modern attack patterns. The collection of this 100 GB of raw data, such as Pcap files, was facilitated by the tcpdump tool. This dataset encompasses nine distinct attack categories: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms. Utilizing the Argus and

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Bro-IDS tools, along with twelve bespoke algorithms, a total of 49 features were extracted and labeled, detailed in the "UNSW-NB15_features.csv" file.Figure 7.

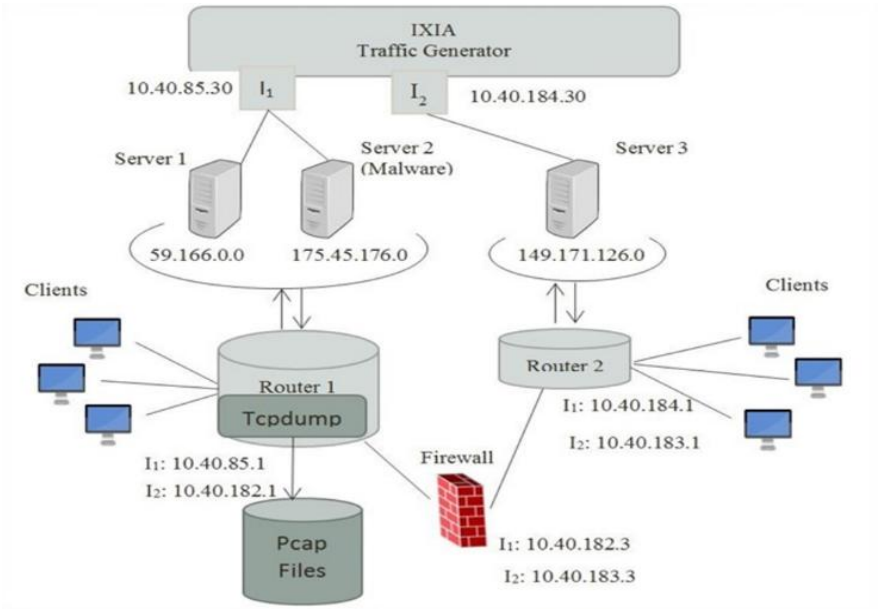

Fig. 4 UNSW-NB 15 dataset creation structure *(Moustafa and Slay 2015)*

The Tcpdump tool captures 100GB of raw traffic (for example, Pcap files) The shortened term (packet capture) is the name of a group of programming libraries that can be used to record network traffic.

This dataset contains nine types of attacks: phasers, analysis, backdoor, DoS, exploit, generic, discovery, shellcode, and worms. Argus uses Bro-IDS tools, and 12 algorithms have been developed to generate 49 class- labeled features. These features are presented in the UNSW-NB15_features.csv file(Moustafa and Slay 2015).

The total number of records is 2,540,044 records. Part of this dataset is set as a training set and test set. The number of records in the training set is 175,341, and 82,332 records with different types of attack and normal records in the test set. Figures 1 and 2 show the regulatory testbed dataset and the UNSW-NB15 feature creation method, respectively(Moustafa and Slay 2015)

B. Experiment setup

The tests conducted on the methods have been performed by MATLAB R2021b software on a processor Intel(R) Core (TM) i5-5300U CPU @ 2.30GHz 2.30 GHz and 8 GB memory. All the tests used the 10-point cross-validation method for evaluation. In this section, the results of the tests are reported. First, the results of experiments without dimensionality reduction on the UNSW-NB 15 dataset with 49 features are presented. The learning results in the group have been analyzed by showing a table and drawing a diagram. Also, the results obtained from the method suggested in the article have been compared.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

C. Metrics

Attack Detection Percentage (ADP): It is equal to the percentage of correctly detected attacks compared to the total number of attacks,

$$\boldsymbol{ADP = \frac{TP + TN}{TP + TN + FP + FN}} \quad (8)$$

TP = correctly diagnosed normal attack – (True Positive)
FP = User request treated as an attack – (False Positive)

Error Warning Percentage (EWP): It is equal to the ratio of the detection of attacks to the sum of the detection of attacks and regular events.

$$\boldsymbol{EWP = \frac{TP}{TP + TN}} \quad (9)$$

D. Results

The proposed GMMES model has been compared with the methods presented in the background with the UNSW-NB 15 dataset based on ADP and EWP criteria of the research including:

Convolution Neural Networks(Kim and Lim 2022) , Decision Tree (DT) (Aswad, Ahmed et al. 2023), Random Forest , Fuzzy Combined C-Means Algorithm (HFCM) (Ismail and Abdullah 2016) & Neural Network Algorithm (NN) (Awujoola Olalekan, Francisca et al. 2020) K_Neighbor (Wang, Wang et al. 2019) and Adaboost Learning(Yang, Liu et al. 2022).

1. COMPARING THE GMMES and CLASSIFICATION METHODS IN THE NORMAL CLASS

Comparing seven classification methods to the GMMES method shows that the proposed method has the highest classification accuracy and speed (Table 1).

TABLE 1
COMPARING THE PERFORMANCE OF THE PROPOSED METHOD (GMMES) WITH OTHER CLASSIFICATION METHODS IN THE NORMAL CLASS

METRICS	CNN	KNN	DT	RF	GB	FCM & NN	ADABOOST	GMMES
ADP	0.75	0.77	0.78	0.77	0.7	0.55	0.8	0.81
EWP	0.76	0.80	0.81	0.80	0.71	0.58	0.82	0.83
TIME	32	10	12.1	17.2	0.46	93	84.2	35

Comparing seven classification methods to the GMMES method shows that the proposed method has the highest classification accuracy and speed.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

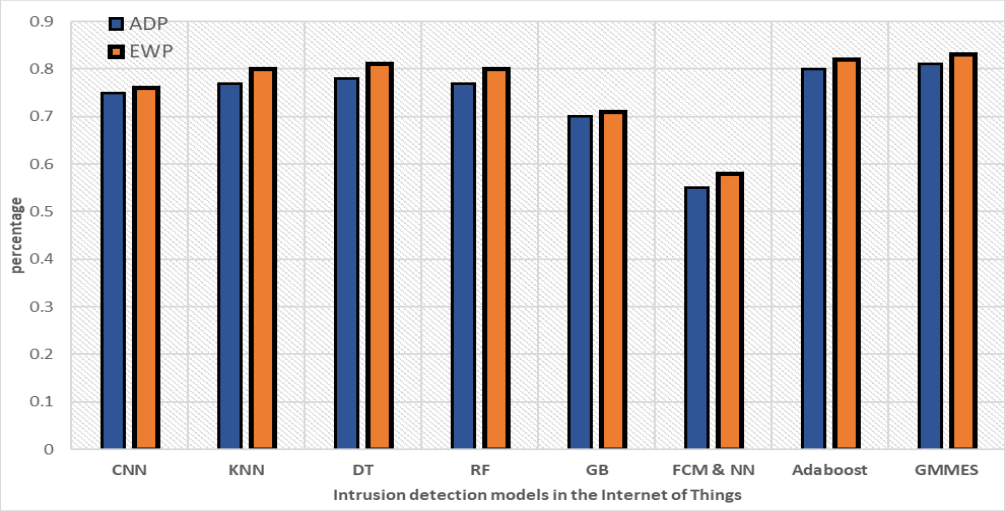

Fig. 5 Comparison diagram of intrusion detection models in normal class.

Each weak learning model randomly selects one of the sets and uses it as training data. Hence, the proposed model is due to the Extraction of the primary dataset having a faster performance than other models.

Based on the obtained results, the accuracy of the convolutional neural network model has reached 0.75%. But due to the complex structure of the convolutional neural network model, training the model in the training data has led to an increase in the calculation time (32 seconds).

The accuracy of the convolutional neural network model in detecting the normal pattern is 76%.

Usually, these models have good accuracy in the training data. Still, in the test data, the model's accuracy has decreased, and due to the high number of features, it faces the problem of overfitting in detecting attacks.

Comparing the evaluated methods in the Normal class based on the learning time shows that the proposed method using the feature selection method based on correlation and parallel processing has reduced the training time in the collective learning algorithm from 22.84 to 35. The proposed method divides the primary data set into subsets using the Gaussian mixture model.

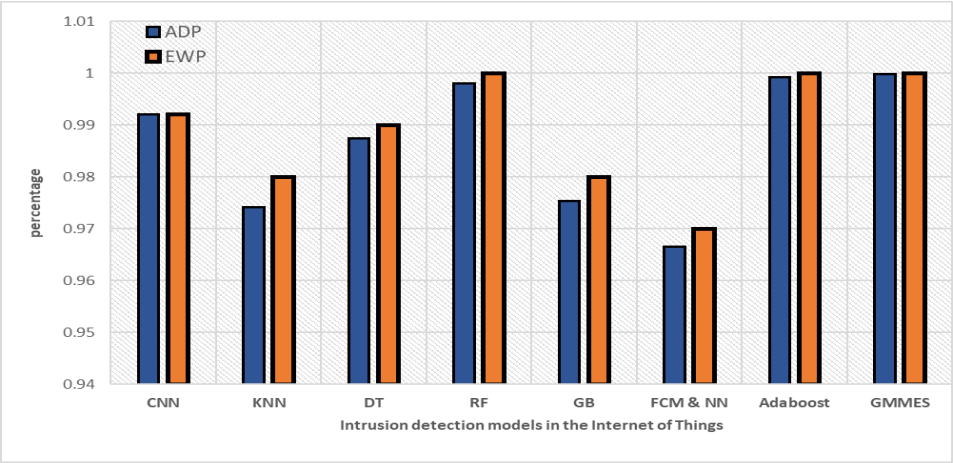

Fig. 6 Comparison time of intrusion detection models in normal class.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

2.COMPARING THE GMMES AND CLASSIFICATION METHODS IN THE Worms CLASS

Table 2 compares the evaluated methods in Worms class based on the evaluation criteria of ADP, EWP, and learning time.

TABLE 2
COMPARING THE PERFORMANCE OF THE PROPOSED METHOD (GMMES) WITH OTHER CLASSIFICATION METHODS IN THE WORM CLASS

METRICS	CNN	KNN	DT	RF	GB	FCM & NN	ADABOOST	GMMES
ADP	0.9921	0.9742	0.9875	0.998	0.9754	0.9665	0.9992	0.9999
EWP	1.00	0.98	0.99	1.00	0.98	0.97	1.00	1.00
TIME	22.41	10.4	5.84	6.7	0.2	23.68	25.29	9.51

The methods evaluated in Worms class show that the classification accuracy of the proposed GMMES method is similar to Adaboost learning (Figure 10).

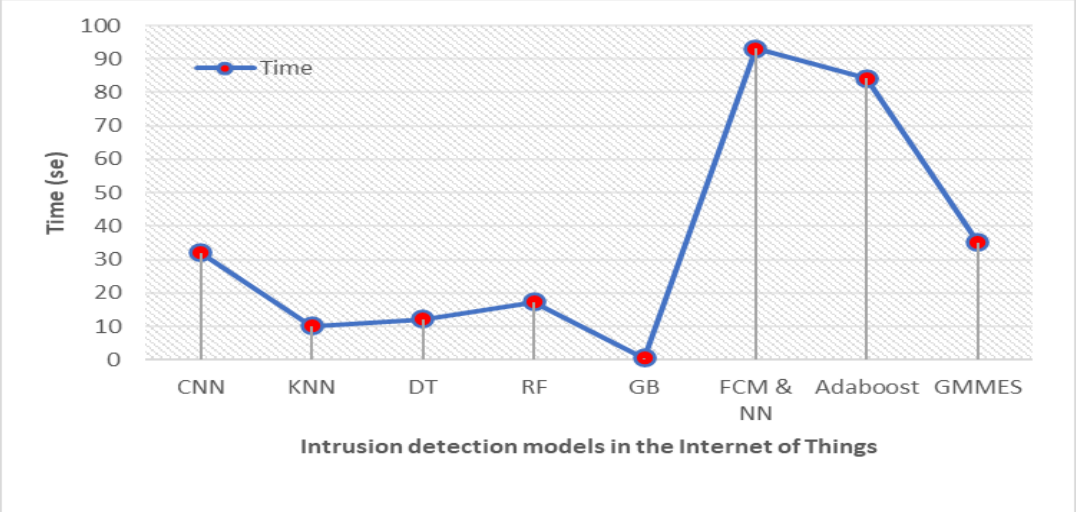

Fig. 7 Comparison diagram of intrusion detection models in Worm class.

The data in figure 11 shows that the GMMES method has reduced the training time in the adaboost learning from 25.29 to 9.51 by using feature selection based on correlation and parallel processing of subsets. The evaluation of the evaluated methods in the Worms class shows that the classification accuracy of the GMMES is similar to that of the Adaboost learning .

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

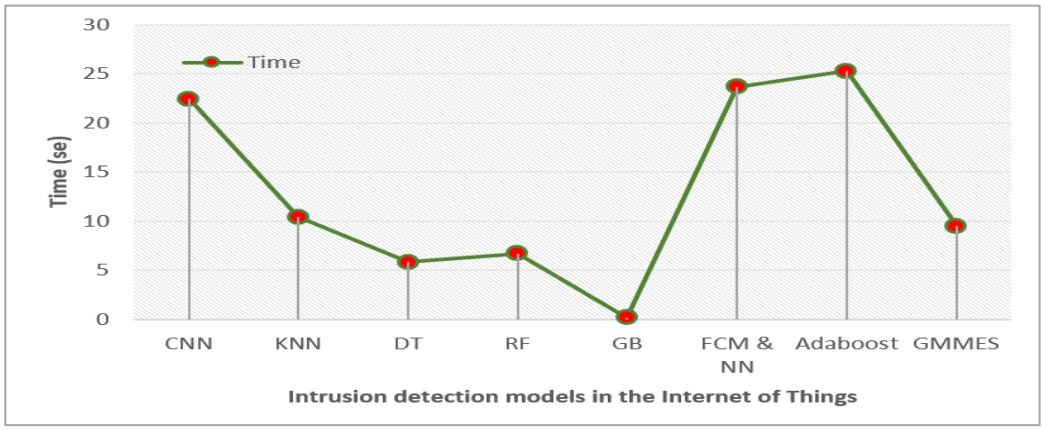

Fig. 8 Comparison time of intrusion detection models in normal class

3.COMPARING THE GMMES AND CLASSIFICATION METHODS IN THE DOS CLASS

Table 3 compares the evaluated methods in Worms class based on the evaluation criteria of ADP, EWP, and learning time.

TABLE 3
COMPARING THE PERFORMANCE OF THE PROPOSED METHOD (GMMES) WITH OTHER CLASSIFICATION METHODS IN THE DOS CLASS

METRICS	CNN	KNN	DT	RF	GB	FCM & NN	ADABOOST	GMMES
ADP	0.9621	0.9712	0.9421	0.951	0.8994	0.5854	0.994	0.9954
EWP	0.9679	0.9855	0.9622	0.9686	0.9326	0.5877	0.9947	0.9957
TIME	29.00	11.09	5.37	6.46	0.28	27.98	32.88	9.06

The evaluated methods in the Worms class show that the classification accuracy of the proposed GMMES method is better than the convolutional neural network and other intrusion detection models (Figure 11).

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

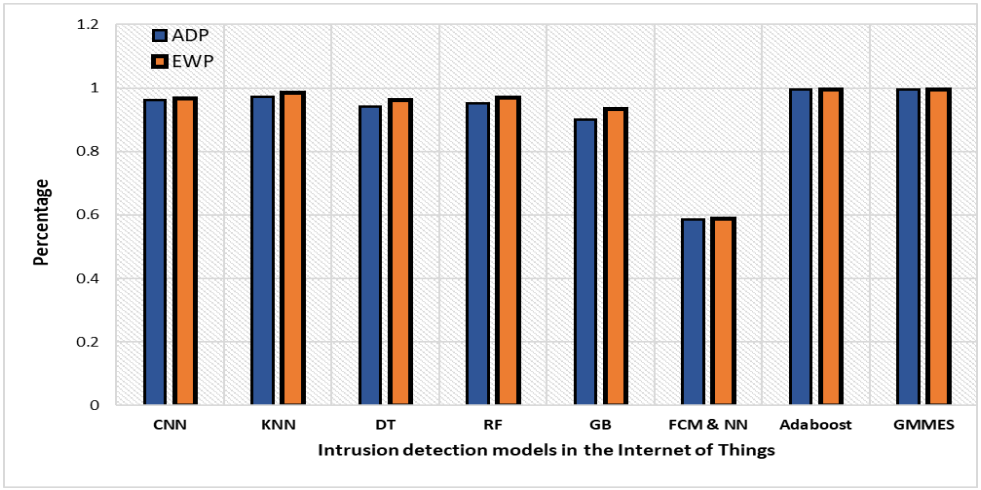

Fig. 9 Comparison diagram of intrusion detection models in Dos class.

The comparison of evaluated methods in Dos class based on learning time shows that the GMMES method has reduced the training time from 32.88 to 9.06 in the collective learning algorithm by using feature selection based on correlation and parallel processing (figure 13).

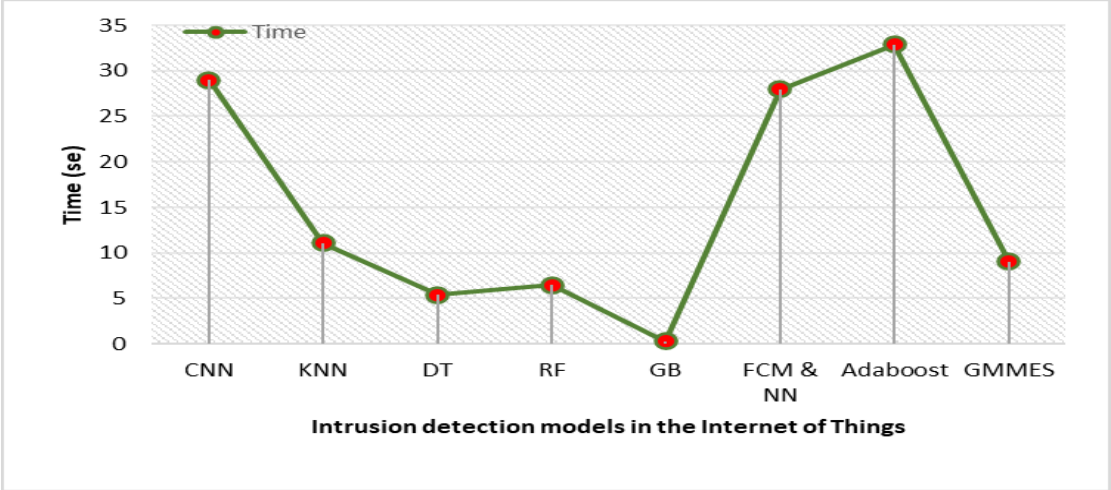

Fig. 10 Comparison time of intrusion detection models in Dos class.

4.COMPARING THE GMMES AND CLASSIFICATION METHODS IN THE GENERIC CLASS

Table 4 compares the evaluated methods in Generic class based on the evaluation criteria of ADP, EWP, and learning time.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

TABLE 4

COMPARING THE PROPOSED METHOD (GMMES) WITH OTHER CLASSIFICATION METHODS IN THE GENERIC CLASS IN THE PERFORMANCE

METRICS	CNN	KNN	DT	RF	GB	FCM & NN	ADABOOST	GMMES
ADP	0.972	0.9943	0.983	0.9911	0.9066	0.529	0.9998	0.9999
EWP	0.9764	0.9574	0.9938	0.9892	0.9174	0.5491	0.9945	0.9975
TIME	28.00	18.96	7.10	8.15	0.36	34.01	39.42	14.40

The evaluated methods in the Generic class show that the classification accuracy of the proposed GMMES method is better than the convolutional neural network and other intrusion detection models (Figure 14).

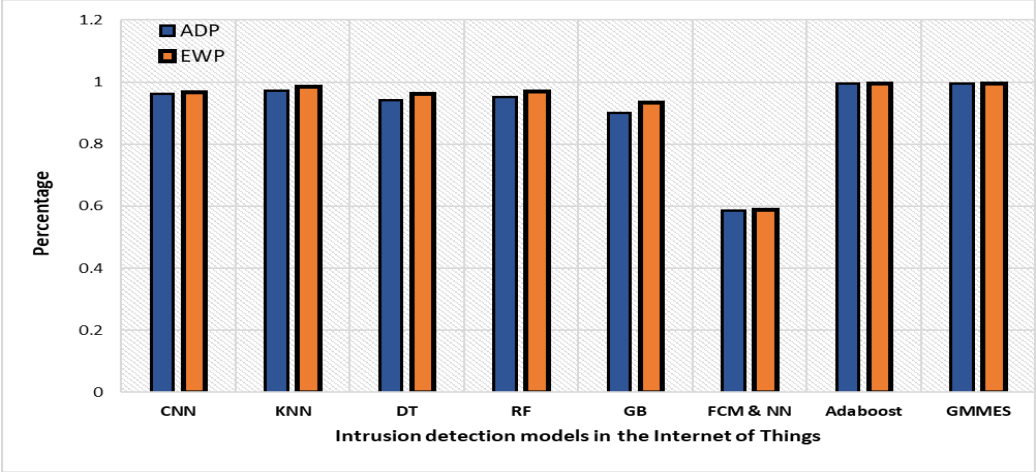

Fig. 11 Comparison diagram of intrusion detection models in Dos class

The comparison of evaluated methods in Generic class based on learning time shows that the proposed method has reduced the training time from 39.42 to 14.40 in the collective learning algorithm by using feature selection based on correlation and parallel processing (figure 15).

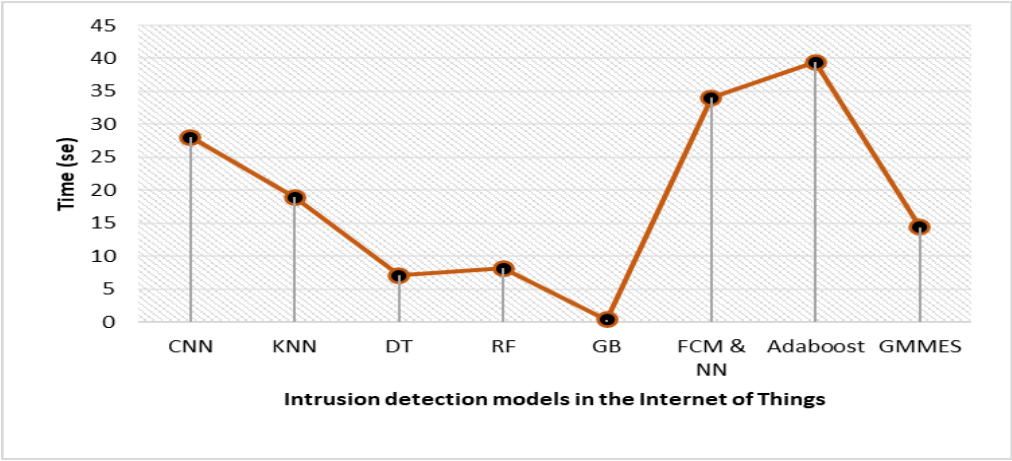

Fig. 12 Comparison time of intrusion detection models in Generic class.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

5.COMPARING THE GMMES AND CLASSIFICATION METHODS IN THE Exploits CLASS

Table 5 compares the evaluated methods in Exploits class based on the evaluation criteria of ADP, EWP, and learning time.

TABLE 5
COMPARING THE PROPOSED METHOD (GMMES) WITH OTHER CLASSIFICATION METHODS IN THE EXPLOITS CLASS IN THE PERFORMANCE

METRICS	CNN	KNN	DT	RF	GB	FCM & NN	ADABOOST	GMMES
ADP	0.9621	0.9712	0.9421	0.951	0.8994	0.5854	0.994	0.9954
EWP	0.9679	0.9855	0.9622	0.9686	0.9326	0.5877	0.9947	0.9957
TIME	29.00	11.09	5.37	6.46	0.28	27.98	32.88	9.06

The evaluated methods in the Exploits class show that the classification accuracy of the proposed GMMES method is better than the convolutional neural network and other intrusion detection models (Figure 16)

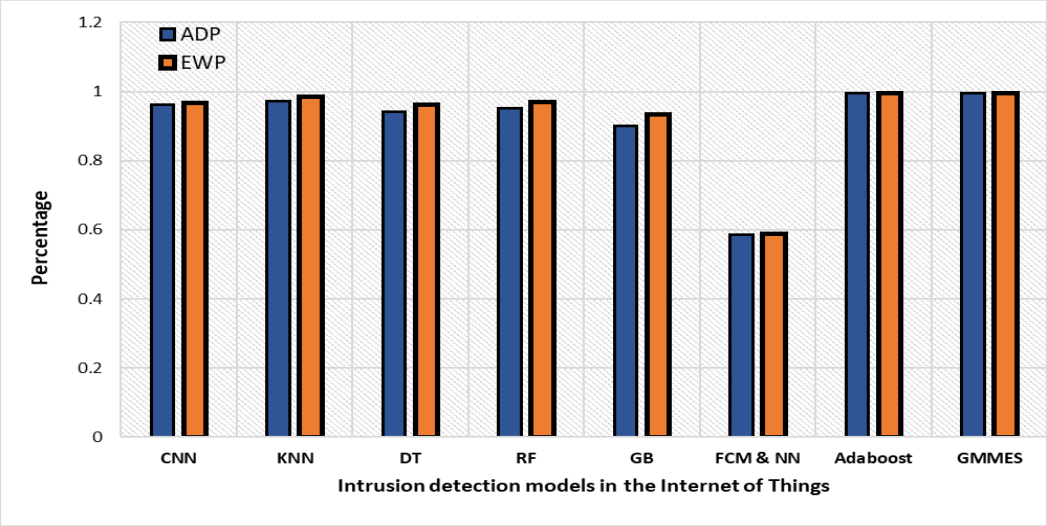

Fig. 13 Comparison diagram of intrusion detection models in Exploits class.

The data in figure 17 shows that the GMMES method has reduced the training time in the adaboost learning from 32.88 to 9.06 by using feature selection based on correlation and parallel processing of subsets. The evaluation of the evaluated methods in the Exploits class shows that the classification accuracy of the GMMES is similar to that of the Adaboost learning.

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

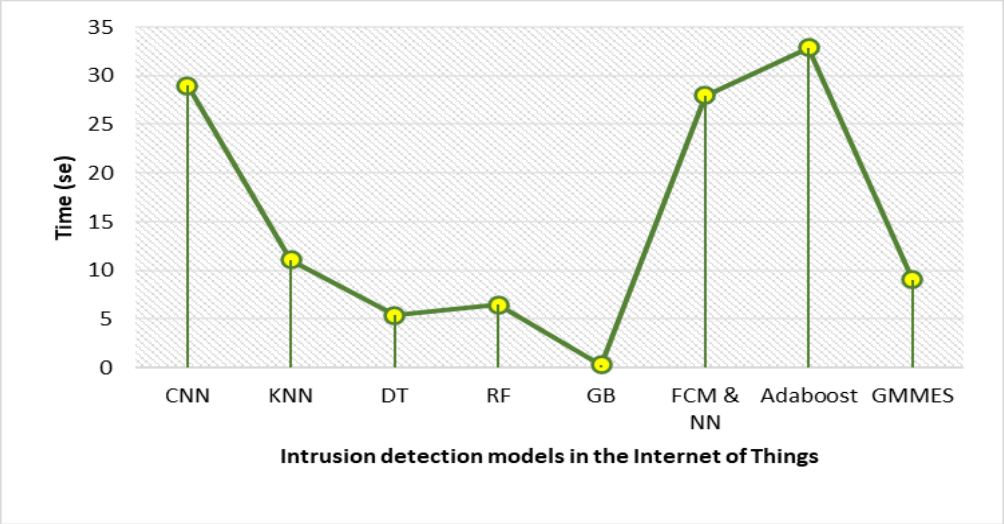

Fig. 14 Comparison time of intrusion detection models in Exploits class

V. ANALYSIS

The proposed method, combining three K-nearest neighbor algorithms, random forest, and support vector machine as a collective learning algorithm, has the highest intrusion detection accuracy among the compared methods. Also, comparing the evaluated methods in attack classes based on learning time shows that the proposed method has increased the prediction accuracy in the normal class from 78% to 81%. Also, the comparison of the evaluated methods in Worms class showed that the classification accuracy of the proposed method is similar to adaboost learning. Still, the proposed method has the highest-class accuracy in other classes, such as Dos, Exploits, and Generic classes. Examining the training time in the studied methods showed:

In Dos, Exploits, and Generic classes, the proposed method has reduced the training time in collective learning algorithms and neural networks by using feature selection based on correlation and parallel processing. Feature selection plays a key role in network intrusion detection systems to select important features and remove unnecessary features. It can help identify legitimate and suspicious samples and improve the overall performance of any NIDS. The purpose of feature selection is to reduce the computational cost of NIDS, remove redundant information, improve the accuracy of NIDS, and help analyze the normality of network data. This article uses the simplest feature selection method, especially the correlation coefficient (CC), which calculates the degree of stability (ability) between several features. The features with the lowest ranking N are selected as the most critical features, which are transferred to the group method to identify the abnormal behaviors of DNS and HTTP instances.

VI. CONCLUSION

Removing inappropriate features from the UNSW-NB 15 dataset is a suitable strategy to reduce the dataset in intrusion detection systems. Today, most approaches in intrusion detection are focused on the problem of extracting essential features. But extracting the features will cause the loss of part of the data. Most current intrusion detection systems use all parameters in network packets to evaluate and discover attack patterns if some of these parameters are irrelevant and redundant. Considering that we are faced with a high number of features in the data set of intrusion detection systems, reducing the

DOI: http://dx.doi.org/10.25098/9.1.31

The Scientific Journal of Cihan University – Sulaimaniya **PP: 118-139**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

dimensions can help to make the analysis easier, increase the separation performance, and remove duplicate and irrelevant information. Reducing the volume of data is possible in two ways. First, it reduces the number of features, eliminating less important features of the intrusion detection system. Second, reducing the number of samples leads to the deletion of records and samples that, by removing them, the intrusion detection system works with better accuracy.

This research presented the group learning framework proposed to detect botnet attacks in the Internet of Things networks through TCP/IP protocols by analyzing MQTT, DNS, and HTTP protocols. This framework consists of three main steps: a feature set, feature selection, and an ensemble method. To begin with, a feature set containing the proposed features, the correlation coefficient, is used to select the features with the lowest correlation with potential characteristics of legitimate and malicious patterns. Finally, the group method is used to classify normal and suspicious samples. Future researchers should consider using deep learning methods in intrusion detection systems, such as Restricted Boltzmann Machine (RBM).

- To present a category of intrusion detection systems for wireless sensor networks and the Internet of Things according to the detection method: misbehavior detection, anomaly detection, and feature-based detection.
- To present a category of intrusion detection systems for wireless sensor networks and the Internet of Things regarding how the IDS agent is deployed in the network.
- Conduct a study on network infrastructure for IDS in MANET networks. Hierarchical intrusion detection systems (multi-layer network infrastructure) in the MANET network can be studied as a research topic.

DOI: http://dx.doi.org/10.25098/9.1.31

REFERENCES

Alghanam, O. A., et al. (2023). "An improved PIO feature selection algorithm for IoT network intrusion detection system based on ensemble learning." Expert Systems with Applications **213**: 118745.

Ananthi, P. and V. Parthipan (2022). "An Innovative Method for Detection of Malicious Behaviours in Automated Vehicle System Using Hybrid Fuzzy C-Means Algorithm with Neural Network Algorithm Based Accuracy and Cost." ECS Transactions **107**(1): 11765.

Asgharzadeh, H., et al. (2023). "Anomaly-based Intrusion Detection System in the Internet of Things using a Convolutional Neural Network and Multi-Objective Enhanced Capuchin Search Algorithm." Journal of Parallel and Distributed Computing.

Aswad, F. M., et al. (2023). "Deep learning in distributed denial-of-service attacks detection method for Internet of Things networks." Journal of Intelligent Systems **32**(1).

Awujoola Olalekan, J., et al. (2020). "Effective and accurate bootstrap aggregating (Bagging) ensemble algorithm model for prediction and classification of hypothyroid disease." Int J Comput Appl **176**(39): 41-49.

Bhattacharya, S., et al. (2021). Investigation of Deep Learning Model Based Intrusion Detection in Traditional and Ad Hoc Networks. 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE.

Bhavani, T. T., et al. (2019). Network intrusion detection system using random forest and decision tree machine learning techniques. First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019, Springer.

Bhutamapuram, U. S. and R. Sadam (2022). "With-in-project defect prediction using bootstrap aggregation based diverse ensemble learning technique." Journal of King Saud University-Computer and Information Sciences **34**(10): 8675-8691.

Elghamrawy, S. M., et al. (2022). An Intrusion Detection Model Based on Deep Learning and Multi-layer Perceptron in the Internet of Things (IoT) Network. The 8th International Conference on Advanced Machine Learning and Technologies and Applications (AMLTA2022), Springer.

Ismail, N. and S. Abdullah (2016). "Principal component regression with artificial neural network to improve prediction of electricity demand." Int. Arab J. Inf. Technol. **13**(1A): 196-202.

Kim, H. and Y. Lim (2022). "Bootstrap aggregated classification for sparse functional data." Journal of Applied Statistics **49**(8): 2052-2063.

DOI: http://dx.doi.org/10.25098/9.1.31

Louk, M. H. L. and B. A. Tama (2023). "Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system." Expert Systems with Applications **213**: 119030.

Mhawi, D. N., et al. (2022). "Advanced feature-selection-based hybrid ensemble learning algorithms for network intrusion detection systems." Symmetry **14**(7): 1461.

Mostafa, S. A., et al. (2023). "Deep learning in distributed denial-of-service attacks detection method for Internet of Things networks."

Moustafa, N. and J. Slay (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). 2015 military communications and information systems conference (MilCIS), IEEE.

Roets, A. and B. L. Tait (2023). IoT-Penn: A Security Penetration Tester for MQTT in the IoT Environment. Cybersecurity in the Age of Smart Societies: Proceedings of the 14th International Conference on Global Security, Safety and Sustainability, London, September 2022, Springer.

Sabitha, R., et al. (2022). "Network Based Detection of IoT Attack Using AIS-IDS Model." Wireless Personal Communications: 1-24.

Shaikh, A. and P. Gupta (2022). Advanced Signature-Based Intrusion Detection System. Intelligent Communication Technologies and Virtual Mobile Networks: Proceedings of ICICV 2022, Springer**:** 305-321.

Talukder, M. A., et al. (2023). "A dependable hybrid machine learning model for network intrusion detection." Journal of Information Security and Applications **72**: 103405.

Thakkar, A. and R. Lohiya (2023). "Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System." Information Fusion **90**: 353-363.

Wang, Y., et al. (2019). "Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection." Applied Soft Computing **77**: 188-204.

Yang, Z., et al. (2022). "A systematic literature review of methods and datasets for anomaly-based network intrusion detection." Computers & Security: 102675.

DOI: http://dx.doi.org/10.25098/9.1.31