International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.2 (2025) pp. 3554-3564
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Optimizing Feature Selection for Machine Learning-Based Intrusion Detection Systems Against Modern Cybersecurity Threats

Ahmed Jamal Alshaibi[1,2,3,*], Jamal Mustafa Al-Tuwaijari[4], Hazim Noman Abed[5], Mohammed Shakir Mohmood[6], Bashar Talib Al-Nuaimi[7]

[1] Department of Computer Science, College of Sciences, University of Diyala, Iraq.
[2] Directorate of Education in Diyala, Ministry of Education ,Iraq.
[3]Department of Complex Information Security of Computer Systems, Faculty of Security, Tomsk State University of Control Systems And Radioelectronics, Tomsk, Russian Federation.
* **Corresponding Author Email:** ahmed.jamal.alshaibi88@gmail.com - **ORCID:**0000-0002-5247-7851

[4]Department of Computer Science, College of Sciences, University of Diyala, Iraq,
Email: dr.altuwaijari@sciences.uodiyala.edu.iq - **ORCID:** 0000-0002-5247-7852

[5] Department of Computer Science, College of Sciences, University of Diyala, Iraq,
Email: hazim_numan@uodiyala.edu.iq – **ORCID:** 0000-0002-5247-7853

[6] Scholarship & Cultural Relations Directorate, Ministry of Higher Education & Scientific Research, Baghdad, Iraq,
Email: mahmood@tut.by – **ORCID:** 0000-0002-5247-7854

[7] Department of Computer Science, College of Sciences, University of Diyala, Iraq,
Email: alnuaimi_bashar@uodiyala.edu.iq – **ORCID:** 0000-0002-5247-7855

Article Info:

DOI: 10.22399/ijcesen.1434
Received : 05 January 2025
Accepted : 17 March 2025

Keywords :

Cybersecurity
Cyberattacks
IoT
Datasets
Anomaly Detection
Machine learning

Abstract:

Technologies like AI, cloud computing, and big data have come a long way and changed a lot for the better. But this rise of cyberattacks to ensure an effective intrusion detection system (IDS). The challenges include lower accuracy from redundant features, less ability to detect new attacks from a single machine learning classifier, high rates of false alarms (FAR), excessive building time of models, etc.This paper introduces a hybrid feature selection approach with an ensemble classifier to select relevant features and give consistent classification of attacks. As the most recent open-access IDS dataset, the CICIDS-2017 dataset has significant promise as a prospective benchmark for IDS of the future since it incorporates contemporary system configurations and threat profiles. Yet in research, especially feature selection, it is still yet to be fully utilized. To overcome these issues, this study introduces a novel IDS framework deploying ensemble-based feature selection to derive a low-dimensional feature subset, and an ensemble-based IDS model for benchmarking on CICIDS-2017. The proposed scheme is a valuable contribution to the research community by integrating the most recent IDS dataset with ensemble methods for feature selection and detection, offering a strong solution for contemporary network security.Keywords: Cyber-Physical Systems (CPS), Intrusion Detection Systems (IDS), Ensemble-Based Methods, Feature Selection, Cybersecurity Threats.

1. Introduction

The widespread use of technologies like big data, IoT, and cloud computing is leading to a global surge in data, driven by the simultaneous generation of vast amounts of data from humans and IoT devices. [1]. In this sense, this shift is changing society and business in an unprecedented way. Big Data and its solution globally would provide a competitive edge to several markets through efficient generation and utilization of rich big data as well as advanced technology in 2. Consequently, this infrastructure not only gains considerable interest from government and business entities, but cybercriminals are also attempting to take advantage of the highly economic and sensitive data it contains Yet, this inflation of data is not an unqualified good: a lot of it is very much incomplete or uncertain, well beyond many of the big data and real-world application areas. This data

can be divided into asymmetric (unbalanced) and symmetric (balanced) distributions You are Symmetric between data types in social networks andymmetric distribution of malicious and regular traffic in networks. In these real-world applications, although some information may be missing, there are still rich and hidden topological structures and patterns hidden. Therefore, it is essential to have a efficient, effective method of searching these relevant patterns [3]. The growing use of [facilities] of the internet is also a persistent threat to computer systems and networks [4] For example, a wide range of different kinds of cyber attacks have undergone significant evolution since the advent of the internet and rapid advances in revolutionary technologies. No matter how hard security experts come up with defense mechanisms, hackers has been and will continue to find a way to escape with specific resources from those most valuable and trusted sources around the world by executing you a robust, distributed and automated cyber attacks. Consequently, many kinds of havoc that were huge for governments, businesses, and even people [5]. In [6], the authors interest-ingly discuss the different types of cyber-attacks and their impacts. The paper also discusses a prediction of USD 6 trillion in cyber-crimes by 2021 and some of the sophisticated global cyber-crimes that may lead to a global loss of USD 1 billion. 1.5 trillion US dollars of cyberattack revenues from the compromise of between two and five million computers a day. This has caused the research community to pay increasingly attention to Intrusion Detection Systems (IDSs), particularly those that can perform real-time detection, over the last few decades. [4,7]. An example of this use is the work of [8] who gave a good overview of what you would pay for security properties in the context of cloud computing platforms. They also presented a new also erudite and efficient 3-layer cloud based IDS based on logic used to symbolize the specification and monitoring properties of Event Calculus. In addition, the proposed method applied the virtualization framework to highlight the supervision of applications at run time and facilitate the automatic reconstructions of the applications. In the end, the paper said it had significantly improved the firewall for cloud-based computing According to [9 -11]. A detailed information about the standard function and roles of intrusion detection systems are shown in [12].

Intrusion Detection and Avoidance There are two types of approaches to detection misuse detection (knowledge-based or signature-based) and anomaly detection. Recent trends have leaned towards in-memory hybrid-based approaches to leverage the strengths of both approaches and achieve a high-performing and efficient system [13]. So MIDS (Misused Intrusion Detection Systems) are signature-based systems that detect intrusions or malicious activities by matching or comparing network events with a database of known attack signatures. MIDS has limitations, though it provides excellent attack detection rates when analyzing known attacks. It has a limitation in its detection of new attacks that have not been previously identified and can be slow in processing and analyzing the massive amounts of data in the signature database so it won't perform as fine while facing new threats. [14–16]. In the sigA [17], Some researchers have effectively proposed using signature-based intrusion detection systems to detect SQL injections in databases. This way only relevant patterns are generated and stored in the signature database which decreases the load for detection. Not however, machine learning is still the best performing AI approach, where it can learn automatically on heaps of unstructured data. In particular, when suitable training datasets are used, machine learning-based IDSs have realized encouraging performance in detection, [7,11]. To conduct research or to develop a practical application, the most commonly used machine learning models are supervised and unsupervised learning, which belong to one of the two main families of machine learning models, respectively, for designing and developing IDSs. The super is a bit on the idea of discovering useful data from labeled datasets. But, manually annotating datasets often is: slow, prone to errors, and costs a lot of money. Hence limits the broader acquisition and implementing of supervised learning based IDs [18], due to restricting access to adequate trained data annually for such models. But their detection accuracy for known attacks is very goodIn addition to these, unsupervised learning does not need labeled datasets, unlike supervised machine learning. In contrast, it can recognize hidden patterns in unlabeled data without supervision from humans, and therefore, the training data are more widely available than in the case of the supervised models [19]. Many research studies came up with machine-based techniques in order to successfully deploy effective and robust Intrusion Detection Systems (IDS) without needing a knowledge of class data or the assumption of a set network context. Clustering algorithms such as Hierarchical Clustering and Gaussian Mixture Model (GMM) are particularly useful in both pre-processing of large and complex unlabeled datasets and also help in reducing several weaknesses. Such a method broadly improves the detection performance while decreasing the high false-negative rate often reported in existing intrusion detection systems

(IDSs). [20]. On one side, [21] utilized classical ML technique to propose a novel semi-supervised anomaly detection framework. In the training phase, this basically means that they can force the normal samples to be a member of a specific cluster using the K-means base algorithm. The threshold value was defined based on the above-mentioned results, as they have the greatest influence towards categorization of samples in the system proposed as being normal or abnormal. The second proposed method successfully detects the anomalies as the cluster center distance is higher than the threshold value obtained. To summarize, they-reecommended their idea to the NSL-KDD dataset, and they obtained the detection accuracy rate of 80.119 %.

We provide the following summary of needful contribution of this research work:

- Researchers implemented and developed an ensemble-based feature selection model based on Information Gain, CFS and PSO which reduces the dimensionality and processing time of the learning model while maintaining high accuracy with fewer descriptors.
- The researchers suggested an ensemble IDS model named KNN, C4. 5, and Random Forest for enhancing multi-class attack detection of CICIDS-2017 dataset and overcome the deficiency of KNN.
- The researchers addressed the issue of the underuse of new datasets in IDS research by using the CICIDS-2017 dataset, demonstrated strong performance of their ensemble IDS model in terms of accuracy, recall, MCC, and ROC.

2 . Analysis of Cybersecurity Datasets

The evolution of intrusion detection system (IDS) technology is not possible without a thorough evaluation of training data from which most IDS are resulted. Over the years, IDS has developed and given rise to a lot of dedicated datasets for individual research and development purposes. These datasets are essential for training and testing the effectiveness of IDS models [11], allowing for optimization and fine-tuning techniques that bolster their resilience in real-world scenarios0[22].

The goal of this section is to analyze the diversity and realism of these datasets, in other words, how well do they model real-world threat scenarios. With this in mind, we have performed an extensive analysis of publicly available cyber security datasets, as presented in Table 1.

Table 1. *Detailed Information of datasets.*

Feature	CSE-CIC-IDS2018	KDDCup 1998	BoT-IoT	UNSWNB15	NSL-KDD	CICIDS2017
Attack Types	Botnet, DoS, DDoS, Web, Infiltration, Brute-force	DoS, Probe, U2R, R2L	IoT-specific attacks (DoS, DDoS, Botnet, Scanning, Malicious Traffic)	Diverse (DoS, DDoS, Botnet, Web, Fuzzers, Exploits)	DoS, DDoS, Probe, U2R, R2L	Diverse (DoS, DDoS, Botnet, Brute-force, Web, Infiltration)
Traffic Capture	PCAP, CSV files, and processed features	PCAP, CSV files, and processed features	PCAP files and processed features	PCAP files and processed features	Preprocessed flow records	PCAP and CSV files
Benign Traffic	Yes	No	Yes	Yes	No	Yes
Attack Scenarios	Real-world attack recordings (malware + network traffic)	Predefined attack types	Simulated and real-world IoT attacks	Diverse real-world attack recordings	Predefined attack types	Simulated realistic scenarios
Number of Instances	2.7 million	4.8 million	3.1 million	2.5 million	489,843	11.8 million
Number of Features	85	41	233	66	41	85
Complexity	Complex (real-world malware + network traffic)	Simple	Moderate (simulated and real-world)	Complex (mix of real-world and simulated)	Simple	Moderate
Strengths	Real-world malware interaction, diverse attacks, labeled data	Large dataset, established benchmark	IoT-specific, a mix of simulated and real-world	Real-world attacks, rich features, labeled data	Large dataset, simple to use	Realistic scenarios, diverse attacks, labeled data
Weaknesses	High computational cost, imbalanced classes	Outdated attacks, unrealistic scenarios	Limited attack types, simulated scenarios	Imbalanced classes, complex features	Outdated attacks, unrealistic scenarios	Scattered presence, huge volume of data, missing values

3. Feature Selection Categories Based on Selection Strategy

Feature selection (FS) is an important dimensionality reduction approach, which comprises four main stages: feature subset generation, evaluation, termination condition, and validation of results. FS methods are classified according to the presence of labels in the training dataset: supervised (labels available), unsupervised (no labels), and semi-supervised. This study supports a supervised model [21].

Furthermore, FS methods can be classified according to the selection strategy into filter, wrapper and embedded methods. Filter methods evaluate feature importance based on intrinsic data characteristics, scoring each feature using univariate or multivariate assessment. Since they don't rely on learning algorithms, they are quicker but sub-optimal for specific algorithms. Modeling and Evaluation of Different Criteria — using discriminative ability (e.g, Fisher Score) and correlation measures (e.g, Pearson's correlation, PCA)

Wrapper methods, on the other hand, evaluate subsets of features using a learning algorithm. These methods keep iterating until a stopping criterion is met e.g the desired number of features or model optimal performance. Wrapper methods have higher accuracy but the search space is too large, making it computationally expensive [23].

Embedded methods achieve a balance between speed and accuracy, as FS is incorporated into the learning algorithm's training process. These methods (for instance, regularization (LASSO, RIDGE) or tree based algorithms (Random Forest, Decision Tree)) are not iterative in nature and are more efficient as they do not require iterative evaluation.The feature selection (FS) process is illustrated in different categories (Filter, Wrapper, and Embedded) and depicted in Figure 1.

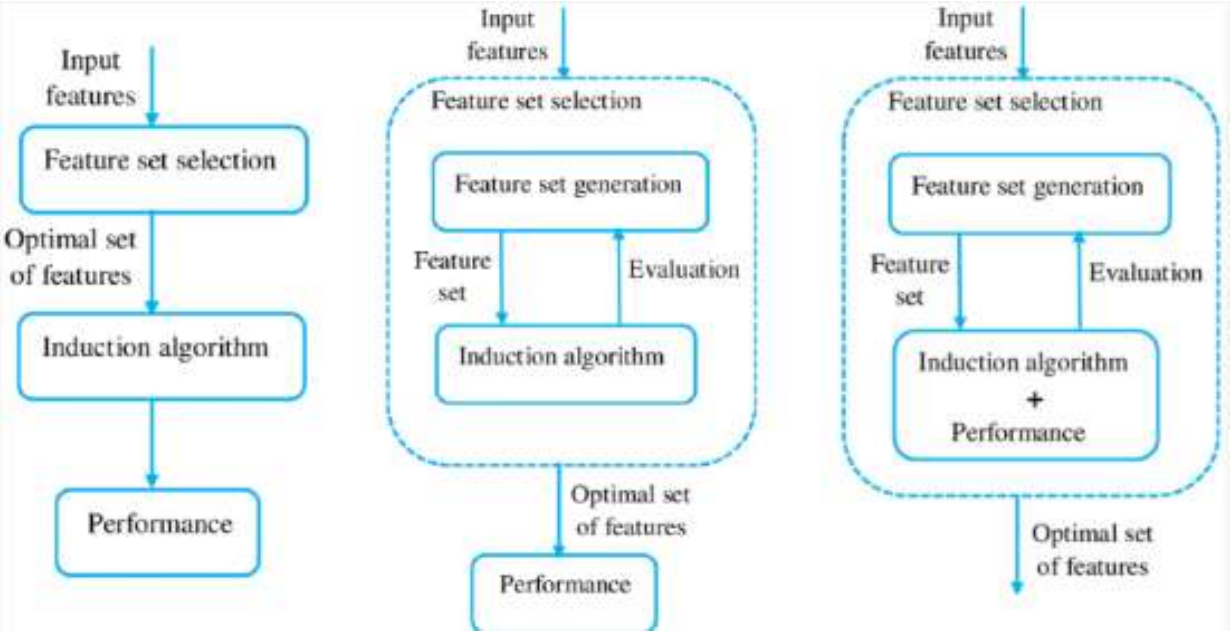

Figure 1. *Schematic representation of the FS process across different categories: Filter, Wrapper, and Embedded*

3.1 The Proposed Ensemble-Based Feature Selection Approach

Currently, there was increasing interest regarding questioning which feature selection method would be optimal for specific datasets and various approaches regarding ensemble feature selection have been introduced. So, the ensemble feature possesses an advantage over others since it combines various results of the feature-selection methods, and hence, a less informative and least relevant feature is dropped out and a diverse feature-subset is retained. With their ability to merge the strengths of different approaches, ensemble methods can enhance the overall performance of intrusion detection systems (IDS), which can be achieved by allowing them to determine and prioritize important features in order to better detect relevant cybersecurity threats [24]. In the proposed Ensemble-Based FS Method model Figure 2, three FS algorithms were used:

Chi-Square (χ^2) Selection: The χ^2 test selects the features that are statistically correlated with the class of the target.

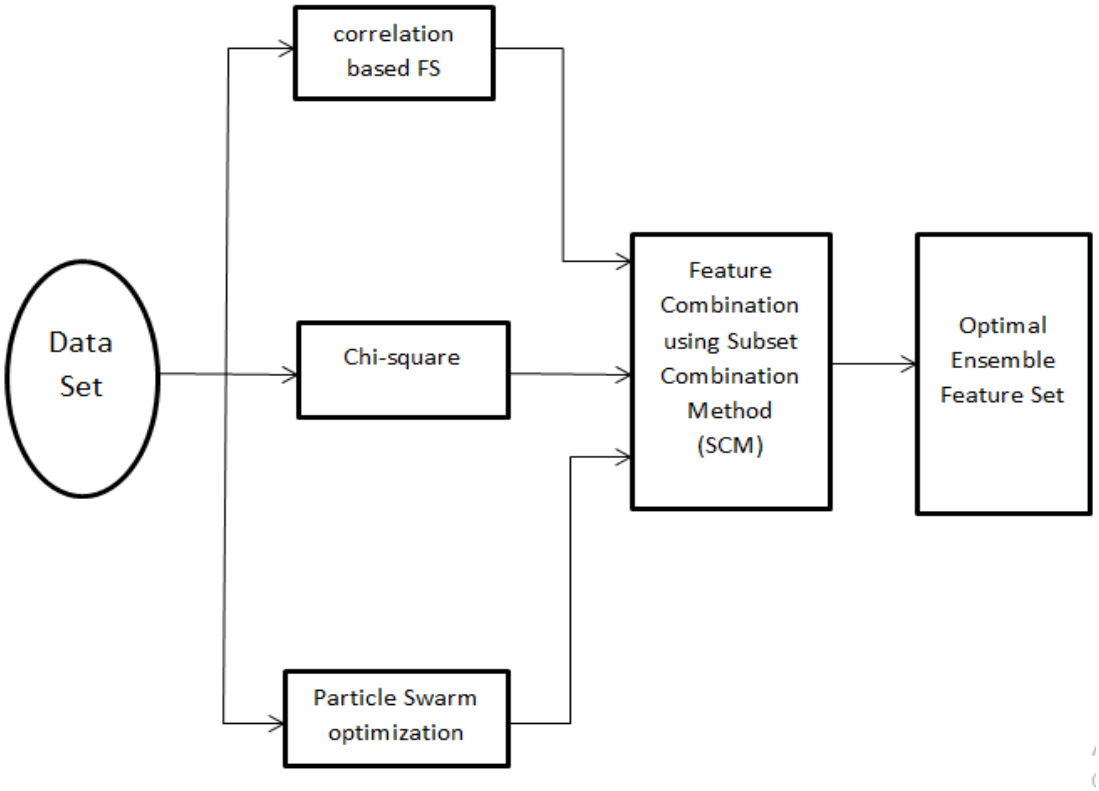

Figure 2. *Ensemble-Based FS Method With ML*

A χ^2 test compares the observed distribution of the data with the expected distribution that would be obtained if they were independent of each other. For feature-level based test (here we handle only categorical data the chi-square statistic will be calculated for each feature for contingency table, by finding observed vs expected values.

Correlation-Based Feature Selection (CFS) : This algorithm selects attributes that are highly correlated with the target class and less correlated with each other. It assesses the quality of a feature subset using the Pearson correlation.

Particle Swarm Optimization (PSO): PSO is based on the natural search process as a population algorithm that fuses multidimensional data into feature subsets. This is commonly used in feature selection problems.

3.2 Dataset Utilized in This Study

The CIC-IDS2017 dataset was used with more data, as it enables risk analysis and behavioral pattern identification as well as the testing of intrusion detection techniques. It is one of the most widely used datasets in current cybersecurity research due to its diverse range of attacks and coverage of various realistic scenarios. A nationallyrecognized dataset — Cyber Intrusion Detection and Evaluation System (CICIDS) was further developed by Canadian Institute for Cybersecurity with University of New Bruns wick, using BProfile system. CICIDS [24] is another dataset that models network traffic generated by users and contains different network protocols.

The distance of entropy between malicious and benign traffic is 0.716, and the distance of entropy between attack types is 0.523 as the statistics show. This suggests that attack types in the dataset are less balanced than benign traffic.

Canadian Institute of Cybersecurity and the University of New Brunswick developed the CICIDS dataset, which is evaluated using the Bprofile system. It simulates network traffic (HTTP, HTTPS, File Transfer Protocol (FTP), Secure Shell Protocol (SSH) and email). It also captures the more transactional behaviors of 25 users. The authors of [25] used the KDD dataset for their work, but decided against using the KDD dataset for their work due to its relatively simple features. Instead, they chose the CI-CIDS-2017 dataset known to have more complex features, more traffic, and more columns applicable to better detect anomalous traffic. Entropy between malign and benign traffic was 0.716 while between attack types: 0.523. This signifies that the attack type proportion was more diverged as compared to the benign and malicious traffic ratio [26]. The statistics of attacks and normal behavior in the CICIDS 2017 dataset is shown in Figure 3.

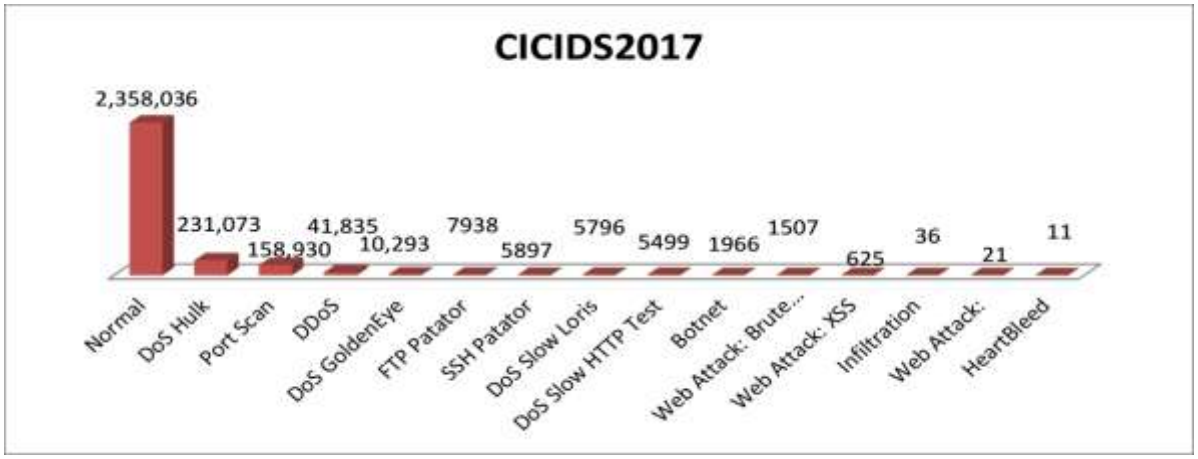

Figure 3. *Statistics of attacks and normal behaviour in the CICIDS 2017 dataset.*

4. Proposed Ml methodology for intrusion detection model

Typically, an IDS comprises a data source, a data preprocessing module, a decision-making mechanism, and a response system to the threats identified [27]. The proposed IDS in this work comprises data pre-processing, an ensemble-based feature selection, and the ensemble-based intrusion detection methods.

In this paper we provide an ensemble of machine learning models based on K-Fold Cross Validation to propose an IDS (Intrusion Detection System) model. This approach decreases the odds of model overfitting, thus providing a more accurate picture of model performance. An ensemble feature selection method combined with K-Fold Cross Validation was utilized to select the most informative features. The base classifiers utilized for the ensemble are KNN, C4. 5, along with Random Forest, which offers a different viewpoint and generalization of the predictions

Research is especially concentrated on data preprocessing, and decision-making mechanism. Figure 4 illustrates the proposed system architecture.

In the study, The IDS model was trained and tested with both original feature selection set and ensemble feature selection using the publicly available dataset CICIDS-2017. As is the custom in this field, K-Fold Cross Validation with K = 10 was Used to assess mdl accuracy.

A voting algorithm based on the average probability rule was used to aggregate individual predictions obtained from the base classifiers. The experimental results indicate that the proposed IDS model based on the ensemble approach and K-Fold Cross Validation methods is very accurate and efficient in detecting attacks both for the original and ensemble feature sets.

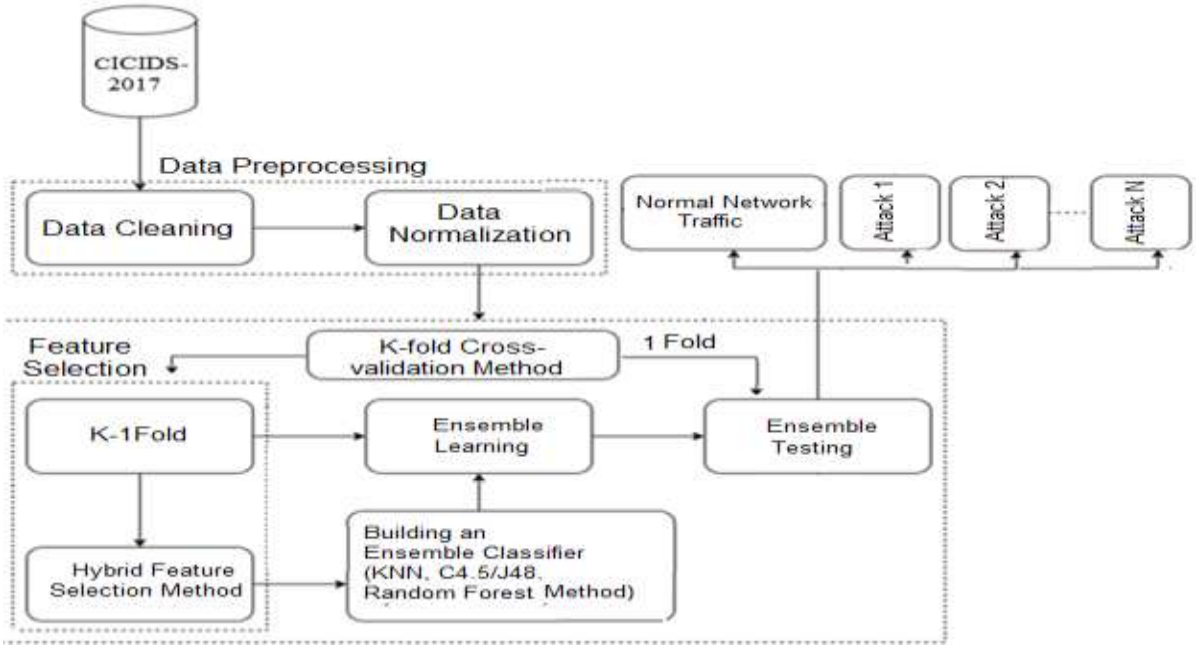

Figure 4. *The architecture of the proposed model*

5. Results and Monitoring

Two models were analyzed in this study, namely the ensemble-oriented feature selection model and the ensemble-oriented intrusion detection model. Methods like KNN and C4 algorithm derivations are in this category. These include: implementation of the methods were J48, 5/J48, and Random forest which have improved detection rate and system efficiency. Its effectiveness for a cyber attack(intrusion detection) was analyzed using CICIDS-2017 dataset. The research covers the preprocessing procedure (data cleaning, feature selection) and comparison of the suggested models and system performance using the CICIDS-2017 dataset.

5 .1 CICIDS-2017 Data Filtering Process and Its Importance

In addition, noises and repeated data should be removed from the data set of CICIDS-2017. By eliminating extraneous attributes and features that do not carry any meaningful information, the number of attributes in the CICIDS-2017 data set can be shortened to 68 columns from 78 to affect an overall cleaning operation for all parameters. Data that did not contain any meaningful information was discarded. As a result, the final dataset has 1,666,532 records, instead of the previous 2,827,876. In order to reduce the entire amount of resources used a 30% sample taken from what was available was obtained. This makes learning and test easier accordingly which can save quite a bit. The min-max of features were done as well to ensure that no one feature dominates another, and in result does not lead to instability or estimation errors.

5.2 Appling proposed FS Process for CICIDS-2017

This hybrid feature selection model is designed to create a whole feature dataset for the 2017 CICIDS dataset. Chi-square, CFS, and Particle Swarm Optimization (PSO) algorithms are combined in the model. 25 features were selected with Chi-square, 21 by CFS, and PSO picked out 13 There was a cutoff of 0.05 for Chi-square, and 0.2 CFS. These features combined into 17 ensemble features (shown as table 2) were critical for model performance under the subset combination method (SCM). The model was evaluated by 10-fold cross-validation, ensuring that each part of data was both training and testing data so as to make evaluations more robust.

Table 2. *Selected Ensemble Features from CICIDS-2017 After Data Preprocessing*

Feature Number	Feature Name
1	Destination Port
7	Fwd Packet Length Max
11	Bwd Packet Length Max
12	Bwd Packet Length Min
13	Bwd Packet Length Mean
18	Flow IAT Std
19	Flow IAT Max
24	Fwd IAT Max
37	Max Packet Length
38	Packet Length Mean
39	Packet Length Std
40	Packet Length Variance
49	Average Packet Size
51	Avg Bwd Segment Size
56	Init_Win_bytes_forward
57	Init_Win_bytes_backward
59	min_seg_size_forward

The individual three algorithms were used and feature selection by aggregation through the KNN algorithm was also performed. Comparison of individual feature selection with ensemble-based features indicates that despite similar performance, the Chi-square features model using 25 features is much higher than ones consisting of 17 ensemble features on average figure 5. Accordingly, the CICIDS-2017 dataset's number of enhanced features was reduced compared to that of one single-feature selection algorithm (Chi-Squared, CFS, and PSO) while maintaining identical error accuracy. In light thereof, we find that our ensemble feature selection method is superior Reminder to independent and isolated individual feature selection approaches.

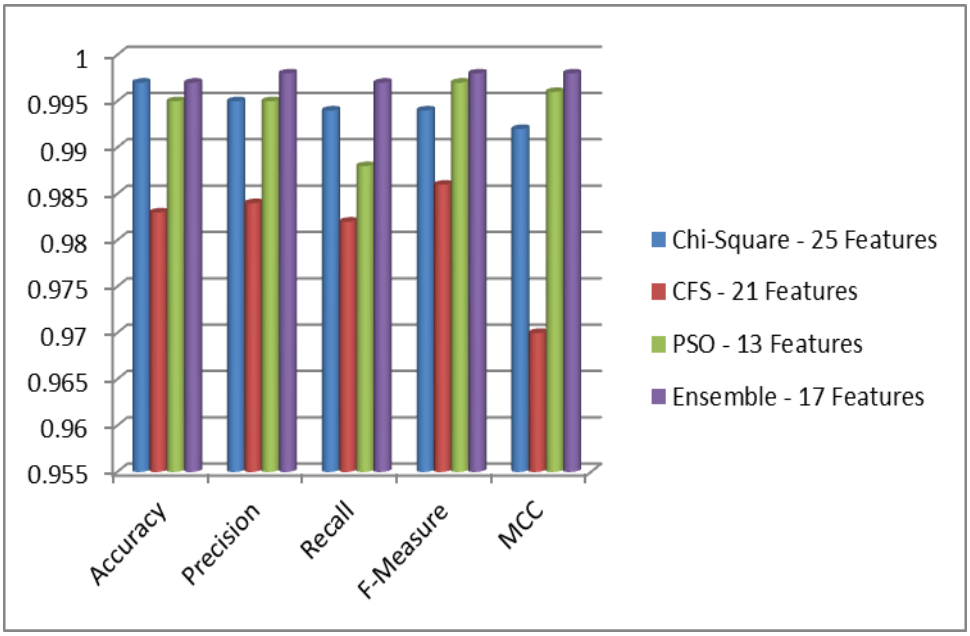

Figure 5. *Comparison of FS results methods for CICIDS – 2017*

Intrusion Detection Systems (IDS) should not be evaluated on the basis of just accuracy, which is capable of misleading in some cases. In addition to the accuracy parameter in an IDS, such additional performance metrics as Precision, Recall, F-measure and MCC should also be used. The accuracy of KNN algorithm was founded superior to the ensemble-based method, as seen in comparison with individual base classifier systems: C4.5 and Random Forest. This indicates that while combining the base-classifier KNN with our proposed method, there is a significant increase in model accuracy. On the other hand, KNN whole classifier performs according to data, Due to the difficulty of accurately identifying minority categories such as SQL Injection, Heart-bleed, and Infiltration attacks, Precision, recall, F-measure, MCC values are all lower. But the ensemble-based model is far more effective in recognizing these attacks; see figure 6 and figure 7.

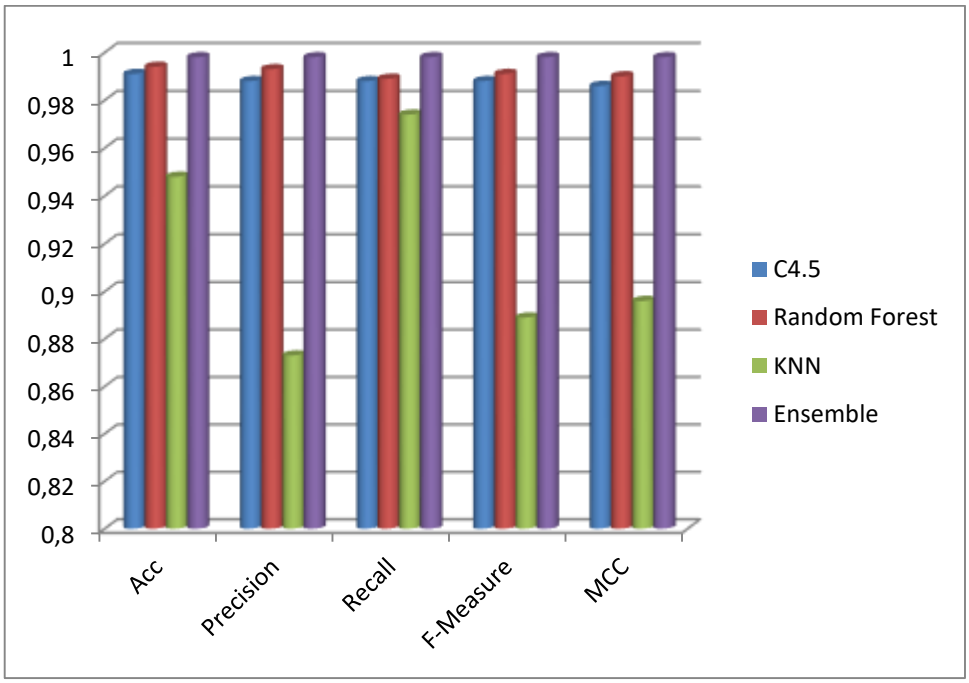

Figure 6. *Ensemble Model Performance with the CICIDS2017 Dataset (68 Features)*

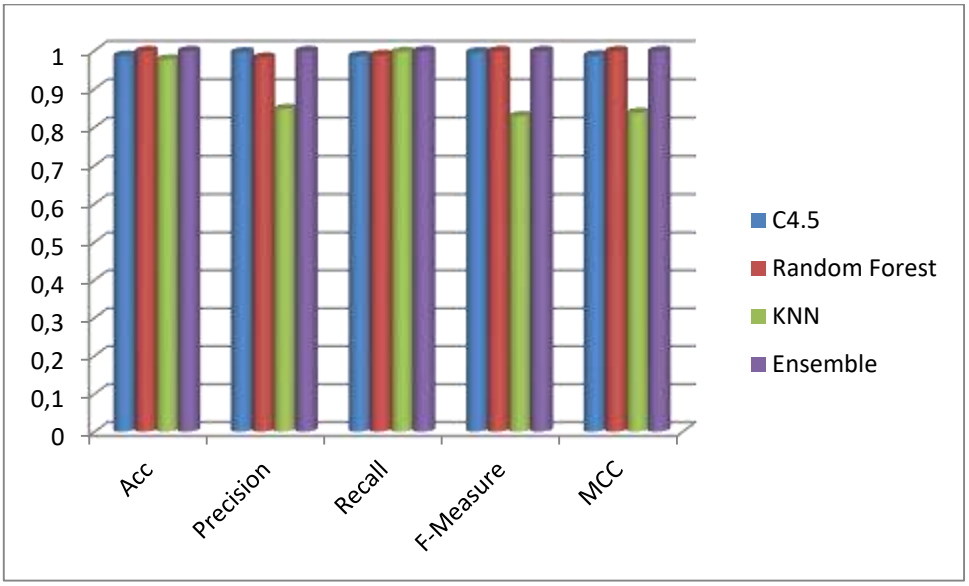

Figure 7. *Ensemble Model Performance with the CICIDS2017 Dataset (17 Features)*

The performance of the proposed ensemble scheme on each single attack category in the CICIDS-2017 dataset, using the 17 derived features, was also analyzed as illustrated in Figure 7. The experimental results show that the proposed ensemble model is robust across various attack types including DDoS attacks, other types of DoS attacks (including Slowloris, Slowhttptest, Hulk, GoldenEye and Heartbleed), Portscan-based attacks and web attacks (including FTP-Patator, SSH-Patator and Brute Force). The proposed model attains consistently high performance in these attack classes including near 100% classification accuracy.

Consequently, a total of 17 ensemble-derived features were utilized by the ensemble intrusion detection model with an overall performance value of 99.922%. It is a good point in time showing how beneficial feature-selection Approach has and tells the applicability of the model, which can lead to a future network security solution for alert a monitoring system.

6. Conclusion

In conclusion, the aforementioned ensemble classifier based model reported an exceptional attack detection and was able to classify multiple classes of attacks with nearly 100% accuracy for the CICIDS-2017 dataset. The ensemble model's capacity to detect poorly represented attacks like SQL Injection, Heart-bleed, and Infiltration was given a better performance than some of the individual algorithms like KNN.

Moreover, the feature selection process was also effective and shortened the time for data process. The CICIDS-2017 dataset are high dimensional input dataset and hence the model without feature selection was obtuse as compared to the above-mentioned feature selection model. But with feature selection, the model processing time was decreased almost by the half see figure 7, which shows that using only relevant feature help in making the intrusion detection process faster.

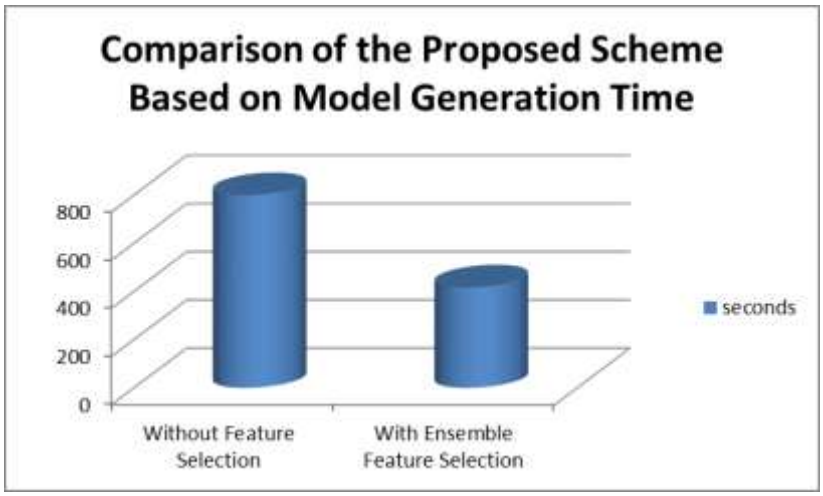

Figure 7. *Comparison of Model Generation Time with and without Ensemble FS.*

Considering all the above advantages, the ensemble-based model with feature extraction, dimensionality reduction, classification, accuracy, detection rate, scalability, robustness, real-time monitoring, cybersecurity and attack detection mentioned in this analysis appears to be an efficient and practical approach for the intrusion detection in computer networks, since, considering the performance and speed compared to other previous works, it can be concluded that it has brought much more advantages to its practical applicability, making it a viable option for the most diverse computer network security monitoring in real time.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Rayes, A., & Salam, S. (2016). *Internet of Things (IoT) overview*. In *Internet of Things from hype to reality* (pp. 1–34).

[2] Alshaibi, A. J., Al-Ani, M. M., & Kadum, J. (2023). The effect of integration and effectiveness of artificial neural networks on information security tasks. *AIP Conference Proceedings, 2591*(1), 030021. https://doi.org/10.1063/5.0121244

[3] Lee, E. (2015). The past, present and future of cyber-physical systems: A focus on models. *Sensors, 15*(3), 4837–4869.

[4] Golani, N., & Rajasekaran, R. (2017). IoT challenges: Security. In *Internet of Things (IoT)* (pp. 211–234).

[5] Gupta, Y., Shorey, R., Kulkarni, D., & Tew, J. (2018). The applicability of blockchain in the Internet of Things. In *2018 10th International Conference on Communication Systems & Networks (COMSNETS)*, 561–564.

[6] Kang, J., Yu, R., Huang, X., Maharjan, S., Zhang, Y., & Hossain, E. (2017). Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains. *IEEE Transactions on Industrial Informatics, 13*(6), 3154–3164.

[7] Rohr, J., & Wright, A. (2019). Blockchains, private ordering, and the future of governance. In *Regulating Blockchain* (43–57).

[8] Zhu, H., Liu, X., Lu, R., & Li, H. (2017). Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM. *IEEE Journal of Biomedical and Health Informatics, 21*(3), 838–850.

[9] Cinque, M., Cotroneo, D., Di Martino, C., Russo, S., & Testa, A. (2009). AVR-Inject: A tool for injecting faults in wireless sensor nodes. In *2009 IEEE International Symposium on Parallel & Distributed Processing* (pp. 1–8).

[10] Sedjelmaci, H., & Feham, M. (2011). Novel hybrid intrusion detection system for clustered wireless sensor network. *International Journal of Network Security & Its Applications, 3*(4), 1–14.

[11] Paul, T., & Rakshit, S. (2021). Big data analytics for marketing intelligence. In *Big Data Analytics* (pp. 215–230).

[12] Gupta, B. B., & Sahoo, S. R. (2021). Machine-learning and deep-learning-based security solutions for detecting various attacks on OSNs. In *Online Social Networks Security* (pp. 57–69).

[13] Thiyagarajan, P. (2020). A review on cyber security mechanisms using machine and deep learning algorithms. In *Handbook of Research on Machine and Deep Learning Applications for Cyber Security* (pp. 23–41).

[14] Gaurav, A., Gupta, B. B., Hsu, C.-H., Yamaguchi, S., & Chui, K. T. (2021). Fog layer-based DDoS attack detection approach for Internet-of-Things (IoTs) devices. In *2021 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1–5).

[15] Promper, C., Engel, D., & Green, R. C. (2017). Anomaly detection in smart grids with imbalanced data methods. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*.

[16] Shekarforoush, S. H., Green, R., & Dyer, R. (2017). Classifying commit messages: A case study in resampling techniques. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 1273–1280).

[17] Ullah, I., & Mahmoud, Q. H. (2017). A hybrid model for anomaly-based intrusion detection in SCADA networks. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 2160–2167).

[18] Beaver, J. M., Borges-Hink, R. C., & Buckner, M. A. (2013). An evaluation of machine learning methods to detect malicious SCADA communications. In *2013 12th International Conference on Machine Learning and Applications* (pp. 54–59).

[19] Ambusaidi, M. A., He, X., Nanda, P., & Tan, Z. (2016). Building an intrusion detection system using a filter-based feature selection algorithm. *IEEE Transactions on Computers*, *65*(10), 2986–2998.

[20] Aminanto, M. E., Choi, R., Tanuwidjaja, H. C., Yoo, P. D., & Kim, K. (2018). Deep abstraction and weighted feature selection for Wi-Fi impersonation detection. *IEEE Transactions on Information Forensics and Security*, *13*(3), 621–636.

[21] Diro, A., & Chilamkurti, N. (2018). Leveraging LSTM networks for attack detection in fog-to-things communications. *IEEE Communications Magazine*, *56*(9), 124–130.

[22] Kolias, C., Kambourakis, G., Stavrou, A., & Gritzalis, S. (2016). Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset. *IEEE Communications Surveys & Tutorials*, *18*(1), 184–208.

[23] Tan, Z., Jamdagni, A., He, X., Nanda, P., & Liu, R. P. (2014). A system for denial-of-service attack detection based on multivariate correlation analysis. *IEEE Transactions on Parallel and Distributed Systems*, *25*(2), 447–456.

[24] Moustafa, N., Turnbull, B., & Choo, K.-K. R. (2019). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things. *IEEE Internet of Things Journal*, *6*(3), 4815–4830.

[25] Ambusaidi, M. A.; He, X.; Nanda, P.; Tan, Z. Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm. *IEEE Transactions on Computers 2016*, 65, 2986–2998.

[26] Jia, Q., Guo, L., Jin, Z., & Fang, Y. (2018). Preserving model privacy for machine learning in distributed systems. *IEEE Transactions on Parallel and Distributed Systems*, *29*(8), 1808–1822.

[27] Feng, P., Ma, J., Sun, C., Xu, X., & Ma, Y. (2018). A novel dynamic Android malware detection system with ensemble learning. *IEEE Access*, *6*, 30996–31011.