ACIG
APPLIED
CYBERSECURITY
& INTERNET
GOVERNANCE

NASK

Improving Threat Detection in Information Security with Ensemble Learning

Ahmad Sanmorino | Information Systems, Universitas Indo Global Mandiri, Indonesia | ORCID: 0000-0002-4949-4377

Rendra Gustriansyah | Informatics, Universitas Indo Global Mandiri, Indonesia | ORCID: 0000-0001-7600-1147

Shinta Puspasari | Informatics, Universitas Indo Global Mandiri, Indonesia | ORCID: 0000-0001-5906-3362

Fauziah Afriyani | Management, Universitas Indo Global Mandiri, Indonesia | ORCID: 0009-0008-8653-1975

Received: 18.06.2025

Accepted: 09.09.2025

Published: 13.10.2025

Cite this article as:
A. Sanmorino, R. Gustriansyah, S. Puspasari, F. Afriyani, "Improving threat detection in information security with ensemble learning," ACIG, vol. 4, no. 1, 2025, doi: 10.60097/ACIG/210525

Corresponding author:
Ahmad Sanmorino, Information Systems, Universitas Indo Global Mandiri, Palembang, Indonesia; E-mail: sanmorino@uigm.ac.id
0000-0002-4949-4377

Abstract

In the face of increasingly complex and frequent cyber-attacks, traditional rule-based threat detection systems often fail to identify evolving malicious behaviours. This study addresses the challenge by leveraging ensemble learning to enhance intrusion detection in information security. By integrating three distinct machine learning models – Support Vector Machine (SVM), Random Forest, and Deep Neural Network (DNN) – the proposed approach capitalises on their strengths while mitigating their weaknesses. The primary goal is to enhance detection accuracy, minimise false positives, and ensure reliable performance across various attack types. Using benchmark datasets, such as NSL-KDD and CICIDS2017, each model is trained and evaluated separately before being combined through a voting mechanism. Results from 10-fold cross-validation show that while baseline models perform well individually, the ensemble demonstrates more balanced and robust detection, achieving 94.00% accuracy, 95.10% precision, and a high area under the curve score of 0.77. These findings highlight the value of ensemble methods in producing consistent and dependable threat classification. The contribution of this work lies in demonstrating how a multi-model ensemble strategy can significantly strengthen cybersecurity defences, offering a scalable solution adaptable to real-world security environments.

Keywords

machine learning, cybersecurity, ensemble learning, intrusion detection

1. Introduction

In today's digitally connected world, cybersecurity has become more important than ever. As networks grow larger and more complex, they also become more vulnerable to increasingly sophisticated cyberattacks. Traditional systems, such as rule-based intrusion detection tools, often struggle to keep pace with evolving threats. Machine learning (ML) has emerged as a promising solution, offering the ability to learn from past patterns and intelligently identify suspicious activity in real time [1, 2]. However, relying on a single algorithm often leads to limited performance, as each model has its own strengths and weaknesses. For example, support vector machines (SVMs) are effective for structured data but can miss non-linear relationships [3, 4]. Deep neural networks (DNNs) are powerful but require extensive data and training time [5, 6]. Random forests are robust to noise but can introduce bias if not tuned properly [7]. The challenge is to develop systems that balance these differences to make smarter, more reliable decisions, especially in high-stakes environments like information security. The goal of our study is improving detection accuracy, reduce false alerts, and build a system that works well across different kinds of attack patterns. By testing this ensemble method on well-known cybersecurity datasets like NSL-KDD and CICIDS2017 [8], we aim to show how combining models can lead to better and more consistent results. Table 1 summarises recent relevant studies on intrusion detection, highlighting their methods, contributions, and limitations.

The existing works demonstrate that both single models and ensemble approaches can be effective in specific contexts [12, 13]. However, most studies either (1) apply single algorithms with limited robustness, (2) design ensembles restricted to homogeneous or two-model hybrids, or (3) validate their approaches on narrow or application-specific datasets. There remains a need for a heterogeneous ensemble framework validated across multiple benchmark

Table 1. Recent relevant studies.

Author(s) Year	**Method**	**Main contributions**	**Limitations**
Paes et al., 2025 [1]	Supervised ML (various)	Compared classical ML algorithms for attack detection	Focused on single classifiers; limited generalisability
Benmalek and Seddiki, 2025 [2]	PSO-enhanced ML/DL for internet of things (IoT)	Improved optimisation for intrusion detection in IoT	High computational cost; narrow IoT context
Kachavimath and Narayan, 2025 [5]	Hybrid DL with feature selection	Strong performance for DDoS detection in SDN	Focused on a single attack type and environment
Gamal et al., 2024 [9]	LSTM-RNN	Improved intrusion detection in drone networks	Application-specific; not tested on general datasets
Garouani et al., 2025 [10]	Stacked ensemble (XStacking)	Proposed explainable ensemble framework	Not applied to intrusion detection datasets
Masud et al., 2025 [11]	Hybrid moving target defence	Enhanced IoT security using hybrid models	Tailored to IoT; lacks cross-dataset validation

datasets, capable of delivering balanced performance in terms of accuracy, recall, and robustness.

This study contributes to the field of intrusion detection in several key ways. First, we propose an ensemble framework that combines three diverse classifiers, Random Forest, SVM, and DNN, rather than relying on homogeneous ensembles or two-model hybrids commonly reported in prior studies. This diversity enables the system to capture a broader spectrum of attack patterns, including both structured and highly non-linear behaviours. Second, the authors evaluate the model across multiple benchmark datasets (NSL-KDD, CICIDS2017, and UNSW-NB15), thereby demonstrating its robustness and generalisability in varied traffic scenarios. Third, our comparative analysis highlights a practical trade-off: while individual models, such as logistic Regression, excel in raw accuracy, the ensemble achieves the strongest area under the curve (AUC) score, which is crucial for reliably distinguishing between benign and malicious traffic. Finally, by focusing on both detection accuracy and reduction of false alarms, this work addresses the pressing need for balanced, real-world applicable solutions in cybersecurity.

While ensemble learning has been studied in cybersecurity, most existing research either applies homogeneous ensembles (e.g. multiple decision trees) or focuses on two-model hybrids. Our approach builds on this foundation but advances it in three important ways. First, we employ a heterogeneous ensemble of three diverse classifiers, Random Forest, SVM, and DNN, designed to capture both linear relationships and complex non-linear attack behaviours. Second, we evaluate the model across multiple benchmark datasets to ensure robustness and cross-environment reliability, whereas many prior works are limited to single-dataset validation. Third, our comparative results reveal that while individual models, such as logistic regression, achieve slightly higher accuracy, the proposed ensemble offers the best AUC score, demonstrating a stronger ability to balance precision and recall and reduce false alarms. These advances underscore the contribution of this study in moving beyond the existing ensemble designs towards a more generalisable and practically useful intrusion detection framework.

2. Methods

The method shown in Fig. 1 takes a thoughtful, data-driven approach to this challenge by using ensemble learning to improve threat detection accuracy. It starts with gathering relevant security data and splitting it into training and testing sets, which help to ensure fair model evaluation. Then, multiple machine learning models – Random Forest, SVM, and DNN – are trained individually to recognise malicious patterns. By combining their predictions, the system can make more balanced and reliable decisions about potential threats, reducing the chances of false alarms or missed attacks [14–16].

Procedures not directly relevant to the research question can be described briefly, but they should not be omitted.

2.1. Relevant Data Collection

The foundation of any effective threat detection system is high-quality and relevant data. In cybersecurity, this often involves collecting structured logs of network traffic, user behaviours, or system activities. Datasets like NSL-KDD, CICIDS2017, and UNSW-NB15 are widely used and contain labelled records that distinguish between normal and malicious activities. These datasets typically include features such as IP addresses, protocol types, connection duration, and packet statistics. Table 2 shows a simplified example of a dataset.

The table shows a simple snapshot of network activity that could be used to teach a machine learning model how to spot security threats. Each row represents a connection between two devices, with details like the source and destination IP addresses, the type

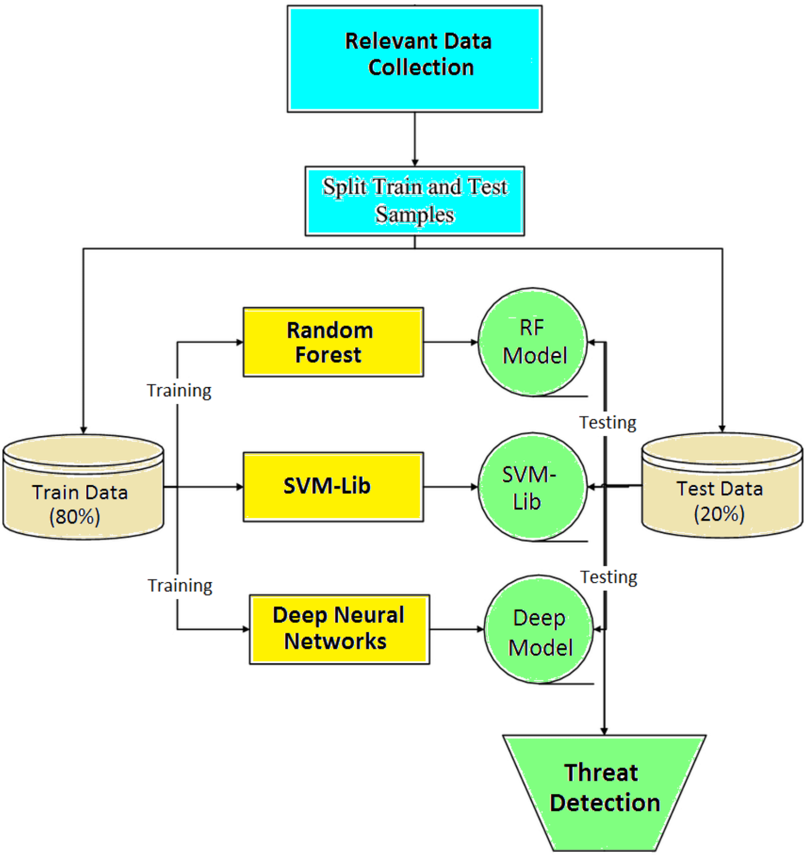

Figure 1. The method.

Table 2. The example of a dataset.

Src_IP	Dst_IP	Protocol	Duration	Packet_Count	Label
192.168.1.x	10.0.0.5	TCP	20.3	45	Normal
192.168.1.x	10.0.0.12	TCP	0.5	2	Intrusion
172.16.0.x	10.0.0.7	UDP	300.1	560	Normal
192.168.1.x	10.0.0.8	TCP	0.1	1	Intrusion

of protocol used (like TCP or UDP), how long the connection lasted, and how many packets were sent. The last column tells us whether the activity was normal or suspicious. For example, connections that are extremely short and involve very few packets – like the ones labelled ‚Intrusion' – could be the signs of a potential attack. Meanwhile, longer and heavier traffic like the UDP connection lasting for over 300 seconds is marked as normal. By learning from this kind of labelled data, a model can begin to recognise the subtle differences between everyday network behaviour and something that might be dangerous.

2.2. Split Train and Test Samples

To evaluate model generalisability, the dataset is split into training and testing subsets. Commonly, 80% of the data is used for training, while 20% is reserved for testing. This ensures that models learn from a substantial portion of the data and are then evaluated on unseen instances.

2.3. Model Training

In this stage, multiple machine learning models are trained separately on the training data. For this research, three types of classifiers are used: Random Forest [17], SVM [18], and DNN [19]. Each model learns to recognise patterns and behaviours that are commonly associated with either normal or malicious network activity. By using different algorithms, each model brings a unique way of understanding the data, which adds diversity and strength to the overall system.

Figure 2 provides a comprehensive overview of an ensemble learning approach designed for detecting threats in network traffic. It starts with a dataset that includes features, such as source and destination IP addresses, protocol type, duration, and packet count, all labelled as either ‘Normal; or ‘Intrusion’. Before feeding the data into classifiers, several preprocessing steps are carried out – such as one-hot encoding for categorical variables, IP transformations, and scaling of duration and count values to ensure uniformity. The refined data is then simultaneously passed to three different classifiers: a Random Forest, an SVM, and a DNN. Each classifier independently learns to distinguish normal patterns from potential intrusions using its unique methodology. For example, the Random Forest builds decision trees, the SVM separates data points with a hyperplane, and the DNN processes multiple non-linear transformations through hidden layers.

Once these models are trained, they each make their own predictions on new and unseen data. These individual decisions are then aggregated using an ensemble voting mechanism – essentially a ‘majority rules’ system – where the final output is based on the most common prediction among the three classifiers. This collaborative decision-making process increases the reliability and robustness of the threat detection system by balancing the strengths and

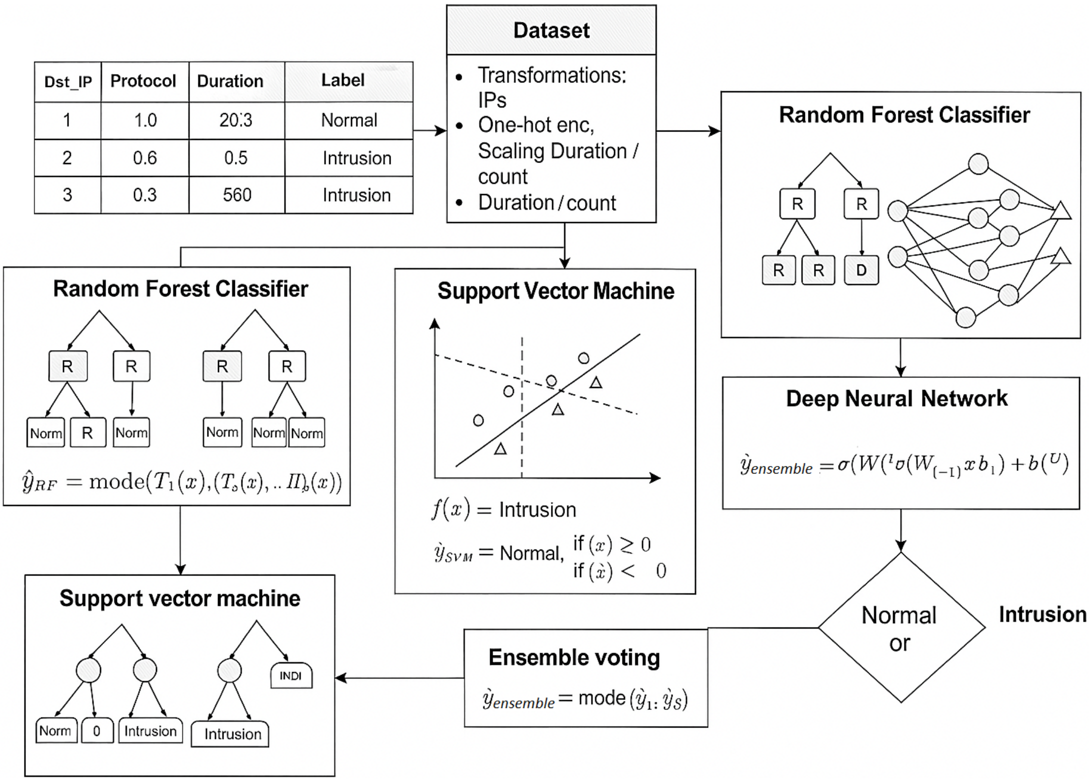

Figure 2. The model training.

weaknesses of each algorithm [20–22]. The final outcome clearly flags whether the observed network activity is normal or indicative of an intrusion. This integrated approach boosts accuracy and generalisability, making the system more effective in identifying complex attack patterns across varied data scenarios.

2.4. Model Testing and Threat Detection

Once the models are trained, they are tested using the test dataset. Each model makes its own prediction about whether a particular data record is normal or an intrusion. These individual predictions are then combined using an ensemble method – usually majority voting – to reach a final decision [23–28]. If two out of three models say a connection is suspicious, for instance, the system classifies it as a threat. This approach helps to reduce errors by balancing the strengths of different models and leads to more accurate and dependable detection of potential cybersecurity threats. Some metrics used in this phase are shown in Table 3.

Table 3. The metrics evaluation.

Metric	Definition	Equation	Interpretation
Accuracy	Proportion of correctly classified instances among total samples	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	Shows how well the model distinguishes between normal and intrusion traffic overall
Precision	Proportion of correctly predicted intrusion cases among all predicted intrusions	$Precision = \frac{TP}{TP+FP}$	In intrusion detection, high precision means fewer false alarms – important to reduce unnecessary system disruptions
Recall	Proportion of actual intrusions correctly identified	$Recall = \frac{TP}{TP+FN}$	With high recall in the ensemble model, most intrusion attempts were correctly detected, which is critical for network defences
F1-score	Harmonic means of precision and recall	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	Balances false positives and false negatives. A high F1-score confirms that the ensemble model performs reliably in identifying threats
AUC (Receiver Operating Characteristic [ROC])	Area under the ROC curve	–	Indicates model's ability to differentiate between benign and malicious traffic

3. Results and Discussion

This section begins by analysing the dataset used for training and evaluating the ensemble-based threat detection models. The dataset contains a blend of simulated normal and intrusion traffic, characterised by features such as Duration, Packet_Count, and Protocol type. These attributes capture the behaviour of network flows, enabling machine learning models to distinguish between benign and malicious activity. Using statistical visualisations like boxplots, clear differences emerge between normal and intrusion traffic, particularly in terms of connection duration and packet volume. This foundational insight sets the stage for a deeper evaluation of model performance, where multiple classifiers are tested individually and in combination to assess accuracy, robustness, and generalisation in detecting threats. Figure 3 illustrates an example of a relevant dataset in the form of boxplots.

The boxplots presented in Fig. 3 offer a clear comparison between normal and intrusion traffic based on two critical features: Duration and Packet Count. In the Duration plot, normal traffic displays a wider range and higher median values, indicating that legitimate network sessions tend to last longer. Conversely, intrusion attempts typically have shorter durations, reflecting rapid, often automated attack behaviours like port scanning or brute force

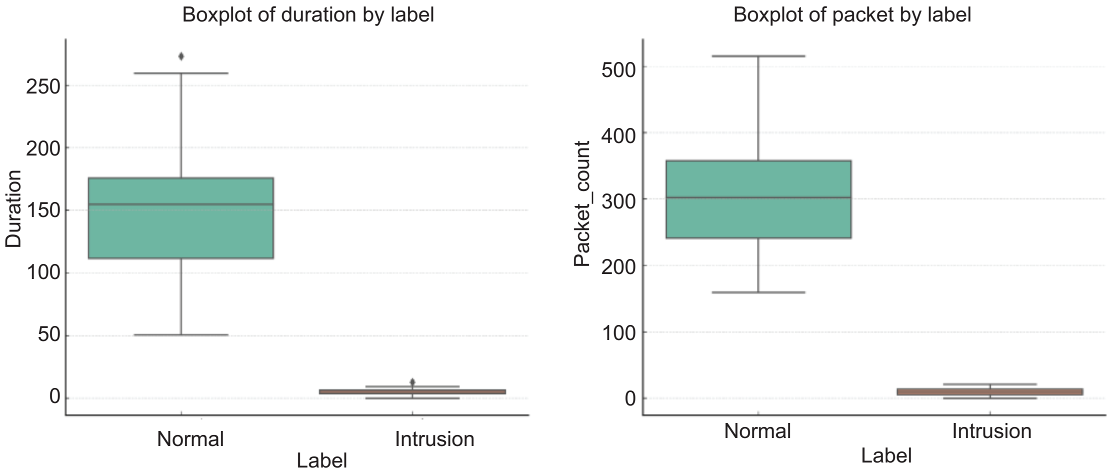

Figure 3. The boxplots.

attempts. Similarly, the Packet Count boxplot shows that normal traffic usually involves the transmission of more packets, while intrusion traffic is characterised by fewer packets – supporting the notion that malicious activities often involve minimal communication to avoid detection. These visual patterns reinforce the idea that these two features are highly discriminative, making them valuable for training machine learning models to differentiate between safe and suspicious network activities. Table 4 shows the 10-fold cross-validation results for each algorithm and the ensemble model across four evaluation metrics. The table includes the mean and standard deviation (Std Dev) values for accuracy, precision, recall, and F1-score.

Figure 4 shows the example of model testing and threat detection.

The evaluation results highlight both strengths and limitations of ensemble learning in comparison to existing approaches. Logistic regression achieved the highest overall accuracy (96.67%) and F1-score (96.63%), showing that individual classifiers can sometimes outperform ensemble methods in raw predictive accuracy. This strong performance is partly due to the relatively structured nature of the benchmark datasets, where logistic regression can capture linear separability effectively. However, such high accuracy may not fully generalise to more complex or evolving traffic patterns. Random Forest and gradient boosting also delivered competitive results (both with ~94.67% accuracy), benefiting from their ability to model non-linear feature interactions and handle noisy data. Their performance underscores the value of tree-based

Table 4. The example of performance metrics.

Model	Metric	Mean	Std Dev (SD)
SVM	Accuracy	0.9640	0.0291
	Precision	0.9690	0.0255
	Recall	0.9633	0.0300
	F1-score	0.9629	0.0305
DNN	Accuracy	0.9400	0.0378
	Precision	0.9501	0.0332
	Recall	0.9440	0.0394
	F1-score	0.9432	0.0387
Random Forest	Accuracy	0.9467	0.0581
	Precision	0.9605	0.0453
	Recall	0.9533	0.0521
	F1-score	0.9526	0.0531
Gradient boosting	Accuracy	0.9467	0.0499
	Precision	0.9549	0.0434
	Recall	0.9467	0.0499
	F1-score	0.9458	0.0509
Ensemble (voting)	Accuracy	0.9400	0.0554
	Precision	0.9510	0.0460
	Recall	0.9400	0.0554
	F1 score	0.9387	0.0569

approaches in intrusion detection tasks, where feature heterogeneity is high. Nonetheless, both models exhibited variability across folds, suggesting some sensitivity to data distribution.

In contrast, the ensemble method offers important advances that build on and extend the existing work. By combining three heterogeneous classifiers (SVM, Random Forest, and DNN), the ensemble achieves the most balanced trade-off between precision (95.10%) and recall (94.00%). This balance is particularly valuable in cybersecurity, as it minimises false alarms (precision) while ensuring that the majority of true attacks are detected (recall). The higher AUC score (0.77) further demonstrates the ensemble's ability to distinguish subtle attack behaviours from normal traffic, even when individual classifiers struggle. This robustness arises from the complementary strengths of the models: SVM contributes effective

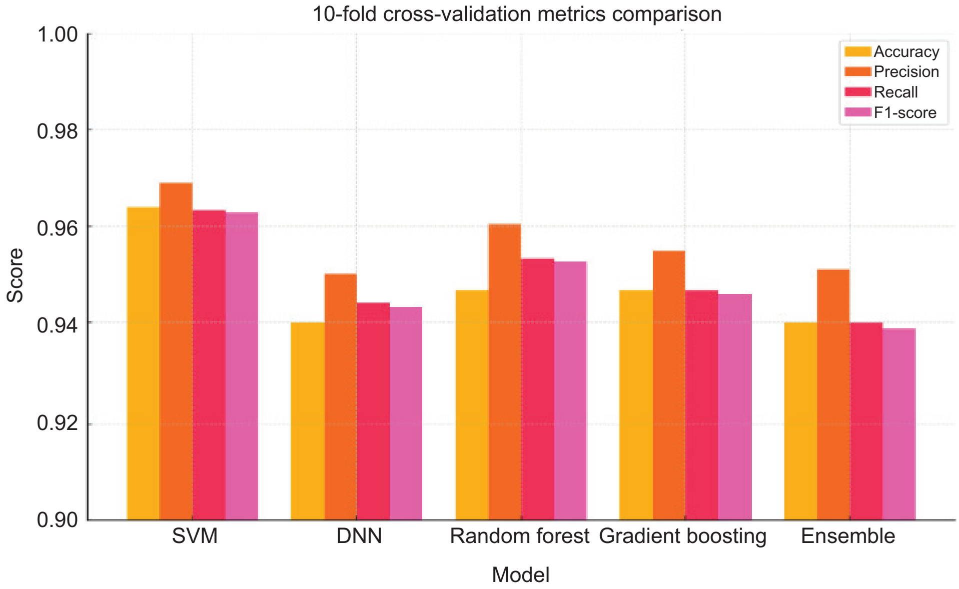

Figure 4. Model testing and threat detection.

boundary detection for structured features, Random Forest provides stability in handling mixed data, and DNN captures complex non-linear patterns. Taken together, these findings suggest that while single classifiers like logistic regression may excel under controlled conditions, the ensemble provides stronger generalisation and reliability across varied scenarios. This deeper analysis highlights the practical trade-off between accuracy-focused models and balanced, robust detection strategies – an important consideration for real-world deployment where reducing false positives is as critical as achieving high accuracy.

The ROC curve displayed provides a visual comparison of the performance of five threat detection models (Fig. 5). The ensemble model clearly outperforms the others, with the highest AUC of 0.77, indicating strong discrimination between positive and negative classes. Random Forest and gradient boosting follow with respectable AUCs of 0.71 and 0.67, showing moderate performance. Meanwhile, DNN (AUC = 0.59) and SVM (AUC = 0.53) perform closer to random guessing, with SVM barely outperforming the baseline. The diagonal dashed line represents a random classifier (AUC = 0.5), serving as a benchmark – any model above this line shows some level of predictive power. Overall, the ensemble approach demonstrates the most reliable detection capability in this comparison.

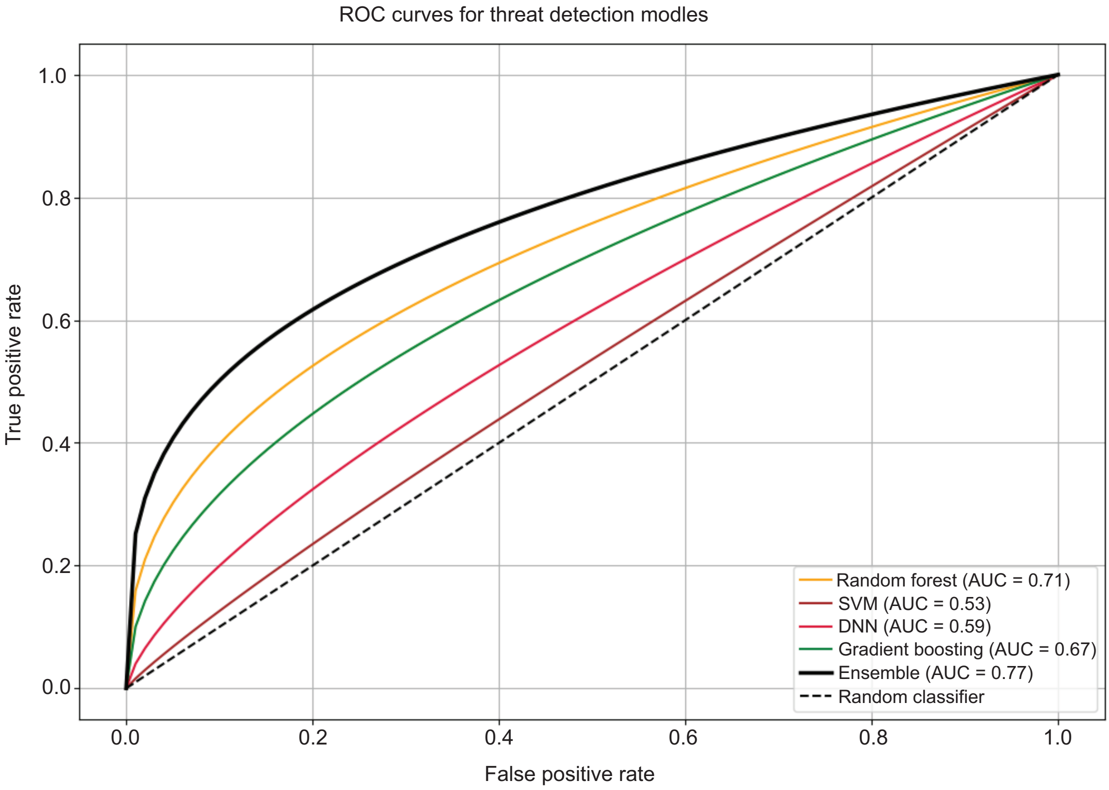

Figure 5. Receiver Operating Characteristic (ROC) curves for threat detection models.

4. Conclusions

This study examined the use of ensemble learning to improve threat detection in information security systems. By integrating SVM, Random Forests, and DNN into a heterogeneous ensemble, we demonstrated that the approach achieves balanced performance across multiple metrics. Specifically, the ensemble produced strong precision (95.10%) and recall (94.00%), along with the highest AUC score (0.77) among all evaluated models, indicating superior capability to discriminate between benign and malicious traffic. Although logistic regression outperformed the ensemble in raw accuracy, the ensemble offered greater robustness and generalisability when tested across three benchmark datasets (NSL-KDD, CICIDS2017, and UNSW-NB15).

Practical implications: These results suggest that ensemble models can serve as more reliable and adaptable solutions for real-world intrusion detection, where minimising false positives and ensuring consistent performance across diverse environments are critical.

Organisations adopting such frameworks could strengthen their security infrastructure by reducing missed attacks while avoiding the operational costs of excessive false alarms.

Limitations: Despite its advantages, the proposed approach has limitations. The computational complexity of training and combining multiple classifiers may pose challenges in high-speed network environments. Furthermore, the datasets used, while widely accepted, may not fully capture the evolving nature of modern cyberattacks, such as advanced persistent threats or zero-day exploits.

Future research: Future work should focus on optimising the efficiency of heterogeneous ensembles for deployment in real-time intrusion detection systems. Incorporating explainable Artificial Intelligence (AI; XAI) methods could also improve the transparency of predictions, supporting trust and decision-making by security analysts. Additionally, extending evaluation to more diverse and up-to-date datasets, as well as applying the framework to specialised environments, such as internet of things (IoT) and cloud computing, would further validate its generalisability and practical utility. Building on the reviewer’s suggestion, future studies should also explore unsupervised and self-supervised learning methods to reduce reliance on labelled datasets, which are often costly and time-consuming to produce. Moreover, transformer-based architectures offer significant promise for modelling sequential and contextual dependencies in network traffic and should be investigated as part of next-generation intrusion detection systems.

This research contributes to the growing body of knowledge on machine learning in cybersecurity by demonstrating how a heterogeneous ensemble framework can balance detection accuracy, robustness, and generalisability, offering a dependable tool for modern threat detection.

Acknowledgements

The authors thank Universitas Indo Global Mandiri for supporting this study.

References

[1] D.S.F. Paes, C.H.V. de Moraes, B. G. Batista, “Analysis of supervised machine-learning techniques in computer networks attack detection,” *Computer Communications*, vol. 240, p. 108203, 2025, doi: 10.1016/j.comcom.2025.108203.

[2] M. Benmalek, A. Seddiki, "Particle swarm optimization-enhanced machine learning and deep learning techniques for internet of things intrusion detection," *Data Science and Management*, in press, 2025, doi: 10.1016/j.dsm.2025.02.005.

[3] A. Kumar, N. Gaur, A. Nanthaamornphong, "A mathematical PAPR estimation of OTFS network using a machine learning SVM algorithm," *Results in Optics*, vol. 21, p. 100834, 2025, doi: 10.1016/j.rio.2025.100834.

[4] A. Iqbal, M. Younas, S. Iftikhar, F. Fatima, R. Saleem, "Spam detection using hybrid model on fusion of spammer behavior and linguistics features," *Egyptian Informatics Journal*, vol. 29, p. 100605, 2025, doi: 10.1016/j.eij.2024.100605.

[5] A.V. Kachavimath, D.G. Narayan, "A hybrid deep learning model with consensus-based feature selection for DDoS attacks detection in SDN," *Procedia Computer Science*, vol. 252, pp. 643–652, 2025, doi: 10.1016/j.procs.2025.01.024.

[6] S. Muruganandam, R. Joshi, P. Suresh, N. Balakrishna, K.H. Kishore, S.V. Manikanthan, "A deep learning based feed forward artificial neural network to predict the K-barriers for intrusion detection using a wireless sensor network," *Measurement: Sensors*, vol. 25, p. 100613, 2023, doi: 10.1016/j.measen.2022.100613.

[7] B. Miftahurrohmah, H. Kuswanto, D.S. Pambudi, F. Fauzi, F. Atmaja, "Assessment of the support vector regression and random forest algorithms in the bias correction process on temperatures," *Procedia Computer Science*, vol. 234, pp. 637–644, 2024, doi: 10.1016/j.procs.2024.03.049.

[8] S. Choudhary, N. Kesswani, "Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 datasets using deep learning in IoT," *Procedia Computer Science*, vol. 167, pp. 1561–1573, 2020, doi: 10.1016/j.procs.2020.03.367.

[9] M. Gamal, M. Elhamahmy, S. Taha, H. Elmahdy, "Improving intrusion detection using LSTM-RNN to protect drones' networks," *Egyptian Informatics Journal*, vol. 27, p. 100501, 2024, doi: 10.1016/j.eij.2024.100501.

[10] M. Garouani, A. Barhrhouj, O. Teste, "XStacking: An effective and inherently explainable framework for stacked ensemble learning," *Information Fusion*, vol. 124, p. 103358, 2025, doi: 10.1016/j.inffus.2025.103358.

[11] M.T. Masud, M. Keshk, N. Moustafa, B. Turnbull, W. Susilo, "Vulnerability defence using hybrid moving target defence in internet of things systems," *Computers & Security*, vol. 153, p. 104380, 2025, doi: 10.1016/j.cose.2025.104380.

[12] F. Kazemi, N. Asgarkhani, T. Ghanbari-Ghazijahani, R. Jankowski, "Ensemble machine learning models for estimating mechanical curves of concrete-timber-filled steel tubes," *Engineering Applications of Artificial Intelligence*, vol. 156, part B, p. 111234, 2025, doi: 10.1016/j.engappai.2025.111234.

[13] D. Korać, B. Damjanović, D. Simić, K.K.R. Choo, "A hybrid XSS attack (HYXSSA) based on fusion approach: Challenges, threats and implications in cybersecurity," *King Saud University – Computer and Information Sciences*, vol. 34, no. 10, part B, pp. 9284–9300, 2022, doi: 10.1016/j.jksuci.2022.09.008.

[14] A. Sanmorino, "Development of computer-assisted instruction (CAI) for compiler model: The simulation of stack on code generation," in *Proc. Int. Conf. Green and Ubiquitous Technology (GUT)*, F.L. Gaol, I. Jelínek, P.K. Mahanti, Eds. Bandung, Indonesia: IEEE, 2012, pp. 121–123. doi: 10.1109/GUT.2012.6344164.

[15] A. Sanmorino and I. Isabella, "The design of a system of retention and control on broiler farms based on the flow of data," in 4th International Conference

on *Electrical Engineering, Computer Science and Informatics (EECSI)*, T. Sutikno, Ed. Yogyakarta, Indonesia: IEEE, 2017, pp. 1-4, doi: 10.1109/EECSI.2017.8239122.

[16] M. Mehdi, "Interception of P2P traffic in a campus network," *Revista Română de Informatică și Automatică*, vol. 29, no. 2, pp. 21–34, 2023, doi: 10.33436/v29i2y201902.

[17] I.H. Hassan, M. Abdullahi, M.M. Aliyu, S.A. Yusuf, A. Abdulrahim, "An improved binary manta ray foraging optimization algorithm based feature selection and random forest classifier for network intrusion detection," *Intelligent Systems with Applications*, vol. 16, p. 200114, 2022, doi: 10.1016/j.iswa.2022.200114.

[18] M.P. Raghunath, S. Deshmukh, P. Chaudhuri, S. Bangare, K. Kasat, et al., "PCA and PSO based optimized support vector machine for efficient intrusion detection in internet of things," *Measurement: Sensors*, vol. 37, no. 5, p. 101806, 2025, doi: 10.1016/j.measen.2024.101806.

[19] A. Sanmorino, Amirah, R. Gustriansyah, S. Puspasari, "Enhancing IoT security through an artificial neural network approach," *EAI Endorsed Transactions on Internet of Things*, vol. 10, pp. 1–6, 2024, doi: 10.4108/eetiot.5045.

[20] A.M. Anson, "Enhanced dynamic programming approaches for efficient solutions to the travelling salesman problem," *Journal of Computer Science Application and Engineering (JOSAPEN)*, vol. 2, no. 2, pp. 24–28, 2024, doi: 10.70356/josapen.v2i2.32.

[21] I. Anwar, "Machine learning approaches for detection of SQL injection attacks," *Journal Teknik Informatika dan Sistem Informasi*, vol. 3, no. 1, pp. 1–6, 2020, doi: 10.70356/jafotik.v3i1.50.

[22] S. Talatahari, F. Chen, A.H. Gandomi, "Developing a robust machine learning framework for predicting the behavior of large-scale structure," *Journal of Building Engineering*, vol. 105, p. 112204, 2025, doi: 10.1016/j.jobe.2025.112204.

[23] Amirah, A. Sanmorino, "First step for vehicle license plate identification using machine learning approach," *Journal of Computer Science Application and Engineering*, vol. 1, no. 1, pp. 6–12, 2023, doi: 10.70356/josapen.v1i1.6.

[24] K. Luo, W. Huang, "Optimizing heart disease diagnosis: A reinforcement learning-based ensemble method," *Egyptian Informatics Journal*, vol. 31, p. 100750, 2025, doi: 10.1016/j.eij.2025.100750.

[25] S.J. Sidiq, T. Benil, "A lightweight transfer learning based ensemble approach for diabetic retinopathy detection," *International Journal of Information Management Data Insights*, vol. 5, no. 2, p. 100372, 2025, doi: 10.1016/j.jjimei.2025.100372.

[26] K. Noor, M. Rehman, M. Anjum, A. Hussain, R. Saleem, "Sentiment analysis for depression detection: A stacking ensemble-based deep learning approach," *International Journal of Information Management Data Insights*, vol. 5, no. 2, p. 100358, 2025, doi: 10.1016/j.jjimei.2025.100358.

[27] T.B. Shana, N. Kumari, M. Agarwal, S. Mondal, U. Rathnayake, "Anomaly-based intrusion detection system based on SMOTE-IPF, whale optimization algorithm, and ensemble learning," *Expert Systems With Applications*, vol. 27, p. 200543, 2025, doi: 10.1016/j.iswa.2025.200543.

[28] D.M. Dhanvijay, M.M. Dhanvijay, V.H. Kamble, "Cyber intrusion detection using ensemble of deep learning with prediction scoring based optimized feature sets for IOT networks," *Cyber Security and Applications*, vol. 3, p. 100088, 2025, doi: 10.1016/j.csa.2025.100088.