Exploring the Application of Generative Adversarial Networks for Encrypted Traffic Classification in SDN-Enabled Home Networks

Gowthami Chopparapu
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
gouthami526@gmail.com (corresponding author)

S .Kavitha
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
Kavithabtech05@gmail.com

Received: 12 July 2025 | Revised: 8 August 2025 | Accepted: 20 August 2025

 DOI: https://doi.org/10.48084/etasr.13342

ABSTRACT

The rapid growth of encrypted network traffic poses significant challenges for traditional classification methods, particularly in Software-Defined Networking (SDN)-enabled home networks, where direct packet inspection is restricted by privacy requirements. To address this, we propose a Generative Adversarial Network (GAN)-based framework that classifies encrypted traffic using only flow metadata and statistical features, without requiring decryption. The proposed model leverages adversarial learning to capture complex traffic patterns and distinguish between benign and malicious flows, ensuring both high accuracy and privacy preservation. Experimental evaluation on the ISCX VPN dataset demonstrates that our approach outperforms conventional classifiers such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Support Vector Machines (SVMs), and Random Forest, achieving 98.8% accuracy, precision, and recall, an Area Under the Curve (AUC) of 0.995, and a low inference time of 2 ms. Furthermore, the model achieves very low false positive and false negative rates (0.006 for each), highlighting its robustness for real-time applications. This framework provides a scalable, efficient, and privacy-preserving solution for encrypted traffic classification in SDN-enabled home networks, offering a promising direction for secure and intelligent network management.

Keywords-Generative Adversarial Networks (GANs); encrypted traffic classification; machine learning; cybersecurity; deep learning; network security

I. INTRODUCTION

The exponential rise of encrypted network traffic in recent years has fundamentally reshaped the landscape of cybersecurity and network management [1]. Encryption, while critical for safeguarding user privacy and ensuring secure communications, acts as a double-edged sword by protecting sensitive data from unauthorized access while rendering conventional traffic monitoring and classification methods ineffective [2]. The traditional reliance on techniques such as Deep Packet Inspection (DPI) has become obsolete in the face of encryption, leaving network administrators and security professionals grappling with new challenges. Addressing this duality, especially in Software-Defined Networking (SDN)-based environments, necessitates innovative and privacy-preserving solutions [3].

SDN has emerged as a revolutionary paradigm for modern network architecture, offering unprecedented flexibility and programmability [4]. By decoupling the control and data planes, SDN enables dynamic traffic management and efficient resource allocation across diverse network applications. However, the increased adoption of encrypted traffic in SDN-enabled networks poses significant challenges for maintaining both security and performance [5]. The inability to directly inspect encrypted flows, combined with the growing traffic volume, highlights the urgent need for advanced, scalable, and privacy-preserving classification techniques [6]. Recent studies have explored a variety of approaches to encrypted traffic classification, including deep learning architectures, generative models for synthetic data generation, anomaly detection mechanisms for SDN environments, and privacy-preserving analytics for network monitoring [7]. While these approaches have shown promising results, most rely on partial payload inspection, are limited to specific network configurations, or struggle with generalization across diverse encrypted traffic patterns [8]. This creates a gap in developing a scalable,

privacy-preserving framework capable of accurately classifying encrypted traffic in real time within SDN-enabled home networks. To address this challenge, we propose a Generative Adversarial Network (GAN)-based framework that classifies encrypted traffic using only observable metadata and statistical features such as packet sizes, flow durations, and inter-arrival times [9, 10]. Unlike conventional methods, the GAN-based approach does not require decrypting the traffic, thereby maintaining confidentiality. Furthermore, GANs excel at learning nuanced patterns from limited datasets, making them well-suited for handling encrypted traffic characterized by sparse or incomplete feature representations [11]. This research applies GANs to a domain traditionally dominated by deterministic and statistical models, offering a promising direction for secure and intelligent network management [12]. Figure 1 presents an overview the overview of GAN structure.

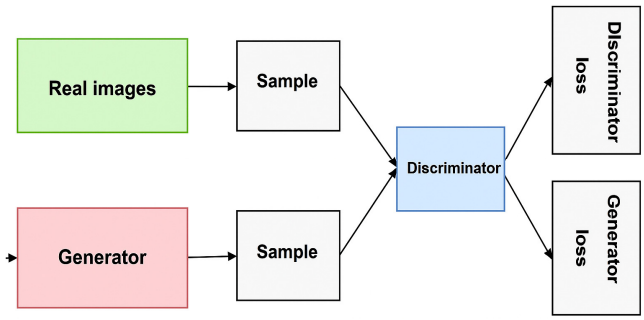

Fig. 1. Overview of the GAN structure.

II. METHODOLOGY

This section delineates a structured methodology employed for the classification of encrypted network traffic within SDN-enabled home network environments using GANs [13]. The approach initiates with the systematic collection and rigorous preprocessing of encrypted network traffic data, placing particular emphasis on preserving user privacy and data confidentiality [14]. Following data acquisition, meticulous feature engineering and statistical analysis are conducted to identify meaningful patterns within the traffic metadata. Subsequently, a GAN model undergoes comprehensive adversarial training, facilitating robust learning and generalization capabilities for distinguishing benign and malicious traffic types [15]. The methodology further involves critical performance evaluation, benchmarking the GAN model against traditional machine learning algorithms to validate its efficacy and superiority [16]. Finally, the procedure concludes with guidelines on effective deployment and real-time operational integration within SDN infrastructures. Figure 2 illustrates the flowchart of the proposed model, including the following steps:

- Encrypted traffic collection: Data are gathered from encrypted communication channels (e.g., HTTPS, SSL/TLS). This phase uses tools such as Wireshark or tcpdump to capture encrypted flows without inspecting payload content. Only metadata and observable traffic patterns are retained for privacy-preserving analysis [17]. The collected data form the foundation for subsequent traffic classification.
- Packet capture: In this step, packet-level data are intercepted using tools like tcpdump or TShark. The captured packets include headers, timestamps, and lengths that are useful for non-payload analysis [18]. The packets are saved in a structured format (e.g., .pcap) for further processing, enabling the extraction of time-based and size-based features for modeling.
- Flow metadata extraction: Flow metadata encompasses essential identifiers such as source IP address, destination IP address, source port, destination port, and the transport-layer protocol. The Flow ID, derived from these fields, acts as a unique identifier for each network session, enabling accurate aggregation and tracking of individual traffic flows. Such metadata are crucial for constructing flow-level features used in network traffic analysis, anomaly detection, and behavioral profiling, without requiring decryption of the actual content. It also provides valuable context for temporal correlation and sequence modeling in traffic classification systems.
- Traffic sampling: This involves selecting a representative subset of the traffic data to reduce volume. It balances data diversity with computational efficiency using techniques such as random or stratified sampling. $S \subseteq T$, where T is the total dataset and S is the sampled subset. Sampling ensures scalability without significant information loss.
- Packet size analysis: The statistical analysis of packet lengths reveals useful patterns about encrypted flows. Key features include average, minimum, and maximum packet sizes within each session. The mean and standard deviation of packet sizes are computed as follows:

$$\mu = (1/n)\sum_{i=1}^{N} P_i \quad (1)$$

$$\sigma = \sqrt{(1/n)\sum_{i=1}^{N}(P_i - \mu)^2} \quad (2)$$

where μ is the mean packet size, n is the total number of packets in the flow, and P_i is the size of the i-th packet in bytes. These values help detect anomalies, such as fixed-size attack packets.

- Entropy computation: Entropy quantifies the randomness or uncertainty in the packet size distribution, which is useful for identifying obfuscated or malicious traffic. The Shannon entropy for a flow is calculated as:

$$H = -\sum_x p(x)\log_2 p(x) \quad (3)$$

where $p(x)$ is the empirical probability of feature x. High entropy typically indicates encrypted or compressed content.

- Flow duration estimation: The duration of a traffic flow helps differentiate short-lived scans from persistent communication. It is computed by subtracting the timestamp of the first packet from that of the last packet in the flow:

$$Duration = T_{end} - T_{start} \quad (4)$$

Flow duration is a key temporal feature for behavioral analysis.

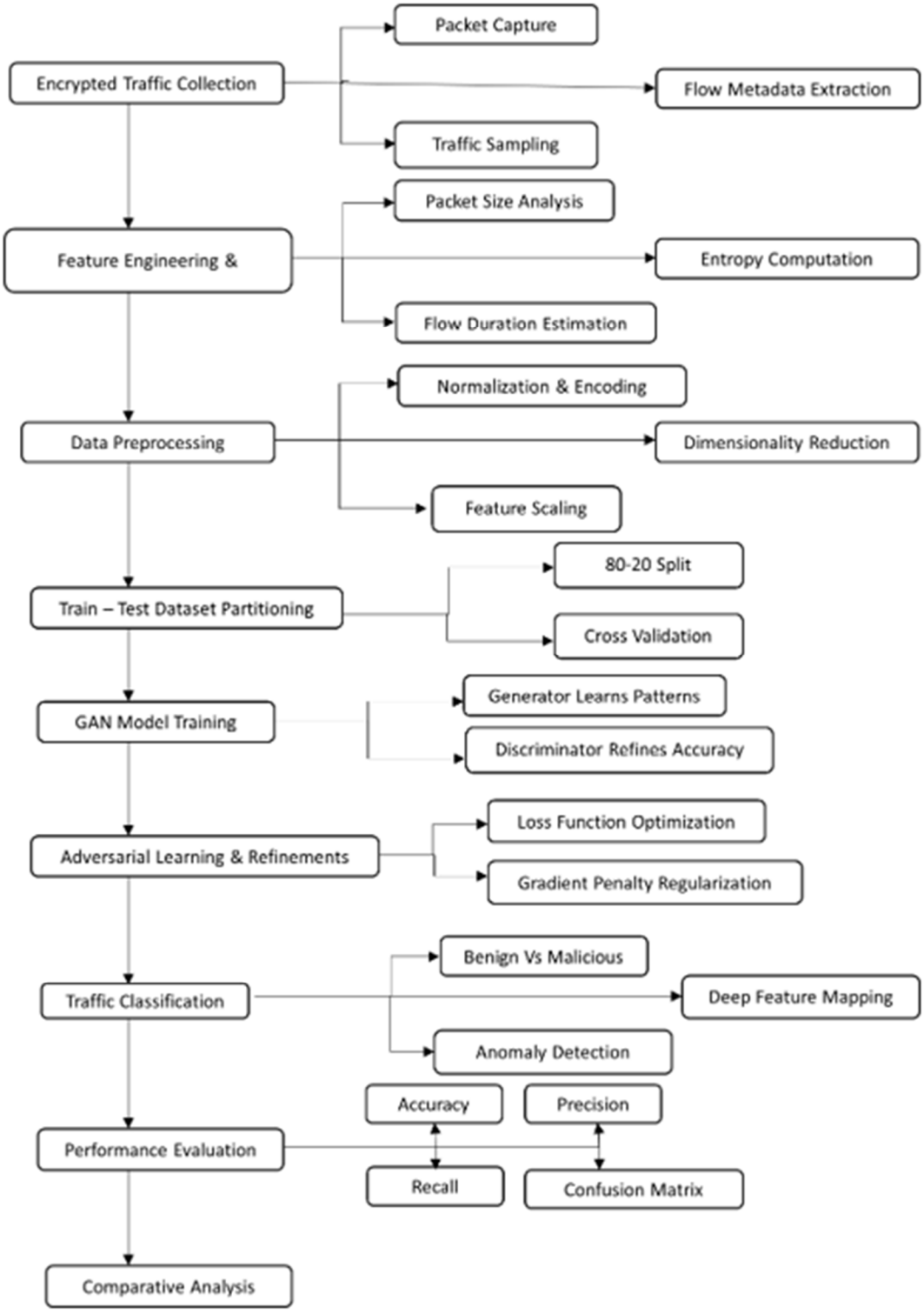

Fig. 2. Flowchart of the proposed GAN-based classification methodology.

- Normalization & encoding: Continuous features are scaled, and categorical values are encoded. Z-score normalization is used to bring all features to a standard scale. Label or one-hot encoding is applied to non-numeric features for model compatibility.
- Dimensionality reduction: Redundant or irrelevant features are removed to improve model performance. Principal Component Analysis (PCA) is used to reduce dimensions while preserving variance:

$$Y = XW \quad (5)$$

where W contains the eigenvectors of the covariance matrix of X. This step reduces overfitting and accelerates training.

- Feature scaling: All features are transformed to a comparable scale to prevent model bias, which is essential for gradient-based optimization algorithms.
- Train–test dataset partitioning: To ensure that the model is trained effectively and evaluated on unseen data, the dataset is divided into training (D_{train}) and testing (D_{test}) subsets, typically using an 80:20 ratio:

$$D_{train} \cup D_{test} = D \quad (6)$$

$$D_{train} \cap D_{test} = \emptyset \quad (7)$$

A stratified split maintains the class distribution in both subsets.

- Cross-validation: A technique to assess model generalizability across different data segments. In k-fold cross-validation, the data are split into k subsets; the model is trained on k-1 folds and tested on the remaining fold. This process is repeated k times and averaged to prevent overfitting and improve robustness.
- GAN model training: A GAN consists of a generator G and a discriminator D. The generator learns to synthesize realistic traffic samples, whereas the discriminator learns to distinguish between real and generated samples. Their loss functions are defined as:

$$\text{Loss}_D = -\text{E}[\log(D(G(z)))] \quad (8)$$

$$\text{Loss}_D = -\text{E}\left[\log\left(D(x)\right)\right] - \text{E}[\log(1 - D(G(z)))] \quad (9)$$

Training proceeds until the discriminator D can no longer distinguish between real and generated samples.

- Adversarial learning and refinements: This step involves iterative tuning of G and D to enhance generation and detection performance. A gradient penalty is used to stabilize GAN training, improving convergence and model stability.
- Traffic classification: Once the GAN-refined features are obtained, the classifier predicts whether a traffic flow is benign or malicious. The prediction is made using a softmax activation over the learned weights W and bias b:

$$\text{Prediction} = \text{argmax}(\text{Softmax}(Wx + b)) \quad (10)$$

Accurate classification is critical for proactive threat detection and real-time network defense.

- Anomaly detection: This step identifies deviations from learned normal traffic behaviors. The GAN generates baseline traffic, and anomalies are detected using distance or reconstruction error metrics: $\text{Anomaly Score} = ||x - G(z)||$ or Mahalanobis distance. This process is crucial for identifying zero-day attacks and novel threats.

III. RESULTS

The proposed GAN-based model for encrypted traffic classification in SDN-enabled home networks was evaluated using the ISCX VPN dataset [19]. The model's performance was compared with other baseline models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Support Vector Machines (SVMs), and Random Forest, based on key evaluation metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). The results are summarized in Table I.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	Dataset	Accuracy	Precision	Recall	F1-score	False positive rate	False negative rate	AUC
CNN	ISCX VPN	0.87	0.85	0.86	0.85	0.1	0.09	0.88
RNN	ISCX VPN	0.84	0.83	0.82	0.82	0.12	0.11	0.85
Random Forest	ISCX VPN	0.81	0.78	0.8	0.79	0.15	0.14	0.83
SVM	ISCX VPN	0.79	0.76	0.77	0.76	0.18	0.16	0.8
GAN-based model (proposed)	ISCX VPN	0.988	0.988	0.988	0.988	0.006	0.006	0.995

The proposed framework was developed and tested using the following hardware and software environment:

- Hardware: Intel® Core™ i7-12700K CPU @ 3.60 GHz, 32 GB DDR4 RAM, NVIDIA GeForce RTX 3060 GPU with 12 GB VRAM, 1 TB NVMe SSD.
- Software: Ubuntu 22.04 LTS (64-bit) OS, Python 3.10, TensorFlow 2.13, Keras 2.13, Scikit-learn 1.3, Pandas, NumPy, Matplotlib, and Wireshark 4.0.7 for packet capture.
- SDN simulation tools: Mininet 2.3.0 for SDN topology emulation, Open vSwitch 2.17.6, and Ryu Controller 4.34 for flow control.

The model operates on flow-level, payload-agnostic features extracted from packet headers and timing metadata. The full feature set includes:

- Size/volume statistics: Packet count, byte count, mean, standard deviation, minimum, and maximum packet size.
- Temporal statistics: Flow duration, mean and standard deviation of inter-arrival time.
- Entropy measures: Shannon entropy of packet sizes and of inter-arrival times.
- Transport metadata: Source port, destination port, and transport protocol (categorical variables encoded via one-hot/label encoding).

The source and destination IPs were used solely to construct the flow identifier and were excluded from the model input to avoid overfitting and preserve privacy. Feature selection was performed by combining univariate mutual-information ranking with cross-validated recursive feature elimination on the training set. The retained subset (used for all reported results) maximized the validation AUC while minimizing redundancy. Continuous features were standardized prior to training, and dimensionality reduction using PCA was explored during ablation but was not applied in the final configuration. Table II presents the training time and inference time comparison of the different models.

TABLE II. COMPARISON OF TRAINING TIME AND INFERENCE TIME COMPARISON FOR DIFFERENT MODELS

Model	Dataset	Training time (s)	Inference time (ms)
CNN	ISCX VPN	150	5
RNN	ISCX VPN	200	7
Random Forest	ISCX VPN	120	4
SVM	ISCX VPN	100	3
GAN-based model (proposed)	ISCX VPN	320	2

The confusion matrix for the proposed GAN-based model, shown in Figure 3, highlights the model's classification accuracy and error distribution. The model correctly classified 12,350 Virtual Private Network (VPN) samples and 12,350 non-VPN samples out of 25,000 total samples. The number of misclassifications was limited to 150 false positives (non-VPN misclassified as VPN) and 150 false negatives (VPN misclassified as non-VPN). The balanced classification across both classes reflects the model's ability to effectively handle class imbalance and accurately distinguish between encrypted VPN and non-VPN traffic.

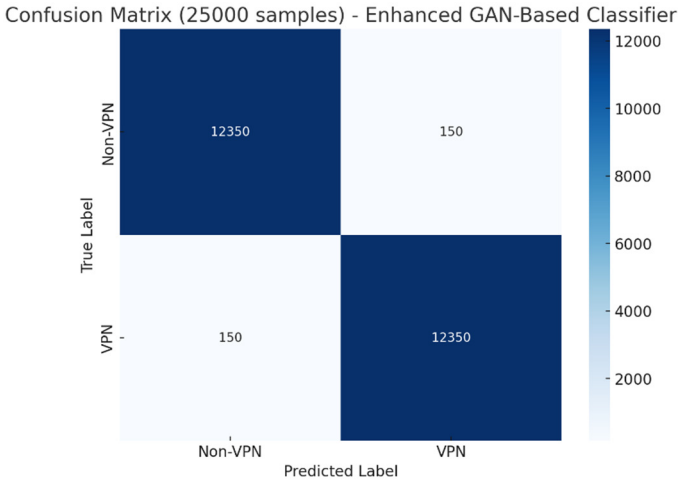

Fig. 3. Confusion matrix of the proposed model for encrypted traffic classification.

Figure 4 depicts the Receiver Operating Characteristic (ROC) curve for the proposed GAN-based model, with an AUC of 0.988, indicating excellent separation between positive and negative classes. The ROC curve approaches the top-left corner of the plot, indicating high sensitivity and specificity across different classification thresholds.

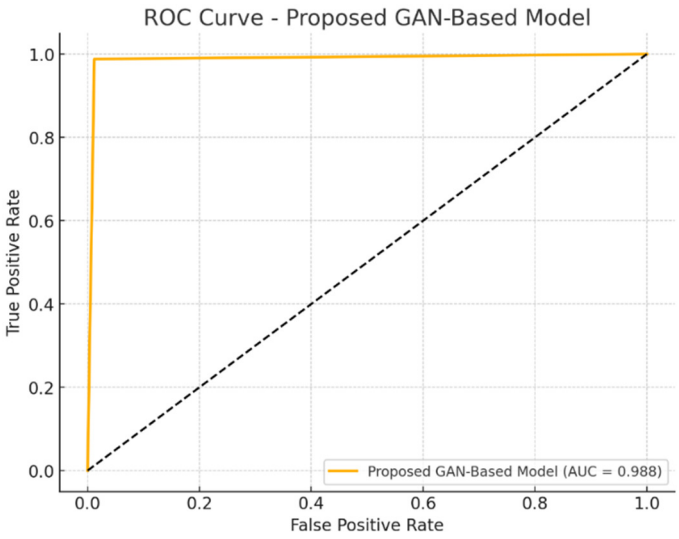

Fig. 4. ROC curve for the proposed GAN-based model.

Figure 5 presents the precision-recall curve, demonstrating that the model maintained high precision and recall across different classification thresholds. The nearly flat curve close to the maximum value indicates that the model achieves high classification accuracy even under varying conditions.

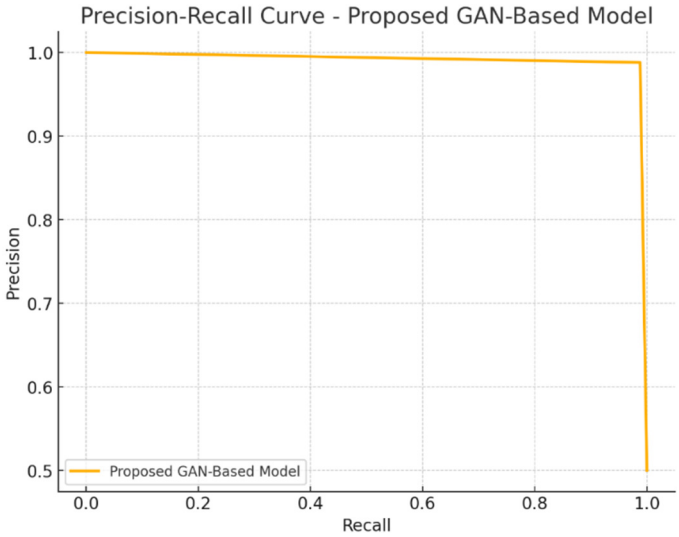

Fig. 5. Precision-recall curve for the proposed GAN-based model.

Figure 6 shows the training and validation accuracy curves, which demonstrate a steady increase over the training epochs. The final validation accuracy of 98.8% closely aligns with the training accuracy, confirming the model's strong generalization capability.

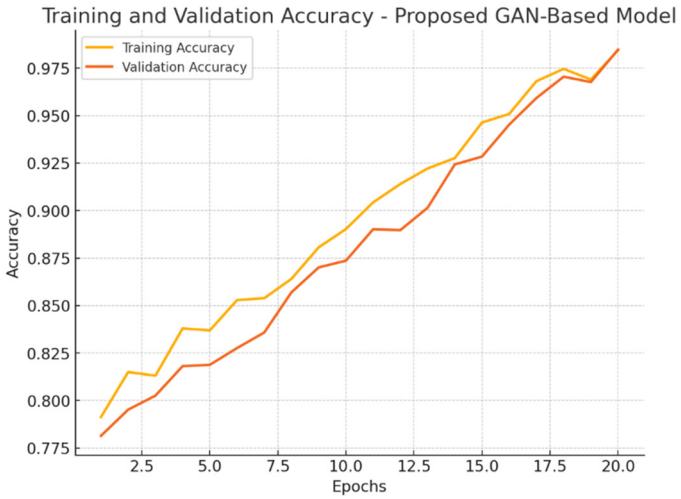

Fig. 6. Training and validation accuracy curves of the proposed GAN-based model.

Figure 7 shows a performance comparison between the proposed GAN-based model for encrypted traffic classification with traditional classifiers, including CNN, RNN, SVM, and Random Forest across five performance metrics: accuracy, precision, recall, F1-score, and AUC. The GAN-based model outperformed all other classifiers for each performance metric, achieving 98.8% for accuracy, precision, and recall. This indicates that the model can accurately classify encrypted traffic as either VPN or non-VPN while minimizing misclassification of undesired traffic.

The model also outperformed the other classifiers in terms of F1-score and achieved an AUC of 0.995, reflecting its strong ability to distinguish between VPN and non-VPN traffic. Overall, these results demonstrate the GAN model's potential to efficiently classify encrypted traffic while preserving user privacy, providing robust protection against malware in SDN-enabled home networks.

Figure 8 compares the false positive and false negative rates of the proposed GAN-based model with traditional classifiers, including CNN, RNN, SVM, and Random Forest. The GAN-based model exhibits the lowest rates, with both false positives and false negatives as low as 0.006, demonstrating high reliability in minimizing misclassifications. This low error rate is crucial in real-world scenarios, where false positives can lead to unnecessary alerts and false negatives can allow malicious traffic to bypass detection.

In contrast, the other models, particularly SVM, show significantly higher rates, with SVM exhibiting the highest. These results indicate that the GAN model is not only more accurate but also more robust in handling misclassifications, ensuring fewer errors in detecting both benign and malicious traffic. Overall, this highlights the GAN's superior performance for real-world, privacy-preserving traffic classification tasks.

Figure 9 compares the training time (s) and inference time (ms) of the proposed GAN-based model with traditional classifiers, including CNN, RNN, SVM, and Random Forest. Although the proposed GAN model requires 320 s for training, which is notably higher than the other models, it compensates with an extremely low inference time of just 2 ms. This result demonstrates the model's efficiency for real-time deployment in network traffic classification, where rapid decision-making is critical.

In contrast, the traditional classifiers exhibit shorter training times but higher inference times, particularly CNN and RNN, which may limit their scalability for real-time deployment in large-scale networks. Despite the slightly longer training duration, the GAN model's rapid inference speed provides a significant advantage, making it a practical and reliable choice for applications requiring fast and accurate classification.

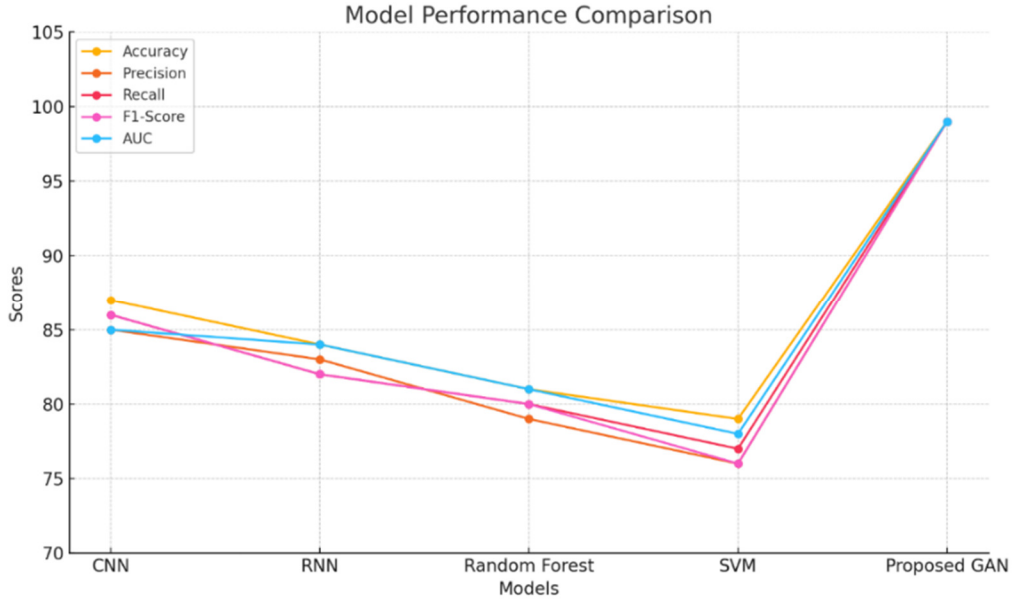

Fig. 7. Model performance comparison.

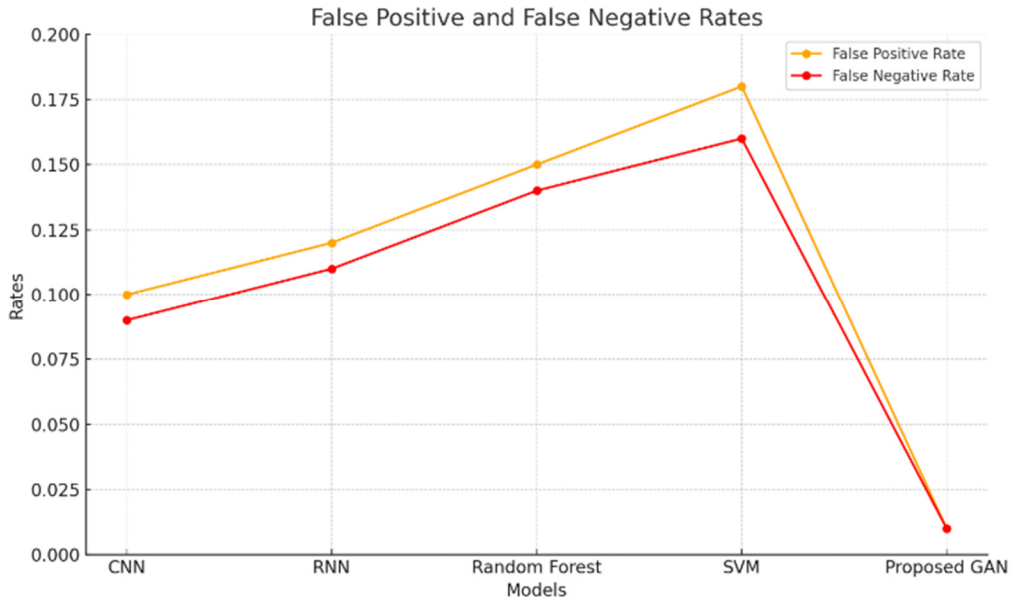

Fig. 8. False positive and false negative rates comparison of different classifiers.

Fig. 9. Training time (s) and inference time (ms) comparison of different classifiers.

IV. CONCLUSIONS

This study presents a robust, privacy-preserving encrypted traffic classification framework based on Generative Adversarial Networks (GANs) for Software-Defined Networking (SDN)-enabled home networks. The proposed model addresses the inherent challenges posed by traffic encryption, adversarial behavior, and dynamic network conditions without compromising data privacy. By leveraging the adversarial training mechanism of GANs, the model effectively learns intrinsic traffic patterns from flow-level metadata, eliminating the need for payload decryption and thereby ensuring compliance with modern data protection standards. Extensive experiments on the ISCX VPN dataset demonstrate that the proposed GAN-based classifier significantly outperforms conventional models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Support Vector Machines (SVMs), and Random Forest. It achieves a superior classification accuracy of 98.8%, precision and recall of 98.8%, and an Area Under the Curve (AUC) of 0.995, alongside minimal false detection rates and a fast inference time of just 2 ms, making it highly suitable for real-time SDN deployments. The model also showcases stable training convergence and strong generalization capabilities, even under limited data conditions and adversarial perturbations.

In addition to its technical effectiveness, the framework is scalable, adaptable to evolving traffic patterns, and deployable across heterogeneous SDN architectures, including Internet of

Things (IoT), smart home, and edge networks. These characteristics make it a compelling solution for modern network security and monitoring applications. Future work will focus on extending the model to support multi-class classification scenarios involving various encrypted traffic types (e.g., VoIP, DNS tunneling, malware), integrating federated GAN architectures for distributed training across privacy-sensitive environments, and exploring explainable GAN frameworks to enhance interpretability and administrator trust. This research lays a strong foundation for the next generation of intelligent, adaptive, and secure traffic classification systems in privacy-driven network ecosystems.

DATA PRIVACY STATEMENT

All data used in this study, including both public and institutional sources, were fully de-identified to ensure privacy. No personally identifiable information was collected or used. Institutional data were accessed under data-sharing agreements that comply with ethical standards for research.

REFERENCES

[1] Y. Zion, C. Hajaj, A. Dvir, G. Ben-Artzi, S. Mahpod, and R. Dubin, "Revolutionizing Our Way to Better Classifiers: Leveraging Synthetic Data with Generative Models for Encrypted Network Traffic Classification." Social Science Research Network, Rochester, NY, Dec. 05, 2023, https://doi.org/10.2139/ssrn.4654236.

[2] P. Wang, Z. Wang, F. Ye, and X. Chen, "ByteSGAN: A Semi-supervised Generative Adversarial Network for Encrypted Traffic Classification of SDN Edge Gateway in Green Communication Network." arXiv, Mar. 09, 2021, https://doi.org/10.48550/arXiv.2103.05250.

[3] Y. Guo, G. Xiong, Z. Li, J. Shi, M. Cui, and G. Gou, "Combating Imbalance in Network Traffic Classification Using GAN Based Oversampling," in *2021 IFIP Networking Conference (IFIP Networking)*, Espoo and Helsinki, Finland, 2021, pp. 1–9, https://doi.org/10.23919/IFIPNetworking52078.2021.9472777.

[4] G. Han, H. Zhang, Z. Zhang, Y. Ma, and T. Yang, "AI-Based Malicious Encrypted Traffic Detection in 5G Data Collection and Secure Sharing," *Electronics*, vol. 14, no. 1, Jan. 2025, Art. no. 51, https://doi.org/10.3390/electronics14010051.

[5] J. Mao *et al.*, "Semisupervised Encrypted Traffic Identification Based on Auxiliary Classification Generative Adversarial Network," *Computer Systems Science and Engineering*, vol. 39, no. 3, pp. 373–390, Aug. 2021, https://doi.org/10.32604/csse.2021.018086.

[6] P. Wang, S. Li, F. Ye, Z. Wang, and M. Zhang, "PacketCGAN: Exploratory Study of Class Imbalance for Encrypted Traffic Classification Using CGAN," in *ICC 2020 - 2020 IEEE International Conference on Communications*, Dublin, Ireland, 2020, pp. 1–7, https://doi.org/10.1109/ICC40277.2020.9148946.

[7] S. M. Rayavarapu, S. P. Tammineni, S. R. Gottapu, and A. Singam, "A Review of Generative Adversarial Networks for Security Applications," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. 14, no. 2, pp. 66–70, Jun. 2024, https://doi.org/10.35784/iapgos.5778.

[8] Y. Duan, L. Wang, D. Liu, B. Deng, and Y. Tian, "Malware Traffic Classification Based on GAN and BP Neural Networks," in *Ubiquitous Security: Second International Conference, UbiSec 2022*, Zhangjiajie, China, 2022, pp. 144–160, https://doi.org/10.1007/978-981-99-0272-9_10.

[9] Z. Tang, J. Wang, B. Yuan, H. Li, J. Zhang, and H. Wang, "Markov-GAN: Markov image enhancement method for malicious encrypted traffic classification," *IET Information Security*, vol. 16, no. 6, pp. 442–458, Jun. 2022, https://doi.org/10.1049/ise2.12071.

[10] H. Sun, C. Peng, Y. Sang, S. Li, Y. Zhang, and Y. Zhu, "Evading Encrypted Traffic Classifiers by Transferable Adversarial Traffic," in *Collaborative Computing: Networking, Applications and Worksharing, 18th EAI International Conference, CollaborateCom 2022*, Hangzhou, China, 2022, pp. 153–173, https://doi.org/10.1007/978-3-031-24386-8_9.

[11] V. K. Jewani *et al.*, "Enhancing Cyber Security Through Generative Adversarial Networks," in *Enhancing Security in Public Spaces Through Generative Adversarial Networks (GANs)*, S. Ponnusamy, J. Antari, P. Bhaladhare, A. Potgantwar, and S. Kalyanaraman, Eds. Hershey, PA, USA: IGI Global Scientific Publishing, 2024, pp. 177–192, https://doi.org/10.4018/979-8-3693-3597-0.ch013.

[12] X. Ma, W. Zhu, Y. Jin, and Y. Gao, "A continual encrypted traffic classification algorithm based on WGAN," in *Third International Seminar on Artificial Intelligence, Networking, and Information Technology*, Shanghai, China, 2022, pp. 460–466, https://doi.org/10.1117/12.2667229.

[13] R. Changala, S. Kayalvili, M. Farooq, L. M. Rao, V. S. Rao, and S. Muthuperumal, "Using Generative Adversarial Networks for Anomaly Detection in Network Traffic: Advancements in AI Cybersecurity," in *2024 International Conference on Data Science and Network Security*, Tiptur, India, 2024, pp. 1–6, https://doi.org/10.1109/ICDSNS62112.2024.10690857.

[14] Y. Ding, G. Zhu, D. Chen, X. Qin, M. Cao, and Z. Qin, "Adversarial Sample Attack and Defense Method for Encrypted Traffic Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18024–18039, Oct. 2022, https://doi.org/10.1109/TITS.2022.3154884.

[15] C. Xu, R. Xia, Y. Xiao, Y. Li, G. Shi, and K.-C. Chen, "Federated Traffic Synthesizing and Classification Using Generative Adversarial Networks," in *ICC 2021 - IEEE International Conference on Communications*, Montreal, Canada, 2021, pp. 1–6, https://doi.org/10.1109/ICC42927.2021.9500866.

[16] X. Wang, W. Wei, X. Yu, D. Zheng, N. Kuma, and L. Liu, "Ensemble Learning-based Traffic Classification with Small-Scale Datasets for Wireless Networks," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops*, Vancouver, Canada, 2024, pp. 1–6, https://doi.org/10.1109/INFOCOMWKSHPS61880.2024.10620836.

[17] M. Mwita, J. Mbelwa, J. Agbinya, and A. E. Sam, "The Effect of Hyperparameter Optimization on the Estimation of Performance Metrics in Network Traffic Prediction using the Gradient Boosting Machine Model," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10714–10720, Jun. 2023, https://doi.org/10.48084/etasr.5548.

[18] V. Ravi and A. S. Poornima, "SecMa: A Novel Multimodal Autoencoder Framework for Encrypted IoT Traffic Analysis and Attack Detection," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23020–23026, Jun. 2025, https://doi.org/10.48084/etasr.10336.

[19] "VPN-nonVPN dataset (ISCXVPN2016)." Canadian Institute for Cybersecurity, UNB. [Online]. Available: https://www.unb.ca/cic/datasets/vpn.html.