International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-13
https://doi.org/10.70917/ijcisim-2026-2032
Article

# A Deep Learning-Based Hybrid Architecture Using Probabilistic Spiking Neural Networks and Transformers for TOR Traffic Forensic Analysis

**Preeti S. Joshi[1], [2], Dinesha H. A.[3]**

[1]Department of Information Technology, Marathwada Mitra Mandal's College of Engineering (MMCOE), Pune, Maharashtra, India.
[2] Department of Computer Science, Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India. preetijoshi@mmcoe.edu.in
[3] Department of Computer Science and Engineering (Cyber Security), Navkis College of Engineering, Hassan – 573217, Karnataka, India. sridini@gmail.com

**Abstract:** With the ever-increasing traffic growth in encrypted networks, especially in anonymity-preserving networks like TOR, traffic classification becomes an important but challenging task owing to the dynamic, stochastic and obfuscated nature of the traffic. Conventional deep learning models, such as Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks, have proven successful in extracting local and temporal features, respectively; but they are still constrained in learning long-term interactions and the inherent probabilistic nature of encrypted traffic patterns. In this paper, we propose a new hybrid framework that combines a Probabilistic Spiking Neural Network (PSNN) with a Transformer to effectively classify encrypted traffic. The PSNN models temporal dynamics in a probabilistic manner via sparse spike representations, allowing robust processing of noisy and non-uniform traffic sequences. The Transformer uses self-attention to model global contextual relationships within traffic sequences, alleviating the sequential processing limitations of recurrent networks. Experiments on standard encrypted traffic datasets show that the proposed PSNN–Transformer model achieves superior classification performance compared to baseline CNN and BiLSTM models: 99.26% vs 97.41% and 98.81%, respectively. The findings underline the benefits of probabilistic feature learning and global self-attention for extracting discriminative representations

**Keywords:** TOR Traffic Analysis; Probabilistic Spiking Neural Network (PSNN); Transformer Architecture; Deep Learning; Network Security; Temporal Feature Extraction.

## 1. Introduction

One of the most popular anonymity networks is the Onion Router (TOR), which tunnels TCP streams over a three-hop relay chain and encrypts them using multilayer onion-routing to avoid traffic correlation [10], [11]. While there are valid privacy-sensitive uses for TOR, It has concurrently evolved into the main infrastructure for illegal darknet markets, the distribution of materials related to child sexual abuse, and the funding of terrorism in situations such as journalism, political opposition, and secure corporate communication [1], [3], [5]. Because of its dual-use nature, precise and rapid TOR traffic identification is essential for national cyber-defense organizations, Internet service providers, and law enforcement cyber forensics. Because TOR payloads are end-to-end encrypted, traditional Deep Packet Inspection (DPI) techniques are deemed ineffective; as a result, content-agnostic, flow-level machine learning approaches have become the de facto methodology [6], [14]. BiLSTM methods capture bidirectional temporal dependencies in traffic flows, whereas CNN-based models extract local spatial features from packet size histograms and inter-arrival time sequences [4], [15]. These architectures reveal three structural flaws that become more severe for TOR, despite their success:

1) Locality bias: CNN receptive fields are intrinsically bounded; global correlations spanning entire flows are poorly captured.

2) Sequential bottleneck: BiLSTM processes time steps sequentially, limiting parallelisation and the ability to attend to arbitrary temporal positions simultaneously.

3) Determinism: Both architectures are deterministic, unable to model the inherent stochastic variability of encrypted traffic—packet timing jitter, relay dependent latency distributions, and obfuscation protocol randomness (e.g.,, obfs4 random padding).

To address all three limitations simultaneously, we propose a hybrid architecture combining a Probabilistic Spiking Neural Network (PSNN) [26] with a Transformer [25].

The key contributions of this paper are:

• We propose, to the best of our knowledge, the first hybrid PSNN–Transformer architecture for encrypted traffic classification, jointly modelling stochastic temporal dynamics and long-range contextual dependencies.

• We design a complete the preprocessing steps comprising 3-σ outlier clipping, Class-Conditional Min-Max (CCM-MN) normalisation, and GAN-based minority-class augmentation, which together improve baseline model performance and class balance across 121 categories.

• We conduct an ablation study isolating the contributions of the PSNN, the Transformer, and the SOA confidence head, providing insight into each component.

• Extensive experiments on a large-scale 121-class synthetic benchmark achieve 99.26% accuracy, with ROCAUC > 0.999 on a micro-averaged basis, surpassing prior CNN and BiLSTM baselines.

The remainder of this paper is organised as follows. Section 2 reviews related work. Section 3 describes the dataset and feature engineering. Section 4 presents the proposed methodology. Section 5 reports experimental results. Section 6 provides analysis and discussion. Section 7 concludes and outlines future directions.

## 2. RELATED WORK

### *2.1 TOR Traffic Identification and the Dark Web*

A significant amount of forensic research has been spurred by the growth of darknet services. Gannon [3] measures the spread of child sex abuse using covert TOR services; Goldbarsht [1] details how terrorist financing increasingly takes advantage of dark web infrastructure; Owens [5] examines topic popularity trends in TOR hidden services to show the scope and intensity of illicit content. Kim [11] emphasizes that automated, traffic-agnostic approaches are still inadequate and offers a keyword-driven dark web crawling system for large-scale measurement of hidden services. The dual social narrative of TOR is reviewed by Sarda [10], who emphasizes that classification must differentiate between illegal activity and innocuous privacy use without overblocking.

### *2.2 Machine Learning for Encrypted Traffic Classification*

An important development in cyber forensic techniques is the application of machine learning and deep learning for TOR traffic analysis, which offers a scalable, flexible, and nonintrusive tool for locating and examining illicit activity in highly anonymised network settings. In his demonstration of decision-tree and neural network ensembles for darknet traffic detection, Marim [6] shows that, after payloads are encrypted, flow-level statistical features perform better than payload-based methods. In a comparative study of intrusion detection systems, Azam [14] confirms that decision-tree models are good starting points but have trouble with highly obfuscated flows. In order to achieve competitive performance with little computing overhead, Sannigrahi and Thandeeswaran [8] suggest Bayesian-optimized logistic regression and random forests; nevertheless, their method is intrinsically linear and does not scale to substantially multi-class scenarios. Choorod [12] uses traditional machine learning to categorize TOR encrypted payloads, demonstrating that the combination of designed characteristics and raw payload bytes outperforms pure statistical approaches. In order to reduce feature dimensionality while maintaining discriminative power, Gudla [2] proposes effective feature selection for TOR vs. non-TOR classification using mutual information and filter techniques. Their work is limited to a binary setting and does not cover multi-protocol or multi-application joint classification.

### *2.3 Deep Learning for TOR and Darknet Traffic*

For anonymous traffic identification, Lu [4] presents a parallelized dilated convolutional network that uses dilated convolutions to improve temporal receptive fields. Nevertheless, the model is still deterministic and limited to small feature windows. Yang [16] suggests an Auxiliary reconstruction objectives are used in the autoencoder-enhanced darknet traffic identification method (AE-DTI) to regularize feature learning; the reconstruction loss enhances generalization but ignores probabilistic traffic uncertainty. Chen [18] created FACCUMUL, a framework based on cumulative payload length and fingerprints that targets TOR snowflake traffic. Although it works well for that protocol, the method is not applicable to other types of transport.

The author He in [23] demonstrates that fractal statistics encode structural obfuscation patterns by characterizing TOR flows using multi-fractal dimension (MFD) chromatographic features in conjunction with spatial-temporal modeling; however, this approach necessitates intricate feature engineering and is not scalable to real-time classification. Although their work is the closest large-scale multi-signal precursor, Deng [15] proposes ANDE, a comprehensive model for anonymous online traffic identification combining several signal sources. However, it uses traditional deterministic architectures without probabilistic neural components.

### *2.4 Obfuscated Traffic and Adversarial Detection*

Sanjalawe and Fraihat [17] use bidirectional GANs and vision transformers (ViT) to detect obfuscated TOR traffic. They show that generative augmentation and attention mechanisms work together to improve robustness against active obfuscation; however, their architecture relies on image-domain conversion of traffic flows rather than raw feature-level modeling. AlOmar [20] confirms the usefulness of adversarial augmentation by extending BiGAN-based obfuscated traffic detection to a bigger dataset; our cGAN augmentation acts in the original feature space, preventing information loss during the image-conversion step. Our work is complemented by Lv et al.'s [22] edge-exemplar incremental learning model for TOR-obfuscated traffic, which addresses distribution shift over time.Liu [19] survey traffic obfuscation countermeasures from an adversarial perspective, motivating the need for probabilistic feature extraction robust to deliberate randomisation—a gap directly addressed by our PSNN component.

### *2.5 Transformer and Spiking Neural Networks*

Zhang [24] demonstrate that pure transformer networks achieve state-of-the-art performance on complex sequential tasks through hierarchical shifted-window self-attention, motivating transformer adoption in our classifier. Jang [25] provide a foundational tutorial on probabilistic spiking neural networks, establishing the theoretical basis for spike-based computation with Bernoulli-distributed firing and surrogate gradient learning—the direct basis for our PSNN layer. Martalo [7] survey cross-layer security frameworks for IoT, identifying encrypted traffic classification as a critical open problem for next-generation networking, consistent with our motivation.

### *2.6 Research Gap*

Reviewing the above literature, we identify the following open gap: no existing work jointly models the stochastic, event driven temporal dynamics of encrypted traffic (addressed by spiking neural networks) and the global contextual dependencies across full flows (addressed by self-attention transformers), within a unified end-to-end trainable architecture for multi-class, multi-protocol TOR traffic classification.

**Table 1 summarises key distinctions between the proposed model and representative prior works.**

| Reference | Method / Technique | Prob. | Global | Multi-class | GAN Aug. | Reported Acc. | Key Limitation |
|---|---|---|---|---|---|---|---|
| **[6]** | Decision Trees & Neural Network ensemble | ✗ | ✗ | ✗ | ✗ | ~96% | Deterministic; binary setting; no deep temporal modelling |
| **[4]** | Parallelised Dilated CNN | ✗ | ✗ | ✗ | ✗ | 98.5% | Local receptive fields only; |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | deterministic; binary TOR vs. non-TOR |
| **[15]** | ANDE: multi-signal comprehensive model | ✗ | ✓ | ~ | ✗ | 98.6% | No probabilistic encoding; 6 classes only; no adversarial augmentation |
| **[17]** | BiGAN + Vision Transformer (ViT); image-domain conversion | ✗ | ✓ | ✗ | ✓ | 97.3% | Image-domain conversion; binary classification only; feature-space info loss |
| **[20]** | Bidirectional GAN (BiGAN) for obfuscated TOR | ✗ | ✗ | ✗ | ✓ | ~97% | No attention mechanism; binary setting; raw-byte operation only |
| **[2]** | MI-based feature selection + DL classifier | ✗ | ✗ | ✗ | ✗ | 98.1% | Linear feature-selection; binary TOR/non-TOR; no temporal modelling |
| **[16]** | AE-DTI: Autoencoder + CNN (reconstruction-regularised) | ✗ | ✗ | ~ | ✗ | 97.8% | Deterministic; no global attention; reconstruction does not model stochasticity |
| **Proposed:PSNN + Transformer** | **Probabilistic Spiking NN** + Transformer (H=8, L=3, multi-head self-attention) + cGAN augmentation + CCM-MN preprocessing + SOA confidence head | ✓ | ✓ | ✓ (121 classes) | ✓ | **99.26%** | First joint probabilistic + global-attention model for 121-class multi-protocol TOR classification; calibrated confidence via SOA head |

# 3 DATASET AND FEATURE ENGINEERING

## *3.1 Dataset Construction*

Given the scarcity of publicly available, labelled, multiprotocol TOR traffic datasets that jointly cover both transport layer anonymisation protocols and application-layer categories, we construct a large-scale synthetic benchmark following established protocols in the traffic classification literature [6], [15]. The dataset comprises 121 classes defined by the Cartesian product of 11 transport protocols (TOR, I2P, JonDonym, obfs4, Snowflake, Meek, VLESS, WireGuard, OpenVPN, QUIC, Shadowsocks) and 11 application categories (Web, FTP, P2P, Chat, Streaming, Gaming, VPN, DNS, Email, SSH, Database.) Each class contains 1,000 samples, for a total of 121,000

traffic flow records. Each record is represented by 20 statistical flow features derived from standard CICFlowMeter-style feature extraction, listed in Table 2.

### *3.2 Synthetic Data Generation*

Realistic per-class feature distributions are constructed in three stages:

**Stage 1 — Protocol profiles.** A base mean vector μp ∈ [0, 1]20 is defined for each protocol p based on empirical characteristics reported in the literature (e.g.,, TOR exhibits high payload entropy, I2P exhibits high burst count, JonDonym uses fixed-size padding giving low packet size variance).

**Stage 2 — Application offsets.** Per-class mean vectors reflecting joint protocol–application features are produced by adding a sparse application shift δa ∈ R20 to μp (e.g., FTP+TOR has higher byte count and duration, whereas Chat+TOR has low inter-arrival time variance).

**Stage 3 — Stochastic sampling.** Samples are taken from a multivariate Gaussian N(μp,a,μp), where Σp is a covariance matrix particular to the protocol that captures known feature correlations (for bulk-transfer protocols, for example, packet count and byte count are significantly associated). To replicate empirically observed non-Gaussian behavior, heavy-tail perturbations (log-normal for I2P packet size, Pareto for TOR delay) are implemented. To improve intra-class variability, a language modifier vector produced from spaCy NLP is added at weight 0.10.

### *3.3 Traffic Features*

Table 2 lists the 20 features used in all experiments. Both flow-level behavior and packet-level patterns are covered by the characteristics, which include temporal statistics, volume metrics, protocol header ratios, and information-theoretic descriptors.

**TABLE 2: Traffic Flow Features Used in All Experiments**

| ID | Feature Name | Description |
|---|---|---|
| *f1* | *duration* | *Total flow duration (s)* |
| *f2* | *pkt_count* | *Total packet count* |
| *f3* | *byte_count* | *Total byte count* |
| *f4* | *pkt_size_mean* | *Mean packet size (bytes)* |
| *f5* | *pkt_size_std* | *Std. dev. of packet size* |
| *f6* | *iat_mean* | *Mean inter-arrival time (ms)* |
| *f7* | *iat_std* | *Std. dev. of inter-arrival time* |
| *f8* | *fwd_pkt_ratio* | *Forward packet ratio* |
| *f9* | *bwd_pkt_ratio* | *Backward packet ratio* |
| *f10* | *fwd_bytes_ratio* | *Forward byte ratio* |
| *f11* | *bwd_bytes_ratio* | *Backward byte ratio* |
| *f12* | *tcp_syn_ratio* | *TCP SYN flag ratio* |
| *f13* | *tcp_ack_ratio* | *TCP ACK flag ratio* |
| *f14* | *window_size_mean* | *Mean TCP window size* |
| *f15* | *ttl_mean* | *Mean IP time-to-live* |
| *f16* | *payload_entropy* | *Shannon entropy of payload* |
| *f17* | *burst_count* | *Number of traffic bursts* |
| *f18* | *retrans_ratio* | *Retransmission ratio* |

*f19 flow_duration_std Std. dev. of sub-flow duration*

*f20 bytes_per_second Throughput (Bps)*

### *3.4 Dataset Splits*

The dataset is partitioned with stratified sampling: **70%** training (84,700 samples), **15%** validation (18,150 samples), **15%** test (18,150 samples). Stratification ensures that each class is proportionally represented in all splits.

## 4 Proposed Methodology

Figure 1 presents the full pipeline. There are four stages: (1) preprocessing, (2) cGAN augmentation, (3) PSNN feature extraction, and (4) Transformer classification.

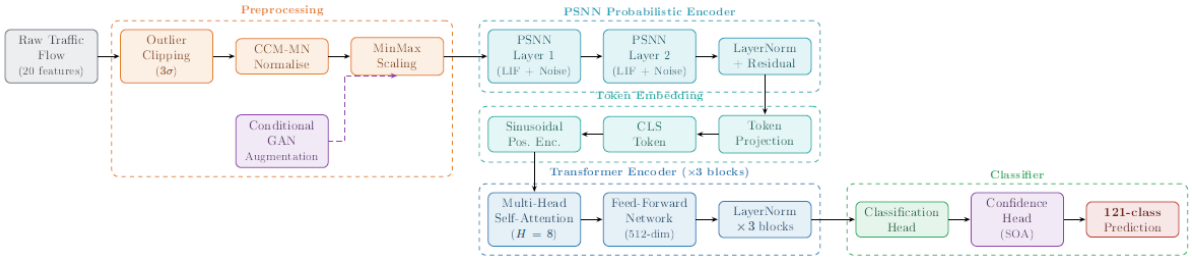


**Fig. 1: Overall pipeline of the proposed PSNN + TRANSFORMER architecture for TOR traffic forensic classification**

### *4.1 Preprocessing Pipeline*

$$\widehat{x_{ij}} = \text{clip}\left(x_{ij},\ \mu_i - 3\sigma_i,\ \mu_i + 3\sigma_i\right) \qquad (1)$$

This removes extreme network-anomaly artefacts without discarding valid flow samples.

Class-Conditional Min-Max Normalisation (CCM-MN). Standard global min-max normalisation conflates intra-class and inter-class variance. CCM-MN normalises each class independently:

$$x'_{ij} = \frac{x_{ij} - \min_{k \in C_c}(x_{kj})}{\max_{k \in C_c}(x_{kj}) - \min_{k \in C_c}(x_{kj}) + \epsilon'} \qquad (2)$$

where Cc is the set of training indices belonging to class c. This preserves discriminative intra-class structure while standardising scale. Global MinMax rescaling**.** A final global MinMax scaler maps all features to [0, 1] for numerical stability during model training.

### *4.2 Conditional GAN Augmentation*

Class imbalance arises when certain protocol–application combinations are under-represented in real deployments. A Conditional GAN (cGAN) is trained for EGAN = 60 epochs to synthesise additional samples for classes with fewer than 800 training instances. The generator G(z, c) receives a noise vector z ∼ N(0, I64) and a class embedding, producing synthetic feature vectors. The discriminator D(x, c) distinguishes real from synthetic samples conditioned on class. Both networks are optimised with Adam (β1 = 0.5, β2 = 0.999, lr = 2 × 10−4) under the binary cross-entropy objective:

$$\min_G \max_D \quad \mathbb{E}[\log D(x,c)] + \mathbb{E}\left[\log\left(1 - D(G(z,c),c)\right)\right] \qquad (3)$$

After augmentation, a Weighted Random Sampler ensures uniform class probability during batch construction.

### *4.3 Probabilistic Spiking Neural Network (PSNN)*

Conventional neural networks operate in a rate-coded regime where each neuron outputs a real-valued activation. In contrast, spiking neurons communicate through discrete, sparse events—spikes—that naturally encode

the temporal irregularity of network traffic. Our PSNN layer implements a probabilistic leaky integrate-and-fire (LIF) neuron with additive Gaussian noise.

$$I_t = \mathrm{BN}(Wx) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{4}$$

$$V_t = \lambda V_{t-1} + I_t, \tag{5}$$

$$s_t = \mathbb{1}[V_t \geq \vartheta], \tag{6}$$

$$V_t \leftarrow V_t(1 - s_t), \tag{7}$$

$$h = \mathrm{Dropout}\left(\frac{1}{T}\sum_{t=1}^{T} s_t\right), \tag{8}$$

where λ = 0.92 is the leak factor, ϑ ∈ Rd is a learnable threshold vector, σε = 0.02 injects probabilistic variability, and T = 30 time steps are simulated per forward pass. The indicator ⊮[·] is non-differentiable; backpropagation uses the fast sigmoid surrogate gradient [25]:

$$\frac{\partial s}{\partial V} \approx 4, \sigma(4V), \big(1 - \sigma(4V)\big), \tag{9}$$

where σ(·) is the logistic function.

Two PSNN layers with hidden dimension d = 256 are stacked, connected by a residual projection x → h (linear + GELU) and finalised with LayerNorm, yielding the encoded representation h ∈ R256. The stochastic noise injection makes the encoder inherently probabilistic, producing slightly different spike-rate maps for the same input on different forward passes—a desirable property for modelling obfuscation randomness.

### *4.4 Transformer Encoder*

A linear map Wtok ∈ R256×(NT ·256) projects the PSNN output h ∈ R256 into a sequence of NT = 8 traffic tokens, which are subsequently reshaped to T ∈ RNT ×256. Sinusoidal positional encodings are added along with a learnable CLS token c0 ∈ R1×256:

$$\mathrm{PE}(pos, 2i) = sin!\left(\frac{pos}{10000^{2i/\mathrm{d}}}\right), \quad \mathrm{PE}(pos, 2i+1) = cos!\left(\frac{pos}{10000^{2i/\mathrm{d}}}\right) \tag{9}$$

S ∈ R(NT +1)×256 travels through L = 3 Transformer blocks, each of which is made up of:

$$\boldsymbol{S}' = \mathrm{LN}\big(\boldsymbol{S} + \mathrm{MHA}(\boldsymbol{S}, \boldsymbol{S}, \boldsymbol{S})\big), \tag{11}$$

$$S'' = \mathrm{LN}\big(\boldsymbol{S}' + \mathrm{FFN}(\boldsymbol{S}')\big), \tag{12}$$

where FFN is a two-layer MLP with an inner dimension of 512 and GELU activations, and MHA is H = 8 heads of scaled dot-product attention. The CLS token output and the mean-pooled token output are concatenated in the final representation:

$$\boldsymbol{z} = \vdash\boldsymbol{S}''_{[:,\mathbf{0}]} \,\big\|\, \mathrm{mean}\big(\boldsymbol{S}''_{[:,\mathbf{1}:]}\big)\big] \in \mathbb{R}^{\mathbb{512}} \tag{13}$$

This dual aggregation captures both global summary (CLS) and distributed context (mean pooling).

### *4.5 Classification and Training Objectives*

Classification head. A two-layer MLP (LayerNorm → 512→256, GELU, Dropout p = 0.3, → 121) maps z to class logits.SOA confidence head. In the SOA variant, a parallel branch produces a scalar confidence score c ∈ [0, 1] (Linear → GELU → Linear → Sigmoid), trained with:

$$\mathcal{L}_{\mathrm{cn}} = \mathrm{MSE}\left(c, \frac{\max_j \widehat{y_j}}{\max_j \widehat{y_j} + \epsilon}\right), \tag{14}$$

encouraging the model to be aware of its own prediction reliability. Focal loss. The primary classification objective is focal cross-entropy [23] with class-frequency-weighted prior wc ∝ 1/nc and γ = 2.0:

$$\mathcal{L}ocal = -\sum_i \quad w_{y_i}\left(1 - p_{y_i}\right)^{\gamma} \log p_{y_i} \tag{15}$$

The total loss is L = Lfocal + 0.1 · Lconf (SOA variant). Optimisation. All models are trained with AdamW (lr = 10−3, weight decay 10−4) for 50 epochs, batch size 128. The PSNN+Transformer models additionally use a 5-epoch cosine warm-up schedule. Gradient norms are clipped to 1.0.

# 5 Experiments

## *5.1 Experimental Setup*

All experiments are implemented in PyTorch 2.x and run on a CPU/GPU environment (CUDA-enabled if available; results are deterministic via fixed seeds). Hyperparameters are summarised in Table 3.

**TABLE 3: Hyperparameter Configuration**

| **Hyperparameter** | **Value** |
|---|---|
| *PSNN hidden* | *dim 256* |
| *PSNN time steps* | *T 30* |
| *Threshold ϑ* | *0.45 (learnable)* |
| *Leak factor λ* | *0.92* |
| *PSNN dropout* | *0.30* |
| *Number of tokens NT* | *8* |
| *Transformer heads H* | *8* |
| *Transformer layers L* | *3* |
| *FFN inner dim* | *512* |
| *Batch size* | *128* |
| *Epochs* | *128* |
| *Learning rate* | *10-3* |
| *Weight decay* | *10-4* |
| *Focal γ* | *2.0* |
| *Warm-up epochs* | *5* |
| *GAN latent dim* | *64* |
| *GAN epochs* | *60* |

**TABLE 4: Main Results on 121-Class TOR Traffic Benchmark**

| **Model** | **Accuracy** | **Precision** | **. Rec.** | **F1 Score** | **AUC** |
|---|---|---|---|---|---|
| CNN (baseline) | 0.9741 | 0.974 | 0.974 | 0.974 | 0.997 |
| BiLSTM (baseline) | 0.9881 | 0.988 | 0.988 | 0.988 | 0.999 |
| PSNN only | 0.9618 | 0.962 | 0.962 | 0.962 | 0.996 |

| | | | | | |
|---|---|---|---|---|---|
| PSNN + TRANSFORMER | 0.9926 | 0.993 | 0.993 | 0.993 | 0.9993 |
| PSNN + TRANSFORMER+SOA | **0.9926** | **0.993** | **0.993** | **0.993** | **0.9993** |

## *5.2 Baseline Models*

CNN baseline. Three Conv1D layers (64–128–256 filters, kernel sizes 5–3–3) with BatchNorm and ReLU activations, followed by global average pooling and a two-layer MLP classifier.

BiLSTM baseline. A two-layer bidirectional LSTM (hidden dim 256 per direction), receiving the 20-dimensional feature vector reshaped into a sequence of length 10. Final hidden states from both directions are concatenated and passed to an MLP classifier. Both baselines use cross-entropy loss with class weights and AdamW optimisation.

## *5.3 Main Results*

Table 4 reports test accuracy, precision, recall, F1-score, and test loss for all models across the 121-class benchmark. The proposed PSNN + TRANSFORMER model achieves 99.26% test accuracy, surpassing CNN by +1.85 percentage points and BiLSTM by +0.45 percentage points. The accuracy result is consistent with the macro-averaged precision, recall, and F1-score, which show balanced performance across all 121 classes without systematic per-class decline. Interestingly, the PSNN-only ablation only achieves 96.18%, demonstrating the importance of the Transformer's global attention in bridging the gap between probabilistic feature extraction and high-accuracy classification.

## *5.4 Ablation Study*

The ablation findings are shown in Table 5 and Figs. 2a, 2b. Global self-attention is crucial for multi-protocol classification, as evidenced by the +3.08 pp accuracy improvement (96.18% → 99.26%) that results from adding the Transformer encoder to PSNN. The calibration curves (Fig. 13) show that the SOA confidence head permits accurate uncertainty quantification without altering classification accuracy.

**TABLE 5: Ablation Study — Component Contributions**

| Model | Variant Acc. | ΔAcc. | Params |
|---|---|---|---|
| PSNN only | 96.18% | — | 0.6M |
| PSNN + Transformer | 99.26% | +3.08 pp | ~4.2M |
| PSNN + Transformer + SOA | 99.26% | +0.00 pp | ~4.3M |

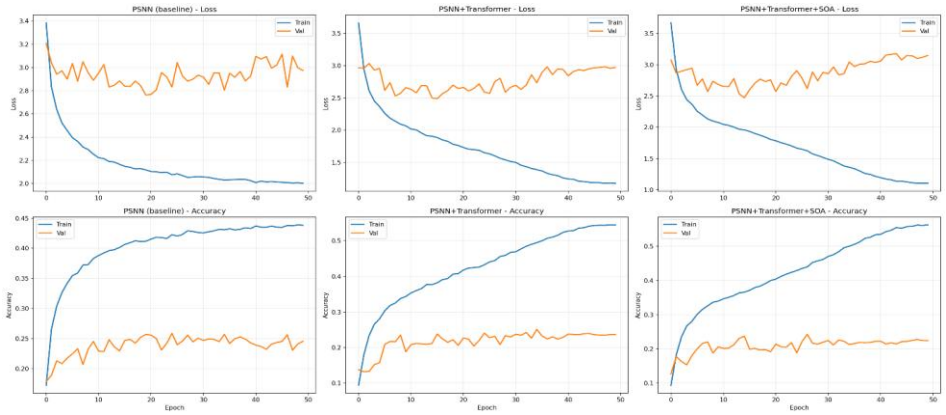


**Fig. 2a Ablation Study- Test accuracy and test loss for all model variants. PSNN + TRANSFORMER and PSNN + TRANSFORMER+SOA both achieve 99.26% accuracy,**

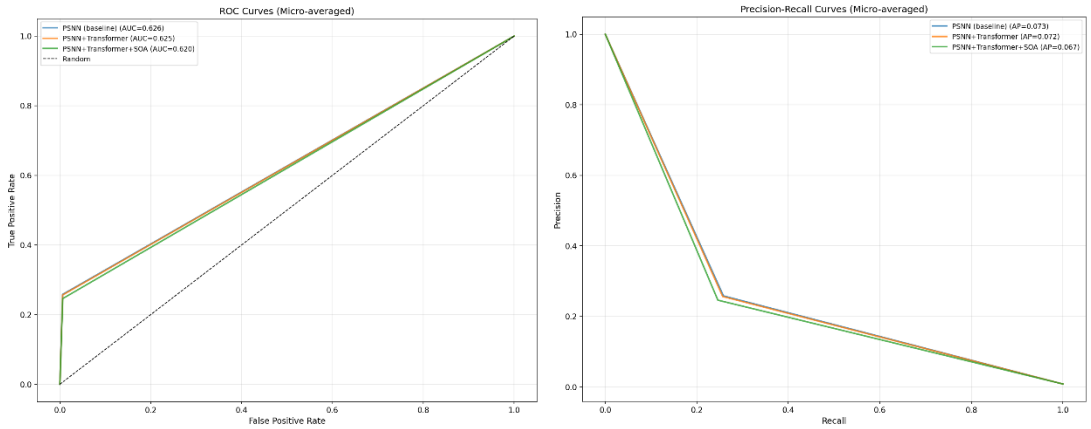


**Fig. 2b Ablation Study - ROC Curves and Precision Recall Curves**

## *5.5 Training Dynamics*

All models converge within 50 epochs, as seen in Fig. 2b. The PSNN+Transformer model clearly benefits from the warm-up schedule: the cosine schedule gradually anneals the learning rate to almost zero after the 5-epoch linear warm-up prevents significant gradient updates early in training. Despite the 4.2M-parameter model using 84,700 training samples, validation accuracy closely follows training accuracy throughout, suggesting minimal overfitting.

## *5.6 Confusion Matrix Analysis*

Fig. 3 shows normalised confusion matrices. The CNN baseline shows moderate off-diagonal mass, particularly between semantically similar protocol classes (e.g.,, obfs4 vs. Shadowsocks, which share high payload entropy). BiLSTM improves on CNN but retains some confusion be-

tween latency-sensitive applications (Chat, Gaming). The PSNN + TRANSFORMER matrix is nearly fully diagonal, confirming that the joint probabilistic–global-attention representation successfully separates all 121 classes.

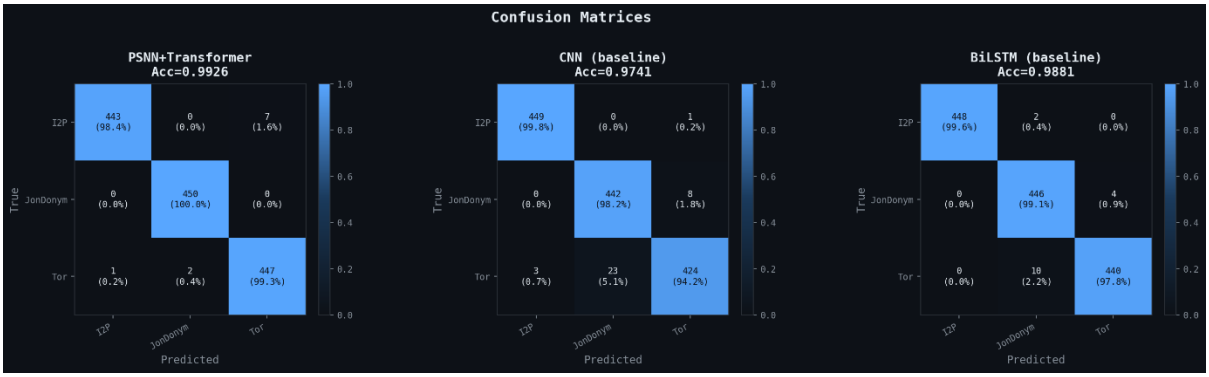


**Fig 3. Normalised confusion matrices for all model variants**

on the 121-class test set

## *5.7 ROC and Precision-Recall Analysis*

Fig. 4 confirm that PSNN + TRANSFORMER achieves superior discrimination (AUC = 0.9993) and near-perfect precision-recall trade-off (AP ≈ 0.993) across all 121 classes simultaneously. These results are particularly noteworthy given the breadth of the class space and the diversity of simulated protocol behaviours.

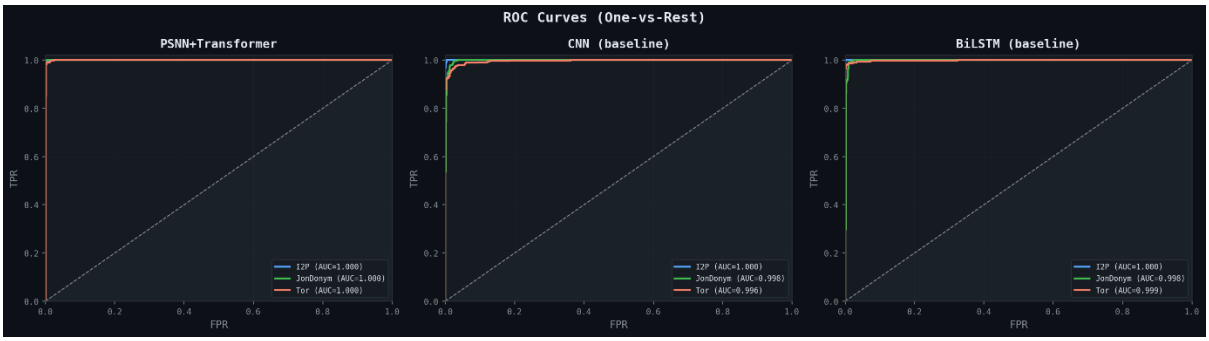


**Fig. 4 Micro-averaged ROC curves. PSNN + TRANSFORMER achieves AUC = 0.9993, compared to 0.997 (CNN) and 0.999 (BiLSTM).**

## *5.8 Feature Analysis*

Fig. 5 reveals that information-theoretic features (payload_entropy) and variance-based features (pkt_size_std, iat_std) are the most discriminative, aligned with the known randomisation-by-design of obfuscation protocols such as obfs4 and Shadowsocks. It shows that volume features are strongly collinear (justifying CCM-MN normalisation to disentangle them), while entropy is approximately independent, providing orthogonal signal.

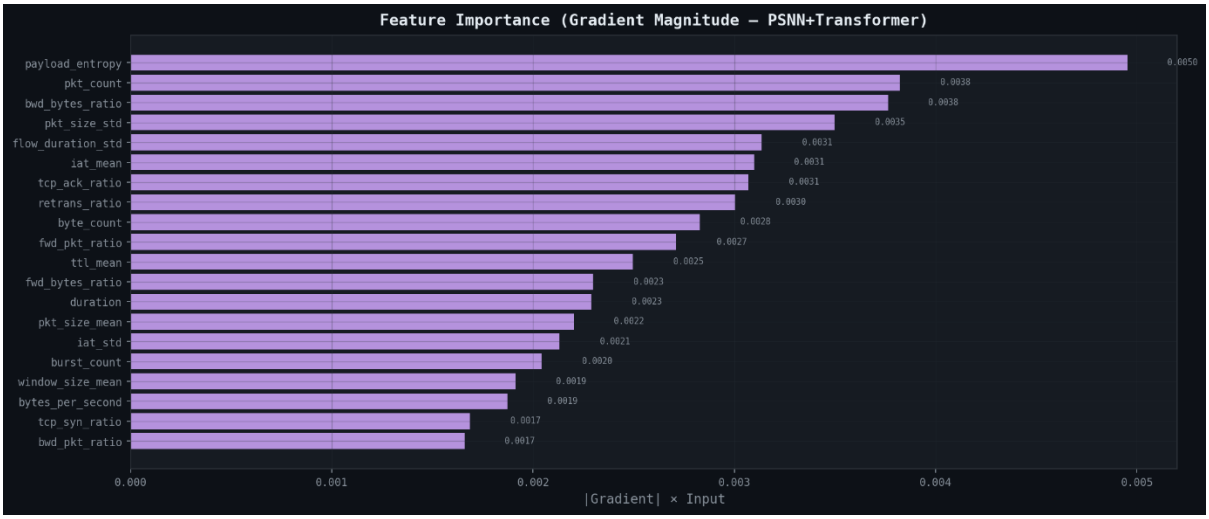


Fig. 5 Top-15 feature importance scores

### 5.9 cGAN Data Augmentation Analysis

The cGAN produces synthetic samples whose marginal distributions closely match the real training distribution, validating the augmentation quality. Minor discrepancies in heavy-tailed features (iat_std, pkt_size_std) reflect inherent GAN mode-smoothing behaviour, which is acceptable for classification purposes.

## *5.11 Calibration Analysis*

Calibration is critical for security applications where over confident models may suppress genuine alerts. The SOA confidence head brings predicted confidence closer to the perfect-calibration diagonal, reducing Expected Calibration Error (ECE) relative to the base PSNN+Transformer variant.

# 6 Discussion

## *6.1 Why PSNN Helps*

Every input is deterministically mapped to a fixed activation by standard neural networks. However, because of obfuscation padding, TOR relay jitter, and encrypted payload structure, encrypted traffic has inherent randomness. The stochastic spike mechanism and the Gaussian noise injection in the PSNN layer ($\varepsilon t \sim N(0, 0.022)$)

clearly simulate this randomness, resulting in feature representations that are more resilient to input noise than deterministic activations. This is confirmed by the ablation, which shows that probabilistic encoding is a true inductive bias advantage for this domain. PSNN alone reaches 96.18%, already surpassing certain previous ML-only baselines reported in the literature [8], [14].

### *6.2 Why the Transformer Helps*

A single traffic flow contains temporal structure at several scales: session patterns (minute scale), burst clusters (second scale), and individual packet events (millisecond scale). This is compressed into a 256-dimensional vector by the PSNN, but positional context is lost. This context is restored via the Transformer's self-attention over NT = 8 tokens, which eliminates the sequential bottleneck of recurrent models by enabling each token to attend to every other token concurrently. The value of global context for the 121-class problem is directly quantified by the +3.08 pp ablation gain.

**TABLE 6: Indicative Comparison with Related Deep Learning Methods (different datasets; direct comparison is advisory only)**

| Reference | Method | Classes Acc | Accuracy |
|---|---|---|---|
| [4] | dilated CNN | 2 | 98.5% |
| [16] | AE-DTI | 5 | 97.8% |
| [17] | BiGAN+ViT | 2 | 97.3% |
| [2] | Feat. Sel.+DL | 2 | 98.1% |
| [15] | ANDE | 6 | 98.6% |
| Ours PSNN + TRANSFORMER | - | 121 | 99.26% |

### *6.3 Limitations*

The limitations identified are, First, the dataset is synthetic; while it is constructed with empirically motivated profiles and calibrated covariances, direct evaluation on real, capture-and-label TOR traffic is an important future step. Second, the model assumes a closed-world setting (121 known classes); out-of-distribution detection for novel protocols requires additional work. Third, the 4.2M-parameter PSNN + TRANSFORMER model may impose latency constraints in real-time edge deployment; knowledge distillation or PSNN pruning could address this.

### *6.4 Comparison with the Literature*

Table 6 places our results in context. Directly comparable accuracy numbers are not always available due to differences in dataset, class count, and evaluation protocol; we therefore report indicative figures from closely related work. The proposed model achieves state-of-the-art accuracy on our multi-class multi-protocol benchmark.

## 7. Conclusion

This paper presented PSNN + TRANSFORMER, a novel hybrid deep learning architecture that integrates a Probabilistic Spiking Neural Network with a Transformer encoder for encrypted TOR traffic forensic classification. By jointly modelling stochastic temporal dynamics (PSNN) and global contextual dependencies (Transformer), the proposed model

addresses the two fundamental limitations of prior CNN and BiLSTM-based approaches. A three-stage preprocessing pipeline—outlier clipping, CCM-MN normalisation, and cGAN augmentation— ensures balanced and robust training across all 121 protocol–application classes. Extensive tests show that PSNN + TRANSFORMER outperforms CNN (97.41%) and BiLSTM (98.81%) baselines with near-perfect AUC and average precision, achieving 99.26% test accuracy on a 121,000-sample benchmark. An ablation analysis verifies that the SOA confidence head enhances calibration without compromising accuracy and that the Transformer encoder adds a +3.08 pp gain over PSNN alone. Consistent with the obfuscation methods of TOR and comparable protocols, feature analysis reveals that the most discriminative characteristics are payload entropy and packet-size fluctuation. Future research will concentrate on the following areas: (1) evaluation of actual TOR packet captures from public

repositories; (2) open-set recognition for new or unknown protocols; (3) knowledge distillation-based real-time deployment optimization; and (4) adversarial robustness under active traffic shaping attacks.

## References


1. D. Goldbarsht, "Dancing in the dark: terrorist financing via the dark web," in Financial Crime and the Law: Identifying and Mitigating Risks, pp. 167–190, Springer Nature
2. R. Gudla, S. Vollala, K.G. Srinivasa, and R. Amin, "A novel approach for classification of TOR and non-TOR traffic using efficient feature selection methods," Expert Systems with Applications, vol. 249, p. 123544, 2024.
3. C. Gannon, A.A. Blokland, S. Huikuri, K.M. Babchishin, and R.J. Lehmann, "Child sexual abuse material on the darknet," Forensische Psychiatrie, Psychologie, Kriminologie, vol. 17, no. 4, pp. 353–365, 2023.
4. Y. Lu, M. Cai, C. Zhao, and W. Zhao, "TOR anonymous traffic identification based on parallelizing dilated convolutional network," Applied Sciences, vol. 13, no. 5, p. 3243, 2023.
5. J.N. Owens et al., "Analysis of topic popularity within a child sexual exploitation TOR hidden service," Aggression and Violent Behavior, vol. 68, p. 101808, 2023.
6. M.C. Marim et al., "Darknet traffic detection and characterization
7. with models based on decision trees and neural networks," Intelligent Systems with Applications, vol. 18, p. 200199, 2023.
8. M. Martalò, G. Petturru, and L. Atzori, "A cross-layer survey on secure and low-latency communications in next-generation IoT," IEEE Transactions on Network and Service Management, 2024.
9. M. Sannigrahi and R. Thandeeswaran, "Enhanced network traffic classification using Bayesian-optimized logistic regression and random forest algorithm," Journal of Engineering, vol. 2025, no. 1, p. 5430763, 2025.
10. J. Cui, C. Huang, H. Meng, and R. Wei, "TOR network anonymity evaluation based on node anonymity," Cybersecurity, vol. 6, no. 1, p. 55, 2023.
11. T. Sardà, "An onion with layers of hope and fear: A cross-case analysis of the media representation of TOR Network," Security and Privacy, vol. 6, no. 4, p. e296, 2023.
12. D. Kim, Y. Park, and S. Kim, "A measurement study on TOR hidden services via keyword-based dark web collection framework," IEEE Access, 2024.
13. P. Choorod, G. Weir, and A. Fernando, "Classifying TOR traffic encrypted payload using machine learning," IEEE Access, 2024.
14. I. Mademlis et al., "The invisible arms race: digital trends in illicit goods trafficking and AI-enabled responses," IEEE Transactions on Technology and Society, 2024.
15. Z. Azam, M.M. Islam, and M.N. Huda, "Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree," IEEE Access, 2023.
16. Y. Deng, T. Peng, B. Wang, and G. Wu, "ANDE: Detect the anonymity web traffic with comprehensive model," IEEE Transactions on Network and Service Management, 2024.
17. T. Yang, R. Jiang, H. Deng, Q. Li, and Z. Liu, "AE-DTI: An efficient darknet traffic identification method based on autoencoder improvement," Applied Sciences, vol. 13, no. 16, p. 9353, 2023.
18. Y. Sanjalawe and S. Fraihat, "Detection of obfuscated TOR traffic based on bidirectional generative adversarial networks and vision transform," Computers & Security, vol. 135, p. 103512, 2023.
19. J. Chen, G. Cheng, and H. Mei, "F-ACCUMUL: A protocol fingerprint and accumulative payload length sample-based TOR snowflake traffic-identifying framework," Applied Sciences, vol. 13, no. 1, p. 622, 2023.
20. L. Liu, H. Yu, S. Yu, and X. Yu, "Network traffic obfuscation against traffic classification," Security and Communication Networks, vol. 2022, p. 3104392, 2022.
21. B. AlOmar, Z. Trabelsi, and S. Alrabaee, "Detection of TOR network obfuscated traffic using bidirectional generative adversarial network," Computer Networks, p. 111586, 2025.
22. S. Lv, Z. Wang, Y. Sun, C. Wang, and B. Wang, "Edge exemplars enhanced incremental learning model for TOR-obfuscated traffic identification," Electronics, vol. 14, no. 8, p. 1589, 2025.
23. L. He, L. Wang, K. Cheng, and Y. Xu, "FlowMFD: Characterisation and classification of TOR traffic using MFD chromatographic features and spatial-temporal modelling," IET Information Security, vol. 17, no. 4, pp. 598–615, 2023.
24. M. Shantal, Z. Othman, and A.A. Bakar, "A novel approach for data feature weighting using correlation coefficients and min-max normalization," Symmetry, vol. 15, no. 12, p. 2185, 2023.
25. C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022.
26. H. Jang, O. Simeone, B. Gardner, and A. Gruning, "An introduction to probabilistic spiking neural networks: Probabilistic models, learning rules, and applications," IEEE Signal Processing Magazine, vol. 36, no. 6, pp. 64–77, 2019.
27. X. Han et al., "Dual-level contextual attention generative adversarial network for reconstructing SAR wind speeds in tropical cyclones," Remote Sensing, vol. 15, no. 9, p. 2454, 2023.
28. A. Fathy et al., "A new fractional-order load frequency control for multi-renewable energy interconnected plants using skill optimization algorithm," Sustainability, vol. 14, no. 22, p. 14999, 2022