Regular Article

HyPoNet: Fine-Grained Sleep Posture Recognition from a Single Abdominal Accelerometer

Hoang-Dieu Vu[1,2], Duc-Nghia Tran[3], Quang-Huy Pham[4], Duc-Tan Tran[1]

[1] Faculty of Electrical and Electronic Engineering, Phenikaa School of Engineering, Phenikaa University, Yen Nghia, Hanoi, Vietnam
[2] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[3] Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[4] Faculty of Information Technology, Electric Power University, Hanoi, Vietnam

Correspondence: Duc-Nghia Tran, nghiatd@ioit.ac.vn
Communication: received 05 July 2025, revised 27 August 2025, accepted 16 September 2025
Online publication: 21 September 2025, Digital Object Identifier: 10.21553/rev-jec.411

Abstract– **Fine-grained sleep posture recognition is essential for the non-invasive management of conditions such as gastroesophageal reflux disease (GERD) and obstructive sleep apnea (OSA). Traditional systems typically recognize only a limited set of coarse sleep positions, thereby restricting their clinical effectiveness in real-world scenarios. This study presents HyPoNet, a lightweight deep learning model designed to classify twelve distinct sleep postures using data from a single wearable sensor system. The proposed hardware platform consists of a tri-axial accelerometer (ADXL345) positioned on the abdomen, interfaced with a low-power microcontroller unit (ESP32) for real-time signal acquisition and wireless data transmission. Acceleration signals along the x, y, and z axes were collected from ten healthy volunteers performing twelve predefined sleep positions under controlled conditions. The collected data were segmented using a sliding window method, and a subject-independent evaluation strategy was applied: data from eight volunteers were used for training and validation (in an 80:20 split), while data from the remaining two volunteers were reserved for testing.**
HyPoNet uses a hybrid neural network architecture that combines bidirectional long short-term memory (BiLSTM) units to capture temporal relationships in acceleration data with one-dimensional convolutional layers for spatial pattern extraction. The system outperformed benchmark models like CNN, GRU, and Transformer variants, achieving an average accuracy of 97.29% and an F1-score of 90.72%. HyPoNet is a promising method for embedded applications in sleep posture monitoring in both home and clinical settings because of its low computational requirements and good classification performance.

Keywords– **Sleep posture recognition, wearable sensors, accelerometer, deep learning,**

1 Introduction

In recent years, wearable technologies have become increasingly vital to the advancement of smart healthcare [1, 2]. These compact, body-worn devices enable real-time, continuous monitoring of physiological signals and human behaviors, offering a practical and scalable solution for both clinical applications and home-based health management. Depending on their design, wearable systems can incorporate various types of sensors—such as photoplethysmography (PPG), electrocardiography (ECG) [3], gyroscopes, and accelerometers—to capture meaningful data related to diverse health conditions [4].

Among these, the **accelerometer** has emerged as one of the most widely used sensors due to its low cost, small size, energy efficiency, and robustness [5–9]. Initially applied to physical activity tracking and fall detection, accelerometers are now increasingly adopted in sleep research, especially for monitoring and classifying **sleep posture**. Sleep posture not only influences sleep quality but also plays a critical role in managing various health conditions.

A promising application area is **positional therapy**—a non-pharmacological treatment approach that encourages or enforces specific sleep orientations to alleviate symptoms [10]. For instance, avoiding the supine position significantly reduces the apnea-hypopnea index (AHI) in OSA patients [11, 12], while sleeping in a left lateral posture or elevating the upper body may reduce GERD symptoms [13]. The ability to monitor and guide sleep posture in real-time is particularly relevant for young adults, who face growing health challenges related to stress, poor sleep hygiene, and sedentary lifestyles [14, 15].

Accelerometer-based wearable devices provide an ideal platform for implementing positional therapy in everyday environments. Their continuous tracking capabilities enable not only long-term retrospective analysis but also real-time interventions—such as vibration alerts to correct harmful postures during sleep. Additionally, the longitudinal data captured by these systems can support clinical assessments and inform personalized treatment strategies [6, 7].

However, many current wearable systems are limited to recognizing only **four coarse-grained postures**: supine, prone, left lateral, and right lateral. This basic classification fails to capture the nuance of real-world sleep behavior, where individuals often adopt **intermediate or transitional positions** (e.g., semi-supine, partial lateral turns). Studies have shown that such subtle variations in posture can influence the severity of symptoms in sleep-related conditions [13, 15], reducing the clinical utility of oversimplified classification schemes.

To address this limitation, we present a novel wearable system that uses a single device with one accelerometer for recognizing **12 distinct sleep positions**. These positions are designed as *fine-grained variations* of the four canonical postures, capturing subtle deviations such as slight body rotations or inclination angles. This richer representation allows for more precise and clinically meaningful analysis of natural sleep behavior without oversimplifying transitional or borderline postures.

At the core of our system is **HyPoNet (Hybrid Position Network)**, a lightweight deep learning architecture designed for efficient and accurate posture classification. HyPoNet combines 1D - convolutional neural networks (1D CNNs) [16] for spatial feature extraction with bidirectional long short-term memory (BiLSTM) layers [17] to capture both forward and backward temporal dependencies in the accelerometer data. This hybrid design enables the model to distinguish subtle differences between similar postures that may exhibit overlapping motion patterns.

Compared to existing models such as **AnpoNet** [18], which was among the first to explore 12-class sleep posture classification using a CNN-LSTM approach, HyPoNet offers significant improvements. Notably, AnpoNet relies on a unidirectional LSTM, limiting its temporal modeling capacity. In contrast, HyPoNet's BiLSTM allows for richer contextual understanding, resulting in more robust and accurate classification.

The contributions of this paper are as follows:

- We develop the wearable system using a **single triaxial accelerometer** to perform **fine-grained sleep posture classification**, recognizing 12 distinct orientations as nuanced variants of the four main sleep positions.
- We propose **HyPoNet**, a hybrid deep learning architecture that combines CNNs and BiLSTM layers to jointly capture spatial patterns and bidirectional temporal dependencies in motion data.
- We conduct extensive experiments comparing HyPoNet to traditional and deep learning baselines, demonstrating its superior accuracy, robustness, and generalizability across multiple users.
- We highlight the practical potential of our system for **long-term, at-home sleep monitoring** and **real-time positional therapy**, supporting broader applications in personalized healthcare and clinical research.

2 Material and Methods

2.1 Data Acquisition and Sleep Posture Annotation

To support the classification of diverse sleep postures, we developed a compact wearable sensing device (Figure 1), designed to be comfortably worn on the abdomen during sleep. The device integrates an ADXL345 tri-axial accelerometer for motion sensing, managed by an ESP8266 microcontroller, and powered by a rechargeable lithium-ion battery. All components are enclosed in a lightweight, durable plastic housing to ensure comfort and protection during prolonged use, particularly overnight.

The hardware design reflects careful consideration of trade-offs between functionality, energy efficiency, and cost. The ADXL345 was selected over more complex inertial measurement units (e.g., 6-axis or 9-axis IMUs) due to its high sensitivity, low noise, and power-efficient performance in static applications. These attributes make it well-suited for capturing posture-related changes without the additional drift or complexity introduced by gyroscopes or magnetometers. Likewise, the ESP8266 microcontroller offers a compact and cost-effective solution with built-in Wi-Fi capabilities. While more powerful alternatives such as the ESP32 provide additional features (e.g., dual-core processing, Bluetooth), these are unnecessary for this application and would compromise battery life. The selected configuration allows the system to operate reliably for 8–10 hours on a single charge—suitable for overnight monitoring.

To enhance robustness and avoid data loss during potential network disconnections, the device supports dual data logging: it streams real-time data via Wi-Fi using UDP in JSON format and simultaneously buffers the data locally to a microSD card. A dedicated mobile application acts as the client interface, supporting live posture monitoring, file management, and session control.

The system was evaluated through a structured data collection experiment involving ten healthy student volunteers (5 male, 5 female), aged between 19 and 24 years. All volunteers were in stable physical condition with no known musculoskeletal or neurological disorders and had relatively uniform anthropometric characteristics in terms of height and weight. Each volunteer wore the device on the abdomen and was instructed to lie on a flat mattress surface in a standardized set of twelve static sleep postures.

Postures were determined from the orientation angle of the abdomen within the transverse plane, with the reference coordinate system defined so that a right lateral body position corresponded to 0°. The 12 positions included the four canonical sleep orientations—right lateral (0°), supine (90°), left lateral (180°), and prone (270°)—along with eight intermediate variations spaced at 30° intervals: right-up (30°), up-right (60°), up-left (120°), left-up (150°), left-down (210°), down-left (240°), down-right (300°), and right-down (330°). These finer-grained postures represent realistic transitional or asymmetric positions

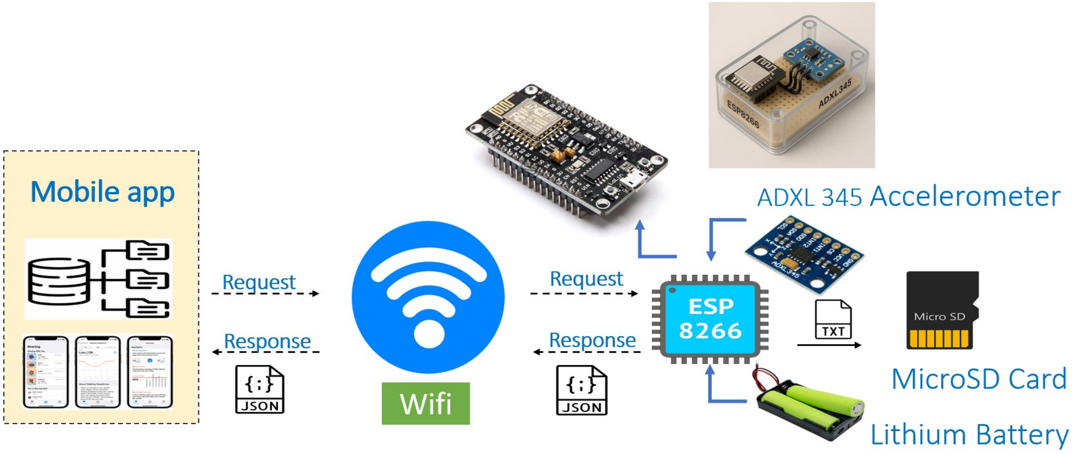

Figure 1. Overview of the wearable device architecture. The illustration shows detailed components of the system, including the ADXL345 accelerometer, ESP8266 microcontroller, rechargeable battery, and microSD storage. It also depicts the data communication flow between components and the wireless transmission of real-time sensor data to the mobile application via a locally hosted Wi-Fi network.

often adopted during natural sleep, which are not captured in conventional four-class classification systems.

Each volunteer remained in one posture for one minute while the acceleration was recorded at a sampling rate of 50Hz, resulting in 3,000 samples per posture. Short rest periods between positions were provided to minimize fatigue. The data were stored using a file-naming convention encoding the volunteer ID and posture label. In total, 120 labeled samples (10 volunteers $\times$ 12 postures) were collected, forming a balanced and diverse dataset for training and evaluating the proposed sleep posture classification model.

2.2 Data Pre-processing

To analyze body posture during sleep using motion data, acceleration signals along the three spatial axes—x, y, and z—denoted as ax, ay, and az, are collected from the wearable tri-axial accelerometer. These signals capture both gravitational orientation and motion-induced dynamics, providing discriminative features for sleep posture classification.

After data acquisition, a preprocessing pipeline is employed to prepare the raw time-series signals for model training. First, a subject-wise data partitioning strategy is adopted to ensure that training, validation, and testing sets contain data from distinct volunteers. This setup enables an evaluation of the model's generalization ability to unseen users, which is crucial for practical deployment in real-world scenarios.

To prepare input samples, the continuous time-series is divided into shorter, fixed-length segments using an overlapping windowing method. Each window captures a short-duration snapshot of the user's acceleration data (ax, ay, az). Importantly, overlapping windows are used, meaning consecutive segments partially share data points. This overlapping technique serves multiple purposes: it increases the number of training samples, preserves the temporal continuity of signals, and enhances the model's ability to detect gradual transitions or boundary cases between postures. Without overlap, subtle posture changes occurring near window edges may be missed or truncated, reducing classification accuracy.

The resulting segments form a structured dataset comprising multivariate time-series windows, which serve as input to the deep learning model. Each window encodes local temporal and spatial patterns that reflect the user's orientation and potential micro-movements, forming the basis for robust posture recognition.

2.3 Classification Module

This study introduces a novel hybrid deep learning architecture named **HyPoNet** (Hybrid Posture Network), which combines the strengths of Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (BiLSTM) networks, as detailed in Figure 2. The architecture is specifically designed for sleep posture classification using triaxial accelerometer data.

HyPoNet starts with a one-dimensional Convolutional layer that extracts local temporal patterns from the input signal, followed by two BiLSTM blocks arranged in sequence. To improve stability during optimization and support better generalization, several **Batch Normalization (BN)** operations are included subsequent to major transformations, while **MaxPooling1D** layers are inserted in between to gradually downsample the temporal dimension. The learned representation is then condensed through **Global Average Pooling**, passed through a **Dropout** layer to mitigate overfitting, and finally mapped to 12 sleep posture categories using a **fully connected (Dense)** layer.

1D Convolutional Layer: The 1D CNN serves as the initial feature extractor. It effectively captures localized patterns in the time series, which are vital for identifying characteristic short-term dynamics of body posture transitions. The convolutional filters slide across the temporal axis of the input, detecting low-level features irrespective of their positions.

BiLSTM Layers: To model longer temporal dependencies and bidirectional context, two stacked BiLSTM layers are utilized. These layers analyze both

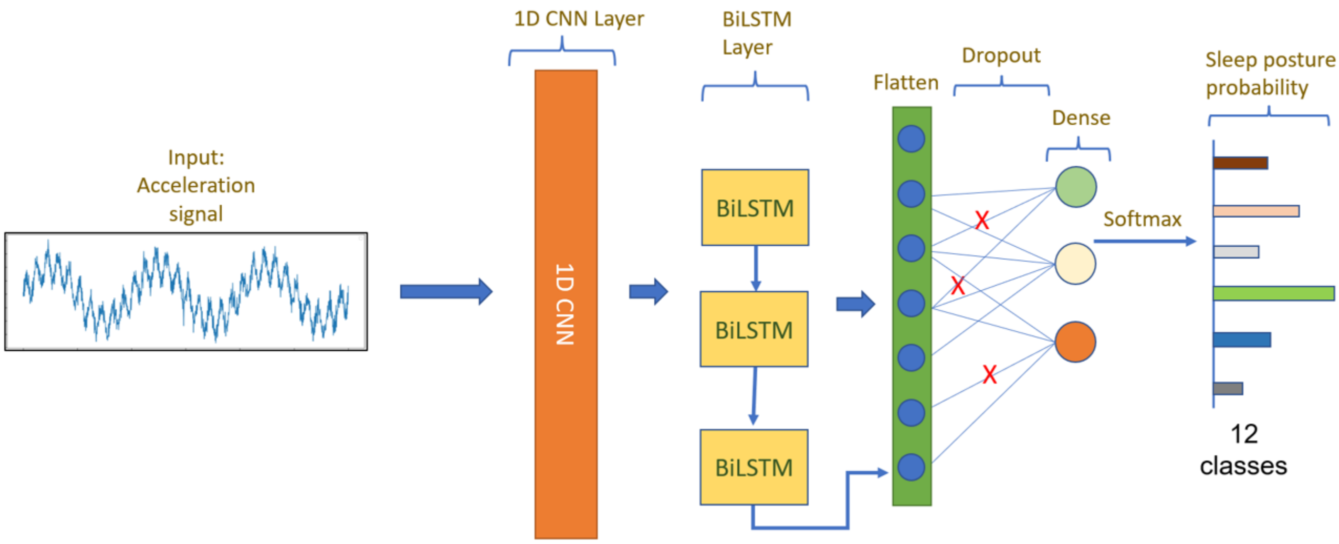

Figure 2. The proposed HypoNet architecture takes preprocessed tri-axial acceleration signals as input and combines 1D Convolutional Neural Networks (1D CNN) with Bidirectional Long Short-Term Memory (BiLSTM) layers to capture both spatial and temporal dependencies, enabling accurate classification of 12 sleep positions.

past and future contexts within the sequence, allowing the network to recognize position changes that evolve over time - a crucial capability for sleep posture classification, where transitions can be gradual and interdependent.

Batch Normalization [19] and Pooling: Each major module is followed by a Batch Normalization layer, which stabilizes and accelerates training by normalizing layer activations. This reduces internal covariate shift and alleviates vanishing/exploding gradient issues. MaxPooling layers are inserted between BiLSTM blocks to reduce the sequence length and introduce a form of time-scale invariance.

Global Pooling and Classification Head: The final sequence output is globally averaged to produce a compact representation of the time series. A Dropout layer follows to mitigate overfitting. Finally, a dense layer maps the features to class probabilities over the 12 sleep positions.

HyPoNet effectively leverages the local feature extraction capability of 1D CNNs and the long-range sequential modeling strength of BiLSTMs. The architecture is designed to be lightweight (only ~21K trainable parameters), enabling efficient deployment on edge devices, while still achieving high accuracy and robustness in sleep posture classification tasks.

3 Experiments

3.1 Comparison models

3.1.1 Models from Previous Studies:

Rawan *et al.* LSTM Model [9]

A recent study by Rawan *et al.* (2023) [9] explored the use of Long Short-Term Memory Networks (LSTMs) for sleep position classification. Their architecture utilized an LSTM unit with a hidden dimension of 25, specifically chosen to capture long-term dependencies in the data. Following the LSTM layer, a fully connected layer with an input dimension of 100 and an output dimension of 12 (increased from the original 4) was implemented. This final layer likely represents the number of sleep positions the model was designed to classify.

The output from the fully connected layer was then processed using a Log Softmax non-linearity function. This function converts the model's raw outputs into probabilities for each sleep position class. The negative log-likelihood loss, derived from these probabilities, was employed as the optimization criterion during training. This loss function measures the discrepancy between the predicted and actual sleep positions, allowing the Adam optimizer [20] to refine the model's weights and improve its classification accuracy.

AnpoNet [18]

Vu *et al.* introduced AnpoNet, a hybrid deep learning architecture designed for analyzing accelerometer data in sleep posture recognition. The model integrates one-dimensional convolutional layers to extract local patterns, followed by batch normalization to enhance training stability and generalization. Temporal dependencies in the signal are captured through an LSTM layer, enabling the model to learn sequential patterns effectively. Specifically, the architecture includes a 1D convolutional layer with a kernel size of 3, stride of 1, and 8 output channels (80 parameters); a 1D batch normalization layer with 8 channels (32 parameters); an LSTM layer with 16 hidden units (1600 parameters); a dropout layer for regularization; a flattening step; and a fully connected layer with 12 output units (204 parameters). In total, the model comprises 1916 trainable parameters, making it lightweight yet capable of capturing both spatial and temporal characteristics in sensor data.

3.1.2 Baseline Models:

1D CNN

The model includes two 1D-CNN layers (8 and 16 filters, $k = 3$), followed by flattening and a 2-layer

MLP (16, 12 units). BN is applied after all Conv/FC layers except the output, and Dropout ($p = 0.4$) is used for regularization.

1D CNN + GRU
The architecture begins with a 1D-CNN (8 filters) followed by BN, MaxPool, and GAP. The resulting features are passed to a GRU layer for temporal modeling, and a final Dense layer (12 units) performs classification, enabling evaluation of the GRU's role in sequence representation.

Transformer [21]
The input data is processed through six transformer encoder blocks, each with 4 attention heads and a head dimension of 256. After flattening, the features pass through a two-layer MLP with 128 and 64 units, including dropout for regularization. The final output layer uses softmax activation with 12 units to predict sleep postures, leveraging self-attention mechanisms to handle long sequences effectively.

3.2 Experimental designs

To ensure the model's ability to generalize and avoid overfitting, the proposed *HypoNet* architecture incorporates a variety of regularization techniques. These include L1 and L2 weight penalties, along with *Dropout* layers distributed throughout the network, which collectively help stabilize learning and improve robustness across volunteers.

An important part of this work was the optimization of temporal segmentation for the input signals. Instead of fixing the number of samples per segment, a time-based sliding window strategy was employed. With a sampling rate of 50 Hz, window durations between 0.5 s and 4.0 s (25–200 samples) were systematically tested. Exploring this wide range made it possible to analyze the impact of window length on the model's ability to recognize and differentiate sleep postures—a factor that has been understudied in earlier research, especially for short and mid-length windows.

In addition, the study examined how different overlap ratios between successive windows affect performance. Overlap levels of 20%, 40%, 60%, and 80% were evaluated to analyze their impact on both accuracy and temporal resolution. Larger overlaps, particularly in combination with shorter windows, provide finer detail and faster response, which are essential characteristics for real-time sleep posture monitoring.

To assess performance in realistic scenarios, a subject-wise cross-validation strategy was adopted. The dataset from 10 participants was split so that records from 8 individuals (≈80%) were allocated to training and validation, while the remaining 2 (≈20%) were reserved for testing only. From the training subset, 20% was further set aside as a validation set for hyperparameter tuning and early stopping.

HyPoNet Architecture. HyPoNet is a hybrid deep learning model that integrates convolutional and recurrent layers to effectively capture both spatial patterns and temporal dependencies in time-series acceleration data. The model takes as input a preprocessed window of tri-axial accelerometer data with shape $(100, 3)$, corresponding to 2 seconds of signal sampled at 50 Hz.

The architecture as in Table I begins with a one-dimensional convolutional layer (**Conv1D**) with 8 filters of size 3 and stride 1, producing an output of shape 100×8. This layer is followed by a **Batch Normalization** layer to stabilize learning.

To model temporal dependencies, a two-stage Bidirectional Long Short-Term Memory (**BiLSTM**) network is employed. The first BiLSTM layer outputs 100×32 features, which are normalized and downsampled using a **MaxPooling1D** layer with pool size 3. The second BiLSTM layer expands the representation to 98×64, followed by additional batch normalization and max pooling operations, resulting in an output of 96×64.

To reduce dimensionality and focus on global temporal features, a **Global Average Pooling** layer is applied, producing a 64-dimensional feature vector. A **Dropout** [22] layer is then used to prevent overfitting, and finally, a fully connected (**Dense**) layer maps the output to 12 units, corresponding to the number of sleep posture classes.

The total number of trainable parameters in HyPoNet is 21,372, making it a compact yet expressive model suitable for embedded and real-time applications.

3.3 Evaluation Metrics

To assess the performance of the proposed classification models, this study adopts two primary evaluation metrics: **Accuracy** and **F1-score**, which are standard in multiclass classification tasks.

The *Accuracy* metric indicates the percentage of correctly predicted instances among the total number of predictions and is calculated as follows

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100\%. \tag{1}$$

Although accuracy provides an overall performance snapshot, it can be misleading in imbalanced datasets. Therefore, the *F1-score* is also employed to account for the balance between *precision* and *recall*, offering a more reliable evaluation when class distributions vary

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

In these formulas, TP, FP, and FN represent the number of true positives, false positives, and false negatives, respectively.

To further interpret the model's classification behavior, a **confusion matrix** is used to provide class-wise prediction outcomes, revealing which classes are commonly confused.

From a system-level perspective, **computational efficiency** is evaluated in terms of:

Table I
LAYER-WISE CONFIGURATION OF **HyPoNet**, A HYBRID CNN-BILSTM MODEL FOR SLEEP POSTURE CLASSIFICATION.

Layer	Kernel / Pool	Output Dim	Stride	#Params
Input (100, 3)	-	-	-	0
Conv1D	3	100 × 8	1	80
BatchNorm1D	-	100 × 8	-	32
BiLSTM-1	-	100 × 32	-	3,200
BatchNorm1D	-	100 × 32	-	128
MaxPooling1D	3	98 × 32	1	0
BiLSTM-2	-	98 × 64	-	16,640
BatchNorm1D	-	98 × 64	-	256
MaxPooling1D	3	96 × 64	1	0
BatchNorm1D	-	96 × 64	-	256
GlobalAvgPooling1D	-	64	-	0
Dropout	-	64	-	0
Dense (FC)	-	12	-	780
			Total Parameters	**21,372**

- **FLOPS** (floating-point operations per second): An indicator of computational complexity.
- **Memory usage** (MB): Total RAM consumed during inference.

4 RESULTS

4.1 Classification Performance

Table II summarizes the performance of HyPoNet compared with baseline and state-of-the-art models on the test set. Reported metrics include mean accuracy and F1-score across all subjects, as well as complexity measures such as parameter size, FLOPs, and memory consumption.

The proposed **HyPoNet** achieved the highest performance across both metrics, with an average **accuracy of 97.29%** and an **F1-score of 90.72%**, while maintaining a moderate model size of **21,372 parameters** and memory footprint of only **152 kB**. These results highlight the model's ability to balance classification accuracy and computational efficiency, making it well-suited for deployment on resource-constrained wearable systems.

Compared to a standard 1D CNN model, HyPoNet significantly outperforms it in both accuracy (by 9.05 percentage points) and F1-score (by 10.38 percentage points), despite having similar memory usage. While the 1D CNN + GRU model achieved a competitive F1-score of 90.52%, its performance was more variable, and its accuracy remained lower than HyPoNet by over 5%.

In contrast, the Transformer-based model, despite its large parameter size (**140,168 parameters**) and high computational cost (5.41 MFLOPs), yielded substantially lower accuracy and F1-score (**57.23%** and **45.48%**, respectively). This suggests that self-attention mechanisms, while powerful, may not be optimal for this specific task or data scale.

Classical models from the literature, including Rawan *et al.*'s LSTM-based approach and AnpoNet, demonstrated modest performance with lower computational requirements. However, both models underperformed compared to HyPoNet, with accuracy drops of over 11% and 20%, respectively.

Overall, the results clearly demonstrate that HyPoNet provides a compelling balance between classification performance and computational cost, reinforcing its suitability for real-time, embedded sleep posture recognition applications.

4.2 Confusion matrix analysis

Figure 3.A presents the confusion matrix for the sleep posture classification task using the proposed model. The results demonstrate that the model achieves outstanding classification performance across most posture categories. Notably, eight out of the twelve classes (`D`, `DL`, `LD`, `LU`, `RD`, `RU`, `U`, `UL`) are classified with perfect accuracy (100%), indicating the model's strong ability to distinguish these postures. Other classes such as `DR` and `UR` also achieve near-perfect performance, with 99 correctly predicted instances and only one misclassification each.

However, the model exhibits a noticeable drop in performance for the `LL` class, with only 78% of instances correctly predicted. A significant portion of the `LL` samples (22%) are misclassified as `LU`, which may be attributed to the subtle variations in body orientation between these two postures, possibly leading to similar signal patterns captured by the wearable sensors. A smaller degree of confusion is also observed for the `LR` class, where 5% of samples are misclassified as `UR`.

4.3 Training Processs

Figure 3.B illustrates the training and validation loss and accuracy curves of **HyPoNet** across 300 training epochs. The model demonstrates rapid convergence, with both training and validation loss sharply decreasing during the initial 50 epochs. After this phase, the

Table II
Performance and Complexity Comparison of Evaluated Models on the Test Set

Model	Accuracy (%)	F1-Score (%)	#Params	#FLOPS (M)	#Memory(kB)
HyPoNet	**97.29 ± 1.29**	**90.72 ± 5.19**	21,372	1.9530	152
1D CNN	88.24 ± 10.97	80.34 ± 18.33	25,436	0.1410	135
1D CNN + GRU	92.15 ± 7.18	90.52 ± 8.68	16,700	1.4633	134
Transformer	57.23 ± 13.77	45.48 ± 8.24	140,168	5.4096	740
Rawan *et al.* (LSTM) [9]	85.50 ± 13.64	82.90 ± 16.12	3,212	0.2806	31
AnpoNet [18]	77.00 ± 18.02	71.86 ± 21.86	1,916	0.1603	34

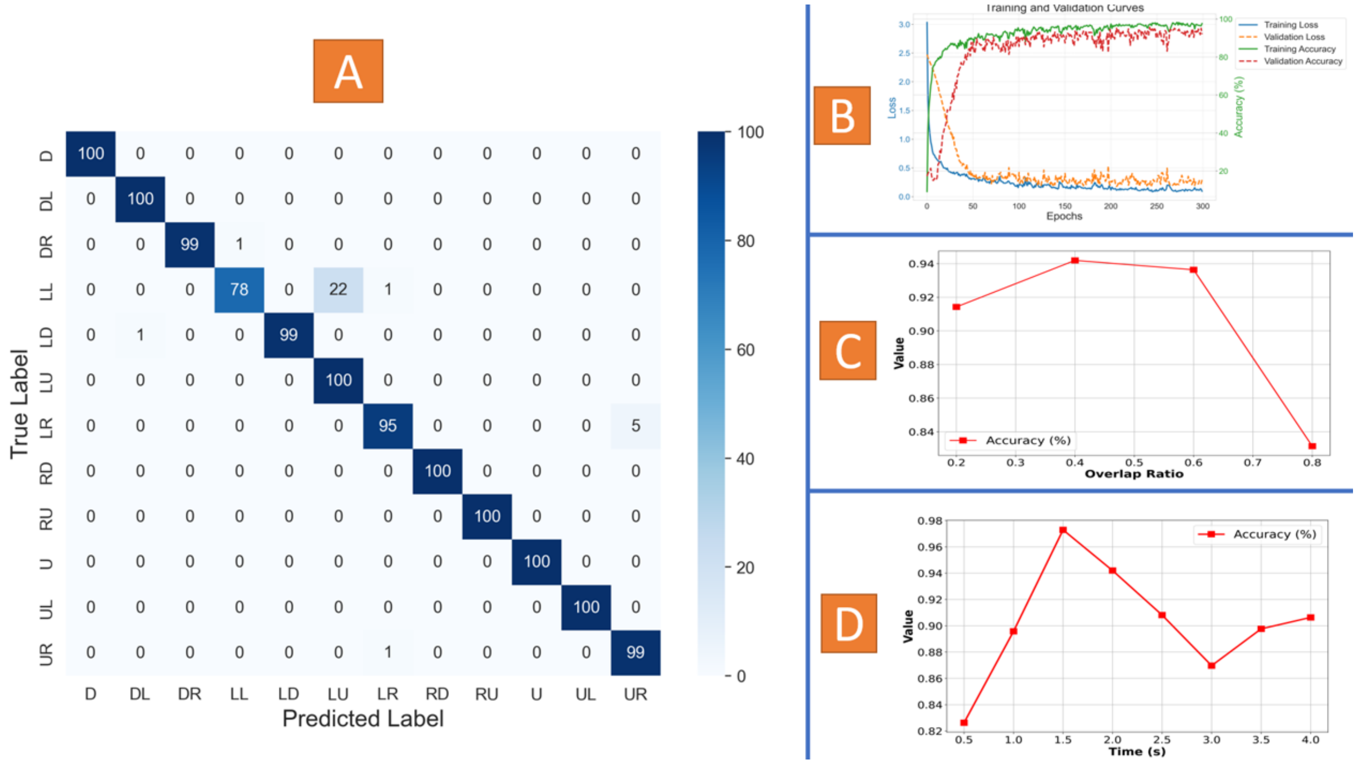

Figure 3. A. Confusion matrices of HypoNet (sleeping position labels: LR – Lateral Right, RU – Right Up, UR – Up Right, U – Supine, UL – Up Left, LU – Left Up, L – Lateral Left, LD – Left Down, DL – Down Left, D – Prone, DR – Down Right, RD – Right Down); B. Training loss and accuracy curves of HypoNet; C. Overlap ratio analysis; D. Impact of varying window sizes (in seconds) on performance.

losses continue to decrease gradually and remain stable, suggesting that **HyPoNet** has effectively captured the underlying temporal and spatial features required for sleep posture classification.

The training accuracy quickly rises and stabilizes above 95% within the first 40 epochs. Meanwhile, the validation accuracy also increases steadily and stabilizes around 90–92% with minimal fluctuations throughout the training process. The small and consistent gap between training and validation accuracy indicates good generalization performance without significant overfitting. Furthermore, the validation loss remains low in the later epochs, further confirming the robustness and stability of the model.

Overall, these results confirm that **HyPoNet** achieves effective learning and generalization, with strong performance on both the training and unseen validation data. The learning curves support the suitability of the proposed hybrid CNN-BiLSTM architecture for the sleep posture classification task.

4.4 Effect of Temporal Segmentation Parameters

Figure 3.C and Figure 3.D illustrate the impact of two key temporal segmentation parameters—window overlap ratio and window size—on the classification performance of the proposed **HyPoNet** model.

As shown in Figure 3.C, increasing the overlap ratio between consecutive windows leads to a noticeable improvement in classification accuracy. This trend suggests that overlapping windows provide beneficial temporal continuity, enhancing the model's ability to capture dynamic transitions in sleep postures. However, performance gains tend to saturate or slightly decline when the overlap ratio becomes excessively high, possibly due to the introduction of redundant information and reduced data diversity. Conversely, using non-overlapping windows results in the lowest accuracy, emphasizing the importance of temporal overlap in sequential data modeling.

Figure 3.D shows that window size also plays a critical role in model performance. The results exhibit

a non-monotonic relationship: while moderate window sizes yield high accuracy, both very short and very long windows degrade performance. Short windows may lack sufficient temporal context for accurate posture recognition, while long windows might include noisy transitions or irrelevant information that negatively affect classification.

Overall, the findings emphasize that segmentation settings play a critical role in determining model effectiveness. The highest classification accuracy was obtained with a window length of **1.5 s** combined with an overlap of **40%**, suggesting that this configuration achieves an effective balance among temporal granularity, contextual coverage, and computational cost.

4.5 Ablation Study

Varying the number of convolutional filters as in Figure 4.A revealed distinct trends in the trade-off between accuracy and F1 score. With 8 filters, the model achieved its highest accuracy (97.3%) and a competitive F1 score of 90.7%, suggesting that a smaller filter set was sufficient to capture the discriminative features in this task. Increasing the number of filters to 16 slightly reduced accuracy to 94.6% but improved the F1 score to 93.0%, indicating a more balanced performance across classes.

However, further increasing the number of filters to 32 resulted in a drop in F1 score to 86.0% while maintaining accuracy at 94.3%, possibly due to overfitting or redundancy in learned feature maps. At 64 filters, the F1 score recovered to 93.5% and accuracy slightly improved to 95.0%, but still did not surpass the best result observed at 8 filters. These results imply that, for this specific task, a relatively small number of filters can achieve strong generalization while avoiding the computational overhead of larger configurations.

The ablation study on different dropout rates as in Figure 4.B revealed a non-linear relationship between regularization strength and model performance. At a low dropout rate of 0.2, the model achieved an accuracy of 79.3% and an F1 score of 68.3%, indicating insufficient regularization and potential overfitting to spurious features. Increasing the dropout rate to 0.4 led to a substantial improvement, with accuracy peaking at 97.3% and the F1 score rising to 90.7%. This result suggests that a moderate level of dropout provides an effective balance between retaining important feature representations and mitigating overfitting.

Increasing the dropout rate to 0.6 led to a substantial decline in performance, with accuracy reduced to 82.7% and the F1-score to 79.9%. At the maximum tested rate of 0.8, the model showed partial recovery, reaching 90.0% accuracy and an F1-score of 82.9%, yet still underperforming compared to the optimal setting of 0.4. The reduced effectiveness at higher dropout values can be attributed to the excessive removal of informative features, limiting the model's capacity to learn discriminative representations. These observations suggest that a dropout probability of 0.4 provides the best trade-off between generalization and predictive strength for this task.

5 Discussion

5.1 Clinical Application

The HyPoNet-based wearable platform carries significant clinical relevance for non-invasive management of sleep-related disorders. Disorders such as OSA and GERD are strongly influenced by body posture during sleep—for instance, supine positioning can aggravate OSA, while GERD symptoms may worsen in the right lateral posture. Positional therapy, which promotes or maintains favorable sleep orientations, has proven effective in symptom reduction without the need for medication.

Traditional diagnostic tools like polysomnography (PSG), while comprehensive, are limited by short monitoring durations and coarse posture categorization. In contrast, the proposed system enables continuous, fine-grained monitoring of 12 distinct body orientations in natural sleeping environments over extended periods. This level of granularity allows for precise identification of posture-induced symptom triggers, making it possible to deliver timely feedback or interventions such as vibrotactile cues to encourage posture correction.

Furthermore, the system's ability to detect intermediate postures enhances diagnostic accuracy and therapy personalization. For instance, a patient frequently rotating between left and supine positions can be more accurately tracked and guided, reducing the risk of undetected harmful postures. The lightweight design and low-power consumption of the wearable make it suitable for long-term use in outpatient settings, while the high classification accuracy of HyPoNet ensures reliability even in the presence of subtle variations in body orientation.

In clinical practice, such a system can serve as a valuable adjunct to PSG by providing longitudinal, real-world data on sleep posture behavior. This information can aid clinicians in evaluating therapy adherence, identifying posture-related symptom patterns, and adjusting treatment plans accordingly. Ultimately, the integration of HyPoNet-based posture monitoring holds promise for advancing personalized, preventive care strategies for sleep-related disorders.

5.2 Limitations and Future Work

Despite its promising performance, HyPoNet remains an early-stage system that requires further development for deployment in practical and clinical settings. The current validation was performed under controlled laboratory conditions with healthy volunteers, focusing on short-term recordings. Such a setup, while useful for initial benchmarking, does not capture the full variability of real-world sleep behaviors influenced by factors such as different bedding, nighttime disturbances, or underlying health conditions. Broader testing—including overnight monitoring and trials involving individuals with sleep-related disorders—will be essential to evaluate generalizability and robustness.

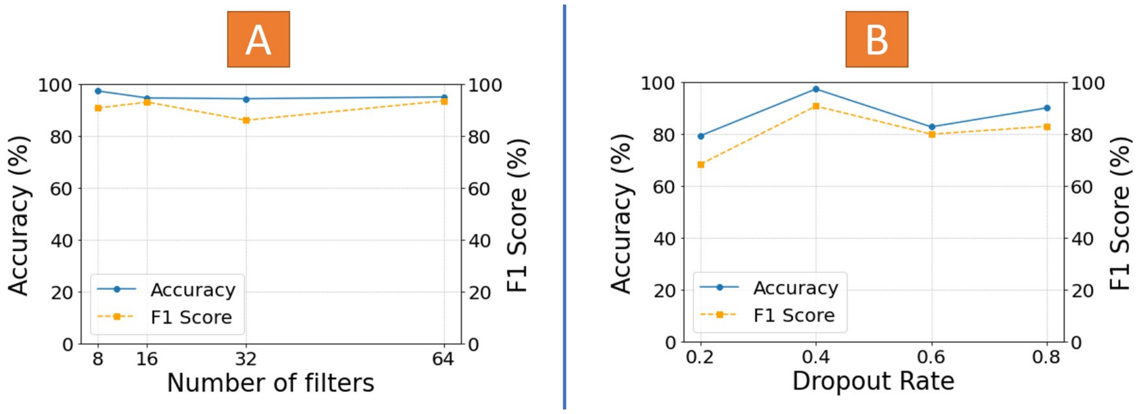

Figure 4. A. Accuracy and F1 Score vs. Number of Filters of HypoNet; B. Accuracy and F1 Score vs. Dropout Rate.;

From a technical standpoint, HyPoNet currently relies solely on acceleration data, which can limit its ability to distinguish between subtle variations in posture, especially among similar angular orientations. While this single-sensor design improves comfort and wearability, future versions could benefit from the integration of gyroscope data to enhance motion characterization. Additional sensing modalities, such as respiratory or cardiac signals, may also support expanded applications beyond posture classification.

Although the number of volunteers in this study is relatively small, the use of sliding-window segmentation (window size 100 samples, 40% overlap) greatly increases the number of training instances, resulting in thousands of labeled windows per class. This ensures that the deep learning model is trained on a sufficiently large dataset to achieve strong performance without heavy reliance on augmentation. Nevertheless, future work will explore standard data augmentation strategies—such as noise injection to simulate sensor variations, time warping for different movement speeds, amplitude scaling to reflect sensor sensitivity changes, and rotational transformations to mimic variations in sensor orientation—which can further improve robustness and reduce potential overfitting to a specific volunteer pool.

Furthermore, while the present work has focused on improving classification accuracy as a foundational step, practical real-time deployment on resource-constrained IoT edge devices will require addressing additional factors beyond memory and FLOPs, including power consumption over extended operation, wireless data transmission reliability, and long-term hardware stability. These engineering considerations will be explored in future development stages when moving from proof-of-concept to large-scale field deployment.

Finally, as sleep monitoring involves the collection of sensitive physiological data, future clinical applications of HyPoNet will need to incorporate robust privacy and security measures. These include compliance with relevant data protection regulations (e.g., HIPAA, GDPR), secure storage and encrypted transmission of data, clear informed consent procedures, and institutional ethical oversight to ensure responsible use of collected information.

6 Conclusion

This work presents **HyPoNet**, a compact deep learning framework designed for detailed classification of twelve distinct sleep postures using signals from a single abdominal tri-axial accelerometer. In contrast to conventional methods that restrict analysis to four broad orientations, HyPoNet can also recognize intermediate postural angles, offering finer granularity that is particularly valuable for positional therapies in disorders such as gastroesophageal reflux disease (GERD) and obstructive sleep apnea (OSA).

The proposed architecture integrates 1D convolutional layers for local feature extraction with a bidirectional LSTM network that captures temporal dependencies in both directions. This hybrid design enables the model to effectively differentiate between subtle variations in body orientation. Evaluated under a subject-independent protocol, HyPoNet achieved an average accuracy of **97.29% ± 1.29** and an F1-score of **90.72% ± 5.19**, outperforming several baseline models while maintaining low computational and memory overhead.

Despite these promising results, several limitations warrant further investigation. The current evaluation was confined to controlled laboratory settings and short-duration recordings from healthy subjects, which may not fully represent the variability of natural sleep conditions. Additionally, the reliance solely on accelerometer data may impede the model's ability to distinguish between closely adjacent postures.

Future work will focus on several key aspects: (i) extending the validation to home-based and overnight recordings, including trials with clinical populations suffering from sleep disorders; (ii) integrating additional sensing modalities, such as gyroscopes, to further enhance motion characterization and classification accuracy; and (iii) developing real-time feedback mechanisms to actively assist users in correcting detrimental sleep postures. The advancement of HyPoNet holds significant potential for non-invasive, continuous sleep monitoring and personalized therapeutic interventions, ultimately contributing to improved clinical outcomes in the management of sleep-related disorders.

ACKNOWLEDGMENTS

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number: NCUD.02-2024.11.

References

[1] K. Zovko, L. Šerić, T. Perković, H. Belani, and P. Šolić, "IoT and health monitoring wearable devices as enabling technologies for sustainable enhancement of life quality in smart environments," *Journal of Cleaner Production*, vol. 413, p. 137506, Aug. 2023.

[2] L. S. Dhingra, A. Aminorroaya, E. K. Oikonomou, A. A. Nargesi, F. P. Wilson, H. M. Krumholz, and R. Khera, "Use of Wearable Devices in Individuals With or at Risk for Cardiovascular Disease in the US, 2019 to 2020," *JAMA Network Open*, vol. 6, no. 6, p. e2316634, Jun. 2023.

[3] D. Jyotishi and S. Dandapat, "An LSTM-based model for person identification using ECG signal," *IEEE Sensors Letters*, vol. 4, no. 8, pp. 1–4, Aug. 2020.

[4] F. Salis, S. Bertuletti, T. Bonci, M. Caruso, K. Scott, L. Alcock, E. Buckley, E. Gazit, C. Hansen *et al.*, "A multisensor wearable system for the assessment of diseased gait in real-world conditions," *Frontiers in Bioengineering and Biotechnology*, vol. 11, p. 1143248, Apr. 2023.

[5] T.-H. Dao, H.-Y. Hoang, V.-N. Hoang, D.-T. Tran, and D.-N. Tran, "Human activity recognition system for moderate performance microcontroller using accelerometer data and random forest algorithm," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 9, no. 4, p. e4, Nov. 2022.

[6] N. T. Thu, T.-H. Dao, B. Q. Bao, D.-N. Tran, P. V. Thanh, and D.-T. Tran, "Real-time wearable-device based activity recognition using machine learning methods," *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 321–333, Jul. 2022.

[7] E. P. Doheny, M. M. Lowery, A. Russell, and S. Ryan, "Estimation Of Respiration Rate And Sleeping Position Using A Wearable Accelerometer," *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC*, pp. 4668–4671, 2020.

[8] R. S. Abdulsadig, S. Singh, Z. Patel, and E. Rodriguez-Villegas, "Sleep Posture Detection Using An Accelerometer Placed On The Neck," in *Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2022, pp. 2430–2433.

[9] R. S. Abdulsadig and E. Rodriguez-Villegas, "Sleep Posture Monitoring Using A Single Neck- Situated Accelerometer: A Proof-of-concept," *IEEE Access*, vol. 11, pp. 17 693–17 706, Jan. 2023.

[10] P. Jeng and L.-C. Wang, "An accurate, low-cost, easy-to-use sleep posture monitoring system," in *Proceedings of the 2017 International Conference on Applied System Innovation (ICASI)*. IEEE, May 2017, pp. 903–905.

[11] P. Vonk, M. Ravesloot, and N. de Vries, "Positional therapy for positional obstructive sleep apnea: What is new?" *Current Sleep Medicine Reports*, vol. 3, no. 3, pp. 113–121, Aug. 2017.

[12] J. P. van Maanen and N. de Vries, "Long-term effectiveness and compliance of positional therapy with the sleep position trainer in the treatment of positional obstructive sleep apnea syndrome," *Sleep*, vol. 37, no. 7, pp. 1209–1215, Jul. 2014.

[13] J. M. Schuitenmaker, T. Kuipers, R. A. Oude Nijhuis, M. P. Schijven, A. J. M. Smout, P. Fockens, and A. J. Bredenoord, "Sleep positional therapy for nocturnal gastroesophageal reflux: A double-blind, randomized, sham-controlled trial," *Clinical Gastroenterology and Hepatology*, vol. 20, no. 12, pp. 2753–2762.e2, Dec. 2022.

[14] P. Dave, "Factors Contributing to Sleep Disorders Among Young Adults," *Asian Journal of Dental and Health Sciences*, vol. 4, no. 2, pp. 26–31, Jun. 2024.

[15] P. R. Srijithesh, R. Aghoram, A. Goel, and J. Dhanya, "Positional therapy for obstructive sleep apnoea," *Cochrane Database of Systematic Reviews*, vol. 2019, no. 11, May 2019.

[16] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, Apr. 2021.

[17] M. Khan, H. Wang, A. Riaz, A. Elfatyany, and S. Karim, "Bidirectional lstm-rnn-based hybrid deep learning frameworks for univariate time series classification," *The Journal of Supercomputing*, vol. 77, no. 7, pp. 7021–7045, Jan. 2021.

[18] H.-D. Vu, D.-N. Tran, H.-H. Pham, D.-D. Pham, K.-L. Can, T.-H. Dao, and D.-T. Tran, "Human sleep position classification using a lightweight model and acceleration data," *Sleep and Breathing*, vol. 29, no. 1, Feb. 2025.

[19] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training By Reducing Internal Covariate Shift," in *Proceedings of the International Conference on Machine Learning*. pmlr, Jul. 2015, pp. 448–456.

[20] D. Kingma and J. Ba, "Adam: A Method For Stochastic Optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diega, CA, USA, Dec. 2015, pp. 1–15.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Proceedings of the 31st International Conference on Neural Information Processing Systems*, p. 6000–6010, Dec. 2017.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

Hoang-Dieu Vu is a lecturer at Phenikaa University's Faculty of Electrical and Electronic Engineering. Before this, he spent six years honing his research skills as a member of the Data Science Lab at Hanoi University of Science and Technology (HUST). His academic achievements are impressive, including an excellent B.Eng degree with a high CPA in Information Technology, the best thesis award in 2018, and a Master of Science in Data Science (2020). Mr. Vu's passion lies in machine learning, deep learning, and continual learning, fueled by his research experience at HUST's Data Science Lab (since early 2016) and a research internship at the Korea Advanced Institute of Science and Technology (KAIST). He serves as a technical committee program member and reviewer of many international conferences.

Duc-Nghia Tran is a PhD at the Institute of Information Technology, Vietnam Academy of Science and Technology (VAST). He received his PhD degree from Sorbonne Paris Cité (France). In his thesis, he focuses on signal processing of EPR spectra for in vivo experiments. His research interests are AI, IoT, Electron Paramagnetic Resonance, parameter estimation, and data analysis. He serves as a technical committee program member and reviewer of many international conferences and journals.

Quang-Huy Pham is a lecturer at the Faculty of Information Technology, Electric Power University, Vietnam. He received his PhD degree in Information Systems from the Graduate University of Science and Technology (GUST), Vietnam Academy of Science and Technology, in 2024. His research interests include network security, digital watermarking, and steganography. He has published many research papers in both national and international journals and conferences.

Duc-Tan Tran is a Professor and Dean of the Faculty of Electrical and Electronic Engineering, Phenikaa School of Engineering, Phenikaa University. He has published over 150 research papers. His main research interests include the representation, processing, analysis, and communication of information embedded in signals and datasets. He serves as a TP Co-chair, technical committee program member, track chair, session chair, and reviewer of many international conferences and journals.