MINISTRY OF EDUCATION
AND TRAINING

VIETNAM ACADEMY OF
SIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SIENCE AND TECHNOLOGY

VU HOANG DIEU

ENHANCING THE EFFECTIVENESS OF SLEEP POSTURE CLASSIFICATION FROM TRIAXIAL ACCELERATION DATA USING HYBRID DEEP LEARNING MODELS

SUMMARY OF DISSERTATION ON COMPUTER SCIENCE
Code: 9 48 01 01

Hanoi - 2026

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy of Science and Technology

Supervisors:

Supervisor 1: Dr. Tran Duc Nghia, Institute of Information Technology, Vietnam Academy of Science and Technology

Supervisor 2: Prof. Tran Duc Tan, Phenikaa University

Referee 1: Assoc. Prof. Tran Giang Son

Referee 2: Dr. Le Hai Anh

The dissertation will be examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology at 9:00 AM on 26 August 2026.

This dissertation can be found at:

1) Graduate University of Science and Technology Library
2) National Library of Vietnam

INTRODUCTION

1. Rationale of the dissertation

Recent progress in wearable sensors and intelligent health-care systems has created favorable conditions for continuous sleep monitoring. Among the available sensors, the triaxial accelerometer is attractive because it is compact, low-cost, low-power, non-invasive, and capable of recording body orientation during sleep without disturbing daily life.

The dissertation therefore addresses the fine-grained classification of twelve static sleep postures using one accelerometer attached to the abdomen. This sensor position is practical for long-term wearable use and reflects whole-body orientation more directly than many peripheral placements.

The research also considers computational constraints. A sleep-posture classifier should be small, fast, and energy-efficient enough for wearable or edge devices. Direct embedded deployment is outside the dissertation scope, but edge feasibility is assessed through parameter count, FLOPs, memory, and inference time.

2. Research objectives of the dissertation

The general objective is to develop and evaluate deep-learning models for classifying twelve sleep postures from triaxial abdominal accelerometer data. The specific objectives are: to formulate the task as a multiclass time-series classification problem; to design AnpoNet by combining one-dimensional convolution and LSTM modules while controlling model size; to design CGB-Net by combining one-dimensional convolution, GRU, and BiLSTM modules for richer spatio-temporal representation; to build a rigorous experimental procedure with subject-independent evaluation, baseline comparison, window analysis, and per-class error analysis; and to clarify the advantages, limitations, and future applicability of the proposed models.

3. Research object and scope

The research object is human sleep-posture recognition from triaxial accelerometer signals. The dissertation focuses on twelve predefined static postures collected in controlled conditions using a single abdominal sensor.

The practical scope is offline analysis of pre-collected and segmented signals. Overnight natural sleep, transitional postures, ambiguous movements, direct clinical validation, and full embedded implementation are left for future research. Therefore, the clinical evidence discussed in the dissertation supports the motivation for fine-grained posture monitoring rather than serving as a direct diagnostic conclusion.

4. Contents of the dissertation

The dissertation first reviews sleep-posture classification and sensor-based approaches. It then describes the accelerometer dataset, the labeling protocol, preprocessing, and subject-independent organization.

5. Research methods

The dissertation combines literature review, experimental dataset design, deep-learning modeling, and quantitative evaluation. The data are collected from a standardized protocol, then normalized and segmented into fixed-length windows. Models are trained directly on time-series accelerometer samples rather than relying only on handcrafted features.

6. Main contributions of the dissertation

The dissertation contributes: (i) AnpoNet, a hybrid deep-learning network for twelve-class sleep-posture classification from triaxial accelerometer data, designed with a small parameter budget for resource-constrained devices; and (ii) CGB-Net, a hybrid network that improves the accuracy and balance of twelve-class classification by combining convolutional, GRU, and bidirectional LSTM mechanisms.

7. Structure of the dissertation

Besides the Introduction, Conclusion and Recommendations, the dissertation has three chapters: Chapter 1 reviews the problem and related methods; Chapter 2 presents dataset design and AnpoNet; and Chapter 3 presents CGB-Net and its experimental assessment.

Chapter 1. OVERVIEW OF SLEEP POSTURE CLASSIFICATION AND SENSOR-BASED APPROACHES

1.1. The sleep posture classification problem

1.1.1. Clinical importance of sleep posture monitoring

Sleep posture is a simple behavioral variable, but it can influence several health-related phenomena during sleep. In obstructive sleep apnea, the severity of events may vary with body orientation.

A conventional four-class description cannot always capture clinically meaningful differences. A person lying on the side at a shallow angle may produce different pressure or respiratory effects from a person lying fully lateral.

The accelerometer is appropriate for this task because, in static conditions, its three axes measure the gravity components projected onto the sensor coordinate system. Body orientation can thus be inferred from stable patterns in the triaxial signal. However, small posture differences create small signal differences, and the measurements may be affected by sensor placement, breathing movement, strap tightness, and individual body morphology.

1.1.2. High-resolution and intermediate postures in clinical studies

The dissertation emphasizes high-resolution posture classes and intermediate orientations. Such postures bridge the gap between simple categories and continuous body-angle descriptions.

1.2. Overview of sleep posture classification systems

Existing systems for sleep-posture classification can be grouped by sensing modality. Camera-based methods can provide rich spatial information, but they raise privacy concerns, require favorable lighting or infrared configuration, and may be obstructed by bedding.

Within wearable systems, sensor location strongly affects performance and usability. Sensors on the chest, waist, abdomen, neck, or limbs capture different aspects of orientation. The abdomen is a practical compromise because it is close to the trunk center and can represent whole-body orientation while remaining acceptable for sleeping.

Prior works differ in the number of classes, data-collection protocol, evaluation strategy, and classification method. Many studies report high performance for three or four postures, but the task becomes more difficult when the number of classes increases or when evaluation excludes test subjects from training. The dissertation identifies three main gaps: limited posture resolution, insufficient subject-independent validation, and limited attention to model complexity for wearable deployment.

1.3. Overview of machine-learning methods for sleep posture classification

1.3.1. Traditional machine-learning approaches

Traditional approaches often extract handcrafted features from each time window, such as mean acceleration, standard deviation, magnitude, signal energy, tilt angle, and correlation between axes. These features are then classified by decision trees, support vector machines, random forests, gradient boosting, or multilayer perceptrons. Such methods can work well when classes are clearly separated and the sensing setup is simple.

However, handcrafted features may not capture subtle temporal and spatial patterns required by twelve-class posture recognition. Treating each window as a static feature vector may ignore the consistency of orientation over time. Moreover, features tuned for one sensor location or group of subjects may not generalize well to new people.

1.3.2. Deep learning for sensor time-series data

Deep learning is suitable for sensor time series because it can learn hierarchical features from data. Convolutional layers detect local patterns and interactions among axes, recurrent layers summarize temporal dependencies, and hybrid structures combine both advantages. For sleep posture, the signal is not highly dynamic, but the time-series structure still matters because stable posture cues must be accumulated and noise must be suppressed.

1.3.3. Convolutional neural networks for sensor signals

1.3.4. Recurrent neural networks for time-series modeling

1.3.5. Hybrid CNN-RNN architectures for sensor data

1.4. Preliminary experiment: feasibility validation

1.5. Evaluation metrics

The dissertation uses standard multiclass metrics: accuracy, precision, recall, F1-score, and the confusion matrix. These metrics are necessary because high overall accuracy can hide weak performance on difficult or minority classes.

1.5.1. Accuracy

1.5.2. Precision

1.5.3. Recall

1.5.4. F1-score

1.5.5. Confusion Matrix

1.6. Summary

Chapter 1 establishes the scientific and technical basis of the dissertation. Sleep posture is clinically meaningful, and twelve-class classification can provide more detailed information than conventional coarse labels.

Chapter 2. PROPOSAL OF ANPONET WITH A CONSTRAINED PARAMETER BUDGET FOR SLEEP POSTURE CLASSIFICATION

2.1. Methodology for dataset design

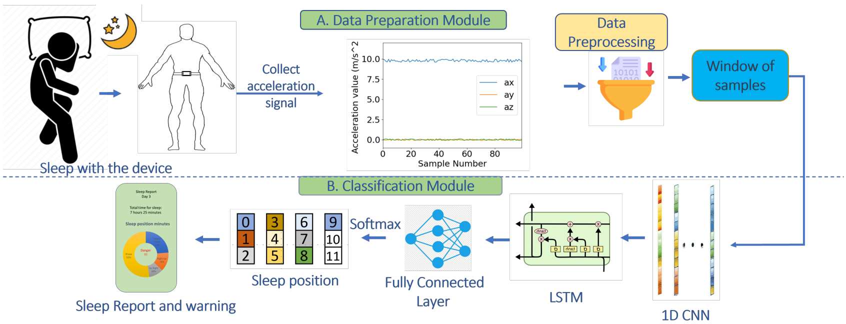

Figure 1: The sleep position classification system

2.1.1. Objectives and scope of the dataset

The dataset is designed to support the classification of twelve static sleep postures from one abdominal triaxial accelerometer. The main objective is to create a controlled and balanced dataset in which posture labels are reliable and the subject-independent split can be applied. The dataset is not intended to represent all natural overnight sleep behaviors.

Each participant performs twelve postures according to a standardized protocol. The posture set includes the main sleeping orientations and intermediate angular variants. This design allows the dissertation to test whether fine-grained differences can be detected by a single accelerometer.

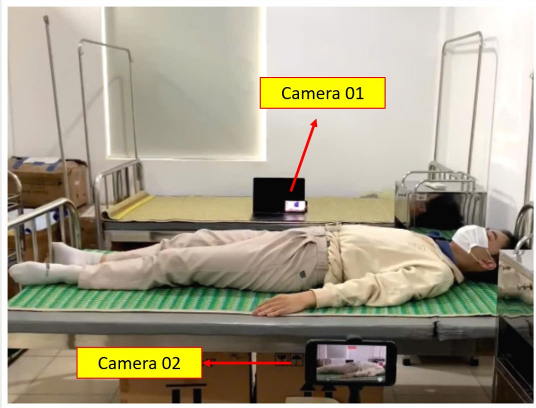

Figure 2: Data collection process

2.1.2. Sensor configuration and hardware setup

The sensing system consists of two modules: an accelerometer-based wearable device and a data-collection interface. The accelerometer records acceleration along the x, y, and z axes. The abdominal placement is selected because it reflects trunk orientation and is suitable for comfortable wear during sleep.

2.1.3. Standardized data-collection protocol

The collection protocol controls posture order, sensor placement, sampling rate, and recording duration. Each posture is maintained for sixty seconds, and the sampling frequency is 50 Hz. Therefore, each posture file contains about 3,000 temporal samples per participant.

The twelve postures are encoded by integer labels from 0 to 11. The classes include supine, prone, left-side and right-side orientations, as well as upward and downward intermediate variants. The posture labels are assigned at the file level because all samples within one recording correspond to the same instructed posture.

2.1.4. Labeling and preprocessing procedure

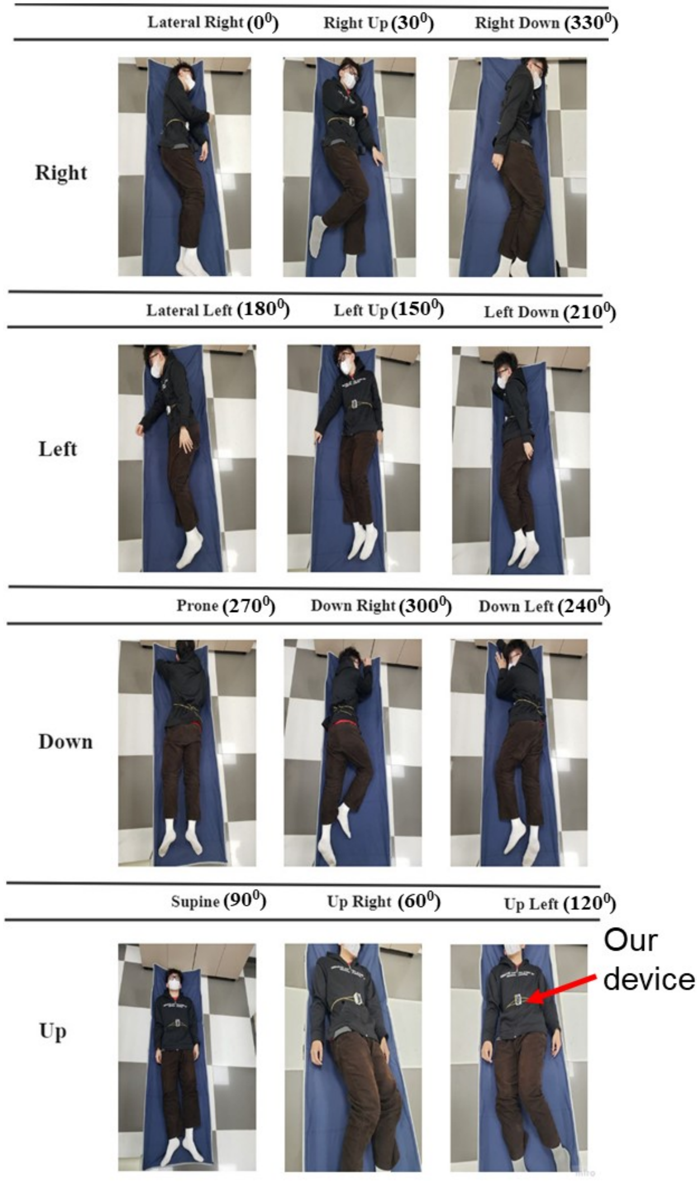

Figure 3: 12 sleep positions

The raw CSV files contain timestamp, acceleration along three axes, and posture label. After reading the files, the data are organized by subject before any segmentation is performed. This is crucial: if segmentation were performed first and windows were split randomly, windows from the same person could appear in both training and test sets.

2.1.5. Datasets

Dataset A contains the participants used for the AnpoNet experiments in Chapter 2. The physical characteristics of participants are recorded to document the variability of height, weight, and waist circumference. Dataset B extends the number of participants to eighteen subjects and is used in Chapter 3.

2.1.6. Experimental setup

Training and evaluation are conducted in two computing environments. Model training and hyperparameter tuning use a high-performance workstation equipped with an NVIDIA RTX A5000 GPU and TensorFlow.

2.1.7. Input data processing and dataset organization

The accelerometer sequence is segmented into sliding windows. A typical configuration uses a fixed window length of 100 samples and an overlap ratio of 0.4. At 50 Hz, a 100-sample window corresponds to 2 seconds.

Each window is represented as a multivariate time series with three channels. The target is one of the twelve posture classes. The training, validation, and test sets are organized so that all data from one participant belong to only one set.

2.2. AnpoNet design toward edge devices

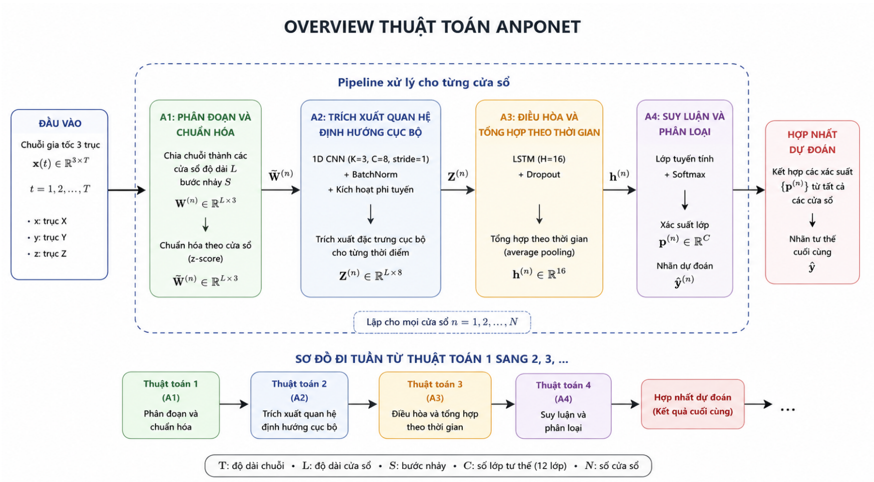

Figure 4: Overview of AnpoNet

Sleep-posture recognition from abdominal acceleration differs from general human-activity recognition. The relevant information is mostly static orientation, not vigorous motion. A compact model must therefore learn subtle and stable patterns without excessive parameters.

2.2.1. AnpoNet approach

AnpoNet follows three principles. First, it must distinguish spatial orientation by learning relationships among the accelerometer axes. Second, it must model temporal stability, because sleep postures remain nearly constant and short-term sensor noise should not dominate predictions.

2.2.2. Problem formulation

Let a segmented accelerometer window be denoted by $X \in \mathbb{R}^{T\times 3}$, where T is the number of time samples and the three columns correspond to the acceleration axes. The classifier learns a mapping $f(X;\theta)$ from the input sequence to one of $C = 12$ posture labels. The output layer produces a probability vector through a softmax function:

$$\hat{y}_c = \frac{\exp(z_c)}{\sum_{j=1}^{C} \exp(z_j)}, \quad c = 1, \ldots, C.$$

The predicted label is the class with maximum probability. The learning problem is to find parameters θ that minimize the discrepancy between predicted probabilities and true labels over the training set while preserving generalization to unseen subjects.

The subject-independent setting can be described as follows. Given a set of subjects S, the subjects are partitioned into disjoint training, validation, and test subsets. If subject s belongs to the test subset, none of that subject's windows are used during training.

2.2.3. Design of the AnpoNet algorithmic framework

The AnpoNet framework begins with a 1D convolutional block. This block extracts local patterns from the triaxial signal and emphasizes relationships among axes. Batch Normalization stabilizes the distribution of activations and improves optimization.

The model is intentionally small. Instead of stacking many convolutional and recurrent layers, AnpoNet uses a minimal structure that matches the information content of the signal. The hypothesis is that posture classification does not require very deep networks if the architecture is aligned with the physics of acceleration: the gravity vector carries orientation, and temporal recurrence confirms consistency.

2.2.4. Training problem and loss function

The model is trained using categorical cross-entropy:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{12} y_{i,c} \log(\hat{y}_{i,c}).$$

Adam is used as the optimizer with an initial learning rate of 0.001, and the experiments use a mini-batch size of 512 and up to 200 epochs.

Multiple random seeds are used to assess stability. The loss function encourages high probability for the true posture class while penalizing incorrect predictions across all classes.

2.2.5. Complexity analysis and algorithmic characteristics of AnpoNet

AnpoNet is analyzed by parameter count, FLOPs, memory use, inference time, and estimated energy. These indicators are more relevant to wearable deployment than accuracy alone. A model with high accuracy but slow inference or large memory could be impractical for embedded systems.

2.3. Experiments and evaluation

2.3.1. Comparison models

AnpoNet is compared with traditional and deep-learning baselines. The traditional group includes Decision Tree, XGBoost, MLP, and ExtraTrees-based approaches. The deep-learning group includes 1D CNN, 1D CNN + GRU, LSTM variants from related work, and an ablated AnpoNet without Batch Normalization.

2.3.2. Experimental design

The experiments use subject-independent splits. Window lengths from 10 to 140 samples and overlap ratios from 10% to 70% are investigated. Accuracy and F1-score are reported as the main metrics, with precision, recall, and confusion matrices for detailed analysis.

2.4. Experimental results and discussion

2.4.1. Performance analysis

The main result is that AnpoNet achieves the strongest performance in Chapter 2. On the test set, it reaches 94.67% accuracy and 92.94% F1-score. This is higher than the best compared deep baseline, 1D CNN + GRU, which reaches 90.81% accuracy but only 72.64% F1-score.

Table 2.6: Comparison of model performance on the test dataset

Model	Accuracy (%)	F1-score (%)
Decision Tree	81.82	78.67
XGBoost	85.34	82.94
MLP	81.25	72.96
1D CNN	87.08	78.19
1D CNN + GRU	90.81	72.64
ExtraTrees baseline	86.50	84.68
LSTM baseline	68.84	65.02
LSTM (CT4)	87.44	72.30
AnpoNet	**94.67**	**92.94**
AnpoNet without BN	71.11	54.88

The ablation result is especially informative. Removing Batch Normalization reduces accuracy to 71.11% and F1-score to 54.88%. This shows that normalization is not a minor accessory in this task; it stabilizes training and improves generalization across subjects.

2.4.2. Evaluation under different random seeds

Random-seed experiments test whether the model is stable under different initializations and data shuffling. AnpoNet maintains high accuracy and F1-score across seeds, demonstrating that the result is not a single lucky run. This matters because small datasets and subject-independent splits can otherwise produce unstable outcomes.

2.4.3. Evaluation based on K-Folds

K-fold subject-independent validation further examines generalization. In each fold, different subjects are held out for testing. The results confirm that AnpoNet can learn features transferable across participants.

2.4.4. Influence of sequence length and overlap ratio on model performance

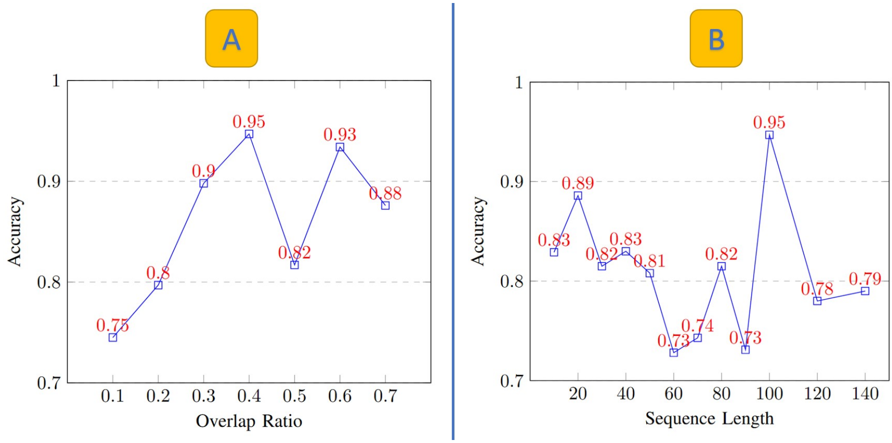

Figure 5: Effect of overlap ratio and sequence length

Window configuration affects the trade-off between accuracy and latency. Very short windows may not provide enough temporal context for stable decisions, while very long windows may include redundant information and increase response delay. The dissertation finds that moderate windows, especially around 100 samples or 2 seconds at 50 Hz, provide a good balance.

2.4.5. Training process analysis

The training curves show effective learning and convergence. Training and validation accuracy increase during the early epochs, while loss decreases. The best validation performance occurs before excessive training, suggesting that early stopping can be useful.

2.4.6. Model complexity and computational efficiency analysis

AnpoNet is the smallest model among the compared architectures, with 1,916 parameters and about 34 kB of memory. Its inference time is 0.081 seconds on the CPU-only test environment, which is faster than all compared models. Although some baselines have low FLOPs, they may require more parameters or memory and still perform worse.

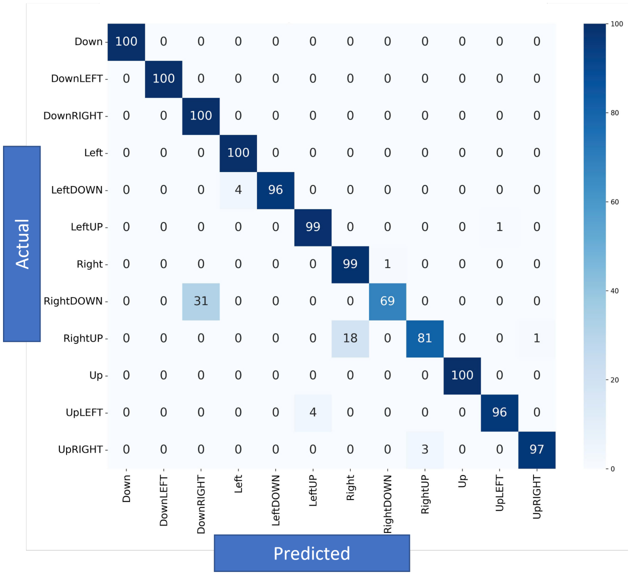

Figure 6: Confusion matrix of AnpoNet

Table 2.7: Model complexity and computational efficiency

Model	Params	FLOPs (M)	Time (s)	Memory (kB)	Energy (mJ)
MLP	22,124	0.0433	0.304	116	0.304
1D CNN	25,436	0.1410	0.204	135	0.989
1D CNN + GRU	16,700	1.4633	1.481	134	10.257
LSTM baseline	3,212	0.2806	0.404	31	1.968
LSTM (CT4)	27,964	0.2548	0.897	154	1.787
AnpoNet	**1,916**	**0.1603**	**0.081**	**34**	**1.122**

2.4.7. Performance comparison with architecturally equivalent models

To isolate the benefit of the architecture from the effect of model size, the dissertation compares AnpoNet with a small LSTM having a comparable number of parameters. AnpoNet reaches 94.67% accuracy and 92.94% F1-score, while the small LSTM reaches 84.28% accuracy and 61.91% F1-score. This confirms that AnpoNet's advantage is not simply due to having more parameters.

2.4.8. Impact of white noise on model robustness

Robustness is tested by adding white noise to the signal. AnpoNet with noise reaches 92.44% accuracy and 80.20% F1-score, compared with 94.67% and 92.94% without noise. The accuracy drop is moderate, which indicates that the model can still recognize many postures under disturbed conditions.

2.4.9. Class-wise performance

The confusion matrix and per-class precision-recall analysis show that AnpoNet performs perfectly on many classes, including several basic and intermediate orientations. Most precision and recall values are near 1.00.

Table 2.10: Precision and recall for each class

Class	Precision	Recall	Class	Precision	Recall
D	1.00	1.00	LD	1.00	0.89
DL	1.00	1.00	LU	0.98	1.00
DR	0.86	1.00	LR (R)	0.91	1.00
LL (L)	0.91	1.00	RD	1.00	0.63
RU	1.00	0.82	U	1.00	1.00
UL	1.00	0.98	UR	1.00	1.00

2.5. Conclusion of Chapter 2

Chapter 2 demonstrates that a compact hybrid architecture can classify twelve sleep postures accurately under subject-independent evaluation. AnpoNet combines 1D convolution, LSTM, and Batch Normalization to achieve high accuracy, balanced F1-score, low parameter count, short inference time, and low memory use. It is therefore a promising candidate for wearable sleep-posture monitoring with limited resources.

Chapter 3. PROPOSAL OF CGB-NET TO ENHANCE ACCURACY IN SLEEP POSTURE CLASSIFICATION

3.1. Problem and research issues

Chapter 3 addresses the need to improve classification accuracy and generalization beyond the compact design of AnpoNet. Although AnpoNet is highly efficient, the twelve-class task remains challenging when the number of subjects increases and inter-subject variability becomes more pronounced. Some posture pairs are still difficult because their angular differences are small and their accelerometer patterns overlap.

The research issue is to design a network that can extract richer spatio-temporal representations while keeping complexity at a reasonable level. The model should learn local orientation features, short- and medium-term temporal dependencies, and bidirectional context within each window. It should also preserve balanced performance across classes, as measured by F1-score, because overall accuracy alone may hide errors in difficult postures.

3.2. Motivation for the proposed model

The motivation for CGB-Net comes from the complementary strengths of three modules. A 1D CNN can identify local patterns and axis relationships in the accelerometer sequence.

CGB-Net is therefore not designed as the smallest possible model. Instead, it seeks a better balance between representational capacity and computational cost. Compared with AnpoNet, it uses more parameters and FLOPs, but it remains much smaller than many deep networks and uses only 9,532 parameters.

3.3. Proposed CGB-Net algorithmic framework

CGB-Net receives a triaxial accelerometer window as input. The first stage applies 1D convolution to extract local spatio-temporal features. These features are then passed to a GRU stage, which summarizes sequential dependencies and captures stable temporal structure

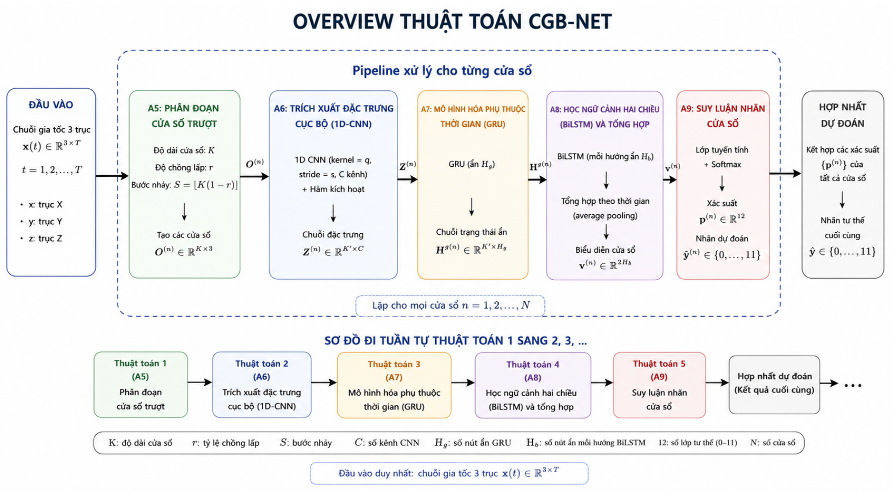

Figure 7: Overview of the CGB-Net

with moderate parameter cost.

The architecture can be interpreted as a staged feature learner. The CNN handles local patterns; the GRU compresses and regularizes temporal dependencies; and the BiLSTM integrates richer context. This arrangement is expected to reduce confusion among adjacent posture classes and improve the macro-level balance of precision and recall.

3.4. Experiments

3.4.1. Comparison models

The Chapter 3 experiments compare CGB-Net with MLP, 1D CNN, 1D CNN + BiLSTM, 1D CNN + GRU, the LSTM model from Rawan et al., AnpoNet, and the proposed CGB-Net. This set of baselines clarifies the incremental value of each architectural component. For example, comparing CGB-Net with 1D CNN tests the value of recurrent context; comparing it with CNN-GRU and CNN-BiLSTM tests the value of combining two recurrent mechanisms; and comparing it with AnpoNet tests whether additional capacity improves accuracy on the expanded dataset.

3.4.2. Model training and optimization

The same general training philosophy as Chapter 2 is retained. The dataset is split by subject, the models are trained with deep-learning frameworks, and evaluation is repeated across different settings. Accuracy and F1-score are the principal metrics.

3.5. Experimental results and discussion

3.5.1. Model performance analysis

CGB-Net achieves the best overall performance across most experimental settings. The four-setting results show that CGB-Net reaches accuracy values of 89.34%, 84.39%, 92.70%, and 83.95%, and F1-score values of 86.24%, 79.41%, 89.71%, and 78.17%. The improvements over AnpoNet are especially clear in F1-score, indicating better balanced recognition across the twelve classes.

Table 3.1: Accuracy (%) of models under four experimental settings

Model	Set 1	Set 2	Set 3	Set 4
MLP	82.02	81.25	88.07	78.18
1D CNN	86.38	77.84	89.52	83.20
1D CNN + BiLSTM	85.99	75.96	88.04	77.91
1D CNN + GRU	85.70	77.07	88.49	79.88
Rawan et al. LSTM	77.45	78.97	90.63	88.60
AnpoNet	85.20	81.63	92.18	81.23
CGB-Net	**89.34**	**84.39**	**92.70**	83.95

The advantage of CGB-Net is not uniform only in accuracy. In several settings, the F1-score gain is larger than the accuracy gain. This means the model is better at recognizing difficult classes rather than merely improving easy classes.

Table 3.2: F1-score (%) of models under four experimental settings

Model	Set 1	Set 2	Set 3	Set 4
MLP	75.86	72.18	77.71	68.68
1D CNN	82.20	66.93	85.57	78.64
1D CNN + BiLSTM	80.11	68.48	83.20	69.20
1D CNN + GRU	79.98	69.07	85.85	69.90
Rawan et al. LSTM	65.87	67.40	85.75	75.80
AnpoNet	76.10	70.54	88.21	72.01
CGB-Net	**86.24**	**79.41**	**89.71**	**78.17**

The cross-validation summary confirms this conclusion. CGB-Net reaches an average accuracy of 87.60% ± 4.19 and an average F1-score of 83.38% ± 5.51. AnpoNet reaches 85.06% ± 5.07 accuracy and 76.72% ± 8.02 F1-score.

Table 3.3: Average performance and computational cost on the test set

Model	Accuracy (%)	F1-score (%)	Params	FLOPs (M)	Memory (kB)
MLP	82.38 ± 4.14	73.61 ± 4.01	22,124	0.0433	116
1D CNN	84.24 ± 4.98	78.34 ± 8.11	25,436	0.1410	135
CNN + BiLSTM	81.98 ± 5.93	75.25 ± 7.51	21,372	1.9530	152
CNN + GRU	82.79 ± 5.23	76.20 ± 8.12	16,700	1.4633	134
LSTM baseline	83.91 ± 6.67	73.71 ± 9.14	3,212	0.2806	31
AnpoNet	85.06 ± 5.07	76.72 ± 8.02	1,916	0.1603	34
CGB-Net	**87.60 ± 4.19**	**83.38 ± 5.51**	9,532	1.72	88

CGB-Net is larger than AnpoNet, but its complexity remains moderate. With 9,532 parameters and 88 kB of memory, it is still compact compared with many deep-learning models. The performance gain is therefore attributable to the architectural combination of CNN, GRU, and BiLSTM rather than uncontrolled model expansion.

3.5.2. Evaluation with random seeds

The random-seed analysis shows that CGB-Net maintains strong performance across different initializations. This is important because the subject-independent protocol can magnify variations caused by data splits and model initialization. Stable results suggest that the model does not depend excessively on a particular random seed.

3.5.3. Confusion matrix analysis

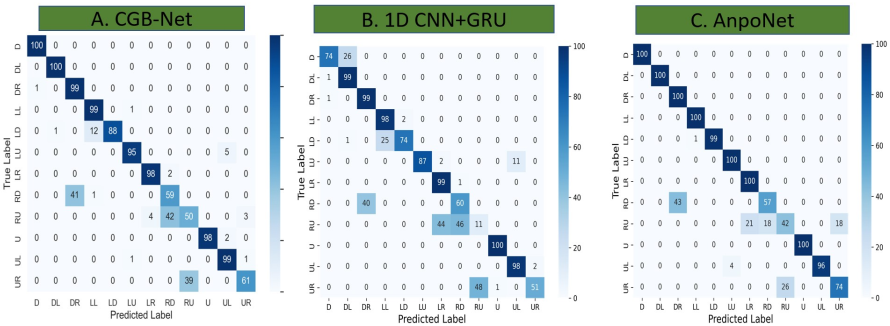

Figure 8: Confusion matrix of three models

The confusion matrices compare CGB-Net, 1D CNN + GRU, and AnpoNet. CGB-Net reduces several off-diagonal errors, especially among postures that are close in orientation.

3.5.4. Training process

The training curves of CGB-Net show a rapid decrease in training loss and a similar trend in validation loss. Training and validation accuracy increase steadily and remain close, with no strong indication of overfitting. The validation accuracy reaches high levels, around 90–95% in the illustrated case.

3.5.5. Influence of sequence length and overlap ratio

Window length strongly affects CGB-Net. In the case study reported in the dissertation, 2.0 seconds gives the best performance, with accuracy 90.77% and F1-score 76.12%. Short windows such as 0.5 or 1.0 seconds do not provide enough temporal context, resulting in lower accuracy and F1-score.

The overlap ratio also matters. An overlap of 0.4 gives the best accuracy in the analyzed case, while lower overlap can reduce temporal continuity and higher overlap can create excessive redundancy. The choice of overlap should therefore balance the need for smooth monitoring, computational efficiency, and independence between samples.

3.6. Conclusion of Chapter 3

Chapter 3 proposes CGB-Net as an accuracy-oriented hybrid architecture for twelve-class sleep-posture recognition. The model integrates 1D CNN, GRU, and BiLSTM to learn richer spatio-temporal features from abdominal accelerometer data.

CONCLUSION AND RECOMMENDATIONS

The dissertation has focused on the problem of fine-grained sleep posture classification from triaxial accelerometer data measured at the abdominal region. Unlike many previous studies that classify only several basic sleep postures, the dissertation addresses the classification of twelve sleep postures, including both basic postures and intermediate postures that differ in body inclination angle. This problem is meaningful for wearable sleep-monitoring systems, especially in applications that require more detailed sleep-posture recognition to support health monitoring and the assessment of users' sleeping behavior. The experiments in the dissertation are designed according to a subject-independent evaluation protocol, in which data from the test subjects do not appear during model training, in order to better reflect practical deployment conditions.

With the objective of improving sleep-posture classification from triaxial accelerometer data, the dissertation has two main contributions.

First, the dissertation proposes AnpoNet for classifying twelve sleep postures from triaxial accelerometer data with a controlled number of parameters. AnpoNet is built by combining the local feature-extraction capability of one-dimensional convolutional neural networks with the temporal-dependency modeling capability of recurrent networks. This approach is intended to exploit the spatio-temporal characteristics of acceleration signals while reducing model complexity so that the model is suitable for wearable sensors or resource-constrained devices. Experimental results show that AnpoNet achieves 94.67% accuracy and a 92.94% F1-score on the test set. In addition to classification performance, the model uses only 1,916 parameters, requires about 34 kB of memory, has a computational cost of 0.1603M FLOPs, an inference time of 0.081 seconds, and an estimated energy consumption of about 1.122 mJ per inference. These results show that AnpoNet achieves a balance between classification accuracy and model compactness, and therefore has value for the future deployment of sleep-posture monitoring systems on wearable devices.

Second, the dissertation proposes CGB-Net to improve accuracy and generalization in the twelve-class sleep posture classification problem. CGB-Net is designed by combining three components: a one-dimensional convolutional neural network, GRU, and BiLSTM. The one-dimensional convolutional network extracts local features from acceleration signals; GRU supports sequential-relation modeling with a reasonable parameter cost; and BiLSTM exploits bidirectional temporal context to strengthen the discrimination between sleep postures with similar characteristics. Experimental results show that CGB-Net achieves an average accuracy of 87.60% and an average F1-score of 83.38% in subject-independent and cross-validation evaluation settings. In the first experimental setting, CGB-Net obtains 89.34% accuracy and 86.24% F1-score, which is about 4.14 percentage points higher than AnpoNet in accuracy and more than 10 percentage points higher in F1-score. This result shows that CGB-Net improves not only the overall correct prediction rate but also the balanced classification ability among sleep-posture classes, especially under inter-subject variation.

In general, the two proposed models play complementary roles. AnpoNet is suitable for compact-model design, with a small number of parameters and low computational cost, oriented toward deployment on resource-constrained devices. Meanwhile, CGB-Net focuses on strengthening spatio-temporal feature representation to improve accuracy and generalization. The obtained results show that using a single triaxial accelerometer attached to the abdomen is feasible for the detailed recognition of twelve sleep postures. This approach is simple, low-cost, less inconvenient for users, and suitable for home sleep-monitoring systems.

Although positive results have been achieved, the dissertation still has several limitations that require further study. The data in the dissertation mainly consist of static sleep postures collected under controlled conditions, and therefore do not fully reflect natural movements, transitional postures, or complex variations during overnight sleep. In addition, edge-device deployment is evaluated only through indirect indicators such as parameter count, computational cost, memory, and inference time; it has not yet been fully verified on embedded hardware or low-power wearable devices in real environments.

In the coming time, future research directions may include expanding the dataset with a larger and more diverse group of subjects, collecting natural overnight sleep data, and considering transitional postures and unstable movements during sleep. In addition, the proposed models can be further optimized through techniques such as quantization, parameter pruning, and model compression, followed by direct deployment and validation on low-power wearable devices to more fully evaluate their practical applicability.

LIST OF THE PUBLICATIONS RELATED TO THE DISSERTATION

[CT1] Hoang-Dieu Vu, Duc-Nghia Tran, Huy-Hieu Pham, Dinh-Dat Pham, Khanh Ly Can, To-Hieu Dao, Duc-Tan Tran. Human sleep position classification using a lightweight model and acceleration data. Sleep and Breathing, Volume 29, 2025, Article 95. (SCIE, Scopus Q1).

[CT2] Hoang-Dieu Vu, Duc-Nghia Tran, Quang-Tu Pham, Ngoc-Linh Nguyen, Duc Tan Tran. CGB-Net: A Novel Convolutional Gated Bidirectional Network for Enhanced Sleep Posture Classification. CMC: Computers, Materials & Continua, vol. 85, no. 2, pp. 2819–2835, 2025. (SCIE, Scopus Q2).

[CT3] Hoang-Dieu Vu, Duc-Nghia Tran, Pham Quang Huy, Duc-Tan Tran. HyPoNet: Fine-Grained Sleep Posture Recognition from a Single Abdominal Accelerometer. REV Journal on Electronics and Communications, vol. 15, no. 3, 2025

[CT4] Hoang-Dieu Vu, Duc-Nghia Tran, Khanh-Ly Can, To-Hieu Dao, Dinh-Dat Pham, Duc-Tan Tran, “Enhancing sleep postures classification by incorporating acceleration sensor and LSTM model,” in 2023 IEEE Statistical Signal Processing Workshop (SSP), 2023, pp. 661–665, doi: 10.1109/SSP53291.2023.10208083. (Scopus)

[CT5] Hoang-Dieu Vu, Duc-Nghia Tran, Quang-Tu Pham, and Duc-Tan Tran, “GA-SPARF: Genetic algorithm-tuned random forest for sleep position monitoring using accelerometer,” Sleep and Breathing, vol. 30, p. 78, 2026, doi: 10.1007/s11325-026-03584-4. (SCIE, Scopus Q1).