Sleep Stage and Apnea Classification from Single-Lead ECG Using Artificial and Spiking Neural Networks

Gergely Imre Biri[1]*, Alexandru Vasilache[1]*, Tao Hu[1], Maximilian Alexander Nicolas Themistocli[1], Sven Nitzsche[1], Jens Juhl[1], Christina Erler[1], Silvester Fuhrhop[2], Wilhelm Stork[3] and Juergen Becker[3]
[1]*FZI Research Center for Information Technology, Karlsruhe, Germany,*
[2]*Corvolution GmbH, Karlsruhe, Germany,*
[3]*Karlsruhe Institute of Technology, Karlsruhe, Germany,*
Email: {biri, vasilache}@fzi.de

2024 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES) | 979-8-3503-8340-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/IECBES61011.2024.10990943

***Abstract*—This study presents a comparative analysis of Artificial Neural Network (ANN) and Spiking Neural Network (SNN) models for sleep stage classification and apnea detection using single-lead electrocardiogram (ECG) signals from the ISRUC-Sleep dataset. Both models leverage a deep convolutional recurrent (DCR) architecture, with the SNN variant using spiking neurons in the convolutional layers. Our findings show that the SNN implementation most often surpasses its ANN counterpart. Specifically, our models achieved state-of-the-art performance for binary (Sleep-Wake), three-class (Wake-NREM-REM), and five-class (Wake-N1-N2-N3-REM) classification with the Macro-F1 scores of 0.9451, 0.8161, 0.6843 respectively, surpassing existing methods in all cases. While our models also demonstrated potential in apnea detection, their performance was constrained by the severe class imbalance present in the dataset.**

***Index Terms*—automatic sleep stage classification, sleep apnea detection, single-lead ECG, spiking neural networks, CNN-RNN, ISRUC-Sleep**

I. Introduction

Sleep disorders, particularly Obstructive Sleep Apnea (OSA), pose a significant global health challenge. An estimated 936 million adults worldwide suffer from mild to severe OSA, which, if left untreated, can result in severe health complications, including cardiovascular, renal, and metabolic disorders [1]. OSA is characterized by recurrent episodes of complete or partial upper airway collapse during sleep, called Obstructive Apnea (OA) and Obstructive Hypopnea (OH), respectively. These apnea events are associated with decreased oxygen saturation or sleep arousal [2]. Consequently, OSA leads to fragmented and non-restorative sleep, significantly impacting overall sleep quality and health.

Based on the Rechtschaffen and Kales standard (R&K) rules [3], the American Academy of Sleep Medicine (AASM) categorizes sleep-wake cycles into awake, non-rapid eye movement (NREM) and rapid eye movement (REM) sleep stages. NREM sleep is further divided into three stages: N1 (drowsiness/transitional sleep), N2 (light sleep), and N3 (deep sleep or slow wave sleep).

Accurately classifying sleep stages and detecting apnea are critical for diagnosing and treating sleep disorders. Polysomnography (PSG), the gold standard for diagnosing sleep disorders, presents significant challenges. A standard PSG measurement requires complex and expensive measurement units to record up to 16 electrode channels. Subsequently, expert clinicians manually classify the collected data, a process that is both time-consuming and potentially subjective. In recent years, approaches based solely on electrocardiography (ECG) have become promising alternatives, since they are convenient to perform and rich in breathing information. Numerous studies have established a strong correlation between sleep and respiratory patterns and have also shown that ECG data offers direct insights into the functioning of the human respiratory and circulatory systems [4]. ECGs have demonstrated particular efficacy in reflecting sleep stages and apnea events among single-lead signals. Specifically, sleep apnea has been associated with irregular cardiac activity [5], and rapid eye movement (REM) sleep has been linked to distinctive changes in respiratory rates [6].

Such ECG-based tests can be classified using machine learning instead of human experts. Various types of Artificial Neural Networks (ANNs) are typically utilized for this purpose, yielding robust classification accuracies. However, recent advances in Spiking Neural Networks (SNNs) make them another viable option. SNNs are a class of neural networks that attempt to closely mimic behavior in biological nervous systems, including the excitation of neurons over time and spike-based communication between neurons [7]. Due to the underlying operating principle of their neurons, SNNs are particularly useful for time-dependent data as measured by an ECG. They also have the potential for significantly higher energy efficiency compared to artificial neural networks if executed on suitable hardware [8], which could be beneficial when deploying such ECG classification in real-world use cases, as they could be integrated directly into the sensor to perform classification in real-time.

Hence, our work investigates sleep stage and apnea classification from single-lead ECG using various neural network architectures, focusing on comparing artificial and spiking neural networks. Our investigations can potentially facilitate the development of more accessible, precise, and efficient sleep monitoring solutions, enabling widespread deployment in clinical and home environments.

II. Related Work

The field of sleep stage detection has seen significant advancements, with numerous studies exploring various signal modalities and machine learning techniques for sleep stage

This research is funded by the German Federal Ministry of Education and Research as part of the project "Privileg", funding no. 16KIS1545.

*These authors contributed equally to this work.

and apnea (Table VI) classification. Traditional methods often utilize PSG data, including electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) signals. In a comprehensive review, Gaiduk *et al.* [9] analyzed the current status of automatic sleep stage scoring methods. The study highlighted that while EEG signals are the most widely used, a combination of PSG signals is also common. The two most frequently used signal sets are EEG alone and EEG+EOG+EMG. This prevalence of EEG-based approaches is due to EEG's key role in sleep stage detection and the crucial information it provides. Besides these three signal types, ECG is used significantly more often than the remaining signals in sleep research [9].

Urtnasan *et al.* [10] proposed a Deep Convolutional Recurrent (DCR) model which obtained 86.4% accuracy for three-class (Wake-NREM-REM) and 74.2% accuracy for five-class (Wake-N1-N2-N3-REM) sleep stage classification with single-lead ECG. Wei *et al.* [11] utilized a Long Short-Term Memory Network (LSTM) approach based on single-lead ECG data, reaching accuracies of 89.84%, 84.07%, and 71.16% for two-class (Sleep-Wake), three-class, and five-class sleep staging, respectively. Their method also demonstrated moderate agreement with clinical analysis, with Cohen's kappa values ranging from 0.52 to 0.58. Pini *et al.* [12] developed an automated heart rate-based algorithm using deep learning, which achieved accuracies of 87.97% for two-class and 80.13% for three-class sleep staging on the Physionet CinC dataset. Earlier work by Wang *et al.* [13] employed decision-tree-based support vector machines (DTB-SVM) on ECG-derived heart rate variability (HRV) features, achieving 73.51% accuracy for three-class sleep staging. Alternative approaches have also been explored. Yücelbaş *et al.* [14] used various feature extraction methods, including Singular Value Decomposition (SVD) and Variational Mode Decomposition (VMD), combined with a Random Forest classifier. They obtained 87.11% accuracy for three-class sleep staging in healthy subjects.

Several deep learning approaches have shown promising performance for obstructive sleep apnea (OSA) detection. Zhang *et al.* [15] proposed a Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM) model that achieved 96.1% accuracy on the Apnea-ECG dataset. Nasifoglu and Erogul [16] explored the use of scalogram and spectrogram representations with various CNN architectures, reaching up to 82.30% accuracy. Mashrur *et al.* [17] developed a scalogram-based CNN (SCNN) that achieved 94.30% accuracy on the Apnea-ECG dataset. Chang *et al.* [18] proposed a 1D CNN model which obtained 87.9% accuracy for per-minute apnea detection.

Recently, spiking neural networks (SNNs) have gained attention for their potential in ECG-based sleep analysis. Tyagi *et al.* [19] proposed an SNN model for sleep apnea detection from single-lead ECG, achieving 94.63% accuracy on the Apnea-ECG database [20]. For sleep stage classification using EEG signals, Jia *et al.* [21] developed a hybrid spiking neural network (HSNN) reaching 76% accuracy on the ISRUC-Sleep [22] dataset, demonstrating the potential of spiking neural networks for sleep analysis tasks.

The literature reveals a trend towards employing advanced machine learning techniques for sleep stage classification and apnea detection, particularly deep learning and neural network architectures. The emerging use of spiking neural networks also shows promise for improving performance in these tasks.

TABLE I
OVERVIEW OF THE ISRUC-SLEEP [22] DATASET

Subgroup	Subjects	Sessions	Subjects	Age
I	100	1	Sleep disorders	20-85
II	8	2	Sleep disorders	26-79
III	10	1	Healthy subjects	30-58

	Mean ± SD	Min	Max
OA (events/h)	1.57 ± 3.89	0	35.66
OH (events/h)	4.23 ± 5.64	0	37.11

III. Methods

Our methodology outlines the design and implementation of Spiking Neural Network (SNN) and Artificial Neural Network (ANN) models. We detail the ISRUC-Sleep dataset, explain our preprocessing methods, describe the ANN architecture, and present the SNN approach.

A. Dataset

We use the open-access ISRUC-Sleep dataset to train, test, and validate our models, allowing comparison with other research. This dataset includes all-night polysomnography (PSG) recordings, approximately eight hours long, acquired non-invasively using the SomnoStar Pro sleep system. The dataset includes three subgroups: (1) 100 subjects with one recording each, (2) 8 subjects with two recordings each, and (3) 10 healthy subjects with one recording each. Two experts scored PSG recordings visually, utilizing frontal, central, and occipital electrodes to detect key sleep features such as K-complexes, spindles, and alpha waves. The agreement between the scores are characterized by Cohen's Kappa (κ) indices of 0.87 ± 0.09 for Subgroup I, 0.82 ± 0.15 for Subgroup II, and 0.9 ± 0.06 for Subgroup III. The annotations are performed according to the AASM criteria and contain five sleep stages, OA, and OH labels. It includes single-channel ECG data recorded at 200 Hz, preprocessed with a notch filter to remove 50 Hz electrical noise. Table I summarizes the dataset, detailing the number of subjects, recording sessions, subject characteristics, and age range, as well as the apnea statistics.

B. Data Preprocessing

Preprocessing steps were applied uniformly for Spiking Neural Networks (SNN) and Artificial Neural Networks (ANN) models. Z-score normalization was applied to reduce the impact of noise and artifacts in the ECG signals. Data points with a z-score exceeding a threshold of 5 were identified as outliers. Linear interpolation replaced these outliers, ensuring a smooth signal without abrupt deviations. The interpolated ECG signals were further processed using the ecg_clean function from the NeuroKit2 library [23]. This step enhanced the signal quality by removing residual noise and artifacts. The cleaned ECG signals were standardized to zero mean and unit variance. The preprocessed data is then split into train (70%), test (15%), and validation (15%) sets. The whole data processing pipeline from the original dataset to the evaluation is presented in Fig. 1.

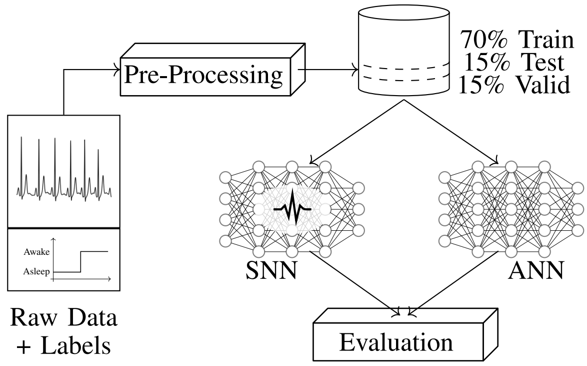

Fig. 1. Data processing pipeline

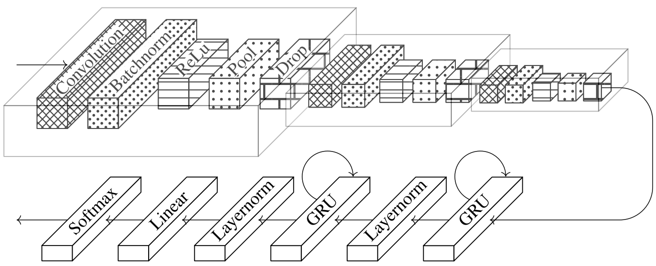

Fig. 2. ANN Architecture

C. Artificial Neural Networks

This section describes the traditional ANN architecture implemented for comparison with our SNN model. In our study, we employed a deep convolutional recurrent (DCR) model to perform automatic sleep stage classification using a single-lead ECG signal, building on the work of Urtnasan *et al.* [10].

We used three convolutional layers in our architecture. The first convolutional layer employs 60 filters with a kernel size of 50. This is followed by a batch normalization layer to standardize the output, a ReLU activation function to introduce non-linearity, a max-pooling layer with a pool size of 2 to reduce the dimensionality of the feature maps, and a dropout layer to prevent overfitting. The second convolutional layer consists of 30 filters with a kernel size of 30. Similar to the first convolutional layer, it is followed by batch normalization, ReLU activation, max-pooling, and dropout layers. The third convolutional layer contains ten filters with a kernel size of 20, followed by batch normalization, ReLU activation, max-pooling, and dropout layers.

For the recurrent part of the model, we included two GRU layers to capture the temporal dependencies of sleep stages. The first GRU layer is followed by a layer normalization step to stabilize the training process and a dropout layer. The second GRU layer is similarly followed by layer normalization and dropout. The specific configuration of hidden nodes and dropout rates is determined by the hyperparameter optimization, as detailed in a later section.

Finally, the output layer is a fully connected multilayer perceptron with a softmax activation function to classify the input signals into the respective sleep stages.

The ANN architecture can be seen in Fig. 2, where different textures represent the convolution, batch normalization, ReLu, and max-pooling layers.

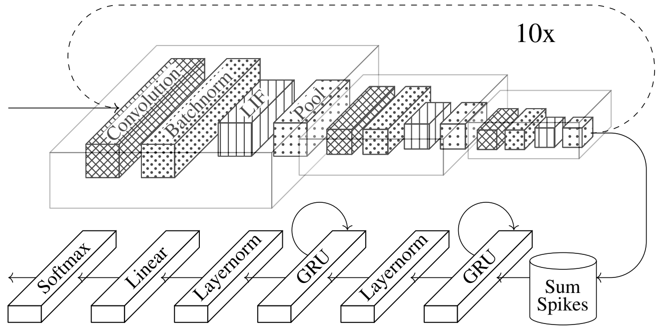

Fig. 3. SNN Architecture

D. Spiking Neural Networks

The SNN architecture is kept as close as possible to that of the ANN, with the most significant differences in the backbone of the network. Here, the ReLU activation functions are replaced by leaky Integrate-and-Fire (LIF) neurons to introduce spikes to the network. The LIF simulates the neuron's membrane potential with a parallel resistor and capacitor circuit. The potential is raised with every incoming spike and decays over time. The neuron will only fire once the threshold is reached. This neuron model is more biologically plausible than an artificial neuron but less computationally complex than the Hodgkin-Huxley neuron model [24].

To train the network faster, the data is not input sequentially, but rather 6000 time steps at once. Therefore, we lose some of the temporal properties of our input. For the LIF neurons, this means that they do not get a sequence of spikes, the neuron has no prior potential before the data is passed through the network, and the output of the LIF is very sparse. To enhance the spike potential and output more spikes, time steps were introduced by letting the data run through the backbone ten times, potentially increasing the memory of the membranes with every run. The dotted line in Fig. 3 represents this procedure. The membrane potential decay rate of the LIF is determined by our experimental hyperparameter optimization performed with Optuna [25]. The threshold of the membrane is learned during the backpropagation step.

Like the ANN, the backbone starts with a convolutional layer containing 60 filters with a kernel size of 50 and is followed by a batch normalization layer, a LIF layer, and a max-pooling layer with a pool size of 2. The second convolutional layer has 30 filters and a kernel size of 30. It is followed by batch normalization, LIF activation, and max-pooling layers. The final convolution layer has ten filters with a kernel size of 20, which is again followed by batch normalization, LIF activation, and max-pooling layers. The output of the final max-pooling layer is appended to a buffer, which is summed across the zeroth dimension once the ten runs are completed to increase the output spikes.

The neck of the network contains two GRU and layer normalization layers to capture temporal dependencies of the different sleep stages. Our hyperparameter optimization determines the hidden layers of the GRUs.

The head of the network consists of a fully connected layer, followed by a softmax function for the classification. Depending on the classification task, the fully connected layer contains either 2, 3, or 5 neurons. The architecture of the SNN can be seen in Fig. 3, where different textures represent

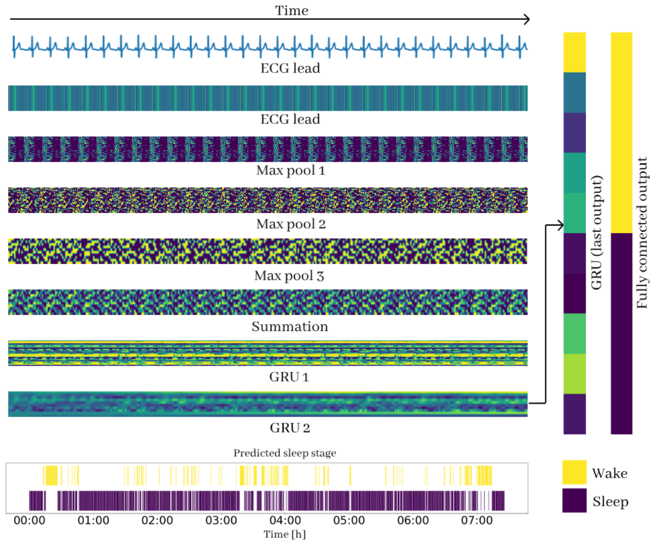

Fig. 4. SNN layer activation and full-night binary sleep stage prediction

the convolution, batch normalization, LIF, and max-pooling layers. Furthermore, to better help visualization, we present the layer activations of the SNN, as well as the output of the model for a full-night sleep recording for the binary sleep-wake classification task (Fig. 4).

IV. Evaluation

A. Hyperparameter Optimization

In our experiments, we utilized the PyTorch framework [26] to implement our models, with PyTorch Lightning [27] serving as a higher-level interface to streamline the training process. By using snnTorch [28] for the LIF layers in our SNN and PyTorch for our other layers, we were able to create comparable networks; PyTorch Lightning was particularly useful in managing the training loop, logging, and ensuring reproducibility across experiments. We employed Optuna [25], a hyperparameter optimization framework, to fine-tune the parameters of both the SNN and ANN models. For both networks, we optimized the learning rate, the amount of hidden layers for the GRUs, and the choice of optimizer. The dropout rate was only optimized in the ANN implementation. For the SNN, some additional parameters needed to be optimized. They were the decay rate and spike gradient for the LIF, the number of repetitions to enhance the spikes, and the type of operation being performed after the simulated time steps. We employed early stopping in the PyTorch Lightning framework to prevent overfitting and pruning to reduce unnecessary computational effort by terminating underperforming Optuna trials early.

For the SNN, we only performed the optimization for the three-class classification problem and used those optimal parameters for the other classification tasks. Optuna yielded the best validation loss for the SNN with the following hyperparameters: For the LIF, the best spike decay rate was 0.5261 with a spike gradient of 25. We also required ten repetitions of our backbone, followed by a summation operation. The GRUs worked best with a hidden size of 23 in the first GRU, followed by a hidden size of 10 in the second GRU. When using the Adam optimizer, the loss rate needed to be set to 1.4857e-05. For the ANN, the computational time was much faster, allowing us to optimize the values for all problems.

TABLE II
SNN and ANN Performance for Binary Sleep-Wake Classification

	SNN	ANN
Precision	0.9109	**0.9770**
Recall	**0.9462**	0.9153
F1	0.9268	**0.9451**
AUC	0.9462	**0.9745**
Accuracy	**0.9469**	0.9169
κ	**0.8538**	0.7751

κ Cohen's Kappa

TABLE III
SNN and ANN Classification Performance for Three-Class (Wake-NREM-REM) Sleep Staging

Model	Scores	Wake	NREM	REM
SNN	**Precision**	**0.7764**	**0.9577**	**0.6105**
	Recall	**0.8915**	**0.8234**	**0.9163**
	F1-Score	**0.8300**	**0.8855**	**0.7327**
	AUC	0.9385	**0.9234**	0.9482
	Accuracy		**0.8496**	
	κ		**0.7258**	
	F1^{W}		**0.8550**	
	F1^{M}		**0.8161**	
ANN	**Precision**	0.7507	0.9429	0.5591
	Recall	0.8910	0.7891	0.8707
	F1	0.8149	0.8591	0.6810
	AUC	**0.9523**	0.9140	**0.9618**
	Accuracy		0.8213	
	κ		0.6782	
	F1^{W}		0.8281	
	F1^{M}		0.7850	

κ Cohen's Kappa, ^{W}F1 (Weighted), ^{M}F1 (Macro)

They are listed in the following order: binary, three-class, five-class, and apnea classification. The first GRU required a hidden size of 49, 22, 25, and 36. For the second GRU, the values were 12, 29, 12, and 36, respectively. The dropout rate was found to be optimal at 0.1434, 0.3600, 0.1562, and 0.1112, respectively. Optuna found that all models performed best with the Adam optimizer and required a learning rate of 1.837e-03, 1.037e-03, 9.655e-04, and 3.089e-04, respectively.

B. Results

We present the outcomes of our experiments, which were conducted using Spiking Neural Networks (SNNs) and Artificial Neural Networks (ANNs) across four distinct classification tasks: binary classification (Table II), three-class classification (Table III), five-class classification (Table IV), and obstructive sleep apnea classification (Table V), all using a single-channel ECG signal. Across all tasks, the performance of the models is evaluated using a set of standard metrics. These include precision, recall, F1-Score, the area under the curve (AUC), accuracy, and Cohen's Kappa.

V. Discussion

Our ANN and SNN models perform similarly in binary classification (Table II), with the SNN achieving higher recall and the ANN better precision. Both models exceed reported results in the literature (Table VI), with the SNN achieving the highest accuracy (0.9469 vs. 0.8948) and the ANN the best Macro-F1 score (0.9451 vs. 0.7940) for Sleep-Wake classification.

In the three-class task (Table III), the SNN outperforms the ANN. Our SNN model has competitive accuracy (0.8496 vs.

0.8711) compared to the state-of-the-art and the best-reported Macro-F1 in the literature (0.8161 vs. 0.7867).

For five-class classification (Table IV), the SNN outperforms the ANN and sets a new state-of-the-art Macro-F1 score (0.6843 vs. 0.6160).

In apnea classification (Table V), the SNN outperforms the ANN, but lags behind the state-of-the-art (Table VI), with lower accuracy (0.8955 vs. 0.9900) and Macro-F1 (0.6832 vs. 0.9285). However, direct comparison is difficult due to differing datasets, with our evaluation performed on full-night recordings where apnea events are only 5% of the total (4602 vs. 89520).

VI. Conclusion and Future Work

Our proposed SNN models demonstrate strong performance across all classification tasks, most often surpassing their ANN counterparts. While both architectures achieved competitive results in binary sleep-wake classification, the SNN models consistently excelled in multi-class sleep-stage scenarios. Notably, our models have established new state-of-the-art Macro-F1 scores for binary (0.9451), three-class (0.8161), and five-class (0.6843) sleep stage classification.

It is important to argue that Macro-F1 is a more reliable metric than accuracy for sleep stage classification as it better handles the inherently uneven class distribution.

Although our apnea classification model performed well, it did not reach state-of-the-art levels, likely due to the dataset's 5% prevalence of apnea events.

Future work will focus on several key areas. We plan to improve the apnea models by training and evaluating them on specialized datasets. Additionally, we will explore embedded implementations of Spiking Neural Networks (SNNs), optimizing them for low power consumption and real-time processing to show their potential for wearable devices. Finally, we aim to enhance sleep stage classification accuracy by incorporating temporal dependencies between samples, using techniques like auto-regressive models to leverage information from previous sleep stages.

TABLE IV
SNN and ANN Classification Performance for Five-Class (Wake-N1-N2-N3-REM) Sleep Staging

Model	Scores	Wake	N1	N2	N3	REM
SNN	**Precision**	**0.8839**	**0.4345**	0.6991	**0.7645**	**0.6363**
	Recall	0.7532	**0.6243**	**0.6212**	0.6993	**0.7972**
	F1	0.8133	**0.5124**	**0.6578**	**0.7305**	**0.7077**
	AUC	0.9395	0.8098	0.8224	0.9015	0.9187
	Accuracy			**0.6877**		
	κ			**0.6019**		
	F1$^{\mathrm{W}}$			**0.6943**		
	F1$^{\mathrm{M}}$			**0.6843**		
ANN	**Precision**	0.8149	0.3881	**0.7164**	0.6185	0.5800
	Recall	**0.8319**	0.4236	0.4147	**0.8424**	0.7907
	F1	**0.8233**	0.4051	0.5253	0.7133	0.6691
	AUC	**0.9628**	**0.8125**	**0.8283**	**0.9319**	**0.9486**
	Accuracy			0.6456		
	κ			0.5852		
	F1$^{\mathrm{W}}$			0.6354		
	F1$^{\mathrm{M}}$			0.6272		

κ Cohen's Kappa, $^{\mathrm{W}}$F1 (Weighted), $^{\mathrm{M}}$F1 (Macro)

TABLE V
Per-segment ANN and SNN Classification Performance for Obstructive Apnea (OA) and Obstructive Hypopnea (OH)

Model	Scores	None	OA	OH
SNN	**Precision**	**0.9927**	**0.4668**	**0.3505**
	Recall	**0.8981**	**0.8797**	**0.8518**
	F1	**0.9430**	**0.6100**	**0.4966**
	AUC	**0.9349**	0.9369	**0.9156**
	Accuracy		**0.8955**	
	κ		**0.5108**	
	F1$^{\mathrm{W}}$		**0.9140**	
	F1$^{\mathrm{M}}$		**0.6832**	
ANN	**Precision**	0.9859	0.1626	0.1951
	Recall	0.8881	0.5686	0.5768
	F1	0.9345	0.2529	0.2916
	AUC	0.9128	**0.9548**	0.8920
	Accuracy		0.8739	
	κ		0.2805	
	F1$^{\mathrm{W}}$		0.9049	
	F1$^{\mathrm{M}}$		0.4930	

κ Cohen's Kappa, $^{\mathrm{W}}$F1 (Weighted), $^{\mathrm{M}}$F1 (Macro)

TABLE VI
Review of ECG-based Sleep Stage and Apnea Classification Methods

Author	Signal	Class	Method	Accuracy	F1$^{\mathrm{M}}$
[11]	ECG^{s}	2-class	ANN	0.8948	0.7595
[12]	ECG/HR	2-class	ANN	0.8797	0.7940
Our work	ECG^{s}	2 class	SNN	**0.9469**	0.9268
Our work	ECG^{s}	2 class	ANN	0.9169	**0.9451**
[12]	ECG/HR	3-class	ANN	0.8013	0.7386
[13]	ECG/HRV	3-class	SVM	0.7351	-
[11]	ECG^{s}	3-class	ANN	0.8407	0.7174
[14]	ECG^{s}	3-class	RF	**0.8711**	-
[10]	ECG^{s}	3-class	ANN	0.8640	0.7867
Our work	ECG^{s}	3-class	SNN	0.8496	**0.8161**
Our work	ECG^{s}	3-class	ANN	0.8213	0.7850
[10]	ECG^{s}	5-class	ANN	**0.7420**	0.6160
[11]	ECG^{s}	5-class	ANN	0.7116	0.5766
Our work	ECG^{s}	5-class	SNN	0.6877	**0.6843**
Our work	ECG^{s}	5-class	ANN	0.6456	0.6272
[18]	ECG^{s}	SA	ANN	0.8790	-
[17]	ECG^{s}	SA	ANN	0.8186	0.6963
[16]	ECG	SA	ANN	0.8230	-
[29]	ECG	SA	ANN	0.8558	0.8467
[15]	ECG^{s}	SA	ANN	0.9610	-
[30]	ECG	SA	ANN	**0.9900**	-
[19]	ECG^{s}	SA	SNN	0.9463	**0.9285**
Our work	ECG^{s}	SA	SNN	0.8955	0.6832
Our work	ECG^{s}	SA	ANN	0.8739	0.4930

^{s}Single-lead ECG, $^{\mathrm{M}}$F1 (Macro)

References

[1] V. Ling, C. Wu, and S. Stiles, "Sleep apnea statistics and facts you should know," *NCOA Adviser*, October 2023. [Online]. Available: https://www.ncoa.org/adviser/sleep/sleep-apnea-statistics/

[2] J. Slowik, A. Sankari, and J. Collen, "Obstructive sleep apnea," *StatPearls*, Mar 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK459252/

[3] T. Hori, Y. Sugita, E. Koga, S. Shirakawa, K. Inoue, S. Uchida, H. Kuwahara, M. Kousaka, T. Kobayashi, Y. Tsuji, M. Terashima, K. Fukuda, N. Fukuda, and S. C. C. of the Japanese Society of Sleep Research, "Proposed supplements and amendments to 'a manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the rechtschaffen & kales (1968) standard," *Psychiatry Clin. Neurosci.*, vol. 55, no. 3, pp. 305–310, 2001.

[4] C. Sun, S. Hong, J. Wang, X. Dong, F. Han, and H. Li, "A systematic review of deep learning methods for modeling electrocardiograms during sleep," *Physiological Measurement*, vol. 43, no. 8, p. 08TR02, 2022.

[5] M. K. Moridani, M. Heydar, and S. S. Jabbari Behnam, "A reliable algorithm based on combination of emg, ecg and eeg signals for sleep apnea detection: (a reliable algorithm for sleep apnea detection)," in *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, 2019.

[6] K. Kido, T. Tamura, N. Ono, M. D. Altaf-Ul-Amin, M. Sekine, S. Kanaya, and M. Huang, "A novel CNN-based framework for classification of signal quality and sleep position from a capacitive ECG measurement," *Sensors (Basel, Switzerland)*, vol. 19, no. 7, p. 1731, 2019.

[7] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 10, no. 9, pp. 1659–1671, 1997.

[8] P. Blouw and et al., "Benchmarking keyword spotting efficiency on neuromorphic hardware," in *Proceedings of the 7th annual neuro-inspired computational elements workshop*, 2019.

[9] M. Gaiduk, Á. Serrano Alarcón, R. Seepold, and N. Martínez Madrid, "Current status and prospects of automatic sleep stages scoring," *Biomedical engineering letters*, vol. 13, no. 3, pp. 247–272, 2023.

[10] E. Urtnasan, J.-U. Park, E. Y. Joo, and K.-J. Lee, "Deep convolutional recurrent model for automatic scoring sleep stages based on single-lead ECG signal," *Diagnostics (Basel, Switzerland)*, vol. 12, no. 5, p. 1235, 2022.

[11] Y. Wei, X. Qi, H. Wang, Z. Liu, G. Wang, and X. Yan, "A multi-class automatic sleep staging method based on long short-term memory network using single-lead electrocardiogram signals," *IEEE Access: Practical Innovations, Open Solutions*, vol. 7, pp. 85 959–85 970, 2019.

[12] N. Pini, J. L. Ong, G. Yilmaz, N. I. Y. N. Chee, Z. Siting, A. Awasthi, S. Biju, K. Kishan, A. Patanaik, W. P. Fifer, and M. Lucchini, "An automated heart rate-based algorithm for sleep stage classification: Validation using conventional polysomnography and an innovative wearable electrocardiogram device," *Frontiers in Neuroscience*, vol. 16, p. 974192, 2022.

[13] J.-S. Wang, G.-R. Shih, and W.-C. Chiang, "Sleep stage classification of sleep apnea patients using decision-tree-based support vector machines based on ECG parameters," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, 2012.

[14] Ş. Yücelbaş, C. Yücelbaş, G. Tezel, S. Özşen, and Ş. Yosunkaya, "Automatic sleep staging based on svd, vmd, hht and morphological features of single-lead ecg signal," *Expert Systems with Applications*, vol. 102, pp. 193–206, 2018.

[15] J. Zhang, Z. Tang, J. Gao, L. Lin, Z. Liu, H. Wu, F. Liu, and R. Yao, "Automatic detection of obstructive sleep apnea events using a deep CNN-LSTM model," *Computational Intelligence and Neuroscience*, vol. 2021, p. 5594733, 2021.

[16] H. Nasifoglu and O. Erogul, "Obstructive sleep apnea prediction from electrocardiogram scalograms and spectrograms using convolutional neural networks," *Physiological Measurement*, vol. 42, no. 6, p. 065010, 2021.

[17] F. R. Mashrur, M. S. Islam, D. K. Saha, S. M. R. Islam, and M. A. Moni, "SCNN: Scalogram-based convolutional neural network to detect obstructive sleep apnea using single-lead electrocardiogram signals," *Computers in Biology and Medicine*, vol. 134, p. 104532, 2021.

[18] H.-Y. Chang, C.-Y. Yeh, C.-T. Lee, and C.-C. Lin, "A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram," *Sensors (Basel, Switzerland)*, vol. 20, no. 15, p. 4157, 2020.

[19] P. K. Tyagi and D. Agrawal, "Automatic detection of sleep apnea from a single-lead ECG signal based on spiking neural network model," *Computers in Biology and Medicine*, vol. 179, p. 108877, 2024.

[20] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ECG database," in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, September 2000, pp. 255–258.

[21] Z. Jia, J. Ji, X. Zhou, and Y. Zhou, "Hybrid spiking neural network for sleep electroencephalogram signals," *Science China Information Sciences*, vol. 65, no. 4, 2022.

[22] S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes, "ISRUC-sleep: A comprehensive public dataset for sleep researchers," *Computer Methods and Programs in Biomedicine*, vol. 124, pp. 180–192, 2016.

[23] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, feb 2021. [Online]. Available: https://doi.org/10.3758%2Fs13428-020-01516-y

[24] S. Valadez-Godínez, H. Sossa, and R. Santiago-Montero, "The step size impact on the computational cost of spiking neuron simulation," in *2017 Computing Conference*. IEEE, 2017, pp. 722–728.

[25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[26] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[27] W. Falcon and The PyTorch Lightning team, "PyTorch Lightning," Mar. 2019. [Online]. Available: https://github.com/Lightning-AI/lightning

[28] J. K. Eshraghian, M. Ward, E. Neftci, X. Wang, G. Lenz, G. Dwivedi, M. Bennamoun, D. S. Jeong, and W. D. Lu, "Training spiking neural networks using lessons from deep learning," *Proceedings of the IEEE*, vol. 111, no. 9, pp. 1016–1054, 2023.

[29] D. Mukherjee, K. Dhar, F. Schwenker, and R. Sarkar, "Ensemble of deep learning models for sleep apnea detection: An experimental study," *Sensors (Basel, Switzerland)*, vol. 21, no. 16, p. 5425, 2021.

[30] U. Erdenebayar, Y. J. Kim, J.-U. Park, E. Y. Joo, and K.-J. Lee, "Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram," *Computer Methods and Programs in Biomedicine*, vol. 180, p. 105001, 2019.