Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

Efficient ECG-Based Sleep Apnea Detection Using CNN-GRU and Sparse Autoencoder

Ramadhian Eka Putra[*1], Sani Muhamad Isa[2]

[1,2]Department of Computer Science, Bina Nusantara University, Indonesia

Email: [1]ramadhian.putra@binus.ac.id

Received : Jul 27, 2025; Revised : Jul 31, 2025; Accepted : Oct 2, 2025; Published : Feb 15, 2026

Abstract

Sleep apnea is a serious and common breathing disorder that occurs during sleep, characterized by repeated pauses in breathing that can increase the risk of hypertension, heart disease, and stroke. Early detection of sleep apnea is crucial, but conventional methods, such as polysomnography, are expensive, complex, and inefficient for mass screening. Therefore, an automated system based on physiological signals such as an electrocardiogram (ECG) is needed for a more practical and efficient approach. This study proposes a sleep apnea classification model utilizing a combination of 1D Convolutional Sparse Autoencoder (1DCSAE), Convolutional Neural Network (CNN), and Gated Recurrent Unit (GRU) architectures, referred to as the SAE-DEEP model. This method is designed to automatically extract features while minimizing the need for preprocessing. Four testing scenarios were conducted to evaluate the impact of signal reconstruction and preprocessing on classification performance. Experimental results show that the CNN-GRU model with signal reconstruction using 1DCSAE achieves an accuracy of 89.8%, a sensitivity of 90.1%, and a specificity of 89.2%, demonstrating balanced and stable classification performance. Additionally, this model was proven to work effectively without complex preprocessing steps, making it a potential solution for efficient sleep apnea detection systems. These findings could contribute to the development of more straightforward, reliable, and clinically viable ECG-based classification systems, as well as wearable devices. In doing so, the proposed model addresses a critical gap in sleep apnea screening, underscoring the urgent need for accessible and cost-effective diagnostic tools.

Keywords : *Classification, CNN, ECG, GRU, Sleep Apnea, Sparse Autoencoder*

1. INTRODUCTION

Sleep is a vital activity that accounts for one-third of human life and plays a significant role in maintaining physical health and quality of life [1]. On average, humans require 7 to 8 hours of sleep per day [2]. When sleep quality is disrupted, the body can experience serious negative effects, both physically and mentally. To date, approximately 84 types of sleep disorders have been identified, and one of the most common and potentially dangerous is sleep apnea [3], [4], [5]. Sleep apnea is a breathing disorder that occurs during sleep, characterized by complete or partial cessation of breathing due to blockage in the upper airway [6], [7]. This disorder is known to have a strong correlation with age [8] and is generally divided into three main categories: (i) Obstructive Sleep Apnea (OSA), (ii) Central Sleep Apnea (CSA), and (iii) Mixed Sleep Apnea (MSA) [9].

The prevalence of sleep apnea continues to rise worldwide. According to a 2007 WHO report, over 100 million people suffer from sleep apnea. This number surged to 1 billion by 2018, with 425 million of them aged 30–69 and experiencing moderate to severe OSA [10]. The incidence of OSA varies between countries, with rates of 8.8% in China, 26% in Brazil, and 49.7% of men and 23.4% of women experiencing sleep-related breathing disorders in Switzerland [11]. In the United States, data indicate that OSA affects 3% of women and 10% of men aged 30–49 years, increasing to 9% of women and 17% of men aged 50–70 years [12]. These facts underscore the urgency of early detection and effective management of sleep apnea globally.

As the prevalence of OSA increases, various researchers have developed several automated approaches to detect sleep apnea from physiological signals, such as EEG, ECG, and SpO2 [13], [14]. Among these, the ECG signal is considered the most promising due to its high precision, comfort of use, and low cost [15], [16], [17], [18], [19]. ECG signals can record heart activity through electrodes placed on the skin and have proven effective in identifying symptoms of sleep apnea [20], [21]. Therefore, ECG signal analysis has become the primary focus in sleep apnea detection research, ranging from conventional machine learning approaches to deep learning.

Initially, sleep apnea detection relied heavily on classical machine learning algorithms [1]. However, due to limitations in extracting features from complex physiological signals, research shifted toward deep learning models [1], [22]. These models have the advantage of automatically learning relevant features without requiring complex manual extraction [23]. Various deep learning models have been applied for ECG-based sleep apnea detection. One such study was conducted by Nasifoglu, who used a CNN with scalogram and spectrogram inputs, achieving an accuracy of 82.30%, sensitivity of 83.22%, and specificity of 82.27% [24]. Another study tested several architectures, including CNN, CNN-LSTM, and CNN-GRU, and found that CNN-LSTM provided the best performance with an accuracy of 89.11% [25]. On the other hand, Wang et al. used a modified version of LeNet-5 to analyze single ECG signals, achieving an accuracy of 87.6% [26].

Bahrami also compared various models such as CNN, LSTM, BiLSTM, GRU, and CNN-LSTM, with CNN-LSTM being the most superior (accuracy of 80.67%) [27]. Another study by Feng et al. employed a combined approach using a Frequential Stacked Sparse Autoencoder (FSSAE) and a Time-Dependent Cost-Sensitive (TDCS) classifier, with FSSAE for feature extraction and TDCS—a combination of a Hidden Markov Model (HMM) and MetaCost—as the classifier. Using the Apnea-ECG dataset (comprising 70 patients), the results yielded an accuracy of 85.1%, a sensitivity of 86.2%, and a specificity of 84.4% [17].

Although deep learning models are promising, their success heavily depends on the quantity and quality of training data [28], [29]. Although previous studies have applied CNN, LSTM, or hybrid models, challenges remain in balancing sensitivity and specificity while reducing dependence on complex preprocessing. Furthermore, the collection of ECG-based sleep apnea data is still limited, which can lead to overfitting and poor model generalization for new patients. In addition, ECG signals are highly susceptible to noise and artifacts caused by movement during sleep or interference from medical devices [30], [31]. Without adequate preprocessing, detection quality can deteriorate significantly due to the difficulty of distinguishing valid apnea signals from external disturbances.

To address these challenges, this study introduces a new hybrid model, SAE-DEEP, which uses 1DCSAE to enrich ECG signal features and combines them with CNN and GRU to balance sensitivity and specificity without requiring complex pre-processing. Furthermore, the CNN-GRU combination was chosen because it can capture the spatial and temporal patterns of ECG signals simultaneously [32]. GRU is known to be effective in processing long-term sequential data such as time series signals [33], [34], [35]. This approach is designed to overcome data limitations and noise interference in ECG signals, which have often been obstacles in previous studies. Thus, this approach can improve model generalization across different patient conditions while reducing dependence on complex pre-processing steps. This makes the method more suitable for broad and real-time applications in ECG-based sleep apnea monitoring systems.

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

2. METHOD

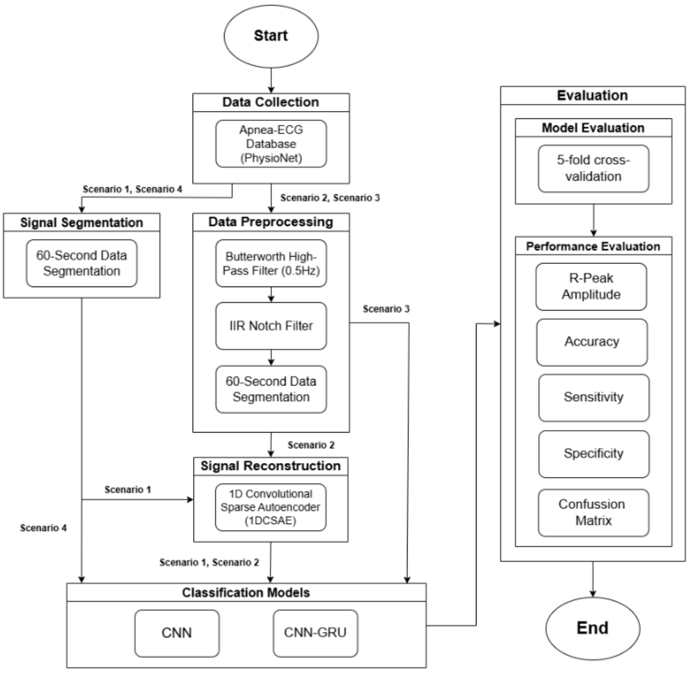
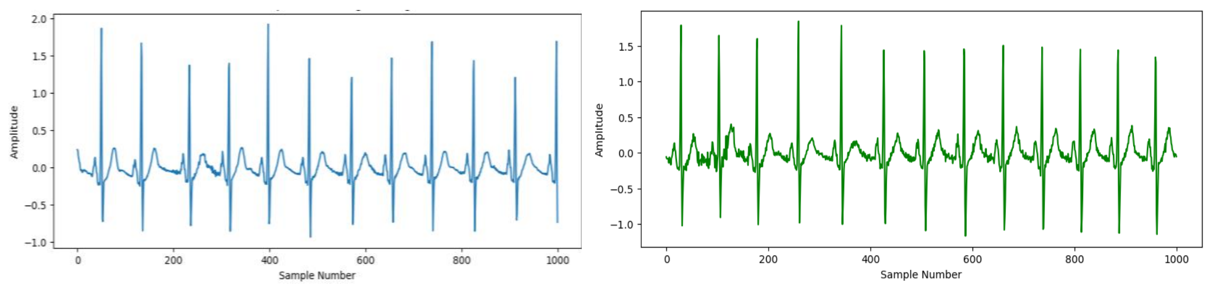


Figure 1. Research Process Flow Diagram

This study aims to evaluate and compare the performance of two deep learning architectures, namely CNN and CNN-GRU, in detecting sleep apnea based on ECG signals. Each method is systematically implemented through the stages of preprocessing, signal segmentation, reconstruction using 1DCSAE, classification, and evaluation. Four testing scenarios were conducted to assess the impact of reconstruction and preprocessing on model performance. This study highlights the advantages of each approach in capturing spatial and temporal patterns, as well as the effectiveness of 1DCSAE in reducing noise without losing important features. The workflow of the methodology is shown in Figure 1, which illustrates the sequential process of each approach for comparative analysis purposes.

2.1. Dataset

This study utilizes the PhysioNet Apnea-ECG dataset, a widely used benchmark for ECG-based OSA detection. The dataset contains 70 overnight single-lead ECG recordings (approximately 7–10 hours each) sampled at 100 Hz. Each recording is annotated by sleep experts on a per-minute basis – if an apnea event occurs in a given minute, that 60-second ECG segment is labeled *"apnea"* (positive), otherwise *"normal"* (negative).



(a) (b)

Figure 2. Input ECG dari dataset Apnea-ECG Database. (a) Normal (b) Apnea

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

As illustrated in Figure 2, the dataset provides two typical patterns of ECG input, where panel (a) depicts a normal signal with relatively stable waveforms and panel (b) shows an apneic signal marked by irregular morphology. This distinction forms the basis for labeling each minute segment and ensures that the dataset can clearly separate normal and apneic events. In total, 34,313 one-minute segments (each 6000 samples) were extracted, comprising 21,249 apnea segments and 13,064 non-apnea segments. The recordings are further categorized by severity using the Apnea-Hypopnea Index (AHI) as follows:

1) Class A (Severe OSA): AHI ≥ 10 (recording contains ≥100 apnea segments)
2) Class B (Mild OSA): 5 ≤ AHI < 10 (recording contains 5–99 apnea segments)
3) Class C (Normal): AHI < 5 (no or very few apnea segments)

For training and evaluation purposes, all segments were randomly divided into two parts, training and testing data, with a ratio of 50:50 to ensure a balanced proportion of apnea and non-apnea segments in both subsets. This division strategy is commonly applied in recent studies to avoid bias toward specific subsets [36]. The Apnea-ECG dataset was chosen due to its availability and reliability as a gold standard in ECG-based apnea research. It provides a comprehensive range of normal and apneic heart patterns, with expert-labeled ground truth for each minute segment, making it ideal for training deep learning models. Indeed, numerous recent studies have adopted this dataset as a benchmark; for example, using all 70 recordings to evaluate various machine learning and deep learning classifiers for OSA detection [19].

2.2. Pre-Processing

Raw ECG signals often experience baseline wander, which can interfere with analysis [37], [38]. To overcome this, a Butterworth high-pass filter with a cutoff frequency of 0.5 Hz is used to eliminate baseline shifts. In addition, ECG signals also contain noise interference from 60 Hz powerline interference, which can reduce signal quality and make it difficult for deep learning models to extract relevant features. Therefore, an IIR notch filter at 60 Hz is applied to eliminate this interference.

In all scenarios, ECG data is segmented into one-minute intervals without overlap. Each segment captures the heart activity pattern within a one-minute time frame, corresponding to the apnea event labeling interval, ensuring more accurate analysis.

2.3. Signal Reconstruction

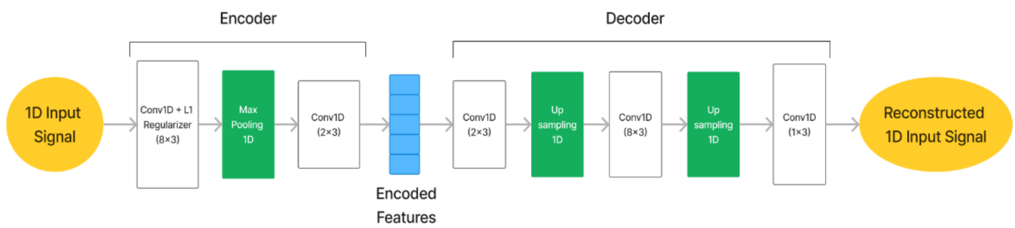


Figure 3. 1D Convolutional Sparse Autoencoder Architecture

Signal reconstruction in this study employs the 1DCSAE approach, which helps reduce disturbances such as noise and baseline wander in ECG signals. As shown in Figure 3, this process utilizes an encoder-decoder architecture comprising Conv1D layers to extract key features and reconstruct signals with improved characteristics. The use of 1DCSAE also aims to minimize the need for complex manual preprocessing by allowing the model to automatically learn relevant features from the raw signal [39]. Although not all distortions such as baseline wander can be eliminated, the visual results show that signal reconstruction produces a smoother and more stable structure. This reinforces the effectiveness of the reconstruction stage as a foundational initial step before the classification process

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

is performed by the CNN or CNN-GRU model, while also reducing the potential loss of important information due to external noise.

2.4. Classification

Table 1. CNN-GRU Architecture Model

Layer	Kernel /Pool Size/Units	Filters	Activation Function	Regularizer	Output Size
Batch Normalization	-	-	-	-	(None, 6000, 1)
Conv1D_1	100	2	ReLU	-	(None, 2951, 2)
Max_Pooling_1	2	2	-	-	(None, 2950, 2)
Conv1D_2	50	60	ReLU	-	(None, 2901, 60)
Max_Pooling_2	2	-	-	-	(None, 1450, 60)
Conv1D_3	30	30	ReLU	-	(None, 1421, 30)
Dropout 25%	3	-	-	-	(None, 1421, 30)
Max_Pooling_3	2	-	-	-	(None, 710, 30)
Conv1D_4	10	20	ReLU		(None, 701, 20)
Max_Pooling_4	2	-	-		(None, 350, 20)
Dropout 25%	-	-	-	-	(None, 350, 20)
Conv1D_5	10	10	ReLU	-	(None, 341, 10)
GRU	128	-	Sigmoid		(128, 1)
Batch Normalization	-	-	-	-	(128, 1)
Dropout 25%	-	-	-	-	(128, 1)
Flatten	-	-	-	-	(128)
Dense_1	-	-	-	-	(100)
Dense_2	-	-	-	-	(30)
Dense_3	-	-	-	-	(10)
Dropout 25%	-	-	-	-	(10)
Dense_4	-	-	-	Kernel = L2 (0.01) Bias = L2 (0.01)	(2)

At the classification stage, two deep learning architectures were applied, namely CNN and CNN-GRU, with the integration of a signal reconstruction stage using 1DCSAE. Before entering the classification models, each ECG signal segment was reconstructed using 1DCSAE to reduce noise artifacts and baseline drift in the signals. This reconstruction process employed a convolutional (Conv1D) encoder–decoder architecture to extract essential features and reconstruct signals with cleaner characteristics. The use of such a convolutional autoencoder is effective for signal denoising and for automatically capturing key morphological patterns [40]. The reconstructed ECG signals, which appear smoother and more stable, are expected to facilitate the subsequent classification stage.

The CNN model was utilized to extract spatial features from the signals through Conv1D operations, enabling the recognition of local patterns such as the QRS complex. Furthermore, the CNN-GRU model was developed by adding a GRU layer after the Conv1D layers to capture temporal patterns from the sequential heart signal data. This combination of CNN and GRU allows the model to more holistically recognize frequency variations and signal structures associated with apnea events [41]. GRU was chosen over LSTM because of its simpler structure with fewer parameters, making the training process more efficient without a significant decrease in detection accuracy [41]. By combining the CNN's capability in local feature extraction with the efficiency of GRU in modeling long-term dependencies, the CNN-GRU architecture can deliver improved apnea detection performance and higher generalization across various patient conditions [41].

In its implementation, the CNN-GRU architecture consists of several Conv1D and pooling layers for feature extraction, followed by a single GRU layer with 128 units for sequence modeling, and finally concluded with dense layers using the softmax activation function to classify the signals into two classes (apnea vs. normal). The complete structure of the CNN-GRU architecture is presented in Table 1.

As shown in the architecture model table in Table 1, Conv1D is used to extract local features from the ECG signal. The convolution process is formulated in equation (1).

$$x_j^{(l)} = f\left(\sum_{i \in M_j} x_i^{(l-1)} * w_{ij}^{(l)} + b_j^{(l)}\right) \qquad (1)$$

Where $x_i^{(l-1)}$ is the input from the previous layer, $w_{ij}^{(l)}$ is the kernel, and f is an activation function such as ReLU. In the output layer, the model uses the softmax function to generate the probabilities of the two classes shown in equation (2).

$$y_k = \frac{e^{a_k}}{\sum_{i=1}^{n} e^{a_i}} \qquad (2)$$

This function ensures that the total probability of both classes (apnea and non-apnea) adds up to 1.

2.5. Testing Scenario

The testing scenarios in this study were conducted through four experimental scenarios to evaluate the effect of signal reconstruction and preprocessing on classification performance. As shown in the workflow of each scenario in Figure 1, in Scenario 1, ECG signal data was segmented into 60-second intervals without overlap, then reconstructed using a Sparse Autoencoder (SAE) before being classified using a CNN and CNN-GRU model. Scenario 2 has a similar workflow but includes a preprocessing step before segmentation and reconstruction. In contrast, in Scenarios 3 and 4, the original ECG signal is used without reconstruction by the SAE. The difference lies in Scenario 3, which involves preprocessing, while Scenario 4 does not.

All scenarios were tested using a *5-fold cross-validation* scheme to ensure stable and representative results. Performance evaluation was conducted by calculating accuracy, sensitivity, and

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

specificity values, as well as analyzing the number of *True Positives*, *True Negatives*, *False Positives*, and *False Negatives* from the classification results. The best model was selected based on its ability to balance the classification of apnea and non-apnea segments, as well as consistent performance across various input configurations.

2.6. Evaluation

The performance evaluation of the model was conducted to assess how accurately the classification system could detect apnea events. As previously described, a 5-fold cross-validation scheme was applied to obtain a more generalized estimate of performance. In this validation method, the dataset is proportionally divided into five folds. The model is trained on four folds and tested on the remaining fold, and this process is repeated until each fold has served as the test set once. This approach ensures that every sample in the dataset is both validated and tested, thereby producing an evaluation that reflects the model's average performance in a more robust manner [42]. To measure performance, three commonly used metrics in medical detection systems were employed: accuracy, sensitivity, and specificity. These metrics were calculated based on the values of *True Positive (TP), True Negative (TN), False Positive (FP),* and *False Negative (FN)*, which were derived from comparing the model's predictions against the ground truth labels in the test data. The formal definitions of these evaluation metrics are as follows:

$$Sensitivity = \frac{TP}{TP+FN} \tag{3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Meanwhile, the confusion matrix used to illustrate the matrix structure in the model evaluation process is shown in Table 2.

Table 2. Confusion Matrix

	Predicted Apnea	Predicted Non-Apnea
Actual Apnea	TP	FN
Actual Non-Apnea	FP	TN

These three metrics were selected because they are widely applied in medical detection evaluations and complement one another in reflecting model performance. Accuracy provides an overall performance measure, but it can be biased when dealing with imbalanced datasets. Sensitivity and specificity, on the other hand, specifically emphasize the model's strength in correctly identifying positive apnea events and avoiding false alarms. In the context of OSA detection, high sensitivity ensures that the majority of apnea episodes are detected (which is crucial for patient safety), while high specificity ensures that normal signals are rarely misclassified as apnea (which is important for reducing *false alarms*). Through the 5-fold cross-validation, stable average values of accuracy, sensitivity, and specificity were obtained for each experimental scenario, thereby confirming the consistency and reliability of the model's performance.

3. RESULT

The results presented encompass the outcomes of preprocessing, including noise reduction and baseline stabilization, as well as signal reconstruction, which is the primary enhancement step before

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

classification. Furthermore, classification results were obtained from two deep learning models, accompanied by a comprehensive evaluation of signal quality and apnea detection performance.

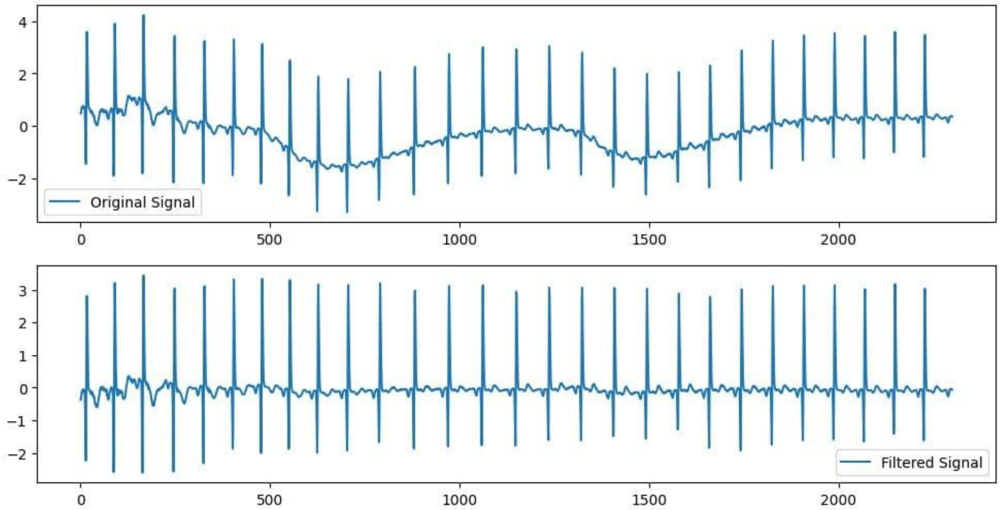


Figure 4. Results of Noise Reduction and Baseline Stabilization

3.1. Preprocessing Result

The preprocessing results illustrated in Figure 4 demonstrate a significant transformation between the raw ECG signal (original signal) and the signal after filtering (filtered signal). In the raw signal, several disturbances are clearly visible, most notably *baseline wander*, which manifests as a slow drift of the baseline, as well as unstable amplitude fluctuations caused by low-frequency voltage shifts. Such disturbances are typically introduced by respiratory activity or body movements during signal acquisition, and they can obscure the main morphological components of the ECG, particularly the R-peak, which is a critical feature in sleep apnea detection. In addition, the raw signal is also affected by 60 Hz powerline interference, producing periodic noise that further degrades signal quality. These conditions not only reduce the reliability of the signal representation but also hinder the ability of deep learning models to accurately capture the spatial and temporal patterns embedded in ECG data.

To address these issues, two primary filtering techniques were applied during preprocessing: a Butterworth high-pass filter with a cutoff frequency of 0.5 Hz to remove *baseline wander*, and an IIR notch filter at 60 Hz to suppress powerline interference. As shown in the lower part of *Figure 4*, the filtered signal exhibits a more stable baseline, more consistent amplitude, and a clearer, more prominent R-peak compared to the raw input. This improvement in signal quality is essential because it produces data that is both more representative and physiologically reliable without discarding crucial diagnostic features. Consequently, the refined signal facilitates more effective feature extraction and classification by CNN and CNN-GRU models. In this context, preprocessing serves not merely as a noise reduction step, but as a foundational stage that ensures stability, readability, and reliability of ECG signals, ultimately supporting the accuracy and robustness of deep learning–based sleep apnea detection systems.

3.2. SAE Input Reconstruction Results

The reconstruction results of ECG signals using SAE demonstrate an improvement in data quality, although with certain limitations. As shown in Figure 5, SAE is capable of reducing noise disturbances and smoothing the signal pattern, thereby making the main morphology—particularly the R-peak—more prominent. The clarity of the R-peak is highly important as it represents one of the key diagnostic features in identifying apnea patterns through the R–R interval. With cleaner signals, both

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

CNN and CNN-GRU models can perform spatial and temporal feature extraction more effectively. This improvement in input quality directly contributes to the increase in specificity, indicating that the model is less likely to misclassify normal segments as apnea. In other words, the reconstruction process using SAE helps the model maintain accuracy in recognizing normal signals, thus reducing the number of false positives and lowering the potential for false alarms.

Nevertheless, Figure 6 illustrates that SAE does not always succeed in reconstructing signals perfectly. In certain segments, the amplitude of the R-peak significantly decreases or even disappears entirely. The loss of such critical morphology can have a serious impact on model performance, particularly in detecting segments associated with apnea events. The absence of R-peak information reduces the temporal cues that CNN-GRU is designed to capture, thereby affecting the model's sensitivity.

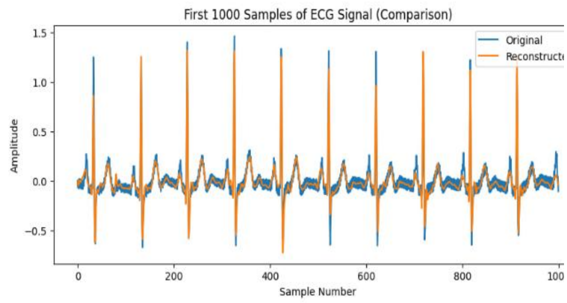


Figure 5. Comparison of ECG signals before and after SAE reconstruction

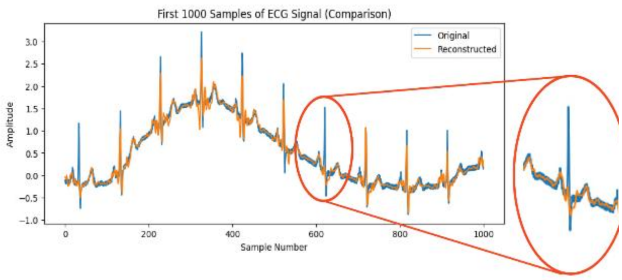


Figure 6. Example of the failure of SAE in reconstructing the R-peak

3.3. Classification and Evaluation Results for Each Scenario

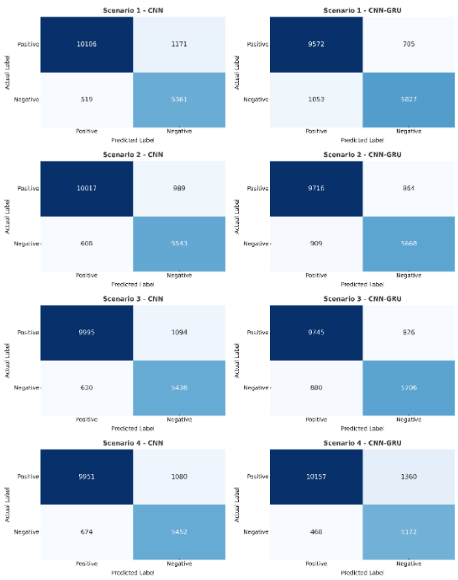


Figure 7. Confussion Matrix

The classification process was conducted across four scenarios, using both the original signals and the reconstructed signals, with the results visualized in the confusion matrix shown in Figure 7. In general, the classification patterns across scenarios exhibit consistency; however, there are notable differences in the number of False Positives (FP) and False Negatives (FN) produced by the models in each scenario. These variations in FP and FN directly influence the sensitivity and specificity values,

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

indicating that the error patterns of the models vary depending on the architecture employed and the condition of the input signals.

Table 3. Classification Results for Each Scenario

Scenario	Input	Model	Accuracy	Sensitivity	Specificity
Scenario 1	Reconstructed Raw	CNN	90,1%	95,1%	82,1%
		CNN-GRU	89,8%	90,1%	89,2%
Scenario 2	Reconstructed Preprocessed	CNN	90,7%	94,3%	84,9%
		CNN-GRU	89,7%	91,4%	86,8%
Scenario 3	Original Preprocessed	CNN	90%	94,1%	83,3%
		CNN-GRU	90,1%	91,7%	87,4%
Scenario 4	Original Raw	CNN	89.8%	93,7%	83,5%
		CNN-GRU	89,3%	95,6%	79,2%

Based on Table 3, the performance differences between the CNN and CNN-GRU architectures range from 0.1% to 5% across the metrics of accuracy, sensitivity, and specificity. For instance, in Scenario 2, the CNN model achieved the highest accuracy (90.7%), whereas the CNN-GRU model attained the highest sensitivity in Scenario 4 (95.6%). Conversely, the highest specificity was obtained by CNN-GRU in Scenario 1 (89.2%). These results indicate that the final performance of the models is influenced not only by the architecture itself but also by the signal preprocessing stages and the use of autoencoders in each scenario. CNN generally exhibited slightly higher accuracy and sensitivity, but was offset by lower specificity compared to CNN-GRU. For example, in Scenario 1, the CNN model achieved a sensitivity of 95.1% and a specificity of 82.1%. In contrast, CNN-GRU in the same scenario yielded a sensitivity of 90.1% and a specificity of 89.2% (see Table 3). This difference suggests that the inclusion of GRU components enhances the model's ability to distinguish normal conditions, thereby reducing false alarms (improving specificity) without significantly compromising sensitivity in apnea detection.

When observing the average performance, Scenarios 1 and 2 (with SAE-DEEP) achieved mean accuracy, sensitivity, and specificity values of approximately 90.0%, 92.7%, and 85.8%, respectively. In contrast, Scenarios 3 and 4 (without autoencoder) obtained averages of 89.8%, 93.8%, and 83.4%. Thus, the application of SAE-DEEP resulted in a slight improvement in accuracy (approximately 0.2%) and specificity (approximately 2.4%), although it led to a small decrease in average sensitivity (approximately 1.1%). This reduction in sensitivity indicates that the autoencoder-based model was slightly less effective in detecting apnea events. This can be attributed to the SAE's inability to consistently preserve crucial R-peak features, as illustrated in Figure 6, leading to some apnea segments being imperfectly detected.

In Scenario 2, where baseline wander reduction was applied before reconstruction, no significant improvement in sensitivity was observed; in fact, specificity slightly decreased compared to Scenario 1. The only improvement was seen in the increased accuracy of the CNN model. These findings suggest that the additional preprocessing stage did not provide substantial benefits within this system. Consequently, SAE-DEEP remained competitive even without extra preprocessing and tended to maintain more stable specificity values.

The most balanced performance was achieved by the CNN-GRU model in Scenario 1, where the differences among accuracy, sensitivity, and specificity were less than 1%. In other words, this model was able to distinguish apnea and non-apnea segments proportionally without bias toward either class. Architecturally, the GRU component enhanced the model’s ability to capture temporal patterns in continuous ECG signals, complementing the spatial feature extraction capability of CNN. This is

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

reflected in the stability of sensitivity and specificity values obtained by CNN-GRU. By contrast, the pure CNN model generally showed higher sensitivity but lower specificity, as demonstrated in the comparison for Scenario 1 in Table 3. Overall, these results indicate that the hybrid CNN-GRU architecture with SAE-DEEP can sustain a more balanced classification performance, maximizing apnea detection while avoiding excessive false alarms on normal segments.

4. DISCUSSIONS

The combined CNN-GRU model with 1DCSAE proposed in this study is effective in detecting sleep apnea from ECG signals. The model achieved an accuracy of 89.8%, sensitivity of 90.1%, and specificity of 89.2% in Scenario 1. Although peak performance was not consistently maintained across all experimental configurations, the CNN-GRU architecture consistently demonstrated good stability in balancing sensitivity and specificity. The GRU component proved effective in recognizing temporal patterns in sequential ECG signals, complementing the spatial feature extraction capabilities of CNN. Furthermore, the use of the 1DCSAE autoencoder for signal reconstruction significantly reduced noise and artifacts without losing critical features. The reconstruction results showed smoother and more stable signal structures with minimal distortion; although signal amplitude slightly decreased due to the encoding-decoding process, essential features such as the R-peak were preserved, allowing classification to proceed optimally.

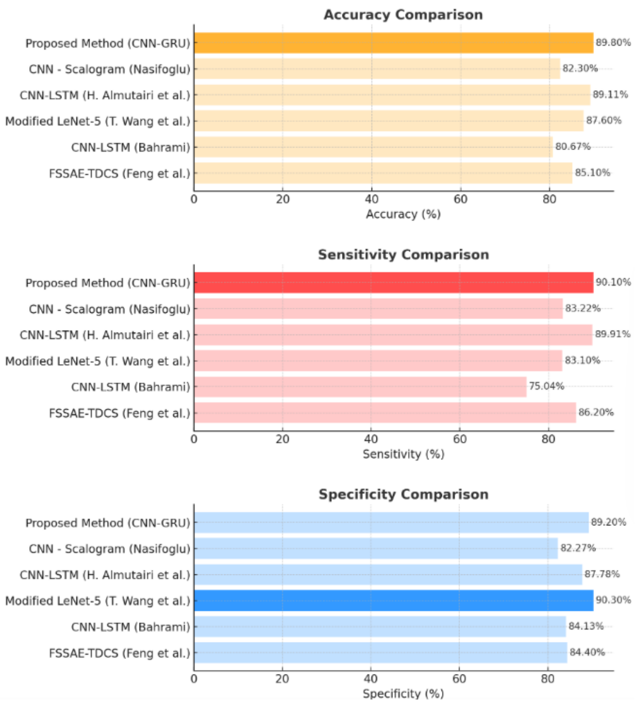


Figure 8. Comparison of the Performance of the Best Model with Other Studies

Compared with previous studies, as illustrated in Figure 8, this model demonstrates competitive advantages. For example, Nasifoglu and Erogul employed a CNN with scalogram and spectrogram inputs to detect apnea, but their model only achieved an accuracy of 82.3%, sensitivity of 83.22%, and specificity of 82.27% [24], which are lower than the performance achieved by the proposed CNN-GRU model. Similarly, Almutairi reported an accuracy of 89.11% and sensitivity of 89.91% using a CNN-LSTM architecture [25], which is comparable in accuracy but less stable in specificity compared with our approach. Wang applied a modified LeNet-5 model that achieved 87.6% accuracy and 90.3% specificity [26]. Although their specificity was relatively high, their model lacked the capability for temporal analysis, unlike CNN-GRU, which captures dynamic signal dependencies. Bahrami and Forouzanfar, who combined CNN and LSTM, reported an accuracy of only 80.67% [27], significantly lower than our results. Meanwhile, Feng [17] proposed a more complex method integrating a Frequential Stacked Sparse Autoencoder (FSSAE) with a Time-Dependent Cost-Sensitive (TDCS) classifier, achieving an accuracy of 85.1%. Although competitive, the architectural complexity poses practical challenges for real-world deployment in resource-constrained environments.

From a practical perspective, the CNN-GRU architecture offers notable computational efficiency. GRU inherently requires fewer parameters and lower computational cost compared to LSTM, making the model more suitable for real-time implementation and deployment on portable (wearable) devices with limited power and computational resources. The integration of 1DCSAE also minimizes the need for complex preprocessing stages, making the overall system simpler and faster. Consequently, the ECG-based apnea detection system can be adapted for continuous monitoring in real-world conditions using wearable devices without significant performance degradation. This efficiency is particularly important as it allows direct processing of raw signals with low energy consumption, thereby increasing the feasibility of applying the system in portable healthcare technologies.

From an academic standpoint, the hybrid CNN-GRU + 1DCSAE design provides new contributions to the field of physiological signal classification. The combination of CNN and GRU enables the simultaneous capture of spatial and temporal patterns, which is essential for understanding the sequential dynamics of cardiac signals. Moreover, the implementation of 1DCSAE within the classification pipeline enhances input quality by reducing noise while preserving key features. This hybrid approach broadens the understanding of integrating classical signal processing techniques with deep learning, demonstrating that signal reconstruction combined with automatic feature extraction can improve model robustness against variations in physiological signals. Thus, this research adds new insights into how hybrid deep learning architectures can be effectively applied to biomedical signal analysis problems.

The proposed model presents several notable advantages. Its stability in sensitivity and specificity is a key strength, as it reduces misclassification, particularly in non-apnea segments. The model is also efficient in classification, as it eliminates the need for manual feature extraction by leveraging 1DCSAE for automated feature learning and reduces reliance on complex preprocessing. The combination of CNN and GRU enables the model to generalize effectively across patient variations, ensuring robust performance under different conditions. With minimal preprocessing required, the architecture offers a lightweight and practical approach without sacrificing accuracy. The ability to directly process raw signals further opens the possibility for widespread application in long-term health monitoring.

Nevertheless, the model has certain limitations. The encoding-decoding process of 1DCSAE introduces a slight reduction in signal amplitude, particularly at the R-peak, as shown in Figure 6. Although the R-peak itself remains preserved, this amplitude distortion may diminish finer details of the signal, potentially affecting classification reliability under specific conditions. Additionally, the model's performance is strongly dependent on the quantity and quality of the training data. The limited availability of sleep apnea ECG data can increase the risk of overfitting and reduce generalizability to new datasets. ECG signals are also susceptible to motion artifacts or external interference, which, without adequate preprocessing, may lower detection accuracy. Therefore, larger and more diverse datasets, along with mechanisms for handling additional noise, are necessary to further evaluate the model's reliability under real-world conditions.

5. CONCLUSION

This study successfully developed a sleep apnea classification model based on ECG signals by integrating the CNN-GRU architecture with 1DCSAE. In its best-performing configuration (Scenario

1), the hybrid model achieved an accuracy of 89.8%, a sensitivity of 90.1%, and a specificity of 89.2%, demonstrating balanced and stable performance in distinguishing between apnea and non-apnea segments. These results highlight the significance of the proposed approach in the field of informatics, particularly in advancing deep learning methods for physiological signal classification. The model showed consistent performance even without complex preprocessing, thereby offering a simpler yet accurate detection system. Practically, the hybrid design leverages a CNN for spatial feature extraction and a GRU for efficient temporal modeling, while the 1DCSAE effectively reduces noise in ECG signals. The computational efficiency of GRU, combined with its minimal preprocessing requirements, makes this model well-suited for deployment in resource-constrained intelligent systems, such as wearable devices for real-time health monitoring, without compromising detection accuracy. However, this model has limitations, namely that the 1DCSAE encoding-decoding process causes amplitude distortion that slightly reduces the ECG R-peak, as well as dependence on limited training data that can limit the model's ability to generalize to a more diverse clinical population. Therefore, further research should explore new architectures (e.g., attention mechanisms) and apply data augmentation strategies to enrich sample variation, thereby improving the strength and reliability of future models.

REFERENCES

[1] M. Bahrami and M. Forouzanfar, “Sleep Apnea Detection From Single-Lead ECG: A Comprehensive Analysis of Machine Learning and Deep Learning Algorithms,” *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3151947.

[2] Y. Lin *et al.*, “Objective Sleep Duration and All-Cause Mortality Among People With Obstructive Sleep Apnea,” *JAMA Netw. Open*, vol. 6, no. 12, p. e2346085, Dec. 2023, doi: 10.1001/jamanetworkopen.2023.46085.

[3] F. F. Karuga *et al.*, “REM-OSA as a Tool to Understand Both the Architecture of Sleep and Pathogenesis of Sleep Apnea—Literature Review,” *J. Clin. Med.*, vol. 12, no. 18, p. 5907, Sep. 2023, doi: 10.3390/jcm12185907.

[4] J. Li, Y. Huang, S. Xu, and Y. Wang, “Sleep disturbances and female infertility: a systematic review,” *BMC Womens. Health*, vol. 24, no. 1, p. 643, 2024, doi: 10.1186/s12905-024-03508-y.

[5] A. M. Das, J. L. Chang, M. Berneking, N. P. Hartenbaum, M. Rosekind, and I. Gurubhagavatula, “Obstructive sleep apnea screening, diagnosis, and treatment in the transportation industry,” *J. Clin. Sleep Med.*, vol. 18, no. 10, pp. 2471–2479, Oct. 2022, doi: 10.5664/jcsm.9672.

[6] B. Xie and H. Minn, “Real-Time Sleep Apnea Detection by Classifier Combination,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 469–477, 2012, doi: 10.1109/TITB.2012.2188299.

[7] S. F. QUAN, J. C. GILLIN, M. R. LITTNER, and J. W. SHEPARD, “Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. Editorials,” *Sleep (New York, NY)*, vol. 22, no. 5, pp. 662–689, 1999.

[8] N. M. Punjabi, “The Epidemiology of Adult Obstructive Sleep Apnea,” *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 136–143, Feb. 2008, doi: 10.1513/pats.200709-155MG.

[9] A. H. Yüzer, H. Sümbül, M. Nour, and K. Polat, “A different sleep apnea classification system with neural network based on the acceleration signals,” *Appl. Acoust.*, vol. 163, p. 107225, Jun. 2020, doi: 10.1016/j.apacoust.2020.107225.

[10] A. Benjafield *et al.*, “Global prevalence of obstructive sleep apnea in adults: estimation using currently available data,” in *B67. Risk and prevalence of sleep disordered breathing*, American Thoracic Society, 2018, pp. A3962–A3962.

[11] A. V Benjafield *et al.*, “Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis,” *Lancet Respir. Med.*, vol. 7, no. 8, pp. 687–698, 2019, doi: https://doi.org/10.1016/S2213-2600(19)30198-5.

[12] S. C. Veasey and I. M. Rosen, “Obstructive Sleep Apnea in Adults,” *N. Engl. J. Med.*, vol. 380, no. 15, pp. 1442–1449, Apr. 2019, doi: 10.1056/NEJMcp1816152.

[13] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, “Detection of apnea events from ECG segments using Fourier decomposition method,” *Biomed. Signal Process. Control*, vol. 61, p.

102005, Aug. 2020, doi: 10.1016/j.bspc.2020.102005.

[14] K. K. Valavan *et al.*, "Detection of Obstructive Sleep Apnea from ECG Signal Using SVM Based Grid Search," *Int. J. Electron. Telecommun.*, vol. 67, no. No 1, pp. 5–12, 2021, doi: 10.24425/ijet.2020.134021.

[15] N. Pombo, B. M. C. Silva, A. M. Pinho, and N. Garcia, "Classifier Precision Analysis for Sleep Apnea Detection Using ECG Signals," *IEEE Access*, vol. 8, pp. 200477–200485, 2020, doi: 10.1109/ACCESS.2020.3036024.

[16] O. Faust, R. Barika, A. Shenfield, E. J. Ciaccio, and U. R. Acharya, "Accurate detection of sleep apnea with long short-term memory network based on RR interval signals," *Knowledge-Based Syst.*, vol. 212, p. 106591, Jan. 2021, doi: 10.1016/j.knosys.2020.106591.

[17] K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A Sleep Apnea Detection Method Based on Unsupervised Feature Learning and Single-Lead Electrocardiogram," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: 10.1109/TIM.2020.3017246.

[18] Q. Shen, H. Qin, K. Wei, and G. Liu, "Multiscale Deep Neural Network for Obstructive Sleep Apnea Detection Using RR Interval From Single-Lead ECG Signal," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: 10.1109/TIM.2021.3062414.

[19] A. Sheta *et al.*, "Diagnosis of Obstructive Sleep Apnea from ECG Signals Using Machine Learning and Deep Learning Classifiers," *Appl. Sci.*, vol. 11, no. 14, p. 6622, Jul. 2021, doi: 10.3390/app11146622.

[20] A. Zarei, H. Beheshti, and B. M. Asl, "Detection of sleep apnea using deep neural networks and single-lead ECG signals," *Biomed. Signal Process. Control*, vol. 71, p. 103125, Jan. 2022, doi: 10.1016/j.bspc.2021.103125.

[21] B. Moradhasel, A. Sheikhani, O. Aloosh, and N. Jafarnia Dabanloo, "Spectrogram classification of patient chin electromyography based on deep learning: A novel method for accurate diagnosis obstructive sleep apnea," *Biomed. Signal Process. Control*, vol. 79, p. 104215, Jan. 2023, doi: 10.1016/j.bspc.2022.104215.

[22] D. Peng, L. Sun, Q. Zhou, and Y. Zhang, "AI-driven approaches for automatic detection of sleep apnea/hypopnea based on human physiological signals: a review," *Heal. Inf. Sci. Syst.*, vol. 13, no. 1, p. 7, 2024, doi: 10.1007/s13755-024-00320-8.

[23] A. Ramachandran and A. Karuppiah, "A Survey on Recent Advances in Machine Learning Based Sleep Apnea Detection Systems," *Healthcare*, vol. 9, no. 7, p. 914, Jul. 2021, doi: 10.3390/healthcare9070914.

[24] H. Nasifoglu and O. Erogul, "Obstructive sleep apnea prediction from electrocardiogram scalograms and spectrograms using convolutional neural networks," *Physiol. Meas.*, vol. 42, no. 6, p. 065010, Jun. 2021, doi: 10.1088/1361-6579/ac0a9c.

[25] H. Almutairi, G. M. Hassan, and A. Datta, "Detection of Obstructive Sleep Apnoea by ECG signals using Deep Learning Architectures," in *2020 28th European Signal Processing Conference (EUSIPCO)*, IEEE, Jan. 2021, pp. 1382–1386. doi: 10.23919/Eusipco47968.2020.9287360.

[26] T. Wang, C. Lu, G. Shen, and F. Hong, "Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network," *PeerJ*, vol. 7, p. e7731, Sep. 2019, doi: 10.7717/peerj.7731.

[27] M. Bahrami and M. Forouzanfar, "Detection of Sleep Apnea from Single-Lead ECG: Comparison of Deep Learning Algorithms," in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2021, pp. 1–5. doi: 10.1109/MeMeA52024.2021.9478745.

[28] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, 2023, doi: 10.1007/s10462-023-10466-8.

[29] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning," *Arch. Comput. Methods Eng.*, vol. 27, no. 4, pp. 1071–1092, 2020, doi: 10.1007/s11831-019-09344-w.

[30] A. Prabakaran and E. Rufus, "Review on the wearable health-care monitoring system with robust motion artifacts reduction techniques," *Sens. Rev.*, vol. 42, no. 1, pp. 19–38, Jan. 2022, doi:

Vol. 7, No. 1, February 2026, Page. 475-489
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5182

10.1108/SR-05-2021-0150.
[31] M. Khalili, H. GholamHosseini, A. Lowe, and M. M. Y. Kuo, “Motion artifacts in capacitive ECG monitoring systems: a review of existing models and reduction techniques,” *Med. Biol. Eng. Comput.*, vol. 62, no. 12, pp. 3599–3622, 2024, doi: 10.1007/s11517-024-03165-1.
[32] J. Yang *et al.*, “Multi-Label Attribute Selection of Arrhythmia for Electrocardiogram Signals with Fusion Learning,” *Bioengineering*, vol. 9, no. 7, p. 268, Jun. 2022, doi: 10.3390/bioengineering9070268.
[33] R. Dey and F. M. Salem, “Gate-variants of Gated Recurrent Unit (GRU) neural networks,” in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, Aug. 2017, pp. 1597–1600. doi: 10.1109/MWSCAS.2017.8053243.
[34] Z. Huang and K. He, “GRU-TSMixers: Sleep Apnea and Hypopnea Detection Based on Multi Scale MLP-Mixers,” in *2024 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jun. 2024, pp. 1–8. doi: 10.1109/IJCNN60899.2024.10650636.
[35] P. Hemrajani, V. S. Dhaka, G. Rani, P. Shukla, and D. P. Bavirisetti, “Efficient Deep Learning Based Hybrid Model to Detect Obstructive Sleep Apnea,” *Sensors*, vol. 23, no. 10, p. 4692, May 2023, doi: 10.3390/s23104692.
[36] H. Almutairi, G. M. Hassan, and A. Datta, “Classification of Obstructive Sleep Apnoea from single-lead ECG signals using convolutional neural and Long Short Term Memory networks,” *Biomed. Signal Process. Control*, vol. 69, p. 102906, Aug. 2021, doi: 10.1016/j.bspc.2021.102906.
[37] S. E. Mathe, N. K. Penjarla, S. Vappangi, and H. K. Kondaveeti, “Advancements in Noise Reduction Techniques in ECG Signals: A Review,” in *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, IEEE, Jul. 2024, pp. 27–33. doi: 10.1109/AIC61668.2024.10730852.
[38] V. Gupta, A. K. Sharma, P. K. Pandey, R. K. Jaiswal, and A. Gupta, “Pre-Processing Based ECG Signal Analysis Using Emerging Tools,” *IETE J. Res.*, vol. 70, no. 4, pp. 4219–4230, Apr. 2024, doi: 10.1080/03772063.2023.2202162.
[39] M. E. Jijón-Palma, C. Amisse, and J. A. S. Centeno, “Hyperspectral dimensionality reduction based on SAE-1DCNN feature selection approach,” *Appl. Geomatics*, vol. 15, no. 4, pp. 991–1004, 2023, doi: 10.1007/s12518-023-00535-6.
[40] A. Shaheen, L. Ye, C. Karunaratne, and T. Seppänen, “Fully-Gated Denoising Auto-Encoder for Artifact Reduction in ECG Signals,” *Sensors*, vol. 25, no. 3, p. 801, Jan. 2025, doi: 10.3390/s25030801.
[41] M. Łępicki *et al.*, “Comparative Evaluation of Sequential Neural Network (GRU, LSTM, Transformer) Within Siamese Networks for Enhanced Job–Candidate Matching in Applied Recruitment Systems,” *Appl. Sci.*, vol. 15, no. 11, p. 5988, May 2025, doi: 10.3390/app15115988.
[42] F. Setiawan and C.-W. Lin, “A Deep Learning Framework for Automatic Sleep Apnea Classification Based on Empirical Mode Decomposition Derived from Single-Lead Electrocardiogram,” *Life*, vol. 12, no. 10, p. 1509, Sep. 2022, doi: 10.3390/life12101509.