Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 1340-1354 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Comparison of Deep Learning Methods for Sleep Apnea Detection Using Spectrogram-Transformed ECG Signals

Sugondo Hadiyoso[1], Inung Wijayanto[2], Ayu Sekar Safitri[2], Thalita Dewi Rahmaniar[2], Achmad Rizal[2], Suman Lata Tripathi[2]

[1] School of Applied Science, Telkom University, Bandung, Indonesia
[2] School of Electrical Engineering, Telkom University, Bandung, Indonesia
[3] Symbiosis Institute of Technology, Symbiosis International Deemed University, Pune, India

Corresponding author: Sugondo Hadiyoso (e-mail: sugondo@telkomuniversity.ac.id), **Author(s) Email**: Inung Wijayanto (e-mail: iwijayanto@telkomuniversity.ac.id), Ayu Sekar Safitri (e-mail: ayusekars@student.telkomuniversity.ac.id), Thalita Dewi Rahmaniar (e-mail: thalitadr@student.telkomuniversity.ac.id), Achmad Rizal (e-mail: achmadrizal@telkomuniversity.ac.id), Suman Lata Tripathi (suman.tripathi@sitpune.edu.in)

Abstract. Sleep apnea is a sleep disorder that occurs when breathing is repeatedly interrupted during sleep. This condition can lead to various serious health problems if left untreated, such as high blood pressure, poor sleep quality, and difficulty concentrating. Sufferers often do not realize they have sleep apnea because it occurs during sleep. Generally, diagnosis is made through interviews with the patient and their family to identify common symptoms such as snoring, and then confirmed through physical examination and Polysomnography (PSG). Since sleep apnea is related to respiratory activity that correlates with changes in cardiac activity, electrocardiogram (ECG) examination during sleep serves as an alternative diagnostic method. Therefore, this study presents a comparative analysis of deep learning models for detecting sleep apnea from spectrogram-based ECG representations. The raw ECG signals were transformed into spectrograms and then saved as images for classification into normal and abnormal categories. Deep learning (DL) methods were applied for the classification of normal and sleep apnea ECGs. EfficientNet, MobileNetV2, DenseNet, AlexNet, and VGG16 were used to evaluate the performance of the proposed method and identify the best-performing model. The evaluation results show that EfficientNet achieved the highest performance with an accuracy of 91.01%, precision of 90.70%, recall of 95.76%, and an F1-score of 92.61%. EfficientNet outperformed the other evaluated models in this study. By utilizing a spectrogram-based approach combined with a scalable architecture, the method demonstrates competitive accuracy for sleep apnea detection. Exploring other approaches to further improve accuracy remains an interesting direction for future research.

Keywords: **Deep Learning; ECG; Sleep Apnea; Spectogram**.

I. Introduction

Frequent pauses in breathing during sleep are the hallmark of sleep apnea, a common sleep disorder that can lead to poor sleep quality and various health problems [1]. Obstructive Sleep Apnea (OSA) and Central Sleep Apnea (CSA) are the two primary types of this condition. The most prevalent type, OSA, is caused by an obstruction of the airway during sleep, whereas CSA results from a failure of the brain to properly signal the respiratory muscles. Healthcare professionals typically use a combination of sleep testing, physical examination, and self-reported symptoms to diagnose sleep apnea [2]. Polysomnography (PSG), which involves monitoring blood oxygen levels, breathing patterns, and heart, lung, and brain activity throughout the night, is considered the gold standard for diagnosing sleep apnea (SA). In some cases, home sleep tests—which monitor blood oxygen levels, breathing patterns, heart rate, and airflow—may serve as a suitable alternative to PSG. One of the parameters recorded during polysomnography is the electrocardiogram (ECG) signal, which provides valuable information for diagnosing sleep apnea. The use of a single-lead ECG for apnea detection has become increasingly popular because it is simpler and more practical than using multiple signals in polysomnography. The ECG signal reflects the presence of sleep apnea through Electrocardiogram-Derived Respiration (EDR), where changes in respiratory patterns affect the ECG waveform [3].

Sleep apnea is typically diagnosed using Polysomnography (PSG), which, although considered the gold standard, is costly, time-consuming, and

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

requires overnight monitoring in specialized facilities. These limitations often lead to long waiting times and reduced accessibility, particularly in resource-limited settings [4]. Consequently, there is an increasing demand for non-invasive, cost-effective, and widely deployable diagnostic alternatives. ECG-based detection offers a promising solution due to its portability, lower cost, and suitability for long-term monitoring [5]. However, existing ECG-based methods often rely on manual feature extraction or simple time-domain analysis, which may overlook subtle temporal and spectral patterns indicative of apnea events [6], [7]. These limitations underscore the need for more advanced feature extraction techniques, such as spectrogram transformations, that can represent the full time–frequency characteristics of ECG signals.

Several methods have been used to detect sleep apnea (SA) via ECG, including heart rate variability (HRV), which begins with detecting the R–R intervals in the ECG signal [8]. Researchers utilize several HRV parameters (e.g., RMSSD, NN50) along with machine learning to detect SA. Other techniques include wavelet analysis [9], fractal analysis [10], Hilbert-Huang transform [11], and morphological analysis of ECG signals [12]. The spectrogram is a signal transformation method frequently employed for SA detection in ECG signals. A spectrogram converts one-dimensional (1D) signals into two-dimensional (2D) representations by displaying information in both the time and frequency domains [13]. Various methods have been explored to use spectrograms as features for SA detection in ECG signals. Ullah et al. used magnified R–R signals, scalograms, and spectrograms for SA detection in single-channel ECG [14]. Combined with a dual convolutional dual attention network (DCDA-Net), their method achieved an accuracy of 98% and an F1-score of 97.5%. Linh et al. proposed a different approach by analyzing the spectrograms of several subbands of the ECG signal [15]. The ECG signal was decomposed using discrete wavelet transform, and their experiments reported that the 8–50 Hz frequency band provided the best accuracy of 98.2% and an F1-score of 0.93. Another variation of spectrogram-based feature extraction was proposed by Gupta et al. [16]. A smoothed Gabor spectrogram (SGS) was combined with SqueezeNet, ResNet50, and a developed deep learning model called Obstructive Sleep Apnea Convolutional Neural Network (OSACN-Net) as a classifier, resulting in an accuracy of 94.81% using tenfold cross-validation. According to the aforementioned studies, the spectrogram requires a classifier to perform effective detection. Since automatic diagnosis of apnea is far more desirable than manual assessment, the use of classifiers in sleep apnea detection is crucial [17]. Classifiers analyze data and generate predictions based on the features extracted from signals, facilitating the automation of SA diagnosis. Deep learning as a classifier in SA detection has been widely adopted, but a comprehensive performance comparison of different models remains lacking.

This study addresses these limitations by employing a spectrogram transformation of ECG signals, enabling the extraction of rich time-frequency features directly from single-lead recordings. While time-frequency analysis has been explored in prior works, its application to spectrogram-transformed ECG data for apnea detection remains underrepresented in the literature. The proposed approach leverages these spectrograms as inputs to deep convolutional neural networks (CNNs), enabling automatic learning of discriminative patterns without manual feature engineering. This not only enhances sensitivity to apnea-related signal variations but also positions the method as a scalable, non-invasive alternative to conventional PSG-based diagnosis.

This study fills this gap by investigating the use of deep learning as a classifier combined with spectrogram as a feature extraction approach for sleep apnea identification. Deep learning is a subset of machine learning that utilizes artificial neural networks to model and solve complex problems. It has demonstrated significant promise in multiple domains, including prediction and diagnosis in medicine. We anticipate that our integration of deep learning with spectrograms will improve the accuracy and reliability of sleep apnea identification, ultimately leading to better patient outcomes and enhanced quality of life. The analysis of ECG spectrogram-based images has the potential to serve as an innovative alternative approach and a benchmark in sleep apnea detection. Specifically, the primary objective of this study is to conduct a comprehensive performance comparison of various deep learning architectures for sleep apnea detection from ECG spectrograms, in order to identify the most effective model for this task. The primary contributions of this work are summarized as follows:

1. Development of an ECG-based sleep apnea detection method. This research proposes a novel approach that utilizes ECG as the primary modality, which is more accessible and cost-effective than standard Polysomnography (PSG).
2. Introduction of a transformation from 1D ECG signals into a 2D spectrogram representation. This transformation enables the extraction of rich time-frequency features, overcoming the limitations of 1D signal analysis and presenting the information in an optimal format for processing by deep learning architectures.
3. Comprehensive comparative evaluation of various CNN architectures. This study does not merely

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

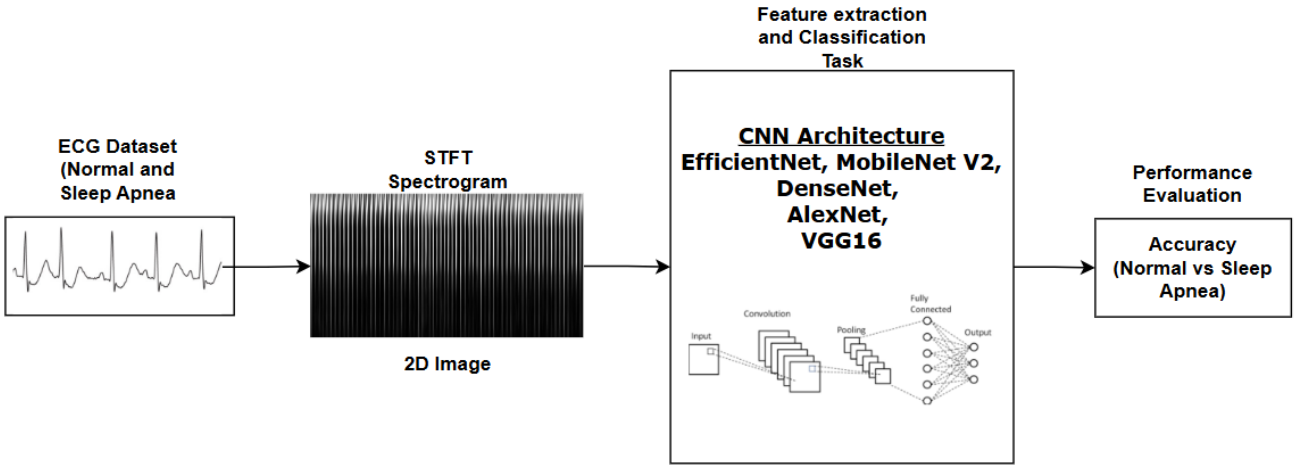

Fig. 1. Propose System for Sleep Apnea Detection based on ECG Spectrogram.

propose a single model but conducts an in-depth comparative analysis to identify the most effective and robust deep learning architecture for classifying apnea-related ECG spectrograms. The remainder of this paper is structured as follows: Section 2 describes the materials and methodologies employed in this research. Sections 3 and 4 present the results and discussion. Finally, Section 5 provides conclusions, acknowledges limitations, and outlines directions for future work.

II. Material and Methods

A. Proposed System

Fig. 1. presents a diagram of the proposed system for sleep apnea detection based on ECG spectrogram analysis. The system utilizes ECG signals (normal and sleep apnea cases) processed through Short-Time Fourier Transform (STFT) to generate spectrograms, which are subsequently converted into 2D images. These images are then input into various CNN architectures (EfficientNet, MobileNetV2, DenseNet, AlexNet, VGG16) for feature extraction and classification, with performance evaluated based on accuracy in distinguishing normal versus sleep apnea cases. Different architectures are employed to identify the highest-performing model.

B. Electrocardiogram

ECG signals have gained significant attention in diagnosing SA as an alternative to Polysomnography (PSG) due to their non-invasive nature and ease of use. Unlike PSG, which can be stressful and requires technical equipment, ECG is more patient-friendly and has a lower technical barrier for use [5]. The ECG signal amplitude of 1–2 mV provides an optimal signal-to-noise ratio among physiological signals, making it ideal for analyzing heart rate variability (HRV) and respiratory changes associated with SA. Additionally, ECG can be utilized to extract respiratory effort curves, known as ECG-derived respiration (EDR), which provide valuable information about the patient's respiratory patterns [18]. SA affects heart rate due to cyclical changes in oxygen levels during apnea or hypopnea episodes, which are reflected in the ECG. These episodes cause variations in heart rate as the body compensates for oxygen reduction by increasing respiration, leading to corresponding changes in ECG waveforms [19].

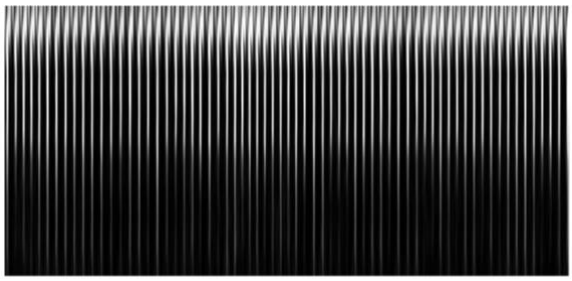

Fig. 2. Example of the Spectrogram on ECG Signal.

C. Dataset

The dataset used in this study was obtained from PhysioNet [20] and consists of 17,010 segments, with 10,496 labeled as “Normal” and 6,514 labeled as “Apnea.” The dataset is publicly available at https://www.physionet.org/content/apnea-ecg/1.0.0/. Each recording ranges from 7 to 10 hours in duration and includes a continuous ECG signal. Annotations for apnea and normal segments were provided by human experts based on simultaneously recorded respiratory and related physiological signals. Recordings with fewer than 5 minutes of disordered breathing were labeled as normal, while those with 100 minutes or more were labeled as apnea. Recordings with 10–96 minutes of disordered breathing were categorized as borderline apnea but were excluded from this study’s analysis.

D. Converting ECG Signal using Short-Time Fourier Transform (STFT)

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

Short-Time Fourier Transform (STFT) is a signal processing technique used to obtain a time–frequency representation of nonstationary signals through windowed power spectral density analysis, such as electrocardiogram (ECG) signals [21]. In this study, STFT was applied to ECG signal segments to visualize frequency information over time, enabling the identification of specific patterns in ECG signals both visually and through image processing algorithms. The ECG data consisted of discrete signals $x[n]$ sampled at $f_s = 100\ Hz$, with each recording having a length of 6000 samples. Each recording was divided into frames of length L = 25 samples with an overlap O = 20 samples, resulting in a hop size expressed by Eq. 1.

$$R = L - O = 5\ samples \quad (1)$$

Each frame was multiplied by a Kaiser window function ($w[n]$) showed in Eq. 2, [22], with a parameter $\beta = 5$ to reduce spectral leakage, where the Kaiser function incorporates the zero-order modified Bessel function I_o.

$$w[n] = \frac{I_o\left(\beta\sqrt{1-\left(\frac{2n}{L-1}-1\right)^2}\right)}{I_o(\beta)} \quad (2)$$

The Short-Time Fourier Transform (STFT) ($X(m,k)$) was then applied for each windowed frame using an FFT length $N_{FFT} = 512$, producing the complex time–frequency representation defined as Eq.3, [23], where X_m is the original signal that sampled or segmented into m overlapping frames. The Kaiser window function is represented with $w[n]$, where the sample index for a single frame is represented with n and k is the frequency bin index, ranging from 0 to $N_{FFT} - 1$. The imaginary unit is represented with j.

$$X(m,k) = \sum_{n=0}^{L-1} X_m[n]w[n]e^{-\frac{j2\pi kn}{N_{FFT}}} \quad (3)$$

The magnitude ($S(m,k)$) spectrum at frame m and frequency bin k was obtained using Eq. 4, [24] and normalized to the range $[0,1]$ using min-max scaling, as defined in Eq. 5, [24], with S_{min} and S_{max} specified in Eq. 6, [24].

$$S(m,k) = |X(m,k)| \quad (4)$$

$$\tilde{S}(m,k) = \frac{S(m,k) - S_{min}}{S_{max} - S_{min}} \quad (5)$$

$$S_{min} = \min_{m,k} S(m,k), S_{max} = \max_{m,k} S(m,k) \quad (6)$$

The normalized spectovram values ($\tilde{S}(m,k)$) then converted into 8-bit grayscale ($I(m,k)$) format according to Eq. 7, [25] which served as the visual representation of the ECG signal. In the resulting image, the horizontal axis corresponds to time with a resolution that is calculated using Eq. 8, [26], while the vertical axis represents frequency, with each bin k mapped using Eq. 9.

$$I(m,k) = \left[255 \cdot \tilde{S}(m,k)\right] \quad (7)$$

$$\Delta t = \frac{R}{f_s} = 0.05\ s/frame \quad (8)$$

$$f_k = k \cdot \frac{f_s}{N_{FFT}} \quad (9)$$

The Δt represent the time resolution of the spectogram image, while R show the hop size between frames as shown in Eq.1. The resulting spectrograms are then used as input representations for deep learning models in the classification of sleep apnea conditions. A visualization example of a spectrogram generated from ECG signal processing is shown in Fig. 2.

E. Convolutional Neural Network Design

Convolutional Neural Networks (CNN) have become a popular choice for analyzing biomedical signals due to their ability to automatically extract meaningful patterns from complex data [27]. This deep learning method is composed of layers such as convolutional, pooling and fully connected layers, which work together to perform task such as classification with high accuracy. In convolutional layers, each filter W_k slides across the input X to compute a feature map by performing a convolution operation, as expressed in Eq. 10, [28].

$$Z_{i,j,k} = (X * W_k)_{i,j} + b_k. \quad (10)$$

The $Z_{i,j,k}$ represent the activation output at (i,j) at the k^{th} filter and the bias term b_k. The output of this process is then passed through a non-linear activation function called Rectified Linear Unit (RelU) that introduced the non-linearity feature and helps the network to learn complex representation. The RelU is defined as in Eq. 11, [28].

$$A_{i,j,k} = \max(0, Z_{i,j,k}). \quad (11)$$

Furthermore, to reduce the dimentionality while preserving the most salient features, then CNN applied max pooling, as shown in Eq. 12, [28].

$$P_{i,j,k} = \max_{m,n \in R} A_{m,n,k}, \quad (12)$$

here the R represent the receptive region over the maximum activation. For the feature extraction process, the produced feature maps are flattened and passed into a fully conencted layers that produce the final output represented by Eq. 13, [28].

$$y = \sigma(W^T x + b). \quad (13)$$

where the x is the input vector, while the W, b and σ represent the weight matrix, bias, and the activation function, respectively.

Convolutional Neural Networks (CNN) have demonstrated significant potential in identifying and classifying patterns [29], highlighting their suitability for

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

the classification of spectrograms derived from ECG signals in the detection of sleep apnea. In this study, several CNN architectures, including EfficientNet, MobileNet V2, DenseNet, AlexNet, and VGG16, are employed to effectively perform sleep apnea classification.

The selection of these five architectures was motivated by their characteristics and relevance to spectrogram-based classification. EfficientNet represents a state-of-the-art scalable architecture that balances accuracy and computational cost, making it suitable for large spectrogram datasets [30]. MobileNet V2 is a lightweight model optimized for efficiency, providing a benchmark for low-resource scenarios [31]. DenseNet facilitates feature reuse through dense connections, which can be beneficial for extracting multi-scale patterns from spectrograms [32]. AlexNet, as one of the pioneering deep CNNs, serves as a baseline for evaluating advances in architectural design. VGG16, known for its depth and uniform convolutional structure, provides a comparison point for deeper but less parameter-efficient networks [33]. Including models with diverse design philosophies allows for a more comprehensive evaluation of performance, efficiency, and generalization in the context of ECG spectrogram classification. In this study, the dataset was divided into 70% for training, 20% for validation, and 10% for testing. This division was intended to adjust hyperparameters and prevent overfitting [34]. Performance evaluation did not employ cross-validation because validation was not conducted as a separate stage in this study [35].

The partitioning process was performed using a random shuffle that maintained class balance between the “Normal” and “Apnea” categories to preserve label distribution. No data augmentation techniques were applied in this study due to the risk of distorting the physiological patterns present in ECG-derived spectrogram images. Since the spectrogram captures subtle time–frequency characteristics of cardiac signals, applying common augmentation methods such as rotation, scaling, or flipping could potentially alter meaningful clinical features and compromise signal integrity. Maintaining the authenticity of the spectrograms was prioritized to ensure that the model learned from accurate and undistorted representations of sleep apnea–related patterns. All training and evaluation processes were conducted using Google Colab, which provides access to cloud-based GPU resources.

The optimizer used in this study is Adam with a learning rate of 0.001, and the model is trained with a batch size of 128, ensuring consistency across all architectures for the classification task. The model was evaluated at 5, 10, 15, and 20 epochs. The choice of the Adam optimizer was motivated by its effectiveness and frequent application in CNN architectures [36], [37], [38]. A learning rate of 0.001 was selected due to its optimal performance, consistent with findings reported in [39], [40]. No explicit regularization techniques, such as L2 weight decay or dropout beyond those built into the architectures, were employed. This decision was primarily influenced by computational limitations, which constrained the exploration of more advanced optimization strategies. These settings were established based on preliminary testing and were kept constant to allow for a fair comparison of architectural performance.

1. EfficientNet

EfficientNet, introduced in 2019, was designed to optimize the scaling of network depth, width, and resolution in a balanced manner [30]. This architecture employs a technique known as “compound model scaling,” which proportionally increases these components to maintain balance between depth and width, allowing for improved accuracy without unnecessary computational overhead [31]. This study uses the base model EfficientNet-B0 with additional layer modifications, as shown in Table 1.

2. MobileNet V2

MobileNet V2 is built on depthwise separable convolution, which decomposes the standard convolution into two parts: a depthwise convolution and a pointwise 1×1 convolution [32]. This method reduces the number of parameters and computational cost while maintaining performance. The depthwise structure of MobileNet enables efficient processing by balancing accuracy and latency through controllable parameters [33]. In this study, the MobileNet V2 architecture is modified by adding several additional layers, as described in Table 2. to enhance its performance for the classification task.

3. DenseNet

DenseNet is a deep learning architecture that introduces the concept of dense connectivity to improve the training of deep networks. Unlike traditional architectures that rely on summing the outputs of previous layers, DenseNet connects each layer to every subsequent layer directly. This dense connectivity allows the network to retain features from earlier layers, making it more efficient at learning complex representations [34]. Table 3 presents the layers of the DenseNet model used in this study.

4. AlexNet

Alexnet used in this study consists of five convolutional layers and three fully connected layers. It is designed to process large image datasets, with the first layer accepting an input image of dimensions 227×227×3 (height, width, and depth for the RGB channels) [35]. AlexNet is well known for its ability to extract complex features from images while balancing speed and accuracy, making it an effective architecture for image classification tasks [36].

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

5. **VGG16**

VGG-16 is a CNN architecture developed by the Visual Geometry Group (VGG) at the University of Oxford. It is an extension of AlexNet and is characterized by its deep structure, consisting of 16 layers. The key features of VGG16 include the use of 3×3 convolutional kernels stacked multiple times and 2×2 pooling layers for feature extraction. This design enables the model to capture more complex patterns in image data [37]. In this study, the base model was used with additional modifications to the architecture by adding extra layers, as outlined in Table 4.

F. Performance Evaluation

The classification performance of each CNN architecture was comprehensively evaluated using accuracy, precision, recall (sensitivity), specificity, F1-score, and the area under the receiver operating characteristic curve (AUC). These metrics were computed at the segment level, where each input corresponds to a single ECG spectrogram segment rather than an aggregated patient-level decision. Each segment was labeled as "A" for apnea or "N" for normal based on ground-truth annotations. Model predictions were obtained as probability scores from the final sigmoid activation layer, and a fixed decision threshold of 0.5 was applied to assign class labels. Segments with predicted probabilities ≥ 0.5 were classified as apnea ("A"), while those below the threshold were classified as normal ("N").

The definitions of each performance metric follow standard binary classification formulations [49]. Accuracy measures the overall proportion of correctly classified segments and is defined in Eq. 14

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

[49]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

Table 1. The Additional Layer for EfficientNet.

No	Layer	Size	Activation
1.	Global Max Pooling 2D	-	-
2.	Dropout	0.3	-
3.	Dense	1024	RelU
4.	Dense	1	Softmax

Table 2. The additional layer for MobileNet V2.

No	Layer	Size	Activation
1.	Global Average Pooling 2D	-	-
2.	Dense	64	RelU
3.	Dropout	0.5	-
4.	Dense	1	Softmax

Table 3. The DenseNet Architecture.

No	Layer	Size	Activation
1.	Conv2D	64	RelU
2.	Max Pooling 2D	3×3	-
3.	Dense	32	RelU
4.	Transition Layer	-	-
5.	Dense	32	RelU
6.	Transition Layer	-	-
7.	Dense	32	RelU
8.	Average Pooling 2D	7×7	-
9.	Flatten	-	-
10.	Dense	1	Softmax

Table 4 The Additional Layer for VGG16.

No	Layer	Size	Activation
1.	Flatten	-	-
2.	Dense	64	RelU
3.	Dropout	0.5	-
4.	Dense	1	Softmax

Precision reflects the proportion of correctly identified apnea segments among all predicted apnea cases (Eq. 15):

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

Recall (or sensitivity) measures the proportion of actual apnea segments correctly identified (Eq. 16):

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

Finally, the F1-score represents the harmonic mean of precision and recall (Eq. 17):

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (17)$$

where TP denotes True Positive, TN denotes True Negative, FP denotes False Positive, and FN denotes False Negative. True Positives (TP) and True Negatives (TN) represent the correct predictions made by the model, while False Positives (FP) and False Negatives (FN) represent the model's misclassifications.

In addition to these threshold-dependent metrics, the AUC was calculated to provide a threshold-independent measure of the model's discriminative ability. The ROC curve was generated by varying the classification threshold from 0 to 1 and plotting the true positive rate (TPR) against the false positive rate (FPR) at each

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

Table 5 Classification results for each architectures.

No	CNN Architecture	Accuracy (%) [95% CI]	Precision (%) [95% CI]	Recall (%) [95% CI]l	F1-Score(%) [95% CI]
1.	**EfficientNet**	**91.01 [89.53–91.53]**	**90.70 [88.32–90.88]**	**95.76 [94.81–96.58]**	**92.61 [91.81–93.36]**
2.	MobileNet V2	84.48 [83.22–85.68]	88.71 [87.26–90.06]	85.76 [84.19–87.22]	87.21 [86.06–88.28]
3.	DenseNet	88.36 [86.99–89.19]	89.62 [88.25–90.88]	91.33 [90.04–92.50]	90.47 [89.54–91.45]
4.	AlexNet	89.04 [87.94–90.07]	92.14 [90.89–93.27]	89.90 [88.53–91.16]	91.01 [90.09–91.91]
5.	VGG16	85.98 [84.77–87.13]	90.23 [88.85–91.49]	86.66 [85.13–88.09]	88.41 [87.36–89.47]

threshold. The AUC value was then obtained by integrating the ROC curve, providing a single scalar value that summarizes performance across all possible thresholds [50]. AUC is particularly informative for imbalanced datasets, as it considers the trade-off between sensitivity and specificity without being dependent on a single cutoff point [50], [51]. To assess the stability and statistical reliability of the reported performance, 95% confidence intervals (CIs) were computed for all evaluation metrics. For accuracy, precision, recall, and specificity, the Clopper-Pearson exact method was applied, as it provides robust interval estimates for proportions, even with moderate sample sizes [52]. The inclusion of these CIs enables a more rigorous comparison between models by quantifying the variability of performance metrics.

III. Results

Table 5 the classification results of the five CNN architectures used for detecting sleep apnea from ECG spectrogram images. Among all models, EfficientNet demonstrated the highest performance with an accuracy of 91.01%, precision of 90.70%, recall of 95.76%, and an F1-score of 92.61%. The narrow confidence intervals (CIs) indicate stable and reliable performance, while the high recall highlights its ability to detect apnea cases with minimal false negatives. The consistent balance across all metrics indicates robustness in both detecting apnea and correctly identifying normal cases.

EfficientNet's superiority is further evidenced by its confusion matrix and ROC curve, as shown in Fig. 3a and Fig. 4a. The confusion matrix reveals that the model makes very few classification errors, with a high number of true positives and true negatives. Its ROC curve exhibits a near-perfect shape, resulting in a high AUC score, which signifies excellent discriminative ability between the two classes. The compound scaling strategy used in EfficientNet likely contributes to its effective balance of depth, width, and resolution, enabling it to extract relevant features more efficiently than the other architectures.

In contrast, MobileNet V2 yielded the lowest performance among the evaluated models, with an accuracy of 84.48% and an F1-score of 87.21%. The confusion matrix in Fig. 3b reveals an increase in false positives and false negatives, indicating that the model struggles to generalize well to unseen data. Fig. 4b also shows a lower AUC on the ROC curve, reflecting a decline in sensitivity and specificity. This could be attributed to the lightweight nature of MobileNet V2, which trades representational power for computational efficiency, making it less suitable for complex biomedical signal classification tasks. DenseNet performed relatively well, achieving an accuracy of 88.36% and an F1-score of 90.47%. Its confusion matrix in Fig. 3c shows a strong ability to correctly classify both normal and apnea events, though some misclassifications still occur. Fig. 4c illustrates a stable ROC curve with an AUC close to that of EfficientNet, reinforcing DenseNet's strong feature extraction capability. The dense connectivity mechanism in DenseNet, which facilitates feature reuse across layers, likely enhances its performance in learning subtle variations present in the spectrograms.

AlexNet, while being one of the older architectures, achieved an accuracy of 89.04% and an F1-score of 91.01%, which is competitive with more modern models. The confusion matrix in Fig. 3d reveals a well-balanced prediction with relatively low false positives and false negatives. Its ROC curve in Fig. 4d also suggests high classification confidence, with an AUC close to that of DenseNet. This performance highlights that, with appropriate modifications, classical CNN models like AlexNet can still be effective for biomedical classification tasks, particularly when the input images are rich in spatial patterns. VGG16 showed moderate performance with an accuracy of 85.98% and an F1-score of 88.41%. As depicted in Fig. 3e, the confusion matrix reflects more frequent misclassifications. Fig. 4e displays a flatter ROC curve with a lower AUC, indicating limited

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

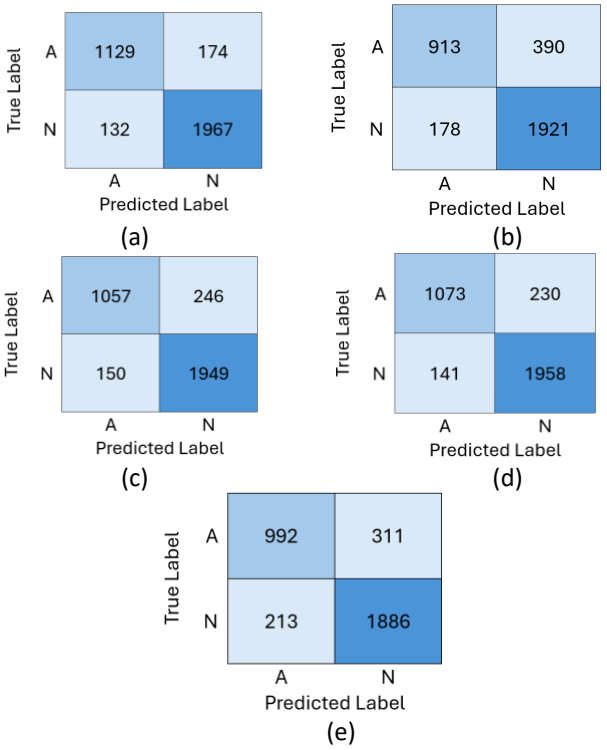

Fig. 3. Confusion Matrix Results (a) EfficientNet, (b) MobineNet V2, (c) DenseNet, (d) Alexnet, (e) VGG16.

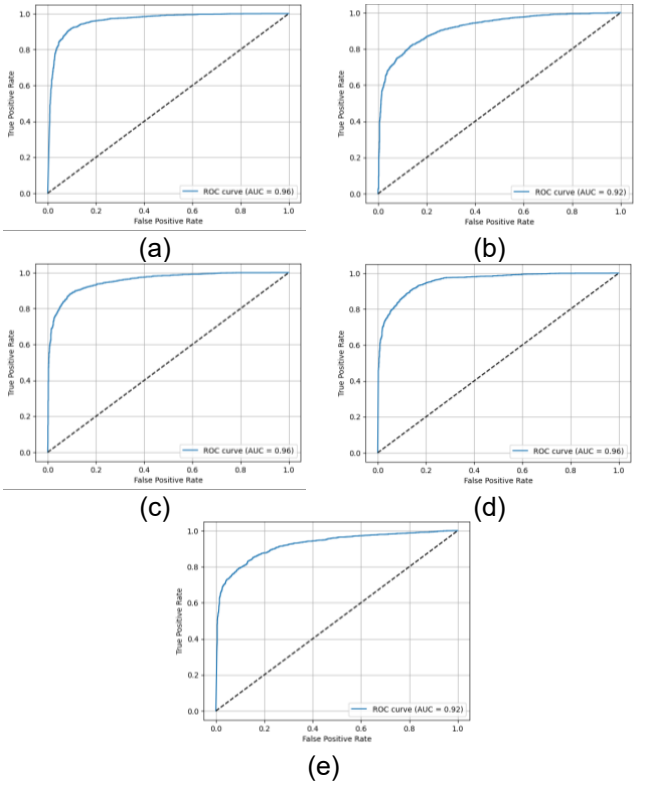

Fig. 4. Receiver Operating Curve (ROC): (a) EfficientNet, (b) MobineNet V2, (c) DenseNet, (d) Alexnet, (e) VGG16.

discriminative ability between apnea and normal conditions. Despite its deeper architecture, the lack of advanced optimization strategies, such as regularization or fine-tuning, may have hindered VGG16's ability to generalize, especially when applied to spectrogram data derived from physiological signals. The inclusion of 95% confidence intervals provides a statistical perspective on model performance stability and comparability.

EfficientNet exhibits consistently narrow CIs across all metrics, indicating low variability and high reliability in classification results. For accuracy, EfficientNet's CI does not overlap with that of MobileNet V2 or VGG16, suggesting a statistically significant advantage over these models. However, its CI overlaps with those of DenseNet and AlexNet, indicating that the observed differences in accuracy among these models may not be statistically significant at the 95% confidence level. Therefore, to determine the best-performing model, further statistical analysis should be conducted using the AUC values. EfficientNet achieved the highest AUC (0.97), with no confidence interval overlap with other models. This provides stronger evidence of its superior discriminative capability. The ROC curves further support this, showing a sharper rise and a larger enclosed area for EfficientNet.

IV. Discussion

EfficientNet achieved the highest overall performance with 91.01% accuracy, 90.70% precision, 95.76% recall, and an F1-score of 92.61%. These results demonstrate its ability to balance sensitivity and specificity more effectively than the other CNN architectures. Importantly, its recall of 95.76% indicates that almost all apnea events were correctly detected, minimizing the risk of missed diagnoses, while maintaining high precision ensures that false positives remain low.

Compared to MobileNet V2 (84.48% accuracy, 85.76% recall) and VGG16 (85.98% accuracy, 86.66% recall), EfficientNet shows a clear advantage in identifying subtle time–frequency variations. DenseNet (88.36% accuracy, 91.33% recall) and AlexNet (89.04% accuracy, 89.90% recall) performed reasonably well, but EfficientNet's superior balance across all metrics resulted in the best F1-score, highlighting its reliability as a decision-support tool. Although AlexNet achieved slightly higher precision (92.14%), its recall was lower than that of EfficientNet, meaning it missed truer apnea cases. In clinical contexts, EfficientNet's higher recall is more valuable, since underdiagnosis poses greater risk than occasional false alarms.

From a clinical perspective, the reported performance metrics carry significant implications. The high recall (95.76%) achieved by EfficientNet suggests

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

Table 6 Comparison of previous studies

Ref	Dataset	Classifier	Result
[13]	Physionet	AlexNet, GoogleNet and ResNet18 models	Scalograms Accuracy:82.30% Sensitivity: 83.22% Specificity: 82.27% Spectrograms Accuracy: 80.13% Sensitivity: 81.99%
[16]	-	Squeeze-Net, Res-Net50, and OSACN-Net	Accuracy: 94.81%
[55]	-	FASSNet	Accuracy: 87.09% Sensitivity: 77.96% Specificity: 91.74%, F1 score: 81.61%
[56]	Physionet	Lightweight CNN	Accuracy: 94.30% Sensitivity: 94.30% Specificity 94.51%
[57]	Physionet	2D-CNN	Accuracy: 92.4%, Recall: 92.3% Specificity: 92.6%,
Proposed Method	Physionet	EfficientNet	Accuracy: 91.01% Precision: 90.70% Recall: 95.76% F1-score: 92.61%

a strong ability to correctly identify apnea events, which is critical in medical diagnostics where missing true positive cases could lead to untreated sleep disorders and subsequent health complications. Conversely, maintaining high precision (90.70%) ensures that false alarms are minimized, preventing unnecessary anxiety or additional diagnostic procedures for patients incorrectly identified as having sleep apnea. In real-world applications, balancing sensitivity and specificity is essential for reliable screening tools. A model with high sensitivity ensures that most apnea cases are detected, aligning with clinical priorities to minimize missed diagnoses [44], [53], [54]. Meanwhile, adequate specificity reduces the likelihood of overdiagnosis or false-positive results. The performance demonstrated by EfficientNet across all four metrics reflects not only strong technical capability but also clinical potential as a decision-support tool in sleep apnea detection.

EfficientNet demonstrates significant advantages in its ability to outperform certain models compared to previous studies listed in Table 6. Compared to the studies in [13] and [55], EfficientNet achieves higher test accuracy by leveraging its scalable architecture and improved feature extraction capabilities. While [13] employed AlexNet, GoogleNet, and ResNet18 for OSA prediction from ECG spectrograms and scalograms, their best spectrogram-based result was 80.13% accuracy, notably lower than our 91.01%. Similarly, [55] proposed the lightweight FASSNet model for wearable-based SA detection, which achieved 87.09% accuracy—still below the performance of EfficientNet in this study. These differences may be attributed to EfficientNet's compound scaling strategy, which balances network depth, width, and resolution to capture discriminative time–frequency patterns more effectively without excessive computational cost.

However, despite these advantages, EfficientNet does not surpass some recent state-of-the-art approaches such as OSACN-Net [16], the hybrid scalogram–spectrogram method in [56], and the fused time–frequency image model in [57], which were optimized for higher accuracy and reported segment-level accuracies exceeding 92%. These higher performances can be linked to specialized optimizations such as combining complementary time–frequency representations, integrating noise-robust preprocessing, or employing hybrid deep learning modules that were not implemented in our approach. In contrast, our method focuses solely on spectrogram images from single-lead ECG data without data fusion or multimodal integration, which may limit absolute accuracy compared to such enhanced methods.

The primary limitation of this study lies in computational constraints, as the large number of spectrogram images processed during training resulted in only 20 epochs being used. This explains why the performance plateaued at an accuracy of 91.01% and an F1-score of 92.61%, rather than approaching the >94% accuracy reported by some state-of-the-art models in Table 6**.** The relatively short training duration likely limited EfficientNet's ability to progressively refine deeper feature hierarchies, especially when dealing with spectrograms that encode intricate time-frequency patterns of ECG signals, where learning subtle temporal variations and frequency shifts is essential for accurate apnea detection.

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

Extending the training duration could have facilitated better convergence and potentially improved the model's ability to generalize to unseen data. Under-optimized models risk overlooking critical features that distinguish pathological from normal patterns, especially in biomedical contexts where minor signal deviations may carry diagnostic importance. The constrained training may therefore have affected the depth and richness of the learned representations, possibly capping the model's full potential. More prolonged training or staged fine-tuning could allow the network to reach a more stable convergence and extract more discriminative features, ultimately improving diagnostic reliability. In addition, applying fine-tuning techniques or incorporating early stopping based on validation loss could help determine whether the model had reached performance saturation or still had room for improvement [13]. Exploring these strategies in future work would enhance the robustness and scalability of the proposed method. However, given the computational limitations in this study, such extensions were not feasible. Nonetheless, certain optimization techniques such as early stopping could be explored to partially address training limitations without incurring significant additional computational cost. By monitoring validation loss or accuracy during training, early stopping can help identify the optimal point where further training would not yield meaningful improvement, thereby improving model generalization without requiring more epochs. This approach may serve as a practical compromise in resource-constrained settings.

While the results obtained from the PhysioNet dataset are promising [56], assessing the model's performance across diverse datasets and real-world scenarios remains an important consideration. In clinical practice, ECG signals may vary substantially due to differences in patient demographics, recording environments, or device specifications. Such heterogeneity can affect classification performance, particularly if the model becomes overly specialized to the training distribution. Broader evaluation across heterogeneous data sources would provide deeper insight into the model's robustness and its readiness for clinical deployment.

Another methodological limitation concerns the potential influence of class imbalance in the dataset, with a higher number of normal segments compared to apnea segments. This imbalance may have contributed to the observed gap between recall (95.76%) and precision (90.70%), indicating that while the model was highly sensitive to apnea events, it also produced more false positives than ideal. Although stratified splitting [56] was used to maintain proportional distribution during training, imbalance may still affect classification performance, particularly metrics such as precision and recall. In such scenarios, models may become biased toward the majority class. Techniques such as class-weighted loss functions, oversampling of the minority class, or synthetic data generation could be explored to improve sensitivity to underrepresented classes without altering the overall dataset composition.

Transfer learning using pre-trained models on related domains could significantly reduce training time while maintaining high accuracy. Transfer learning allows a model to leverage pre-trained weights from large-scale datasets, enabling more efficient feature extraction even when domain-specific data are limited [58]. Fine-tuning selected layers of these pre-trained models can adapt them to the unique characteristics of ECG spectrograms, thereby improving sensitivity to subtle apnea-related patterns without requiring excessively large training datasets [13], [58]. Exploring these methods would not only improve performance but also ensure scalability when applied to larger or more heterogeneous datasets.

IV. Conclusion

Studies on computer-based sleep apnea diagnosis systems using deep learning methods have been widely conducted. Most of these methods have limitations in feature quality, leading to a shallow analysis of the obtained results. Time–frequency analysis has been explored in some prior works and can present optimal information for deep learning architectures. This study reinforces the potential of deep learning applied to spectrogram-transformed ECG signals as a viable solution for automated sleep apnea detection. By utilizing time–frequency representations of ECG signals, the proposed method enables meaningful feature extraction that supports accurate classification of sleep apnea conditions. EfficientNet demonstrated notable advantages due to its compound scaling strategy by obtaining accuracy, precision, recall and F1-score of 91.01%, 90.70%, 95.76%, and 92.61%, respectively. This method is able to balance the depth, width, and resolution to enhance learning efficiency. Its ability to process spectrogram features efficiently underlines the importance of model scalability and architectural optimization in biomedical signal classification tasks. Despite these strengths, the study was limited by computational constraints, particularly the short training duration, which may have hindered the model's full optimization. Future studies should consider extending training duration and applying fine-tuning techniques to enhance feature learning and generalization capability. To address computational limitations and improve model efficiency, several optimization strategies can be explored in future research.

Acknowledgments

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

The authors would like to express their sincere gratitude to the Directorate of Research and Community Service (PPM), Telkom University, for their valuable support and facilitation throughout this research. The assistance provided in the form of funding, guidance, and administrative support has been instrumental in ensuring the smooth execution of this study.

Conflict of Interest Statement

The authors declare that they have no competing interests. All authors have read and agreed to the published version of the manuscript.

Author Contribution

Sugondo Hadiyoso conceptualized and designed the study and participated in data analysis and interpretation. Inung Wijayanto contributed to the simulation and editing of the manuscript. Ayu Sekar Safitri performed the classification simulation, interpretation, and analysis. Thalita Dewi Rahmaniar compared the performance results and assisted with the analysis. Achmad Rizal collected the dataset and, Suman Lata Tripathi provided critical feedback on the manuscript. All authors have read and approved the final version of the manuscript and agree to its publication.

Funding

This study was conducted without any specific financial support.

Data Availability

The dataset supporting the findings of this study is publicly available at: https://www.physionet.org/content/apnea-ecg/1.0.0/.

Declarations

Ethical Approval

This study used existing, publicly available datasets. The original data collection was conducted in accordance with ethical standards and was approved by the appropriate institutional review boards.

Competing Interests

The authors have no competing interests to disclose.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The authors declare that during the preparation of this study, ChatGPT (GPT-5.0) was used to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final version of the publication.

References

[1] A. Qureshi and R. D. Ballard, “Current reviews of allergy and clinical immunology Obstructive sleep apnea,” 2003, doi: 10.1067/mai.2003.1813.

[2] A. P. Razi, Z. Einalou, and M. Manthouri, “Sleep Apnea Classification Using Random Forest via ECG,” *Sleep Vigil*, vol. 5, no. 1, pp. 141–146, Jun. 2021, doi: 10.1007/s41782-021-00138-4.

[3] R. J. Thomas *et al.*, “Relationship between delta power and the electrocardiogram-derived cardiopulmonary spectrogram: possible implications for assessing the effectiveness of sleep,” *Sleep Med*, vol. 15, no. 1, pp. 125–131, Jan. 2014, doi: 10.1016/j.sleep.2013.10.002.

[4] C. Guilleminault, A. Tilkian, and W. C. Dement, “The Sleep Apnea Syndromes,” *Annu Rev Med*, vol. 27, no. 1, pp. 465–484, Feb. 1976, doi: 10.1146/annurev.me.27.020176.002341.

[5] T. Wang, C. Lu, and G. Shen, “Detection of Sleep Apnea from Single-Lead ECG Signal Using a Time Window Artificial Neural Network,” *Biomed Res Int*, vol. 2019, pp. 1–9, Dec. 2019, doi: 10.1155/2019/9768072.

[6] A. Zarei and B. M. Asl, “Automatic classification of apnea and normal subjects using new features extracted from HRV and ECG-derived respiration signals,” *Biomed Signal Process Control*, vol. 59, p. 101927, May 2020, doi: 10.1016/j.bspc.2020.101927.

[7] P. Kumar Tyagi and D. Agrawal, “Automatic detection of sleep apnea from single-lead ECG signal using enhanced-deep belief network model,” *Biomed Signal Process Control*, vol. 80, p. 104401, Feb. 2023, doi: 10.1016/j.bspc.2022.104401.

[8] A. Rizal, F. D. A. A. Siregar, and H. T. Fauzi, “Obstructive Sleep Apnea (OSA) Classification Based on Heart Rate Variability (HRV) on Electrocardiogram (ECG) Signal Using Support Vector Machine (SVM),” *Traitement du Signal*, vol. 39, no. 2, pp. 469–474, Apr. 2022, doi: 10.18280/ts.390208.

[9] A. Rizal, S. Hadiyoso, H. Fauzi, and R. Widadi, “Obstructive Sleep Apnea Detection based on ECG Signal using Statistical Features of Wavelet Subband,” *International journal of electrical and computer engineering systems*, vol. 13, no. 10, pp. 877–884, Dec. 2022, doi: 10.32985/ijeces.13.10.13.

[10] M. Bsoul, H. Minn, and L. Tamil, “Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 3, pp. 416–427, May 2011, doi: 10.1109/TITB.2010.2087386.

[11] G. B. Papini *et al.*, “On the generalizability of ECG-based obstructive sleep apnea monitoring: merits and limitations of the Apnea-ECG database,” in

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Jul. 2018, pp. 6022–6025. doi: 10.1109/EMBC.2018.8513660.

[12] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. Van Huffel, "A Novel Algorithm for the Automatic Detection of Sleep Apnea From Single-Lead ECG," *IEEE Trans Biomed Eng*, vol. 62, no. 9, pp. 2269–2278, Sep. 2015, doi: 10.1109/TBME.2015.2422378.

[13] 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.

[14] N. Ullah, T. Mahmood, S. G. Kim, S. H. Nam, H. Sultan, and K. R. Park, "DCDA-Net: Dual-convolutional dual-attention network for obstructive sleep apnea diagnosis from single-lead electrocardiograms," *Eng Appl Artif Intell*, vol. 123, p. 106451, Aug. 2023, doi: 10.1016/j.engappai.2023.106451.

[15] T. T. D. Linh, N. T. H. Trang, S. Lin, D. Wu, W. Liu, and C. Hu, "Detection of preceding sleep apnea using ECG spectrogram during CPAP titration night: A novel machine-learning and bag-of-features framework," *J Sleep Res*, vol. 33, no. 3, May 2024, doi: 10.1111/jsr.13991.

[16] K. Gupta, V. Bajaj, and I. A. Ansari, "OSACN-Net: Automated Classification of Sleep Apnea Using Deep Learning Model and Smoothed Gabor Spectrograms of ECG Signal," *IEEE Trans Instrum Meas*, vol. 71, pp. 1–9, 2022, doi: 10.1109/TIM.2021.3132072.

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

[18] N. Salari *et al.*, "Detection of sleep apnea using Machine learning algorithms based on ECG Signals: A comprehensive systematic review," *Expert Syst Appl*, vol. 187, p. 115950, Jan. 2022, doi: 10.1016/j.eswa.2021.115950.

[19] A. Hossen and S. Qasim, "IDENTIFICATION OF OBSTRUCTIVE SLEEP APNEA USING ARTIFICIAL NEURAL NETWORKS AND WAVELET PACKET DECOMPOSITION OF THE HRV SIGNAL," *The Journal of Engineering Research [TJER]*, vol. 17, no. 1, p. 24, May 2020, doi: 10.24200/tjer.vol17iss1pp24-33.

[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, pp. 255–258. doi: 10.1109/CIC.2000.898505.

[21] O. Alan V. and S. Ronald W., *Discrete-Time Signal Processing*. Prentice-Hall, 1999.

[22] S. Sultan Qurraie and R. Ghorbani Afkhami, "ECG arrhythmia classification using time frequency distribution techniques," *Biomed Eng Lett*, vol. 7, no. 4, pp. 325–332, Nov. 2017, doi: 10.1007/s13534-017-0043-2.

[23] P. Ewert, B. Wicher, and T. Pajchrowski, "Application of the STFT for Detection of the Rotor Unbalance of a Servo-Drive System with an Elastic Interconnection," *Electronics (Basel)*, vol. 13, no. 2, p. 441, Jan. 2024, doi: 10.3390/electronics13020441.

[24] C. Mateo and J. A. Talavera, "Short-Time Fourier Transform with the Window Size Fixed in the Frequency Domain (STFT-FD): Implementation," *SoftwareX*, vol. 8, pp. 5–8, Jul. 2018, doi: 10.1016/j.softx.2017.11.005.

[25] K. Avci, "Fractional Kaiser Window With Application to Finite Impulse Response Digital Filter Design," *IEEE Access*, vol. 12, pp. 155549–155563, 2024, doi: 10.1109/ACCESS.2024.3484930.

[26] U. R. Acharya *et al.*, "Automated characterization of coronary artery disease, myocardial infarction, and congestive heart failure using contourlet and shearlet transforms of electrocardiogram signal," *Knowl Based Syst*, vol. 132, pp. 156–166, Sep. 2017, doi: 10.1016/j.knosys.2017.06.026.

[27] B. K. Nugraha, Q. D. Amalia, A. S. Safitri, A. Rizal, and H. T. Fauzi, "Comparison Analysis for Life-Threatening Arrhythmia Classification from ECG Data Using Machine Learning and Deep Learning Methods," in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ICITISEE63424.2024.10730398.

[28] C. Cao *et al.*, "Deep Learning and Its Applications in Biomedicine," *Genomics Proteomics Bioinformatics*, vol. 16, no. 1, pp. 17–32, Feb. 2018, doi: 10.1016/j.gpb.2017.07.003.

[29] B. M. Mathunjwa, Y.-T. Lin, C.-H. Lin, M. F. Abbod, and J.-S. Shieh, "ECG arrhythmia classification by using a recurrence plot and convolutional neural network," *Biomed Signal Process Control*, vol. 64, p. 102262, Feb. 2021, doi: 10.1016/j.bspc.2020.102262.

[30] Y. A. Saadoon, M. Khalil, and D. Battikh, "Predicting Epileptic Seizures Using EfficientNet-B0 and SVMs: A Deep Learning Methodology for EEG Analysis," *Bioengineering*, vol. 12, no. 2, p. 109, Jan. 2025, doi: 10.3390/bioengineering12020109.

[31] F. A. D. Putra and U. N. Wisesty, "Pneumonia Detection Based on Chest X-Ray Images with MobileNetV2 Architecture," in *2024 International*

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), IEEE, Dec. 2024, pp. 748–753. doi: 10.1109/ICICYTA64807.2024.10913362.

[32] S. U. R. Khan and Z. Khan, "Detection of Abnormal Cardiac Rhythms Using Feature Fusion Technique with Heart Sound Spectrograms," *J Bionic Eng*, Apr. 2025, doi: 10.1007/s42235-025-00714-8.

[33] A. Nair, H. Vadher, P. Patel, T. Vyas, C. Bhatt, and A. Bruno, "Lung sound disease detection using attention over pre-trained efficientnet architecture," *Multimed Tools Appl*, vol. 84, no. 22, pp. 25485–25519, Aug. 2024, doi: 10.1007/s11042-024-20078-1.

[34] J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning," *Genet Program Evolvable Mach*, vol. 19, no. 1–2, pp. 305–307, Jun. 2018, doi: 10.1007/s10710-017-9314-z.

[35] T.-T. Wong and P.-Y. Yeh, "Reliable Accuracy Estimates from *k* -Fold Cross Validation," *IEEE Trans Knowl Data Eng*, vol. 32, no. 8, pp. 1586–1594, Aug. 2020, doi: 10.1109/TKDE.2019.2912815.

[36] I. Amelia Dewi and M. A. Negara Ekha Salawangi, "High performance of optimizers in deep learning for cloth patterns detection," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 3, p. 1407, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1407-1418.

[37] Y. Shao *et al.*, "An Improved BGE-Adam Optimization Algorithm Based on Entropy Weighting and Adaptive Gradient Strategy," *Symmetry (Basel)*, vol. 16, no. 5, p. 623, May 2024, doi: 10.3390/sym16050623.

[38] Y. Shao *et al.*, "An Improvement of Adam Based on a Cyclic Exponential Decay Learning Rate and Gradient Norm Constraints," *Electronics (Basel)*, vol. 13, no. 9, p. 1778, May 2024, doi: 10.3390/electronics13091778.

[39] C. Peng, "Comprehensive Analysis of the Impact of Learning Rate and Dropout Rate on the Performance of Convolutional Neural Networks on the CIFAR-10 Dataset," *Applied and Computational Engineering*, vol. 102, no. 1, pp. 183–192, Nov. 2024, doi: 10.54254/2755-2721/102/20241161.

[40] J. Chi, Y. Liu, V. Wang, and J. Yan, "Performance Analysis of Three kinds of Neural Networks in the Classification of Mask Images," *J Phys Conf Ser*, vol. 2181, no. 1, p. 012032, Jan. 2022, doi: 10.1088/1742-6596/2181/1/012032.

[41] N. Tauhid Dinitra, I. Wijayanto, F. Akhyar, A. Sekar Safitri, and R. Satria Firdhaust, "Enchancing Skin Cancer Classification Using EfficientNet-Based Architectures," in *2024 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, Jul. 2024, pp. 415–420. doi: 10.1109/ICoDSA62899.2024.10652183.

[42] M.-H. Liu *et al.*, "EfficientNet-based machine learning architecture for sleep apnea identification in clinical single-lead ECG signal data sets," *Biomed Eng Online*, vol. 23, no. 1, p. 57, Jun. 2024, doi: 10.1186/s12938-024-01252-w.

[43] H. Pan, Y. Yu, J. Ye, and X. Zhang, "MobileNetV2: A lightweight classification model for home-based sleep apnea screening," Dec. 2024, [Online]. Available: http://arxiv.org/abs/2412.19967

[44] 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.

[45] C. Zhang *et al.*, "ResNet or DenseNet? Introducing Dense Shortcuts to ResNet," Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.12496

[46] D. Gupta, B. Bajpai, G. Dhiman, M. Soni, S. Gomathi, and D. Mane, "Review of ECG arrhythmia classification using deep neural network," *Mater Today Proc*, May 2021, doi: 10.1016/j.matpr.2021.05.249.

[47] A. Kumar M. and A. Chakrapani, "Classification of ECG signal using FFT based improved Alexnet classifier," *PLoS One*, vol. 17, no. 9, p. e0274225, Sep. 2022, doi: 10.1371/journal.pone.0274225.

[48] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, and K.-W. Huang, "An Improved VGG16 Model for Pneumonia Image Classification," *Applied Sciences*, vol. 11, no. 23, p. 11185, Nov. 2021, doi: 10.3390/app112311185.

[49] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Oct. 2020.

[50] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit Lett*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[51] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS One*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.

[52] C. J. CLOPPER and E. S. PEARSON, "THE USE OF CONFIDENCE OR FIDUCIAL LIMITS ILLUSTRATED IN THE CASE OF THE BINOMIAL," *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934, doi: 10.1093/biomet/26.4.404.

[53] M.-H. Liu *et al.*, "EfficientNet-based machine learning architecture for sleep apnea identification in clinical single-lead ECG signal data sets," *Biomed Eng Online*, vol. 23, no. 1, p. 57, Jun. 2024, doi: 10.1186/s12938-024-01252-w.

[54] T. Gomes, A. Benedetti, A.-L. Lafontaine, R. J. Kimoff, A. Robinson, and M. Kaminska, "Validation

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

of STOP, STOP-BANG, STOP-BAG, STOP-B28, and GOAL screening tools for identification of obstructive sleep apnea in patients with Parkinson disease," *Journal of Clinical Sleep Medicine*, vol. 19, no. 1, pp. 45–54, Jan. 2023, doi: 10.5664/jcsm.10262.

[55] Y. Yu, Z. Yang, Y. You, and W. Shan, "FASSNet: fast apnea syndrome screening neural network based on single-lead electrocardiogram for wearable devices," *Physiol Meas*, vol. 42, no. 8, p. 085005, Aug. 2021, doi: 10.1088/1361-6579/ac184e.

[56] F. R. Mashrur, Md. 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," *Comput Biol Med*, vol. 134, p. 104532, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104532.

[57] S. M. I. Niroshana, X. Zhu, K. Nakamura, and W. Chen, "A fused-image-based approach to detect obstructive sleep apnea using a single-lead ECG and a 2D convolutional neural network," *PLoS One*, vol. 16, no. 4, p. e0250618, Apr. 2021, doi: 10.1371/journal.pone.0250618.

[58] A. Kolesnikov *et al.*, "Big Transfer (BiT): General Visual Representation Learning," May 2020.

Author Biography

Sugondo Hadiyoso received his Doctorate in Electrical Engineering from the Bandung Institute of Technology (ITB), Indonesia, in March 2023. Since 2010, he has served as a lecturer in the Diploma Program of Telecommunication Technology, School of Applied Science, Telkom University. He is also currently a member of the Biomedical Instrumentation Research Group at Telkom University. His research interests include wireless sensor networks, embedded systems, logic design on FPGA, and biomedical engineering. His doctoral research focused on signal processing and the analysis of EEG signals. He can be contacted via email at sugondo@telkomuniversity.ac.id.

ORCID: https://orcid.org/0000-0002-2086-2156

INUNG WIJAYANTO received his Bachelor's and Master's degrees in Telecommunications Engineering from Telkom University, Bandung, Indonesia, in 2008 and 2011 , respectively. He received his Doctorate in Electrical Engineering from the Faculty of Engineering, Gadjah Mada University, Yogyakarta, Indonesia, in 2022.

Since 2010, he has been a teaching staff at the School of Electrical Engineering, Telkom University. His current research interests include audio and image processing, biomedical signal and image processing and analysis, computer vision, and medical instrumentation. ORCID: https://orcid.org/0000-0003-1412-0428

Ayu Sekar Safitri completed her Bachelor's degree in Telecommunication Engineering and her Master's degree in Electrical Engineering at Telkom University in 2023 and 2025, respectively. She is currently pursuing a Doctoral degree in Electrical Engineering at the same institution. Her studies focus on image processing, biomedical signals, and artificial intelligence, with a particular emphasis on EEG signal processing in the biomedical field. Her research is dedicated to integrating artificial intelligence to advance both image processing and biomedical applications. ORCID: https://orcid.org/0009-0008-0231-849X

Thalita Dewi Rahmaniar received her Bachelor's degree in Telecommunication Engineering from Telkom University in 2023. She completed her Master's degree in Electrical Engineering at Telkom University in 2025, specializing in image processing, biomedical signal processing, and artificial intelligence. Her research interests focus on the development and application of artificial intelligence in biomedical and image processing domains, with a particular emphasis on low-cost, data-driven healthcare innovations. ORCID: https://orcid.org/0009-0002-0010-3432

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967

Achmad Rizal received his Bachelor's degree in Telecommunication Engineering from STT Telkom (now Telkom University), Bandung, Indonesia, in 2000. He received his Master's degree in Biomedical Engineering from the Bandung Institute of Technology (ITB), Bandung, Indonesia, in October 2006, and his Ph.D. degree from Gadjah Mada University, Yogyakarta, Indonesia. His research interests include biomedical signal processing, biomedical image processing, biomedical instrumentation, and telemedicine. He is currently a Professor at the School of Electrical Engineering, Telkom University. ORCID: https://orcid.org/0000-0001-9712-965X

Suman Lata Tripathi completed her Ph.D in Microelectronics and VLSI at MNNIT, Allahabad. She received her M.Tech. in Electronics Engineering from Uttar Pradesh Technical University, Lucknow, and her B.Tech. in Electrical Engineering from Purvanchal University, Jaunpur. She is affiliated with the Department of Electronics and Communication Engineering, Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed University), Pune, India. She has published more than 35 research papers in peer-reviewed journals and conferences. She has organized several workshops, summer internships, and expert lectures for students. She has also served as a session chair, conference steering committee member, editorial board member, and reviewer for international and national IEEE journals and conferences.

Manuscript received June 5, 2024; Revised September 20, 2025; Accepted October 10, 2025; date of publication October 25, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.967