IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 January 2025, accepted 5 March 2025, date of publication 13 March 2025, date of current version 2 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3550949

RESEARCH ARTICLE

Convolutional Autoencoder With Sequential and Channel Attention for Robust ECG Signal Denoising With Edge Device Implementation

ALIF WICAKSANA RAMADHAN[1], SYIFA KUSHIRAYATI[1], SALSABILA AURELLIA[1], MGS M. LUTHFI RAMADHAN[1], MUHAMMAD HANNAN HUNAFA[1], MUHAMMAD FEBRIAN RACHMADI[1], APRINALDI JASA MANTAU[1], (Member, IEEE), SITI NURMAINI[2], (Member, IEEE), SATRIA MANDALA[3,4], (Member, IEEE), AND WISNU JATMIKO[1], (Senior Member, IEEE)

[1]Faculty of Computer Science, University of Indonesia, Depok 16424, Indonesia
[2]Intelligent System Research Group, Sriwijaya University, Palembang 30139, Indonesia
[3]Human Centric (HUMIC) Engineering, Telkom University, Bandung 40257, Indonesia
[4]School of Computing, Telkom University, Bandung 40257, Indonesia

Corresponding author: Alif Wicaksana Ramadhan (alifwr@outlook.com)

This work was supported by the Katalis Program from the Ministry of Education, Culture, Research, and Technology (Kemendikbud) under Grant NKB-1010/UN2.RST/HKP.05.00/2024.
This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Electrocardiograms (ECG) are vital for diagnosing various cardiac conditions but are often corrupted by noise from multiple sources, which can hinder accurate interpretation. Denoising ECG signals is particularly challenging because noise usually overlaps with the frequency range of the signal of interest. We proposed a convolutional autoencoder with sequential and channel attention (CAE-SCA) to address this issue. Sequential attention (SA) is based on long short-term memory (LSTM), which captures causal-temporal relationships. Meanwhile, channel attention (CA) is used to emphasize important features within channels. SA is applied to the skip connection of each encoder block, and CA is applied after each decoder block. We validated the CAE-SCA using the MIT-BIH and SHDB-AF databases as clean ECG signals, with the MIT-BIH Noise Stress Test Database as the noise source. Experimental results give an average SNR value of 16.187 dB, RMSE of 0.059, and PRD value of 18.529 in the MIT-BIH database. While in the SHDB-AF dataset, the model obtained 15.308 dB of SNR, 0.049 of RMSE, and 19.220 of PRD. These results demonstrate our CAE-SCA outperforms all the state-of-the-art methods across all tested metrics. For efficiency, CAE-SCA achieved competitive results in the metrics of floating-point operations (FLOPs), inference time, and total parameters. This allowed CAE-SCA to be implemented in edge devices as we tested using our custom ECG acquisition circuit. A significance test further confirms a statistically significant improvement in SNR values achieved by the CAE-SCA compared to baseline models, suggesting the CAE-SCA's potential for advancing ECG processing in healthcare applications.

INDEX TERMS Autoencoder, convolutional neural networks (CNN), ECG, signal denoising, LSTM.

I. INTRODUCTION

The rising incidence of heart disease in Indonesia has led to substantial advances in health screening equipment. The Indonesian Ministry of Health reported that the number of heart disease cases is increasing every year [1]. As a

The associate editor coordinating the review of this manuscript and approving it for publication was R. K. Tripathy.

result, there is still a great demand for heart diagnostic tools, which motivates vigorous research [2], [3], [4]. According to the World Health Organization (WHO), cardiovascular illnesses were responsible for over 38% of Indonesia's total fatalities in 2019 [5], [6]. This highlights the urgent need for precise, effective, and affordable diagnostic methods are needed. For people at high cardiovascular risk due to conditions such as hypertension, diabetes, and hyperlipidemia, early identification is essential to enable prompt intervention [7].

By capturing the electrical activity of the heart, electrocardiograms (ECGs) can be used for the diagnosis of cardiac conditions. Important clinical information for identifying arrhythmias is provided by key ECG components, including the P wave, QRS complex, T wave, and U wave [8], [9]. To improve arrhythmia detection, the Holter monitor, which is the gold standard ambulatory electrocardiogram equipment, provides extended monitoring periods (24-72 hours) [7]. Nevertheless, recording outside of a hospital increases the risk of noise from electrical interference, muscle activity, or electrode movement, all of which can lower the quality of the diagnosis [9]. Accurate recordings depend on the use of efficient noise reduction techniques, such as ECG denoising and acceptability evaluation [10].

Many deep learning-based denoising approaches have been proposed using various models. A popular model for this task is the denoising autoencoder (DAE) [11]. An autoencoder comprises two parts: an encoder and a decoder. The encoder reduces the complexity of the data while retaining important features. The decoder works oppositely to the encoder. The compressed data between the encoder and decoder is referred to as the latent space. By compressing and decompressing the input into a new representation, the model retains the important features in the latent space required to reconstruct the new representation. Many researchers have proposed denoising autoencoders, with various improvements, from simple, fully connected neural networks to deep neural networks [12], [13] and not limited to one-dimensional signal data but also for image data [14], [15].

Another challenge in implementing a deep learning-based approach to Holter monitors is device constraints. The analysis can be performed from a recorded ECG signal, which can be performed anywhere; there are no limitations in the device specifications. Meanwhile, a Holter monitor is a small device that can accompany the patient under any condition for 24 to 72 hours. It is impossible to use such a device with super specifications. To handle this situation, we must develop a deep learning model while maintaining a trade-off between model inference speed and accuracy.

To explore various approaches to signal denoising tasks, we focus on two main model architecture types. These two groups are architecture based on convolutional neural networks (CNN) and architecture based on recurrent neural networks (RNN) [2], [16], [17], [18], [19], [20], [21]. It is commonly known that the RNN-based model is slower due to its way of sequentially processing the input [22]. The CNN-based model is designed to process data in a parallel way. As stated previously, we created a lightweight deep-learning model while maintaining accuracy; thus, we focused only on CNN-based architecture.

Chiang et al. proposed a fully convolutional denoising autoencoder (FCNDAE) for a simple CNN-based model for the signal denoising task [16]. Their proposed autoencoder was constructed using convolutional blocks as the encoder and convolutional blocks as the decoder. An improvement to this was proposed by Chen et al. by adding skip connections between each encoder and decoder block [23]. Later improvement has also been shown by [24] in their work [24]. They added a convolutional block attention module (CBAM) to the autoencoder network. The results demonstrate improvement compared to the vanilla UNet-based autoencoder.

Furthermore, [19] proposed another approach by adding the transformer encoder as the bottleneck of the autoencoder, and custom other modules named adaptive parametric ReLU module (AP-ReLU) and dynamic feature aggregation module (DAM) [19]. The proposed AP-ReLU was inspired by the ReLU activation function. The DAM network was also inspired by the CBAM proposed in several modules in the encoder and decoder blocks.

To obtain insight into the previously available denoising architecture model, we conducted preliminary experiments on it. The environments are different from each other; thus, we rebuilt them in our environment configurations. In the RNN-based model, the LSTM model achieved the best performance values. In addition, in the CNN-based models, we found that by adding a skip connection to the CNN-based autoencoder, there was a significant performance improvement. Of course, it was not free. The number of computations required for the model also increased, thereby increasing the inference's latency higher. In addition, the addition of attention blocks improves the performance of the UNet-based model.

We aim to create a new model based on autoencoder architectures by combining the advantages of some baseline methods. Specifically, we would like to combine the advantages of the channel attention module and LSTM to capture important features both through channels and sequences. The channel attention module demonstrates how humans emphasize useful information while giving less attention to unimportant details, which improves model efficiency and accuracy. In addition, LSTM networks can capture long-term causal relationships in sequential data, such as ECG.

This study aims to develop a more efficient denoising autoencoder model to optimize its functionality in embedded devices so that it can be implemented directly in Holter monitor systems used by medical practitioners. This technology is expected to enhance early detection and intervention for heart

disease, especially in areas with limited access to medical facilities. The main contributions of this study are as follows:

- We introduced a convolutional autoencoder architecture with sequential and channel attention (CAE-SCA) for denoising ECG signals by combining the 1D-UNet-based network, LSTM-based RNN, and channel attention module.
- Our CAE-SCA outperformed all accuracy metrics while maintaining competitive performance in terms of computational efficiency metrics.

The remainder of this paper is organized as follows. Section II provides related work. Section III explains the details of the proposed CAE-SCA. Section IV gives a brief overview of the dataset used in this study. Section V explains the experimental design and experimental results. Section VI concludes this paper.

II. RELATED WORKS

Several researchers have addressed the task of denoising ECG signals using an autoencoder (AE) [16], [19], [24], [25]. Research by Chiang et al. proposed a fully convolutional network autoencoder (FCNDAE) [16]. This model consists of 6 convolution layers on the encoder side; each encoder block consists of a convolutional, batch normalization, and activation Exponential Linear Units (ELU) layer. On the decoder side, 7 decoder blocks built with one deconvolution layer were used. The results demonstrate that FCNDAE outperforms traditional DNN and CNN approaches. This is demonstrated by a higher SNR value, a lower RMSE value, and a PRD value on testing using the MIT-BIH arrhythmia database [26] with noise added from the MIT-BIH noise stress test database [27].

Several studies have added an attention mechanism on the decoder side and a skip connection between the output for each encoder-decoder block [19], [24], [25] like U-Net Architecture [28]. The attention block demonstrates how humans emphasize useful information while giving less attention to unimportant details. The attention-based Convolutional Denoising Autoencoder (ACDAE) used the CNN layer followed by the leaky rectified linear unit (ReLU) activation function and the maximum pooling layer for each of the 4 encoders and the transposed CNN layer followed by the Leaky ReLU activation function, the maximum pooling layer, and the channel attention module for each of the 4 decoder blocks. The results show that the skip connection mechanism can help recover the detailed features of the ECG signal that might be lost when compressed with an encoder block. In addition, the channel attention module hardly improved the model's performance if placed on the last layer of the decoder [25].

Another model with a similar architecture as ACDAE [25] is Convolutional Denoising Autoencoder with Block Attention Module (CDAE-BAM) [24]. The significant difference is the additional spatial attention module in the attention block, where the attention block is placed on every last layer of the encoder and decoder block. The final result shows that this model outperforms the other models, FCNDAE [16] and ACDAE [25], on the sum of squared distances (SSD), maximum absolute distance (MAD), and cosine similarity metrics. The models were tested on the denoising ECG task using the Medical Information Mart for Intensive Care (MIMIC-III) Database [29] with noise added from the MIT-BIH noise stress test database [27].

While ACDAE [25] and CDAE-BAM [24] tried to give global attention using channel and spatial attention mechanisms. On the other hand, study [19] used a transformer layer in the bridge of the encoder-decoder block. They also used a custom activation network in place of a classical activation function called adaptive parametric ReLU (AP-ReLU) [30] for better involvement of negative value results for each output of the network layer. They also used the Dynamic Aggregation Module (DAM) on the last layer of the decoder block, whose mechanism is similar to the channel and spatial attention module used in [24]. research. The proposed method is APtrans-CNN. This research gives a result by combining an AE-based CNN with transformers and DAM and using AP-ReLU as an activation network to perform better denoising tasks, instead of only one or two combinations of them, tested on the MIT-BIH arrhythmia [26] and atrial fibrillation database [31] with noise added from the MIT-BIH noise stress test database [27].

From the related works on denoising tasks using autoencoders, we hypothesize that autoencoder performance can be significantly improved by incorporating skip connections between encoder and decoder blocks and adding local or global attention mechanisms. This research aims to combine these strategies while optimizing for efficient resource utilization.

III. PROPOSED METHOD

Fig. 1 explains the overall system design of our ECG device and the proposed deep learning-based ECG signal denoising filter. In addition to the deep learning-based filter, we also used an analog filter for the ECG signal acquisition circuit. This is a common implementation in any signal-acquisition device, including ECG.

The recorded raw signal, $x(n)$ has two main components: the signal of interest or clean signal, $s(n)$, and the noise or unwanted signal $\eta(n)$, represented in (1). The signal-denoising task can be considered a system with the same input and output domains. The input is a noised ECG signal $x(n)$, and the output is a predicted clean ECG signal $\hat{s}(n)$. Both signals are ECG signals in the time domain.

$$x(n) = s(n) + \eta(n), \quad n \in \mathbb{Z}_{\geq 0} \tag{1}$$

The output of the denoised signal is visualized in end-user devices, such as smartphones and desktops. The results are sent to the Raspberry Pi 5 as the core processor device through a Wi-Fi connection.

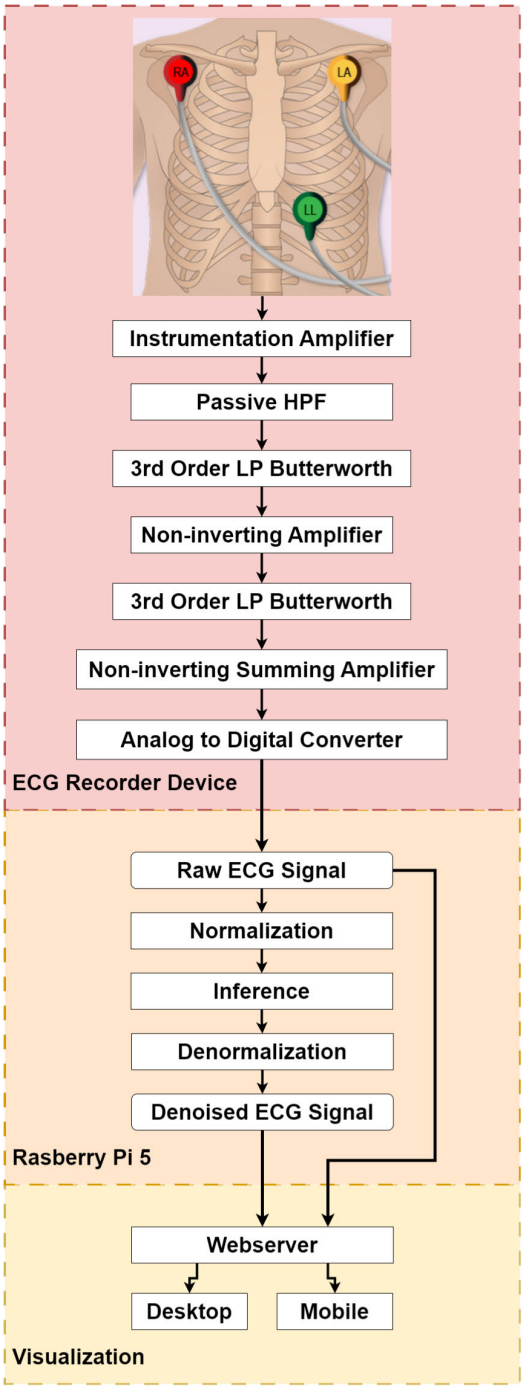


FIGURE 1. System design.

A. ELECTRONIC CIRCUIT DESIGN

The upper side of Fig. 1 shows the ECG signal-acquisition circuit design. This circuit is used to amplify a very low ECG signal to the desired range so that it can be measured by an analog-to-digital converter (ADC) accurately. Even though we have already implemented a deep learning-based filter, we still need an analog filter in this circuit. First, a deep learning-based filter is a digital filter with limitations such as, may experience aliasing. In addition, adding an analog filter also helps optimize the digital filter's performance.

An instrumentation amplifier is commonly used for ECG signal conditioning [32], [33], [34]. ECG signals are susceptible to interference from common-node noise sources, such as power-line noise, motion artifacts, and other electrical disturbances. The instrumentation amplifier was used for its high gain with low noise, high input impedance, and excellent common-mode rejection ratio (CMMR). The driven right leg (DRL) was also used to optimize the use of the instrumentation amplifier [35]. The DRL circuit actively reduces common-mode noise at the source by driving the common-mode voltage (measured across the body) to near zero. This is achieved by injecting an inverted version of the common-mode signal back into the patient through the right leg electrode, creating a feedback loop that cancels the noise.

We also added a passive high-pass filter (HPF) to reduce noise produced by the patient's body which has a low frequency. Other noise sources also come at high frequencies. Aliasing can cause a high-frequency fold in the interest signal and distort the digitized signal [36]. To avoid this, the captured signal should have a range of frequencies below half the sampling rate used by the ADC, according to the Nyquist theorem [37]. This is achieved by adding a low-pass filter (LPF) to the ECG acquisition circuit. In our circuit, we use 3_{rd} order Butterworth filter rather than a common LPF. The Butterworth filter is designed to have a maximally flat response in the pass band, meaning that there are no ripples or variations in the amplitude of the signal within the pass band [38]. This ensures that the desired ECG signal is not distorted or attenuated.

B. DIGITAL FILTER DESIGN

A digital system can only process distinct signal data. For a continuous signal, such as an ECG signal in our case, it is common to divide the signal into a short time frame or window to make it distinct data. Each resulting frame reflects a segment of the signal sequence in a short period. The signal is then processed frame-by-frame and treated as a sequence. This approach is called windowing, which allows us to analyze the local characteristics of the signal over time and has also been implemented in many diverse research studies [3], [39], [40].

In this paper, we have proposed a deep learning-based filter due to its success in various problems. This deep learning-based filter can process each segment of the ECG signal. The output of each filter operation is then arranged to form a new continuous sequence of filtered ECG signals. Fig. 2 shows the visualization of these processes.

C. MODEL ARCHITECTURE

This study proposes a combination of a CNN-based autoencoder with LSTM (Long Short-Term Memory) on the encoder gate at the skip connection and a channel attention block on the decoder gate at the skip connection. The basic mechanism

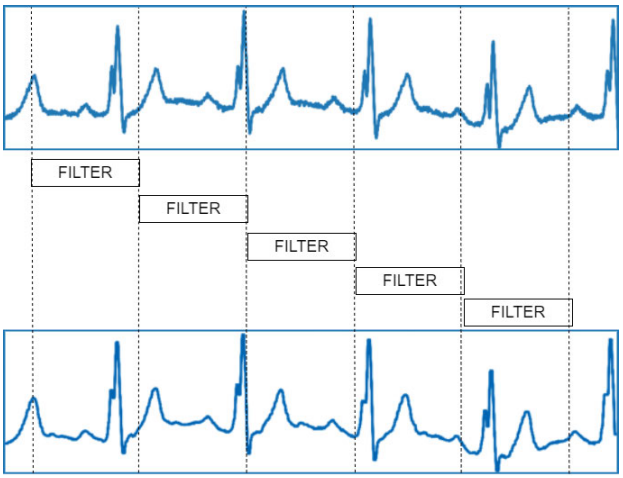


FIGURE 2. Windowing filter process.

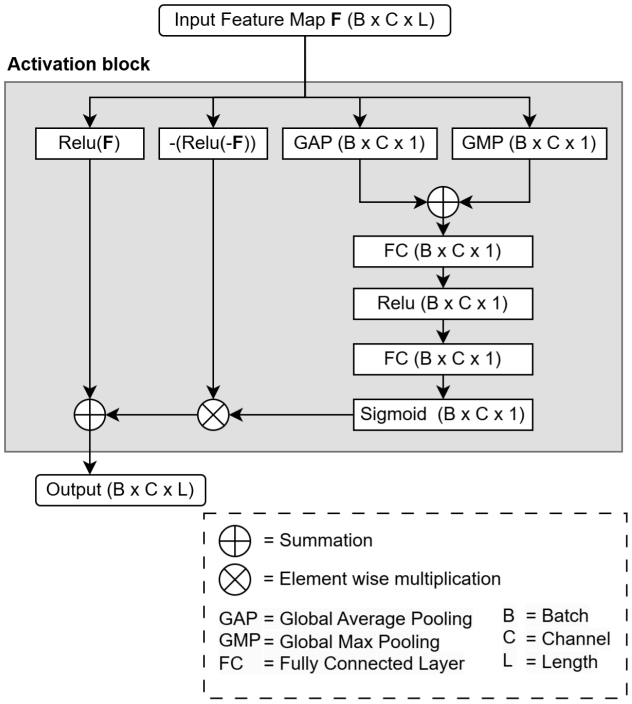


FIGURE 3. Activation block architecture.

of the autoencoder is mapping the input $\hat{\mathbf{x}}$ into several latent variables or hidden representations $\mathbf{z}$ through deterministic function $f(\hat{\mathbf{x}})$, and mapping back to the reconstructed vector $\hat{\mathbf{s}}$ through deterministic function $g(\mathbf{z})$ [11]. In denoising, the vector $\hat{\mathbf{x}}$ is a segmented, noised ECG signal, and the reconstructed vector $\hat{\mathbf{s}}$ is a predicted, segmented, clean ECG signal. The autoencoder of the proposed model can be seen in Fig. 4. The function $f(\hat{\mathbf{x}})$ is equal to a sequence of several encoder blocks, and function $g(\mathbf{z})$ is equal to a sequence of several decoder blocks. The encoder block comprises a CNN followed by a batch normalization, activation block, dropout, and maximum pooling layer. The decoder block comprises transposed 1D CNN followed by a batch normalization block, activation block, and dropout layer. The encoder and decoder blocks contain five blocks. The architecture of the activation block is illustrated in Fig. 3.

Bidirectional LSTM inspired the architecture to play a spatial or temporal attention role in skip connection for denoising PCG (phonocardiogram) signals [41]. The channel attention module at the end of each decoder block before entering the skip connection was adopted from the usage on [25]. Then, to better capture the negative value from each output of the network layer, we use an activation block adopted from [30] instead of a single activation function. The overall architecture can be seen in Fig. 4.

An ECG signal is a quasi-periodic signal characterized by a periodic pattern with slight variations in both time and amplitude [42]. Based on this characteristic we deliver three strategies. First, we use a CNN-based autoencoder as the foundational framework of the proposed model. The proposed framework is designed to generalize patterns across segmented ECG sequences [16], [19], [24], [25]. Second, we apply attention mechanisms to mitigate the loss of detailed information at the autoencoder's bottleneck. The first attention mechanism incorporates LSTM networks to enhance the skip connection input from the encoder by capturing long-term causal relationships in sequential data, such as ECG signals. The LSTM architecture used in this study employs a unidirectional LSTM-based RNN (Recurrent Neural Network) by [43]. Third, we employ a channel attention block, illustrated in Fig. 5. This module is applied at the end of each decoder block layer to emphasize essential information across the channel dimension. LSTM focuses on sequential causal-time information, the channel attention block generalizes critical features across channel or feature dimensions. The combined architecture of these strategies, depicted in Fig. 4, effectively harnesses both sequential and channel-specific information to optimize ECG signal denoising.

IV. DATASET

A. MIT-BIH ARRHYTHMIA DATABASE AND SHDB-AF DATABASE

The MIT-BIH Arrhythmia Database is a well-known database for biomedical applications, specifically for the detection and classification of arrhythmias in electrocardiogram (ECG) signals [3], [18], [44], [45]. This database was built by the Massachusetts Institute of Technology (MIT) [26]. The dataset contains 48 records of two-channel ambulatory ECG signals, each lasting an hour for each. The records were obtained from 47 subjects recorded between 1975 and 1979, representing a range of cardiac events that are mandatory for arrhythmia clinical assessment. Each record in the MIT-BIH Arrhythmia Database was acquired at a sampling rate of 360 Hz and a resolution of 11 bits.

Meanwhile, another ECG dataset was obtained from Japan. In contrast with the MIT-BIH Arrhythmia database,

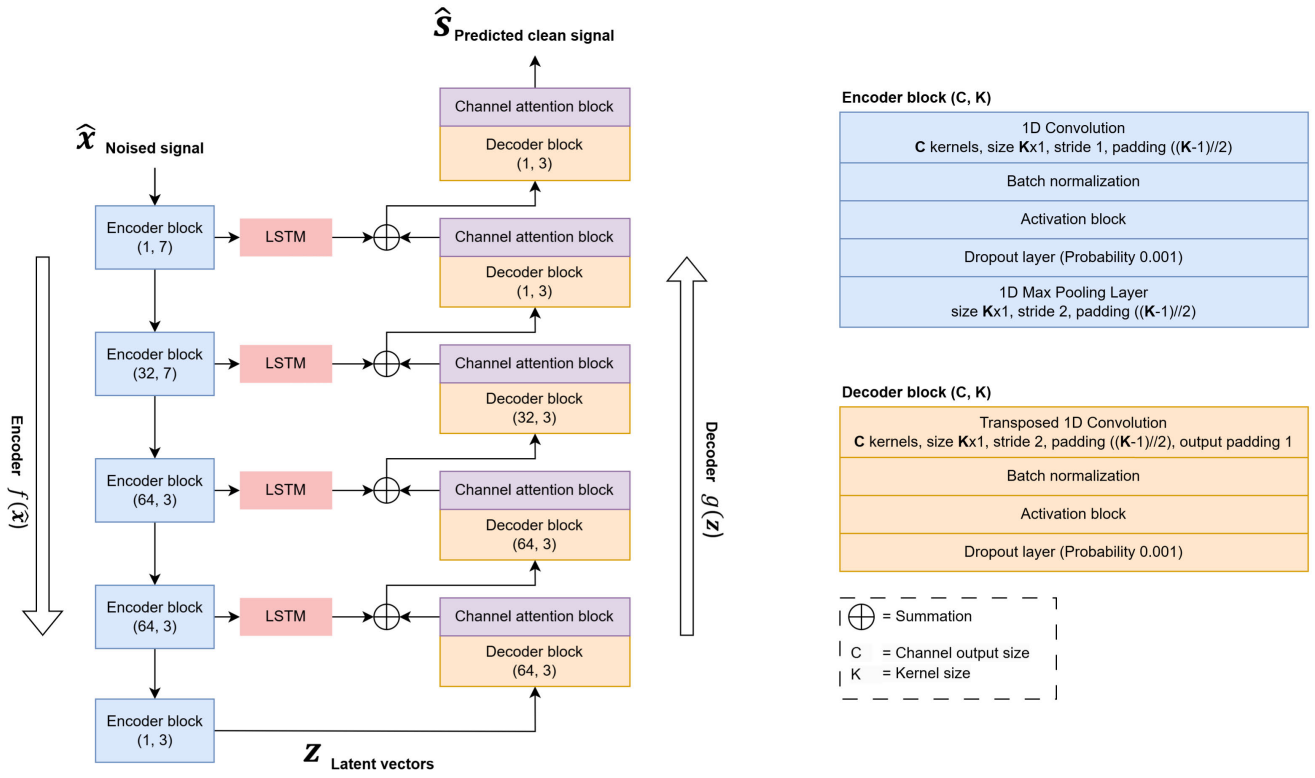


FIGURE 4. CAE-SCA architecture.

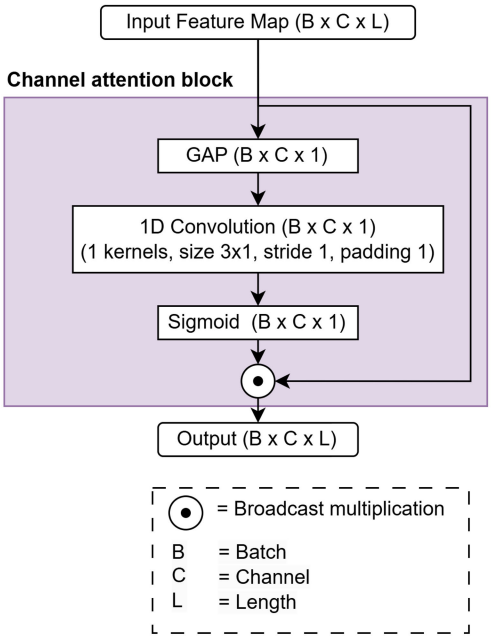


FIGURE 5. Channel attention block architecture.

SHDB-AF provides newer Holter ECG records, which particularly target atrial fibrillation (AF) detection [46]. The data were recorded with a sampling rate of 200 Hz from 100 patients with paroxysmal AF. The proposed dataset was developed to evaluate the robustness of deep learning models across various demographic characteristics, including sex, age, and ethnicity. The SHDB-AF fills a significant void in international ECG research databases because it is one of the only datasets that includes Japanese patients. In this study, both the MIT-BIH Arrhythmia and SHDB-AF databases were used for benchmarking.

B. MIT-BIH NOISE STRESS TEST DATABASE

The MIT-BIH Arrhythmia and SHDB-AF databases are well-known arrhythmia detection and classification databases. Meanwhile, this research aims to develop a novel ECG signal-denoising method. Hence, pairs of clean and noise signals are required. In this research, instead of generating noise signals synthetically, we used the MIT-BIH Noise Stress Test Database [27] since it was recorded under real-noise conditions using ambulatory electrocardiogram devices. This noise-signal generation approach has also been proposed in several other studies [3], [47], [48].

This database incorporates 12 records of ECG signals with a duration of 30 minutes each and three records of typical noise in ambulatory ECG recordings with the same duration. The noise signals were obtained using physically active participants and standard ECG devices, including recorders, electrodes, and leads. The electrodes were configured on the limbs in positions where the subjects' ECG readings could not be seen. Each of the three noise signals contains different dominant noise components, which are baseline

wander (BW), motion artifact (MA), and electrode motion (EM). These three noises are the most typical in ECG signal recordings. BW noise is caused by electrode misplacement, unwell-prepared skin, subject breathing movement, and insufficient contact between electrode cables [49]. These types of noise range from 0.05 to 3 Hz. In addition, MA noise is usually generated by the patient's body movement; thus, it interferes with the ECG signal features [50]. Finally, EM noise is commonly caused by the movement of the electrodes. This is often considered the most problematic noise because it appears as an ectopic beat. This makes EM noise more difficult to remove by simple filters than other types of noise.

C. NOISED ECG DATASET GENERATION

The MIT-BIH Arrhythmia dataset and the MIT-BIH Noise Test Stress dataset have a sampling rate of 360 Hz. The SHDB-AF dataset was sampled at 200 Hz; thus, the signal in that dataset was resampled to the MIT dataset sampling rate using the Fourier method [51]. This synchronized step ensures the temporal consistency of the input to the noise reduction model. After the sampling rates were synchronized, we generated a noised ECG signal by the Algorithm 1. This algorithm was inspired by the use of the MIT-BIH Noise Test Stress dataset in their original paper [27]. While the original approach to generate a noise signal was by defining a certain SNR value, we randomized the SNR between a certain range. This was done to obtain more diverse and various types of noise signals.

From the dataset, we obtained BW (η_{BW}), MA (η_{MA}), and EM (η_{EM}) noise. Inspired by this research [24], [49], three types of noises with specific weight a, b, c into a single noise signal. The sum of a, b, c is equal to 1. Then, the combined noised signal was merged into the clean ECG signal from the MIT-BIH Arrhythmia or SHDB-AF dataset, so the combined signal had a signal-to-noise ratio (SNR) value between -5 dB and 10 dB as in [9]. This was achieved using the following equation:

$$P_{\text{clean}} = \frac{1}{N}\sum_{i=1}^{N} s_i^2 \tag{2}$$

$$P_{\text{noise}} = \frac{1}{N}\sum_{i=1}^{N} \eta_i^2 \tag{3}$$

$$K = \sqrt{\frac{P_{\text{clean}}}{10^{\frac{\text{SNR}}{10}} \cdot P_{\text{noise}}}} \tag{4}$$

$$x(n) = s(n) + K\left[a \cdot \eta_{\text{BW}}(n) + b \cdot \eta_{\text{MA}}(n) + c \cdot \eta_{\text{EM}}(n)\right] \tag{5}$$

Algorithm 1 Noise Addition to Signal Clips

Input: clean ECG signal $s(n)$, $0 < n < N$
Output: noised ECG signal $x(n)$, $0 < n < segment\_size$

```
for i = 0 to (N−segment_size) step stride=512 do
    signal_clip ← signal[i : i + segment_size, :]
    em_clip ← em_noise[i : i + segment_size, :]
    ma_clip ← ma_noise[i : i + segment_size, :]
    bw_clip ← bw_noise[i : i + segment_size, :]
    if clip_size < segment_size then
        break
    end if
    r1, r2 ← two random numbers with sum equal to 1
    a ← r1
    b ← r2 − r1
    c ← 1 − r2
    snr ← random.randint(−5, 10)
    noise_clip ← (a × em_clip) + (b × ma_clip) + (c × bw_clip)
    power_sclean ← equation (2)
    power_snoise ← equation (3)
    noise_coef ← equation (4)
    noised_signal_clip ← equation (5)
    Save noised_signal_clip and signal_clip
end for
```

V. EXPERIMENT RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP AND DESIGN

1) EXPERIMENTAL SAMPLES

For the experiment in this research, we use 2 datasets as resources of clean ECG signals: the MIT-BIH Arrhythmia [26] dataset and the SHDB-AF dataset [46]. After all the samples from the mentioned dataset have been processed using the process in subsection IV-C, we choose all the records with channel MLII from the MIT-BIH Arrhythmia dataset and 50 records (001-050 record) with channel ECG1 from the SHDB-AF dataset to be segmented and added with some noise using Algorithm 1. The length of each noised-clean ECG segmented signal was 1024 data points, or equal to 2.84 seconds in duration. We selected this number based on the following considerations:

- The lower limit of frequency noise we want to reduce is 0.05 Hz.
- The lowest heart rate is around 0.67 Hz, but in practice, if we use frequency domain analysis, 0.05 Hz is recommended as a cut-off frequency of a high-pass filter [52].
- The larger the last segment, the higher the time and computational complexity.

The total number of segments generated from the MIT-BIH Arrhythmia dataset was 119,192, and from the SHDB-AF dataset was 63,400. The generated noise-free ECG signals from the MIT-BIH Arrhythmia dataset are used as data for training, validation, and testing at a ratio of 8:1:1. In contrast, the generated signal from the SHDB-AF dataset was used for testing only. The proposed approach was designed to evaluate the generalization capabilities of CAE-SCA, specifically, its ability to perform effectively on unseen data that were not included during the training phase.

TABLE 1. Experimental result model training (MIT-BIH Dataset).

Model	Train Loss	Val Loss MSE
CAE-SCA (Proposed)	1.505	0.005
APTrans [20]	0.842	0.0128
CNN-Attention Block [25]	1.69	0.00671
DRNN [21]	3.08	0.021
FCNDAE [16]	6.32	0.0631
GRU [22]	2.24	0.0114
LSTM [21]	1.78	0.00858
UNet [24]	1.85	0.0125

TABLE 2. Testing result on MIT-BIH dataset.

Model	SNR (dB)	RMSE	PRD (%)
CAE-SCA (Proposed)	**16.187±5.128**	**0.059±0.041**	**18.529±12.306**
APTrans [20]	11.989 ± 2.817	0.091 ± 0.048	26.578 ± 9.647
CNN-Attention Block [25]	14.089 ± 4.868	0.073 ± 0.045	23.227 ± 14.931
DRNN [21]	9.219 ± 4.592	0.125 ± 0.062	40.114 ± 24.457
FCNDAE [16]	3.533 ± 0.760	0.228 ± 0.093	66.837 ± 5.812
GRU [22]	11.806 ± 4.779	0.0934 ± 0.0501	30.035 ± 18.825
LSTM [21]	13.65 ± 5.23	0.078 ± 0.048	24.98 ± 16.87
UNet [24]	12.80 ± 5.41	0.086 ± 0.052	27.79 ± 18.88

Before the segmented, noise-free ECG signal enters the model, a minimum-maximum normalization will be applied to the segmented signals with the following equation:

$$s_i^* = \frac{s_i - s_{\min}}{s_{\max} - s_{\min}} \tag{6}$$

where s_i^* is the normalized signal of the segmented input signal, $s_{\min}$, and $s_{\max}$ are the minimum and maximum values of the segmented input signal, respectively. This normalization improves the convergence speed and stability of the model during the learning period [3].

2) LOSS FUNCTION

A distance loss function was utilized to evaluate global similarity across the entire signal and the maximum loss function to evaluate local differences [24], [53]. (7) shows the distance loss function.

$$l_{\text{dist}} = \sqrt{\sum_{i=1}^{N} \left|\hat{s}_i - s_i\right|^2} \tag{7}$$

$$l_{\max} = \max\left(|\hat{s}_1 - s_1|, |\hat{s}_2 - s_2|, \ldots, |\hat{s}_N - s_N|\right) \tag{8}$$

where l_{dist} is the loss value using the maximum loss function; it $\hat{s}_i$ is the ground truth signal's sample at time/step i, and s_i is the predicted clean signal output's sample at time i. The maximum loss value generated in (8) is notated as l_{max}, which is calculated by selecting the highest ground truth distance and predicted clean signal.

$$Loss = \lambda l_{\text{dist}} + \beta l_{\max} \tag{9}$$

These two loss functions are then combined into a single loss function by weighting them, as shown in (9), with the first λ as the weight of the distance loss and the second β as the weight of the maximum loss. The value of λ and β were configured to be 0.7 and 0.3, respectively, according to the proposed configuration used in original research [53]. We should select a value β that is not too high because it can cause the denoised signal to become unsmooth. Then, mean square error (MSE) is applied. The MSE loss is computed as the square of the RMSE, as shown in (10).

TABLE 3. Testing result on SHDB-AF dataset.

Model	SNR (dB)	RMSE	PRD (%)
CAE-SCA (Proposed)	**15.308±4.129**	**0.049±0.025**	**19.220±9.681**
APTrans [20]	13.606 ± 2.809	0.060 ± 0.032	22.031 ± 7.657
CNN-Attention Block [25]	14.317 ± 4.136	0.053 ± 0.025	21.595 ± 11.221
DRNN [21]	10.613 ± 4.341	0.083 ± 0.037	33.803 ± 20.391
FCNDAE [16]	3.688 ± 0.905	0.175 ± 0.056	65.761 ± 7.010
GRU [22]	12.68 ± 4.37	0.064 ± 0.028	26.55 ± 15.50
LSTM [21]	12.28 ± 4.11	0.067 ± 0.030	27.37 ± 14.96
UNet [24]	10.45 ± 4.72	0.082 ± 0.030	34.91 ± 20.51

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\hat{s}_i - s_i]^2} \tag{10}$$

3) PERFORMANCE METRICS

Researchers commonly use several performance measurement parameters, such as SNR, root mean square error (RMSE), and percent root mean square difference (PRD) [18], [45], [53] to assess the performance of noise reduction methods. SNR measures the ratio between the clean signal and the noise in the signal. RMSE measures the variance between the denoised and noise-free signals, and PRD measures the quality of the signal recovery. This research aimed to achieve high SNR, low RMSE, and low PRD value. The equations for SNR, RMSE, and PRD are in (11), (10), and (12) respectively.

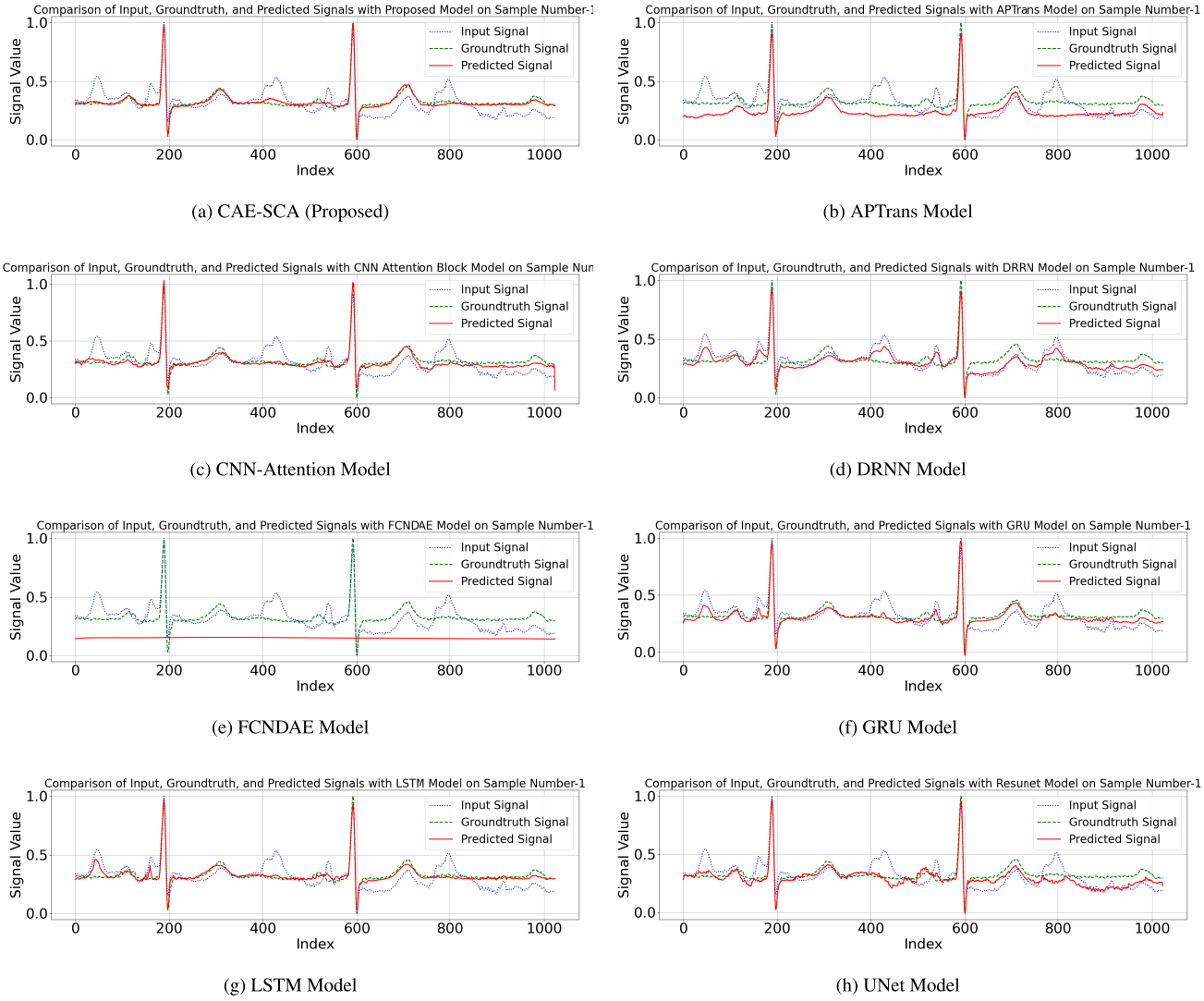
$$SNR_{\text{output}} = 10 \cdot \log_{10}\left(\frac{\sum_{i=1}^{N} s_i^2}{\sum_{i=1}^{N}(\hat{s}_i - s_i)^2}\right) \tag{11}$$

$$PRD = 100 \cdot \sqrt{\frac{\sum_{i=1}^{N}[\hat{s}_i - s_i]^2}{\sum_{i=1}^{N} s_i^2}} \tag{12}$$

In addition to model performance, computational costs were used to evaluate models that deliver good performance with efficient computation. To assess computational cost, we used FLOPs (floating point operations), memory size, and inference time [54], [55]. FLOPs estimate the total computational load by summing all arithmetic operations, such as addition, subtraction, division, and multiplication, in the algorithm on the model. In contrast, the inference time measures the actual time taken for model execution. The memory size is the estimated CPU memory consumption. The inference time and memory size are evaluated while running the model on a Raspberry Pi device to simulate edge computing deployment scenarios.

4) TEST OF SIGNIFICANCE

A test of significance was used to gather evidence from a sample of the population to evaluate a hypothesis regarding



(a) CAE-SCA (Proposed)

(b) APTrans Model

(c) CNN-Attention Model

(d) DRNN Model

(e) FCNDAE Model

(f) GRU Model

(g) LSTM Model

(h) UNet Model

FIGURE 6. **Signal denoising result visualizations.**

TABLE 4. **Performance comparison.**

Model	SNR MIT-BIH (dB)	SNR SHDB-AF (dB)	FLOPs	Total Params	Inference Raspi (ms)
CAE-SCA (Proposed)	16.187 ± 5.128	15.308 ± 4.129	437.029 M	117.83 K	345.125
APTrans [20]	11.989 ± 2.817	13.606 ± 2.809	12.497 G	6.1 M	963.174
CNN-Attention Block [25]	14.089 ± 4.868	14.317 ± 4.136	166.453 M	76.24 K	380.911
DRNN [21]	9.219 ± 4.592	10.613 ± 4.341	828.375 M	25.54 K	214.394
FCNDAE [16]	3.533 ± 0.760	3.688 ± 0.905	509.803 M	80.223 K	289.449
GRU [22]	11.806 ± 4.779	12.68 ± 4.37	6.522 G	199.69 K	2248.219
LSTM [21]	13.65 ± 5.23	12.28 ± 4.11	9.065 G	277.62 K	8836.908
UNet [24]	12.80 ± 5.41	10.45 ± 4.72	13.4825 G	1.84 M	4382.789

the population. In this study, the population consisted of ECG signals, and the sample is the test dataset selected randomly, as described in V-A1. The size of the test dataset is 11,920 which is sufficiently large to apply the central limit theorem. The theorem states that as the sample size increases, the sampling distribution of the sample mean approaches a normal distribution [56]. To strengthen the assumption of a normal distribution, the graphical distribution of the SNR values obtained from the CAE-SCA and the comparison model is examined. Based on these considerations, a right-sided t-test was applied to determine whether the CAE-SCA leads to significantly higher SNR than another model.

5) GENERAL HYPERPARAMETERS AND THE COMPUTING PLATFORM

Several hyperparameters were used for all deep learning models in this research. Each model was optimized using the Adam optimizer with a learning rate of 0.0001 and batch size of 32. The number of epochs for each training session depends on the change in validation MSE loss between the current and previous epochs. If the change is less than 0.0001 10 times in a row, the training is halted. All experiments, including the training, validation, and testing of the models, were conducted on an NVIDIA Tesla V100. We used Python 3.10 and PyTorch 2.5 as the machine learning framework.

B. MODEL EVALUATION

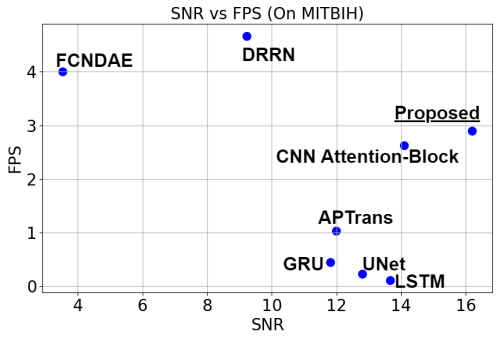
1) TRAINING RESULT

Table 1 is the overview of the training results for each tested model. The results indicate varying levels of effectiveness among different model architectures for ECG denoising on the MIT-BIH dataset. The UNet model, which incorporates residual connections into its autoencoder architecture, demonstrated lower training and validation losses than the FCNDAE model, which lacks residual connections. Among the autoencoder-based models with residual connections, excluding CAE-SCA, the CNN-Attention Block model achieved the lowest validation loss, outperforming UNet, APTrans, and DRNN. These results indicate that the integration of residual connections and attention mechanisms in both the channel and spatial dimensions significantly contributes to improving the performance of denoising models. The recurrent models, including GRU and LSTM, demonstrate reasonably balanced performance, with LSTM achieving the lowest validation loss within this group. This suggests that recurrent models can capture temporal or spatial dependencies in ECG signals, which is crucial for effective denoising.

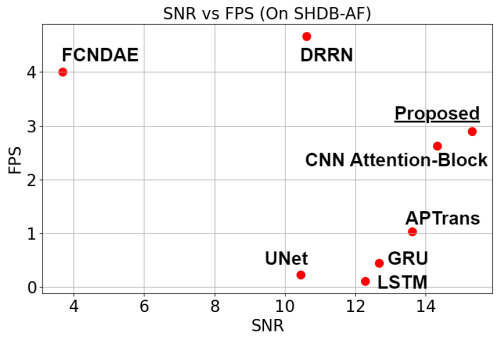
The proposed CAE-SCA, which combines CNN, LSTM, and channel attention block architectures, achieved the lowest loss values overall, which indicates that it is the most effective method for ECG denoizing in this study. Combining the spatial feature extraction capabilities of CNNs with the LSTM's ability to capture long-term dependencies enables the CAE-SCA to effectively address the complexities of ECG signals, leading to superior denoising performance. In addition, incorporating a channel attention block alongside LSTM within skip or residual connections allows the model to effectively handle both temporal and channel dependencies. The CAE-SCA model is the best-performing model, followed by CNN-Attention Block and LSTM, which also demonstrate strong generalization.

2) TESTING RESULT

To investigate generalization capabilities, we used a reserved 10% of the MIT-BIH dataset and the unseen SHDB-AF dataset. CAE-SCA performed better than other models, as evidenced by the results in Table 2 and Table 3,



(a) On MITBIH Dataset



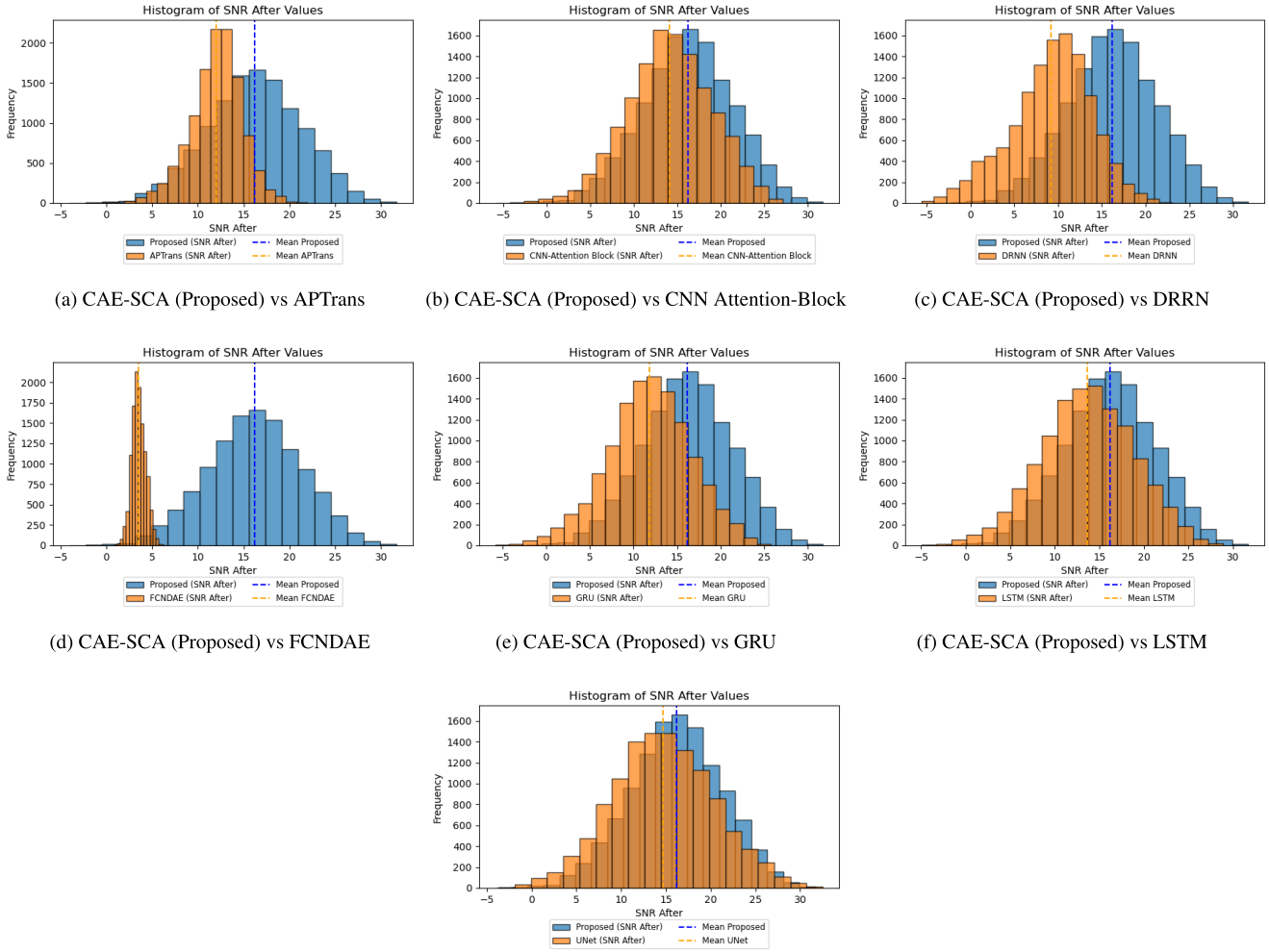
(b) On SHDB-AF Dataset

FIGURE 7. Contribution's position visualization.

which indicate an SNR of 16.187 ± 5.128 dB, RMSE of 0.059 ± 0.041, and PRD of 18.529 ± 12.306% on MIT-BIH. The performance of CNN-Attention Block and LSTM was marginally worse. The performance of APTrans, DRNN, FCNDAE, GRU, and UNet was below average. The best SNR, RMSE, and PRD values were obtained by CAE-SCA on SHDB-AF, demonstrating its resilience and versatility in unknown datasets. Fig. 6 displays the example segmented signal following denoising using the tested models.

Each model's predicted signal was compared against the input and ground truth signals to assess the accuracy of peak detection, noise suppression, and overall fidelity. The CAE-SCA demonstrated superior performance, closely aligning with the ground truth signal while effectively capturing peaks and subtle variations. This model excelled at reducing noise and maintained consistency even in regions of sudden signal spikes, highlighting its robustness and adaptability to dynamic signal changes.

Despite experiencing slight lags in high-noise regions, particularly around sample number 800, the U-Net model effectively detected peaks and preserved the overall signal structure, with notable accuracy observed near sample numbers 200 and 600, as illustrated in Fig. 6h. Although it had trouble with abrupt changes, APTrans produced smooth forecasts with good noise reduction, as shown in Fig. 6b. Accurate peak detection and balanced performance were attained by the CNN attention block, whereas minor



(a) CAE-SCA (Proposed) vs APTrans (b) CAE-SCA (Proposed) vs CNN Attention-Block (c) CAE-SCA (Proposed) vs DRRN

(d) CAE-SCA (Proposed) vs FCNDAE (e) CAE-SCA (Proposed) vs GRU (f) CAE-SCA (Proposed) vs LSTM

(g) CAE-SCA (Proposed) vs UNet

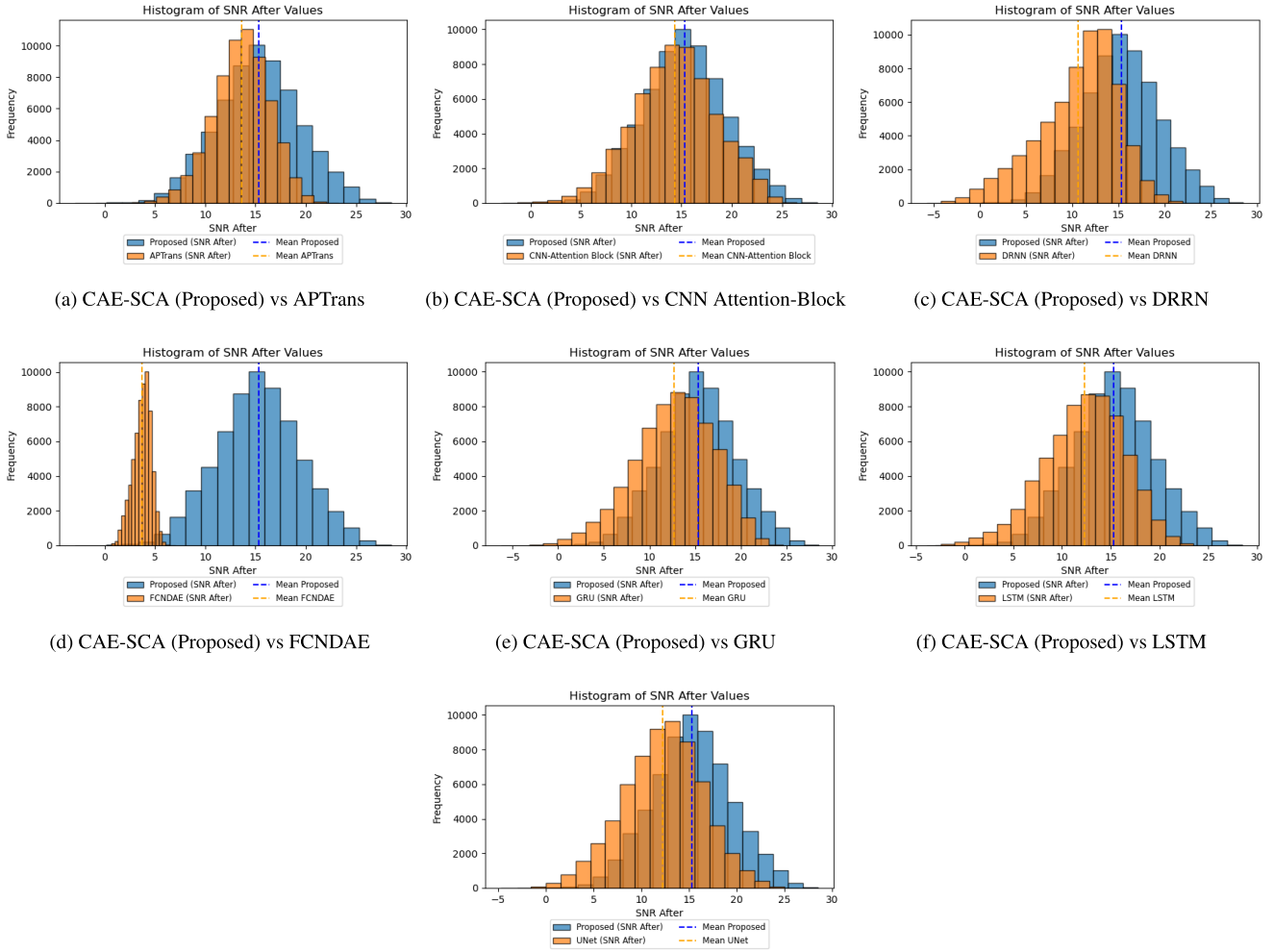
FIGURE 8. **Histogram of SNR Values for MIT-BIH Test dataset.**

variations were observed in low-amplitude areas close to sample number 700 as visualized in Fig. 6c. Although it created artifacts in low-amplitude regions, the DRRN model successfully recorded peaks, as shown in Fig. 6d. FCN-DAE, on the other hand, did not perform well, generating outputs that were flat and did not include important peaks, as visualized in Fig. 6e. Peak reconstruction was a strong suit for both the GRU and LSTM models, with the GRU handling noise well and the LSTM performing marginally better in dynamic areas, as shown in Fig. 6f and Fig. 6g.

Overall, the CAE-SCA emerged as the most reliable choice for signal reconstruction, as shown in Fig. 6a, closely followed by GRU and LSTM, which also demonstrated strong capabilities. U-Net and CNN attention blocks demonstrated promise but exhibited minor weaknesses in noise-sensitive regions, whereas the FCN-DAE model struggled significantly with the task. These findings underscore the importance of model selection in signal reconstruction applications, particularly when accuracy and noise resilience are critical. Future research could focus on hybrid approaches that integrate the strengths of these models to further improve the model performance.

3) COMPUTATIONAL COST EVALUATION

On the MIT-BIH and SHDB-AF datasets, a performance comparison was performed to assess the computational efficiency and SNR. With SNR values of 16.187 (MIT-BIH) and 15.308 (SHDB-AF), CAE-SCA is perfect for real-time edge applications. The proposed method requires only 437.029 MFLOPs, 117.83 K parameters, and 345.125 ms inference time. Although APTrans provides competitive SNR values (11.989 and 13.606), its practicality is limited by its requirements of 12.497 G FLOPs, 6.1 M parameters, and 963.174 ms inference time. The CNN-Attention Block balances efficiency and performance by achieving SNR values of 14.089 and 14.317 with 166.453 M FLOPs, 76.24 K parameters, and 380.911 ms inference time. Although they



(a) CAE-SCA (Proposed) vs APTrans (b) CAE-SCA (Proposed) vs CNN Attention-Block (c) CAE-SCA (Proposed) vs DRRN

(d) CAE-SCA (Proposed) vs FCNDAE (e) CAE-SCA (Proposed) vs GRU (f) CAE-SCA (Proposed) vs LSTM

(g) CAE-SCA (Proposed) vs UNet

FIGURE 9. **Histogram of SNR Values for SHDB-AF Test dataset.**

struggle with complexity, GRU (6.522 G FLOPs, 199.69 K parameters, 2248.219 ms) and LSTM (9.065 G FLOPs, 277.62 K parameters, 8836.908 ms) provide moderate performance.

The other models with poor performance included DRNN, FCNDAE, and UNet. With 828.375 M FLOPs, 25.54 K parameters, and a quick inference time of 214.394 ms, the proposed DRNN obtained SNR values of 9.219 and 10.613. With 509.805 M FLOPs, 80.223 K parameters, and 289.449 ms inference, FCNDAE exhibits poor SNR values (3.533 and 3.688). Despite its quick inference (4382.789 ms) and low parameters (1.84 M), UNet yields moderate outcomes but requires 13.482 G FLOPs. With all the factors considered, CAE-SCA exhibits the best balance between accuracy, efficiency, and inference speed, which makes it ideal for use in settings with limited resources.

Fig. 7 represents the position of the CAE-SCA's contribution compared to other models by visualizing the relation between the obtained SNR value and calculated frames per second (FPS) value based on inference time in Raspberry Pi 5 in both MITBIH and SHDB-AF datasets, respectively. The word frame refers to the segment for each input fed to the model in each inference. The FPS value is calculated by dividing one by the inference time, resulting the total inference frequency happened in one second. These results shed light on the trade-off between denoising performance (SNR) and computational efficiency. On both datasets, the CAE-SCA achieved a notable balance, exhibiting high SNR while maintaining competitive FPS. Overall, the CAE-SCA obtained the highest SNR value compared to the other methods. Although it is still behind FCNDAE and DRRN in terms of FPS, it still achieves commendable performance with approximately 3 FPS, which suggests its suitability for real-time application in embedded devices.

4) TEST OF SIGNIFICANCE EVALUATION

The proposed CAE-SCA achieved the best average value across all metrics in the model performance evaluation.

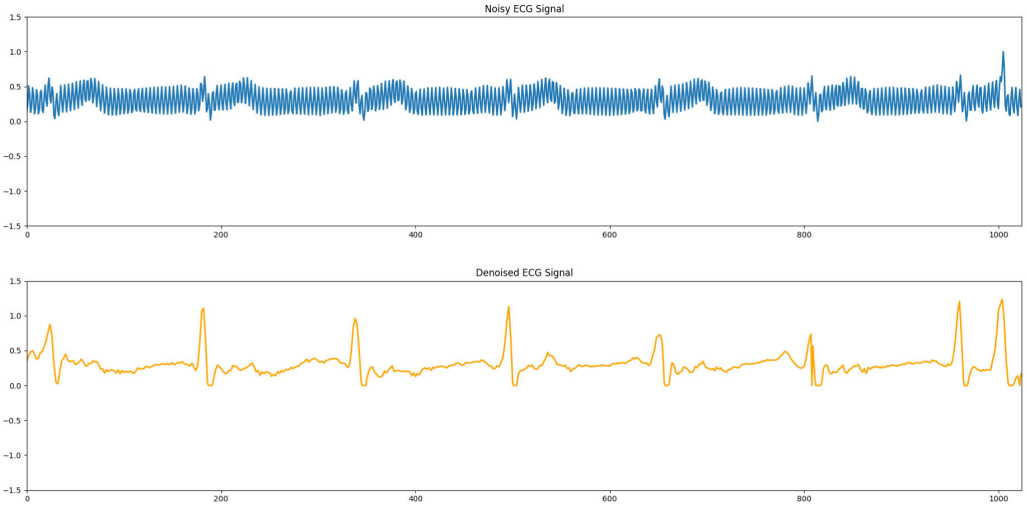


FIGURE 10. Hardware implementation test result.

TABLE 5. Significance test result.

Model A	Model B	p value (MIT-BIH)	p-value (SHDB-AF)
CAE-SCA (Proposed)	APTrans [20]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
CAE-SCA (Proposed)	CNN-Attention Block [25]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
CAE-SCA (Proposed)	DRNN [21]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
CAE-SCA (Proposed)	FCNDAE [16]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
CAE-SCA (Proposed)	GRU [22]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
CAE-SCA (Proposed)	LSTM [21]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$
CAE-SCA (Proposed)	Unet [24]	$< 1 \times 10^{-10}$	$< 1 \times 10^{-10}$

We conducted a significance test to provide evidence supporting the superiority of the CAE-SCA over the compared models. Fig. 8 and Fig. 9 illustrate the differences in histograms between the CAE-SCA and other models, revealing that the distribution of SNR values for each model approximates a normal distribution; however, the CAE-SCA exhibits a higher mean. This allowed us to perform a right-sided independent sample t-test. The results, shown in Table 5, indicate that all p-values for SNR comparisons between the CAE-SCA and each baseline model are well below the chosen significance level of 0.001. This confirms the statistically significant difference in the mean SNR values between the CAE-SCA and the other models. In particular, CAE-SCA leads to significantly higher SNR than the alternative models.

C. REAL HARDWARE IMPLEMENTATION TEST

As stated previously, we would like to implement the proposed model in an embedded Holter monitor. We powered our ECG acquisition circuit with two 9-volt batteries for the positive and negative signal ranges. The Raspberry Pi was also powered as the main processor with a power bank. In the test scenario, we used a smartphone as the end-user device to visualize the signal. Fig. 10 shows the visualization of the real raw captured signal from our ECG acquisition device and the filtered signal using our proposed ECG denoising model. As shown in Fig. 10, the raw signal (above) contains a very noisy signal. We cannot find the ECG signal features P, Q, R, S, and T easily. Meanwhile, the filtered signal (bottom) shows a clear visualization of a common ECG signal, allowing the signal to be analyzed more easily.

VI. CONCLUSION

This study aims to develop a novel, strong, and efficient denoising autoencoder architecture for edge devices. The performance experiments were conducted on both a cloud environment and Raspberry Pi 5, with the latter serving as a representative edge device for testing in the context of Holter monitor implementation. While maintaining competitive computational efficiency, the CAE-SCA has outperformed other state-of-the-art models in terms of model performance metrics. The incorporation of LSTM in the skip connections as a sequential attention mechanism enables CAE-SCA to effectively capture the causal dependencies of ECG signals. The CNN-based channel attention mechanism and the overall autoencoder architecture address the quasi-periodic nature of ECG signals, thereby enhancing the model's performance in denoising tasks. In the future, it will be beneficial to test the ECG acquisition circuit under different user activities and disturbances. For instance, testing during walking, sitting, standing, and other activities would provide a more comprehensive validation of the CAE-SCA's robustness and reliability.

ACKNOWLEDGMENT

The authors would like to express their gratitude to ChatGPT from OpenAI for providing valuable assistance in refining their manuscript. The tool's suggestions for language clarity, structure, and coherence significantly enhanced the quality of the study. Additionally, they extend their sincere appreciation to the Katalis Program from the Ministry of Education,

Culture, Research, and Technology of Indonesia for their generous support. While the final interpretations and analyses remain their own, support from both sources was instrumental in the completion of this research.

REFERENCES

[1] *BMC Public Health—Bmcpublichealth.biomedcentral.com*. Accessed: Dec. 17, 2024. [Online]. Available: https://bmcpublichealth.biomedcentral.com/

[2] A. Rasti-Meymandi and A. Ghaffari, "A deep learning-based framework for ECG signal denoising based on stacked cardiac cycle tensor," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103275.

[3] X. Wang, B. Chen, M. Zeng, Y. Wang, H. Liu, R. Liu, L. Tian, and X. Lu, "An ECG signal denoising method using conditional generative adversarial net," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 7, pp. 2929–2940, Jul. 2022.

[4] A. Azzouz, B. Bengherbia, P. Wira, N. Alaoui, A. Souahlia, M. Maazouz, and H. Hentabeli, "An efficient ECG signals denoising technique based on the combination of particle swarm optimisation and wavelet transform," *Heliyon*, vol. 10, no. 5, Mar. 2024, Art. no. e26171.

[5] *Cardiovascular Diseases (CVDs)—Who.int*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[6] W. S. P. Harmadha, F. R. Muharram, R. S. Gaspar, Z. Azimuth, H. A. Sulistya, F. Firmansyah, C. E. C. Z. Multazam, M. Harits, and R. M. Putra, "Explaining the increase of incidence and mortality from cardiovascular disease in Indonesia: A global burden of disease study analysis (2000–2019)," *PLoS ONE*, vol. 18, no. 12, Dec. 2023, Art. no. e0294128.

[7] S.-K. Chua, L.-C. Chen, L.-M. Lien, H. Lo, Z.-Y. Liao, S.-P. Chao, C.-Y. Chuang, and C.-Z. Chiu, "Comparison of arrhythmia detection by 24-hour Holter and 14-day continuous electrocardiography patch monitoring," *Acta Cardiologica Sinica*, vol. 36, no. 3, pp. 251–259, May 2020.

[8] P. M. Tripathi, A. Kumar, R. Komaragiri, and M. Kumar, "A review on computational methods for denoising and detecting ECG signals to detect cardiovascular diseases," *Arch. Comput. Methods Eng.*, vol. 29, no. 3, pp. 1875–1914, May 2022.

[9] S. Chatterjee, R. S. Thakur, R. N. Yadav, L. Gupta, and D. K. Raghuvanshi, "Review of noise removal techniques in ECG signals," *IET Signal Process.*, vol. 14, no. 9, pp. 569–590, Dec. 2020.

[10] P. Bing, W. Liu, Z. Zhai, J. Li, Z. Guo, Y. Xiang, B. He, and L. Zhu, "A novel approach for denoising electrocardiogram signals to detect cardiovascular diseases using an efficient hybrid scheme," *Frontiers Cardiovascular Med.*, vol. 11, Apr. 2024, Art. no. 1277123.

[11] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2008, pp. 1096–1103.

[12] D. Chen, S. Shi, X. Gu, and B. Shim, "Robust DoA estimation using denoising autoencoder and deep neural networks," *IEEE Access*, vol. 10, pp. 52551–52564, 2022.

[13] M. Scarpiniti, S. S. Ahrabi, E. Baccarelli, L. Piazzo, and A. Momenzadeh, "A novel unsupervised approach based on the hidden features of deep denoising autoencoders for COVID-19 disease detection," *Expert Syst. Appl.*, vol. 192, Apr. 2022, Art. no. 116366.

[14] C. Gheller and F. Vazza, "Convolutional deep denoising autoencoders for radio astronomical images," *Monthly Notices Roy. Astronomical Soc.*, vol. 509, no. 1, pp. 990–1009, Oct. 2021.

[15] Z. Shang, L. Sun, Y. Xia, and W. Zhang, "Vibration-based damage detection for bridges by deep convolutional denoising autoencoder," *Struct. Health Monitor.*, vol. 20, no. 4, pp. 1880–1903, Jul. 2021.

[16] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise reduction in ECG signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60806–60813, 2019.

[17] H. Rahman, T. F. N. Bukht, A. Imran, J. Tariq, S. Tu, and A. Alzahrani, "A deep learning approach for liver and tumor segmentation in CT images using ResUNet," *Bioengineering*, vol. 9, no. 8, p. 368, Aug. 2022.

[18] Y. Hou, R. Liu, M. Shu, and C. Chen, "An ECG denoising method based on adversarial denoising convolutional neural network," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104964.

[19] J. Wang, S. Pei, Y. Yang, and H. Wang, "Convolutional transformer-driven robust electrocardiogram signal denoising framework with adaptive parametric ReLU," *Math. Biosci. Eng.*, vol. 21, no. 3, pp. 4286–4308, 2024.

[20] K. Antczak, "Deep recurrent neural networks for ECG signal denoising," 2018, *arXiv:1807.11551*.

[21] M. Dias, P. Probst, L. Silva, and H. Gamboa, "Cleaning ECG with deep learning: A denoiser tested in industrial settings," *Social Netw. Comput. Sci.*, vol. 5, no. 6, p. 699, Jul. 2024.

[22] I. Banerjee, Y. Ling, M. C. Chen, S. A. Hasan, C. P. Langlotz, N. Moradzadeh, B. Chapman, T. Amrhein, D. Mong, D. L. Rubin, O. Farri, and M. P. Lungren, "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artif. Intell. Med.*, vol. 97, pp. 79–88, Jun. 2019.

[23] Z. Chen, M. Wang, M. Zhang, W. Huang, H. Gu, and J. Xu, "Post-processing refined ECG delineation based on 1D-UNet," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104106.

[24] W. Chorney, H. Wang, L. He, S. Lee, and L.-W. Fan, "Convolutional block attention autoencoder for denoising electrocardiograms," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105242.

[25] P. Singh and A. Sharma, "Attention-based convolutional denoising autoencoder for two-lead ECG denoising and arrhythmia classification," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[26] G. B. Moody and R. G. Mark, "MIT-BIH arrhythmia database," Tech. Rep., 1992.

[27] G. B. Moody, W. E. Muldrow, and R. G. Mark, "The MIT-BIH noise stress test database," Tech. Rep., 1992.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.

[29] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, May 2016, Art. no. 160035.

[30] M. Zhao, S. Zhong, X. Fu, B. Tang, S. Dong, and M. Pecht, "Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2587–2597, Mar. 2021.

[31] G. B. Moody and R. G. Mark, "MIT-BIH atrial fibrillation database," Tech. Rep., 1992.

[32] L. Xiu and Z. Li, "Low-power instrumentation amplifier IC design for ECG system applications," in *Proc. Eng.*, vol. 29, pp. 1533–1538, Jan. 2012.

[33] A. Wong, K.-P. Pun, Y.-T. Zhang, and C.-S. Choy, "An ECG measurement IC using driven-right-leg circuit," in *Proc. IEEE Int. Symp. Circuits Syst.*, Apr. 2006, pp. 345–348.

[34] W. Wang, "Design of low-noise low-power ECG amplifier circuit with high integration level," *J. Phys., Conf. Ser.*, vol. 2649, no. 1, Nov. 2023, Art. no. 012062.

[35] B. B. Winter and J. G. Webster, "Driven-right-leg circuit design," *IEEE Trans. Biomed. Eng.*, vol. BME-30, no. 1, pp. 62–66, Jan. 1983.

[36] V. K. Singh, W.-G. Ho, and R. Gharpurey, "A frequency-folded ADC channelizer with digital equalization and relaxed anti-alias filtering," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 7, pp. 2304–2317, Jul. 2018.

[37] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.

[38] C. Chang, "Maximally flat amplitude low-pass filter with an arbitrary number of pairs of real frequency transmission zeros," *IEEE Trans. Circuit Theory*, vol. CT-15, no. 4, pp. 465–467, Apr. 1968.

[39] A. W. Ramadhan, A. Wijayanto, and H. Oktavianto, "Implementation of audio event recognition for the elderly home support using convolutional neural networks," in *Proc. Int. Electron. Symp. (IES)*, Sep. 2020, pp. 91–95.

[40] N. P. Martono and H. Ohwada, "Evaluating the impact of windowing techniques on Fourier transform-preprocessed signals for deep learning-based ECG classification," *Hearts*, vol. 5, no. 4, pp. 501–515, Oct. 2024.

[41] S. N. Ali, S. B. Shuvo, Md. I. S. Al-Manzo, A. Hasan, and T. Hasan, "An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise," *IEEE Access*, vol. 11, pp. 87887–87901, 2023.

[42] Alsaade and Bhuiyan, "On development of fuzzy filter for de-noising ECG signal," *Int. J. Comput. Sci. Netw. Secur.*, vol. 20, no. 12, pp. 252–257, Dec. 2020.
[43] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," 2014, *arXiv:1402.1128*.
[44] H. Li and P. Boulanger, "A survey of heart anomaly detection using ambulatory electrocardiogram (ECG)," *Sensors*, vol. 20, no. 5, p. 1461, Mar. 2020.
[45] I. Hermawan, N. Sevani, A. F. Abka, and W. Jatmiko, "Denoising ambulatory electrocardiogram signal using interval dedependent thresholds based stationary wavelet transform," *JOIV, Int. J. Informat. Visualizat.*, vol. 8, no. 2, p. 742, May 2024.
[46] K. Tsutsui, S. B. Brimer, N. Ben-Moshe, J. M. Sellal, J. Oster, H. Mori, Y. Ikeda, T. Arai, S. Nakano, R. Kato, and J. A. Behar, "SHDB-AF: A Japanese Holter ECG database of atrial fibrillation," 2024, *arXiv:2406.16974*.
[47] X. Gu, J. Hu, L. Zhang, J. Ding, and F. Yan, "An improved method with high anti-interference ability for R peak detection in wearable devices," *IRBM*, vol. 41, no. 3, pp. 172–183, Jun. 2020.
[48] H. Li, G. Ditzler, J. Roveda, and A. Li, "DeScoD-ECG: Deep score-based diffusion model for ECG baseline wander and noise removal," *IEEE J. Biomed. Health Informat.*, vol. 28, no. 9, pp. 5081–5091, Sep. 2024.
[49] F. P. Romero, D. C. Piñol, and C. R. Vázquez-Seisdedos, "DeepFilter: An ECG baseline wander removal filter using deep learning techniques," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 102992.
[50] S. Banerjee and G. K. Singh, "Monte Carlo filter-based motion artifact removal from electrocardiogram signal for real-time telecardiology system," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
[51] W. G. Hawkins, "Fourier transform resampling: Theory and application," *IEEE Trans. Nucl. Sci.*, vol. 44, no. 4, pp. 1543–1551, Aug. 1997.
[52] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis, O. Pahlm, P. Rautaharju, and G. S. Wagner, "Recommendations for the standardization and interpretation of the electrocardiogram," *Circulation*, vol. 115, no. 10, pp. 1306–1324, Mar. 2007.
[53] J. Wang, R. Li, R. Li, K. Li, H. Zeng, G. Xie, and L. Liu, "Adversarial denoising of electrocardiogram," *Neurocomputing*, vol. 349, pp. 212–224, Jul. 2019.
[54] R. Desislavov, F. Martínez-Plumed, and J. Hernández-Orallo, "Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning," *Sustain. Computing: Informat. Syst.*, vol. 38, Apr. 2023, Art. no. 100857.
[55] C. Scheerlinck, H. Rebecq, D. Gehrig, N. Barnes, R. E. Mahony, and D. Scaramuzza, "Fast image reconstruction with an event camera," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 156–163.
[56] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 9th ed., Boston, MA, USA: Cengage Learning, 2016.

ALIF WICAKSANA RAMADHAN received the bachelor's and master's degrees from the Department of Electronics Engineering, Electronic Engineering Polytechnic Institute of Surabaya. He is currently a Research Assistant with the Faculty of Computer Science, University of Indonesia. His research interests include computer vision, robotics, and machine learning.

SYIFA KUSHIRAYATI received the bachelor's degree in biomedical engineering from Bandung Institute of Technology, Indonesia. She is currently a Research Assistant with the Faculty of Computer Science, University of Indonesia. Her research interests include digital signal processing and artificial intelligence.

SALSABILA AURELLIA received the bachelor's degree in biomedical engineering from Airlangga University and the master's degree in biomedical engineering from the University of Indonesia. She is currently a Research Assistant with the Faculty of Computer Science, University of Indonesia. Her research interests include artificial intelligence and biomedical signal and image processing.

MGS M. LUTHFI RAMADHAN received the bachelor's degree from the Faculty of Computer Science, Sriwijaya University, and the master's degree from the Faculty of Computer Science, University of Indonesia. He is currently a Research Assistant with the Faculty of Computer Science, University of Indonesia. His research interests include deep learning, pattern recognition, and computer vision.

MUHAMMAD HANNAN HUNAFA received the bachelor's degree from the Faculty of Informatics, Telkom Institute of Technology Purwokerto. He is currently pursuing the master's degree with the Faculty of Computer Science, University of Indonesia. His research interests include machine learning, deep learning, and computer vision.

MUHAMMAD FEBRIAN RACHMADI is currently a Faculty Member with the Faculty of Computer Science, University of Indonesia. He is also affiliated as a Postdoctoral Research Scientist with the Brain Image Analysis (BIA) Unit, RIKEN Center of Brain Science (RIKEN CBS), Wako, Japan, working together with Dr. Henrik Skibbe, and the Team Leader of the BIA Unit, to develop machine/deep learning techniques for processing and analyzing marmoset brain image data. His main research interests include medical image analysis and computation using data-driven methods, such as deep learning algorithms.

APRINALDI JASA MANTAU (Member, IEEE) received the Ph.D. degree in computer science and systems engineering from Kyushu Institute of Technology, Japan, in 2024. He is currently a Lecturer with the Faculty of Computer Science, University of Indonesia. He is a Researcher and a Lecturer in the field of computer science. His research interests include machine learning, computer vision, robotics, data mining, swarm intelligence, and swarm robotics.

SITI NURMAINI (Member, IEEE) received the master's degree in control systems from Institut Teknologi Bandung (ITB), Indonesia, in 1998, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia (UTM), in 2011. She is currently a Professor with the Faculty of Computer Science, Sriwijaya University. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotics.

SATRIA MANDALA (Member, IEEE) received the Bachelor of Science degree in electrical engineering from Brawijaya University, Indonesia, in 1997, and the master's degree in computer science with a specialization in network security and the Ph.D. degree in computer science, with a focus on wireless network security from Universiti Teknologi Malaysia (UTM), in 2006 and 2011, respectively. He has also held the position of the Dean of the Faculty of Information and Industrial Technology, Telkom Institute of Technology Surabaya. He is currently the Director of the Human-Centric Engineering, Telkom University, Indonesia. He is also a Lecturer with the School of Computing, Telkom University.

WISNU JATMIKO (Senior Member, IEEE) received the B.S. degree in electrical engineering and the M.Sc. degree in computer science from the University of Indonesia, Depok, Indonesia, in 1997 and 2000, respectively, and the Dr.Eng. degree from Nagoya University, Japan, in 2007. He is currently a Full Professor with the Faculty of Computer Science, University of Indonesia. His research interests include autonomous robots, optimization, real-time traffic monitoring systems, machine learning, and artificial intelligence. From 2019 to 2020, he was the Chairman of the IEEE Indonesia Section.