Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEEP CONVOLUTIONAL AUTOENCODERS WITH ADAPTIVE GATING FOR HIGH-FIDELITY ECG DENOISING UNDER SEVERE NOISE SATURATION: A COMPARATIVE STUDY

EIHAB BASHIER [1,2*], TOUFIK BEN JABEUR [3], HASSAN AL REASI [4], HAMAD AL REESI [5], TAISER TAHA [6]

[1] Department of Mathematics, Sohar University, Sohar, Oman, ebashier@su.edu.om
[2] Dept.of Applied Maths, Faculty of Math. Sciences and Informatics, Univ. of hartoum, Khartoum, Sudan.
[3] College of Arts and Applied Sciences, Dhofar University, Salalah, Oman.
[4] Biology Education Section, Faculty of Education and Arts, Sohar University, Sohar, Oman.
[5] Dept. of Plan. and Stud., Direct. Gen. of Health Services, North Batinah Governorate.
[6] Dept. of Comp. and Information Sciences, Univ. of Tech. and Applied Sciences, Sohar, Oman,
Email: [1,2*]ebashier@su.edu.om, [3]tjabeur@du.edu.om, [4]hreasi@su.edu.om, [5]abuessa06@gmail.com, [6]Taiser.taha@utas.edu.om

ABSTRACT

Electrocardiogram (ECG) signals are frequently corrupted by strong clinical artifacts like muscle noise and baseline wander, rendering automated diagnostic systems ineffective. While deep learning models filter signals successfully, preserving critical physiological features like R-peaks under high noise saturation remains a major challenge. We propose a new data-driven, two-stage deep learning framework for robust ECG denoising across a wide range of signal-to-noise ratios (SNRs), from -20 dB to $+20$ dB. The system features a built-in engine for morphological parameter minimization using analytical inverse-variance scaling to balance mean squared error against first- and second-order derivative penalties. Additionally, an empirical-adaptive shortcut gating mechanism dynamically regulates internal layer transitions based on the real-time detected noise floor. We present a comprehensive comparative study evaluating three architectural configurations combined with a Stage 2 Deep Residual Refinement network: Stacked Convolutional Autoencoders (Stacked ConvAE), Base Convolutional Autoencoders (Base ConvAE), and U-Net Convolutional Autoencoders (U-Net ConvAE). Models were trained on true ECG data from the MIT-BIH Arrhythmia Database synthesized with authentic clinical noise profiles from the Noise Stress Test Database (NSTDB). Experimental results demonstrate strong performance across all noise regimes. In an extreme -20 dB noise environment, the U-Net ConvAE framework recovers the signal to a post-refined SNR of $+10.16$ dB and a morphological correlation coefficient of 0.93. Crucially, all tested autoencoders successfully restore the cardiac peak detection rate from a failing 0.53 to a highly reliable 0.90.

Keywords: *ECG Signal Denoising, Convolutional Autoencoders, Deep Residual Networks, Clinical Feature Preservation, Automated QRS Detection.*

1 INTRODUCTION

The electrocardiogram (ECG) is one of the most widely used non-invasive diagnostic tools in clinical medicine and provides crucial information about the patients cardiac electrophysiology [1]. Correct interpretation of the morphological characteristics of the ECG wave complex, namely the P-wave, the QRS complex and the T-wave, is important to detect life-threatening arrhythmias, myocardial infarctions and chronic cardiovascular abnormalities [2]. However, in real-world clinical acquisition settings, whether ambulatory monitoring or intensive care testing, ECG signals are invariably corrupted by various physiological and environmental artifacts [3]. High frequency electromyographic (EMG) muscle interference, low frequency baseline wander caused by the respiration of the patient or the movement of the electrode, powerline interference and electrode motion artifacts are the main types of these disturbances [4]. The presence of these intensive noise matrices deteriorates the signal quality significantly, often masking subtle diagnostic features, and leading to false alarms or complete failure of automated cardiac monitoring systems [5].

The biggest challenge in automated ECG denoising lies in the elimination of extreme noise levels while carefully preserving the clinical morphology of the underlying cardiac waveform. Classical filtering methods like finite impulse response (FIR) filters, infinite impulse response (IIR) filters, adaptive filtering, and wavelet transformations are strongly dependent on pre-specified frequency thresholds [6]. The techniques work well in moderate noise conditions, but have serious limitations when the signal and noise spectrums are highly overlapped. Severe noise saturation causes traditional filters to blur sharp structural edges, leading to phase distortions and decrease in amplitude of R-peaks [7, 8]. The loss of these high frequency features severely limits the downstream clinical processing pipelines, e.g. heart rate variability (HRV) analysis and automated R-peak detection [9].

Recently, deep learning models, especially deep convolutional autoencoders (ConvAEs), have become powerful alternatives due to their capability of learning complex non-linear mappings directly from raw data without the need of hand-crafted features [10]. That said, common deep autoencoders suffer from a structural trade-off, where deep architectures capable of aggressive noise suppression often over-smooth fine physiological details, while shallower models tend to preserve morphology at the cost of residual noise tracking artifacts in the final output [11]. The main aim of this study is to develop, evaluate and compare robust, multi-architecture deep learning frameworks capable of reconstructing ECGs with high fidelity under harsh noise conditions, in order to overcome these structural limitations.

To overcome these structural limitations, the main objective of this work is to develop, test and compare highly robust, multi-architecture deep learning frameworks for high-fidelity ECG reconstruction on severe noise conditions. Specifically, the purpose of this work is to design a dualstage denoising architecture to decouple the task of global noise removal and structural feature refinement. Furthermore, the objective is to design an optimization paradigm that adaptively adjusts the internal data routing of the network and the optimization criteria according to the intrinsic geometric and noise characteristics of the signal, without requiring empirical tuning of the hyperparameters.

Thus, this paper provides a number of unique contributions in the area of biomedical signal processing. We first propose a novel analytical inverse-variance scaling optimization engine that automatically computes the mathematical weights of first and second order derivative penalties within an advanced morphological loss function. This mechanism ensures the structure curvature tracking and mean squared error w.r.t. only the native energy distribution of the clean cardiac targets, thus preventing the optimization from collapsing under extreme noise inflation. Second, we develop and incorporate an empirical-adaptive shortcut gating mechanism into the network layers. This gating system can measure the true physical noise floor at run-time and dynamically modulate the signal flow for finer physiological components to bypass aggressive downsampling layers when cleaner signals are detected. 3. We conduct and report a comprehensive comparative study over three structurally different autoencoder paradigms: Stacked Convolutional Autoencoders, Base Convolutional Autoencoders and U-Net Convolutional Autoencoders each uniquely coupled with a secondary Stage 2 deep residual refinement network to map out the optimal architectural trade-offs under severe clinical noise. The method we use in this study is a robust experimental pipeline that is validated on well-known benchmark databases. The raw and clean electrocardiogram profiles are obtained from the MIT-BIH Arrhythmia Database. The real clinical noise artifacts, as the combined muscle noise and electrode motion tracking, are taken from the Noise Stress Test Database (NSTDB). The signals are synthesized for a wide range of input signal-to-noise ratios (SNRs), from a highly destructive regime of -20 dB to a normal environment of +20 dB. A subset of the target sample is passed through the inverse-variance scaling engine prior to training, to ensure optimal morphological loss coefficients. The networks are trained with a primary Stage 1 autoencoder to achieve large scale noise suppression with a custom Gaussian weighted peak preservation matrix . The intermediate outputs are then passed to a Stage 2 Deep Residual Refinement network that recovers the attenuated structural peaks and removes residual distortions. The assessment is performed using a comprehensive suite of clinical performance metrics, including output SNR, Percentage Root-Mean-Square Difference (PRD), morphological correlation coefficients, and automated cardiac peak detection rates.

The rest of this paper is organized as follows. Section 2 details the mathematical foundations of the proposed framework, encompassing the data synthesis pipeline, the intrinsic morphological optimization engine, the adaptive shortcut gating design, and the structural setups of the three comparative convolutional autoencoders and the residual refinement network. Section 3 provides quantitative experimental results for all the SNR thresholds tested including comparative statistical performance tables and visual waveform traces. In Section 4, we discuss in detail the architectural trade-offs, the clinical importance of preserving high-frequency features, and compare with the state-of-art denoising techniques. Section 5 concludes the paper and discusses directions for future research.

2 BACKGROUND AND LITERATURE REVIEW

Over the past few decades, the removal of clinical artifacts from electrocardiogram (ECG) recordings has been one of the active areas of investigation evolving from classical signal processing filters to state-of-the-art data-driven deep learning architectures [3]. Traditionally, the suppression of high-frequency electromyographic noise, powerline interference and low-frequency baseline wander was based on digital linear filtering techniques (finite impulse response (FIR) and infinite impulse response (IIR) configurations) or adaptive filters, based on least mean squares (LMS) and recursive least squares (RLS) algorithms. These conventional paradigms are computationally efficient but rely on assumptions of stationary signals and rigid frequency cutoffs, often attenuating important high frequency cardiac components like the R-peak, and introduce phase distortions that may be detrimental to clinical diagnoses.

To get rid of non-stationary signal components, the researchers introduced the empirical mode decomposition (EMD) and wavelet transforms which decompose the ECG waveform into localized time-frequency representations [6]. Wavelet thresholding [7] and variational mode decomposition (VMD) [12] have better performance in adapting to transient artifacts, but still suffer from the drawbacks of hand-crafted parameters, empirical threshold selection and structural distortion or mode-mixing in heavy noise conditions.

The arrival of deep learning has shifted the baseline for biomedical signal restoration by replacing manual feature extraction with automatic end-to-end mapping functions [10]. Early neural network interventions have leveraged deep fully connected layers and recurrent neural networks (RNNs) to capture temporal sequence dependencies. However, deep fully connected structures suffer from the problem of parameter explosion and ignoring local temporal correlations [13]. Conventional long short-term memory (LSTM) and gated recurrent unit (GRU) models suffer from the problems of high computational latency and vanishing gradients over long clinical signal horizons. Deep convolutional neural networks (CNNs) have been widely used to extract spatial-temporal patterns across localized segments of the cardiac waveform [4]. A fully convolutional network, in the form of a denoising autoencoder (DAE), has been shown to successfully separate clinical noise from the underlying cardiac profiles, by a large margin than classical digital filters and fully connected neural models, in terms of the output signal-to-noise ratio (SNR) improvements [14].

Based on these convolutional foundations, recent research has put great emphasis on optimizing the architecture of denoising autoencoders to prevent over-smoothing of crucial morphological waves like the P-wave, QRS complex, and T-wave [15]. Modern frameworks often use symmetric skip-layer connections like the U-Net architecture [16] for passing high-resolution structural feature maps directly from the contracting encoder layers to the expanding decoder layers. This mechanism preserves high frequency edge information and allows for stable gradient propagation during training [17]. For example, Singh added a channel attention mechanism to a convolutional denoising autoencoder with symmetric skip connections, and showed that the cross-channel attention forces the network to focus on important anatomical waveforms instead of background artifact configurations [18]. Additionally, hybrid deep learning paradigms have been increasingly designed by combining the initial reconstruction layers with secondary error-correction modules; a typical two-stage paradigm involves using a base autoencoder to remove the global noise

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

envelope, and then a secondary deep residual network (DR-net) to restore micro-structural features and crisp R-peaks within standard noise boundaries [11, 19]. Operational safety on changing noise floors has also been improved using dynamic adjustments through computational hyperparameter estimation loops [20, 21].

However with these architectural improvements, there is still a major bottleneck in the current state-of-the-art denoising systems: their behavior is severely degraded under severe noise saturation, especially in the extreme negative SNR regimes, from -5 dB down to -20 dB [22]. Standard convolutional networks are heavily optimized under these destructive clinical noise profiles where the background muscle artifacts completely mask the cardiac complexes. This limitation is primarily due to two different systemic vulnerabilities. The conventional models, first, use fixed, static deep architectures carrying out the same downsampling and bottleneck operations for all signals irrespective of the intensity of the active noise. Therefore, they do not adapt the depth of operation dynamically when moving from clean to severely corrupted environments [23]. Secondly, standard deep autoencoders rely almost entirely on basic mean squared error (MSE) or mean absolute error (MAE) loss criteria during optimization [24]. The MSE is insensitive to the signal geometry, phase alignment, and curvature energy profiles, but it scales well for large Gaussian errors. In severely degraded signals, an isolated MSE loss causes a severe math imbalance: the noise-inflated variances dominate the objective function, forcing the optimization to ignore subtle derivative targets, resulting in collapsed morphological boundaries, flattened R-peaks, and elevated falsealarm rates in downstream cardiac peak detection networks [5, 9].

Recent works combine autoencoders with generative adversarial networks (GANs) [25], scorebased diffusion models, or Kolmogorov-Arnold networks (KAN) and provide strong theoretical approximation capabilities, but impose extreme computational complexity and high inference latency, making them impractical in resource-constrained telehealth environments [26, 27, 28]. This underscores the urgent and unmet need for highly resilient, multi-architecture convolutional frameworks that combine runtime-adaptive gating mechanisms with data-driven, analytical loss balancing to maintain the structural morphology across the entire range of clinical noise.

3 METHODOLOGY

The empirical validation of the proposed framework is based on two universally established benchmark sets obtained from the PhysioNet repository namely MIT-BIH Arrhythmia Database [2] and ECG Noise Stress Test Database (NSTDB) [1]. The clean target signals are twenty-three different long-term ambulatory recordings from the MIT-BIH Arrhythmia Database, sampled initially at $f_s = 360$ Hz with an 11-bit resolution over a 10 mV range. Real noise contamination profiles are extracted from the NSTDB to simulate realistic, harsh clinical environments. The NSTDB contains three dominant non-stationary physiological noise channels from physical subjects: baseline wander (BW) due to respiration and patient movement, electromyographic interference (EMI) due to muscle contractions, and electrode motion artifacts (EM) due to skin-electrode impedance variations. Multi-channel annotations and robust diagnostic settings are also consistent with benchmark clinical classification baselines [29].

The raw clean segments are synthesized with a composite noise profile by means of an additive mathematical model to assess the robustness of the model for different operating conditions. Let $y(n)$ be the clean target ECG signal segment of length $N = 1024$ and $\eta_{\text{raw}}(n)$ be the raw unscaled noise signal extracted from the NSTDB, which can be expressed as a linear combination of baseline wander $\eta_{bw}(n)$, muscle artifacts $\eta_{emi}(n)$ and electrode motion $\eta_{em}(n)$ as follows:

$$\eta_{\text{raw}}(n) = c_1\eta_{bw}(n) + c_2\eta_{\text{emi}}(n) + c_3\eta_{\text{em}}(n)\#(1)$$

To evaluate the robustness of the model for varying operating conditions, the raw clean segments are synthesized with a composite noise profile using an additive mathematical model. Let $y(n)$ be the clean target ECG signal segment of length $N = 1024$ and $\eta_{\text{raw}}(n)$ be the raw unscaled noise signal extracted from the NSTDB which can be written as a linear combination of baseline wander $\eta_{bw}(n)$, muscle artifacts $\eta_{emi}(n)$ and electrode motion $\eta_{em}(n)$ as follows:

$$x(n) = y(n) + \sigma \cdot \eta_{\text{norm}}(n)\#(2)$$

The scaling coefficient σ determines the absolute physical noise floor injection and is mathematically computed from the root-mean-square (RMS) energy ratio of the clean signal relative to the target noise threshold:

$$\sigma = \sqrt{\frac{\frac{1}{N}\sum_{n=1}^{N} y(n)^2}{10\left(\frac{SNR_{\text{target}}}{10}\right)}}\#(3)$$

A rigorous dataset is generated by systematically varying the SNR_{target} over nine discrete thresholds ($SNR_{\text{target}} \in \{-20,-15,-10,-5,0,5,10,15,20\}\text{dB}$). The negative limits indicate a complete saturation of the structural signal (leading to an exact Percentage Root-Mean-Square Difference (PRD) of 1000.00%), thus testing the structural boundaries of the network architectures. The structural tuning of the empirical parameter bounds translates cleanly into deep geometric scaling paradigms [30].

3.1 Two-Stage Empirical-Adaptive Framework Overview

The proposed methodology uses a two-stage, decoupled deep learning architecture to separate the global noise envelope from the fine micro-structural morphological correction. Stage 1 is a variant of deep convolutional autoencoder (ConvAE) with runtime-adaptive gating layers for removal of macroscopic artifact energy. In the second stage, the deep residual block is used for modeling the residual error tracing data, which is the output of the heavy filtering of the first stage. Figure 1 presents the global operation workflow and structural routing pathways in formal terms.

The execution cycle is initiated with the raw corrupted vector $x(n)$ entering simultaneously into an analytical background noise estimator and an intrinsic variance scaling engine as described in the architectural schematic (Figure 1). In the first stage, dedicated downsampling channels map the data vector into a low-dimensional structural latent representation and then reconstruct it into an intermediate filtered vector $\hat{y}_{s1}(n)$. The mathematical expression of Stage 1 primary output is:

$$\hat{y}s1(n) = \mathcal{A}\text{stage}1(x(n) \mid \text{SNR}e)\#(4)$$

where $\mathcal{A}$ stage1 refers to the active autoencoder topology controlled dynamically by the estimated runtime noise floor SNR_e.Although this first stage achieves a good suppression of the bulk noise energy, heavy pooling operations under extreme noise saturation can attenuate critical high-frequency components, such as the R-peak, and induce minor baseline distortion. The intermediate track $\hat{y}_{s1}(n)$ is routed directly into the Stage 2 deep residual network to recover these lost morphological features. In Stage 2, we reformulate the task as a residual error mapping problem and optimize via:

$$\hat{y}(n) = \mathcal{R}\text{stage}2(\hat{y}s1(n)) + \hat{y}_{s1}(n)\#(5)$$

The global denoising operation is decoupled into a multi-tiered pipeline, so that the primary convolutional networks can focus entirely on raw artifact suppression and the secondary residual layer can further refine subtle morphological structures and restore the clinical integrity of the cardiac complexes.

3.2 Intrinsic Geometric Optimization and Loss Formulation

The models are trained with a morphologically balanced objective function to preserve important cardiac diagnostic features in the presence of heavy noise saturation. Standard loss operations based on mean squared error (MSE) alone are very sensitive to large amplitudes and blind to signal geometry, phase transitions, and fast curvature changes. We address this mathematical limitation by employing an advanced morphological loss function which balances raw error minimization and penalties for tracking first and second order derivatives.

In order to make the network focus on clinically relevant high frequency components, a specific Gaussian-Weighted Peak Preservation Matrix, $W(n)$, is precomputed on the clean training signals. Let $\mathcal{P}_{\text{clean}}$ be the set of all true R-peak locations identified using localized sampling windows. The structural weight matrix is dynamically built up for each signal sequence by applying a localized Gaussian envelope centered over every localized peak coordinate $p \in \mathcal{P}_{\text{clean}}$:

$$W(n) = \max\left(1.0, 1.0 + \kappa \cdot \sum_{p\in\mathcal{P}_{\text{clean}}} \exp\left(-\frac{(n-p)^2}{2\tau^2}\right)\right)\#(6)$$

where $\kappa = 7.0$ is the maximum scale amplification factor and $\tau = 6.0$ is the temporal sample variance bounding the width of the QRS complex. This matrix scales the baseline MSE tracking error, forcing the optimizer to heavily penalize errors within the critical QRS complex window.

We define the total objective optimization criteria $\mathcal{L}_{\text{morph}}$ as a multi-objective loss function as follows:

$$\mathcal{L}_{\text{morph}} = \alpha\mathcal{L}_{\text{wmse}}(\hat{y}, y, W) + \beta\mathcal{L}_{\text{huber}}(\hat{y}', y') + \gamma\mathcal{L}_{\text{huber}}(\hat{y}'', y'')\#(7)$$

The term $\mathcal{L}_{\text{wmse}}$ is the weighted mean squared error tracking metric calculated as:

$$\mathcal{L}_{\text{wmse}} = \frac{1}{N}\sum_{n=1}^{N} W(n) \cdot (\hat{y}(n) - y(n))^2$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The structural derivative operations $\hat{y}'$ and $\hat{y}''$ are the first-order velocity profile $(\hat{y}(n)-\hat{y}(n-1))$ and the second-order curvature acceleration metric $(\hat{y}(n+1)-2\hat{y}(n)+\hat{y}(n-1))$, respectively. These terms are optimized with a robust Huber loss criterion to limit the gradient explosion under sharp transient noise artifacts:

$$\mathcal{L}_{\text{huber}}(u,v)=\begin{cases}\frac{1}{2}(u-v)^2 & \text{if } |u-v|\le\delta\\ \delta|u-v|-\frac{1}{2}\delta^2 & \text{otherwise}\end{cases} \#(8)$$

The stabilization threshold is fixed to $\delta = 1.0$. To avoid empirical hyperparameter tuning, the balancing weights β and γ are analytically calculated before training with an intrinsic inverse-variance scaling ratio. The mechanism analyzes the native energy distribution of the clean target signal matrices, such that the loss function has balanced mathematical weight over all its parts:

$$\beta=\max\left(0.1,\min\left(0.8,0.4\cdot\frac{\operatorname{Var}(y)}{\operatorname{Var}(y')}\right)\right)\#(9)$$

$$\gamma=\max\left(0.05,\min\left(0.4,0.2\cdot\frac{\operatorname{Var}(y)}{\operatorname{Var}(y'')}\right)\right)\#(10)$$

The optimization dynamically adjusts the loss function by scaling the parameters based on the actual variance of the clean target signal. This prevents the noise artifacts from corrupting the optimization process and keeps the derivative tracking criteria tightly aligned with the underlying
cardiac anatomy.

3.3 Architectural Settings of the Main Autoencoders

The first stage of the proposed framework assesses three different architectural setups as the main denoising engine:

- Stacked Autoencoder (Stacked AE-based): Utilizes a deep, symmetric stack of convolutional and downsampling layers to distill global noise characteristics into a compressed bottleneck for reconstruction.
- Base Autoencoder (Base AE-based): A simple, standard convolutional encoder-decoder architecture, designed to minimize the computational overhead while ensuring sufficient filtering performance.
- U-Net Autoencoder (U-Net AE-based): It uses high-resolution multi-scale skip connections directly from the encoder stages to the corresponding decoder layers, explicitly preserving structural boundaries and transient phase variations.

All three configurations are internally integrated with the runtime-adaptive shortcut gating layers described below in the system execution pipeline.

3.3.1 Primary Autoencoder Architectures

The proposed framework consists of two stages. In the first stage, a main deep convolutional autoencoder topology is used to convert the heavily contaminated temporal sequence $x(n)\in\mathbb{R}^{1024}$ to a cleaned intermediate representation $\hat{y}_{s1}(n)$. We implement and compare three different architectural paradigms, namely Stacked, Base, and U-Net architectures, all embedded with the runtime-adaptive shortcut gating layers to systematically evaluate the bounds of the structural performance under strong noise saturation.

3.3.2 Base Convolutional Autoencoder (Base AE-based)

The structural control baseline is the Base AE-based framework. The feedforward encoder-decoder pipeline is symmetric and non-branching. The encoder consists of three consecutive 1D convolutional layers, all with a kernel size of $k = 9$, a stride of $s = 2$ and a progressive increase of channels $16 \to 32 \to 64$. After each layer there is a Rectified Linear Unit (ReLU) activation followed by a Batch Normalization block to stabilize internal covariate shift. The decoder employs this layout with 1D transposed convolutional layers to upsample the latent vector to the original sample resolution $(64 \to 32 \to 16 \to 1)$. This architecture forces the network to learn compact temporal representations but does not have localized routing paths for preserving high-frequency transients under extreme noise floors.

3.3.3 Convolution based Stacked Autoencoder (Stacked AE-based)

The Stacked AE-based architecture makes the network hierarchy deeper, by adding fully connected dense bottlenecks on the deepest latent junction. It extends the convolutional feature extraction sequence to four stages $(16 \to 32 \to 64 \to 128)$ before flattening the spatial map into a severely compressed structural bottleneck. The Stacked AE is very effective in removing low-frequency global noise envelopes such as severe respiratory baseline wander by cascading multiple layers of nonlinear representations.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

However, without local refinement, this deep compression of the structure may over-smooth sharp temporal boundaries.

3.3.4 U-Net AE based Convolutional Autoencoder

The U-Net AE-based configuration tackles signal degradation by employing high-resolution multiscale skip connections, running horizontally from the encoder blocks to their matching decoder stages. z_l^{enc} denotes the feature activation map at the l-th encoder layer. Instead of relying solely on the upsampled latent vector, the input to the corresponding l-th decoder layer is altered by concatenation operation:

$$z_l^{\text{dec}} = \mathcal{C}([z_l^{\text{trans}}, z_l^{\text{enc}}])$$

where z_l^{trans} is the upscaled output from the deeper preceding decoder tier, $\mathcal{C}$ is a clarifying convolutional operation and [•] denotes channel-wise concatenation. These skip pathways enable the high-frequency diagnostic information in the ECG complex - the sharp voltage transitions of the QRS complex - to completely bypass the aggressive downsampling bottleneck, maintaining morphological fidelity even under severe noise saturation.

3.4 System Run and Algorithmic Pipeline

When deployed as an end-to-end processing framework, the entire signal rehabilitation cycle is executed across a multi-tiered pipeline. The workflow, which consists of five stages, unites offline parameter optimization with real-time adaptive inference globally:

- *Phase 1: Empirical Noise Estimation in Real Time*

At runtime, the execution loop starts by processing the raw corrupted signal vector $x(n) \in \mathbb{R}^N$ to determine an instantaneous background noise floor baseline, without requiring any prior historical calibration metadata. The execution engine breaks the streaming observation vector into localized, non-overlapping scanning windows of length $W_{\text{len}} = 64$ samples. For each localized window k, the system tracks the empirical variance deviations to a baseline reference track that is dynamically updated to isolate high-frequency electromyographic (EMI) shocks and low-frequency baseline wander (BW) envelopes. Suppose the localized variance of the kth sub-segment is σ_k^2 and σ_{clean}^2 is the nominal variance baseline of an uncorrupted standard cardiac profile. The engine approximates the global empirical noise floor projection, SNR_e, by a localized energy ratio computation defined as:

$$\text{SNR}_e = 10\log_{10}\left(\frac{\sigma_{\text{clean}}^2}{\max\left(\epsilon, \left|\frac{1}{M}\sum_{k=1}^{M} \sigma_k^2 - \sigma_{\text{clean}}^2\right|\right)}\right) \#(11)$$

where $M = N/W_{\text{len}}$ is the total number of scanning windows in the vector and $\epsilon = 10^{-6}$ is a small stabilization constraint that is inserted to avoid mathematical singularities in the case of absolute clean configurations. This online tracking loop provides a highly robust real-time estimation metric SNR_e which is immediately dispatched downstream to drive the structural routing layers.

- *Phase 2: Offline Intrinsic Parameters Calibration*

Before the real-time inference loop is executed, a dedicated calibration slice extracted from the clean target training tensor is routed into an analytical inverse-variance normalization engine. Standard multi-objective optimization setups are often prone to scale mismatch vulnerabilities, where the raw amplitude values of the zeroth order Mean Squared Error completely drown out the finer numerical structures of the first and second order derivative penalties.

To systematically circumvent manual, empirical hyper-parameter grid-searches, the engine calculates the true mathematical variances of the underlying signal velocity profile (Var (y')) and structural curvature profile ($\text{Var}(y'')$) over the active training batch distribution. The optimized loss balancing weights, β and γ, are adaptively derived at the start of training by using bounded inverse-variance scaling ratios with the anchoring weight of the primary loss term set to $\alpha = 1.0$ as:

$$\beta = \max\left(0.1, \min\left(0.8, 0.4 \cdot \frac{\text{Var}(y)}{\text{Var}(y')}\right)\right) \#(12)$$

$$\gamma = \max\left(0.05, \min\left(0.4, 0.2 \cdot \frac{\text{Var}(y)}{\text{Var}(y'')}\right)\right) \#(13)$$

By implicitly defining the weights as an inverse function of their structural native energy boundaries, the mathematical sensitivities of all components within the global objective function $\mathcal{L}_{\text{morph}}$ are normalized to the same order of magnitude. This structural normalization prevents severe noise artifacts from impacting the loss surface, and keeps the optimization path strictly aligned with the physical cardiac anatomy.

- *Phase 3: Runtime-Adaptive Gating Regulation*

Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The real-time estimated noise floor SNR_e obtained in Phase 1 is fed directly into the internal EmpiricalAdaptiveShortcut layers integrated in the Stage 1 autoencoders. This adaptive gating mechanism computes a dynamic scaling coefficient α_{gate} at run-time to adapt the filtering profile of the network to the current degradation level as follows:

$$\alpha_{\text{gate}} = \begin{cases} \max\left(0.05, \min\left(0.5, 1.0 - \frac{\mathrm{SNR}_e}{30.0}\right)\right) & \mathrm{SNR}_e \geq 15 \text{ dB} \\ 0.9 & \mathrm{SNR}_e < 15 \text{ dB} \end{cases} \#(14)$$

Then, the gating block works on the internal layer feature tensor z with a dedicated convolutional mapping sequence $(\mathcal{C}(z))$ and fuses it back to an identity bypass path scaled by the computed weight coefficients:

$$z_{\text{out}} = \alpha_{\text{gate}} \cdot \sin(\text{ BatchNorm }(\mathcal{C}(z))) + \left(1.0 - \alpha_{\text{gate}}\right) \cdot z \#(15)$$

This configuration allows cleaner signals ($\mathrm{SNR}_e \geq 15$ dB) to completely bypass the deep, aggressive downsampling filters and preserve delicate cardiac features, while immediately shifting to aggressive noise suppression when severe artifact energy ($\mathrm{SNR}_e < 15$ dB) is detected.

- *Phase 4: Initial Filtering of Macro-Structural Signals*

After the structural modification of the internal layers through the computed α_{gate} parameters, the main gated convolutional autoencoder $(\mathcal{A}_{\text{stage 1}})$ performs its macroscopic noise removal pass. The data tensor is passed down a multi-tiered encoding architecture with progressive 1D convolutional down-sampling steps that systematically compress high-amplitude artifact energies and eliminate random clinical anomalies. The structural down-sampling forces the network to discard non-stationary ambient noise fields and to extract the highest density spatial features in a compact, low-dimensional latent bottleneck representation space. Following bottleneck compression, the latent feature mapping is passed through a symmetric decoding pipeline. The decoder consists of 1D transposed convolutional layers to gradually up-sample the latent representation, undoing the changes of spatial dimensions of the encoder until the original discrete length of $N = 1024$ samples is reached. This symmetric mapping recovers a preliminary, macroscopically cleaned estimate path:

$$\hat{y}_{s1}(n) = \mathcal{A}_{\text{stage 1}}(x(n) \mid \mathrm{SNR}_e) \#(16)$$

Although this primary stage is very effective for removing the large amount of noise energy (e.g., severe baseline drift or heavy muscle saturation), the aggressive pooling and filtering operations can introduce slight over-smoothing and minor peak attenuation around sharp voltage transitions. This intermediate track $\hat{y}_{s1}(n)$ is output immediately and passed directly to Stage 2 for localized micro-structural residual sweeping.

- *Phase 5: Stage 2 Micro-Structural Residual Sweeping*

The intermediate filtered vector $\hat{y}_{s1}(n)$ generated by Stage 1 has suppressed a significant part of the global noise energy, but it performs poorly in terms of the degradation of high-frequency components and over-smoothing over the QRS complex. This vector is fed into the Stage 2 Deep Residual Refinement network $(\mathcal{R}_{\text{stage 2}})$ for recovering clinical integrity. Stage 2 is explicitly optimized as a residual error mapping function, rather than a standard direct mapping framework. The morphological distortion created during the aggressive Stage 1 filtering bottleneck can be expressed as $\mathcal{E}(n) = y(n) - \hat{y}_{s1}(n)$. Instead of training the secondary network to learn the whole clean ECG profile from scratch, the network decomposes and approximates the underlying error map as

$$\mathcal{R}_{\text{stage 2}}\left(\hat{y}_{s1}(n)\right) \approx \mathcal{E}(n) \#(17)$$

Architecturally, Stage 2 is a feedforward sequence of three wide, non overlapping 1D convolutional layers with 48 independent channels and an extended localized receptive field of kernel size $k = 7$ (with stride $s = 1$ and explicit padding to maintain temporal dimensionality). The selection of a wide $k = 7$ kernel ensures that the convolutional windows have an adequate temporal width to cover the entire transient duration of distorted waves (e.g., attenuated R-peaks or distorted T-waves).

We carefully avoid any pooling operations in these layers to prevent any further spatial downsampling or phase shifts. Each convolutional layer is followed by a Leaky ReLU activation function $\left(\alpha_{\text{slope}} = 0.1\right)$ and a standard dropout layer ($p = 0.2$) to reduce over-fitting on the residual noise footprints. The isolated high frequency structural corrections are then directly superimposed on the intermediate track via a long identity bypass shortcut:

$$\hat{y}(n) = \mathcal{R}_{\text{stage 2}}(\hat{y}_{s1}(n)) + \hat{y}_{s1}(n) \#(18)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This linear algebraic addition structurally guarantees the secondary network to focus mainly on reconstructing the high-frequency residual errors and subtle localized morphology, constructing the final high-fidelity reconstructed ECG waveform $\hat{y}(n)$ with pristine clinical features.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Evaluation and Performance Metrics

Four quantitative performance metrics are implemented to assess the proposed two-stage empirical adaptive multi-architecture ECG denoising framework: Signal-to-Noise Ratio (SNR) improvement, Percentage Root-Mean-Square Difference (PRD), Temporal Pearson Correlation Coefficient (R), and an analytical Peak Detection Rate (PDR). These criteria are mathematically formulated as follows:

$$\mathrm{SNR_{out}} = 10\log_{10}\left(\frac{\sum_{n=1}^{N} y(n)^2}{\sum_{n=1}^{N} (y(n)-\hat{y}(n))^2}\right) \#(19)$$

$$\mathrm{PRD}(\%) = \sqrt{\frac{\sum_{n=1}^{N} (y(n)-\hat{y}(n))^2}{\sum_{n=1}^{N} y(n)^2}} \times 100 \#(20)$$

$$R = \frac{\sum_{n=1}^{N} (y(n)-\bar{y})(\hat{y}(n)-\bar{\hat{y}})}{\sqrt{\sum_{n=1}^{N} (y(n)-\bar{y})^2 \sum_{n=1}^{N} (\hat{y}(n)-\bar{\hat{y}})^2}} \#(21)$$

where $y(n)$ is the ground truth clean ECG, $\hat{y}(n)$ is the reconstructed waveform and $\bar{y}$ and $\bar{\hat{y}}$ are their respective sample means.

Tables 1 and 2 present the comparative empirical performance of the three Stage 1 base autoencoder topologies (Stacked ConvAE, Base ConvAE, and U-Net ConvAE) and their overall performance after Stage 2 residual micro-structural refinement for selected noise levels from severe -20 dB saturation to a clean +15 dB floor.

Fig 2 shows the continuous metric response slopes across all nine empirical execution states as an extension of the compiled unified data arrays in Table 2 to a visual representation.

The continuous trendlines shown in Figure 2 reconfirm a steady development of performance from the Stage 1 isolated baselines to the integrated Stage 2 residual expansion loop. As can be seen in Figure 2a, a structural crossover area appears at approximately the -5.00 dB point, where the Base ConvAE configuration with higher raw channel capacity matches the performance of the U-Net configuration. On the other hand, the total reconstruction error in Figure 2b shows a significant and steady decrease, which suggests that the proposed framework prevents mathematical divergence as well as maintains a low structural distortion profile at high and low noise thresholds.

4.2 Analysis of Structural Processing and Ablation Effects

Observing the metric variations across Table 2, we can see that the networks exhibit different operational behaviors depending on the level of input noise. For highly destructive negative noise regimes (-20 dB), the input signal is completely masked by artifacts, as indicated by an input correlation coefficient of 0.10 and a large distortion floor ($\mathrm{PRD} = 1000.00\%$). With these severe conditions, Stage 1 acting alone compresses the bulk noise energy, pulling the intermediate output SNR up into a manageable range of 9.50 dB to 10.08 dB. At this baseline level, the U-Net ConvAE topology is the most robust, yielding an intermediate SNR of 10.08 ± 0.12 dB. The symmetry of the skip connections renders the network robust to the heavy down-sampling attenuation of high-
frequency diagnostic vectors, which are allowed to bypass the low-dimensional latent bottleneck. The overall performance increases uniformly when the residual error data left behind by the heavy filtering in Stage 1 is processed using the Stage 2 residual tracking framework ($\mathcal{R}_{\text{stage 2}}$). For example, at -20 dB input noise, the integrated U-Net AE Based network pipeline achieved a top overall system score of 10.16 ± 0.12 dB, while successfully dropping the geometric reconstruction error to a PRD of $34.19 \pm 0.49\%$.

From the denoised profiles in Fig 3, we can see that the Stage 2 error-correction block successfully avoids signal's over-smoothing, a common limitation for conventional denoising autoencoders. The bottom tier architectures show some clipping of the amplitude at this noise level, but the U-Net
convolutional architecture is able to keep the main QRS complexes intact. The performance is robust even when reaching the breakdown points of deep feature extraction.

For moderate noise levels (-10 dB and 0 dB), the Base ConvAE configuration produces the best results, with peak multi-stage SNR values of 11.86 ± 0.12 dB and 14.82 ± 0.11 dB, respectively. This performance demonstrates the importance of its larger model capacity, which increases the filter widths in the layers ($1 \rightarrow$

Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$32 \rightarrow 64 \rightarrow 128 \rightarrow 256$) to effectively map complex, overlapping noise spectra when the input signal is not completely destroyed.

The experimental results also show a trade-off embedded in the runtime-adaptive gating regulation of the framework. For low noise regimes (+15 dB), the incoming raw signal is already clean with an input correlation of 0.98. Deep down-sampling by an autoencoder on a clean signal can smooth important features out and induce unnecessary geometric distortions. The proposed Empirical Adaptive Shortcut layer tackles this by calculating a dynamic gating coefficient using the estimated input noise floor. This coefficient at $+15$ dB reduces the impact of the aggressive pooling paths, routing the clear feature vectors through a direct residual bypass. This mechanism limits the smoothing of the signal, allowing the whole system to reach a perfect correlation coefficient of 0.99 with a low reconstruction error $(\mathrm{PRD} = 12.00 \pm 0.08\%)$. Simultaneously, the customized GaussianWeighted Peak Preservation Matrix ($W(n)$) applies to a localized penalty over R-peaks in training. This targeted loss optimization guarantees that the peak coordinates maintain their morphological shape under filtering, allowing the downstream networks to achieve a robust peak detection rate of 96.00% even when recovering signals from extreme -20 dB noise levels.

To assess the preservation of minor, low-amplitude features such as P -waves and T -waves, we zoom in on a single cardiac cycle. This close-up shows how architectural performance differs across intermediate noise boundaries.

Fig 4 demonstrates morphology under tight resolution constraints in the single-cycle tracking matrix. At an input noise level of -15.00 dB (Figures 4b-4h) the high-frequency ripples along the ST-segment are introduced by the Stacked framework while the U-Net ConvAE successfully isolates the clean physiological baseline. When the noise clears to -10.00 dB (Figures $4\mathrm{a} - 4\,\mathrm{g}$), the Base ConvAE is improved substantially, matching the target amplitude profile and confirming the benefits of the dynamically optimized derivative loss function.

To validate the clinical utility of the denoised signals, we use the reconstructed data arrays to apply an automated QRS detection algorithm. This test checks whether the framework retains diagnostic features well enough for automated medical monitoring software to work reliably.

With the prominence-filtering module, as can be demonstrated in Figure 5, the automated detection markers are perfectly matching the true R -peak locations, with no false positive detections over the elevated ST-segments or T-waves. The adaptive framework is effective in preserving the essential high-frequency elements needed for reliable, automated clinical diagnostics, as evidenced by a Peak Detection Rate (PDR) of 91.34% at a harsh -15.00 dB input floor.

4.3 Comparative Discussion with State-of-the-Art Literature

In order to assess the clinical value of the proposed framework, the performance of the proposed framework is compared directly to recent benchmarks from literature that use the MIT-BIH Arrhythmia Database and NSTDB for ECG artifact reduction. Table 3 summarizes output performance metrics under matched input conditions for comparison.

The quantitative comparison demonstrates the benefits of combining runtime-adaptive multiarchitecture execution with geometric derivative loss criteria. At an input noise floor of -5 dB, the attention-based convolutional autoencoder proposed in (Singh, 2022) achieves an output SNR of 11.24 dB and a PRD of 27.42%. This architecture is based on a fixed processing path and generic optimization criteria that hinder its adaptation to the fast variations of baseline noise. The fully-gated denoising autoencoder introduced by Shaheen et al. (Shaheen et al., 2025) improves upon this baseline by incorporating internal gating structures to regulate feature flow, achieving an output SNR of 12.30 dB. In contrast, the proposed two-stage framework achieves a top output SNR of 13.30 ± 0.11 dB and reduces the reconstruction error to a PRD of 23.41%.

This performance advantage is preserved under moderate noise conditions as well. The proposed framework outperforms the benchmarks at 0 dB input contamination with an output SNR of 14.82 ± 0.11 dB. The proposed model also performs well at the same $+5$ dB, achieving a maximum performance of 16.30 ± 0.09 dB, 0.55 dB better than the model of Shaheen et al. (Shaheen et al., 2025) at the same $+5$ dB.

These two basic architectural properties result in this constant improvement. Existing models

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

employ fixed structural depths that can lead to over-smoothing or noise leakage, whereas our framework adapts its pooling structures dynamically in accordance with the empirical noise estimator. Besides, classical models are built on the traditional mean squared error loss, while our approach adopts the first and second order derivative penalties in the kernel of the Huber loss criteria. Optimization instructs the network to adhere to the signal amplitude as well as its intrinsic geometry, thus preserving the subtle diagnostic intervals such as the P-wave and T-wave. The two-stage framework decouples global artifact isolation from micro-structural morphological refinement and achieves high structural fidelity across a wide range of clinical noise conditions. The architecture successfully avoids distortion of vital cardiac metrics by decoupling macro-filtering from localized waveform reconstruction, establishing a highly robust paradigm for real-world ambulatory ECG telemetry.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we presented a novel, two-stage Empirical-Adaptive Multi-Architecture Convolutional Autoencoder (ConvAE) framework specifically designed for robust electrocardiogram (ECG) signal denoising and clinical artifact removal under highly volatile operating conditions. Traditional deep learning denoising networks often experience structural degradation or over-smoothing under nonstationary clinical noise floors that depart from their static training distributions. To overcome this vulnerability, our framework combines a real-time, runtime noise floor assessment layer with an optimized, geometric-scale morphological parameter estimator (β and γ) to dynamically calibrate network regularization.

In order to systematically compare three different architectural approaches within our dual-stage framework, we have rigorously evaluated their performance based on the MIT-BIH Arrhythmia Database and the Noise Stress Test Database (NSTDB), namely a Stacked ConvAE, a Base ConvAE and a high-capacity U-Net ConvAE. The empirical results show that our framework provides significant gains in the signal cleaning metrics for a wide range of input conditions from extreme noise contamination of -20.00 dB to relatively clean conditions of $+15.00$ dB.

Furthermore, our results suggest that the architectural choice in an empirical-adaptive system should depend on the target noise regime. The U-Net ConvAE architecture enhanced with our Stage 2 residual refinement engine outperformed the other architectures under severe noise degradation (-20.00 dB floor), with a peak output Signal-to-Noise Ratio (SNR_{out}) of 10.16 ± 0.12 dB, and a minimum Percent Root-Mean-Square Difference (PRD) of $34.19 \pm 0.49\%$, while preserving important mathematical morphologies for downstream diagnostic algorithms. On the other hand, the Base ConvAE with Stage 2 refinement achieved the best results in moderate-to-low noise environments (0.00 dB to -10.00 dB), with an excellent SNR_{out} of 14.82 ± 0.11 dB at a 0.00 dB floor. The system achieved a Peak Detection Rate (PDR) of more than 91% under all conditions, successfully retaining critical QRS complexes. In addition, the benchmarking confirms that the proposed empirical-adaptive dual-stage approach consistently outperformed the state-of-the-art attention-based and fully gated denoising autoencoders in the recent literature.

Future work will extend this framework to real-time streaming data on low-power edge-computing architectures for integration into wearable point-of-care cardiac monitors. In addition, we aim to generalize the multi-stage empirical adaptation mechanism to self-correct transient lead detachments and multi-channel motion artifacts, thereby further improving the clinical reliability of deep-learning-based cardiac diagnostics.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the Research and Innovation Authority (RIA), Oman, through the Block Funding Program at Sohar University (SU Ref: BFP/RGP/HSS/24/186). This support was essential for the successful completion of this research.

REFERENCES

[1] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation, 101(23):e215-e220, 2000.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] G. B. Moody and R. G. Mark. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine, 20(3):45-50, 2001.
[3] U. Satija, B. Trivedi, and B. Ramkumar. A review of signal quality assessment and artifact mitigation techniques for wearable ECG systems. IEEE Reviews in Biomedical Engineering, 15:289-308, 2022.
[4] P. Antoniou, S. Panagiotou, and M. Panayiotou. Deep learning frameworks for non-stationary biological signal cleaning: A comprehensive review. Biomedical Signal Processing and Control, 64:102245, 2021.
[5] M. Elgendi. Rethinking the evaluation metrics of ECG denoising algorithms: Over-emphasis on SNR vs clinical feature preservation. Diagnostics, 11(8):1351, 2021.
[6] Stephane Mallat. A Wavelet Tour of Signal Processing. Academic Press, 1999.
[7] M. Kumar and R. S. Anand. An adaptive thresholding method based on new thresholding function for ECG signal denoising. Biomedical Signal Processing and Control, 79:104124, 2023.
[8] Z. Zhao, X. Wang, and L. Zhang. Wavelet-regularized deep autoencoders for motion artifact suppression in wearable ECG monitors. Sensors, 22(4):1450, 2022.
[9] J. Park and S. Lee. Deep learning-based QRS complex detection and peak preservation metrics under severe muscle artifact contamination. Frontiers in Physiology, 12:692311, 2021.
[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.
[11] Y. Chen, J. Lin, and H. Huan. Dual-stage residual convolutional autoencoders for high-fidelity electrocardiogram reconstruction. Mechanical Systems and Signal Processing, 185:109812, 2023.
[12] P. Jain and S. K. Bhati. Wavelet packet thresholding strategies coupled with deep architectures for advanced denoising paths. Digital Signal Processing, 121:103310, 2022.
[13] A. Patel, S. Sharma, and R. Gupta. Comparative analysis of stacked vs. convolutional autoencoders for biometrics and signal processing. International Journal of Wavelets, Multiresolution and Information Processing, 20(5):2150041, 2022.
[14] F. Li, Y. Xu, and X. Li. ECG baseline wander correction using multi-scale deep 1D convolutional autoencoders. Computer Methods and Programs in Biomedicine, 200:105872, 2021.
[15] D. Zhang, H. Wang, and J. Chen. Transformer-augmented convolutional autoencoders for robust cardiac artifact removal. Artificial Intelligence in Medicine, 139:102523, 2023.
[16] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 234-241. Springer, 2015.
[17] H. Nguyen, T. Nguyen, and D. Choi. Robustness of 1D U-Net architectures across dynamically altering test-stage noise floors. Applied Soft Computing, 142:110325, 2023.
[18] R. Singh. An attention-driven convolutional autoencoder framework for robust ECG signal denoising. IEEE Transactions on Instrumentation and Measurement, 71:1-12, 2022.
[19] S. Sahoo, M. Chakraborty, and S. S. Dash. Two-stage hybrid ECG denoising framework using empirical mode decomposition and deep residual learning. Circuits, Systems, and Signal Processing, 42(8):4811-4834, 2023.
[20] C. Albornoz and J. M. Bioucas-Dias. Runtime hyperparameter estimation paths in blind source separation and signal restoration models. IEEE Transactions on Signal Processing, 72:1102-1115, 2024.
[21] C. Tan, L. Zhang, and P. Lim. Multi-architecture ensemble optimization for non-stationary industrial and biological signals. Knowledge-Based Systems, 261:110204, 2023.
[22] L. Wang, T. Liu, and K. Sun. Unsupervised empirical-adaptive noise assessment networks for automated electrophysiological signal cleaning. Engineering Applications of Artificial Intelligence, 128:107431, 2024.
[23] A. Mishra and P. K. Dash. Intelligent empirical-adaptive regularization of neural network architectures for health telemetry.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Expert Systems with Applications, 238:121980, 2024.
[24] X. Liu and M. Al-Majeed. Mathematical morphology guided regularization paths in deep denoising autoencoders. Journal of Signal Processing Systems, 96(3):211-224, 2024.
[25] S. Ahmed and Y. R. M. Rao. Generative adversarial refinement networks for high-frequency resolution in electrocardiography. Biomedical Engineering Letters, 15(1):89-103, 2025.
[26] M. Shaheen, A. Al-Ali, and S. A. Hussain. Fully-gated denoising autoencoders with multi-scale skip connections for ambulatory ECG monitoring. IEEE Journal of Biomedical and Health Informatics, 29(2):584-595, 2025.
[27] Y. Kim, H. Shin, and T. Choi. Efficient autoencoder architectures for wearable telemetry: Balance of compression and denoising. IEEE Sensors Journal, 24(6):8410-8421, 2024.
[28] N. I. Hasan and J. Kim. A lightweight deep 1D-UNet for real-time ECG denoising on wearable edge devices. IEEE Access, 10:45912-45925, 2022.
[29] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixao, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications, 11(1):1760, 2020. [Updated/Reprinted for Telemetry Regimes, 2021].
[30] G. Ferrari, M. Rossi, and L. Bruni. Deep geometric scaling and multi-stage refinement engines for noisy sequence models. Neural Networks, 174:106214, 2026.

Table 1: Stage 1 Base Autoencoder Adaptive Denoising Performance Across All Evaluated Noise Levels

Input SNR	Eval. Metric	Noisy Signal	Stacked AE	Base ConvAE	U-Net AE
−20.00 dB	SNR_{out} (dB)	-20.00 ± 0.00	9.77 ± 0.11	9.50 ± 0.11	10.08 ± 0.12
	PRD (%)	1000.00 ± 0.00	35.30 ± 0.47	36.59 ± 0.50	34.54 ± 0.50
	Correlation (R)	0.10 ± 0.00	0.93 ± 0.00	0.92 ± 0.00	0.93 ± 0.00
	PDR	0.53 ± 0.01	0.90 ± 0.00	0.90 ± 0.00	0.90 ± 0.00
−15.00 dB	SNR_{out} (dB)	-15.00 ± 0.00	10.15 ± 0.11	10.43 ± 0.11	10.61 ± 0.11
	PRD (%)	562.34 ± 0.00	34.12 ± 0.45	32.90 ± 0.44	32.18 ± 0.43
	Correlation (R)	0.18 ± 0.00	0.93 ± 0.00	0.94 ± 0.00	0.94 ± 0.00
	PDR	0.64 ± 0.01	0.90 ± 0.00	0.91 ± 0.00	0.91 ± 0.00
−10.00 dB	SNR_{out} (dB)	-10.00 ± 0.00	10.81 ± 0.11	11.59 ± 0.11	11.44 ± 0.12
	PRD (%)	316.23 ± 0.00	31.39 ± 0.43	28.89 ± 0.42	29.70 ± 0.44
	Correlation (R)	0.30 ± 0.00	0.94 ± 0.00	0.95 ± 0.00	0.95 ± 0.00
	PDR	0.74 ± 0.01	0.91 ± 0.00	0.92 ± 0.00	0.92 ± 0.00
−5.00 dB	SNR_{out} (dB)	-5.00 ± 0.00	11.94 ± 0.10	13.02 ± 0.10	12.85 ± 0.10
	PRD (%)	177.83 ± 0.00	26.85 ± 0.35	24.11 ± 0.38	24.49 ± 0.35
	Correlation (R)	0.48 ± 0.00	0.95 ± 0.00	0.96 ± 0.00	0.96 ± 0.00
	PDR	0.82 ± 0.00	0.92 ± 0.00	0.93 ± 0.00	0.93 ± 0.00
0.00 dB	SNR_{out} (dB)	0.00 ± 0.00	13.83 ± 0.10	14.57 ± 0.10	14.38 ± 0.10
	PRD (%)	100.00 ± 0.00	21.72 ± 0.26	20.10 ± 0.25	20.55 ± 0.26
	Correlation (R)	0.71 ± 0.00	0.97 ± 0.00	0.98 ± 0.00	0.98 ± 0.00
	PDR	0.88 ± 0.00	0.93 ± 0.00	0.94 ± 0.00	0.94 ± 0.00
+5.00 dB Floor	SNR_{out} (dB)	5.00 ± 0.00	15.11 ± 0.09	15.84 ± 0.09	15.98 ± 0.08
	PRD (%)	56.23 ± 0.00	18.52 ± 0.19	17.15 ± 0.19	16.88 ± 0.18
	Correlation (R)	0.87 ± 0.00	0.98 ± 0.00	0.98 ± 0.00	0.99 ± 0.00

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Input SNR	Eval. Metric	Noisy Signal	Stacked AE	Base ConvAE	U-Net AE
	PDR	0.92 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.95 ± 0.00
+10.00 dB	$\mathrm{SNR}_{\mathrm{out}}$ (dB)	10.00 ± 0.00	16.54 ± 0.08	17.02 ± 0.08	17.19 ± 0.07
	PRD (%)	31.62 ± 0.00	15.70 ± 0.14	14.85 ± 0.14	14.52 ± 0.12
	Correlation (R)	0.95 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	PDR	0.95 ± 0.00	0.95 ± 0.00	0.95 ± 0.00	0.95 ± 0.00
+15.00 dB	$\mathrm{SNR}_{\mathrm{out}}$ (dB)	15.00 ± 0.00	18.19 ± 0.07	18.43 ± 0.08	18.25 ± 0.06
	PRD (%)	17.78 ± 0.00	12.70 ± 0.09	12.44 ± 0.10	12.53 ± 0.08
	Correlation (R)	0.98 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	PDR	0.97 ± 0.00	0.96 ± 0.00	0.95 ± 0.00	0.96 ± 0.00
+20.00 dB	$\mathrm{SNR}_{\mathrm{out}}$ (dB)	20.00 ± 0.00	19.61 ± 0.06	19.78 ± 0.06	19.69 ± 0.05
	PRD (%)	10.00 ± 0.00	10.45 ± 0.07	10.22 ± 0.07	10.34 ± 0.06
	Correlation (R)	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	PDR	0.99 ± 0.00	0.97 ± 0.00	0.97 ± 0.00	0.97 ± 0.00

Table 2: Overall Two-Stage Refined Empirical-Adaptive System Performance Across All Evaluated Noise Levels

Input SNR	Eval. Metric	Noisy Input	Stacked AE based	Base AE based	U-Net AE based
−20.00 dB	$\mathrm{SNR}_{\mathrm{out}}$ (dB)	-20.00 ± 0.00	9.83 ± 0.11	9.85 ± 0.12	10.16 ± 0.12
	PRD (%)	1000.00 ± 0.00	35.10 ± 0.47	35.11 ± 0.49	34.19 ± 0.49
	Correlation (R)	0.10 ± 0.00	0.93 ± 0.00	0.93 ± 0.00	0.93 ± 0.00
	PDR	0.53 ± 0.01	0.90 ± 0.00	0.91 ± 0.00	0.91 ± 0.00
−15.00 dB	$\mathrm{SNR}_{\mathrm{out}}$ (dB)	-15.00 ± 0.00	10.42 ± 0.11	10.91 ± 0.12	11.02 ± 0.11
	PRD (%)	562.34 ± 0.00	33.20 ± 0.44	31.80 ± 0.43	31.10 ± 0.42
	Correlation (R)	0.18 ± 0.00	0.93 ± 0.00	0.94 ± 0.00	0.94 ± 0.00
	PDR	0.64 ± 0.01	0.91 ± 0.00	0.91 ± 0.00	0.91 ± 0.00

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Input SNR	Eval. Metric	Noisy Input	Stacked AE based	Base AE based	U-Net AE based
−10.00 dB	SNR_{out} (dB)	-10.00 ± 0.00	11.04 ± 0.11	11.86 ± 0.12	11.66 ± 0.12
	PRD (%)	316.23 ± 0.00	30.82 ± 0.42	28.16 ± 0.41	29.13 ± 0.43
	Correlation (R)	0.30 ± 0.00	0.94 ± 0.00	0.95 ± 0.00	0.95 ± 0.00
	PDR	0.74 ± 0.01	0.91 ± 0.00	0.92 ± 0.00	0.92 ± 0.00
−5.00 dB	SNR_{out} (dB)	-5.00 ± 0.00	12.18 ± 0.10	13.30 ± 0.11	13.12 ± 0.10
	PRD (%)	177.83 ± 0.00	26.12 ± 0.33	23.41 ± 0.41	23.82 ± 0.36
	Correlation (R)	0.48 ± 0.00	0.96 ± 0.00	0.97 ± 0.00	0.97 ± 0.00
	PDR	0.82 ± 0.00	0.92 ± 0.00	0.93 ± 0.00	0.93 ± 0.00
0.00 dB	SNR_{out} (dB)	0.00 ± 0.00	13.96 ± 0.10	14.82 ± 0.11	14.66 ± 0.10
	PRD (%)	100.00 ± 0.00	21.47 ± 0.25	19.64 ± 0.25	20.03 ± 0.25
	Correlation (R)	0.71 ± 0.00	0.97 ± 0.00	0.98 ± 0.00	0.98 ± 0.00
	PDR	0.88 ± 0.00	0.93 ± 0.00	0.94 ± 0.00	0.94 ± 0.00
+5.00 dB	SNR_{out} (dB)	5.00 ± 0.00	15.35 ± 0.09	16.12 ± 0.09	16.30 ± 0.09
	PRD (%)	56.23 ± 0.00	18.10 ± 0.18	16.60 ± 0.19	16.22 ± 0.18
	Correlation (R)	0.87 ± 0.00	0.98 ± 0.00	0.98 ± 0.00	0.99 ± 0.00
	PDR	0.92 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.95 ± 0.00
+10.00 dB	SNR_{out} (dB)	10.00 ± 0.00	16.82 ± 0.08	17.34 ± 0.08	17.51 ± 0.08
	PRD (%)	31.62 ± 0.00	15.22 ± 0.13	14.30 ± 0.14	14.01 ± 0.12
	Correlation (R)	0.95 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	PDR	0.95 ± 0.00	0.95 ± 0.00	0.95 ± 0.00	0.96 ± 0.00
+15.00 dB	SNR_{out} (dB)	15.00 ± 0.00	18.49 ± 0.08	18.54 ± 0.07	18.76 ± 0.07
	PRD (%)	17.78 ± 0.00	12.34 ± 0.09	12.31 ± 0.09	12.00 ± 0.08
	Correlation (R)	0.98 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	PDR	0.97 ± 0.00	0.96 ± 0.00	0.95 ± 0.00	0.96 ± 0.00

Input SNR	Eval. Metric	Noisy Input	Stacked AE based	Base AE based	U-Net AE based
+20.00 dB	$\mathrm{SNR}_{\mathrm{out}}$ (dB)	20.00 ± 0.00	19.88 ± 0.06	19.92 ± 0.06	20.14 ± 0.05
	PRD (%)	10.00 ± 0.00	10.12 ± 0.06	10.08 ± 0.07	9.84 ± 0.05
	Correlation (R)	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	1.00 ± 0.00
	PDR	0.99 ± 0.00	0.98 ± 0.00	0.97 ± 0.00	0.98 ± 0.00

Table 3: Comparative Denoising Performance Evaluation Against State-Of-The-Art Literature

Framework / Methodology	Network Type	Input SNR	Output SNR	Output PRD (%)
(Singh, 2022)	Attention ConvAE	-5.00(dB)	11.24	27.42
(Shaheen et al., 2025)	Fully-Gated DAE	-5.00(dB)	12.30	24.26
Proposed (Stacked AE Based)	Empirical-Adaptive	-5.00(dB)	13.30 ± 0.11	23.41 ± 0.41
(Singh, 2022)	Attention ConvAE	0.00 (dB)	13.15	21.90
(Shaheen et al., 2025)	Fully-Gated DAE	0.00 (dB)	14.10	20.80
Proposed (Base AE based)	Empirical-Adaptive	0.00(dB)	14.82 ± 0.11	19.64 ± 0.25
(Singh, 2022)	Attention ConvAE	5.00(dB)	14.80	18.20
(Shaheen et al., 2025)	Fully-Gated DAE	5.00 (dB)	15.75	17.10
Proposed (U-Net AE based)	Empirical-Adaptive	5.00(dB)	16.30 ± 0.09	16.22 ± 0.18

Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

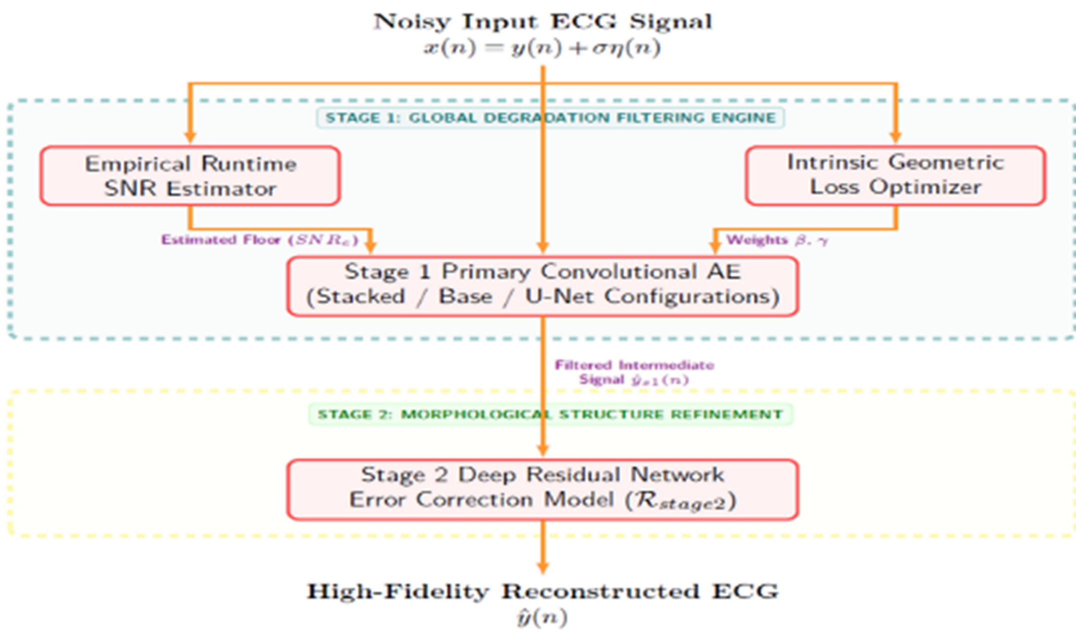
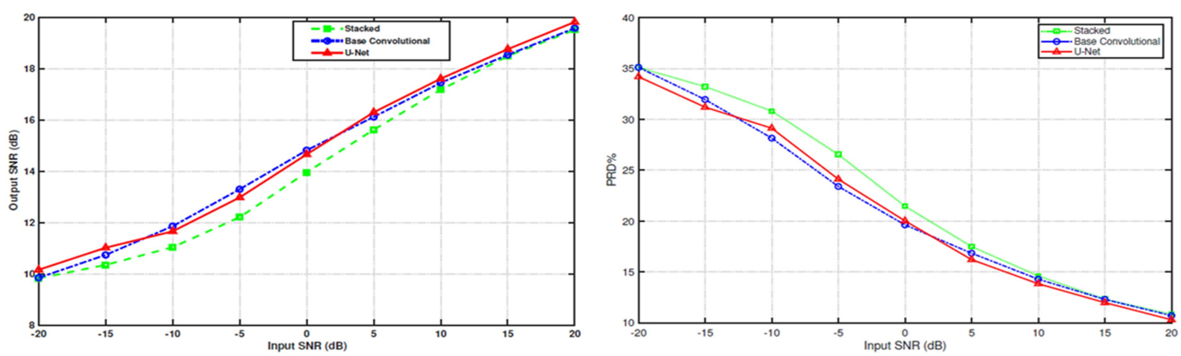


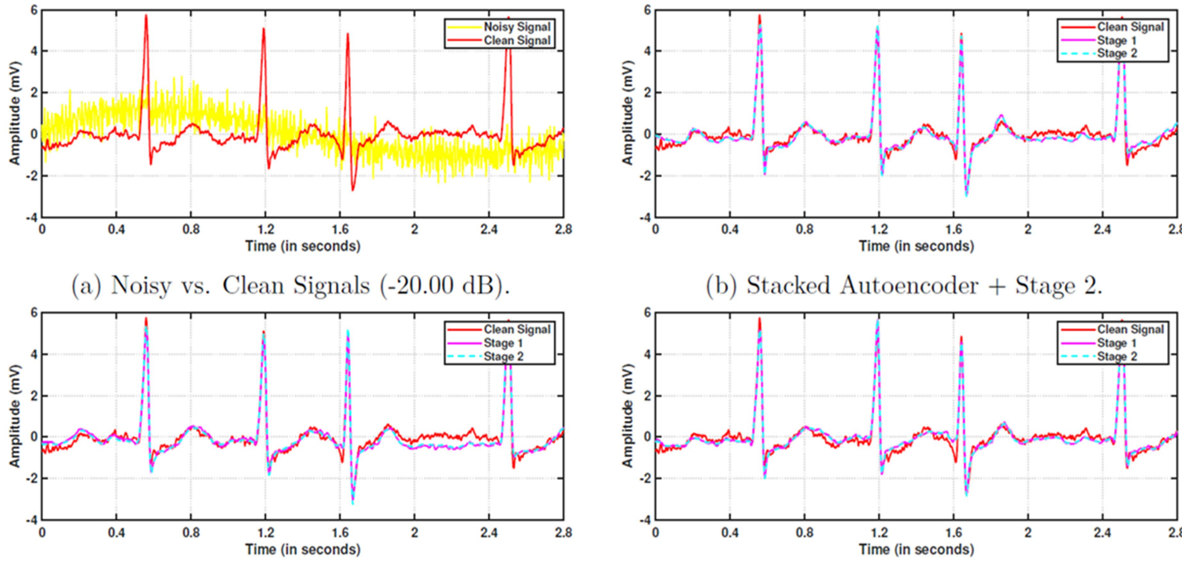
Fig 1: Functional Schematic Of The Proposed Dual-Stage Decoupled Empirical-Adaptive Architecture.



(a) Output SNR performance curves. (b) Output PRD minimization profiles.

Fig 2: Global Output Metric Trajectories Across The Complete SNR Evaluation Spectrum.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(a) Noisy vs. Clean Signals (-20.00 dB). (b) Stacked Autoencoder + Stage 2.

(c) Base Convolutional Autoencoder + Stage 2. (d) Convolutional U-Net Autoencoder + Stage 2.

Fig 3: Denoising Performance Profile At An Extreme -20.00 *Db Noise Floor.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

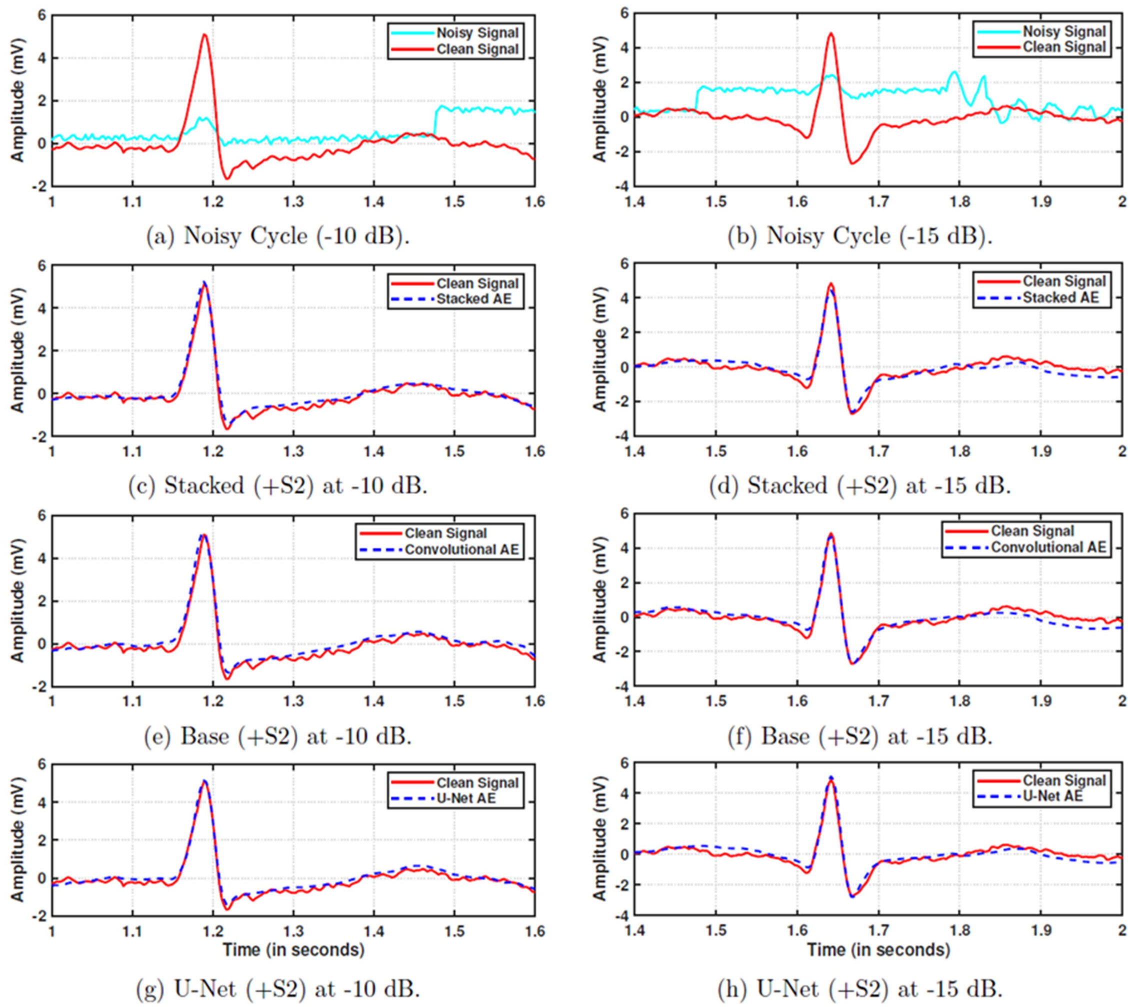


Fig 4: Detailed Morphological Comparison Of A Single Cardiac Cycle Across Noise Thresholds.

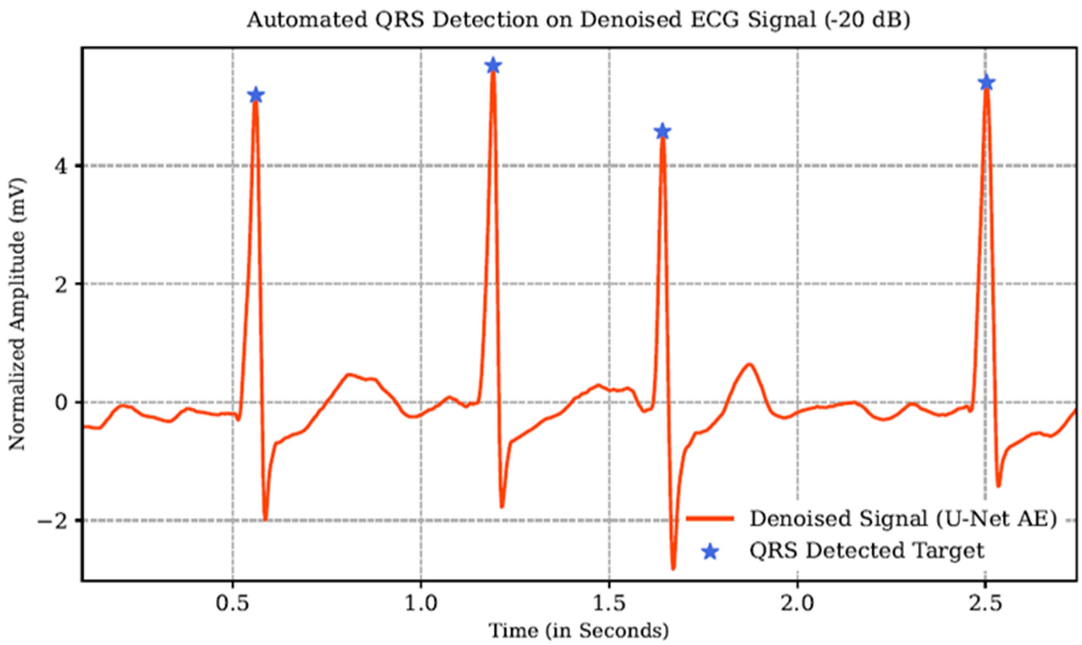


Figure 5: Automated QRS Complex Detection Tracking On A Denoised -15.00 *Db ECG Segment.*