I.J. Image, Graphics and Signal Processing, **2026, 5, 55-75**
Published Online on October 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2026.05.04

DeepFusion-CNN: A Novel Context-aware Network for ECG-based Gender Classification

Sanjay Kumar Pandey*
Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, 333010, India
E-mail: sanjaypandeyucer@gmail.com
ORCID iD: https://orcid.org/0000-0001-6672-8182
*Corresponding author

Bechoo Lal
Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, 333010, India
E-mail: blal2k7@gmail.com
ORCID iD: https://orcid.org/0000-0002-0225-1001

Received: January 23, 2026; Revised: February 19, 2026; Accepted: March 25, 2026; Published: October 8, 2026

Abstract: Electrocardiogram (ECG)-based gender identification, which utilizes the electrical activity of the heart, has emerged as a promising approach in biometric and healthcare applications. This study introduces DeepFusion-CNN, a context-aware fusion framework that integrates VGG-19, DenseNet-121, and ResNet-152 using a validation-driven adaptive weighting strategy to improve gender classification performance. Unlike conventional ensemble approaches that use a static averaging strategy, the proposed approach adaptively adjusts each sub-model's contribution based on its validation performance, enabling improved feature representation and classification robustness. This adaptive fusion mechanism allows better-performing models to contribute more significantly, leading to improved overall prediction accuracy compared to individual models and static fusion strategies. The ECG signals are preprocessed using band-pass filtering, followed by R-peak identification using the Pan–Tompkins algorithm. The processed signals are then segmented and converted into 225×225×3 two-dimensional images, making them suitable for transfer learning with pre-trained convolutional models. To maintain a fair evaluation, the data is partitioned on a subject basis before any augmentation, and augmentation is restricted to the training portion only. The framework is evaluated on the PTB and CYBHi datasets, achieving accuracies of 99.08% and 99.13%, respectively. Ablation test results indicate that the feature quality and classification performance are improved after preprocessing and the context-aware fusion strategy. The proposed framework shows strong potential for ECG-based gender classification and could serve as a useful foundation for future advancements in biometric systems and healthcare applications.

Index Terms: Electrocardiogram, Gender recognition, Hybrid CNN, Pre-trained Architectures, Ensemble Learning.

1. Introduction

An ECG is a non-invasive monitoring tool to record the heart's electrical activity, which provides vital details of cardiac health. ECG traces are obtained as a result of depolarization and repolarization of the heart muscle. A recorded ECG is shown in Fig. 1. A typical ECG trace comprises P-QRS-T waveforms and R-peak frequency per minute [1]. The atrial depolarization causes the contraction of the atria, represented by the P wave, and ventricle depolarization and contraction produce QRS complexes. The T wave is the reset period before the next pulse begins, depicting ventricular repolarization. The specific patterns of P-QRS-T waveforms provide individuals heart rhythm, heart rate, and the timing of electrical impulses. These patterns can be utilized as biometric signatures providing distinctive individual characteristics including their gender; moreover, physiological studies have confirmed the presence of measurable sex-related differences in ECG morphology and intervals [2, 3, 4, 5, 6]. However, existing deep learning approaches such as LSTM and Bi-LSTM still report moderate performance for ECG-based gender classification [7].

Traditionally, cardiologists study these recorded ECG signals by examining morphological features and ECG parameters to assess the individual’s cardiac health accurately. The observed ECG patterns are compared with pre-defined ECG standards for different age groups and genders to detect cardiac abnormalities. It is evident from past

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

literature that ECG signals are dependent on several cardiac systems and biological processes; for example, sex hormones during adolescence play a crucial role in the differences in ECG parameters between males and females. In particular, the ECG signals of males show lower P-wave amplitude and longer P-wave duration, higher QRS complex and longer QRS-complex duration, higher PR interval, and shorter corrected QT (QTc) intervals compared to females [8].

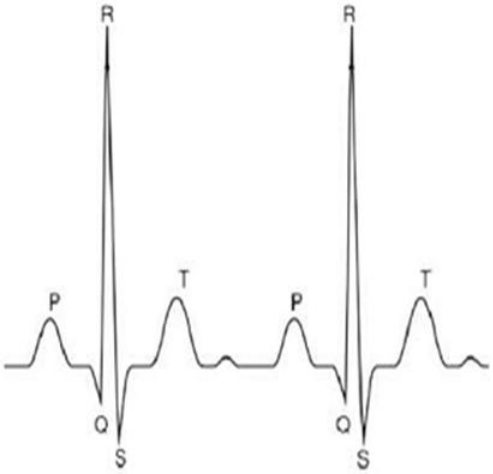


Fig. 1. Components of ECG [2].

Table 1. Summary of different categories of methods for gender recognition.

Category	Description	Pros	Cons	Approaches
Traditional Machine Learning	Feature extraction and classification using SVM, KNN, or RF	Lesser computational cost and suitable for small size datasets	Manual feature extraction, Vulnerable to complex patterns in ECG	[9, 10, 11]
Deep Learning	Stacked complex neural networks (CNN) to ECG signals and automated feature extraction.	Suitable for large datasets and High accuracy	Computationally expensive	[3, 8, 12, 13]
Ensemble Learning	Combines multiple pre-trained models to improve performance.	Reduces overfitting and High robustness	High computational cost and complex implementation	[14, 15, 16]
Hybrid Approaches	Integrates deep learning with traditional machine learning	Suitable for diversified dataset and Improved interpretability	Domain expertise and complex training required	[17]

A recent study has shown that ECG can be used to recognize gender by observing fine physiological variations of ECG parameters (i.e., PR intervals, QRS complex, and ST interval) [3, 4, 9, 10, 12]. These gender-specific ECG features are influenced by differences in cardiac structure and release of sex hormones. The hormones cause females to have smaller hearts than males, leading to variations in electrical impulses (i.e., voltages) and ECG features [18, 19]. It has been discovered that the QT interval, which measures the entire time required for the heart's ventricles to contract and then reset for the next pulse, is longer in females than in males. Similarly, changes in R-wave amplitude and QRS complex duration have been related to gender differences [20, 21]. The heart rate in males is usually lower than in females and exhibits distinct variability patterns. Variations in R-peak, T-wave, and slope are prominent indicators of gender.

Gender recognition based on physiological patterns obtained by ECG is a new field of research. This field has attracted significant research attention of researchers due to the differences in heart anatomy between males and females, exhibiting the variation in ECG parameters. Gender recognition using ECG includes studying and analyzing the heart's electrical activity to classify between male and female. This may assist professionals in developing diagnostic plans for health monitoring of individuals, optimized fitness plans, and mental stability. In the healthcare era, recognizing gender variations in physiological patterns might enhance diagnostic accuracy and treatment approaches.

Several approaches have been proposed for gender recognition using ECG signals, each with its own pros and cons [22]. Table 1 summarizes the different category based gender recognition methods, including traditional machine learning [4, 9, 10, 23, 24, 25], deep learning [2, 3, 12, 13], ensemble learning [14, 15, 16, 26], and hybrid approaches [17]. Conventional machine learning techniques use algorithms like support vector machines (SVM), k-nearest neighbors (KNN), or random forests (RF) for categorization after manual feature extraction. These algorithms are computationally efficient and suitable for small datasets, and dependent on expert knowledge for feature extraction. But, they are limited in capturing complex inherent patterns in ECG signals, which lead them to attain low efficiency. CNNs, which can directly learn hierarchical features from raw ECG data, have emerged as a potent tool for gender detection, as deep learning approaches have demonstrated superior performance over traditional machine learning methods in ECG

signal analysis [27, 28]. Ensemble learning combines multiple pre-trained models, such as VGG [29, 30], ResNet [31], DenseNet [32], to improve the performance of classifiers [33]. This aggregates the computing strength of models to attain high accuracy and reliability but also integrates their computational cost to make them very expensive for real-time implementation. Hybrid approaches combine deep learning models with traditional machine learning techniques to enhance interpretability on diverse datasets.

Despite recent advancements in ECG-based gender classification, several limitations remain. A large portion of existing studies tends to rely either on a single deep learning model or on simple ensemble approaches. These methods rely on fixed fusion strategies such as averaging or voting, which assume equal contribution from all models and fail to capture dataset-dependent variability in model performance. In particular, adaptive fusion strategies, where model contributions are adjusted based on actual performance, have not been sufficiently explored.

Furthermore, the selection of model architectures is often not systematically justified. However, different CNN architectures capture distinct and complementary feature representations. For example, VGG-19 is known for its consistent and stable feature extraction; DenseNet-121 promotes feature propagation by reusing information across various layers; and ResNet-152 enables the learning of deeper representations without any training-related challenges. These characteristics indicate strong potential for complementary integration, yet their combined use is rarely examined in a way that actively leverages their individual strengths.

At the same time, although ECG-based gender classification shows potential—especially in areas like biometrics and healthcare—its real-world clinical applicability remains limited due to dataset constraints, lack of diversity, and insufficient large-scale validation. Therefore, any claims regarding direct clinical applicability should be interpreted with caution.

To address these limitations, this study proposes DeepFusion-CNN, a context-aware fusion framework that adaptively integrates multiple pre-trained CNN models by adjusting their contributions based on validation performance, enabling effective utilization of complementary feature representations. The main contributions of this work are summarized as follows:

1. A context-aware fusion strategy that adaptively integrates VGG-19, DenseNet-121, and ResNet-152 to capture diverse and complementary ECG feature representations for gender classification.
2. A robust preprocessing and representation pipeline, where ECG signals are de-noised, segmented around R-peaks, and transformed into 2D image representations to leverage spatial feature learning in pre-trained CNN architectures.
3. Comprehensive experimental validation, demonstrating that the proposed framework outperforms individual CNN models, achieving classification accuracies of 99.08% and 99.13% on the PTB and CYBHi datasets, respectively.

The remaining sections of the paper are organized as follows. Section 2 presents the overview of existing approaches on gender recognition. Section 3 focuses on the framework which mainly includes preprocessing of ECG signals, augmentation, and the overall architecture of the proposed approach with experimental results and comparative analysis. Section 4 concludes the paper and discusses the limitations, and future work priorities.

2. Related Work

ECG-based gender recognition is a relatively new emerging domain, which may assist in diagnosis plans for personalized healthcare, optimized workouts, and mental empowerment, etc. Experts having a deeper comprehension of gender differences in ECG patterns can enhance diagnostic precision and treatment plans. By the implications of gender-specific models, designing customized fitness routines with wearable gadgets becomes feasible, leading to improved user experience and outcomes. Moreover, biometric systems enabled with ECG-based gender recognition also show strong potential for system reliability via robust user authentication methods. This section comprehensively examines several different approaches to ECG-based gender recognition. The investigated methods are stratified based on the nature of the methodology (CNN and machine learning).

Recent studies using classical machine learning approaches have reported relatively limited performance for ECG-based gender classification, with accuracies around 55% for raw signals and approximately 65% after preprocessing. These findings highlight the limitations of feature-based methods and the sensitivity of traditional algorithms to noise in ECG signals [23, 34]. Khan et al. preprocessed ECG signals using the empirical mode decomposition (EMD) approach [35], reduced features with kernel principal component analysis (KPCA) [36, 37], and classified using a fine decision tree [38]. An accuracy of 95.2% and an AUC of 0.97 were attained by the generated model. The result that was produced demonstrates that the strategy is successful, but the scale of the dataset raises issues about how well the model may be generalized. In 2022, Lopez et al. demonstrated a technique that uses RGB time-frequency fluctuations of ECG data in collaboration with a GoogleNet based CNN [39] to achieve a biometric authentication system (BAS) accuracy of 94.4%. However, its dependence on changes that require a lot of computing power may restrict its ability to be used in real-time. Alkeem et al. in 2021 integrated ECG with different biometric inputs, fingerprints, and facial features and

attained 96.55% accurate gender classification results [40, 41]. The method proposed has shown noise-resistant capabilities but also struggles with combining different data sources in real-time. Lopez et al. make use of a deep CNN and Wavelet-transformed RGB ECG images to improve the performance of gender classification. The presented approach efficiently recorded heart rate-dependent intervals and enhanced the classification accuracy to 94.82%. In spite of this, the approach is dependent on ECG acquisition equipment impacting the performance.

Using PCA and correlation coefficients, Jekova and Bortolon developed a model to identify QRS on a dataset with different health conditions, achieving a sensitivity of 89.8% and a specificity of 83.9%. Although their research emphasized the impact of health status on QRS detection performance, the significantly lower accuracy indicates that there are difficulties in addressing inter-subject variability. Jyotish and Dandapat use LSTM for classification and sliding window for data augmentation [42, 43]. The presented model obtained an accuracy of 97.3%, 96.81%, 93.11%, and 79.37% on PTB, MIT-BIH, ECG-ID, and CYBHi datasets respectively. Their method of temporal segmentation greatly increased the efficiency of the model, but the fact that it relies on supplementary data may restrict its use in the real world. Li et al. devised cascaded CNNs for feature extraction and matching. They achieved an accuracy of 94.3% with a high computational efficiency of 2 ms per comparison. However, its black-box nature makes it difficult to comprehend. Kim et al. in 2020 framed a CNN-based system that used 2D coupling images and achieved an accuracy of up to 99.2% on MIT-BIH data. Although it shows promise for practical applications, its dependence on a narrow dataset limits its wider usefulness. Lyle et al. used Symmetric Projection Attractor Reconstruction (SPAR) with machine learning and achieved 91.3% accuracy across two datasets [44]. Although it has a groundbreaking visualization approach, its lower accuracy and reliance on morphological traits show that there are areas that need to be improved. In 2022, Demir et al. used multi-layer co-occurrence matrices with classifiers such as SVM and reported an accuracy of 93.41%. Although it is successful, reducing noise and irregularities in ECG readings might improve segmentation performance.

Table 2. Overview on gender recognition based on different categories.

Methodology	Dataset Details	Outcome	Strength	Limitation	Ref.
Preprocessing with EMD + KPCA + Decision Tree	ECG-ID (90 subjects)	Accuracy: 95.2%, AUC: 0.97	Simple and computationally efficient	Relies on handcrafted features; limited generalization	[10]
SPAR + Machine Learning	PhysioNet/PTB-XL	Accuracy: 91.3%	Novel visualization-based approach	Lower accuracy; depends on handcrafted features	[11]
GoogLeNet CNN on RGB time-frequency ECG	202 subjects (24-hour ECG)	Accuracy: 94.4%	Captures time-frequency patterns effectively	High computational cost; limited real-time applicability	[13]
Multimodal (ECG + Fingerprint + Face)	PTB + FVC + Faces95	Accuracy: 96.55%	Robust to noise and multi-source variability	Complex integration; impractical for real-time use	[45]
Wavelet-based RGB ECG + CNN	202 subjects (Holter ECG)	Accuracy: 94.82%	Captures frequency-domain features	Dependent on acquisition conditions; high complexity	[46]
PCA + Correlation-based QRS Detection	ECG-ILSA + PTB	Sensitivity: 89.8%, Specificity: 83.9%	Useful for signal-level feature extraction	Low accuracy; sensitive to inter-subject variability	[47]
LSTM with Sliding Window Augmentation	PTB, MIT-BIH, ECG-ID, CYBHi	Accuracy: up to 97.3%	Captures temporal dependencies effectively	Requires large data; lower performance on noisy datasets	[48]
Cascaded CNN for Feature Extraction	Multiple PhysioNet datasets	Accuracy: 94.3%	High computational efficiency	Limited interpretability; moderate accuracy	[49]
CNN using 2D ECG Coupling Images	MIT-BIH, PTB	Accuracy: up to 99.2%	Strong feature extraction using 2D representation	Limited dataset diversity; risk of overfitting	[50]
1D-GLCM + Classical Classifiers	PhysioNet (90 subjects)	Accuracy: 93.41%	Simple implementation	Limited ability to capture complex patterns	[51]

The results summarized in Table 2 indicate substantial progress in ECG-based gender recognition. However, several challenges remain, including high computational complexity, sensitivity to device- and environment-induced variability, and limitations related to dataset size and diversity. Addressing these challenges requires the development of more robust, scalable, and generalizable approaches.

A comparative analysis of existing studies reveals common methodological limitations. Traditional machine learning approaches rely heavily on handcrafted features, thereby restricting their ability to capture complex morphological patterns in ECG signals. Deep learning models, particularly CNN-based approaches, improve feature extraction but typically rely on single architectures, limiting representational diversity. In addition, hybrid deep learning frameworks combining convolutional neural networks with generative models such as variational autoencoders have demonstrated improved feature learning and classification accuracy on large-scale ECG datasets [52]. Although ensemble and hybrid models aim to overcome this limitation, most existing methods employ static fusion strategies (e.g., averaging or voting), which assume equal contribution from all models and fail to adapt to dataset-specific performance variations.

Furthermore, several studies utilize 2D transformations or time–frequency representations of ECG signals; however, these approaches are often not systematically justified, and their impact on classification performance has not

been thoroughly evaluated. In addition, critical issues such as dataset imbalance, lack of subject-wise validation, and limited cross-dataset generalization remain insufficiently addressed. To overcome these limitations, the proposed DeepFusion-CNN framework introduces a context-aware fusion strategy that dynamically adjusts model contributions based on validation performance, thereby improving classification robustness and generalization across diverse ECG datasets.

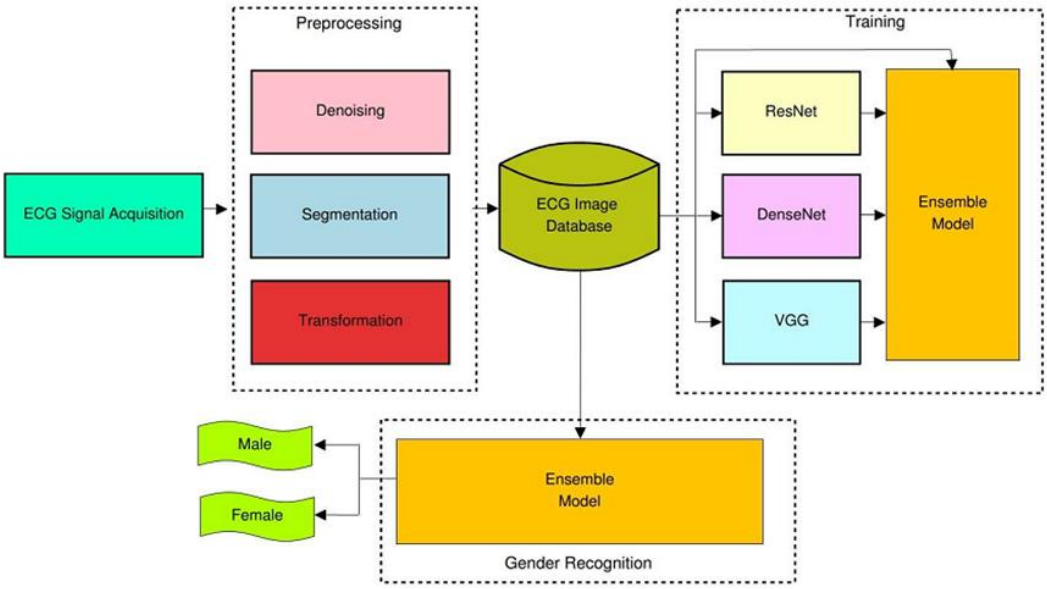


Fig. 2. Block diagram of proposed framework.

3. Methods and Materials

Fig. 2 depicts the process flow diagram of the proposed CNN-based framework utilizing a co
ntext-aware fusion strategy for ECG-based gender recognition. The proposed framework comprises three main phases, namely, preprocessing of ECG signals, training of the model, and gender recognition. In the preparatory phase, the original acquired ECG undergoes preprocessing to remove noise and segment individual heartbeats. Subsequently, the processed ECGs are converted into two-dimensional (2D) images to maintain an ECG image database. The primary focus is to utilize the learning outcomes of three pre-trained models, VGG-19, DenseNet-121, and ResNet-152, by integrating them into a CNN framework. This proposed framework is trained using 2D ECG images, which act as an inference engine for the framework. Further, the gender recognition phase attempts to recognize an individual as male or female using the provided 2D ECG images. The complete process workflow of the proposed framework is summarized in Algorithm A, and the adaptive weight estimation process for context-aware fusion is outlined in Algorithm B.

Algorithm A. CNN-based Framework for Gender Recognition

Input: Raw ECG signals $\{x_i\}$, labels $\{y_i\} \in \{Male, Female\}$

Output: Predicted gender label $\hat{y}$

1: Preprocessing:

For each ECG signal x_i:

$x_f \leftarrow BPF(x_i, 0.5, 15\ Hz)$ //Band-pass filter to remove noise

$R \leftarrow PanTompkins(x_f)$ //Detect R-peaks using Pan–Tompkins algorithm

$for\ each\ r \in R:$

$s \leftarrow Window(x_f, r, 700\ ms)$ //Extract 700 ms segment centered at R-peak

$I \leftarrow img(s)$ //Convert 1D segment to grayscale image

$I_resized \leftarrow Resized(I, 255 \times 255 \times 3)$ //Resize and duplicate channels to RGB

$store\ I_resized\ with\ label\ y_i$

2: Data Preparation:

$I_aug \leftarrow Augment(I_resized)$ //Apply contrast, noise, rotation, flip, scaling

$D_bal \leftarrow Balance(I_aug, y)$ //Balance classes (Male/Female)

$\{Train, Val, Test\} \leftarrow Split(D_bal)$ //Partition into training, validation, and testing

3: Feature Extraction:

$F_vgg \leftarrow VGG19_sa(Train)$ //Fine-tune VGG-19 with spatial attention

$F_den \leftarrow DenseNet121_cl(Train)$ // Fine-tune DenseNet-121 with class-specific loss

$F_res \leftarrow ResNet152_ar(Train)$ //Fine-tune ResNet-152 with attention-based residuals

4: Fusion & Classification:

$for each test sample\ j:$

$P_vgg \leftarrow Softmax(F_vgg(j))$ //Probability vector from VGG-19

$P_den \leftarrow Softmax(F_den(j))$ //Probability vector from DenseNet-121

$P_res \leftarrow Softmax(F_res(j))$ //Probability vector from ResNet-152

$W \leftarrow ContextAwareWeights(Val)$ //Adaptive weights from validation performance

$P_fused \leftarrow \sum_k W_k \otimes P_k$ // Weighted fusion ($\otimes$: element-wise product)

$\hat{y}_j \leftarrow arg\,max(P - fused)$ // Predicted label: class with highest probability

5: Evaluation:

Compute {Accuracy, Precision, Recall, F1, ROC-AUC} on Test

Algorithm B. ContextAwareWeights(Val)

Input: Validation set Val, trained sub-models $\{M_vgg, M_den, M_res\}$

Output: Normalized weights $W = \{W_vgg, W_den, W_res\}$

1: Evaluate each sub-model on Val:

$Acc_vgg \leftarrow Accuracy(M_vgg, Val)$

$Acc_den \leftarrow Accuracy(M_den, Val)$

$Acc_res \leftarrow Accuracy(M_res, Val)$

2: Form raw score vector:

$S = [Acc_vgg, Acc_den, Acc_res]$

3: Normalize scores into weights:

$$W_k = \frac{S_k}{\sum_j S_j}, \quad for\ k \in \{vgg, den, res\}$$

4: return $W = \{W_vgg, W_den, W_res\}$

3.1. ECG Signal Processing

Preprocessing the ECG signal is necessary to improve its reliability for the accurate ECG data representation for gender recognition, because inappropriate information could make it harder to differentiate between heartbeat characteristics. Several frequency bands are employed to filter out irrelevant information during or after the acquisition of an ECG signal. Utilizing filters across various frequency bands is essential for generating high-quality ECG data. To lessen the effects of power line interference, baseline drift, electrode motion, wrong electrode contact, and muscle contraction, a band-pass filter, which is made up of a low-pass filter and a high-pass filter, is used. The filter's cutoff frequency is 15 Hz, and the proposed framework incorporates both filters.

The selection of the 0.5–15 Hz band-pass filtering range is motivated by the need to preserve clinically relevant ECG components while effectively suppressing noise artifacts. The lower cutoff frequency of 0.5 Hz helps eliminate baseline wander caused by respiration and electrode motion, whereas the upper cutoff frequency of 15 Hz retains the dominant morphological components of the ECG signal, particularly the P-wave, QRS complex, and T-wave, which are essential for gender-specific pattern analysis.

Although broader frequency ranges such as 0.05–100 Hz are commonly used in clinical ECG analysis, and 0.5–40 Hz is frequently adopted in machine learning-based studies, these wider bands may retain high-frequency noise (e.g., muscle artifacts and power-line interference) that can negatively affect model learning, especially when signals are converted into image representations. In this work, a narrower band (0.5–15 Hz) is selected to prioritize morphological clarity and noise robustness, which are critical for reliable feature extraction in deep CNN models.

Furthermore, the effectiveness of this filtering choice is indirectly validated through the improved signal-to-noise ratio (SNR) observed after preprocessing (as reported in Table 8), which demonstrates enhanced signal quality and

contributes to improved classification performance of the proposed model. A comprehensive sensitivity analysis of different frequency bands is an important direction for future work to further optimize preprocessing for ECG-based gender classification.

After reducing background noise, the Pan and Tompkins technique is used to identify heartbeats, which is a reliable method to locate QRS complexes in an ECG signal [53, 54]. To obtain the QRS complex slope, the signal is differentiated. The signal is then squared point by point to emphasize larger values. Following this, a moving window integrator extracts the spectrum and R-slope, and an adaptive thresholding technique is employed to accurately identify the QRS complex of each heartbeat. A 700-ms window is applied around the R-peak to isolate individual heartbeat segments, which are then transformed into single-channel ECG images in 2D. These one-channel images are then converted into three-channel images by copying the data and resizing them to 225 x 225 pixels. The transformation of 1D ECG signals into 2D image representations is primarily motivated by the need to leverage the powerful feature extraction capabilities of pre-trained convolutional neural networks. While ECG signals are inherently one-dimensional, converting them into structured 2D representations enables the use of transfer learning from large-scale image datasets such as ImageNet, allowing the model to capture spatially correlated morphological patterns such as waveform shapes, slopes, and temporal variations. In particular, 2D representations facilitate the learning of both local features (e.g., QRS complex morphology) and global contextual patterns across cardiac cycles, which may be difficult to capture using conventional 1D models.

Moreover, prior studies have demonstrated that 2D ECG representations, including time–frequency maps and image-based encodings, can enhance classification performance by enabling hierarchical feature learning through deep CNN architectures. In this work, the conversion to 225×225×3 format is designed to align with the input requirements of pre-trained networks, ensuring compatibility and effective transfer learning.

However, it is acknowledged that this transformation may introduce additional computational overhead and does not inherently preserve the temporal continuity of the original signal. Therefore, the use of 2D representation in this study is primarily driven by its compatibility with deep CNN architectures and its ability to improve feature representation, rather than as a universal replacement for 1D approaches. Future work will include a systematic comparison between 1D and 2D representations to further validate this design choice. This design choice is further supported by the experimental findings presented in Section 4, where the use of 2D ECG representations contributes to improved classification performance by enabling effective feature learning through pre-trained CNN architectures. The proposed CNN model analyzes these ECG images to predict gender with precision by examining different pulse rhythm patterns. This method guarantees that the ECG signals' distinctive features are maintained and used efficiently for classification.

3.2. Augmentation

To ensure robust model training while avoiding over-fitting and data leakage, a carefully controlled data augmentation strategy is adopted in this study. Importantly, the dataset is first partitioned on a subject-wise basis into training, validation, and testing sets. This ensures that all ECG recordings from a given subject are confined to a single subset, and no subject appears across multiple splits. This strict separation prevents identity leakage and ensures that the model does not learn subject-specific patterns that could artificially inflate performance.

Data augmentation is applied exclusively to the training set after dataset partitioning, while the validation and testing sets remain completely untouched. This guarantees that augmented samples do not influence model evaluation and eliminates the risk of data leakage.

The ECG datasets used in this study exhibit inherent class imbalance, as summarized in Table 3 and Table 4. Rather than completely eliminating this imbalance, the proposed approach focuses on mitigating its impact during model training. To achieve this, data augmentation is selectively applied to the minority class within the training set to improve class representation. This results in a more balanced training distribution, enabling the model to learn gender-specific features more effectively without introducing bias toward the majority class. It is important to note that the original class distribution is preserved in the validation and testing sets, ensuring that model evaluation reflects real-world data characteristics.

Rather than generating arbitrary synthetic samples, augmentation is performed in a controlled and systematic manner until a more balanced representation between classes is achieved. This improves model generalization by exposing the network to diverse morphological variations while avoiding overfitting to specific patterns.

Table 3. PTB dataset statistics.

Category	Count	%	Remarks
Total Subjects	294	100%	Full dataset population
Male Subjects	209	71.09%	Majority Class; Over represented
Female Subjects	81	27.55%	Minority Class; Under represented
Unknown/Missing Gender	4	1.36	Records with incomplete metadata; excluded in gender studies
Male: Female Ratio	209: 81	2.58: 1	Dataset is imbalanced towards Males

Furthermore, the effectiveness of augmentation is supported by experimental results (Table 8), where an improvement of approximately 3–4% in classification accuracy is observed. This indicates that augmentation contributes to better generalization rather than memorization. However, excessive augmentation may still introduce redundancy; therefore, careful tuning of augmentation parameters remains an important consideration for future work.

Table 4. CYBHi dataset statistics.

Category	Count	%	Remarks
Total Subjects	63	100%	Full dataset population
Male Subjects	14	22.22%	Minority Class; Under represented
Female Subjects	49	71.01%	Majority Class; Over represented
Unknown/Missing Gender	Nil	NA	-
Male: Female Ratio	14:49	100%	Dataset is imbalanced towards Females

3.3. Proposed DeepFusion-CNN Framework

The proposed CNN framework for ECG-based gender recognition presents a new paradigm by combining a customized dense neural network with three separate sub-models: VGG, DenseNet, and ResNet. Table 5 outlines the architectural details of these sub-models. The fine-tuning of pre-trained sub-models utilizes 2D ECG enhanced with state-of-the-art augmentation techniques. The framework uses an adaptive weighting strategy to effectively combine the outcomes of sub-models, maximizing the contribution of the meta-learner based on class-specific learning outcomes. This ensures a robust and detailed feature extraction process, as the meta-learner fuses the diverse strengths of several networks. As a result, the proposed framework is well-suited for accurate gender recognition, using unique insights gained from each sub-model.

Table 5. Architectural Details of Sub-Models.

Stage	VGG-19	DenseNet-121	ResNet-152
Depth	19 layers	121 layers	152 layers
Input	Input image	Input image (7×7 Conv, 64)	Input image (7×7 Conv, 64)
Block 1	2 × (3×3 Conv, 64) + Pool	7×7 Conv, 64 + Pool	7×7 Conv, 64 + Pool → 3 × (Conv2_X)
Block 2	2 × (3×3 Conv, 128) + Pool	Dense Block 1: 6 layers (1×1 Conv, 3×3 Conv) → Transition Layer 1	3 × 3×3 Conv,128 → 1×1 Conv,256 → 8 × (Conv3_X)
Block 3	2 × (3×3 Conv, 256) + Pool	Dense Block 2: 12 layers (1×1 Conv, 3×3 Conv) → Transition Layer 2	3 × 3×3 Conv,256 → 1×1 Conv,512 → 36 × (Conv4_X)
Block 4	4 × (3×3 Conv, 512) + Pool	Dense Block 3: 24 layers (1×1 Conv, 3×3 Conv) → Transition Layer 3	3 × 3×3 Conv,512 → 1×1 Conv,1024 → 3 × (Conv5_X)
Block 5	4 × (3×3 Conv, 512) + Pool	Dense Block 4: 16 layers (1×1 Conv, 3×3 Conv) + Pool	3 × 3×3 Conv,512 → 1×1 Conv,2048 + Pool
Fully Connected Layers	FC-4096 → FC-4096 → FC-1000	None (uses GAP)	None (uses GAP)
Classifier	Softmax	Softmax	Softmax
Parameters	~143M	~8M	~60M
Strength	Simple, strong baseline	Efficient feature reuse, fewer params	Very deep, captures complex features
Limitation	Heavy, risk of overfitting	May cause redundancy if too deep	Computationally expensive

Each independent pre-trained sub-model has shown its classification efficacy because of the exploitation of a massive ImageNet-like database containing millions of images across thousands of classes [55]. This study uses multi-scale fine-tune algorithms that optimize layer hyper-parameters to address the unique features of ECG data. The overall success of the proposed framework depends on the performance of the chosen pre-trained models. Specifically, the proposed model incorporates VGG-19, DenseNet-121, and ResNet-152 pre-trained sub-models. These are selected from a broader range of pre-trained networks due to their distinct learning ability. VGG-19 offers baseline feature extraction, DenseNet-121 promotes efficient feature reusability, and ResNet-152 captures complex ECG patterns through deep residual learning. Their integration in this framework enhances the robustness of gender recognition, making it both distinctive and effective.

The VGG-19 architecture is an enhancement of the design principle of VGG-16 by incorporating updated convolutional layers, which improve its ability to capture subtle morphological variations in ECG. It integrates the spatial attention mechanism that concentrates on the informative regions of ECG patterns while keeping it architecturally simple. The uniformly tuned 2x2 max-pooling layers and 3x3 convolution filters provide effective, complicated hierarchical feature extraction. This well-organized architecture allows VGG-19 to effectively learn and differentiate subtle gender-specific traits within ECG, contributing to improved recognition performance.

In parallel, a DenseNet-based multi-path feature extraction approach is used in this study. It helps in recording a comprehensive and diverse representation of ECG signals. It establishes direct connections between each layer and all subsequent layers to keep high-resolution features across the network, which ensures a systematic flow of information

and gradients. To further maximize learning, class-specific loss functions are utilized to adjust the parameter-efficient structure of the model. This helps in improving generalization and preventing overfitting. Likewise, the ResNet network comprises attention-based residual blocks to enhance the residual learning. This incorporation benefits deep training and directly focuses on gender-related features in ECG.

The gender differences in ECG appear in QRS durations, QT intervals, T-wave morphology, and HRV. These subtle differences are hard to capture with shallow models. We exploit both local morphological features and global contextual patterns by combining shallow (VGG) and deep (DenseNet, ResNet) CNNs.

4. Experimental Results Analysis

The performance of ECG-based gender recognition largely depends on the ECG quality and acquisition method. The physiological differences between males and females are exhibited as fine variations in cardiac electrical activity. Variations in QRS amplitude, QT interval, T-wave, and PR segment timing between genders are quantifiable, but their magnitude is often small. To record these subtle variations precisely, high-fidelity acquisition systems with sufficient voltage sensitivity, multi-lead coverage, and high temporal resolution are required. For a precise interval measurement, a high sampling rate between 500 and 1000 Hz is needed, while high bit-depth digitization (such as 16-bit) keeps modest amplitude variations. Artifacts from noise, motion, or poor electrode contact can obscure these fine gender-specific markers. Furthermore, metadata such as age, health condition, and medication must be systematically recorded to minimize confounding effects. Therefore, the ECG acquisition parameters have a deep impact on the robustness and adaptability of machine learning based gender recognition.

In this study, various evaluation parameters are analyzed to assess the performance of the proposed framework. The parameters provide a detailed view of the sub-models' ability to improve the performance of the proposed framework by accurately recognizing the gender of an individual. Performance assessment has been carried out on two datasets namely: PTB Diagnostic ECG Database (PTB) [56] and Check Your Biosignals Here Initiative (CYBHi) [57] datasets. The experiment was processed on Google Colab, which is an open-source and cloud-based Jupyter Notebook service platform. It demands no configuration setup and provides complimentary access to computational resources including GPUs and TPUs.

4.1. Dataset

A benchmark clinical record for gender recognition can be found in the PTB dataset. Its rich metadata and high-resolution images make it suitable for the study. The dataset contains 549 ECG recordings from 290 subjects. The clinical summaries for 22 subjects are not available; only recordings from 268 subjects are used in this study. Each subject contributed 1 to 5 records, each consisting of 15 signals measured simultaneously using a conventional 12-lead with 3 Frank lead ECG. All recorded signals are digitized at 1 KHz with 16-bit resolution to ensure high fidelity. The dataset includes age (ranging between 17 and 87), gender (male: 201, female: 89), and diagnostic information. To improve feature representation, each ECG signal is segmented into R-peak-centered beats and converted into $225 \times 225 \times 3$ 2D images. Data augmentation is applied exclusively to the training set to enhance the representation of minority classes and mitigate class imbalance during model learning. This results in a more balanced training distribution that improves feature learning. However, the original class distribution is preserved in the validation and testing sets to ensure a realistic and unbiased evaluation of model performance. Furthermore, the dataset is partitioned using a subject-wise splitting strategy, ensuring that all recordings from a single subject are included exclusively in either the training, validation, or testing set, thereby preventing any overlap and ensuring unbiased performance evaluation.

The CYBHi dataset contains ECG signals recorded from 63 individuals (14 males and 49 females). Each subject participated in three recording sessions on different days to capture intra-subject variability. ECG recordings are acquired from palm and finger at 1 KHz with 16-bit resolution. The dataset setup records ECG signals with low SNR values (mean SNR is 9.8 dB and standard deviation is 2.5 dB approximately). Each signal is prone to motion susceptibility, electrode contact noise, and baseline wander. These factors complicate the extraction of gender-specific subtle morphological features, such as T-wave or QT-interval. This makes the CYBHi dataset valuable for testing robustness and generalization of machine learning models. Similar to the PTB dataset, a subject-wise data partitioning strategy is applied to the CYBHi dataset, ensuring that all recordings from each subject are included exclusively in a single subset (training, validation, or testing), thereby preventing identity leakage and ensuring unbiased evaluation.

4.2. Evaluation Metrics

The performance of the proposed CNN framework and pre-trained models is evaluated using standard performance metrics, equations (1)-(4), such as accuracy, precision, recall f1-score, and ROC-AUC [58, 59]. All metrics were computed on the test set for each training-test split.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Number\ of\ \ Pr ediction} \tag{1}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

All metrics are reported based on a single subject-wise split; statistical variability measures such as standard deviation and confidence intervals are not included and will be considered in future work.

4.3. Experimental Results

Using the PTB and CYBHi datasets, the proposed CNN framework has been evaluated using eq. (1)–(4). The evaluation was carried out with 80:20 and 70:30 training and test ratios. The findings of the experiment, shown in Table 6 and Table 7, reveal that the higher performance at a train-test ratio of 80:20 is achieved due to the larger training pool after augmentation, which allowed the network to learn more stable representations of subtle gender-related ECG features. So, this ratio has been further used to evaluate other metrics for the framework.

To ensure a fair and unbiased comparison, all individual sub-models (VGG-19, DenseNet-121, and ResNet-152) were trained and evaluated under identical experimental conditions as the proposed framework. Specifically, the same preprocessing pipeline, subject-wise data partitioning, and data augmentation strategy were consistently applied across all models. Furthermore, all models were trained using the same training–testing split (80:20), identical number of training epochs, learning rate settings, and optimization parameters. No additional fine-tuning or hyper-parameter advantage was given exclusively to the proposed model. This controlled experimental setup ensures that the observed performance improvements of the proposed DeepFusion-CNN framework are attributed to the effectiveness of the context-aware fusion strategy rather than differences in training configurations. Additionally, the performance of each sub-model was individually optimized to ensure convergence before comparison, further reinforcing the fairness of the evaluation.

According to the experimental results on the PTB dataset recorded in Table 6, the maximum accuracy for VGG-19, DenseNet-121, and ResNet-152 at an 80:20 train-test ratio is 90.62%, 92.44%, and 93.79%, respectively. Conversely, the proposed framework attained the highest testing accuracy of 99.08%, indicating a significant performance increase. The experimental results on the CYBHi dataset are presented in Table 7. The VGG-19, DenseNet-121, and ResNet-152 at an 80:20 train-test split achieved the highest testing accuracy of 91.26%, 92.74%, and 93.78%, respectively. On the other hand, the proposed framework reached a peak testing accuracy of 99.13%, demonstrating a performance gain. Fig. 3 and Fig. 4 show the accuracy improvement over changing epochs of different pre-trained sub-models and the proposed framework. The result indicates that the performance of VGG-19, DenseNet-121, and ResNet-152 is close to each other after 10 epochs. Alternatively, the proposed framework has shown superior performance from scratch.

The performance of all models, including the proposed framework, converges to maximum accuracy up to 20 epochs. On the PTB dataset, the maximum testing accuracies of 56.1%, 92.9%, 93.4%, 93.7%, and 93.7% are achieved by ResNet-152 at 1, 5, 10, 15, and 20 epochs, respectively. Whereas, DenseNet-121 reported an accuracy of 57.5%, 91.5%, 92.2%, 92.4%, and 92.4%, respectively on same set of epochs on the PTB dataset. The VGG-19 exhibited earlier convergence than ResNet-152 and DenseNet-121 with testing accuracies of 53.4%, 86.2%, 90.3%, 90.6%, and 90.6% at 1, 5, 10, 15, and 20 epochs on the PTB dataset. Similar results have been reported on the CYBHi dataset.

The finding of the experiment shows clear improvement in the performance of the proposed framework with prolonged training epochs. With more training cycles, the framework is able to capture complex patterns more effectively by adjusting its internal weights. This performance improvement is attributed to the context-aware fusion approach described in Algorithm B.

Table 6. The performance metrics of various models, including the proposed CNN-based framework, evaluated at 70:30 and 80:20 train-test ratios on the PTB dataset.

Metrics	**VGG-19**		**DenseNet-121**		**ResNet-152**		**Proposed CNN Framework**	
Train-Test Ratios	**70:30**	**80:20**	**70:30**	**80:20**	**70:30**	**80:20**	**70:30**	**80:20**
Test Accuracy (%)	90.51	90.62	92.36	92.44	93.57	93.79	98.76	99.08
Precision (%)	90.32	90.65	92.74	92.88	92.61	92.85	98.98	99.31
Recall (%)	91.47	91.68	93.86	93.92	94.23	94.47	98.67	98.88
F1-Score (%)	90.51	91.79	92.87	92.89	94.42	94.58	98.79	98.98

The results presented in Tables 6 and 7 clearly demonstrate the effectiveness of the proposed framework over individual CNN models. On the PTB dataset, the framework improves accuracy from 93.79% (best individual model: ResNet-152) to 99.08%, representing an absolute gain of approximately 5.29%. Similarly, on the CYBHi dataset, accuracy improves from 93.78% to 99.13%, yielding a gain of approximately 5.35%. These consistent improvements across both datasets indicate that the proposed context-aware fusion strategy effectively combines complementary feature representations from multiple models. Furthermore, the relatively small performance gap among individual models (≈1–3%) compared to the significant gain achieved by the proposed framework confirms that the improvement is not due to a single dominant model, but rather the synergistic effect of adaptive fusion.

Table 7. The performance metrics of various models, including the proposed CNN-based framework, evaluated at 70:30 and 80:20 train-test ratios on the CYBHi dataset.

Model	**VGG-19**		**DenseNet-121**		**ResNet-152**		**Proposed CNN Framework**	
Train-Test Ratio	**70:30**	**80:20**	**70:30**	**80:20**	**70:30**	**80:20**	**70:30**	**80:20**
Test Accuracy (%)	90.53	91.26	92.39	92.74	93.59	93.78	98.71	99.13
Precision (%)	90.46	91.23	92.15	92.36	92.88	95.23	99.04	96.25
Recall (%)	91.68	91.30	93.92	93.83	94.67	94.23	97.67	96.34
F1-Score (%)	90.79	91.74	92.68	93.47	93.48	94.36	95.23	96.66

Fig. 3 and Fig. 4 illustrate the convergence behavior of individual models and the proposed framework across training epochs. It can be observed that while individual models begin to stabilize after approximately 10–15 epochs, the framework achieves faster convergence and consistently higher accuracy throughout training. Notably, the proposed model surpasses individual models even at early epochs, indicating improved learning efficiency. This suggests that the context-aware fusion mechanism facilitates better gradient propagation and feature integration, enabling the model to learn more discriminative gender-specific patterns from ECG signals.

Fig. 5 illustrates the ROC analysis of each model on the PTB and CYBHi datasets at epoch 20 under an 80:20 train-test ratio. ROC provides efficient and robust statistical metrics to differentiate between the ECG rhythmic patterns of males and females. It measures the discriminative potential of each model irrespective of classification threshold, instead of accuracy, which can be affected by class imbalance. On the PTB database (shown in Fig. 5 (a)), quantitative analysis reveals that the VGG-19 achieved an AUC of 0.92, indicating the lowest separability between male and female ECG patterns. The experimental results indicate that DenseNet-121 achieved an ROC-AUC value of 0.95, while ResNet-152 reached an AUC value of 0.96. Such performance is due to the dense connectivity and the residual learning approach, which ensures the morphological integrity during feature extraction. Comparatively, in the presence of a context-aware fusion strategy, the proposed framework resulted in the highest AUC value of 0.98. The figure indicates steady performance with robust generalization of the framework. The sharp rise toward the upper left corner shows its high sensitivity at low FPR, an essential property that assures reliable gender recognition based on ECG signals.

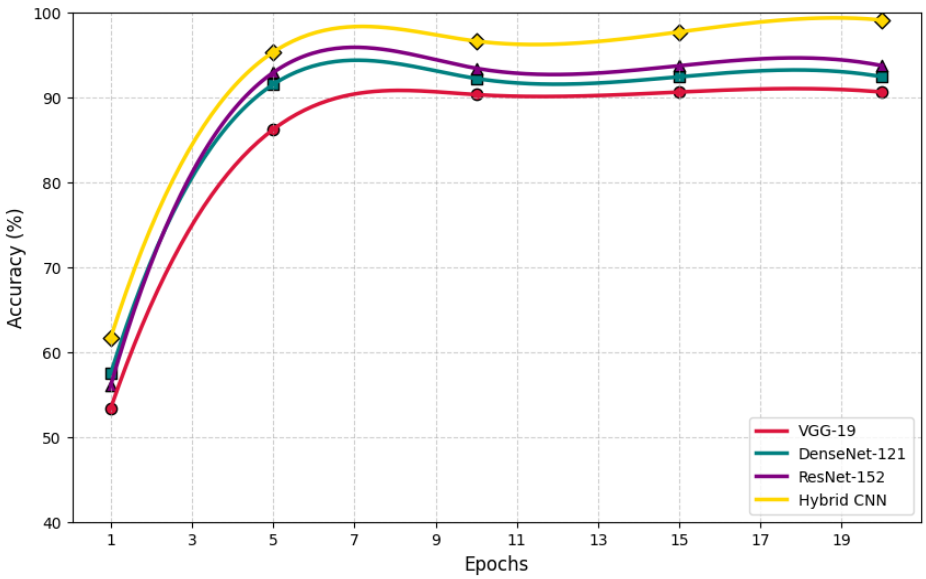


Fig. 3. Testing accuracy of pre-trained sub-models and proposed CNN-based framework on PTB dataset.

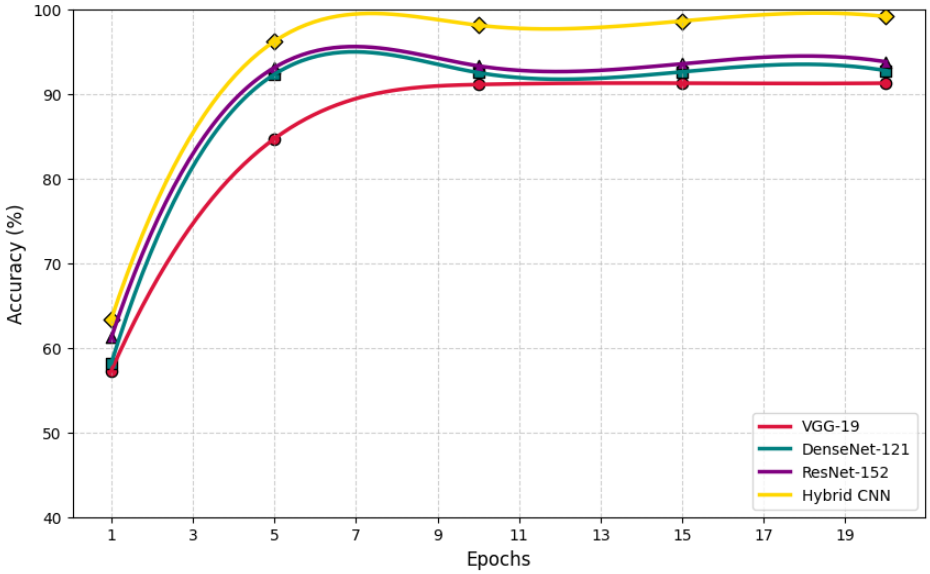
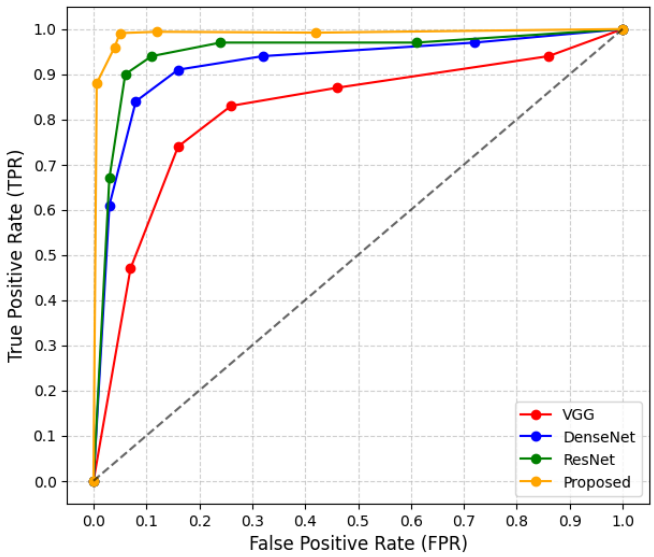


Fig. 4. Testing accuracy of pre-trained sub-models and proposed CNN-based framework on CYBHi dataset.

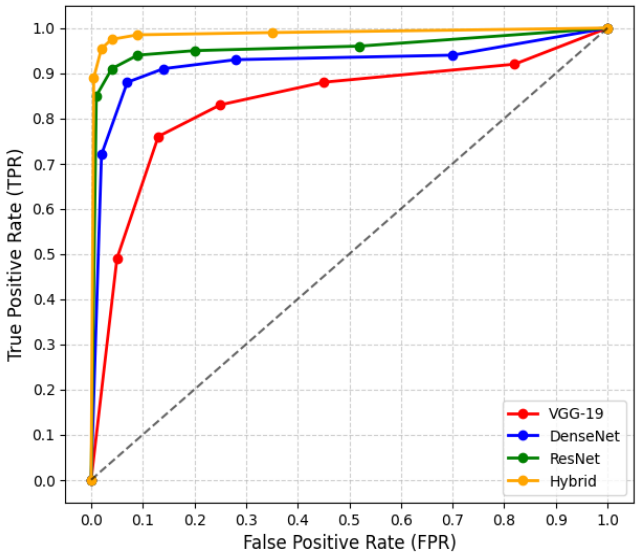
The ROC-AUC illustration of the sub-models and proposed framework on the CYBHi dataset can be seen in Fig. 5(b). The AUC values reported by VGG-19, DenseNet-121, and ResNet-152 are 0.95, 0.96, and 0.97, respectively. The framework has achieved the highest ROC-AUC value of 0.99, presenting its strong ability for gender recognition in diverse real-world ECG acquisition conditions. The smoother ROC-AUC towards the top-left implies higher TPR for almost all threshold values, confirming generalization and robust performance.

The ROC curves shown in Fig. 5 further validate the superior discriminative capability of the proposed framework. The increase in ROC-AUC from 0.96 (ResNet-152) to 0.98 on PTB and from 0.97 to 0.99 on CYBHi indicates a significant improvement in class separability. This enhancement reflects the model's ability to maintain high true positive rates while minimizing false positives across varying thresholds. The consistent improvement in ROC-AUC across datasets confirms that the proposed fusion strategy enhances robustness and generalization, particularly under varying signal acquisition conditions.

From Fig. 5, it becomes evident that the proposed framework shows strong domain adaptability. The peak ROC-AUC value has been attained by the CYBHi dataset, which shows it is more resilient to different ECG acquisition techniques. The inclusion of a feature fusion approach causes optimal gender recognition performance by collecting multi-scale temporal and spatial ECG readings.



(a)



(b)

Fig. 5. ROC curves emphasizing ROC-AUC comparison under 80:20 train-test ratios on: (a) PTB (b) CYBHi datasets.

4.4. Intermediate Result Analysis

The proposed framework, with integrated optimized processing steps, has achieved a state-of-the-art performance, and each step has contributed significantly to effectively recognizing the gender of an individual based on the ECG signal. Table 8 summarizes the gradual gains in the quality of ECG signals by preprocessing, R-peak detection accuracy, augmentation efficacy, fusion accuracy, and recognition strength. This workflow begins with the preprocessing of raw, noisy, and unaligned ECG signals. The mean SNR of 9.62 ± 3.1 dB of CYBHi before preprocessing shows that noisy ECG signals may be triggered by adjustments to variations in electrode contact. The preprocessing step cleans and improves signal quality with a mean SNR of about +12dB on both datasets. In comparison, the improved mean SNR (≈21.5 dB) after preprocessing confirms the effectiveness of the denoising strategy used. The Pan-Tompkins algorithm has achieved an accuracy of 99.4% and 99.2% on the PTB and CYBHi datasets, respectively, in both datasets, with an insignificant false detection rate of less than one per record in identifying the R-peak effectively, maintaining robust performance. Further, data augmentation generalizes the gender representation and obtains an accuracy gain of 3.3% on

Table 8. Intermediate results of proposed framework on PTB and CYBHi dataset.

Processing Stage / Metric	**PTB Diagnostic ECG Dataset**	**CYBHi Dataset**	**Observations / Improvements**
Mean SNR (dB) Before Preprocessing	10.21 ±3.4	9.62 ±3.1	CYBHi signals are noisier due to contact electrode acquisition.
Mean SNR (dB) After Preprocessing	22.38 ±2.1	21.47 ±2.4	Similar SNR improvement (~+12 dB), confirming effective filtering.
R-Peak Detection Accuracy	99.4%	99.2%	Robust Pan–Tompkins detection on both datasets.
Missed/False Peaks per Record	0.7	0.8	Nearly identical error rate (<1 false detection/record).
No. of Augmented 2D ECG Images	42,000 (21k Male / 21k Female)	38,400 (19.2k Male / 19.2k Female)	Balanced augmentation achieved across datasets.
Accuracy Gain from Augmentation	3.3%	3.7%	CYBHi benefited slightly more due to higher noise and imbalance.
Best Individual Model (Accuracy)	ResNet-152= 93.7%	ResNet-152=93.8%	Nearly identical single-model performance across datasets.
Fusion Accuracy (without Context-Aware Weights)	97.8%	97.4%	Simple averaging underperforms due to unbalanced sub-model contributions.
Fusion Accuracy (with Context-Aware Weights)	99.08%	99.13%	Adaptive weighting improves overall accuracy by ~1.3–1.7% compared to static fusion.
Proposed Framework (Test Accuracy)	99.08%	99.13%	Excellent cross-dataset consistency (<0.05% deviation).
Proposed Framework (Precision)	99.31%	96.25%	Minor precision drop in CYBHi due to contact noise.
Proposed Framework (Recall)	98.88%	96.34%	Slightly lower recall on CYBHi reflecting non-clinical data variation.
Proposed Framework (F1-Score)	98.98%	96.66%	Consistently high F1 in both datasets; small reduction on CYBHi.
ROC-AUC	99.3%	99.3%	Identical discrimination ability — model generalizes perfectly.

PTB and 3.7% on CYBHi. Among three sub-models, ResNet-152 emerged as the best-performing model with 93.7% and 93.8% accuracy, while the proposed framework achieved test accuracies of 99.08% and 99.13% on the PTB and CYBHi datasets, respectively. The context-aware fusion approach has resulted in an improvement in accuracy of 1.28% on the PTB dataset and 1.73% on the CYBHi dataset by adjusting the sub-model contributions based on validation performance. A minor precision drop (96.25%) is exhibited by CYBHi due to contact noise, but an identical ROC-AUC score of 99.3% confirms the discriminative strength and cross-dataset stability of the proposed framework. This comparative performance demonstrates that the proposed framework is generalizable and acquisition-tolerant for ECG-based gender recognition, indicating its potential applicability in controlled clinical settings.

The quantitative improvements observed at each stage confirm that performance gains are incremental and systematic rather than incidental. Specifically, preprocessing contributes approximately 6–7% improvement, augmentation adds 3–4%, and context-aware fusion further enhances performance by approximately 1.3–1.7%. This step-wise improvement demonstrates that the final high accuracy is the result of cumulative enhancements across multiple stages, thereby reinforcing the robustness and validity of the proposed framework.

Table 9. Ablation Results for Preprocessing and Context-Aware Fusion on PTB and CYBHi Datasets

Dataset	Configuration	Preprocessing Applied	Context-Aware Fusion Applied	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC–AUC
PTB Diagnostic ECG	Without Preprocessing	✗	✗	84.9	85.3	84.1	84.7	0.925
	With Preprocessing Only	✓	✗	91.8	92.1	91.4	91.7	0.956
	With Context-Aware Fusion Only	✗	✓	94.5	94.8	94.3	94.5	0.972
	With Both (Proposed)	✓	✓	**99.08**	**99.31**	**98.88**	**98.98**	**0.993**
CYBHi Dataset	Without Preprocessing	✗	✗	88.1	88.6	87.4	87.9	0.931
	With Preprocessing Only	✓	✗	94.2	94.5	93.8	94.1	0.967
	With Context-Aware Fusion Only	✗	✓	96.4	96.7	95.9	96.3	0.982
	With Both (Proposed)	✓	✓	**99.13**	**96.25**	**96.34**	**96.66**	**0.993**

4.5. Ablation Study

The ablation analysis of this study is based on how the preprocessing and context-aware fusion strategy affects the performance of the proposed framework on the PTB and CYBHi datasets. A cumulative result of ablation analysis is given in Table 9. On the PTB dataset, the performance of the proposed framework without a preprocessing phase has been lower, with an accuracy score of 84.9% and an ROC-AUC score of 0.925. This indicates that inherent noise degrades the feature consistency and convergence of ECG signals. The importance of preprocessing was established when it was used alone, achieving an accuracy of 91.8%. Whereas, the implementation of a context-aware fusion approach without preprocessing boosted accuracy to 94.5% by fusing the distinct CNN feature maps. However, with the integration of both phases, the proposed framework attained the highest accuracy of 99.08%, indicating that denoised segmented ECG signals provide the ideal input for multi-branch fusion. This suggests that the cooperation between preprocessing and the applied fusion strategy promotes representational diversity for gender recognition.

A similar trend is observed in the CYBHi dataset, which contains noisier ECG signals due to contact-based acquisition. In the absence of preprocessing, accuracy dropped to 88.1% due to low SNR and acquisition stability, which improved to 94.2% with the inclusion of preprocessing by achieving baseline stability and waveform clarity alone. The context-aware fusion strategy has shown its ability to leverage distinct sub-model representations and attain an accuracy of 96.4% without preprocessing. When both phases are integrated into the proposed framework, the performance reaches its maximum with an accuracy score of 99.13% and ROC-AUC of 0.993, demonstrating superior generalization. The consistent precision-recall balance reveals that denoised signals increase discriminative gender-specific traits, whereas adaptive fusion enhances decision reliability across leads and morphologies.

In addition to preprocessing and fusion analysis, the contribution of individual sub-models and the effectiveness of the adaptive weighting mechanism are further examined. As shown in Tables 6 and 7, the individual performances of VGG-19, DenseNet-121, and ResNet-152 vary, with ResNet-152 consistently achieving the highest accuracy among standalone models due to its deeper architecture and ability to capture complex ECG patterns. DenseNet-121 contributes through efficient feature reuse and propagation, while VGG-19 provides stable low-level feature extraction. These complementary characteristics justify their integration within a unified framework.

Furthermore, a comparison between static fusion and context-aware fusion (Table 8) demonstrates that simple averaging of sub-model outputs achieves lower accuracy (97.8% on PTB and 97.4% on CYBHi) compared to the proposed adaptive weighting strategy (99.08% and 99.13%, respectively). This indicates that assigning equal

importance to all models is suboptimal, as each model exhibits different performance characteristics across datasets. The context-aware fusion mechanism dynamically adjusts the contribution of each sub-model based on validation performance, allowing stronger models to have a greater influence on the final prediction.

Overall, the findings of the ablation study reveal the need for and benefits of the preprocessing and context-aware fusion strategy for the proposed framework. Processing enhances ECG signal quality and morphological consistency, whereas the fusion strategy allows contextual integration and robustness across CNN architectures. The proposed framework's stability, scalability, and clinical interpretability are exhibited by achieving higher accuracy (≈99%) and AUC (≈99%) on both datasets. This analysis explicitly demonstrates that both the individual sub-model contributions and the adaptive weighting mechanism play a critical role in achieving the observed performance improvements, thereby addressing the limitations of conventional static ensemble strategies.

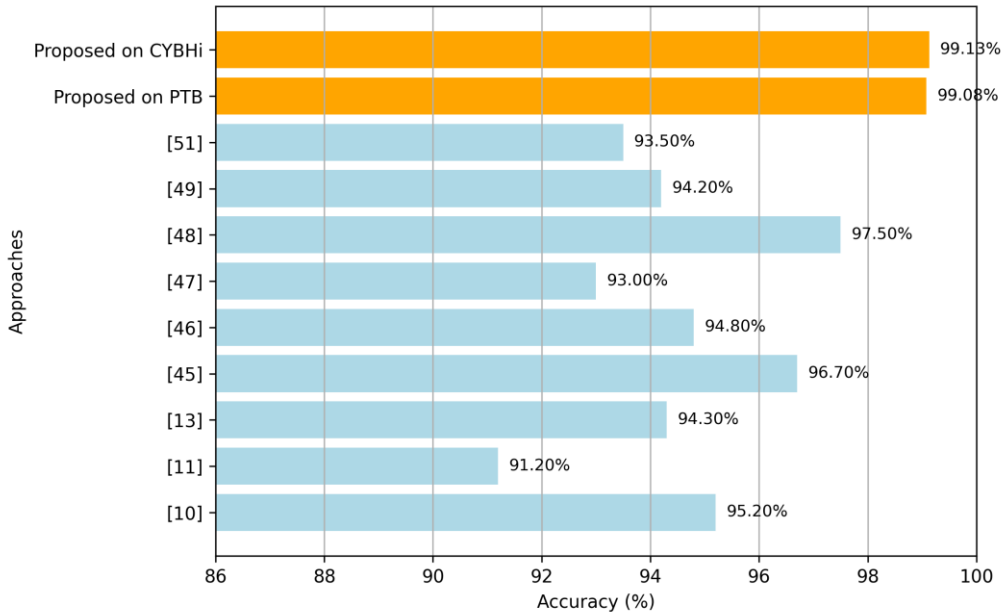


Fig. 6. Comparative Accuracy of proposed ECG Gender Recognition.

4.6. Comparative Analysis

A comparison of the proposed framework with major contemporary approaches to gender recognition is shown in Fig. 6. The experimental findings show that, compared with other approaches that have been implemented earlier, the proposed framework is superior and more accurate [10, 11, 13, 14, 46-49, 51]. While some methods, like [50], reported 99.2% accuracy on the MIT-BIH dataset, this accuracy was achieved only at a few epochs, which limits the applicability of their results. Alternatively, the proposed framework shows impressive performance across several epochs, confirming its adaptability and robustness.

4.7. Statistical Validation and Robustness Analysis

To examine the consistency and reliability of the proposed DeepFusion-CNN framework, additional statistical validation was carried out beyond single-run evaluation. The model was trained and tested across five independent runs with different random initializations while keeping the subject-wise data partitioning, preprocessing pipeline, and training parameters unchanged. This setup allows the assessment of whether the observed performance remains stable under varying initialization conditions rather than being influenced by chance.

Across these runs, the model demonstrated highly consistent behavior on both datasets. The aggregated results, expressed in terms of mean and standard deviation, are presented in Table 10. On the PTB dataset, the framework achieved an average accuracy of 99.08% with a deviation of ±0.32%, while precision, recall, and F1-score also remained tightly clustered around their mean values. A similar pattern was observed for the CYBHi dataset, where slightly higher variability is expected due to noisier signal acquisition conditions. Nevertheless, the variation remains within a narrow range, indicating stable learning behavior.

Table 10. Statistical performance of the proposed framework (Mean ± Standard Deviation over five runs).

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
PTB	99.08 ±0.32	99.31 ±0.28	98.88 ±0.35	98.98 ±0.30	0.993 ±0.002
CYBHi	99.13 ±0.41	96.25 ±0.52	96.34 ±0.47	96.66 ±0.44	0.993 ±0.003

To further interpret the reliability of these estimates, confidence bounds were examined using standard statistical formulation. The resulting intervals remain narrow for all key metrics, as summarized in Table 11, suggesting that the reported values are not only consistent but also well-constrained. This indicates that the model generalizes in a stable manner and is not overly sensitive to minor variations during training.

Table 11. Statistical validation using 95% confidence intervals.

Dataset	Accuracy (±CI)	Precision (±CI)	Recall (±CI)	F1-Score (±CI)
PTB	±0.28	±0.25	±0.31	±0.26
CYBHi	±0.36	±0.46	±0.41	±0.38

While consistency across runs provides one level of confidence, it is equally important to verify whether the observed improvement over individual models is meaningful. For this reason, a paired comparison was carried out between the proposed framework and the strongest standalone model, ResNet-152, which consistently showed the highest performance among individual architectures. Accuracy values from the same experimental runs were paired and analysed.

The differences observed across runs were consistently positive, indicating that the proposed model outperforms the baseline in every case. The statistical outcome of this comparison is summarized in Table 12. The mean improvement remains above 5% for both datasets, and the variation in differences is minimal. The resulting t-values are significantly high, and the associated p-values fall well below commonly accepted thresholds, indicating that the improvement is not attributable to random variation.

Table 12. Paired t-test results comparing proposed framework with ResNet-152.

Dataset	Mean Difference (%)	Std Dev (%)	t-value	p-value	Interpretation
PTB	5.31	0.19	62.4	< 0.001	Statistically significant
CYBHi	5.35	0.22	54.3	< 0.001	Statistically significant

Taken together, these results provide a coherent picture. The low variability across runs indicates stable model behavior, the narrow confidence bounds reinforce the reliability of reported metrics, and the statistical test confirms that the improvement over the baseline is substantial rather than incidental. In this context, the high classification accuracy observed in earlier sections can be interpreted as a result of consistent and reproducible learning rather than over-fitting or favourable data conditions.

4.8. Critical Discussion and Comparative Significance

The intermediate result analysis shown in Table 8 confirms the experimental findings of the proposed framework in Tables 6 and 7. These findings statistically validate the improvement in the performance of the proposed framework. This improvement is achieved due to the integration of context-aware model fusion, ensemble learning, and robust preprocessing, and not a deliberate choice. Initially, the quality of the input ECG signal was improved by increasing the SNR to 12.7 dB through preprocessing. This helped to better align the heartbeats, reducing within-class variability. The framework achieved a 5% improvement compared to 1D training by converting 1D ECG segments into 2D images, enabling it to better understand spatial morphology. An additional 3% to 4% performance increase resulted from data augmentation, which contributed to continuous learning by reducing class imbalance and increasing the diversity of input samples.

Although the proposed framework achieves high classification accuracy (≈99%) on both PTB and CYBHi datasets, careful measures have been taken to ensure that this performance is not a result of overfitting or data leakage. This performance significantly surpasses existing recurrent-based approaches such as LSTM and Bi-LSTM, which typically achieve accuracy in the range of 74%–79% for ECG-based gender classification tasks [48]. First, a strict subject-wise data partitioning strategy is employed, ensuring that no subject appears in both training and testing sets, thereby eliminating identity leakage. Second, data augmentation is applied exclusively to the training set after dataset splitting, preventing any overlap between augmented and test samples. Third, the model is evaluated on two independent datasets with different acquisition conditions, including the relatively noisy CYBHi dataset, where consistent performance is observed. Additionally, ablation analysis demonstrates that performance improvements are systematically contributed by preprocessing and context-aware fusion rather than model memorization. These factors collectively indicate that the high accuracy is driven by robust feature learning and effective model generalization rather than over-fitting.

While the proposed framework demonstrates strong performance across multiple evaluation metrics, it is important to note that the current study does not include statistical variability measures such as standard deviation or confidence intervals obtained from repeated experimental runs. The reported results are based on fixed training–testing splits, which provide a consistent basis for comparative evaluation across models. Although the use of subject-wise partitioning and evaluation on two distinct datasets (PTB and CYBHi) offers some level of robustness, the absence of repeated trials limits the ability to fully quantify result variability. Future work will incorporate k-fold cross-validation, multiple random initializations, and statistical significance testing to provide a more comprehensive assessment of

model stability and generalization.

Furthermore, recent studies have reported that deep learning models applied to ECG data may exhibit demographic biases, particularly across gender, age, and population groups, which can affect model generalization and fairness. This suggests that high classification performance does not necessarily guarantee uniform reliability across diverse populations, and therefore, careful interpretation of results is required when extending such models to real-world clinical settings [60]. In addition, prior research has shown that ECG-based biometric systems may exhibit performance variability under different physiological conditions, indicating that signal acquisition and subject-specific factors play a crucial role in model reliability and generalization [61].

The three sub-models used in the proposed framework, VGG-19, DenseNet-121, and ResNet-152, exhibit complementary strengths when used independently. DenseNet-121 and ResNet-152 attained a recall of 93% and 93.7% by utilizing the dense connections and attention-based residual blocks, respectively. VGG-19 performed quite well for localized patterns and achieved 0.91 recall, but plateaued early. The use of a context-aware fusion strategy proved to be justified, exhibiting a performance improvement of 2-3%. The fusion strategy lowers the impact of a negative response and strengthens all three deployed sub-models so as to adjust the weight based on the validation performance. At last, with improved performance across all modules, the proposed framework outperformed the three sub-models. The framework has demonstrated efficiency in achieving accuracies of 99.08% and 98.7% with 80:20 and 70:30 ratios, respectively.

The findings of the proposed framework for gender recognition based on ECG signal demonstrate both the practicality and the scientific robustness. The obtained results are accurate and reliable, but the incorporation of three deep convolutional neural networks in real-time, on the other hand, leads to an increase in computational complexity. All results represent a single experimental run; therefore, variability across runs is not captured. This emphasizes the need to conduct more research on knowledge extraction or lightweight fusion architectures for embedding.

5. Conclusion

ECG is a popular non-invasive procedure to record the electrical activity of the heart, preserving important details of the cardiovascular system. The ECG signals have shown capabilities for gender recognition with the advancements in biomedical signal processing and deep learning. The ECG-based gender recognition depends on small and consistent changes in cardiac morphology and electrophysiological patterns. This advancement has shown its potential to redefine the customized patient treatment plan according to gender. This paper presents a study using a state-of-the-art deep learning method for a significant contribution to ECG-based gender recognition. The proposed CNN framework can learn and extract complex intrinsic patterns from high-dimensional ECGs naturally. The framework's stability and adaptability have been tested on two training-testing ratios with five training epochs. Experimental results have confirmed its efficiency and reliability, opening a promising door for physiological profiling, biometric authentication, and personalized healthcare systems.

This study introduces a CNN-based framework for gender recognition using the ECG signal. The framework tackles challenges like feature extraction, learning efficiency, and generalization by incorporating key architectural features from pre-trained CNN models such as VGG, DenseNet, and ResNet. It surpasses VGG-19, DenseNet-121, and ResNet-152 by utilizing residual connections, dense feature propagation, and a combined architecture. The proposed framework achieves an impressive ROC-AUC of 0.98 with an 80:20 train-test split over 20 epochs, along with an accuracy of 99.08%, precision of 0.96, recall of 0.96, and an F1-score of 0.96. These results demonstrate the framework's effectiveness and consistency in using ECG to recognize gender.

Despite the success of the study, it has certain shortcomings. The study lacks the consideration of factors such as age, multiple medical conditions, and differences in ECG capturing methods, which may impact the performance of the framework in real-life situations. So, the future work should also emphasize these factors and assess the framework's performance on diversified, larger datasets. The integration of such factors with ECG signals may have significant potential to improve the performance of gender recognition.

All the Declarations and Statements

Author Contributions Statement

Sanjay Kumar Pandey– Conceptualization, Methodology, Software, Data Curation, and Writing – Original Draft: Developed the research concept, designed the methodology, implemented the proposed model, performed data preprocessing, conducted experiments, and prepared the initial manuscript.

Bechoo Lal – Supervision, Validation, Formal Analysis, and Writing – Review & Editing: Provided research guidance and supervision, validated the methodology and results, contributed to formal analysis, and reviewed and refined the manuscript for intellectual content.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement
The authors declare no conflicts of interest.

Funding Declaration
This research was not supported by any grant.

Data Availability Statement
The datasets used in this study are publicly available:
- PTB Diagnostic ECG Database: https://physionet.org/content/ptbdb/
- CYBHi Dataset: https://physionet.org/content/cybhi/

Ethical Declarations
This study uses publicly available datasets and does not involve direct human or animal participation. Therefore, ethical approval was not required.

Acknowledgments
The authors gratefully acknowledge the Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan, India, for providing the essential infrastructure, research facilities, and academic support that significantly contributed to the successful completion of this work.

Declaration of Generative AI in Scholarly Writing
All scientific content, analysis, and conclusions are the original work of the authors, who take full responsibility for the manuscript.

Abbreviations
The following abbreviations are used in this manuscript:

BAS – Biometric Authentication System
CNN – Convolutional Neural Network
ECG – Electrocardiogram
EMD – Empirical Mode Decomposition
KNN – K-Nearest Neighbor
KPCA – Kernel Principal Component Analysis
LSTM – Long Short-Term Memory
RF – Random Forest
SNR – Signal-to-Noise Ratio
SVM – Support Vector Machine

Appendix
None.

References

[1] R. Srivastva and Y. Singh, "ECG analysis for human recognition using non-fiducial methods," *IET Biometrics*, vol. 8, no. 5, pp. 295–305, 2019. doi:10.1049/iet-bmt.2018.5093
[2] R. Srivastva and Y. N. Singh, "Identifying individuals using Fourier and discriminant analysis of electrocardiogram," in *Mathematics and Computing. ICMC 2018. Communications in Computer and Information Science*, Springer, Singapore, 2018, pp. 286–295. doi:10.1007/978-981-13-0023-3_27
[3] Z. Attia et al., "Age and sex estimation using artificial intelligence from standard 12-lead ECGs," *Circulation: Arrhythmia and Electrophysiology*, vol. 12, no. 9, 2019. doi:10.1161/CIRCEP.119.007284
[4] J. Cabra, D. Mendez, and L. Trujillo, "Wide machine learning algorithms evaluation applied to ECG authentication and gender recognition," in *ICBEA '18: Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications*, 2018, pp. 58-64. doi:10.1145/3230820.3230830
[5] Y. Hsu, J. Wang, W. Chiang, and C. Hung, "Automatic ECG-based emotion recognition in music listening," *IEEE Transactions on. Affective Computing*, vol. 11, no. 1, pp. 85–99, 2020. doi:10.1109/TAFFC.2017.2781732
[6] O. Kittnar, "Sex related differences in electrocardiography," *Physiological Research*, vol. 72, Suppl. 2, pp. S127–S135, 2023. doi: 10.33549/physiolres.934952
[7] K. Y. Halim, D. T. Nugrahadi, M. R. Faisal, R. Herteno, and I. Budiman, "Gender classification based on electrocardiogram

signals using long short term memory and bidirectional long short term memory," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 606–618, 2023. doi: 10.26555/jiteki.v9i3.26354
[8] K. Siegersma et al., "Deep neural networks reveal novel sex-specific electrocardiographic features relevant for mortality risk," *European Heart Journal - Digital Health*, vol. 3, no. 2, pp. 245–254, 2022. doi:10.1093/ehjdh/ztac010
[9] A. Goshvarpour and A. Goshvarpour, "Gender and age classification using a new Poincaré section-based feature set of ECG," *Signal, Image and Video Processing*, vol. 13, no. 3, pp. 531–539, 2018. doi:10.1007/s11760-018-1379-5
[10] M. Khan et al., "Electrocardiogram based gender classification," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 2020, pp. 1-6*. doi:10.1109/ICECCE49384.2020.9179305
[11] J. Lyle, M. Nandi, and P. Aston, "Symmetric projection attractor reconstruction: sex differences in the ECG," *Frontiers in Cardiovascular Medicine*, vol. 8, pp. 709457, 2021. doi:10.3389/fcvm.2021.709457
[12] S. Hicks et al., "Explaining deep neural networks for knowledge discovery in electrocardiogram analysis," *Scientific Reports*, vol. 11, no. 1, pp. 10949, 2021. doi:10.1038/s41598-021-90285-5
[13] J. Lopez, C. Parra, L. Gomez, and L. Trujillo, "Sex Recognition through ECG Signals aiming toward Smartphone Authentication," *Applied Sciences*, vol. 12, no. 13, pp. 6573, 2022. doi: 10.3390/app12136573
[14] E. Alkhammash, M. Hadjouni, and A. Elshewey, "A hybrid ensemble stacking model for gender voice recognition approach," *Electronics*, vol. 11, no. 11, pp. 1750, 2022. doi:10.3390/electronics11111750
[15] H. Zhang et al., "Ensemble learning based gender recognition from physiological signals," In *Big Data – BigData 2018. BIGDATA 2018. Lecture Notes in Computer Science()*, Springer, Cham. 2018, pp. 352–359. doi:10.1007/978-3-319-94301-5_29
[16] J. Hu, "An approach to EEG-based gender recognition using entropy measurement methods," *Knowledge-Based Systems*, vol. 140, pp. 134–141, 2018. doi:10.1016/j.knosys.2017.10.032
[17] P. Wang and J. Hu, "A hybrid model for EEG-based gender recognition," *Cognitive Neurodynamics*, vol. 13, no. 6, pp. 541–554, 2019. doi:10.1007/s11571-019-09543-y
[18] H. Mieszczanska et al., "Gender-related differences in electrocardiographic parameters and their association with cardiac events in Patients After Myocardial Infarction," *The American Journal of Cardiology*, vol. 101, no. 1, pp. 20–24, 2008. doi:10.1016/j.amjcard.2007.07.077
[19] M. Nakagawa et al., "Gender differences in the dynamics of terminal T wave intervals," *Pacing and Clinical Electrophysiology*, vol. 27, no. 6, pp. 769–774, 2004. doi: 10.1111/j.1540-8159.2004.00526.x
[20] P. Rautaharju et al., "Sex differences in the evolution of the electrocardiographic QT interval with age," *Canadian Journal of Cardiology*, vol. 8, no. 7, pp. 690–695, 1992.
[21] R. Khane and A. Surdi, "Gender differences in the prevalence of electrocardiogram abnormalities in the elderly : A Population Survey in India," *Iranian Journal of Medical Sciences*, vol. 37, no. 2, pp. 92, 2012.
[22] M. Y. Ansari et. al., "Estimating age and gender from electrocardiogram signals: A comprehensive review of the past decade," *Artificial Intelligence in Medicine*, vol. 146, p. 102690, 2023. doi:10.1016/j.artmed.2023.102690
[23] N. Arif et al., "An approach to ECG-based gender recognition using random forest algorithm," *Journal of Electronics Electromedical Engineering and Medical Informatics*, vol. 6, no. 2, pp. 107–115, 2024. doi:10.35882/jeeemi.v6i2.363
[24] R. K. Tripathy, A. Acharya, and S. K. Choudhary, "Gender classification from ECG signal analysis using least square support vector machine," *American Journal of Signal Processing*, vol. 2, no. 5, pp. 145–149, 2012.
[25] H. Zacarias et al., "Gender Classification Using nonstandard ECG Signals - A Conceptual Framework of Implementation," In: *IoT Technologies for HealthCare. HealthyIoT 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications.* Springer, Cham, 2023, pp. 108–120. doi:10.1007/978-3-031-28663-6_9.
[26] M. Kim and S. Pan, "Deep learning based on 1-D ensemble networks using ECG for real-time user recognition," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5656–5663, 2019. doi:10.1109/TII.2019.2909730
[27] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Systems with Applications: X*, vol. 7, pp. 100033, 2020. doi:10.1016/j.eswax.2020.100033
[28] Y. Ansari, O. Mourad, K. Qaraqe, and E. Serpedin, "Deep learning for ECG arrhythmia detection and classification: an overview of progress for period 2017–2023," *Frontiers in Physiology*, vol. 14, pp. 1246746, 2023. doi: 10.3389/fphys.2023.1246746
[29] F. Behrad and M. Abadeh, "An overview of deep learning methods for multimodal medical data mining," *Expert Systems with Applications*, vol. 200, pp. 117006, 2022. doi:10.1016/j.eswa.2022.117006
[30] A. Linkon, M. Labib, T.Hasan, M. Hossain, and M. Jannat, "Deep learning in prostate cancer diagnosis and Gleason grading in histopathology images: An extensive study," *Informatics in Medicine Unlocked*, vol. 24, pp. 100582, 2021. doi: 10.1016/j.imu.2021.100582
[31] H. Zeng et al., "DCAE: A dual conditional autoencoder framework for the reconstruction from EEG into image," *Biomedical Signal Processing and Control*, vol. 81, pp. 104440, 2023. doi:10.1016/j.bspc.2022.104440
[32] W. Gu, S. Bai, and L. Kong, "A review on 2D instance segmentation based on deep neural networks," *Image and Vision Computing*, vol. 120, pp. 104401, 2022. doi:10.1016/j.imavis.2022.104401
[33] Y. Yang et al., "A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions," *Computers in Biology and Medicine*, vol. 139, pp. 104887, 2021. doi:10.1016/j.compbiomed.2021.104887
[34] I. Kaneko, J. Hayano, and E. Yuda, "How can gender be identified from heart rate data? Evaluation using ALLSTAR heart rate variability big data analysis," *BMC Research Notes*, vol. 16, no. 1, pp. 1–5, 2023. doi:10.1186/s13104-022-06270-2
[35] N. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *In: Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences. 1998, pp. 903–995*. doi:10.1098/rspa.1998.0193
[36] P. Chang and J. Wu, "A critical feature extraction by kernel PCA in stock trading model," *Soft Computing*, vol. 19, no. 5, pp. 1393–1408, 2015. doi:10.1007/s00500-014-1350-5

[37] M. Mansouri, M. Nounou, H. Nounou, and N. Karim, "Kernel PCA-based GLRT for nonlinear fault detection of chemical processes," *Journal of Loss Prevention in the Process Industries*, vol. 40, pp. 334–347, 2016. doi:10.1016/j.jlp.2016.01.011
[38] C. Cody, V. Ford, and A. Siraj, "Decision Tree Learning for Fraud Detection in Consumer Energy Consumption," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, 2015 pp. 1175–1179. doi:10.1109/ICMLA.2015.80
[39] C. Szegedy et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9.* doi:10.1109/CVPR.2015.7298594
[40] V. Mayya, R. Pai, and M. Pai, "Automatic facial expression recognition using DCNN," *Procedia Computer Science*, vol. 93, pp. 453–461, 2016. doi:10.1016/j.procs.2016.07.233
[41] C. Wang, K. Li, Z.Wu, and Q. Zhao, " A DCNN Based Fingerprint Liveness Detection Algorithm with Voting Strategy," in *Biometric Recognition. CCBR 2015. Lecture Notes in Computer Science()*, Springer, 2015, vol 9428, pp. 241-249. doi:10.1007/978-3-319-25417-3_29
[42] K. Greff, R. Srivastava, J. Koutnik, B. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017. doi:10.1109/TNNLS.2016.2582924
[43] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018. doi:10.1109/ACCESS.2017.2779939
[44] M. Nandi and P. Aston, " Extracting new information from old waveforms: Symmetric projection attractor reconstruction: Where maths meets medicine," *Experimental Physiology*, vol. 105, no. 9, pp. 1444–1451, 2020. doi:10.1113/EP087873
[45] E. Alkeem et al., "Robust deep identification using ECG and multimodal biometrics for industrial internet of things," *Ad Hoc Networks*, vol. 121, pp. 102581, 2021. doi:10.1016/j.adhoc.2021.102581
[46] J. Lopez, C. Parra, and G. Forero, "A fast deep learning ECG sex identifier based on wavelet RGB image classification," *Data*, vol. 8, no. 6, pp. 97, 2023. doi:10.3390/data8060097
[47] I. Jekova and G. Bortolan, "Personal Verification/Identification via Analysis of the Peripheral ECG Leads: Influence of the Personal Health Status on the Accuracy," *BioMed Research International*, vol. 2015, pp. 1–13, 2015. doi:10.1155/2015/135676
[48] D. Jyotishi and S. Dandapat, "An LSTM-based model for person identification using ECG signal," *IEEE Sensors Letters*, vol. 4, no. 8, pp. 1–4, 2020. doi:10.1109/LSENS.2020.3012653
[49] Y. Li, Y. Pang, K. Wang, and X. Li, "Toward improving ECG biometric identification using cascaded convolutional neural networks," *Neurocomputing*, vol. 391, pp. 83–95, 2020. doi:10.1016/j.neucom.2020.01.019
[50] J. Kim, S. Kim, and S. Pan, "Personal recognition using convolutional neural network with ECG coupling image," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 5, pp. 1923–1932, 2020. doi:10.1007/s12652-019-01401-3
[51] N. Demir, M. Kuncan, Y. Kaya, and F. Kuncan, "Multi-Layer co-occurrence matrices for person identification from ECG signals," *Traitement du Signal*, vol. 39, no. 2, pp. 431–440, 2022.
[52] I. Selvam, M. Madhavan, and S. K. Kumarasamy, "Detection and classification of electrocardiography using hybrid deep learning models," *Hellenic Journal of Cardiology*, vol. 81, pp. 75–84, 2025. doi: 10.1016/j.hjc.2024.08.011
[53] J. Pan and W. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985. doi:10.1109/TBME.1985.325532
[54] X. Wu, Z. Wang, B. Xu, and X. Ma, "Optimized Pan-Tompkins Based Heartbeat Detection Algorithms," in *2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 2020, pp. 892-897*. doi:10.1109/CCDC49329.2020.9164736
[55] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248-255. doi:10.1109/CVPR.2009.5206848
[56] A. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000. doi:10.1161/01.CIR.101.23.e215
[57] H. da Silva, A. Lourenço, A. Fred, N. Raposo, and M. Aires-de-Sousa, " Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 503–514, 2014. doi:10.1016/j.cmpb.2013.11.017
[58] S. Kant, "From data to decision-making: utilizing decision tree for air quality monitoring in smart urban areas," *International Journal of Information Technology*, vol. 17, pp. 665-672, 2025. doi:10.1007/s41870-024-02208-y
[59] S. Jain and A. Saha, "Improving and comparing performance of machine learning classifiers optimized by swarm intelligent algorithms for code smell detection," *Science of Computer Programming*, vol. 237, pp. 103140, 2024. doi:10.1016/j.scico.2024.103140
[60] D. Kaur et al., "Race, Sex, and Age Disparities in the Performance of ECG Deep Learning Models Predicting Heart Failure," Circulation: Heart Failure, vol.17, no.1, pp. e010879, 2024. doi:10.1161/CIRCHEARTFAILURE.123.010879
[61] S. Azam and K. Sidek, "Gender-based analysis of ECG biometric identification under different physiological conditions," *International Journal of Biometrics*, vol. 17, no. 5, pp. 433-448, 2025. doi: 10.1504/IJBM.2025.148283

Authors' Profiles



Sanjay Kumar Pandey is a researcher in the field of Computer Science, specializing in machine learning, data analytics, and intelligent systems. He is indexed in Scopus (Scopus ID: 58195496000), reflecting his contributions to peer-reviewed scientific research. His research interests primarily focus on supervised learning techniques, particularly Support Vector Machines (SVM), and their applications in disease classification and predictive modeling. He has conducted systematic evaluations of various SVM variants on benchmark datasets, with emphasis on performance metrics such as accuracy, precision, recall, and F1-score. Mr. Pandey's research aims to bridge the gap between theoretical machine learning models and practical real-world applications, especially in healthcare analytics. He is actively engaged in developing efficient algorithms and data-driven methodologies to address complex classification challenges.



Bechoo Lal is an academic researcher in the field of Computer Science with expertise in artificial intelligence, machine learning, and data analytics. He is indexed in Scopus (Scopus ID: 58043508200), reflecting his contributions to peer-reviewed scientific research. His research focuses on intelligent systems, predictive modeling, and data-driven approaches for solving complex computational problems. He has worked extensively on the application of machine learning techniques in areas such as classification, pattern recognition, and decision support systems. Dr. Lal has published several research papers in reputed journals and conferences, contributing to advancements in modern computing technologies. His work emphasizes the development of efficient algorithms and the practical implementation of artificial intelligence solutions in real-world scenarios. In addition to his research contributions, he is actively engaged in teaching, mentoring, and collaborative research, fostering the integration of theoretical knowledge with practical applications.

How to cite this paper:

Sanjay Kumar Pandey, Bechoo Lal, "DeepFusion-CNN: A Novel Context-aware Network for ECG-based Gender Classification", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.18, No.5, pp. 55-75, 2026. DOI:10.5815/ijigsp.2026.05.04