El Cezeri
Journal of Science and Engineering
ISSN: 2148-3736

Research Article

Heart failure detection using deep learning and Gradient Boosting classifier

Ahmad Mahmoud Ahmad [1a]
[1] Department of Computer Systems and Networks, Faculty of Informatics Engineering, Tishreen University, Latakia, Syria.
ahmad.m.ahmad@tishreen.edu.sy

DOI : 10.31202/ecjse.1476222
Received: 03.05.2024 Accepted: 14.01.2025
How to cite this article:
Ahmad Mahmoud Ahmad, " Heart failure detection using deep learning and Gradient Boosting classifier ", El-Cezeri Journal of Science and Engineering, Vol: 12, Iss: 1, (2025), pp.(1-8).
ORCID: [a]0009-0006-6571-636X.

Abstract : Heart failure (HF) is marked by a diminished capacity of the heart to effectively pump blood. Traditionally, the electrocardiogram (ECG) has served as a non-invasive diagnostic tool, gauging the heart's electrical activity and rhythm. Recent advancements have leveraged machine learning (ML) and deep learning (DL) techniques to automate the identification and classification of HF types from ECG data. This study introduces a novel deep learning architecture, blending the efficacy of a convolutional neural network (CNN) for feature extraction with an eXtreme Gradient Boosting (XGBoost) layer for final classification. The first CNN model operates on ECG segments in the time domain, while the second CNN processes the Continuous Wavelet Transform (CWT) of the same segments. This composite model offers superior automatic HF detection, particularly with 2-second ECG fragments, by capturing intricate features from both time and frequency domains. Training and testing utilize datasets from the MIT-BIH, BIDMC, and PTB Diagnostic ECG databases. Through 10-fold cross-validation, the proposed approach attains remarkable accuracy, sensitivity, and F1-score, all surpassing 99.9%. This modality represents a significant stride in DL applications for ECG diagnosis, holding promise for enhanced clinical utility.

Keywords : Convolutional neural network, deep learning, electrocardiogram, eXtreme Gradient Boosting, heart failure.

1 Introduction

Heart failure (HF) is characterized by a diminished capacity of the heart to pump blood [1]. This condition commonly arises from reduced left ventricular function and structural or functional defects within the myocardium, impeding either ventricular filling or blood ejection. Factors such as increased hemodynamic overload and ischemia-related dysfunction also contribute significantly to HF pathogenesis. Furthermore, HF stands as a principal cause of morbidity and mortality [2]. For decades, the electrocardiogram (ECG) has served as a pivotal non-invasive diagnostic tool for assessing the heart's electrical and rhythmic activity [3]. Its sensitivity to detecting HF renders it indispensable for predictive monitoring. However, cardiologists face a significant challenge in swiftly and accurately interpreting ECG signals, especially during prolonged monitoring sessions [4]. To surmount this challenge, various clinical decision support systems (CDSS) have emerged over the past decade, ranging from rudimentary rules-based systems to sophisticated algorithms rooted in machine learning (ML) and deep learning (DL) [5]–[7]. ML algorithms are primarily categorized into supervised, unsupervised, and reinforcement learning, depending on how they are initialized and trained [8]. Unsupervised learning leverages unlabeled datasets, while supervised learning relies on labeled data, where training samples and datasets are pre-classified and categorized. The Convolutional Neural Network (CNN), a type of deep learning neural network, is employed to classify data. Within a CNN, deep features are extracted from input images using convolution and pooling, computational load is reduced through downsampling, and final predictions are generated by fully connected layers [9]. Recent endeavors have focused on utilizing ML and DL methodologies to automatically identify and classify different types of HF from ECG data [10]–[12].
Asyali [13] explored the discriminatory power of nine commonly used long-term HRV measures, aiming to develop Bayesian classifiers. Sensitivity and specificity rates of 81.8% and 98.1% were achieved, respectively, depending on all normal-to-normal beat intervals' standard deviations. Jin et al. [14] proposed a wearable, cell phone-based platform capable of continuous real-time monitoring and recording of ECG data to immediately recognize abnormal cardiovascular disease (CVD) conditions. Their approach integrates an adaptive artificial neural network (ANN)-based hybrid strategy, combining patient-specific training

methods with established medical database training techniques. The results demonstrated 99% accuracy in detecting normal heartbeats and 92% accuracy in identifying premature ventricular contractions (PVCs). Chen et al. [15] utilized RR interval segments and sparse auto-encoders (SAE) to detect heart failure (HF), achieving an accuracy of 72.44%, a sensitivity of 50.93%, and a specificity of 80.93%. Masetic et al. [16] presented a method involving auto-regressive parish feature extraction and subsequent classification, resulting in 100% accuracy, sensitivity, and specificity in detecting HF. Liu and Kim [17] proposed employing Long Short-Term Memory (LSTM) and Symbolic Aggregate approximation (SAX) for categorizing heart disease using ECG signals, achieving 98.4% accuracy. Wang et al. [18] integrated a CNN module and LSTM network for HF detection, obtaining 86.42% accuracy, 74.91% sensitivity, and 91.21% specificity. Acharya et al. [18] categorized ECG signals using an 11-layer CNN, achieving 99.99% accuracy, 98.87% sensitivity, and 99.01% specificity. Cheng et al. [19] combined a 24-layer DCNN with Bidirectional LSTM for hierarchical and time-sensitive feature mining in ECG data, achieving an F1 score of 89% and an accuracy of 89.3% with 10-fold cross-validation. Padmavathi et al. [20] introduced an 11-layer CNN for HF detection, with a specificity rate of 79.30%, sensitivity of 81%, and accuracy of 80.10%. Lih et al. [21] developed a 16-layer CNN-LSTM design, achieving 97.89% specificity, 99.3% sensitivity, and 98.5% accuracy. Zhang et al. [22] enhanced the DenseNet model for HF detection using 2-second ECG fragments, achieving 89.38% sensitivity, 99.50% specificity, and 94.97% accuracy. Kusuma and Jothi [23] identified congestive heart failure (CHF) using an automated diagnosis system based on LSTM architecture and Deep CNN, achieving 99.52% accuracy. Botros et al. [24] proposed a CNN with a Support Vector Machine (SVM) layer and an integrated classification layer, achieving over 99% accuracy, sensitivity, and specificity with blindfold cross-validation. Rawi et al. [25] introduced a CNN with eXtreme Gradient Boosting (XGBoost) feature extraction, achieving 99.38% accuracy and 98.36% F1-score. Wang et al. [26] suggested a continuous wavelet transform (CWT) and CNN-based automatic ECG classification method, achieving 67.47% sensitivity, 68.76% F1-score, and 98.74% accuracy overall. Mogili and Narsimha [27] proposed a hybrid model combining a CNN for automatic ECG feature extraction with XGBoost for arrhythmia classification. Tested on the MIT-BIH Arrhythmia database, the model achieved an accuracy of 99.84% for 11 arrhythmia types and 99.69% for 5 AAMI standard classes, demonstrating its robustness with high sensitivity and specificity. Premalatha and Bai [28] developed a deep CNN-based model to classify cardiac dysrhythmia using oversampled datasets to address class imbalance. Coupled with XGBoost for structured prediction, their approach was validated on a real-time IoT dataset of elderly heart patients, achieving a recall of 100%, an F1-score of 94.8%, a precision of 98%, and an accuracy of 98%, outperforming traditional classifiers like decision trees, random forests, and SVM. Khan et al. [29] employed the MIT-BIH ARR dataset and a 1-D ResNet model, achieving an impressive accuracy of 98.63%. However, they noted that the performance of the F class still requires improvement. Al-Jibreen et al. [30] utilized the MIT-BIH Arrhythmia dataset for signal segmentation and classification using cosine wavelet transforms and a lightweight CNN with depth-wise separable convolution. Their approach achieved a classification accuracy of 99.28% for normal beats and 93.81% for abnormal beats. Majhi and Kashyap [31] proposed tree-based classifiers, Random Forest (RF) and XGBoost, for heart disease detection using three major ECG datasets: Physionet Challenge 2016, PASCAL Challenge, and MIT-BIH. Pre-processing techniques like filtering and denoising were applied, followed by feature extraction using DWT, IDWT, and EWT. SHAP analysis identified critical features impacting model predictions. Their results showed EWT with XGB achieving superior AUCs of 97.44% and 98.25% on the Physionet and MIT-BIH datasets, respectively, outperforming other feature-model combinations.

The synthesis of existing literature underscores the pivotal role of robust models and effective feature extraction in creating comprehensive feature extraction and classification systems. Recent studies have highlighted the efficacy of deep neural networks in interpreting ECG signals within both the time and time-frequency domains. However, challenges persist, particularly concerning the low sensitivity observed when employing CWT as input for CNNs.

In response to these challenges, this paper proposes a novel deep learning model that capitalizes on the strengths of CNNs for feature extraction and leverages the XGBoost classifier for end-of-model classification. The proposed model integrates two CNNs: the first processes ECG segments in the time domain, while the second operates on the CWT of the same segments. By combining features extracted from both temporal and spectral ECG data, the proposed model achieves enhanced accuracy in automatic HF detection using 2-second ECG fragments. The evaluation of the proposed model utilizes three prominent ECG databases—MIT-BIH, BIDMC, and PTB Diagnostic—for both training and testing purposes. The paper meticulously outlines the methodologies and materials employed, detailing the database descriptions, preprocessing procedures, and the proposed approach's implementation. Additionally, the results and discussion section thoroughly analyze the obtained outcomes using various performance metrics, providing valuable insights into the model's efficacy and potential areas for improvement.

2 Materials and Methods

2.1 Databases Description

Three ECG databases sourced from literature were utilized in this study:

1) BIDMC Database [32]: This database comprises ECG signals from 15 patients diagnosed with Congestive Heart Failure (CHF). The patients include 11 men and 4 women aged between 54 and 63 years. The signals were sampled at a

frequency of $fs = 250Hz$.

2) MIT-BIH Database [33]: This dataset consists of ECG signals from 18 healthy individuals exhibiting Normal Sinus Rhythm (NSR). The cohort includes 13 women aged between 20 and 50 years and 5 men aged between 26 and 45 years. The signals were recorded using ambulatory Holter and ECG recorders, with a sampling frequency of $fs = 360Hz$. Each segment of the signals spans approximately 20 hours and has a resolution of 250 points.
3) Physionet PTB Diagnostic ECG Database [34]: This database comprises 549 recordings obtained from 290 individuals aged between 17 and 87 years. Each recording contains 15 signals measured simultaneously. The sampling frequency for each signal is $fs = 1000Hz$, and they are represented with a 16-bit resolution ranging around ± 16.384 mV.

2.2 Pre-Processing

To maintain uniformity in sampling frequency across all ECG indicators, the recordings from the BIDMC database undergo initial down-sampling to 250 Hz. Subsequently, an ordinary filter with a 20-millisecond window is applied to smooth the signals. These ECG signals are then partitioned into small labeled segments, each sized 2 seconds, for subsequent processing with the CNN model. In total, the dataset comprises 500,000 segments, with half belonging to the HF group and the other half to the healthy (good) group. The overall properties are summarized in Table 1

2.3 Proposed Approach

The suggested approach is depicted in Fig. 1. The methodology begins with loading the combined dataset and subsequently implementing the requisite pre-processing steps. The deep learning model comprises two CNN models—one for processing the raw ECG signal and the other for its CWT profile. The proposed DL model synergizes the effectiveness of CNNs as feature extraction tools with the XGBoost layer for classification at the model's conclusion. Each component of the proposed model is elaborated in the following subsections. The pseudo-code of proposed approach is provided in Algorithm 1.

Algorithm 1. Algorithmic workflow for the proposed method

```
1. Load the combined dataset.
2. Implement the following preprocessing steps:
  a. Preprocess the data as required.
  b. Segment the ECG signals into 2-second segments comprising 250 points each.
3. Define the structure of the CNN models for feature extraction:
  a. Construct two separate CNN models: one for processing ECG segments in the time domain and the other
for segments represented in CWT form.
  b. for each CNN model:
    i. Initialize the CNN model.
    ii. for each convolutional layer in the model:
      - Add convolutional layer with specified parameters.
    iii. Add max-pooling layer with specified parameters.
    iv. Add another convolutional layer with specified parameters.
    v. Repeat convolutional and pooling layers as needed to achieve desired depth or feature richness.
4. Extract deep features:
  a. for each CNN model:
    i. for each pooled feature map:
      - Pass through additional convolutional layers followed by pooling operations.
      - Repeat this process until desired depth or feature richness is achieved.
5. Combine the outputs of the final pooling layers from different CNN models:
  a. Initialize an empty array to store combined deep features.
  b. for each CNN model:
    i. Extract deep features obtained from the final pooling layer.
    ii. Concatenate or average the feature vectors obtained from each model.
    iii. Append the combined deep features to the array.
6. Use the XGBoost layer for classification based on the deep features obtained from the combined CNN
models.
7. Train and evaluate the DL model.
8. Optionally, fine-tune the model based on performance evaluation.
```

2.4 Continuous Wavelet Transform (CWT)

In order to improve feature extraction for efficient use of CNN model, the ECG signal can be transformed to the time-frequency domain because it is made up of various frequency components. The most widely used time-frequency research tool is CWT, which decomposes a signal over the course of time by using different wavelet functions. CWT develops and inherits the short-

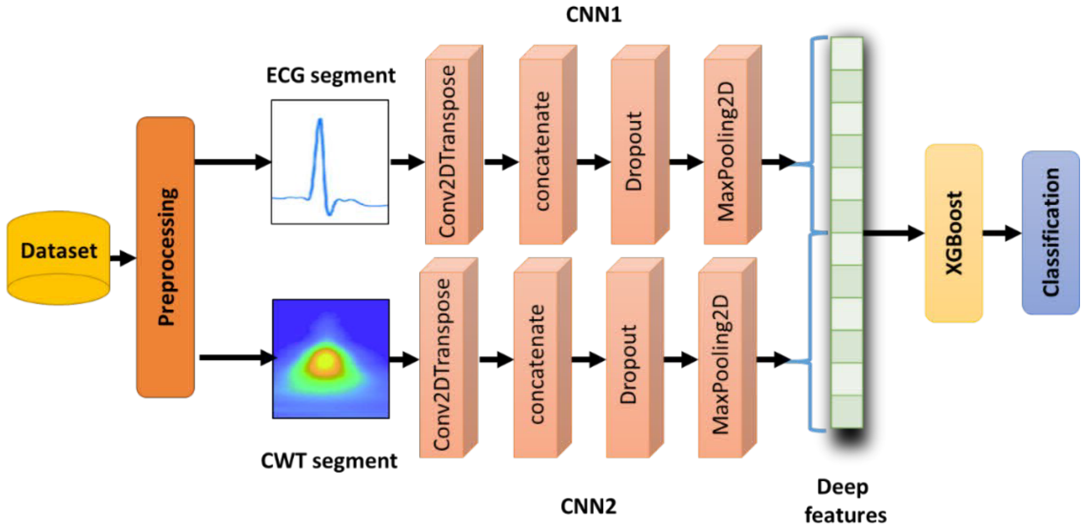

Figure 1: Proposed DL approach for ECG diagnosis

Table 1: Attribute definitions of datasets before and after pre-processing.

Dataset	Original Attributes	Pre-Processing Steps	Attributes After Pre-Processing
BIDMC [32]	- 15 patients with CHF - 11 men, 4 women - Sampling frequency: 250 Hz	- Signals partitioned into 2-second segments - Smoothing with 20 ms filter	- Down-sampled to 250 Hz - 2-second segments - 250 points/segment
MIT-BIH [33]	- 18 healthy individuals with NSR - 13 women, 5 men - Sampling frequency: 360 Hz	- Down-sampling to 250 Hz - Signals segmented into 2-second windows	- Down-sampled to 250 Hz - 2-second segments - 250 points/segment
Physionet PTB [34]	- 290 individuals, 549 recordings - Sampling frequency: 1000 Hz - 15 signals/recording - Resolution: $\pm$16.384 mV	- Down-sampling to 250 Hz - Signals segmented into 2-second windows	- Down-sampled to 250 Hz - 2-second segments - 250 points/segment
Combined Dataset	-	-	**Total: 500,000 segments (balanced)**

time Fourier transform's (STFT) localization concept. The CWT of x(t) signal is computed using Eq. 1 [26]:

$$C_a(b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t).\varphi\left(\frac{t-b}{a}\right) dt \tag{1}$$

Where the wavelet function is ϕ(t), the translation parameter is b, and the scale parameter is a. To convert the scale into frequency, Eq. 2 is implemented where Fc is the mother wavelet's center frequency and fs is the sampling frequency of signal x(t). The wave coefficients of the signal at various scales are obtained by using various CWT scale factors. A 2D scalogram of an ECG signal in the time-frequency site can be created using these wave coefficients.

$$F = \frac{F_c * f_s}{a} \tag{2}$$

2.5 Deep Feature Extraction Using CNN

In this study, two CNN models [32-34] are proposed for deep feature extraction, each consisting of three convolutional layers and one pooling layer. Given the focus on ECG segments within both the time and time-frequency domains, two separate CNN models are employed—one for processing ECG segments in the time domain and the other for segments represented in CWT form. Each CNN model receives a 2-second ECG segment comprising 250 points. Both CNN models share the same structural configuration, detailed as follows:

1) The first convolutional layer (CL) utilizes five 1x14 filters with a stride of 1.
2) The subsequent CL employs three 9x9 filters, also with a stride of 1.
3) The convolution stage produces three feature maps by combining various filters with the 250-point ECG signal.

4) Following the convolution stage, the max-pooling layer reduces the dimensions of the feature maps. It employs a pool size of two and a stride of four, enhancing the model's resilience to changes in feature position.
5) Subsequently, another CL is applied with ten filters of size 1x9 and a stride of 1.
6) After multiple convolutional and pooling layers, deep features are extracted by passing the pooled feature maps through additional convolutional layers followed by pooling operations. This process continues until the desired depth or feature richness is achieved.
7) Finally, the outputs of the final pooling layers from different CNN models are combined into a row of deep features. This combination can be achieved by concatenating or averaging the feature vectors obtained from each model.

2.6 XGBoost Classifier

XGBoost is a potent regression-and-classification technique [35]. Based on the gradient improving framework, XGBoost continuously enhances learner performance and efficiency by adding new decision trees to fit a value with leftover multiple iterations. In contrast to Friedman's curve boosting [36], XGBoost approximates the loss function using a Taylor expansion. The model also has better tradeoff bias and variance and typically uses fewer decision trees to achieve higher accuracy. A second-order Taylor expansion is carried out on the square loss function in XGboost, a more potent version of the Gradient Boosting Decision Tree (GBDT) algorithm, to improve accuracy. The following is the main definition of XGBoost [37]:

$$XG^{(t)} \cong \sum_{i=1}^{n}[g_i f_i(x_i) + \frac{1}{2} h_i f_t^{i}(x_i)] + \Omega(f_t) \tag{3}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

First and second-order gradient statistics for the loss function are shown here as gi and hi. The sample numbers are represented by n. The regression tree functions at the t-th iteration are represented by ft (xi). The number of leaves on a tree is represented by t. L2 average of leaf scores is represented by w2j. Based on the model's complexity, the regularization term ω(ft) effectively avoids overfitting. To increase the algorithm's statement and learning speed, XGboost uses shrinking and column subsampling techniques.

2.7 Performance Metrics

Since the proposed approach is dedicated to classifying ECH into healthy or HF cases, the performance of the method is measured using the formula of accuracy (Eq. 5), sensitivity (Eq. 6) and F1-score (Eq. 7). Where True Positive (TP) refers to the number of correctly classified data as HF indicating the actual HF case; False Positive (FP): The number of incorrectly categorized data as HF that is not indicative of the correct case; True Negative (TN): The number of data classified as healthy case and indicating that actual healthy case; False Negative (FN): The number of data classified as healthy not where the actual one is HF.

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

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$F1-score = \frac{2 \cdot \Pr ecision \cdot \text{Sensitivity}}{\Pr ecision + \text{Sensitivity}} \tag{7}$$

Where:

$$\Pr ecision = \frac{TP}{TP + FP} \tag{8}$$

3 Results and Discussion

The rigorous evaluation of the proposed model necessitated a meticulous approach to dataset partitioning and model assessment. Firstly, the entire dataset was stratified into three distinct subsets: the training set, employed for training the CNN model; the validation set, utilized for fine-tuning hyperparameters; and the testing set, crucial for evaluating the ultimate performance of the model. In order to ensure the robustness and reliability of the implemented approach, 75% of the 2-second healthy segments, alongside an equivalent proportion of 75% of HF segments, were randomly selected for inclusion in the training and validation

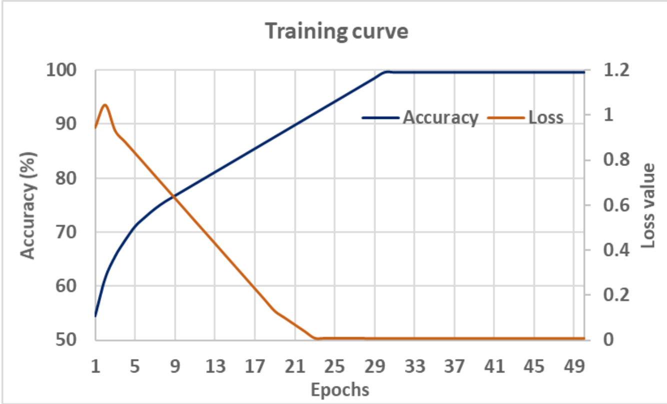

Figure 2: Training performance of CNN model.

sets. Subsequently, the validation and testing sets were constructed from the remaining data through a randomized selection process. Furthermore, to mitigate potential biases and ensure an equitable distribution of data across training and validation folds, the dataset was stratified into equal-sized folds. This stratification facilitated the proportional representation of each class within every fold, as necessitated by the principles of stratified cross-validation. In each iteration of the training process, one-fold was designated for validation while the remaining folds were utilized for model training. This iterative approach enabled comprehensive model evaluation across different subsets of the data, ensuring robustness and generalizability. Specifically, a 10-fold stratified cross-validation methodology was adopted in this study to systematically assess the performance of the proposed model. The training process of the CNN model, depicted in Figure 2, showcases the progression of model performance over epochs. Notably, the optimal performance metrics were attained after 30 epochs of training, indicating the convergence of the model towards an optimal solution. This meticulous methodology not only enhances the rigor of our experimental design but also underscores the reliability and validity of the reported findings.
As delineated in Table 2, the CNN-XGBoost hybrid model demonstrated superior performance across multiple evaluation metrics. Specifically, the hybrid model achieved an average accuracy of 99.95%, an average sensitivity of 99.96%, and an F1-score of 99.94%. Conversely, when utilizing the CNN as an independent classifier, the average accuracy, average sensitivity, and F1-score were notably lower, standing at 97.43%, 97.12%, and 92.22%, respectively.

Table 2: Ten-fold performance metrics (average $\pm$ standard deviation) of the CNN model.

Classifier	Accuracy (%)	Sensitivity (%)	F1-score (%)
Direct CNN classifier	97.43 ± 1.53	97.12 ± 1.65	92.22 ± 1.61
CNN-XGBoost (proposed method)	99.95 ± 0.32	99.96 ± 0.33	99.94 ± 0.34

The reduced standard deviation observed in the CNN-XGBoost model compared to the direct CNN model across all assessment metrics, as depicted in Table 2, is indicative of greater consistency in performance. Specifically, the lower standard deviations of 1.53 for accuracy and 1.65 for sensitivity in the direct CNN models contrast with the notably diminished standard deviations of 0.32 and 0.33, respectively, in the CNN-XGBoost model. This disparity in standard deviations underscores the CNN-XGBoost model's heightened stability and robustness in classification tasks. The diminished variability suggests that the performance metrics of the CNN-XGBoost model exhibit closer proximity to the average performance across multiple evaluations, thereby implying a more reliable and consistent classification outcome. In comparing the performance of the proposed CNN-based classification techniques against existing methods, a comprehensive analysis reveals their efficacy in medical diagnostics in Table 3. Zhang et al. [22] utilized a DenseNet model trained on a combination of BIDMC, MIT-BIH, and PTB datasets, achieving an accuracy of 94.97% and a sensitivity of 89.38%. Despite using a deep network architecture, their model's performance is limited, likely due to insufficient feature extraction capabilities or imbalanced training data. Wang et al. [26] incorporated CWT with CNN for feature extraction and achieved an accuracy of 98.74%, though sensitivity remained significantly low at 67.47%. The proposed approach overcomes these limitations by combining raw ECG signals and their transformed CWT profiles, ensuring comprehensive feature extraction and improved model generalization. Botros et al. [24] proposed an SVM layer integrated with a CNN model, achieving over 99% accuracy and sensitivity on the MIT-BIH and BIDMC datasets. While their results are competitive, the lack of reported F1-scores limits a full evaluation of the model's balance between precision and recall. Khan et al. [29] applied a 1-D ResNet model on the MIT-BIH dataset, achieving 98.63% accuracy and 92.41% sensitivity. However, the performance plateaued without additional enhancements like ensemble strategies or advanced feature extraction techniques. The work of Rawi et al. [25] is particularly noteworthy, as it combines CNN and XGBoost for classification using the MIT-BIH and PTB datasets. Their model achieved an accuracy of 99.38%, a sensitivity of 98.37%, and an F1 score of 99.11%, demonstrating the effectiveness of combining CNN-based feature extraction with XGBoost for robust classification. However, the novelty of the proposed method lies in its dual CNN architecture, where features are extracted not only from raw ECG signals but also from their corresponding CWT profiles. This dual representation enhances the richness of the extracted features, enabling the XGBoost classifier to achieve superior discriminative performance. By leveraging this enhanced feature set, the proposed approach achieves a remarkable accuracy of 99.95%, a sensitivity of 99.96%, and an F1 score of 99.96%, outperforming Rawi et al.'s results across all evaluation metrics. In comparison to Mogili and Narsimha [27], who also employed a CNN-XGBoost hybrid model on the MIT-BIH dataset with an accuracy of 99.84% and sensitivity of 92.61%, the proposed method demonstrates clear improvements. The significantly higher sensitivity and F1-score of the proposed approach are attributed to the multi-representation strategy and careful preprocessing, which unify datasets from BIDMC, MIT-BIH, and PTB. This integration creates a more diverse and balanced training set, further enhancing model generalization. Premalatha and Bai [28] proposed a CNN-XGBoost model with oversampling techniques to address class imbalance issues in a real-time IoT elderly patient dataset. Their model achieved 98% accuracy and a recall (sensitivity) of 100%. However, oversampling can introduce biases and overfitting risks, as indicated by their relatively lower F1-score of 94.8%. In contrast, the proposed method maintains a balanced performance without relying on oversampling techniques, achieving higher precision, recall, and F1-score simultaneously. Al-Jibreen et al. [30] presented a lightweight CNN model with separable convolution on the MIT-BIH Arrhythmia dataset, achieving modest results with an accuracy of 93.64% and a notably low F1-score of 53%, highlighting its limitations in handling class imbalances and complex features.

Table 3: Comparison of related work with the proposed approach.

Ref.	Method	Dataset	Accuracy	Sensitivity	F1-score
[22]	DenseNet model	BIDMC+ MIT-BIH+ PTB	94.97%	89.38%	-
[26]	CWT and CNN model	MIT-BIH	98.74%	67.47%	68.76%
[24]	SVM layer with CNN model	MIT-BIH+ BIDMC	$> 99\%$	$> 99\%$	-
[25]	CNN with XGBoost	MIT-BIH+ PTB	99.38%	98.37%	99.11%
[27]	CNN with XGBoost	MIT-BIH	99.84%	92.61%	95.99%
[28]	CNN-XGBoost model with oversampling	IoT Elderly Dataset	98%	100%	94.8%
[29]	1-D ResNet model	MIT-BIH	98.63%	92.41%	92.63%
[30]	Lightweight CNN with separable convolution	MIT-BIH Arrhythmia	93.64%	93.8%	53%
[31]	XGB Explainer with EWT features	MIT-BIH	98.14%	98.14%	98.10%
Proposed method	CNN-XGBoost model	BIDMC+ MIT-BIH+ PTB	99.95%	99.96%	99.96%

However, this study is limited in certain respects and it is important to acknowledge them. To begin with, our investigation scope has been confined by the lack of datasets in literature which limits the size and heterogeneity of the dataset that was employed for classification purposes. Secondly, we have used only two categories of ECG signals: normal and HF (abnormal) samples when classifying them. This dichotomous categorization framework may not be able to capture the subtleties involved in finer classifications of ECG diagnoses as a result undermining the generalizability of our proposed approach.

In future, further research efforts need to broaden these limitations through examining data consisting more diverse kinds regarding ECG abnormalities and subclasses. Moreover, such inclusion of different datasets from different sources will enable us to better model and thus understand what really happens in the ECG signals on a patient or population level. Also, broadening this classification system so as to accommodate fine specifications within various classes among other ECG subcategories would bring about an improved insight towards heart health and disease process.

4 Conclusion

The paper introduces a novel deep learning model that combines the efficacy of CNNs for feature extraction with XGBoost for classification, aiming to enhance automatic detection of HF using 2-second ECG fragments. The model consists of two CNNs: one processing ECG segments in the time domain and the other processing the CWT of the same ECG segment. By leveraging CNNs to extract deep features from both time and frequency domains, the proposed model achieves more accurate HF detection. To evaluate the model's performance, datasets from the MIT-BIH, BIDMC, and PTB Diagnostic ECG databases are utilized for training and testing. Results demonstrate that the proposed CNN-XGBoost model significantly outperforms using CNN alone as an independent classifier. Specifically, the CNN-XGBoost model achieves an impressive average accuracy of 99.95%, average sensitivity of 99.96%, and F1-score of 99.94%. In contrast, the direct utilization of CNN yields lower performance metrics, with an average accuracy of 97.43%, average sensitivity of 97.12%, and F1-score of 92.22%. Overall, the proposed model represents a promising advancement in the field of deep learning for ECG diagnosis. By combining CNNs with XGBoost, it offers a robust and accurate approach to HF detection, demonstrating its potential to enhance clinical decision-making and patient care in cardiac health.

Authors' Contributions

The paper is entirely authored by Ahmad Ahmad.

Competing Interests

The author declare that he has no conflict of interest.

References

[1] A. A. Inamdar and A. C. Inamdar, "Heart failure: diagnosis, management and utilization," *Journal of Clinical Medicine*, vol. 5, no. 7, p. 62, Jun 2016.
[2] P. A. H. et al., "2022 american college of cardiology/american heart association/heart failure society of america guideline for the management of heart failure: executive summary," *Journal of Cardiac Failure*, vol. 28, no. 5, pp. 810–830, May 2022.
[3] N. R. A. H. Kashou and P. A. Noseworthy, "Ecg interpretation: clinical relevance challenges and advances," *Hearts*, vol. 2, no. 4, pp. 505–513, Nov 2021.
[4] J. Schläpfer and H. J. Wellens, "Computer-interpreted electrocardiograms: benefits and limitations," *Journal of the American College of Cardiology*, vol. 70, no. 9, pp. 1183–1192, Aug 2017.
[5] W. B. A. et al., "Implementing machine learning in interventional cardiology: the benefits are worth the trouble," *Frontiers in Cardiovascular Medicine*, vol. 8, p. 711401, Dec 2021.
[6] S. M. et al., "Artificial intelligence for clinical decision support for monitoring patients in cardiovascular icus: A systematic review," *Frontiers in Medicine*, vol. 10, p. 1109411, Mar 2023.
[7] M. A. Rahman and A. Tumian, "Variables influencing machine learning-based cardiac decision support system: A systematic literature review," *Applied Mechanics and Materials*, vol. 892, pp. 274–283, Jul 2019.
[8] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan 2020.
[9] S. S. et al., "Convolutional neural networks for radiologic images: a radiologist's guide," *Radiology*, vol. 290, no. 3, pp. 590–606, Mar 2019.
[10] J. Xue and L. Yu, "Applications of machine learning in ambulatory ecg," *Hearts*, vol. 2, no. 4, pp. 472–494, Oct 2021.
[11] I. U. H. I. Haq and B. Xu, "Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging," *Cardiovascular Diagnosis and Therapy*, vol. 11, no. 3, p. 911, Jun 2021.
[12] L. C. Y. Q. B. J. Schmidt and G. W. Wei, "Review of applications and challenges of quantitative systems pharmacology modeling and machine learning for heart failure," *Journal of Pharmacokinetics and Pharmacodynamics*, pp. 1–2, Feb 2022.
[13] M. H. Asyali, "Discrimination power of long-term heart rate variability measures," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 1, Sep 2003, pp. 200–203.
[14] Z. J. Y. Sun and A. C. Cheng, "Predicting cardiovascular disease from real-time electrocardiographic monitoring: An adaptive machine learning approach on a cell phone," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep 2009, pp. 6889–6892.
[15] W. C. G. L. S. S. Q. Jiang and H. Nguyen, "A chf detection method based on deep learning with rr intervals," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul 2017, pp. 3369–3372.
[16] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 54–64, Jul 2016.
[17] M. Liu and Y. Kim, "Classification of heart diseases based on ecg signals using long short-term memory," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul 2018, pp. 2707–2710.
[18] U. R. A. et al., "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ecg signals," *Applied Intelligence*, vol. 49, pp. 16–27, Jan 2019.
[19] J. C. Q. Zou and Y. Zhao, "Ecg signal classification based on deep cnn and bilstm," *BMC Medical Informatics and Decision Making*, vol. 21, pp. 1–2, Dec 2021.
[20] C. Padmavathi and S. V. Veenadevi, "Heart disease recognition from ecg signal using deep learning," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 2303–2316, 2020, online.
[21] O. S. L. et al., "Comprehensive electrocardiographic diagnosis based on deep learning," *Artificial Intelligence in Medicine*, vol. 103, p. 101789, Mar 2020.
[22] Y. Zhang and M. Xia, "Application of deep neural network for congestive heart failure detection using ecg signals," in *Journal of Physics: Conference Series*, vol. 1642, no. 1, Sep 2020, p. 012021.
[23] S. Kusuma and K. R. Jothi, "Ecg signals-based automated diagnosis of congestive heart failure using deep cnn and lstm architecture," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 247–257, Jan 2022.
[24] J. B. F. Mourad-Chehade and D. Laplanche, "Cnn and svm-based models for the detection of heart failure using electrocardiogram signals," *Sensors*, vol. 22, no. 23, p. 9190, Nov 2022.
[25] A. A. R. M. K. Elbashir and A. M. Ahmed, "Ecg heartbeat classification using convxgb model," *Electronics*, vol. 11, no. 15, p. 2280, Jul 2022.
[26] T. W. et al., "Automatic ecg classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, Jan 2021.
[27] R. Mogili and G. Narsimha, "Detection of cardiac arrhythmia from ecg using cnn and xgboost," *International Journal of Intelligent Engineering & Systems*, vol. 15, no. 2, 2022.
[28] G. Premalatha and V. T. Bai, "Design and implementation of intelligent patient in-house monitoring system based on efficient xgboost-cnn approach," *Cognitive Neurodynamics*, vol. 16, no. 5, pp. 1135–1149, 2022.
[29] F. K. X. Y. Z. Yuan and A. U. Rehman, "Ecg classification using 1-d convolutional deep residual neural network," *Plos One*, vol. 18, no. 4, p. e0284791, 2023.
[30] A. A.-J. S. A.-A. S. Islam and A. M. Artoli, "Person identification with arrhythmic ecg signals using deep convolution neural network," *Scientific Reports*, vol. 14, no. 1, p. 4431, 2024.
[31] B. Majhi and A. Kashyap, "Explainable ai-driven machine learning for heart disease detection using ecg signal," *Applied Soft Computing*, vol. 167, p. 112225, 2024.
[32] D. S. B. et al., "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, no. 3, pp. 661–670, Mar 1986.
[33] A. L. G. et al., "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun 2000.
[34] R. B. D. Kreiseler and A. Schnabel, "Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet," *Biomed. Tech.*, vol. 40, pp. 317–318, 1995.
[35] Y. C. C. K. H. Chang and G. J. Wu, "Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Applied Soft Computing*, vol. 73, pp. 914–920, Dec 2018.
[36] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct 2001.
[37] W. T. et al., "Xgboost prediction model based on 3.0 t diffusion kurtosis imaging improves the diagnostic accuracy of mri birads 4 masses," *Frontiers in Oncology*, vol. 12, p. 833680, Mar 2022.