CAT-TODNet: A Contextual Transformer-Based Optimized Deformable Convolution Framework for Efficient ECG-Based Heart Failure Detection

Vinitha V, V. Parthasarathy, R. Santhosh
Department of Computer Science and Engineering,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

***Abstract*—Heart Failure detection using Electrocardiogram (ECG) signals is a critical clinical task, as continuous analysis of cardiac waveforms supports early diagnosis and effective intervention. Despite advancements in machine learning and deep learning techniques, existing approaches often suffer from limited contextual representation, sensitivity to noise, and inadequate handling of non-stationary temporal deformations in ECG signals, which restrict diagnostic reliability. To address these challenges, this study introduces a novel deep learning framework termed Contextual Auxiliary Transformer with Triple Stacked Optimized Deformable Convolution Network (CAT-TODNet) for accurate heart failure detection from ECG signals. ECG recordings acquired from the MIT-BIH Arrhythmia Database are initially subjected to three-stage preprocessing, including denoising, signal smoothing, and Power Line Interference (PLI) removal, to enhance signal quality. The Contextual Auxiliary Transformer (CAT) module explicitly captures both static and dynamic contextual dependencies, enabling robust contextual feature extraction. These context-aware features are subsequently processed through triple stacked deformable convolution layers with adaptive receptive fields. To ensure stable offset estimation under non-stationary ECG conditions, the Al-Biruni Earth Radius (ABER) optimization algorithm is employed to optimize deformable convolution offsets, overcoming the limitations of gradient-based learning. Experimental results demonstrate that CAT-TODNet achieves an accuracy of 98.88.**

***Keywords*—Heart Failure (HF); Electrocardiogram (ECG); Artificial Intelligence (AI); Contextual Auxiliary Transformer (CAT); deformable convolution; optimization algorithm**

I. Introduction

Heart Failure (HF) is better understood as a complicated condition with a range of clinical indications rather than just one illness [1]. Heart failure is caused by a variety of circumstances, such as structural heart disease as well as functional changes in the heart. Poor circulation is the result of the heart's inability to pump blood efficiently throughout the body, which causes this ailment. People who have heart failure usually have a number of obvious symptoms, including exhaustion, edema in the ankles, and dyspneal [2]. In addition, patients could show symptoms from abnormalities in cardiac or noncardiac tissues, such as peripheral edema, lung fissure, and increased jugular vein pressure. Researchers have been looking into a variety of strategies for predicting cardiac conditions [3]. However, it has proven difficult to anticipate these conditions precisely in their early phases due to a number of factors, including complexity, implementation time, and the precision of the selected method. It is crucial to stress that early diagnosis and appropriate treatment can make a substantial difference in saving lives [4]. According to an international survey conducted by the European Society of Cardiology (ESC), 4.2 million new instances of heart failure were identified globally every year on top of the 30 million people who currently have the condition. Regretfully, a substantial proportion of patients among one fifth and half do not live past the first year following the onset of heart failure, and almost all patients experience a mortality rate within the period of five years [5].

Concerning the diagnosis of heart failure (HF), the electrocardiogram (ECG) is the non-invasive test that is most frequently used in the medical industry to record and capture a patient's cardiac activity [6]. However, the lack of sensitivity and specificity of classic manual analysis approaches employed to diagnose HF through the examination of ECG signals is frequently observed. ECG is a widely used diagnostic instrument that measures changes in the electrical activity of the heart to determine the condition of the heart [7]. It provides physicians with a clear-cut and simple-to-understand clinical signal. Although people with and without heart failure exhibit discernible differences in their ECG patterns, these differences are not very specific [8]. Moreover, it's critical to recognize that individual variations may be significant while manually examining ECG readings in order to diagnose heart failure. To be clearer, Cardiologists use visual analysis of ECG recordings to identify abnormalities [9]. However, this approach can take a long time, particularly when evaluating multiple ECG readings from different patients [10], [11]. Furthermore, the consistency of manual ECG signal interpretation might be affected by differences in interpretation between observers. Because of this unpredictability and time-consuming nature, developing and implementing Computer-Aided Diagnosis (CAD) systems becomes essential to improving the reliability of diagnosing heart failure using ECG records [12], [13].

There are numerous computer-assisted approaches for diagnosing heart failure have been developed as a result of the development of machine learning techniques [14]. These systems use automated feature extraction to determine heart problems with high accuracy. The most sophisticated instrument in the realm of biomedical engineering applications is Deep Learning (DL), a cutting-edge subset of Machine Learning (ML) [15]. It demonstrates its efficacy in this field by integrating hierarchical neural networks. The common ML models utilized are Naïve Bayes (NB), Gradient Boosting Classifier (GBC), Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting model (AdaBoost), and Stochastic Gradient Descent (SGD)

[16], [17]. Those ML classifiers are utilized for classification purpose of HF. The common DL models utilized are Long Short-Term Memory (LSTM), Convolution Neural Network (CNN), Gated Recurrent Unit (GRU), Deep Neural Network (DNN)., Auto Encoder (AE) etc.., [18]. Those DL models are utilized for feature extraction and feature analysis. Nonetheless, conventional ML/DL techniques necessitate manually designed features, which demand a trial-and-error approach to attain the highest classification accuracy [19]. To achieve effective and precise classification, it's essential to identify the most suitable visual characteristics for extracting ECG signals. Consequently, there's a need for further exploration into ECG signal classification without the need for feature extraction [20].

Despite recent advances in machine learning, deep learning, and transformer-based approaches for ECG-based heart failure detection, several critical limitations remain unresolved. Existing convolutional and recurrent models primarily focus on local temporal patterns and often fail to capture long-range contextual dependencies inherent in ECG signals. Although transformer-based models improve global representation, most rely on implicit self-attention mechanisms that lack explicit contextual supervision and remain sensitive to signal noise and inter-patient variability. In addition, conventional deformable convolution networks estimate offset parameters using gradient-based learning, which leads to unstable receptive field adaptation under non-stationary ECG deformations. These limitations reveal a clear research gap in developing a unified framework that explicitly models contextual information while enabling robust, optimized receptive field learning for reliable heart failure detection. This work introduces a tightly coupled framework that jointly integrates contextual transformer learning, deformable convolution, and metaheuristic offset optimization for ECG-based heart failure detection. Unlike existing transformer-based ECG models that rely on implicit self-attention or token-wise context aggregation, the Contextual Auxiliary Transformer (CAT) explicitly models static contextual dependencies prior to self-attention and enforces contextual consistency using auxiliary supervision. Furthermore, conventional deformable convolution layers estimate offsets through local gradient-based learning, which becomes unstable for ECG signals exhibiting non-stationary temporal deformations. To address this limitation, the offset estimation process is formulated as an optimization problem and solved using the Al-Biruni Earth Radius (ABER) algorithm, enabling robust exploration-exploitation balance for offset refinement. The unified interaction between CAT-derived contextual features and ABER-optimized deformable convolution distinguishes the proposed framework from existing transformer or deformable CNN architectures applied independently.

II. Related Works

In study [21] authors combine a variety of data mining approaches in conjunction with the SMOTE to improve the accuracy of forecasting the survival of patients with heart failure. In this study authors uses sophisticated data mining techniques to increase prediction accuracy and emphasizes to address class imbalance concerns in the dataset. In study [22], authors investigate the use of LSTM and CNN for automated congestive heart failure detection based on ECG readings. It emphasizes how well deep learning methods handle ECG data and produce precise diagnostic findings. A method that uses ECG signals to automatically diagnose heart failure was presented in study [23]. With an emphasis on facilitating prompt and accurate diagnostics, it probably involves techniques for feature extraction and classification to identify heart failure disorders. In order to improve the identification of cardiac disorders from ECG data, this work highlights the application of deep learning techniques in the form of ensemble models [24].

It draws attention to the possibility of enhancing overall detection performance by mixing several deep learning models. In [25], this research emphasizes the use of ECG readings in a time-adaptive manner to identify cardiovascular illnesses. It probably discusses the dynamic nature of ECG data and looks into techniques for adaptive or real-time detection, which could offer insightful information about how a disease develops. Authors in [26], designs a novel lightweight deep learning framework that uses heart sound recordings to classify cardiovascular illnesses. The special characteristics and architecture while emphasizing its potential for use in medical applications. In [27] authors provide an end-to-end deep learning system for improving the identification of cardiac illness by denoising heart sounds in real time.

It probably investigates how deep learning might improve the accuracy of disease identification by managing noise and variability in heart sound recordings. Authors in study [28], explores the creation of an explainable deep learning method for multi-label ECG data classification. It probably looks on techniques that enhance comprehension and openness by classifying ECG signals and offering comprehensible insights into the classification process. Authors in this work [29] focuses on using machine learning and deep learning techniques to analyze ECG images in order to diagnose cardiovascular disorders. It could go over feature extraction, image processing, and classification techniques, highlighting the importance of visual information in diagnosis. In study [30], authors investigate the use of deep learning neural networks for categorizing ECG signals in order to identify congestive heart failure and arrhythmia. It emphasizes how crucial it is to classify data accurately and effectively in order to recognize these serious cardiac problems.

In study [31], authors study probably presents an ECG-convolution-vision transformer network for inter-patient congestive heart failure identification. It emphasizes the creation of specialized networks for difficult jobs like diagnosing heart failure in various people. In study [32], explores a thorough method for the multilevel categorization and identification of cardiac arrhythmias. In order to increase the precision of arrhythmia diagnosis, it most likely blends deep convolutional neural networks with signal processing methods such as the superlet transform. Authors focuses on creating a wearable electrocardiogram device that leverages deep learning methods to identify cardiovascular illness in real time [33]. It emphasizes how important wearable technology is for ongoing surveillance and early detection.

The use of hybrid CNN-SVM deep neural networks in conjunction with Long Short-Term Memory (LSTM) networks to classify ECG signals into normal sinus rhythm, aberrant arrhythmia, and congestive heart failure [34]. It probably shows how useful it is to combine various deep learning architectures

to achieve precise classification. In study [35], authors draw attention to the effective use of machine learning algorithms in the prediction of cardiovascular illness. It probably highlights how crucial feature selection methods like Relief and LASSO are to raising the precision and effectiveness of prediction models. In study [36], this research presents a novel model that uses ECG signals to categorize heartbeats in order to diagnose heart illness.

The GOA in conjunction with the hybrid convolutional neural network presents a novel strategy that holds potential to improve the efficiency and accuracy of heart disease detection. Authors in study [37], centered on the use of ECG data in a multimodal fusion strategy for the diagnosis of cardiovascular illnesses. It most likely looks at ways to combine several information streams and enhance the accuracy of disease diagnosis by integrating multiple data sources or modalities. In study [38], authors discuss a DL method that uses empirical mode decomposition to extract ECG features in order to categorize cardiovascular illnesses. This approach is probably going to emphasize how crucial signal processing methods are to improving the data that is taken from ECG recordings in order to classify diseases. Authors in [39] highlight the creation and implementation of an autonomous heart failure prediction machine learning model. It might go over the usefulness of applying machine learning models in actual healthcare settings and show how they could be used to detect heart failure early on. In study [40], authors introduce a low-power ensemble network that uses ECG data to identify heart illness. It might concentrate on the effectiveness and computational viability of ensemble models, offering a method for detecting cardiac disease that is accurate and resource-efficient. Recent advances in transformer-based object detection have significantly improved contextual modeling, convergence efficiency, and real-time performance. Grounding DINO 1.5 [41] extends the DETR framework to open-set object detection by incorporating grounded representations and large-scale pretraining, enabling strong generalization to unseen categories. Relation-DETR [42] introduces explicit positional relation priors within the attention mechanism to enhance spatial reasoning and accelerate convergence. RT-DETR [43] focuses on real-time detection by eliminating non-maximum suppression and employing a hybrid encoder with optimized query selection, achieving superior speed–accuracy trade-offs compared to YOLO-based detectors. Despite these advancements, most existing DETR variants primarily target visual object localization and lack explicit contextual supervision or optimization-driven adaptive feature learning, motivating the development of more robust and context-aware detection frameworks.

III. Proposed Model

This section explains the proposed model with suitable mathematical and diagrammatic explanations. Fig. 1 represents the architectural pipeline of the proposed CAT-TODNet. The proposed research model is composed of three major stages: 1) Triple-Way Pre-Processing, 2) Transformer-based Feature Extraction, and 3) Optimized Heart Failure Detection. The overall pipeline of the proposed model is illustrated in Fig. 1. A detailed explanation of each stage is provided below.

A. Triple-Way Pre-Processing

In the Triple-Way Pre-Processing stage, the acquired ECG signals from the MIT-BIH database are pre-processed through three distinct operations: 1) denoising, 2) signal smoothing, and 3) removal of power-line interferences. The primary objective of this stage is to minimize the computational burden on subsequent Transformer and convolutional layers used for feature extraction and feature processing.

Formally, let the acquired ECG signal be represented as $s(t)$. The pre-processed signal $s'(t)$ can be defined as:

$$s'(t) = f_3(f_2(f_1(s(t)))) \tag{1}$$

where $f_1(\cdot)$, $f_2(\cdot)$, and $f_3(\cdot)$ represent the denoising, smoothing, and power-line interference removal functions, respectively.

This triple-stage processing enhances the quality of the input signal by suppressing noise components that are irrelevant to heart failure detection. Consequently, the removal of signal distortions improves both the feature representation and the overall performance of the Transformer and convolution operations in identifying the rate and severity of heart failures.

For denoising the acquired ECG signals, the Discrete Wavelet Transform (DWT) is employed since it effectively captures the signal information in both time and frequency domains. Let the original signal be represented as $s(t)$; it can be decomposed for denoising as follows:

$$s(t) = \sum_{k} c_{j_0}(k)\, \phi_{j_0,k}(t) + \sum_{j=j_0}^{J} \sum_{k} d_j(k)\, \psi_{j,k}(t) \tag{2}$$

In the above equation, $c_{j_0}(k)$ and $d_j(k)$ denote the approximation (scaling) and detail (wavelet) coefficients, respectively. The functions $\phi_{j_0,k}(t)$ and $\psi_{j,k}(t)$ represent the scaling and wavelet functions. The first term represents the coarse approximation of the signal, while the second term captures the detailed variations and noise components across different scales.

The Daubechies wavelet (DB) with a selected scale number is used as the mother wavelet to perform decomposition up to level five. The resulting coefficients—denoted as WC1–WC5 and WAC5—represent the wavelet and approximating coefficients, respectively. The noise components are eliminated by thresholding small-magnitude coefficients (setting them to zero), resulting in a denoised version of the ECG signal, $s_d(t)$, free from redundant and fluctuating artifacts.

1) Signal smoothing: After denoising, the ECG signal undergoes a smoothing process to further eliminate low-bit noise fluctuations. Let the denoised signal be $s_d(t)$. The signal length is computed and divided into N equal segments to evaluate local variations between neighboring peaks. The smoothing operation determines the local maxima deviations and removes minimal fluctuations as noise, resulting in the smoothed signal $s_s(t)$. The process can be expressed as:

$$s_s(t) = s_d(t) - \Delta n(t) \tag{3}$$

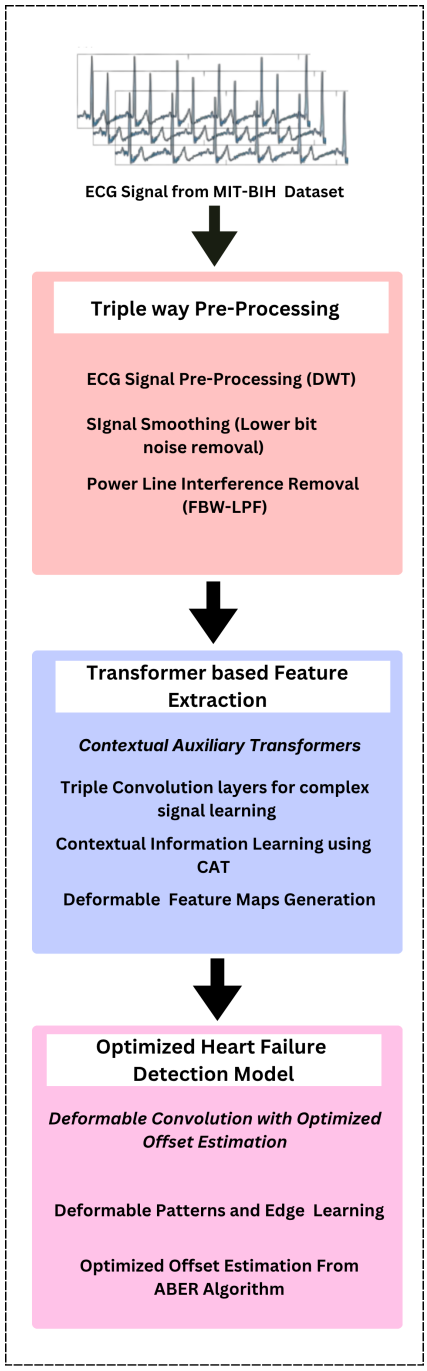

Fig. 1. Pipeline of the proposed work.

where $\Delta n(t)$ denotes the estimated low-bit noise component. This step refines the ECG waveform by preserving essential physiological characteristics while suppressing high-frequency artifacts.

2) Removal of power line interferences: To eliminate power line interference (PLI), typically occurring at 50 Hz, a Fractional Butterworth Low-Pass Filter (FBW-LPF) is adopted. The magnitude response of the FBW-LPF is mathematically formulated as:

$$|H(j\omega)| = \frac{1}{\sqrt{1 + \left(\frac{j\omega}{\omega_c}\right)^{2\alpha}}} \tag{4}$$

where ω represents the analog angular frequency (in rad/s), ω_c is the cutoff frequency (set at 50 Hz), and α defines the fractional order of the filter. This filter efficiently suppresses frequency components below the 50 Hz range, resulting in a clean waveform $s_f(t)$ suitable for the subsequent Transformer-based feature extraction stage.

B. Transformer-Based Feature Extraction

To address the limitations of conventional self-head attention Transformers in handling complex contextual dependencies, a Contextual Auxiliary Transformer (CAT) is adopted to enhance feature extraction and contextual representation. The CAT model extends the standard Transformer by integrating contextual learning within the self-attention mechanism, thereby reducing computational redundancy and improving discriminative representation of the ECG signals.

The pre-processed ECG signal $s_f(t)$ is given as the input to the CAT module, which operates based on the query (Q), key (K), and value (V) matrices. These matrices encode relationships across temporal and spatial contexts of the signal sequences. The CAT model reformulates these parameters into contextual groups and grids to efficiently capture cross-key dependencies. The contextual input y is derived from the feature representations, and the contextual interaction is modeled as the combination of keys and local attention matrices through three convolutional layers:

$$B = \text{Conv}_3(K \otimes y) \tag{5}$$

where, $\text{Conv}_3(\cdot)$ denotes a three-layer convolutional operation, $\otimes$ represents element-wise multiplication, and B refers to the local attention matrix incorporating contextual relationships. The matrix B assists in describing localized dependencies at each spatial position based on neighboring queries and keys, avoiding isolated query-key evaluations common in traditional Transformer-based models.

The resulting contextual attention allows the generation of a feature map F by integrating the value matrix V with the attention weights as follows:

$$F = B \cdot V \tag{6}$$

Here, F captures both static and dynamic feature interactions, making it highly capable of representing signal contexts with varying temporal behavior. The final CAT module output, denoted by X, integrates both static and dynamic contextual representations through the attention mechanism:

$$X = F_{static} \oplus F_{dynamic} \tag{7}$$

where, $\oplus$ denotes the element-wise summation of feature maps arising from static and dynamic contexts. The inclusion of deformable contextual maps enables the CAT network to

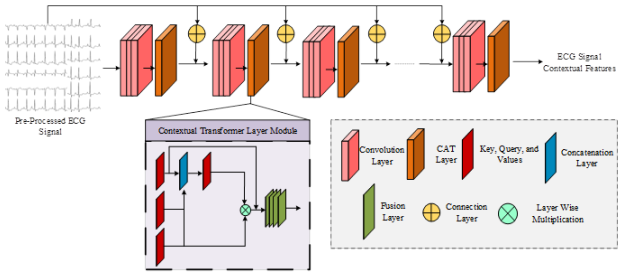

Fig. 2. Contextual Auxiliary Transformer (CAT) based feature extraction module.

model fine-grained local variations essential for accurate heart failure detection.

Finally, the CAT module is trained using a combined loss function that incorporates both the primary and auxiliary objectives, expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{main} + \lambda \mathcal{L}_{aux} \tag{8}$$

where, $\mathcal{L}_{main}$ and $\mathcal{L}_{aux}$ denote the general and auxiliary losses, respectively, and λ is the weighting coefficient controlling the contribution of the auxiliary task. Fig. 2 illustrates the architectural layout of the proposed Transformer-based feature extraction stage.

Conventional transformer architectures model contextual relationships implicitly through query-key similarity, treating each token independently during attention computation. In contrast, the Contextual Auxiliary Transformer employed in this work introduces a dual-context mechanism that combines static contextual representations with dynamic self-attention. The static context is extracted through grouped contextual convolutions applied to neighboring signal segments, enabling explicit modelling of local temporal dependencies prior to attention calculation. This contextual guidance stabilizes attention responses for ECG signals affected by noise, baseline drift, and waveform deformation. Moreover, the auxiliary loss enforces consistency between contextual and attention-driven representations, which is absent in standard self-attention or auxiliary attention models. As a result, CAT produces deformable-aware feature maps that are more suitable for subsequent offset-adaptive convolution.

The output feature map X obtained from the Transformer-based feature extraction module is provided to the optimized heart failure detection network, which comprises a sequence of deformable convolutional layers integrated with an optimized offset estimator. The deformable convolution is used to accurately capture complex and subtle ECG feature variations that are often missed by standard convolution operations.

1) Deformable convolution operation: The conventional convolution operation on a feature map x with kernel weights w and sampling locations p_n can be formulated as:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \tag{9}$$

where p_0 denotes the reference location on the output feature map, and R represents the regular grid kernel (e.g., 3×3).

In a deformable convolution layer, additional offsets Δp_n are introduced to adaptively sample feature locations, resulting in enhanced modeling of geometric variations in ECG waveforms. The deformable convolution operation is expressed as:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{10}$$

Here, Δp_n represents the learnable offset parameters, dynamically adjusted during training. These offsets are computed using bilinear interpolation to ensure differentiability. Although deformable convolutions enhance adaptability, they may also produce redundant or noisy features. To alleviate this, a weighted coefficient sampling method ω is employed to selectively suppress irrelevant responses as:

$$x'(p) = \omega(p) \cdot x(p) \tag{11}$$

This weighting scheme ensures that only informative local features contribute to the heart failure classification, thereby improving detection precision and generalization.

In deformable convolution networks, offset parameters are commonly learned using back-propagation alongside convolution weights. While effective for spatial image data, this strategy becomes sub-optimal for ECG signals due to irregular temporal deformations, inter-patient variability, and sensitivity to noise. Gradient-based offset learning is prone to local minima and unstable convergence under such conditions. To overcome this limitation, offset estimation is reformulated as a global optimization problem, where the objective is to maximize feature similarity alignment across deformable receptive fields. The Al-Biruni Earth Radius (ABER) algorithm is employed to optimize offset vectors due to its balanced exploration and exploitation capability, enabling robust search in highly non-convex offset spaces. Unlike conventional optimizers such as PSO or gradient descent, ABER maintains adaptive search radius control, which improves convergence stability and prevents premature stagnation in offset learning.

2) Optimized offset estimation using ABER algorithm: To further enhance the offset estimation, the Al-Biruni Earth Radius (ABER) optimization algorithm is integrated as an intelligent offset estimator. The ABER algorithm finds optimal offsets by maximizing similarity between reference and input ECG feature maps. Let the inputs to the estimator be denoted as F_r (reference features) and F_i (input features); the optimization problem can be defined as:

$$\min_{\Delta p} f(F_r, F_i, \Delta p) \tag{12}$$

where, $f(\cdot)$ represents the fitness function evaluating the similarity between input and reference features. The search space dimension is denoted as D, and the candidate solution at iteration t is expressed as X_t.

The ABER algorithm consists of two primary phases:

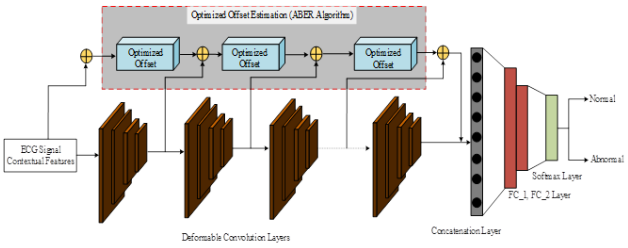

Fig. 3. Illustration of optimized heart failure detection model.

3) Non-Local feature enhancement: To capture long-range dependencies within the ECG features, non-local operations are employed after offset optimization. This operation enhances the model's capacity to represent both global and local characteristics. Mathematically, the non-local operation between an input feature f_i and an output position f_o is defined as:

$$f_o(i) = \frac{1}{C(i)} \sum_{\forall j} \theta(f_i(i)) \cdot \phi(f_i(j)) \cdot g(f_i(j)) \tag{13}$$

where, $\theta(\cdot)$, $\phi(\cdot)$, and $g(\cdot)$ denote embedding, pairwise, and transformation functions, respectively, and $C(i)$ is a normalization factor. The combined use of deformable convolution, ABER-optimized offset estimation, and non-local contextual modeling effectively refines the heart failure detection mechanism.

Fig. 3 illustrates the architecture of the proposed Optimized Heart Failure Detection Model.

a) Exploitation phase: In this phase, the primary objective is to enhance the performance of the previously explored candidate solutions by refining them toward the optimal offset estimator. The formulation for the exploitation-based ABER can be expressed as:

$$X_t^{\text{new}} = X_{\text{best}} + \gamma \cdot (R_d \cdot \text{rand}()) \tag{14}$$

where, X_t^{new} represents the updated offset vector at iteration t, X_{best} is the best solution obtained from the exploration phase, R_d is the search area diameter [from Eq. (13)], and γ is the adaptive learning rate controlling step-size toward the best candidate.

To further refine the best offset estimator within the exploitation region, the updated position is evaluated using:

$$X_{t+1} = X_t + \phi \cdot (X_{\text{best}} - X_t) + \delta \cdot \text{rand}() \tag{15}$$

where, ϕ and δ denote adaptation coefficients balancing exploitation intensity and stochastic perturbation, respectively. The random vector $\text{rand}() \in [0, 1]$ introduces controlled randomness to avoid premature convergence and improve the stability of offset estimation.

Finally, the best candidate solution during dual iteration is updated as:

Algorithm 1 ABER-Based Optimized Offset Estimation for Deformable Convolution

Require: Reference feature F_r, Input feature F_i, Parameters R_d, β, ϕ, δ
Initialize candidate solutions X_i randomly within the search space.
for $t = 1$ to $T_{\max}$ **do**
  Evaluate fitness $f(X_i)$ for all candidates.
  Update best solution: $X_{\text{best}} = \arg\min f(X_i)$.
  {Exploration Phase}
  Compute: $R_d = 2E_r \cdot \text{rand}() - E_r$.
  Update candidate position: $X_t^{\text{new}} = X_t + \beta \cdot R_d$.
  {Exploitation Phase}
  Update position:

$$X_{t+1} = X_t + \phi(X_{\text{best}} - X_t) + \delta \cdot \text{rand}()$$

  if $f(X_{t+1}) < f(X_{\text{best}})$ **then**
    $X_{\text{best}} = X_{t+1}$.
  end if
  Gradually reduce R_d to ensure convergence.
end for
return Optimized offset vector: $X_{\text{optimal}} = X_{\text{best}}$.

$$X_{\text{optimal}}^{t+1} = \begin{cases} X_{t+1}, & \text{if } f(X_{t+1}) < f(X_{\text{best}}) \\ X_{\text{best}}, & \text{otherwise} \end{cases} \tag{16}$$

where, $f(\cdot)$ denotes the fitness function evaluating the similarity and feature matching accuracy between the reference and input ECG signals.

C. Experimental Evaluation

This section extensively details the results achieved through our novel approach, considering both quantitative and qualitative evaluations in contrast to established methodologies. Furthermore, it thoroughly assesses how our method performs across three separate datasets when compared to existing approaches. It initiates by outlining the dataset characteristics, proceeds with an in-depth analysis of results under various conditions, and concludes with a meaningful discussion, providing valuable insights based on the findings.

The electrocardiogram (ECG) signals used in this study are obtained from the MIT-BIH Arrhythmia Database, which is publicly available through PhysioNet. The database consists of annotated ECG recordings collected from 48 subjects, with each record sampled at 360 Hz and containing two leads. The signals are segmented into fixed-length windows to facilitate uniform analysis and model training.

For experimental evaluation, ECG segments are categorized into heart failure and non–heart failure classes based on the provided annotations. A patient-wise data split strategy is employed to avoid data leakage, ensuring that samples from the same subject do not appear in more than one subset. The dataset is divided into training, validation, and testing sets using a fixed ratio.

The heart failure detection model, referred to as CATTODNet, is implemented using MATLAB. The model is trained for 100 epochs with a batch size of 32 and a learning rate of 0.001, using the Adam optimization algorithm. All

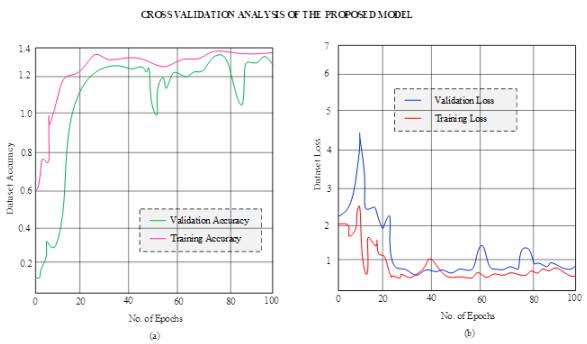

Fig. 4. (a) Shows the training and validation accuracy, (b) shows the training and validation loss.

experiments are conducted on an HP Pavilion laptop equipped with an AMD Ryzen 5 5600H processor with Radeon Graphics operating at 3.30 GHz, 8 GB of Random-Access Memory (RAM), and the Windows 11 Operating System. The same training configuration is maintained across all experiments to ensure reproducibility and fair performance evaluation.

The performance metrics utilized for validating the proposed work can be formulated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \tag{19}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

where, TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative rates, respectively. The AUC-ROC curve offers a comprehensive assessment of the classifier's performance across a range of possible thresholds. A higher AUC signifies superior performance by the classifier.

Fig. 4(a) and (b) represent the training and validation accuracy, and training and validation loss of the proposed work, respectively.

The following state-of-the-art models were selected for comparison against the proposed CAT-TODNet architecture:

- SMOTE-HF [21]: This study adopts ECG data for analysing the heart failure by comparing the nine benchmark machine learning algorithms.
- DeepCNN-LSTM [6]: This study also utilized ECG signals for diagnosing the heart disease using combine deep learning algorithm.

TABLE I. Quantitative Evaluation of Pre-Processing Across Methods

Model	Accuracy	Precision	Recall	F1-Score
CAT-TODNet (w/ Pre-processing)	**0.9899**	**0.9903**	**0.9800**	**0.9876**
CAT-TODNet (w/o Pre-processing)	0.9435	0.9323	0.9326	0.9300
DeepCNN-LSTM [22]	0.9267	0.9309	0.9215	0.9276
CHF-CVTN [31]	0.9154	0.9149	0.9184	0.9126

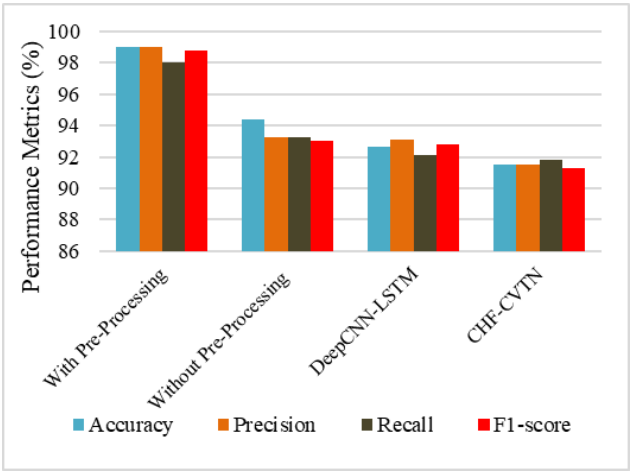

Fig. 5. Evaluation of pre-processing process.

- Ensemble-DL [24]: In this research, an ensemble of DL models is utilized for detecting the heart disease from two benchmark ECG database.
- Heart-CVDD [25]: This research adopts two stage classification algorithms for cardiovascular disease detection from ECG signals.
- CHF-CVTN [31]: In this study the heart failure disease are diagnosed by adopting convolution based transformer network using ECG signals.
- CHF-DL-ML [34]: The diagnosis of congenital heart failure disease gets diagnosed by adopting combined DL and ML algorithms named LSTM-CNN-SVM from ECG signals.

IV. Experimental Evaluation

Fig. 5 represents the comparison of the proposed pre-processing process and its impacts with the existing works. In order to employ the better pre-processing examination process, we also split the proposed CAT-TODNet into two scenarios such as with and without pre-processing respectively. In addition to that, we have compared the proposed CAT-TODNet with two existing works that adopts pre-processing such as DeepCNN-LSTM [22], and CHF-CVTN [31] respectively. The quantitative results of the pre-processing are shown in Table I below.

From Fig. 5 and Table I, it is seen that the proposed scenario with pre-processing achieves higher results whereas the proposed scenario without pre-processing lesser than the with pre-processing scenario. On the other hand, the existing works which uses pre-processing in which the DeepCNN-LSTM

TABLE II. COMPARISON OF CONVENTIONAL MODELS AND PROPOSED FEATURE EXTRACTION MODEL

Model	Accuracy	Precision	Recall	F1-Score	AUC
RF	0.8478	0.8598	0.8502	0.8487	–
DT	0.8646	0.8734	0.8615	0.8711	–
SVM	0.9156	0.9110	0.9199	0.9197	–
DNN	0.9465	0.9451	0.9595	0.9537	–
AEs	0.9552	0.9509	0.9698	0.9689	–
Proposed (CAT-TODNet)	**0.9777**	**0.9888**	**0.9810**	**0.9836**	**–**

TABLE III. COMPARISON OF EXISTING METHODS AND PROPOSED FEATURE EXTRACTION MODEL

Model	Accuracy	Precision	Recall	F1-Score	AUC
SMOTE-HF [21]	0.8889	0.8809	0.8813	0.8898	–
DeepCNN-LSTM [22]	0.9046	0.9045	0.9026	0.9022	–
Ensemble-DL [24]	0.9367	0.9321	0.9300	0.9308	–
Heart-CVDD [25]	0.9576	0.9562	0.9698	0.9648	–
CHF-CVTN [31]	0.9763	0.9720	0.9708	0.9790	–
Proposed (CAT-TODNet)	**0.9988**	**0.9999**	**0.9921**	**0.9948**	**–**

utilizes only normalization procedure whereas the CHF-CVTN resampling, normalization, and segmentation respectively. The quantitative results of the existing works are significantly lower than the proposed scenarios. The reason for such higher results of the proposed work is that, we perform signal denoising, signal smoothing, and powerline interference removal which effectively suppress noise components in the ECG signal thereby providing denoised signal for further processing that also improve the performance results. Whereas the existing works are not focused on resolving noises whereas they only normalize them to certain limit which resolves the complexity but reduces the performance of the overall model. In detailed manner, the proposed with pre-processing scenario achieves higher accuracy, precision, recall, and F1-score rate of (4.64%–7.45%), (5.8%–7.54%), (4.74%–6.16%), and (5.76%–7.5%) than the existing works.

A. Evaluation of Feature Extraction

The validation of proposed feature extraction model named Contextual Auxiliary Transformer (CAT) to the conventional model and existing works are detailed in this section. The conventional models includes Machine Learning (ML) and Deep Learning (DL) models such as Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Deep Neural Network (DNN), Auto Encoders (AEs) whereas the existing feature extraction models includes SMOTE-HF [21] which uses ML algorithms, DeepCNN-LSTM [22] which uses CNN for feature extraction, Ensemble-DL [24] uses self-organizing maps for feature extraction, Heart-CVDD [25] uses time based dynamic CNN for feature extraction, and CHF-CVTN [31] uses transformer based convolution model for feature extraction. The quantitative results of the feature extraction in terms of conventional models and existing research works are shown in Table II and Table III below.

The graphical representation depicted in Fig. 6(a) and 6(b), and Fig. 7(a) and 7(b) compares the proposed Feature Extraction (FE) approach with both traditional models and previous studies, evaluating their performance metrics and AUC-ROC values. Specifically, Fig. 6(a) and 6(b) analyse the performance metrics of conventional models and prior works, respectively. The illustration indicates that the proposed model outperforms both the traditional models and existing works. Regarding the quantitative outcomes, the accuracy, precision, recall, and F1-score achieved by the proposed FE are 2.55%–12.99%, 3.79%–2.9%, 1.12%–13.08%, and 1.47%–13.49% higher than the conventional ML/DL models, whereas for existing works the proposed work achieved are 2.25%–10.99%, 2.79%–11.9%, 2.13%–8.95%, and 1.58%–10.5% higher. Similarly, Fig. 7(a) and 7(b) present the AUC-ROC evaluation for conventional models and existing works, respectively. The AUC-ROC is higher for the conventional models compared to existing works. This superior performance of the proposed FE stems from its utilization of an attention-based transformer model, enabling robust extraction of signal context features by analysing intricate signal patterns. Conversely, the shortcomings in conventional and existing models include higher tendencies toward overfitting, training complexity, adaptability issues, and respectively lower accuracy in feature extraction.

TABLE IV. COMPARISON OF EXISTING WORKS AND PROPOSED OPTIMIZED CLASSIFICATION MODEL

Model	Accuracy	Precision	Recall	F1-score	AUC
SMOTE-HF [21]	0.8555	0.8500	0.8526	0.8523	–
DeepCNN-LSTM [22]	0.8857	0.8845	0.8816	0.8812	–
Ensemble-DL [24]	0.9078	0.9032	0.9022	0.9019	–
Heart-CVDD [25]	0.9387	0.9373	0.9317	0.9359	–
CHF-DL-ML [34]	0.9663	0.9610	0.9609	0.9690	–
DeepECG-Net [44]	0.9820	0.9820	0.9680	0.9750	–
Proposed (CAT-TODNet)	**0.9888**	**0.9899**	**0.9821**	**0.9847**	**–**

B. Evaluation of Heart Disease Classification Model

This section thoroughly outlines the validation process for the proposed optimized heart failure detection model concerning existing methods. The current classification models under scrutiny encompass diverse approaches: SMOTE-HF [21], employing nine ML algorithms; DeepCNN-LSTM [22], utilizing LSTM; Ensemble-DL [24], employing autoencoders; Heart-CVDD [25], utilizing Bi-directional LSTM; and CHF-DL-ML [34], utilizing support vector machines. The quantitative outcomes derived from these models concerning heart disease classification are tabulated below in Table IV for a comprehensive comparison (also see, Fig. 8 and 9).

V. DISCUSSION AND LIMITATIONS

This section explores the overall summary of the proposed work with its implementation results. From the implementation summary, the results achieved demonstrate superior performance than the existing works. We also compare the proposed implementation model with the conventional ML and DL models respectively. It is clearly depicted that in the above tables and figures that the results of the proposed work are within 97%–100%. The near-saturation accuracy is achieved under a strict patient-wise evaluation protocol, which mitigates overfitting and ensures reliable performance estimation. The major difference among the proposed and existing works are that, most of the existing works utilized CNN and ML algorithms as their FE and classification model

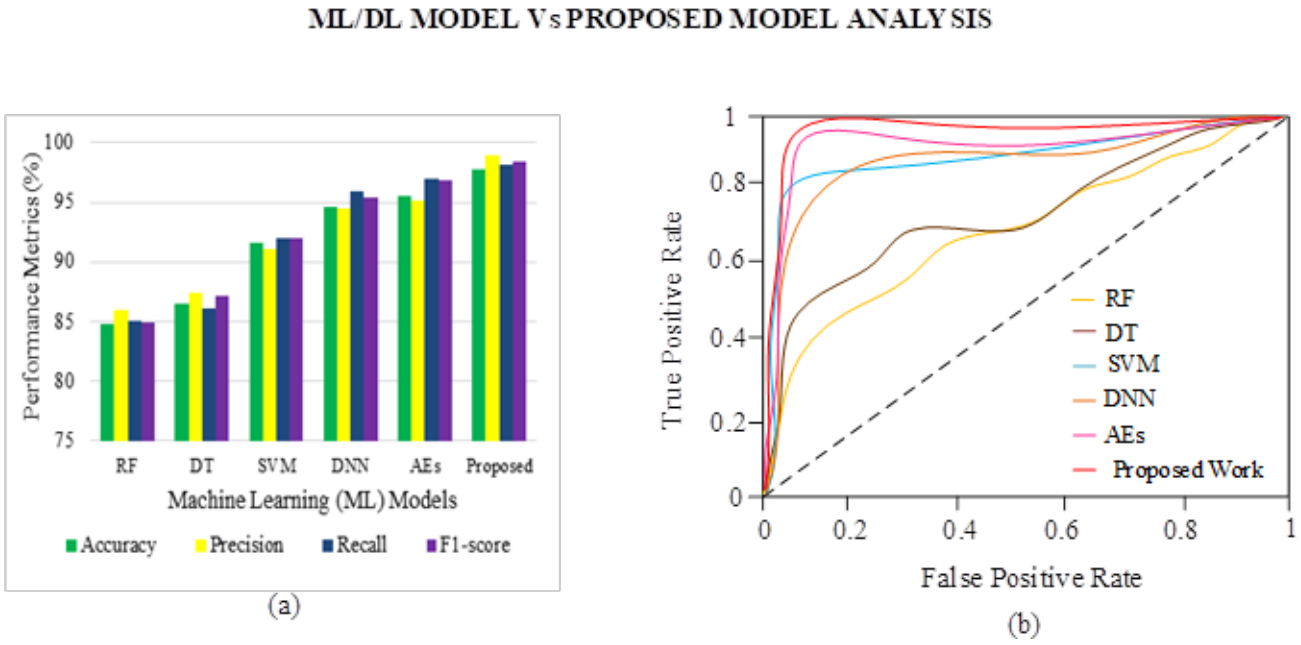

Fig. 6. (a) Performance metrics analysis of proposed work vs. ML/DL Models, (b) AUC comparison of proposed vs. ML/DL models.

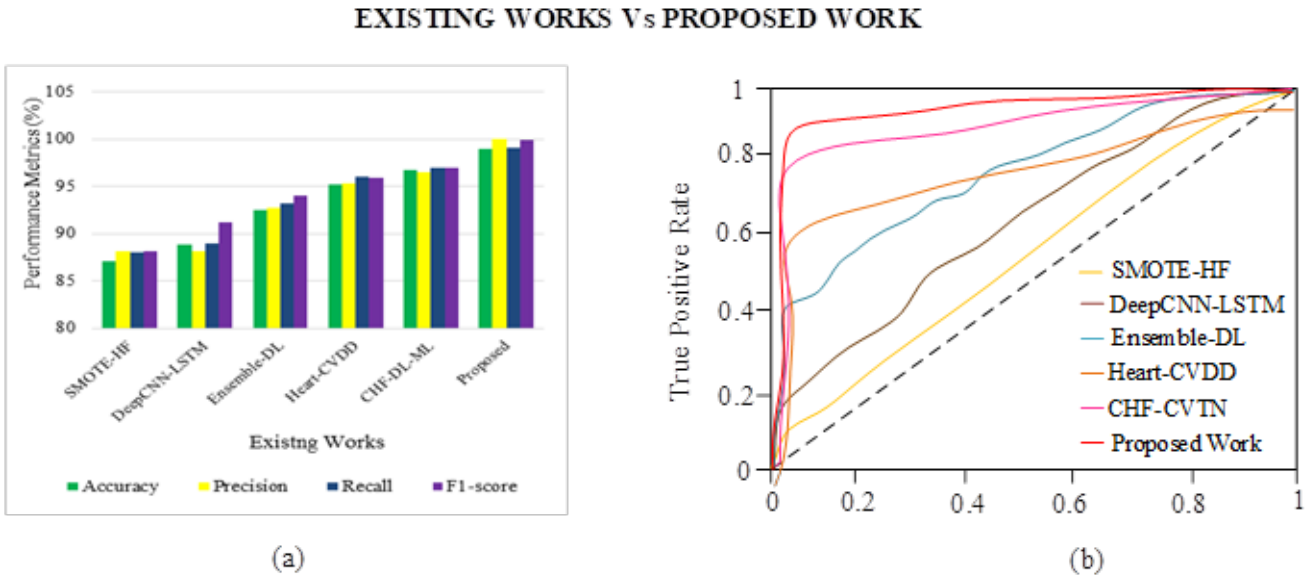

Fig. 7. (a) Performance metrics analysis of proposed vs. Existing models, (b) AUC comparison of proposed vs. Existing models.

and there were no improvements over that models which leads to higher complexity, time consumption, and poor accuracy in detecting heart disease respectively. On the whole, the major contribution of the proposed CAT-TODNet is that we utilize the attention based contextual transformer model for feature extraction which extracts the signal contextual features with the signal characteristics. Furthermore, those contextual features are provided to the optimized offset estimator which dynamically set the offset value to the deformable convolution layers to capture the feature interdependencies and provide the accurate output.

On the other hand, one of the major limitations of our study is that, the proposed research is not suitable for remote detection/prediction of heart disease detection. To be clearer, the designed model is centralized and did not adopt for distributed remote diagnosis for heart disease detection. On the other hand, we utilize only limited data to train the proposed model which make it to provide only limited results.

VI. Conclusion

The detection of Heart Failure (HF) through continuous monitoring of heart wave patterns is pivotal in reducing mortality rates. Utilizing Electrocardiogram (ECG) alongside Artificial Intelligence (AI) has been explored in previous

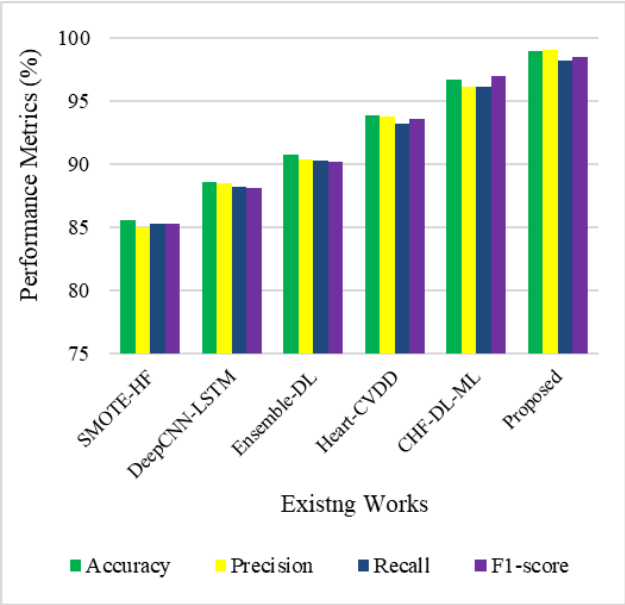

Fig. 8. Evaluation of heart failure detection model for proposed vs. Existing works.

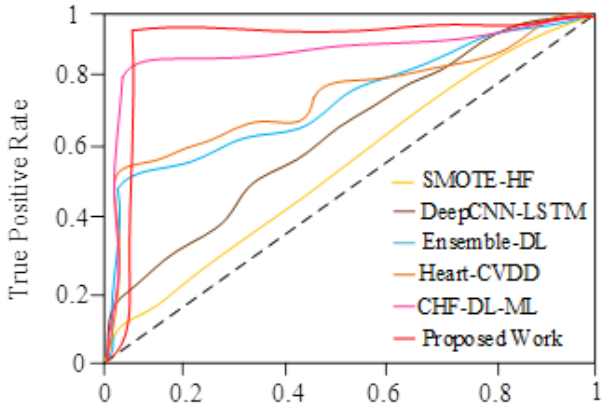

Fig. 9. AUC analysis of proposed heart failure detection model.

research, albeit with issues of accuracy and complexity. To address these shortcomings, our innovation, CAT-TODNet, integrates a stacked and optimized deformable convolution with a Contextual Attention Transformer (CAT) for HF detection from ECG signals. By preprocessing data from the MIT-BIH dataset through denoising, signal smoothing, and power line interference removal, we optimize feature extraction through CAT. This model learns contextual features and generates deformable feature maps crucial for HF detection. Our optimized HF model incorporates dynamic offset estimation in deformable convolution layers, enhancing the interpretation of contextual features. The integration of the Al-Biruni Earth Radius (ABER) optimization algorithm dynamically adjusts the offset estimator, improving feature similarities and ensuring precise and reliable outputs.

MATLAB simulations validate our approach, exhibiting superior performance compared to existing methods across multiple evaluation metrics such as accuracy, recall, precision, F1-score, and AUC. This research underscores the potential of CAT-TODNet in significantly enhancing HF detection accuracy, thereby contributing to reducing HF-related mortality rates. Future work could delve deeper into real-time implementation and explore the scalability of the proposed model across diverse datasets for robust and widespread applicability in clinical settings.

REFERENCES

[1] M. R. Maurya, N. U. S. Riyaz, M. S. B. Reddy, H. C. Yalcin, H. M. Ouakad, I. Bahadur, and K. K. Sadasivuni, "A review of smart sensors coupled with Internet of Things and Artificial Intelligence approach for heart failure monitoring," *Medical & Biological Engineering & Computing*, pp. 1–19, 2021.

[2] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1329–1333.

[3] A. Gahane and C. Kotadi, "An Analytical Review of Heart Failure Detection based on IoT and Machine Learning," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 2022, pp. 1308–1314.

[4] M. O. Butt, A. U. Rehman, S. Javaid, T. M. Ali, and A. Nawaz, "An Application of Artificial Intelligence for an Early and Effective Prediction of Heart Failure," in *2022 Third International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT)*, 2022, pp. 1–6.

[5] A. Newaz, F. S. Haq, and N. Ahmed, "A case study on risk prediction in heart failure patients using random survival forest," in *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2021, pp. 1–6.

[6] A. Kumar, K. Rathor, S. Vaddi, D. Patel, P. Vanjarapu, and M. Maddi, "ECG Based Early Heart Attack Prediction Using Neural Networks," in *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2022, pp. 1080–1083.

[7] R. K. Mahendran, P. Velusamy, and P. Pandian, "An efficient priority-based convolutional auto-encoder approach for electrocardiogram signal compression in Internet of Things based healthcare system," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 1, p. e4115, 2021.

[8] R. Avanzato and F. Beritelli, "Heart disease recognition based on extended ECG sequence database and deep learning techniques," in *2022 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, 2022, pp. 117–121.

[9] S. N. Singh and M. Bhushan, "Smart ECG monitoring and analysis system using machine learning," in *2022 IEEE VLSI Device Circuit and System (VLSI DCS)*, 2022, pp. 304–309.

[10] I. H. Tsai and B. I. Morshed, "Beat-by-beat Classification of ECG Signals with Machine Learning Algorithm for Cardiac Episodes," in *2022 IEEE International Conference on Electro Information Technology (eIT)*, 2022, pp. 311–314.

[11] R. R. Lopes *et al.*, "Improving electrocardiogram-based detection of rare genetic heart disease using transfer learning: An application to phospholamban p.Arg14del mutation carriers," *Computers in Biology and Medicine*, vol. 131, p. 104262, 2021.

[12] D. R. Krithika and K. Rohini, "Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics," in *2021 International Conference on Computing Sciences (ICCS)*, 2021, pp. 42–46.

[13] A. Al Ahdal, M. Rakhra, S. Badotra, and T. Fadhaeel, "An integrated machine learning techniques for accurate heart disease prediction," in *2022 International Mobile and Embedded Technology Conference (MECON)*, 2022, pp. 594–598.

[14] Z. Sun *et al.*, "Comparing machine learning models and statistical models for predicting heart failure events: a systematic review and meta-analysis," *Frontiers in Cardiovascular Medicine*, vol. 9, p. 812276, 2022.

[15] M. A. M. Hasan, J. Shin, U. Das, and A. Srizon, "Identifying prognostic features for predicting heart failure by using machine learning algorithm," in *2021 11th International Conference on Biomedical Engineering and Technology*, 2021, pp. 40–46.

[16] J. Botros, 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, 2022.

[17] H. R. Al-Absi *et al.*, "Risk factors and comorbidities associated to cardiovascular disease in Qatar: A machine learning based case-control study," *IEEE Access*, vol. 9, pp. 29929–29941, 2021.

[18] M. F. Aslan, K. Sabanci, and A. Durdu, "A CNN-based novel solution for determining the survival status of heart failure patients with clinical record data: numeric to image," *Biomedical Signal Processing and Control*, vol. 68, p. 102716, 2021.

[19] S. Hussain *et al.*, "Novel deep learning architecture for predicting heart disease using CNN," in *2021 19th OITS international conference on information technology (OCIT)*, 2021, pp. 353–357.

[20] V. Jahmunah *et al.*, "Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals," *Computers in Biology and Medicine*, vol. 134, p. 104457, 2021.

[21] A. Ishaq *et al.*, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021.

[22] 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*, 2022.

[23] L. Chen, H. Yu, Y. Huang, and H. Jin, "ECG Signal-Enabled Automatic Diagnosis Technology of Heart Failure," *Journal of Healthcare Engineering*, 2021.

[24] A. Rath, D. Mishra, G. Panda, S. C. Satapathy, and K. Xia, "Improved heart disease detection from ECG signal using deep learning-based ensemble model," *Sustainable Computing: Informatics and Systems*, vol. 35, p. 100732, 2022.

[25] M. S. Haleem *et al.*, "Time adaptive ECG driven cardiovascular disease detector," *Biomedical Signal Processing and Control*, vol. 70, p. 102968, 2021.

[26] S. B. Shuvo *et al.*, "CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings," *IEEE Access*, vol. 9, pp. 36955–36967, 2021.

[27] S. N. Ali *et al.*, "An End-to-End Deep Learning Framework for Real-Time Denoising of Heart Sounds for Cardiac Disease Detection in Unseen Noise," *IEEE Access*, vol. 11, pp. 87887–87901, 2023.

[28] E. Sala and S. Ursprung, "Artificial Intelligence in Radiology: The Computer's Helping Hand Needs Guidance," *Radiology: Artificial Intelligence*, vol. 2, no. 6, p. e200207, 2020.

[29] M. Abubaker and B. Babayiğit, "Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods," *IEEE Transactions on Artificial Intelligence*, vol. 4, pp. 373–382, 2023.

[30] S. Krishnakumar *et al.*, "Detection of Arrhythmia and Congestive Heart Failure Through Classification of ECG Signals Using Deep Learning Neural Network," in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 2021, pp. 1–7.

[31] T. Liu *et al.*, "Inter-Patient Congestive Heart Failure Detection Using ECG-Convolution-Vision Transformer Network," *Sensors*, vol. 22, 2022.

[32] P. M. Tripathi, A. Kumar, M. Kumar, and R. S. Komaragiri, "Multi-level Classification and Detection of Cardiac Arrhythmias With High-Resolution Superlet Transform and Deep Convolution Neural Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[33] P. Wang *et al.*, "A Wearable ECG Monitor for Deep Learning Based Real-Time Cardiovascular Disease Detection," 2022.

[34] A. C. Cinar and S. A. Tuncer, "Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 24, pp. 203–214, 2020.

[35] P. Ghosh *et al.*, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021.

[36] A. Tyagi and R. Mehra, "Intellectual heartbeats classification model for diagnosis of heart disease from ECG signal using hybrid convolutional neural network with GOA," *SN Applied Sciences*, vol. 3, 2021.

[37] G. Song *et al.*, "A Multimodel Fusion Method for Cardiovascular Disease Detection Using ECG," *Emergency Medicine International*, vol. 2022.

[38] Y. Li *et al.*, "A deep learning approach to cardiovascular disease classification using empirical mode decomposition for ECG feature extraction," *Biomedical Signal Processing and Control*, vol. 79, p. 104188, 2023.

[39] K. A. Pandian, T. S. Sai Kumar, S. P. Dhandare, and S. Thabasum Aara, "Development and Deployment of a Machine Learning Model for Automatic Heart Failure Prediction," in *2021 Asian Conference on Innovation in Technology (ASIANCON)*, 2021, pp. 1–6.

[40] S. Shin *et al.*, "Lightweight Ensemble Network for Detecting Heart Disease Using ECG Signals," *Applied Sciences*, vol. 12, no. 20, p. 10562, 2022.

[41] T. Ren *et al.*, "Grounding DINO 1.5: Advancing the edge of open-set object detection," *arXiv preprint arXiv:2405.10300*, 2024.

[42] X. Hou *et al.*, "Relation DETR: Exploring explicit position relation prior for object detection," in *European Conference on Computer Vision (ECCV)*, pp. 89–105, Springer Nature Switzerland, 2024.

[43] Y. Zhao *et al.*, "DETRs beat YOLOs on real-time object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16965–16974, 2024.

[44] M. Alghieth, "DeepECG-Net: A hybrid transformer-based deep learning model for real-time ECG anomaly detection," *Scientific Reports*, vol. 15, no. 1, p. 20714, 2025.