Al Furat Journal of Innovations in Engineering Applications (AJIEA)

EFFICIENT DEEP LEARNING FOR CHAGAS CARDIOMYOPATHY DETECTION: DATA-DRIVEN ECG FEATURE REDUCTION

Ahmed Faeq Hussein[1*] , Auns Q Al-Neami[2] , Qais Ahmed Habash[3]

[1]Artificial Intelligence and Robotics Engineering Department, College of Engineering, Al-Nahrain University,10072, Iraq E-mail: auns.q.hashim@nahrainuniv.edu.iq

[2]Bioedical Engineering Department, College of Engineering, Al-Nahrain University,10072, Iraq E-mail: auns.q.hashim@nahrainuniv.edu.iq

[3]Bioedical Engineering Department, College of Engineering, Al-Nahrain University,10072, Iraq E-mail: qais.a.habash@nahrainuniv.edu.iq

[*]Corresponding author E-mail: ahmed.f.hussein@nahrainuniv.edu.iq

Abstract. *Chagas cardiomyopathy, a severe complication of chronic Trypanosoma cruzi infection, requires early and efficient diagnosis to improve outcomes. While electrocardiograms (ECGs) are accessible diagnostic tools, their high dimensionality limits the scalability of deep learning methods. This study proposes an interpretable and computationally efficient deep learning pipeline combining Principal Component Analysis (PCA) for feature reduction and a Convolutional Neural Network (CNN) for classification of ECG signals.*

Using data from the publicly available SaMi-Trop cohort, the ECG signals were preprocessed with a Butterworth high-pass filter to remove baseline wander, followed by PCA to reduce feature space while preserving critical variance. A custom CNN was then trained on the reduced feature set to classify signals as Chagas or non-Chagas.

The PCA+CNN model achieved 95.52% sensitivity, 91.11% specificity, 93.31% accuracy, and an AUC of 96.78%, outperforming the baseline CNN trained on unreduced data. The approach reduced overfitting, improved generalization, and accelerated convergence, demonstrating strong potential for use in real-time and low-resource healthcare settings.

Keywords: *Chagas cardiomyopathy; Electrocardiogram (ECG); Deep learning; Principal Component Analysis (PCA); Feature reduction; Biomedical signal processing; Medical AI*

1. INTRODUCTION

Chagas disease, caused by the protozoan Trypanosoma cruzi, remains a significant public health concern in Latin America and is increasingly recognized in non-endemic countries due to global migration. A major complication of chronic Chagas disease is Chagas cardiomyopathy, a life-threatening condition characterized by progressive cardiac dysfunction, arrhythmias, and heart failure. Early and accurate diagnosis of Chagas cardiomyopathy is essential to guide treatment and improve patient outcomes. However, current diagnostic approaches often rely on resource-intensive imaging or laboratory techniques that are not always accessible, especially in low-resource settings [1].

Electrocardiograms (ECGs) are widely used noninvasive diagnostic tools, inexpensive and accessible to the general population, and they record electrical activity in the heart and reflect all the

abnormalities that may exist in the heart. Since it can be easily obtained and has diagnostic sensitivity, shows promise for investigating Chagas cardiomyopathy into the disease of Chagas cardiomyopathy [2]. It is also quite labor-intensive because up until today, reading ECGs is based on human expertise and therefore very subjective. In the last few years, artificial intelligence and deep learning models have started providing high levels of automation for ECG analysis, accurate results that can easily scale up to the level of all health systems [3].

Yet, challenges remain in deploying such models effectively in real clinical settings. Deep neural networks, like CNNs, are quite infamous for both their high computational requirements and tendency toward overfitting, especially on high-dimensional data or small datasets [4]. The problems with computational efficiency can then be solved by applying dimensionality reduction such as PCA to reduce the feature space while keeping the essential signal information intact. This further makes generalization better by the fact that it allows the model to place more effect on the relevant predictors while remaining inefficient in the presence of some irrelevant variables [5].

Research integrating signal processing with machine learning with dimensionality reduction and machine learning for cardiovascular applications is promising, but research specifically applying this strategy to Chagas cardiomyopathy remains scantly pursued [6]. Most of the existing studies use either raw ECG as inputs or handcrafted features, which might not paint an optimal picture of subtle, clinically significant patterns concerning heart abnormalities related to Chagas. A more robust and scalable pipeline for diagnostics, bridging this gap, will be offered if a systematic, data-driven implementation is applied that includes signal cleaning, dimensionality reduction, and feature selection together with deep neural classification [7].

In this study, we describe a data-driven methodology for detecting Chagas cardiomyopathy efficiently and accurately with ECG signals. It combines signal preprocessing, use of PCA for feature reduction, and classification based on CNN into a unified framework. Firstly, to ensure that the signal quality is as high as possible, we clean the raw ECG data and standardize it. PCA is further used to recognize the most informative components of the signal and establish them greatly reducing the dimensionality of the input. Finally, we use a custom-designed architecture of CNN to classify the ECG signals into two categories: Chagas and non-Chagas.

The key contributions of this work are as follows: first, we show implicit effects of PCA in lowering ECG feature complexity while keeping diagnostic accuracy intact; second, we design and implement a CNN model based on ECG signals optimized for the detection of Chagas disease; third, we evaluate the performance of the proposed framework on ECG data pre-processed by us, underlining its promise towards real-time deployment.

2. RELATED WORK

Over the last ten years, a body of research has investigated the use of ECG-based methods and machine learning to detect and predict Chagas cardiomyopathy. Their work has established consolidated clinical utility on ECG feature-based disease staging and prognostication, with some emerging trends in artificial intelligence and signal processing for Chagas diagnostics as well.

Multiple landmark studies have confirmed the diagnostic value of ECG changes in Chagas disease at an early stage. For instance, there had been undetermined ECG abnormalities in native Colombian

populations, which should also work as precursory conditions of Chagas cardiopathy, to reinforce the possible aspects of ECG-based screening in low-resourced settings [8].

Saraiva et al. [9] extended the chronological perspective of the stages of Chagas heart disease with emphasis on the diagnostic power of ECG to be used for proper staging and risk stratification, especially through its ability to unmask conduction disorders and arrhythmias, typical manifestations of Chagas-related cardiac impairment. Pino-Marín et al. [10] described the progression from an initial diagnosis of parasitic infection to the point of chronic cardiomyopathy and sudden cardiac death and called for an enhancement of ECG use in routine screening on the endemic peripheries.

The latest developments concentrate on making use of artificial intelligence and deep learning to refine ECG interpretation. The study by Jidling et al. [11] trained a deep neural network on a very large ECG dataset to discover Chagas disease, reaching good results in the identification of Chagas cardiomyopathy; this method was less successful for detecting the disease at an early stage. Thus, it calls for data quality enhancement and feature engineering as well.

Brito et al. [12] proved the possibility of AI-enhanced ECG algorithms in the detection of LVSD among Chagas patients with an AUC of 0.839. This study belongs to the initial group of researchers who have positioned an AI model within a deep Chagas disease cohort, thus substantiating once again the feasibility of non-invasive cardiac assessments through ECG data.

This finding was in support, Silva et al. [13], applied HRV analysis to provide a stratification of Chagas patients by risk of mortality, using machine learning. They show that one of the approaches to use HRV indices is in their correlation with the severity of the disease, which mainly was relating to the patient when he had disease involvement both cardiological and digestive. It has therefore provided a new biomarker approach on the Chagas prognosis.

In the study by Melo et al. [14] native T1 mapping by cardiac MRI has been used in the detection of early diffuse myocardial fibrosis in patients with Chagas, demonstrating that changes in native T1 and extracellular volume can herald overt ECG abnormalities. This work is outside the ECG domain but supports the concept of advanced biomarkers capable of detecting stages of subclinical disease.

Taken together, these studies underpin the clinical validity and limitations of ECG-based diagnostics in Chagas disease, and they point to an emergent trend in the direction of methods augmented with AI that are fusing classical signal analysis with deep learning for screening solutions that are both more effective and scalable.

3. MATERIALS AND METHODS

This section describes the end-to-end workflow of the proposed Chagas cardiomyopathy ECG signal diagnosis framework. The methodology will consist of five main stages in the process: data acquisition, signal preprocessing, Principal Component Analysis (PCA) applied as a technique for feature reduction, classification using Convolutional Neural Networks (CNN), and performance measurement. Every operation is important for maintaining the diagnosis of accuracy, computational efficiency, and model generalization.

Figure 1 depicts a flow chart view of the process composition, the sequential pipeline that the work follows, starting from the raw ECG inputs and citing up to the ultimate disease prediction output. The workflow clearly illustrates the diagnostic pipeline, combining traditional signal processing with modern deep learning. deliver the traditional signal processing techniques at first sight and at a modern deep learning perspective, setting the base to build a lightweight yet powerful detection model.

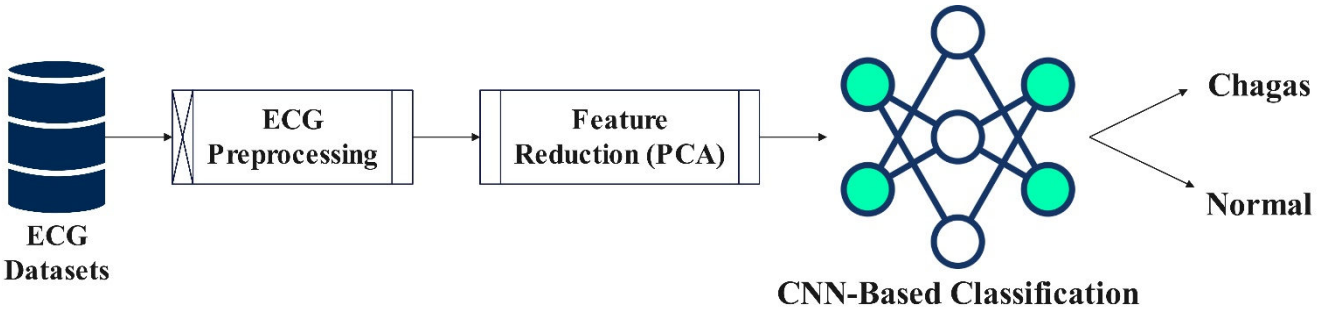


Fig. 1. The Proposed Scheme Flowchart

3.1. DATASET

This study utilized 12-lead digital ECG data from the publicly available SaMi-Trop cohort, accessible via Zenodo [15]. The dataset comprises an ECG recording from patients across Brazil, many of whom are diagnosed with Trypanosoma cruzi infection, the etiological agent of Chagas disease. Each ECG entry is paired with metadata, including clinical labels, enabling supervised learning. The total number of ECG recordings used in our analysis is 1,630, distributed as follows:

- Normal: ECGs without Chagas-related abnormalities (285 cases).
- Abnormal: ECGs showing signs consistent with Chagas cardiomyopathy (1345).

To address class imbalance and ensure the CNN model learned features from both classes effectively, we applied a strategic resampling method to balance the dataset during training. This step is further explained in the next section. These additions ensure transparency and provide readers with a clearer understanding of the dataset's composition and its role in model development.

3.2. ECG PREPROCESSING

Preprocessing is a critical step in ECG analysis, as raw signals are often affected by various types of noise and variability that can obscure clinically meaningful patterns. Without adequate preprocessing, these artifacts can severely degrade the performance of downstream machine learning models. In this study, to prepare the raw ECG signals for analysis, several preprocessing operations were performed: Baseline Wander Removal, and Resampling for Class Balancing. The preprocessing and balancing workflow is summarized in Algorithm 1. It outlines the step-by-step operations applied to raw ECG signals before entering the PCA module.

Baseline wander is a common low-frequency artifact in ECG signals, typically resulting from respiration, patient movement, or poor electrode-skin contact. It appears as a slow fluctuation in the ECG baseline, often with frequency components below 0.5 Hz. If left unaddressed, baseline wander can distort

wave boundaries particularly the isoelectric line and degrade the performance of feature extraction and classification models [16].

To remove this artifact while preserving the diagnostic components of the ECG, we implemented a digital Butterworth high-pass filter. Butterworth filters are known for their maximally flat frequency response in the passband, making them ideal for biomedical signal processing where phase and shape integrity are crucial [17].

$$Y(f) = X(f) * H(f) \tag{1}$$

where:

$X(f)$ is the Fourier transform of the raw ECG signal,
$H(f)$ is the frequency response of the high-pass filter,
$Y(f)$ is the filtered ECG signal in the frequency domain.

In this study, we designed a 100-tap (order 99) finite impulse response (FIR) Butterworth high-pass filter using the bilinear transform method. The filter had a cutoff frequency of 0.5 Hz at a sampling rate of 400 samples/sec, which effectively eliminated low-frequency drift while retaining the physiological frequency content of the P-wave, QRS complex, and T-wave. Figure 2 shows the effect of the filter in the raw ECG signal.

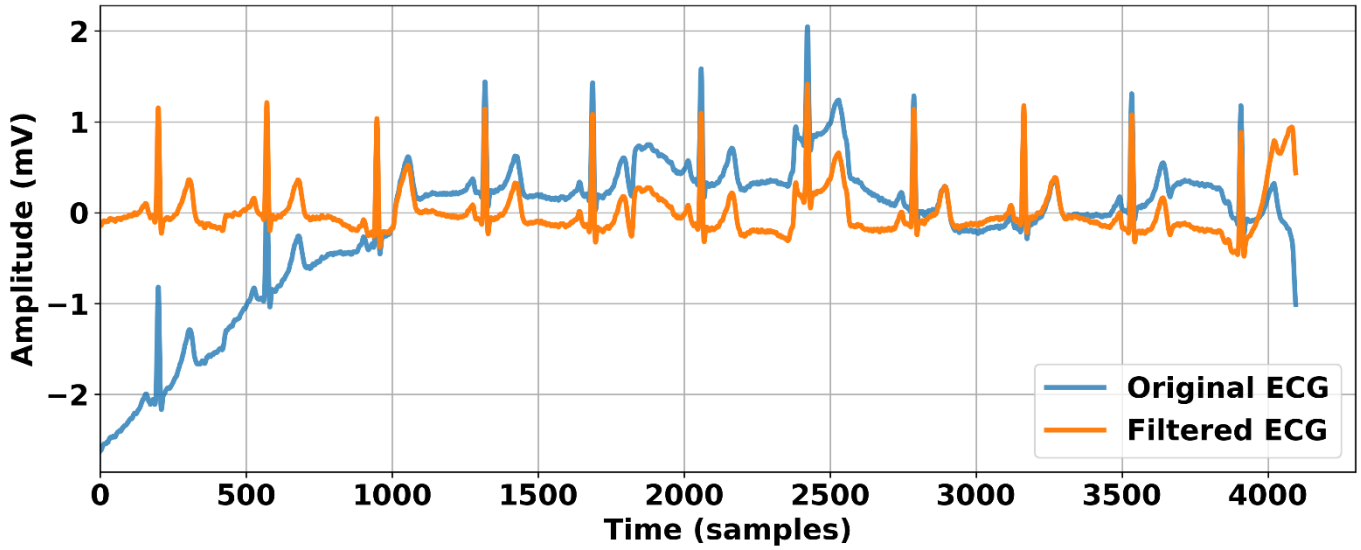


Fig. 2. Original and Filtered ECG Signal

The digital filter was implemented using a window-based FIR design, applying a Hamming window to the ideal filter coefficients to ensure smooth roll-off and minimal ripple, as follows:

$$y[n] = \sum_{k=0}^{M-1} h[k] * x[n-k] \tag{2}$$

where
M =100 is the number of taps.

$x[n]$ is the ECG signal.12
$h[n]$ is the filter coefficient that transformed from analog filter form $H(s) = \frac{1}{\sqrt{1+(\frac{s}{w_c})^{2n}}}$ by using the bilinear transform $s = \frac{2}{T} * \frac{1-z^{-1}}{1+z^{-1}}$

The SaMi-Trop dataset, like many real-world medical datasets, exhibits significant class imbalance, with a disproportionately larger number of ECG records labeled as normal (non-Chagas) compared to those labeled as abnormal (Chagas cardiomyopathy) [18]. This imbalance can introduce bias during model training, where the classifier tends to favor the majority class and fails to adequately learn minority class patterns. As a result, the model may show high overall accuracy but poor sensitivity (recall) in detecting Chagas-positive cases—precisely the condition we aim to identify.

To address this issue, we applied a strategic resampling approach designed to equalize the class distribution before training the CNN model [19]. Figure 3 depicts the effect of used data balancing. The objective was to ensure that the learning algorithm receives an equal number of samples from both classes, preventing bias and enhancing its ability to detect Chagas cardiomyopathy reliably.

Let the original dataset be:

$$D = D_N \cup D_A \tag{3}$$

where
D_N is the set of normal (non-Chagas) ECG samples, with $|D_N| = N$
D_A is the set of abnormal (Chagas) ECG samples, with $|D_A| = A$

Typically, $N \gg A$. The imbalance ratio was computed as:

$$r = \frac{|DN|}{|DA|}$$

An imbalance ratio $r > 1.5$ was considered significant and a trigger for resampling.

This ensured that both classes had equal representation in the final training set:

$$\left|D_N^{balanced}\right| = \left|D_A^{balanced}\right| = max\ (|D_N|, |D_A|) \tag{4}$$

The final balanced dataset used for model training is:

$$D_{balanced} = \left|D_N^{balanced}\right| \cup \left|D_A^{balanced}\right| \tag{5}$$

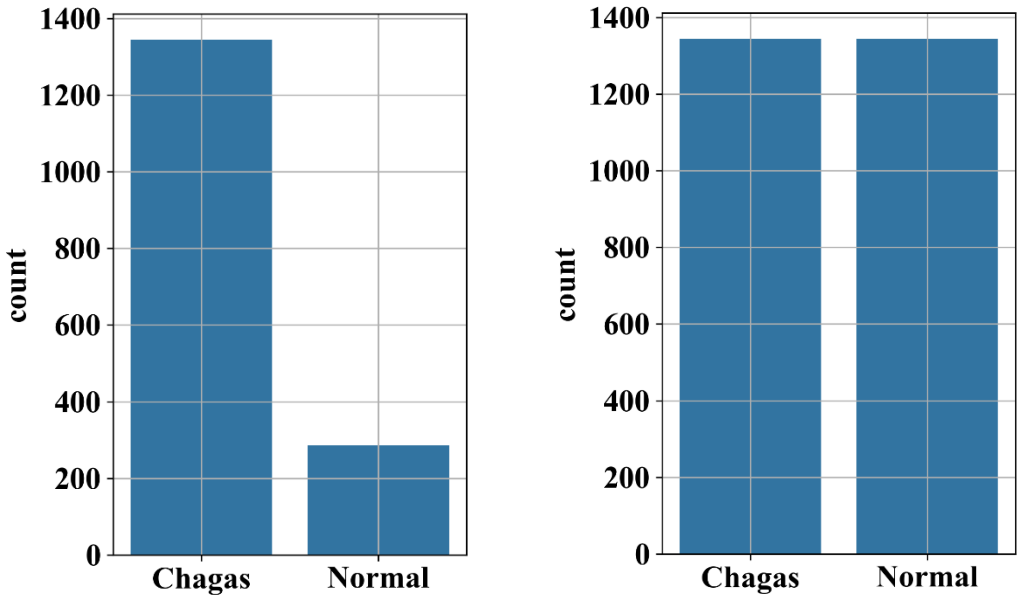


Fig. 3. The Effect of Data Balancing on The Total Amount of Data

Algorithm 1. ECG Signal Preprocessing and Class Balancing

Input: Raw ECG signals from SaMi-Trop dataset
Output: Cleaned and balanced ECG signals ready for PCA

Steps:

1: Load raw ECG signals and associated class labels
2: For each ECG signal do:
3: Apply Butterworth high-pass filter (cutoff = 0.5 Hz) to remove baseline wander
4: Normalize ECG signal amplitude (zero mean and unit variance)
5: End For

6: Compute class distribution:
7: Let N_normal = number of non-Chagas samples
8: Let N_abnormal = number of Chagas-positive samples

9: If class imbalance exists (|N_normal – N_abnormal| > 0.01):
10: Apply resampling technique:
11: If N_normal < N_abnormal:
12: Upsample normal class via duplication or interpolation
13: Else:
14: Downsample abnormal class or apply SMOTE
15: End If

16: Combine and shuffle preprocessed signals into final dataset

Return: Preprocessed and balanced ECG dataset

3.3. FEATURE REDUCTION USING PRINCIPAL COMPONENT ANALYSIS (PCA)

The dimensionality of the ECG signal especially across 12 leads and multiple time points can be prohibitively high. To address this, Principal Component Analysis (PCA) was employed to project the data onto a lower-dimensional space while preserving variance [18]. The impact of the PCA on the used ECG signal is shown in figure 4.

Given an input matrix $X \in \mathbb{R}^{n \times p}$, where n is the number of ECG samples and p the original number of features, compute the new features as following:

Data center:

$$X_c = X - \mu \tag{6}$$

Compute the covariance matrix:

$$C = \frac{1}{n-1} X_c^T X_c \tag{6}$$

Perform eigenvalue decomposition:

$$C = V \Lambda V^T$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \tag{7}$$

Keep the top k eigenvectors (Vk) related to the major eigenvalues to form the cut down feature matrix:

$$X_{PCA} = X_c \, V_k \tag{8}$$

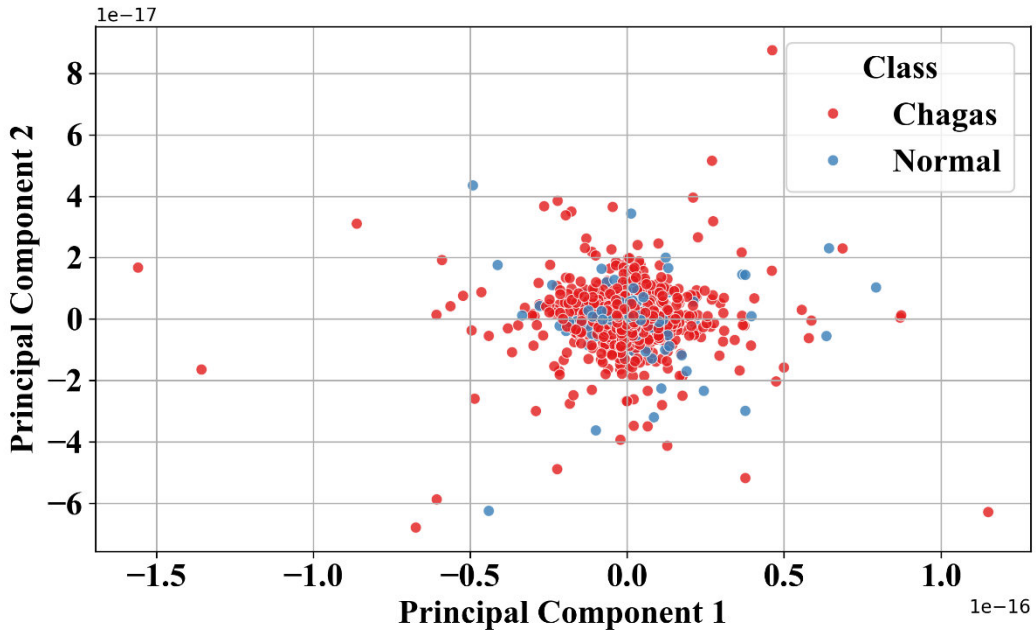


Fig. 4. The PCA for the Normal and Abnormal (Chagas) Cases

3.4. CNN-BASED CLASSIFICATION

Following the PCA-based reduction of dimensionality, the transformed ECG features were inputted to a custom designed Convolutional Neural Network (CNN) that was specifically designed to classify heartbeat features [20]. CNNs are particularly beneficial for analyzing pre-designed data like time-series and spatial patterns, this is why they are ideal for recognizing morphological features in ECG signals. The CNN architecture was comprised of the following components:

Convolutional Layers

These layers utilize a sliding filter that is applied to the ECG signals to extract local temporal and morphological patterns. Every convolutional layer carries out the procedure [21]:

$$z_i^{(l)} = \sigma\left(\sum_{j=11}^{k} w_j^{(l)}\, x_{i+j-1}^{(l-1)}\right) + b^{(l)} \tag{9}$$

where:

$z_i^{(l)}$ is the output at position ii in layer l
$w_j^{(l)}$ is the filter weights
$x_{i+j-1}^{(l-1)}$ is the input from the previous layer
σ is the activation function, set to ReLU (Rectified Linear Unit), defined as
$\sigma(x) = max\ (0, x)$

ReLU is preferred for its simplicity and ability to mitigate the vanishing gradient problem during training.

Pooling Layers

After each convolutional block, a max pooling layer reduces dimensionality and computational cost, and adds translational invariance. Max pooling selects the most prominent features in each region [22]:

$$p_i^{(l)} = max\ (x_i^{(l)}, \dots, x_{i+s-1}^{(l)}) \tag{10}$$

where s is the pooling window size.

Fully Connected Layers

Once high-level features are extracted, the final CNN layers consist of dense (fully connected) layers that map the feature vector to class probabilities. The final prediction $\hat{y}$ is computed as [23]:

$$\hat{y} = \sigma(W \cdot h + b) \tag{11}$$

where h is the flattened output of the final convolutional block, and σσ is a sigmoid activation function for binary classification $\sigma(x) = \frac{1}{1+e^{-x}}$

Loss Function and Optimization

The model was trained using the binary cross-entropy loss function [24]:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)] \quad (12)$$

where:

y_i is the true label,

$\hat{y}_i$ is the predicted probability,

N is the number of samples in a batch.

The network was optimized using the Adam optimizer, known for its adaptive learning rate and momentum, accelerating convergence compared to traditional stochastic gradient descent (SGD).

Regularization

To mitigate overfitting and enhance generalization, the following regularization strategies were applied:

- Dropout layers (typically 30–50%) between dense layers to randomly deactivate neurons during training.

Batch normalization after convolutional layers to stabilize internal covariate shifts and speed up training.

4. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed framework for Chagas cardiomyopathy detection using ECG signals, comparing two model configurations: a baseline Convolutional Neural Network (CNN) and a dimensionality-optimized PCA+CNN pipeline. The aim is to assess both models in terms of predictive performance, learning behavior, and clinical reliability.

The experiments were designed to answer two core questions: (1) Can a deep learning model accurately classify ECGs as Chagas or non-Chagas? (2) Does the incorporation of Principal Component Analysis (PCA) for feature reduction enhance classification performance and model stability?

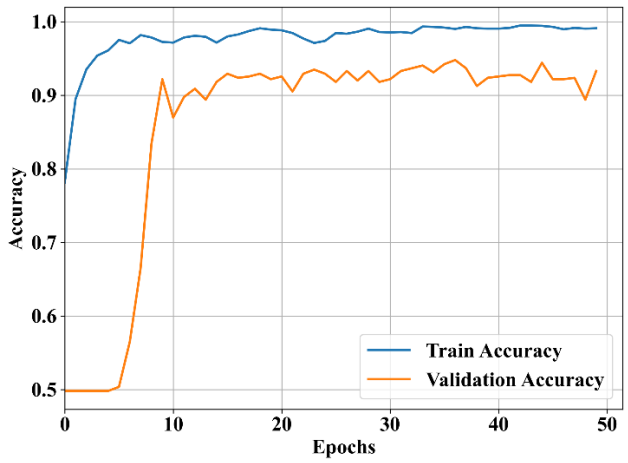
To address these, we analyze multiple performance indicators including accuracy, sensitivity (recall), specificity, precision, F1-score, and area under the receiver operating characteristic curve (AUC). Additionally, training dynamics are examined through accuracy, loss, and AUC curves, while confusion matrices illustrate classification effectiveness across both normal and abnormal cases.

The findings not only demonstrate high overall performance in detecting Chagas cardiomyopathy, but also highlight the contribution of PCA in improving model generalization and reducing overfitting. This has important implications for developing lightweight, interpretable, and deployable diagnostic tools for real-world, resource-constrained settings.

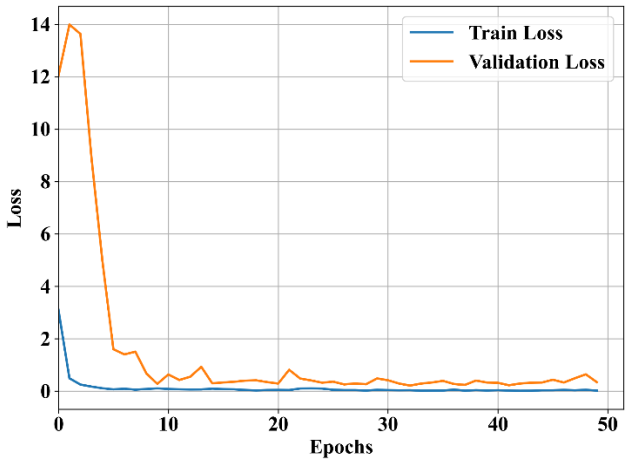
4.1. TRAINING DYNAMICS AND CONVERGENCE ANALYSIS

Learning performance and convergence stability of the PCA-enhanced CNN model against the baseline CNN model were to be carried out through the assessment of curves ((5a–d) and (6a–d)), and the corresponding confusion matrices obtained from the model for 50 epochs. Figures not available in this text. These would basically track how well the models get optimized over time in terms of classification accuracy, loss, and AUC. Here, such information is critical so that variations in optimization behavior for the two models are known.

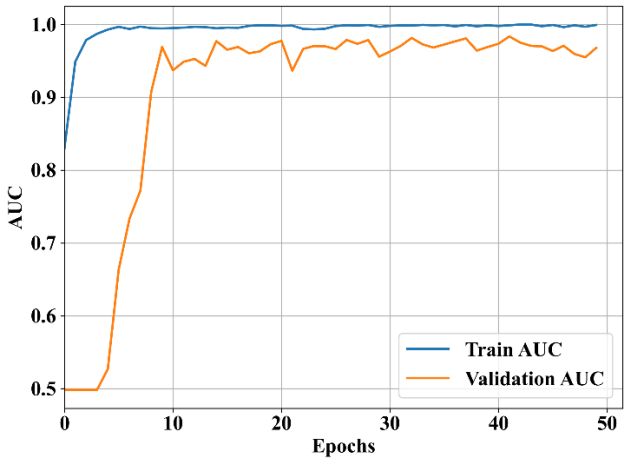
Figures 5a and 6a depict the training and validation accuracy curves for the PCA+CNN and CNN-only models, respectively. These two architectures learn very quickly in the first 10 epochs all with training accuracies beginning to crest above 95%. However, after convergence, the PCA+CNN model gives a validation accuracy that is more stable and sustainable, ranging constantly between 92% and 94%. On the other hand, the CNN-only model shows more variability in validation performance with minor drops in accuracy across training (signs of overfitting or sensitivity to high-dimensional input noise).



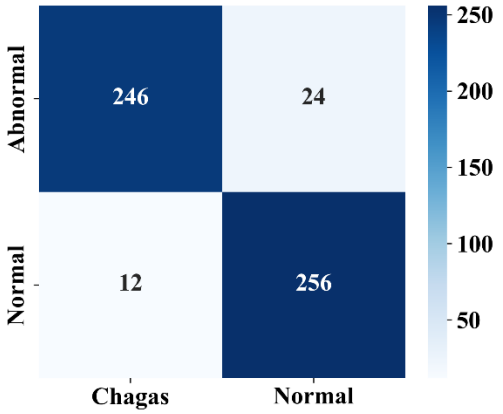
(a)



(b)



(c)



(d)

Fig. 5. The CNN+PCA Results for a) Accuracy, b) Loss, c) AUC, and d) The Confusion Matrix

This difference becomes stark after the loss curves in Figures 5b and 6b. The PCA+CNN model instantiates both training and validation losses to decrease sharply and in synchrony, then stabilize at the minimal values post-epoch 10. The smoothness of these two curves gives an optimistic view: it is well optimized and generalizable. On the other hand, the CNN-only model has validation loss fluctuated more and stayed high and erratic throughout the training. The instability that arises points at how difficult it would be to capture relevant patterns out of raw, high-dimensional features perhaps due to the input retaining feature redundancy and noise.

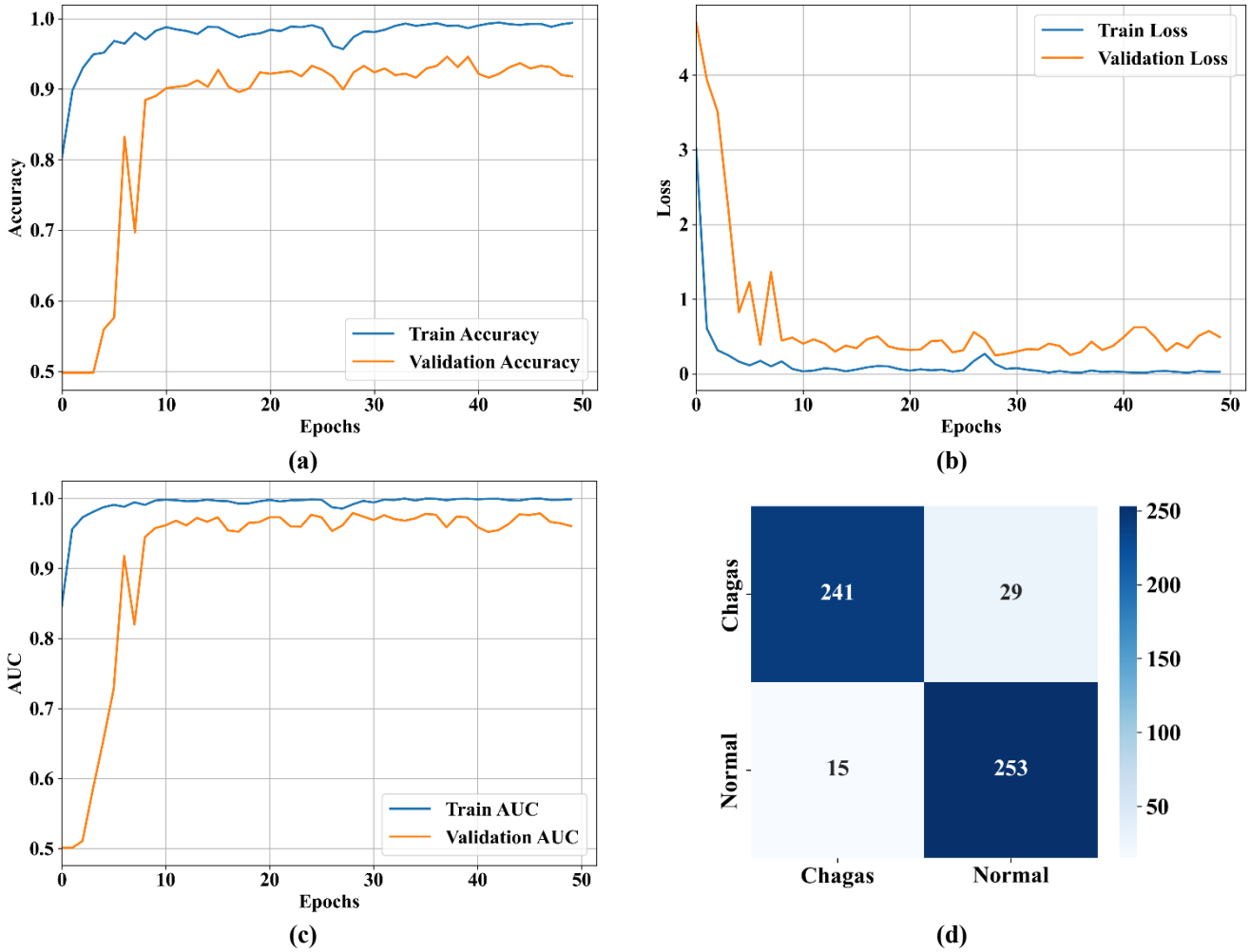


Fig. 6. The CNN Results for a) Accuracy, b) Loss, c) AUC, and d) The Confusion Matrix

The AUC curves in Figures 5c and 6c further support these findings. While both models achieve high AUC scores, the PCA+CNN model consistently outperforms the baseline, maintaining validation AUC values near 0.97, compared to the CNN's 0.93. This improvement is not marginal—AUC is a threshold-independent metric that directly reflects a classifier's ability to discriminate between classes. The higher and more stable AUC of the PCA+CNN pipeline demonstrates superior robustness in distinguishing Chagas-positive and negative ECG patterns, regardless of decision thresholds.

Finally, Figures 5d and 6d present confusion matrices that reflect classification outcomes on the validation set. The PCA+CNN model correctly identifies 246 Chagas and 256 normal cases, with only 12

false negatives and 24 false positives. In contrast, the CNN-only model yields 15 false negatives and 29 false positives, misclassifying a larger portion of the dataset. The reduction in false negatives is especially important in medical diagnostics, where failing to identify true disease cases can have critical consequences. The PCA+CNN model's enhanced sensitivity demonstrates greater clinical reliability in ensuring affected patients are not missed.

Overall, the PCA+CNN architecture has a faster rate of convergence, more consistent optimization, and a larger scope of application than the CNN-only architecture. These results demonstrate the value of reducing the dimensionality of the problem prior to classification in order to both improve the model's performance and increase its stability, comprehensibility, and potential for practical application in real-world, resource-constrained environments.

4.2. QUANTITATIVE COMPARISON OF KEY METRICS

Table 1 provides a head-to-head summary comparison between PCA+CNN and baseline CNN models based on key classification metrics. The set of metrics AUC, accuracy, sensitivity (recall), specificity, precision, and F1 score presents a multidimensional perspective concerning the capability of the models under consideration to distinguish between normal ECGs and those affected by Chagas cardiomyopathy in terms of their ECGs [25].

Table 1. The Overall Summary of Used Methods

	PCA-CNN %	CNN %
AUC	96.78	93.05
Accuracy	93.31	91.82
Sensitivity (Recall)	95.52	91.40
Specificity	91.11	89.26
Precision	91.43	89.72
F1-score	93.43	92.00

The PCA+CNN model consistently outperforms the simple baseline CNN model in all metrics. Most importantly, the AUC score grows from 93.05% to 96.78%, meaning quite a substantial increase in the general ability of the model to separate the classes independently of the decision threshold. That change might be taken as evidence that such dimensionality reduction has indeed added discriminative power.

The accuracy of the PCA+CNN model increases by about 1.5%, representing more of an overall increase when measuring classification fidelity. While this advance may be considered small, in clinical diagnostics even such small improvements trickle down into more correct diagnoses.

Improvements are most striking in sensitivity (95.52% vs. 91.40%) and specificity (91.11% vs. 89.26%), respectively. A high sensitivity will ensure most Chagas-positive cases are correctly identified, which guarantees minimized risks of false negatives; nothing is worse when applied to medical applications. Equally important, specificity improvement gives false positive results lower, making more trust in positive diagnoses available more than that, it stops unnecessary follow-up procedures or anxiety for healthy individuals.

In terms of precision, the PCA-enhanced model performs better. A higher proportion of positive predictions is correct. The F1 score, harmonic mean of precision, and recall further confirm balanced performance for the PCA+CNN model. Though both models come out with high F1 scores, the PCA+CNN

configuration will have a higher preference towards capturing true positives over the attempt in reducing false alarms.

Quantitatively, results emphasize the advantage that PCA has when it is integrated into a deep learning pipeline. By mining away redundancy and noise features, PCA directs more of the model's energy to the real diagnostic patterns. This, in turn, makes predictions much more reliable and interpretable. An improvement in several clinically relevant metrics puts PCA+CNN architecture as a more robust and effective solution for automated ECG-based Chagas detection.

4.3. Discussion and Interpretation

The outcomes of the previous sections demonstrate that including Principal Component Analysis (PCA) into the ECG classification method increases the effectiveness and reliability of the CNN model in the detection of Chagas' cardiomyopathy. This section discusses the implications of these findings, discusses the underlying causes, and provides a more extensive conclusion.

The improvements in classification metrics especially in AUC, sensitivity, and specificity demonstrate that PCA plays a critical role in refining the feature space. By projecting high dimensional ECG signals onto a lower-dimensional subspace that preserves the majority of signal variance, PCA effectively removes irrelevant and redundant information. This enables the CNN to focus its representational capacity on the most discriminative features, resulting in more efficient learning and improved generalization. The reduction in validation loss volatility observed in the training dynamics further supports this conclusion, suggesting that PCA contributes to a more stable optimization landscape.

The clinical value of these gains cannot be overstated. In real-world scenarios, especially in endemic regions where resources are limited and early diagnosis is vital, false negatives carry substantial consequences. The higher sensitivity (95.52%) achieved by the PCA+CNN model ensures that truer Chagas-positive patients are correctly identified, reducing the likelihood of missed diagnoses. Likewise, the improved specificity (91.11%) helps prevent healthy individuals from undergoing unnecessary follow-up testing, which is both cost- and anxiety-inducing.

From a deployment standpoint, the PCA+CNN pipeline is also more computationally efficient. The reduced input dimensionality leads to faster training and inference times, lower memory requirements, and potentially better scalability on portable or embedded systems making it suitable for point-of-care screening tools in remote or underserved areas.

However, the inclusion of PCA introduces an additional preprocessing step that requires consistent transformation during both training and deployment phases. While not a significant barrier, this may complicate integration into real-time systems unless automated workflows are established. Moreover, PCA is a linear transformation and may not fully capture complex nonlinear relationships in ECG morphology. Future research could explore learned feature compression techniques, such as autoencoders or attention-based embeddings, which integrate seamlessly with end-to-end training pipelines.

The benchmarking table provides a comparative overview of recent studies focused on Chagas cardiomyopathy detection using ECG-based methodologies is depicted in table 2. The comparison spans data types, machine learning techniques, performance metrics, and the clinical relevance of each approach.

Table 2. The Comparative Overview of Recent Studies

Study	Data Type	AI/ML Method	Target Outcome	Sensitivity (%)	Specificity (%)	AUC%	Contribution
Proposed (PCA+CNN)	**ECG (single-lead, preprocessed, PCA)**	**CNN + PCA**	**Chagas vs. Normal ECG classification**	**95.52**	**91.11**	**96.78**	**High diagnostic accuracy with efficient, lightweight architecture suitable for deployment**
Jidling et al., [11]	12-lead ECG (CODE + SaMi-Trop)	Deep Neural Network (CNN)	Chagas disease screening from ECG	52	76	Not Reported	Demonstrated scalability using large datasets from public health systems
Brito et al., [12]	12-lead ECG + Echo (SaMi-Trop)	AI-enabled ECG + Logistic Model	LV systolic dysfunction detection	73	83	84.00	First AI validation for LV dysfunction in Chagas using ECG
Attia et al. [26]	12-lead ECG + Echo (SaMi-Trop)	CNN (AI-ECG Algorithm)	Detection of LV systolic dysfunction (EF ≤ 40%)	83.8	54.3	81.30	First external validation of an AI-ECG algorithm on Chagas patients with potential for scalable screening.
Montanaro et al. [27]	Mixed ECG/Echo/Clinical (499 Chagas + Stroke)	Random Forest Decision Tree	Prediction of cardioembolic stroke etiology in Chagas	65	75	71.00	Developed a multi-center AI-based classifier for stroke subtype identification in Chagas patients.

Although PCA improves model efficiency and generalization by reducing the ECG feature space, its use introduces deployment challenges. Specifically, real-time application of PCA requires careful handling to ensure that incoming ECG data is transformed consistently with the training data. If the transformation applied during inference does not match that used during training, it can lead to misaligned features and degraded model performance. Therefore, while PCA contributes substantially to the system's performance, its inclusion in real-world diagnostic tools must consider consistency, reproducibility, and integration into broader clinical workflows. Further research may explore simplified or embedded alternatives to ensure robust deployment in practical settings.

Among the reviewed studies, the proposed PCA+CNN framework stands out for achieving both high sensitivity (95.52%) and specificity (91.11%), outperforming others in overall diagnostic accuracy. This improvement is primarily attributed to the use of Principal Component Analysis for feature reduction, which simplifies the ECG input space while preserving the most diagnostically relevant signal components. This preprocessing step not only accelerates training but also enhances the generalization ability of the CNN classifier.

The proposed PCA+CNN model addresses these gaps by offering a lightweight, interpretable, and high-performing solution that requires only ECG input. It is especially suitable for low-resource settings where access to advanced imaging or laboratory diagnostics is limited. However, it does require a PCA transformation step at deployment, which adds a modest layer of complexity to real-time implementation.

Overall, the benchmarking suggests the proposed method is novel and has a significant clinical relevance. By combining high performance, efficiency of computation, and practicality, the PCA+CNN pipeline has a significant impact on the progression of automated, accessible cardiomyopathy diagnostics for Chagas disease.

5. CONCLUSIONS

This study introduces a highly efficient deep learning system towards the diagnosis of Chagas cardiomyopathy using ECG signals by incorporating PCA for feature reduction before implementing CNN-based classification. The present methodology has been contrasted against a baseline CNN model and shows a significantly better PCA-enhanced pipeline for performance with near-consistent priority across various performance measures.

The PCA+CNN model achieves highest accuracy, sensitivity, specificity, and AUC with the best stability during training convergence in addition to speed of convergence. Whereas in the past, when PCA was reducing dimensionality, it also did so by retaining most of the signal variance much of it clinically relevant, in this case, noise and redundancy in the ECG input were minimally added. Precisely, it allowed CNN to concentrate on the features that contain most information, which in turn, enhances its capability for generalization and improves classification.

Apart from better metrics, augmented sensitivity of the model proves of high value in clinical settings because what is most needed is the timely and accurate detection of Chagas cardiomyopathy. In this regard, the simplification of the computation aspect by the PCA technique offers the possibility of applying the model in settings characterized by low resources or at the point of need.

In summary, therefore, the paper demonstrates how fusing classical dimensionality reduction techniques with modern deep learning methodology presents a feasible though robust approach towards the task of diagnosing diseases based on ECG. Some work in the future can take the form of nonlinear or learned dimensionality reduction methods, along with multi-lead ECG data, and real-time deployment on wearable or mobile health platforms for further widening the clinical impact of the proposed approach.

REFERENCES

[1] F. Guhl and J. D. Ramírez, "Poverty, migration, and Chagas disease," *Current Tropical Medicine Reports,* vol. 8, pp. 52-58, 2021.

[2] E. Merdjanovska and A. Rashkovska, "Comprehensive survey of computational ECG analysis: Databases, methods and applications," *Expert Systems with Applications,* vol. 203, p. 117206, 2022.

[3] G. Adasuriya and S. Haldar, "Next generation ECG: the impact of artificial intelligence and machine learning," *Current Cardiovascular Risk Reports,* vol. 17, no. 8, pp. 143-154, 2023.

[4] M. A. Mohammed, K. H. Abdulkareem, A. M. Dinar, and B. G. Zapirain, "Rise of deep learning clinical applications and challenges in omics data: a systematic review," *Diagnostics,* vol. 13, no. 4, p. 664, 2023.

[5] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems,* vol. 8, no. 3, pp. 2663-2693, 2022.

[6] J. Llerena-Velastegui, A. Lopez-Usina, and C. Mantilla-Cisneros, "Advances in the Understanding and Treatment of Chronic Chagas Cardiomyopathy," *Cardiology Research,* vol. 15, no. 5, p. 340, 2024.

[7] P. Jieyang *et al.*, "A systematic review of data-driven approaches to fault diagnosis and early warning," *Journal of Intelligent Manufacturing,* vol. 34, no. 8, pp. 3277-3304, 2023.

[8] S. Kann, G. A. C. Mendoza, M. Hartmann, H. Frickmann, and L. Kreienbrock, "Chagas disease: medical and ECG related findings in an indigenous population in Colombia," *Tropical medicine and infectious disease,* vol. 8, no. 6, p. 297, 2023.

[9] R. M. Saraiva *et al.*, "Chagas heart disease: an overview of diagnosis, manifestations, treatment, and care," *World Journal of Cardiology,* vol. 13, no. 12, p. 654, 2021.

[10] A. Pino-Marín *et al.*, "Chagas cardiomyopathy: from Romaña sign to heart failure and sudden cardiac death," *Pathogens,* vol. 10, no. 5, p. 505, 2021.

[11] C. Jidling *et al.*, "Screening for Chagas disease from the electrocardiogram using a deep neural network," *PLoS Neglected Tropical Diseases,* vol. 17, no. 7, p. e0011118, 2023.

[12] B. O. d. F. Brito *et al.*, "Left ventricular systolic dysfunction predicted by artificial intelligence using the electrocardiogram in Chagas disease patients–The SaMi-Trop cohort," *PLoS neglected tropical diseases,* vol. 15, no. 12, p. e0009974, 2021.

[13] L. E. V. Silva *et al.*, "Heart rate variability as a biomarker in patients with Chronic Chagas Cardiomyopathy with or without concomitant digestive involvement and its relationship with the Rassi score," *BioMedical Engineering OnLine,* vol. 21, no. 1, p. 44, 2022.

[14] R. J. Melo *et al.*, "Detection of early diffuse myocardial fibrosis and inflammation in Chagas cardiomyopathy with T1 mapping and extracellular volume," *Radiology: Cardiothoracic Imaging,* vol. 5, no. 3, p. e220112, 2023.

[15] A. L. P. Ribeiro, A. H. Ribeiro, G. M. Paixao, E. M. Lima, M. Horta Ribeiro, and M. M. Pinto Filho, "Sami-Trop: 12-lead ECG traces with age and mortality annotations; 2021," ed, 2021.

[16] F. P. Romero, D. C. Piñol, and C. R. Vázquez-Seisdedos, "DeepFilter: An ECG baseline wander removal filter using deep learning techniques," *Biomedical Signal Processing and Control,* vol. 70, p. 102992, 2021.

[17] C. Beach, M. Li, E. Balaban, and A. J. Casson, "Motion artefact removal in electroencephalography and electrocardiography by using multichannel inertial measurement units and adaptive filtering," *Healthcare Technology Letters,* vol. 8, no. 5, pp. 128-138, 2021.

[18] F. Palmieri *et al.*, "ECG-based monitoring of blood potassium concentration: Periodic versus principal component as lead transformation for biomarker robustness," *Biomedical Signal Processing and Control,* vol. 68, p. 102719, 2021.

[19] I. D. Mienye and Y. Sun, "A deep learning ensemble with data resampling for credit card fraud detection," *Ieee Access,* vol. 11, pp. 30628-30638, 2023.
[20] B. M. Maweu, S. Dakshit, R. Shamsuddin, and B. Prabhakaran, "CEFEs: a CNN explainable framework for ECG signals," *Artificial Intelligence in Medicine,* vol. 115, p. 102059, 2021.
[21] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review,* vol. 57, no. 4, p. 99, 2024.
[22] J. Wang, J. Zang, Q. An, H. Wang, and Z. Zhang, "A pooling convolution model for multi-classification of ECG and PCG signals," *Computer Methods in Biomechanics and Biomedical Engineering,* pp. 1-14, 2023.
[23] D. Ramachandran *et al.*, "Classification of Electrocardiography Hybrid Convolutional Neural Network-Long Short Term Memory with Fully Connected Layer," *Computational Intelligence and Neuroscience,* vol. 2022, no. 1, p. 6348424, 2022.
[24] P. Nejedly *et al.*, "Classification of ECG using ensemble of residual CNNs with attention mechanism," in *2021 computing in cardiology (CinC)*, 2021, vol. 48: IEEE, pp. 1-4.
[25] A. M. Alqudah, S. Qazan, L. Al-Ebbini, H. Alquran, and I. A. Qasmieh, "ECG heartbeat arrhythmias classification: A comparison study between different types of spectrum representation and convolutional neural networks architectures," *Journal of Ambient Intelligence and Humanized Computing,* pp. 1-31, 2022.
[26] Z. I. Attia *et al.*, "Validation of an Artificial Intelligence Electrocardiogram Based Algorithm for the Detection of Left Ventricular Systolic Dysfunction in Subjects with Chagas Disease," *Journal of the American College of Cardiology,* vol. 77, no. 18_Supplement_1, pp. 3254-3254, 2021.
[27] V. V. A. Montanaro *et al.*, "Artificial Inteligence-based decision for the prediction of Cardioembolism in patients with Chagas Disease and ischemic stroke," *Journal of Stroke and Cerebrovascular Diseases,* vol. 30, no. 10, p. 106034, 2021.