Congestive Heart Failure Detection in ECG Using LSTM and CNN

Luiz Ribeiro[1,2][0009-0002-2452-7264], Nathan Guerreiro[1,2][0009-0007-2879-2248], Mohamed Khalil Chaabani[1,3][0009-0006-1813-6944], Luiz E. Luiz[1,2][0000-0001-9305-091X], André Eugenio Lazzaretti[2][0000-0003-1861-3369], and João Paulo Teixeira[1][0000-0002-6679-5702]

[1] CeDRI, SusTEC, Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal
{a58865, a61448, a60546}@alunos.ipb.pt
{luiz.luiz, joaopt}@ipb.pt
https://cedri.ipb.pt/

[2] Graduate Program in Electrical and Computer Engineering, Federal University of Technology - Paraná, Brazil
{luizotavioribeiro, nathanguerreiro, lluiz}@alunos.utfpr.edu.br
lazzaretti@utfpr.edu.br
https://www.utfpr.edu.br/

[3] Private Higher Polytechnic Institute, Private University of Tunis, 32 Bis, 1002 Avenue Kheired-dine Pacha, Tunis, Tunisia
khalil.chaabani@yahoo.com

Abstract. Congestive Heart Failure (CHF) is a chronic condition in which the heart does not pump blood efficiently. This pathology causes fatigue, dyspnea, oedema, nausea, and memory problems, affecting patients' quality of life. The causes include coronary artery disease, cardiomyopathy, arterial hypertension, and myocarditis. The diagnosis is usually based on the patient's medical history, physical exams, echocardiogram, electrocardiogram, and other methods. Aiming to improve diagnostic tools, this study proposes an artificial intelligence model based on deep learning to classify pathological ECG signals indicative of CHF. The selected models were LSTM and CNN. The training was conducted using a personalised dataset created from the public databases BIDMC Congestive Heart Failure and PTB Diagnostic ECG from Physionet. ECG data from 28 individuals aged 22 to 71 were selected, including 14 with severe CHF (NYHA class 3 and 4) and 14 control samples without ECG abnormalities. The database architecture was designed so that the input to the neural networks was raw ECG signals without filtering or feature extraction. The results showed an accuracy of 98.21% for the CNN model and 92.26% for the LSTM model.

Keywords: Digital Health · Deep Learning · ECG Classification

1 Introduction

Cardiac pathologies correspond to approximately 31% of global deaths, standing out as one of the leading causes of mortality, with 17.9 million fatalities in

2016 [14]. Congestive Heart Failure (CHF) is a chronic condition that compromises the heart's ability to pump blood efficiently, leading to reduced blood flow to organs and tissues. The main causes of CHF include coronary artery disease (CAD) and conditions such as infections and chemotherapy treatments. The most common symptoms include dyspnea, persistent fatigue, swelling, nausea, and confusion.

The diagnosis of CHF involves a variety of tests, with the electrocardiogram (ECG) being one of the most widely used methods for assessing cardiac function. The ECG records electrical variations during the cardiac cycle and can expose abnormalities associated with CHF, such as prolongation of the QRS complex, inversion of the T wave, and alterations in the PR and ST segments.

Given the complexity of diagnosing CHF, implementing artificial intelligence models, especially deep learning techniques, emerges as a promising helping tool. These models can process large volumes of data, aiding in detecting characteristics associated with CHF in ECG signals.

Therefore, this proposed work intends to compare LSTM and CNN models to accurately detect CHF, reinforcing the capability of deep learning models applied in digital health systems. To this end, the paper is structured as follows: an introductory section, followed by a brief review of similar works, an explanation of the methodology with corresponding results and discussions, and finally, a comparison of the results with related studies.

2 Related Work

The concept of artificial intelligence is broad, and various techniques can be employed for pathologies classification using biosignals. In the work developed by [3], Convolutional Neural Networks (CNN) are employed to classify different types of myocardial infarction. Similarly, [9] utilizes Long Short-Term Memory (LSTM) networks to classify chronic laryngitis while [16] applies Artificial Neural Networks (ANN) to classify electroencephalogram (EEG) signals. In the work by [11], LSTM networks combined with an Autoencoder (AE) are used to classify cardiac arrhythmias.

Further research includes the study by [6], which focuses on classifying ECG signal components such as the QRS complex, P waves, and T waves. Meanwhile, [1] highlights using ANN to classify cardiac arrhythmias. In the study carried out by [10], Generative Adversarial Networks (GAN) are employed to generate synthetic ECG signals to improve training data, and [4] uses a Multilayer Perceptron (MLP) to detect atrial fibrillation by extracting Jitter and Shimmer parameters.

There are multiple approaches, from model selection to how the biosignals are processed to serve as input for training. Therefore, to delimit the scope of this paper only works that employed CNN and/or LSTM for CHF classification using ECG signals will be covered.

The work by [13] proposes a CNN-LSTM architecture for CHF diagnosis using RR intervals as input. The CNN extracts features through convolution and pooling, while the LSTM captures temporal dependencies. An attention mechanism is applied to the LSTM's output to assess the importance of each sequence, followed by binary CHF classification. This method achieved 99.52% accuracy, 99.31% sensitivity, 99.28% specificity, a 98.94% F-score, and a 99.9% AUC.

The research conducted by [7] proposes using recurrent neural networks based on LSTM for CHF classification. The methodology involves filtering the signals using the Wavelet transform, partitioning the signals into small intervals ranging from 1 to 4 seconds. A total of 24 LSTM models were trained with different hyperparameters, and the best classification result was an accuracy of 99.86%, sensitivity of 99.85%, specificity of 99.85%, precision of 99.87%, and an F1-score of 99.86%.

The work [12] employs a 1D CNN structure with a Gabor filter to classify ECG signals into four classes: healthy, CHF, myocardial infarction, and coronary artery disease. The input to the network consists of 2-second-long ECG signals, which first pass through the Gabor filter and are then fed into the CNN. The network classification accuracy reached 98.74%.

In the work [15], a 1D CNN methodology is used for CHF classification. The input data consists of heartbeats, meaning the sample size corresponds to one ECG cycle rather than a predefined time window. To split the heartbeats, the signal was extracted around the R peak: 235 ms before the peak and 390 ms after. With this methodology, an accuracy of 97.8% is achieved, based exclusively on ECG sample classification. This work also highlights the higher efficiency of this method compared to HRV (Heart Rate Variability) feature extraction.

None of the referenced studies directly compare the performance of different neural networks using the same architecture and dataset. Each focuses on a single model, like CNN, LSTM, or a combination, without comparing them under identical conditions. This work addresses that gap by analyzing two networks on the same dataset and architecture to determine which is more effective for CHF classification. The comparison offers insights into the strengths and weaknesses of each model, providing a clearer understanding of which performs better for this task.

3 Methodology

This section overviews the approaches and models employed to detect CHF from ECG signals. The methodology covers the construction of the dataset, the architecture of the LSTM and CNN models, and the training and evaluation process.

3.1 Dataset

The dataset was constructed using data from the BIDMC Congestive Heart Failure Database [2] and the PTB Diagnostic ECG Database [5]. Both datasets are from PhysioNet [8].

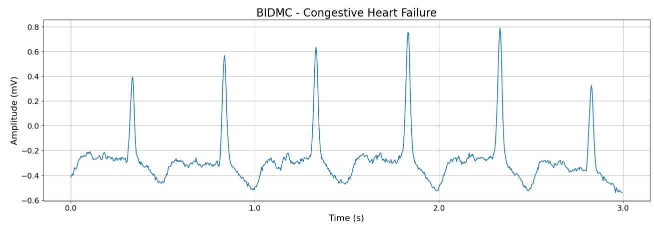
The BIDMC database provides high-quality ECG signals from patients with congestive heart failure, while the PTB database offers control signals. Due to the absence of control signals in the BIDMC dataset and differences in sample rates between the two databases, a custom dataset was created. This custom dataset includes 840 samples from 28 individuals (30 samples per individual). Among these 28 individuals, 14 are diagnosed with severe CHF classified as NYHA class 3–4, while the other 14 are healthy controls from PTB.

To mitigate training bias related to gender and age characteristics, each pathological sample is paired with a corresponding control signal that matches both gender and approximate age. In other words, for every pathological subject, there is a control subject with the same gender and a similar age. This balance is illustrated in Table 1, while Figure 1 compares a healthy control sample and a pathological sample.

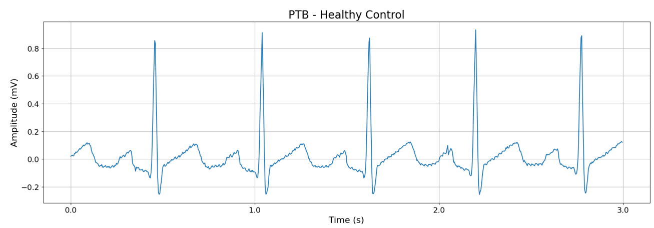
Table 1: Pairs formed based on gender and age, showing the correspondence between pathological and control signals.

Pair	**Gender (CHF)**	**Gender (Control)**	**Age (CHF)**	**Age (Control)**
Pair 1	Male	Male	71	68
Pair 2	Female	Female	61	67
Pair 3	Male	Male	63	64
Pair 4	Male	Male	54	54
Pair 5	Female	Female	59	57
Pair 6	Male	Male	48	37
Pair 7	Male	Male	51	50
Pair 8	Female	Female	63	69
Pair 9	Male	Male	22	26
Pair 10	Female	Female	54	54
Pair 11	Male	Male	61	59
Pair 12	Male	Male	63	68
Pair 13	Female	Female	61	52
Pair 14	Male	Male	53	55

The ECG signals' sampling rate was standardized at 250 Hz and underwent no noise filtering. Each of the 840 samples used for training consists of 3 seconds, resulting in a total of 750 time steps per sample. Normalization was conducted using Min-Max normalization, and no additional feature extraction was performed, meaning that only the amplitude of the ECG was considered for training the models. In addition to the ECG, the training data includes labels and the age and gender information for each sample. The final dataset was partitioned into 80% for training and 20% for testing, with 20% of the training set allocated for validation.



(a) Pathological ECG from BIDMC database.



(b) Healthy control ECG from PTB database.

Fig. 1: Comparison between a pathological and a healthy ECG sample: (a) ECG signal from a patient with CHF and (b) ECG signal from a healthy control.

3.2 LSTM Architecture

The LSTM model consists of three bidirectional LSTM layers with 100 units each, processing sequences of ECG signals in both forward and backward directions. This bidirectional architecture captures temporal dependencies effectively, enhancing the model's understanding of the data.

Each LSTM layer is followed by a dropout layer set to 30% to prevent overfitting. The final LSTM layer outputs a vector that consolidates the temporal information and concatenates with age and gender features. A dense layer with two units (one for each class) utilizes sigmoid activation for binary classification. The model employs binary cross-entropy as the loss function, Adam optimizer for weight correction, and is trained for 150 epochs with an early stopping criterion to retain the best model based on validation loss.

3.3 CNN Architecture

The CNN model employs three one-dimensional convolutional layers to extract local features from the ECG signals. The first layer applies 64 filters with a kernel size of 3, followed by max pooling with a pool size of 2, which reduces dimensionality while preserving key features. A dropout layer with a 30% rate is also included to enhance generalization.

The second convolutional layer uses 128 filters, while the third layer uses 256 filters, each followed by max pooling and dropout. The data dimensionality is progressively reduced, allowing the model to focus on the most relevant features. Figure 2 shows a graphic representation of the convolutional layers. Lastly, the dense layer acts as a classifier, returning the probability of each sample belonging to the pathological or non-pathological class.

Both architectures utilize binary cross-entropy for loss computation, with similar training configurations and model-saving strategies. This ensures that the best-performing model is retained for evaluation.

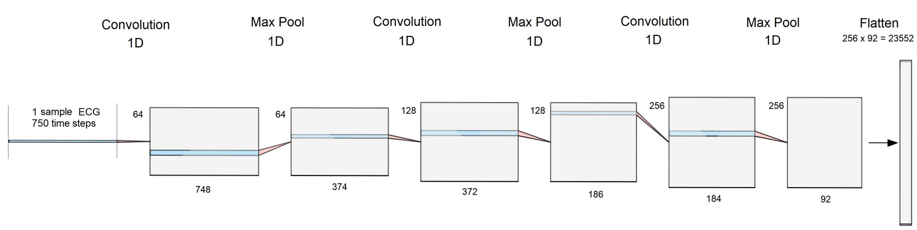


Fig. 2: Representation of the dimensions and operations in the 1D CNN. The figure highlights the operations performed within the convolutional layers, illustrating how the input data is processed through each layer.

4 Results & Discussion

This section presents the outcomes of the experiments, comparing the performance of the CNN and LSTM models in detecting CHF from ECG data. It also includes an analysis of the results, emphasizing the strengths and weaknesses of each model and a comparison with state-of-the-art methods.

4.1 CNN Model Results

The CNN model achieved superior performance compared to LSTM, with a loss of 0.0977 and an accuracy of 98.21%. The model's sensitivity was 96.10%, indicating that it correctly identified 96.10% of the pathological cases. It demonstrated perfect specificity (100%), correctly classifying all non-pathological cases. Precision was also 100%, highlighting that all instances predicted as pathological were correct. The F1-score, which balances precision and sensitivity, reached 0.9801, showcasing the model's ability to maintain high precision and sensitivity.

The confusion matrix for the CNN model further illustrates its robust performance. Out of 91 non-pathological cases, all were correctly classified, with zero false positives. In contrast, of the 77 pathological cases, 74 were correctly classified, and only three were misclassified, resulting in a very low false-negative rate.

Table 2: Confusion Matrix for CNN Model

	Predicted Non-Pathological	Predicted Pathological
Non-Pathological	91	0
Pathological	3	74

4.2 LSTM Model Results

The LSTM model, while still providing solid results, underperformed compared to the CNN. It produced a higher loss of 0.2367 and an accuracy of 92.26%. The model's sensitivity was 92.21%, meaning it detected 92.21% of the pathological cases, a drop compared to the CNN. Specificity reached 92.31%, indicating a reasonable classification ability for non-pathological cases. The precision was recorded at 91.03%, and the F1-score was 0.9161, reflecting a good overall performance yet still trailing behind the CNN results.

The confusion matrix for the LSTM model reveals a higher incidence of false negatives and false positives when compared to CNN. Of the 91 non-pathological cases, 84 were correctly classified, resulting in 7 false positives. In contrast, of the 77 pathological cases, 71 were accurately identified, with 6 misclassifying as non-pathological.

Table 3: Confusion Matrix for LSTM Model

	Predicted Non-Pathological	Predicted Pathological
Non-Pathological	84	7
Pathological	6	71

4.3 Comparative Analysis

Overall, the CNN demonstrated superior performance, especially in sensitivity and F1-score, making it a better candidate for real-world applications where high precision and sensitivity are critical. The LSTM model, while still accurate, exhibited lower sensitivity, which could result in more pathological cases going undetected.

These results highlight the advantage of convolutional architectures in capturing spatial patterns in ECG signals. At the same time, LSTM, despite its ability to model temporal dependencies, may struggle with the inherent noise and variability of ECG data without further preprocessing or tuning. Figure 3 illustrates the loss progression of both CNN and LSTM models over the training epochs. The graph highlights a significant contrast in the convergence behaviour between the two models, where the CNN model demonstrates a faster and more stable reduction in loss.

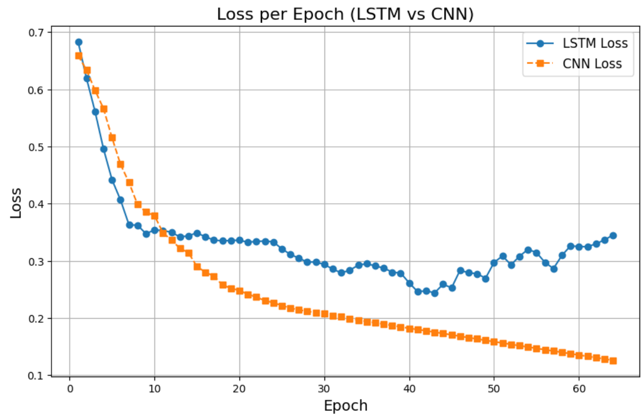


Fig. 3: Comparison Of Train Loss For CNN And LSTM Models

4.4 Comparison with Related-Work Methods

Table 4 compares our results and those of similar works from the literature, focusing on deep learning models for CHF classification.

Table 4: Comparison of results with state-of-the-art methods for CHF classification

Method	**Acc. (%)**	**Sens. (%)**	**Spec. (%)**	**Prec. (%)**	**F1 Score**
Proposed CNN	98.21	96.10	100.00	100.00	0.9801
Proposed LSTM	92.26	92.21	92.31	91.03	0.9161
[7] (LSTM)	99.86	99.85	99.85	99.87	0.9986
[12] (CNN)	98.74	-	-	-	-
[13] (CNN+LSTM)	99.52	99.31	99.28	-	0.9894
[15] (CNN)	97.80	-	-	-	-

As shown in Table 4, the proposed CNN model achieved an accuracy of 98.21%, which is comparable to other state-of-the-art methods, such as [12] with 98.74% and [15] with 97.80%. However, the sensitivity of our CNN model (96.10%) is slightly lower than the 99.31% reported in [13] but demonstrates excellent specificity and precision of 100.00%. The LSTM model in our work achieved an accuracy of 92.26%, which is lower than the models based on LSTM from [7], but it still demonstrates a good specificity (92.31%) and precision (91.03%).

The results indicate that while our CNN-based model performs competitively, especially in specificity and precision, the LSTM model could benefit from further optimization. Compared to [7], where multiple LSTM models were trained and fine-tuned, further hyperparameter tuning or additional layers might improve our LSTM model's performance.

5 Conclusion

This study demonstrated that deep learning techniques are robust and effective for detecting pathologies in ECG signals. The CNN outperformed the LSTM in all metrics for CHF detection, especially when considering the samples used in this study (3 seconds), highlighting its superior ability to classify CHF from ECG data.

This result occurs because the LSTM's strength in capturing long-term dependencies is not fully utilized with shorter 3-second samples. At the same time, CNN excels with smaller inputs by effectively extracting spatial features through convolutional layers. As a result, the CNN makes better use of the shorter sample length, leading to superior performance in detecting CHF in this context.

Classification accuracies in this and related studies typically range from 97% to 99%, largely due to the nature of ECG signals, which can be divided into many smaller segments. While a few segments may be misclassified, the majority are correctly identified. Implementing thresholds based on the classification of multiple segments can improve overall accuracy by allowing a more reliable inference about the entire signal, thereby enhancing diagnostic confidence and aiding clinical decision-making.

Future work could explore using hybrid CNN-LSTM architectures, fragmenting the signal into both small and large samples to maximize the strengths of each network, ensuring that smaller samples capture only heartbeats to avoid bias—extracting just the heartbeat eliminates the issue of fixed-window samples containing varying heart rates—and applying thresholds, where if a certain percentage of samples from a larger signal are classified into the same class, the overall signal can confidently be inferred to belong to that class.

Acknowledgment

This work was supported by national funds through FCT/MCTES (PIDDAC): CeDRI, UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and UIDP/05-757/2020 (DOI: 10.54499/UIDP/05757/2020); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020).

References

1. Anuradha, B., Reddy, V.C.V.: Ann for classification of cardiac arrhythmias. ARPN Journal of Engineering and Applied Sciences **3**(3), 1–6 (2008)
2. Baim, D.S., Colucci, W.S., Monrad, E.S., Smith, H.S., Wright, R.F., Lanoue, A., Gauthier, D.F., Ransil, B.J., Grossman, W., Braunwald, E.: Survival of patients with severe congestive heart failure treated with oral milrinone. Journal of the American College of Cardiology **7**(3), 661–670 (1986). https://doi.org/10.1016/S0735-1097(86)80478-8

3. Baloglu, U.B., Talo, M., Yildirim, O., Tan, R.S., Acharya, U.R.: Classification of myocardial infarction with multi-lead ecg signals and deep cnn. Pattern Recognition Letters **122**, 23–30 (2019). https://doi.org/10.1016/j.patrec.2019.02.016
4. Borghi, P.H., Borges, R., Teixeira, J.P.: Atrial fibrillation classification based on mlp networks by extracting jitter and shimmer parameters. In: CENTERIS/ProjMAN/HCist 2020 - Procedia Computer Science. vol. 181, pp. 931–939 (2021). https://doi.org/10.1016/j.procs.2021.01.249
5. Bousseljot, R., Kreiseler, D., Schnabel, A.: Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. Biomedizinische Technik / Biomedical Engineering **40**(s1), 317–318 (1995)
6. Costa, R., Winkert, T., Manhães, A., Teixeira, J.P.: Qrs peaks, p and t waves identification in ecg. In: CENTERIS/ProjMAN/HCist 2020 - Procedia Computer Science. vol. 181, pp. 957–964 (2021). https://doi.org/10.1016/j.procs.2021.01.252
7. Darmawahyuni, A., Nurmaini, S., Yuwandini, M., Rachmatullah, M.N., Firdaus, F., Tutuko, B.: Congestive heart failure waveform classification based on short time-step analysis with recurrent network. Informatics in Medicine Unlocked **21**, 100441 (2020). https://doi.org/10.1016/j.imu.2020.100441
8. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P.C., Mark, R., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation [Online] **101**(23), e215–e220 (2000)
9. Guedes, V., Junior, A., Teixeira, F., Fernandes, J., Teixeira, J.P.: Long short term memory on chronic laryngitis classification. In: CENTERIS/ProjMAN/HCist 2018 - Procedia Computer Science. vol. 138, pp. 250–257 (2018). https://doi.org/10.1016/j.procs.2018.10.036
10. Hossain, K.F., Kamran, S.A., Tavakkoli, A., Pan, L., Ma, X., Rajasegarar, S., Karmaker, C.: Ecg-adv-gan: Detecting ecg adversarial examples with conditional generative adversarial networks. In: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 50–56 (2021). https://doi.org/10.1109/ICMLA52953.2021.00016
11. Hou, B., Yang, J., Wang, P., Yan, R.: Lstm-based auto-encoder model for ecg arrhythmias classification. IEEE Transactions on Instrumentation and Measurement **69**(4), 1232–1240 (2020). https://doi.org/10.1109/TIM.2019.2910342
12. Jahmunah, V., Ng, E.Y.K., Tan, R.S., Acharya, U.R.: Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using gaborcnn model with ecg signals. Computers in Biology and Medicine **134**, 104457 (2021). https://doi.org/10.1016/j.compbiomed.2021.104457
13. Kusuma, S., Jothi, K.R.: Ecg signals-based automated diagnosis of congestive heart failure using deep cnn and lstm architecture. Biocybernetics and Biomedical Engineering **42**(1), 247–257 (2022). https://doi.org/10.1016/j.bbe.2022.02.003
14. Organização Pan-Americana da Saúde: Doenças cardiovasculares. Online (2023), https://www.paho.org/pt/topicos/doencas-cardiovasculares, accessed: Jul. 30, 2024
15. Porumb, M., Iadanza, E., Massaro, S., Pecchia, L.: A convolutional neural network approach to detect congestive heart failure. Biomedical Signal Processing and Control **55**, 101597 (2020). https://doi.org/10.1016/j.bspc.2019.101597
16. Rodrigues, P.M., Teixeira, J.P.: Classification of electroencephalogram signals using artificial neural networks. In: Proceedings of 3rd International Conference on BioMedical Engineering and Informatics (BMEI'10). Yantai, China (October 2010). https://doi.org/10.1109/BMEI.2010.5639941