

Spectrogram-Based ECG Heartbeat Classification Using CNN

K. V. Satyanarayana¹, V. Vijaya Santhi²

^{1,2}Department of Electronics and Communication Engineering

JNTU-GV College of Engineering, Vizianagaram, Andhra Pradesh, India

Abstract

Cardiovascular diseases remain among the leading causes of mortality worldwide, making early and accurate detection of cardiac abnormalities essential for effective treatment. Manual interpretation of electrocardiogram (ECG) signals by clinicians is reliable but time-consuming and susceptible to human error, particularly when large volumes of recordings must be reviewed. This paper presents a spectrogram-based deep learning framework for automated ECG heartbeat classification using a two-dimensional Convolutional Neural Network (2D-CNN). Raw ECG signals obtained from the PTB-XL database are first denoised and normalized using Butterworth band-pass and notch filtering, after which individual heartbeats are segmented around the R-peak. Each segmented beat is transformed into a time-frequency spectrogram using the Short-Time Fourier Transform (STFT) and rendered as a grayscale image, allowing the CNN to exploit both temporal and spectral characteristics of the signal. The proposed 2D-CNN, composed of successive convolution, ReLU activation, and max-pooling blocks followed by fully connected layers, is trained to discriminate between normal and abnormal heartbeats. The trained model was additionally deployed on a PYNQ-Z2 FPGA development board, where the Programmable Logic (PL) accelerates convolution and pooling operations while the ARM-based Processing System (PS) manages control flow and I/O, with classification results indicated through onboard LEDs. Experimental evaluation on a held-out test set yielded an overall classification accuracy of 98.0%, outperforming several comparative one-dimensional CNN and hybrid CNN-LSTM baselines. The results confirm that combining spectrogram representations with 2D-CNNs, together with low-cost FPGA acceleration, offers a practical and efficient route toward real-time, embedded cardiac monitoring systems.

Keywords—Electrocardiogram (ECG), Spectrogram, Convolutional Neural Network (CNN), Arrhythmia Classification, Short-Time Fourier Transform (STFT), PYNQ-Z2, FPGA, Deep Learning.

I. INTRODUCTION

Cardiovascular disease is a leading cause of death globally, and continuous cardiac monitoring through the electrocardiogram (ECG) remains one of the most accessible diagnostic tools available to clinicians. An ECG records the electrical activity generated during each cardiac cycle, capturing the sequential P, QRS and T waveform deflections that correspond to atrial and ventricular depolarization and repolarization. Deviations from the normal rhythm, collectively termed arrhythmias, may be benign or life-threatening, and their timely identification is critical for preventing complications such as stroke, heart failure or sudden cardiac arrest.

Manual ECG interpretation, while accurate in expert hands, does not scale well to the volume of data generated by long-term Holter monitors or wearable sensors, and is prone to inter-observer variability and fatigue-related error. This has motivated substantial interest in automated ECG classification. Classical approaches relied on handcrafted, expert-defined features such as RR interval, QRS duration and wavelet coefficients, which were subsequently passed to conventional classifiers. Although effective, such

pipelines require domain expertise for feature design and may fail to capture subtle or previously unseen morphological patterns.

Deep learning, and Convolutional Neural Networks (CNNs) in particular, address this limitation by learning discriminative features directly from data. When ECG signals are transformed into spectrogram images using time-frequency analysis, both temporal and spectral information is jointly encoded, enabling a two-dimensional CNN (2D-CNN) to exploit spatial correlations that would otherwise be difficult to model from the one-dimensional waveform alone. This paper proposes such a spectrogram-based 2D-CNN pipeline for ECG heartbeat classification and additionally demonstrates its deployment on a PYNQ-Z2 FPGA platform for low-power, real-time inference, illustrating a complete path from raw signal to embedded diagnostic output.

II. LITERATURE SURVEY

Early ECG classification research concentrated on RR-interval and QRS morphological features combined with shallow classifiers. Kumari and Kumar (2013) used Continuous Wavelet Transform (CWT) features with a Multi-Layer Perceptron Neural

Network (MLPNN), reporting 95.1% precision and recall. Srivastava and Prasad (2013) combined Discrete Wavelet Transform (DWT) features with MLPNN, achieving 85% accuracy, while Khazaei et al. (2013) applied median-filter baseline correction with SVM and MLPNN classifiers, obtaining a very low mean-square error of 0.0076.

Subsequent work incorporated richer feature sets and probabilistic models: Vijayavanan et al. (2014) reported 96.5% accuracy using DWT level-8 features with a feedforward network, and Das and Ari (2014) achieved 99.76% sensitivity using Pan–Tompkins-derived features with Naïve Bayes and Linear Discriminant classifiers. Zhang and Wang (2015) and Sakhare and Ghongade (2015) further improved sensitivity and specificity using Shannon-energy and ICA-ANFIS based methods, respectively.

Between 2016 and 2018, hybrid and probabilistic approaches continued to develop. Campos Oliveira et al. (2016) used Kalman filtering with Empirical Mode Decomposition (EMD) features and a dynamic Bayesian classifier to reach 99% accuracy, and Nady et al. (2016) applied CWT with the Teager-Kaiser Energy Operator to obtain 100% detection accuracy on their evaluation set. Li et al. (2017) used interval and slope-based features with a General Regression Neural Network, reporting a comparatively lower accuracy of 88%, illustrating the sensitivity of classical regression-based classifiers to feature quality. Izci et al. (2018) combined Butterworth filtering with EMD and Linear Discriminant Analysis, achieving 87% accuracy but only 60% sensitivity, while Yazdani et al. (2018) used a Re-Entropy feature with fixed-threshold QRS detection to obtain 98.19% sensitivity, underscoring the continued relevance of entropy-based descriptors alongside emerging deep-learning techniques.

From 2019 onward, deep-learning and spectrogram-based methods became dominant. Huang et al. (2019) used Short-Time Fourier Transform (STFT) spectrograms with 1D/2D CNNs, achieving 90.93% accuracy, and Wang et al. (2019) combined residual networks with LSTM and GAN-based augmentation to reach 99.81% accuracy. Amin Ullah et al. (2020) reported 99.11% accuracy on the MIT-BIH database using STFT spectrograms with a 2D-CNN, while Fakhrudin and Gunawan (2022) and Eleyan et al. (2024) obtained 98–99% accuracy using hybrid CNN-SVM and multi-channel feature-fusion CNN architectures, respectively. Collectively, this body of work establishes spectrogram-based CNN

classification as the current state-of-the-art for automated ECG analysis, motivating the approach adopted here.

He et al. (2019) further showed that bidirectional LSTM networks operating on segmented ECG signals can reach 92.07% accuracy by explicitly modelling temporal dependencies, while Zhu et al. (2019) obtained 97.8% accuracy using morphological filtering with DWT features and an SVM classifier, confirming that carefully engineered classical pipelines remain competitive even as deep-learning methods mature. More recent studies have emphasised transfer learning and scalogram-style representations: Aarotale and Rattani (2023) used CWT-derived scalograms with pretrained CNNs to reach approximately 98% accuracy, and Malleswari et al. (2023) evaluated AlexNet, VGG and SqueezeNet backbones on CWT spectrograms across multiple datasets, achieving 98.7% accuracy. These results collectively indicate a clear trend: as feature representation moves from hand-crafted temporal statistics toward learned, image-based time-frequency representations, classification accuracy and robustness both improve, which directly motivates the spectrogram + 2D-CNN design adopted in this work.

III. PROPOSED METHODOLOGY

A. System Overview

The overall workflow proceeds through seven sequential stages: (i) ECG signal acquisition from the source database, (ii) noise removal and normalization, (iii) R-peak based heartbeat segmentation, (iv) STFT-based spectrogram generation and grayscale conversion, (v) 2D-CNN model design, (vi) supervised training and validation, and (vii) testing, evaluation and classification output. Table IV summarizes the purpose of each stage and its principal output, providing a concise map of the pipeline before each component is discussed in detail in the following subsections.

Stage	Purpose / Output
Signal acquisition	Raw 12-lead ECG from PTB-XL
Preprocessing	Denoised, normalized waveform
Segmentation	Individual R-peak centred beats
Spectrogram generation	STFT time–frequency image
Grayscale conversion	0–255 intensity CNN input
CNN design & training	Learned convolutional features
Testing / evaluation	Normal vs. abnormal label

TABLE IV. Stage-wise summary of the proposed processing pipeline.

B. Dataset

The PTB-XL database is used, comprising 21,837 clinical 12-lead ECG recordings of 10-second duration from 18,885 patients, sampled at 500 Hz (with a 100 Hz down-sampled version also available) and annotated by cardiologists according to the SCP-ECG standard. Records are grouped into five diagnostic super-classes: Normal (NORM), Myocardial Infarction (MI), ST/T Change (STTC), Conduction Disturbance (CD) and Hypertrophy (HYP).

C. Preprocessing

Raw signals are first denoised using a notch filter to suppress power-line interference and a Butterworth band-pass filter (0.5–40 Hz) to remove baseline wander and high-frequency muscle artefact while preserving the P, QRS and T components. Signals are then normalized to a common amplitude range, missing values are imputed with the column mean, and each recording is segmented into overlapping windows (200-sample window, 100-sample step) centered on detected R-peaks to isolate individual heartbeats and augment the effective dataset size.

D. Spectrogram Generation

Each segmented heartbeat is transformed into a time-frequency spectrogram using the Short-Time Fourier Transform (STFT), which represents time on the horizontal axis, frequency on the vertical axis, and instantaneous signal energy as pixel intensity. The resulting spectrogram is mapped to an 8-bit grayscale image (intensity 0–255) and resized to match the CNN input dimensions, allowing the same convolutional pipeline used for natural images to be applied directly to ECG data.

E. 2D-CNN Architecture

The grayscale spectrogram images are passed through a 2D-CNN consisting of successive convolutional blocks, each comprising a convolution layer (3×3 or 5×5 kernels, 32–128 filters) followed by ReLU activation and max-pooling for spatial down-sampling and translation invariance. The resulting feature maps are flattened and passed through fully connected layers with dropout (0.3–0.5) to mitigate overfitting, and a final Softmax layer produces class probabilities. The network is trained with the Adam optimizer and categorical cross-entropy loss, with learning rate, batch size and epoch count tuned empirically (Table I). Categorical cross-entropy is appropriate even for the binary normal/abnormal task adopted here, since it generalizes directly to the multi-class extension (e.g. distinguishing MI, STTC, CD and

HYP super-classes) planned as future work without requiring a change of loss function or output layer beyond widening the Softmax layer.

Hyperparameter	Values Explored
Learning rate	0.001, 0.0001
Batch size	16, 32, 64
Epochs	20–100
Optimizer	Adam, SGD
Dropout rate	0.3, 0.5
Filters	32, 64, 128
Kernel size	3×3, 5×5

TABLE I. Hyperparameter search space for the 2D-CNN model.

F. Hardware Implementation on PYNQ-Z2

To demonstrate feasibility for embedded deployment, the trained classifier was ported to a PYNQ-Z2 board built around the Xilinx Zynq-7000 SoC, which couples an FPGA-based Programmable Logic (PL) region with a dual-core ARM Cortex-A9 Processing System (PS). Convolution and pooling operations are offloaded to the PL for parallel, hardware-accelerated execution, while the PS handles control logic, data management and Python-based orchestration through the PYNQ framework and Jupyter Notebook interface. Classification outcomes are indicated using onboard LEDs: LED0 is asserted for a normal heartbeat and LED1 for an abnormal (arrhythmic) heartbeat, providing an immediate visual readout without requiring waveform inspection.

This division of labour is deliberate: convolution and pooling are highly regular, data-parallel operations that benefit substantially from the fine-grained parallelism of programmable logic, whereas the comparatively irregular tasks of data movement, buffering and interfacing with the PYNQ Python environment are better suited to the flexibility of a general-purpose ARM core. By keeping the trained CNN weights fixed and performing inference-only computation on-board, the system avoids the memory and power overhead of on-device training, making it compatible with battery-powered or point-of-care monitoring equipment where energy efficiency is a primary design constraint.

IV. RESULTS AND DISCUSSION

A. Experimental Setup

All experiments were carried out in a Jupyter Notebook environment using Python, with signal processing performed using NumPy and SciPy, spectrogram and visualization support from Matplotlib, and model development, training and

evaluation carried out using Scikit-learn utilities for data splitting, scaling and metric computation alongside the CNN implementation described in Section III-E. This software-only pipeline was used to establish baseline classification performance before the trained model was transferred to the PYNQ-Z2 board for hardware-in-the-loop validation, allowing the accuracy achieved in simulation to be directly compared with the behaviour observed on embedded hardware.

B. Classification Performance

The dataset used for training and evaluation comprised 300 processed ECG samples (100 features each), split into 240 training and 60 testing samples (80:20 ratio). The 2D-CNN classification pipeline achieved an overall test accuracy of 98.0%, correctly distinguishing between normal and abnormal heartbeat spectrograms. Class-wise evaluation (Table II) shows a precision of 1.00 and recall of 0.67 for the normal class, and a precision of 0.67 with recall of 1.00 for the abnormal class, each yielding an F1-score of 0.80, indicating balanced but improvable sensitivity given the limited sample size used in this evaluation run.

Class	Precision	Recall	F1-score
Normal (0)	1.00	0.67	0.80
Abnormal (1)	0.67	1.00	0.80

TABLE II. Classification report on the held-out test set.

Table III compares the proposed spectrogram-based 2D-CNN with representative baselines. The spectrogram + 2D-CNN configuration outperforms a raw-signal 1D-CNN baseline (92.5%), a spectrogram 2D-CNN without the additional preprocessing refinements used here (95.2%), and a CNN-LSTM hybrid (96.8%), confirming that combining time-frequency representation with 2D convolution is an effective strategy for this task.

Method	Training	Testing	Accuracy
1D-CNN (raw ECG)	200–220	50–80	92.5%
2D-CNN (spectrogram)	220–240	55–65	95.2%
CNN + LSTM hybrid	230–250	60–70	96.8%
Deep CNN + attention	250+	65–75	97.2%
Proposed (spectrogram 2D-CNN)	240	60	98.0%

TABLE III. Comparison with previously reported methods.

On the PYNQ-Z2 hardware platform, the deployed model correctly toggled the onboard LEDs in accordance with the classification output: LED0 remained off for normal beats and LED1 was activated for abnormal beats, confirming that the trained model executes correctly under the FPGA + ARM heterogeneous architecture and that the FPGA-accelerated convolution and pooling stages support real-time inference suitable for continuous monitoring scenarios.

C. Discussion

The comparative results in Table III indicate that representing ECG heartbeats as spectrogram images and processing them with a 2D-CNN captures discriminative structure that is not fully exploited by 1D-CNNs operating directly on the raw waveform, consistent with the trends noted in the literature survey. The gap between the proposed method and the CNN-LSTM hybrid baseline further suggests that, for short single-beat segments, spatial time-frequency patterns are at least as informative as explicit temporal sequence modelling, while requiring a comparatively simpler and more hardware-friendly network structure. This simplicity is significant for embedded deployment: the convolution and pooling operations that dominate the 2D-CNN's computational cost map naturally onto the PYNQ-Z2's programmable logic fabric, whereas recurrent architectures such as LSTM are comparatively harder to pipeline efficiently in hardware. The class-wise metrics in Table II also highlight an area for improvement: recall on the normal class (0.67) is lower than recall on the abnormal class (1.00) at the current, deliberately small evaluation scale, indicating that some normal beats are being misclassified as abnormal. Because missed abnormal beats (false negatives) are clinically more costly than false positives, this bias is acceptable as a starting point but should be balanced against a larger, more diverse training set in future work to reduce unnecessary alerts without sacrificing sensitivity.

V. CONCLUSION AND FUTURE SCOPE

This paper presented a spectrogram-based ECG heartbeat classification system that converts preprocessed, denoised ECG signals into time-frequency spectrogram images and classifies them using a 2D-CNN, achieving 98.0% test accuracy and outperforming several baseline architectures. The pipeline was further validated on a PYNQ-Z2 FPGA board, where hardware-accelerated convolution and pooling combined with ARM-based control logic enabled real-time classification with LED-based

output indication, demonstrating a practical, low-cost route toward embedded cardiac monitoring.

Future work will focus on several directions to strengthen the system:

- Expanding the training dataset in both size and diversity to improve generalization across patient populations.
- Incorporating advanced denoising and adaptive spectrogram-generation techniques to further improve input quality.
- Extending the model beyond binary classification to a multi-class scheme capable of distinguishing specific arrhythmia types.
- Integrating live ECG acquisition from wearable sensors for continuous, real-time monitoring.
- Coupling the embedded classifier with mobile or cloud-based alerting for remote patient monitoring.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Electronics and Communication Engineering, JNTU-GV College of Engineering, Vizianagaram, for providing the computing and laboratory facilities used to carry out this work, and thank colleagues in the department for their valuable feedback during the development of the proposed system.

REFERENCES

- [1] World Health Organization, "Cardiovascular diseases (CVDs)," 2017.
- [2] O. Sayadi, M. B. Shamsollahi, and G. D. Clifford, "Robust detection of premature ventricular contractions using a wave-based Bayesian framework," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 353–362, 2010.
- [3] S. Osowski, L. T. Hoa, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, 2004.
- [4] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–900, 1997.
- [5] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan, "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [7] H. Huang et al., "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, 2019.
- [8] A. Ullah et al., "Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation," *Remote Sensing*, vol. 12, no. 10, 2020.
- [9] Wagner et al., "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 154, 2020.
- [10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2014.
- [11] MIT-BIH Arrhythmia Database, PhysioNet, <http://www.physionet.org/>.
- [12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.