

CardioView: a framework for detection Premature Ventricular Contractions with eXplainable Artificial Intelligence

Giuseppe Arienzo^{1,†}, Alessia Auriemma Citarella^{1,†}, Fabiola De Marco^{1,*,†}, Anna Maria De Roberto^{1,†}, Luigi Di Biasi^{1,†}, Rita Francese^{1,†} and Genoveffa Tortora^{1,†}

¹Computer Science Department of University of Salerno, Fisciano, SA 84084 IT

Abstract

Artificial Intelligence plays a vital role in disease diagnosis, but effectively classifying diverse Premature Ventricular Contraction (PVC) subtypes remains a challenge. While computer-aided systems demonstrate high performance, human oversight remains crucial for reliability. This study introduces Explainable AI algorithms utilizing the GRADient-weighted Class Activation Mapping algorithm, as part of the proposed framework CardioView, providing insights into the diagnosis process. With high accuracy, recall, precision, and AUC (96.21%, 98.09%, 94.74%, 99.28% respectively), the system enhances understanding of PVC classification. CardioView allows individuals to gain insights into the discrimination process, revealing its operations and visualizing the components of the electrocardiogram waveform that aid in distinguishing between PVC and non-PVC classes, as well as within various PVC subclasses. Furthermore, CardioView integrates a human-in-the-loop approach, ensuring active involvement of cardiologists throughout the diagnostic process and reinforcement learning mechanisms.

Keywords

Computer-aided diagnosis system, Convolutional Neural Networks, eXplainable Artificial Intelligence, Premature Ventricular Contractions, Symbiotic Artificial Intelligence

1. Introduction

In the era of medical development, Computer-Aided Diagnosis (CAD) systems may become a crucial component of the healthcare revolution, mainly when combined with Artificial Intelligence (AI) tools. In particular, these systems may improve diagnostic accuracy, expedite workflow, enable early detection, offer decision support, help individualized medication, and act as teaching aids. CAD systems can potentially improve comprehension of patient conditions, enhance patient outcomes, and streamline healthcare delivery [1]. While AI capabilities

INI-DH 2024: Workshop on Innovative Interfaces in Digital Healthcare, in conjunction with International Conference on Advanced Visual Interfaces 2024 (AVI 2024), June 3–7, 2024, Arenzano, Genoa, Italy (2024)

*Corresponding author.

†These authors contributed equally.

✉ g.arienzo99@gmail.com (G. Arienzo); aauriemmacitarella@unisa.it (A. Auriemma Citarella); fdemarco@unisa.it (F. De Marco); m.derob89@gmail.com (A. M. De Roberto); ldibiasi@unisa.it (L. Di Biasi); francese@unisa.it (R. Francese); tortora@unisa.it (G. Tortora)

🆔 0000-0002-6525-0217 (A. Auriemma Citarella); 0000-0003-4285-9502 (F. De Marco); 0000-0001-6201-5193 (A. M. De Roberto); 0000-0002-9583-6681 (L. Di Biasi); 0000-0002-6929-0056 (R. Francese); 0000-0003-4765-8371 (G. Tortora)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

continue to advance, its inherent “black box” characteristic restrains its widespread adoption in the digital health sector. Specifically, many AI techniques are not transparent in their underlying methodology responsible for their exceptional performance, as documented in scientific literature [2]. Human operators still need to learn to rely on CAD systems results without a comprehensive understanding of the underlying processes. Establishing trust in AI requires an elucidation explanation of the mechanisms utilized to generate its outcomes. This objective can be achieved through the eXplainable Artificial Intelligence (XAI) approach [3]. XAI refers to developing AI systems that can explain their decision-making processes clearly and comprehensively. Common XAI approaches include feature importance analysis, decision tree analysis, and visualization tools such as heat maps and saliency maps, highlighting the most significant portions of the data [4]. Particular relevant use of XAI is to highlight the portions of a medical image that are more relevant for classification tasks, improving our understanding of disease development and the ability to discover novel imaging biomarkers, disease patterns, and abnormalities [5] [6].

Physiologically, in the cardiac system, electrical signals follow a defined path during a typical cardiac cycle, initiating contraction sequentially: impulse for cardiac rhythm originates from the sinoatrial node that represents the “cardiac pacemaker”, and from that situs, it spreads along cardiac atria. Subsequently, this impulse travels down through the conduction pathways and causes ventricular depolarization and contraction essential for pumping out blood. The characteristic elements of a normal heartbeat consist of the P wave, the QRS complex (comprising the Q, R, and S waves), and the T wave. The QRS complex represents the electrical activity linked to ventricular contractions. Premature Ventricular Contraction (PVC) is an ectopic arrhythmia originating outside the sinoatrial node, disrupting normal cardiac rhythm. Symptoms include palpitations, skipped heartbeats, dizziness, and breathlessness. Ectopic beats can arise from atrial or ventricular myocardium and manifest as premature heartbeats, differing from sustained ectopic rhythms. PVCs result from abnormal myocardial depolarization, leading to premature contractions and altered cardiac output. They can originate from various ventricular regions, each generating distinct PVC morphologies based on the impulse origin. Among the diagnostic criteria of PVC analysis we can find:

- premature QRS compared to what is expected from previous cardiac rhythm;
- different morphology of QRS compared to previous QRS analysis (broad QRS complex \geq 120 ms);
- discordant ST segment and T wave changes;
- PVC is usually followed by a compensatory pause;
- ectopic beats originating from the right ventricle have a left bundle block morphology pattern, while ectopic beats originating from the left ventricle have a right bundle block morphology pattern.

PVCs are common in the general population though their reported prevalence ranges widely, between 1 and 40% [7] - [8]. This large variation likely depends on the different study populations and different ways of recording and storing ECG [9]. The significance of PVC in the general health population is controversial [10], while the association between PVCs and negative prognosis in patients with structural heart disease is well established [11] - [12].

This research employs AI and XAI techniques to investigate PVCs, supporting cardiologists in detecting PVC classes. Integrating XAI features into the model aims to enhance its transparency and comprehensibility for cardiologists, enabling them to gain insights into PVC physiology and potentially elevating diagnostic accuracy.

This study analyzed images from the MIT-BIH Arrhythmia Database, a widely accepted resource in arrhythmia detection [13]. The chosen images were employed for training a Convolutional Neural Network (CNN) for the PVCs classification. In addition, XAI features were incorporated using the Grad-CAM approach, which generated images highlighting the specific areas of the heartbeats considered crucial by the CNN for its decisional process [14]. Subsequently, clustering techniques, such as K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [15], identified three distinct clusters based on the Grad-CAM-generated images. This process contributes to the development of a framework, called *CardioView*.

Our key contributions are:

- We use the images generated by an XAI algorithm for improving the performance of a CNN-based PVC classifier competitive with the state-of-art;
- We apply clustering algorithms to the images generated by the XAI algorithm and identify three PVC clusters useful for identifying patterns in PVC sub-classes.
- We propose a visual framework, which incorporate the human in the loop approach, ensuring active involvement and oversight of cardiologists throughout the PVC detection process.

CardioView is designed to gather information through surveys administered to cardiologists. Collecting responses assists us in enhancing the algorithms and performing reinforcement learning (RL). In this preliminary study, we focus solely on the initial phase: the classification of PVCs, the application of XAI algorithms, and their clusterization.

The paper organization follows: in Section 2, we reported the state of the art of using the XAI in PVC detection. Section 3 describes the research design, materials, and methods employed in the study. Section 4 provides the results of our work and related discussion. Section 5 summarizes the main findings of the study, implications, and future directions.

2. Related works

In this section, we provide a concise overview of existing research and current studies that involve the classification of PVCs using XAI techniques. The necessity of explainable PVC detection lies in the critical domain of medical diagnostics, particularly in cardiology. The interpretability of the AI models helps to clarify the rationale behind the findings, improving the care of patients. In [16], to show and visualize the saliency, the authors introduced a special implementation of Grad-CAM for the CNN model and a second method, which involves learning the input deletion mask for the LSTM model. In PVC, the LSTM network gives more weight to the aberrant PVC rhythm than the CNN network does to the proximal beats. The authors reported the following overall performance for eight arrhythmia classification tasks. In particular, the best results are for [17]: 98% of Precision, 97% of recall and F1 score, and an accuracy of 97%.

In [18], the authors proposed a novel deep learning approach for multilabel classification of ECG signals for the CPSC-2018 dataset. They reached an overall precision of 98.6%, a recall of 94.9%, an F1 score of 96.7% and 96.2 in accuracy. Additionally, class activation maps obtained from the Grad-CAM enhance the model interpretability. The XAI framework ensured that the CNN learned appropriate features, mitigating its black-box nature and instilling confidence in the model's results. Specifically for the PVC class, ECGs with PVC are distinguished by a vast QRS complex. In [19], the authors classified 12-lead ECG recordings from the CPSC-2018 dataset with a deep neural network. For the PVC class, they obtained 86.9% Precision, 83.9% of recall, 85.1% of F1 score, an AUC of 97.6%, and an accuracy of 97.1%. To evaluate the behaviour of the model at the patient and population levels, they used the Shapley Additive exPlanations approach (SHAP). In [20], to extract features from individual ECG leads, the authors used three different One-dimensional Convolutional Neural Networks (1D-CNNs), introducing a novel lead-wise attention module that combines the outputs from the three backbones to produce a more reliable representation. Moreover, they emphasized the XAI framework significantly, enabling our system to provide more meaningful and clinically relevant explanations for its predictions. The suggested approach, known as LightX3ECG, reached for PVC class in the CPSC-2018 dataset, a precision of 87.96%, a recall of 71.97%, an F1 score of 79.17% and 96.37% of accuracy. Also, the authors make innovative use of the XAI-generated images for improving the CNN classifier and for using clustering algorithms on the XAI dataset to identify PVC sub-patterns.

3. Methods

3.1. Working hypothesis & Workflow

PVCs can indicate underlying heart conditions like cardiomyopathy or coronary artery disease [21]- [22]. Several research suggest multiple distinct categories of PVCs may exist, as they vary in their associations with cardiac outcomes [23]. This observation leads to the formulation of the following initial working hypotheses (WHs):

WH1 If multiple PVC outcomes exist, then multiple PVC classes exist.

WH2 If multiple classes of PVC exist, it is possible to partition an ECG dataset composed of PVC and non-PVC ECG acquisition in multiple clusters, each related to a likely outcome.

A natural inference drawn from the WHs is that merely classifying ECG tracks into PVC and NON-PVC classes during the examination of the clinicians may prove insufficient, as reported in [24]. Furthermore, a crucial need arises to distinguish between different PVC classes has a preventative measure against further cardiac risks. Given the potential implications of PVC classes, developing precise and efficient methods for their detection and diagnosis is imperative.

Figure 1 illustrates the workflow of the proposed framework. The methodology initiates with dataset analysis and preprocessing to minimize signal noise and enhance quality, utilizing techniques tailored for both black and white (BW) and color (RGB) datasets. It comprises two main phases: initially, the BW dataset trains a CNN model employing classic and k-fold training, followed by feature extraction post-training; subsequently, the RGB dataset undergoes

training using the classic method only, with the resulting model applied to the Grad-CAM algorithm for creating a modified Grad-CAM version of the dataset. This version then retrains the model. Features are extracted similarly to the first phase. Finally, K-means and DBSCAN clustering algorithms are applied to the extracted features to identify potential PVC patterns. CardioView is structured to facilitate the incorporation of surveys obtained from cardiologists, which constitutes a crucial element of the human-in-the-loop process integral to RL.

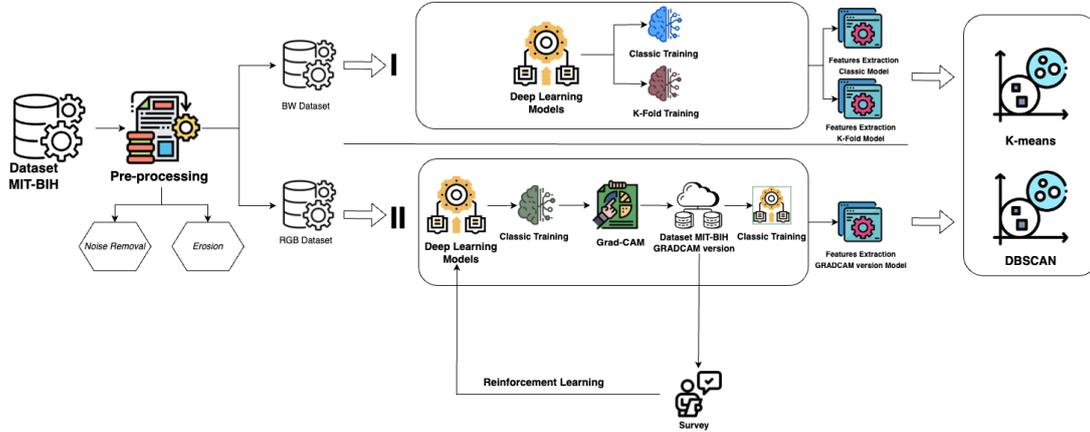


Figure 1: CardioView: Proposed Framework

3.2. Dataset & Pre-processing

This work primarily relies on the MIT-BIH Arrhythmia Database, containing 48 half-hour extracts from two-channel ECG recordings from 47 subjects, allowing 24-hour continuous monitoring. With data digitized at 360 samples per second per channel and 11-bit resolution, it serves as a benchmark for arrhythmia detection algorithms and ML/DL models. Our dataset includes 7130 elements for PVC QRS complexes and 75048 for non-PVC QRS complexes, with subjects aged 23 to 89, 60% hospitalized, and 40% outpatients. In this work, we utilized a dataset subset of a total of 14123 images. Pre-processing images is crucial for data quality improvement, ensuring suitability for analysis by removing noise and enhancing accuracy. In this study, raw data from the MIT-BIH Arrhythmia Database was converted into color images, and the imperfections were removed with adaptiveThreshold function in the CV2 library. Three datasets were employed: the BW and RGB version dataset and the dataset obtained by Grad-CAM visualization.

3.3. CNN design

This work explores various models, selecting the best-performing one to ensure optimal Grad-CAM representations. However, as networks become deeper, Grad-CAM representations may become less reliable due to increased noise. The network architecture includes layers for image normalization, convolution, pooling, upsampling, dropout, flattening, and fully connected layers. A CNN is trained using k-fold cross-validation with 4 epochs and 5 folds, enhancing

performance evaluation and generalization. Data is split into training, validation, and test sets, with an 80-10-10 ratio for standard training and a 90-10 ratio for k-fold validation. This approach improves model robustness and generalization, ensuring effective training and evaluation of the CNN network architecture.

3.4. XAI and clustering algorithms

The Grad-CAM algorithm enhances XAI by highlighting significant features contributing to model predictions. It operates by selecting the target layer, calculating gradients, and computing weights for activation maps. These maps are then aggregated and overlaid on the original image, emphasizing regions of interest. Grad-CAM aids model validation and fine-tuning, aligning with human intuition and expertise, improving model accountability, and building trust. This study employs two clustering algorithms: K-means and DBSCAN. K-means is an unsupervised method that partitions data into k clusters by iteratively updating cluster centroids. However, it requires a predefined number of clusters, is sensitive to initial centroid selection, and struggles with overlapping clusters. In contrast, DBSCAN does not require a predefined number of clusters and group points based on spatial density, making it more adaptable to various cluster shapes and sizes while effectively handling noise. Hyperparameter selection is crucial, impacting clustering model performance significantly. The Elbow method was used to determine the optimal number of clusters for K-means, with a knee point identified at k=3.

3.5. Proposed survey

The human-in-the-loop (HITL) and RL phase will incorporate a dynamic survey involving experienced cardiologists. This survey allows them to express their confidence levels regarding the outcomes generated by the Grad-CAM algorithm, fostering iterative refinement and improving collaborative decision-making in PVC diagnosis from the perspective of symbiotic AI (SAI) systems. Professionals will learn about XAI's role in early PVC detection from cardiac images. Participation will be voluntary, with optional personal data collection. Subsequently, participants will examine 20 images (10 randomly selected between PVC and non-PVC). They will identify in each images the presence or absence of PVCs based on their experience (Figure 2).

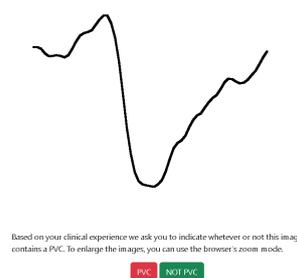


Figure 2: Survey step 1

If their response aligns with the truth matrix, they will receive the corresponding Grad-CAM

image to assess its efficacy in identifying crucial points. They will rate their confidence in the XAI algorithm on a Likert scale from 0 to 10 [25] and can leave comments for each Grad-CAM image. Participants disagreeing with the Grad-CAM result can suggest an alternative image zone for reinforcement learning. This step is graphically represented in Figure 3, where the result of applying Grad-CAM algorithm on a non-PVC image is shown.

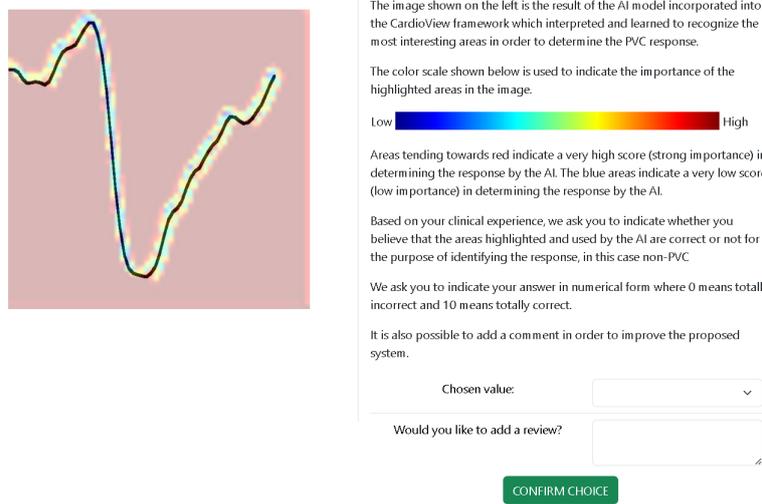


Figure 3: Survey steps

3.6. Performance measures

Performance evaluation is essential to assess the effectiveness of different ML and DL models. This work uses the following metrics: (Equations 1-3):

$$Accuracy (ACC) = \frac{TP + TN}{TN + FP + FN + TP} \quad (1)$$

$$Precision (PRE) = \frac{TP}{TP + FP} \quad (2)$$

$$Recall (RE) = \frac{TP}{TP + FN} \quad (3)$$

where TP are the *True Positives*, the total number of samples correctly classified as positive, FP are *False Positive*, the total number of items incorrectly classified as positive. TN (*True Negatives*) and FN (*False Negatives*) represent the total number of samples appropriately defined and incorrectly selected as negative, respectively.

The work assesses clustering algorithms using Silhouette Score [26], Calinski-Harabasz Index [27], and Davies-Bouldin Index [28] to measure clustering quality.

The Silhouette Score (see Equation 4), which ranges from -1 to 1, evaluates how well a data point matches its cluster in comparison to other clusters.

$$\text{Silhouette Score} = \frac{(b - a)}{\max(a, b)} \quad (4)$$

The Calinski-Harabasz Index (CH Index) (see Equation 5) measures the within-cluster to between-cluster dispersion ratio.

$$\text{Calinski - Harabasz Index} = \frac{B}{W} * \frac{(N - k)}{(k - 1)} \quad (5)$$

The Davies-Bouldin Index (DB Index) (see Equation 6) considers the internal similarity of the cluster and calculates the average similarity between each cluster and its most similar cluster.

$$\text{Davies - Bouldin Index} = \frac{1}{k} * \sum_{i=1}^k \max_{(j \neq i)} \left(\frac{s(i) + s(j)}{d(i, j)} \right) \quad (6)$$

4. Results

Table 1 shows the CNN classification results on the test sets for the original dataset (CNN) and the Grad-CAM version of the dataset (CNN-GC) with k-fold and classic training. Overall, CNN-GC with classic training reaches the best results on the classification test (96.21% of ACC, 99.28 of AUC, 98.09% of RE and 94.74% of PRE). On the other hand, the results on the test set with k-fold cross-validation are also high, reaching an ACC of 96.77%, 99.46% of AUC, 98.11% of RE and 95.33% of PRE.

Table 1
Results of CNN

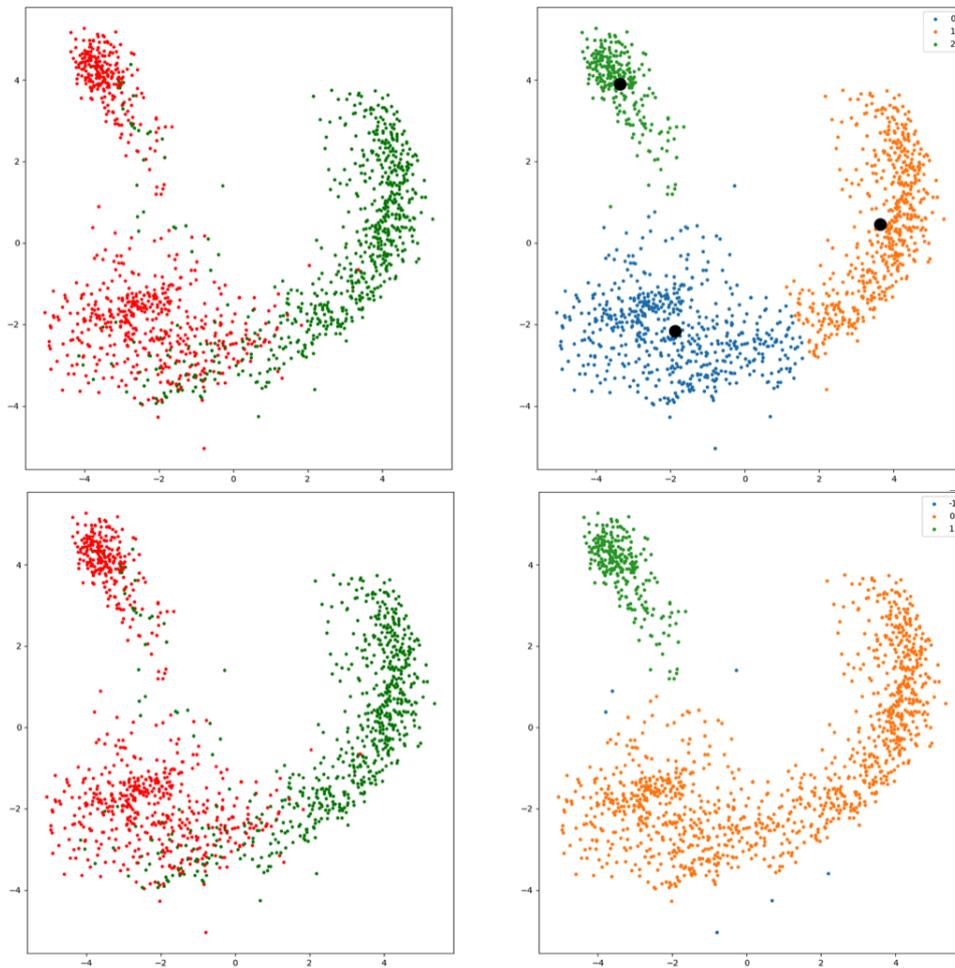
Model	ACC (%)	AUC (%)	RE (%)	PRE (%)
CNN (test)	95.64	98.91	97.13	94.17
CNN-GC (test)	96.21	99.28	98.09	94.74
K-fold				
CNN (test)	96.77	99.46	98.11	95.33

Table 2 shows Silhouette Score, CH, and DB Indexes on MIT-BIH dataset for test set. K-means partitions data into clusters based on centroid-based distance minimization and assumes spherical clusters of equal size. In contrast, DBSCAN identifies clusters based on data density without needing a predefined cluster number. This leads to different cluster results. Scatter plots depict positive (PVC) and negative (NON-PVC) classified images in red and green respectively. Generated clusters are shown in different colors. The best values of the three indexes are reached using the Grad-CAM version of the dataset and the classic training. For binary classification (see images on the left for Figure 4), the results of K-means and DBSCAN are very similar. On the GRADCAM version of the dataset, DBSCAN turns out to be less suitable in identifying intraclass variation of PVCs, despite detecting three clusters, since it cannot discriminate subpatterns within PVCs, unlike k-means.

Table 2

Values of Silhouette Score, CH and DB indexes k-means and DBSCAN on test dataset

Train	Silhouette Score	CH Index	DB Index
k-means (CL)	0.63	3219.87	0.50
K-means (KF)	0.58	2583.99	0.58
DBSCAN (CL)	-0.39	75.66	6.97
DBSCAN (KF)	0.07	270.75	2.53
Test set (Grad-CAM)			
k-means (CL)	0.64	3352.70	0.49
DBSCAN (CL)	0.22	419.80	2.60

**Figure 4:** K-Means (up) and DBSCAN (bottom) on test set GRADCAM version with classic training

5. Conclusions

In this study, the MIT-BIH dataset of BW images trained a CNN model using classic and K-fold training, followed by processing the Grad-CAM dataset of RGB images solely with classic training. Trained models extracted essential features, inputted into k-means and DBSCAN clustering algorithms to identify PVC and NON-PVC patterns for early diagnosis. Future works plan to integrate surveys for the human-in-the-loop phase to explain CNN workings, increasing trust in medical AI. Grad-CAM facilitated visualization of ECG waveform segments distinguishing PVC from NON-PVC. Clustering uncovered potential multiple PVC classes, suggesting distinct outcomes. The proposed CNN achieved 96.21% accuracy, 98.09% recall, 94.74% precision, and 99.28% AUC on the GRADCAM dataset, showcasing promising results for PVC detection.

Future studies aim to incorporate surveys into the human-in-the-loop phase to guarantee the reinforcement learning of the proposed framework. By leveraging the insights and expertise of medical professionals through surveys, we not only improve the interpretability and transparency of AI models but also foster a synergistic partnership that ensures the development of more reliable and trustworthy medical SAI systems.

6. Acknowledgments

This study was carried out within the FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1555 11/10/2022, PE00000013).

References

- [1] H. Dogan, R. O. Dogan, A comprehensive review of computer-based techniques for r-peaks/qrs complex detection in ecg signal, *Archives of Computational Methods in Engineering* (2023) 1–19.
- [2] B. H. Van der Velden, H. J. Kuijff, K. G. Gilhuijs, M. A. Viergever, Explainable artificial intelligence (xai) in deep learning-based medical image analysis, *Medical Image Analysis* 79 (2022) 102470.
- [3] C.-S. Lin, Y.-C. F. Wang, Describe, spot and explain: Interpretable representation learning for discriminative visual reasoning, *IEEE Transactions on Image Processing* 32 (2023) 2481–2492. doi:10.1109/TIP.2023.3268001.
- [4] A. Saranya, R. Subhashini, A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends, *Decision analytics journal* (2023) 100230.
- [5] Y. Zhang, D. Hong, D. McClement, O. Oladosu, G. Pridham, G. Slaney, Grad-cam helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging, *Journal of Neuroscience Methods* 353 (2021) 109098.

- [6] R.-K. Sheu, M. S. Pardeshi, A survey on medical explainable ai (xai): Recent progress, explainability approach, human interaction and scoring system, *Sensors* 22 (2022) 8068.
- [7] P. Cheriyaath, F. He, I. Peters, X. Li, P. Alagona Jr, C. Wu, M. Pu, W. E. Cascio, D. Liao, Relation of atrial and/or ventricular premature complexes on a two-minute rhythm strip to the risk of sudden cardiac death (the atherosclerosis risk in communities [aric] study), *The American journal of cardiology* 107 (2011) 151–155.
- [8] H. L. Kennedy, J. A. Whitlock, M. K. Sprague, L. J. Kennedy, T. A. Buckingham, R. J. Goldberg, Long-term follow-up of asymptomatic healthy subjects with frequent and complex ventricular ectopy, *New England Journal of Medicine* 312 (1985) 193–197.
- [9] R. Scorza, M. Jonsson, L. Friberg, M. Rosenqvist, V. Frykman, Prognostic implication of premature ventricular contractions in patients without structural heart disease, *Europace* 25 (2023) 517–525.
- [10] F. Ataklte, S. Erqou, J. Laukkanen, S. Kaptoge, Meta-analysis of ventricular premature complexes and their relation to cardiac mortality in general populations, *The American journal of cardiology* 112 (2013) 1263–1270.
- [11] A. P. Maggioni, G. Zuanetti, M. G. Franzosi, F. Rovelli, E. Santoro, L. Staszewsky, L. Tavazzi, G. Tognoni, Prevalence and prognostic significance of ventricular arrhythmias after acute myocardial infarction in the fibrinolytic era. gissi-2 results., *Circulation* 87 (1993) 312–322.
- [12] R. J. Myerburg, K. M. Kessler, A. Castellanos, Sudden cardiac death: epidemiology, transient risk, and intervention assessment, *Annals of internal medicine* 119 (1993) 1187–1197.
- [13] G. B. Moody, R. G. Mark, The impact of the mit-bih arrhythmia database, *IEEE engineering in medicine and biology magazine* 20 (2001) 45–50.
- [14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [15] A. Ram, A. Sharma, A. S. Jalal, A. Agrawal, R. Singh, An enhanced density based spatial clustering of applications with noise, in: *2009 IEEE International Advance Computing Conference*, IEEE, 2009, pp. 1475–1478.
- [16] S. Vijayarangan, B. Murugesan, R. Vignesh, S. Preejith, J. Joseph, M. Sivaprakasam, Interpreting deep neural networks for single-lead ecg arrhythmia classification, in: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 300–303.
- [17] B. Murugesan, V. Ravichandran, K. Ram, S. Preejith, J. Joseph, S. M. Shankaranarayana, M. Sivaprakasam, Ecgnet: Deep network for arrhythmia classification, in: *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2018, pp. 1–6.
- [18] M. Ganeshkumar, V. Ravi, V. Sowmya, E. Gopalakrishnan, K. Soman, Explainable deep learning-based approach for multilabel classification of electrocardiogram, *IEEE Transactions on Engineering Management* (2021).
- [19] D. Zhang, S. Yang, X. Yuan, P. Zhang, Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram, *Iscience* 24 (2021).
- [20] K. H. Le, H. H. Pham, T. B. Nguyen, T. A. Nguyen, T. N. Thanh, C. D. Do, Lightx3ecg: A lightweight and explainable deep learning system for 3-lead electrocardiogram classification, *Biomedical Signal Processing and Control* 85 (2023) 104963.

- [21] S. Van Duijvenboden, J. Ramírez, M. Orini, N. Aung, S. E. Petersen, A. Doherty, A. Tinker, P. B. Munroe, P. D. Lambiase, Prognostic significance of different ventricular ectopic burdens during submaximal exercise in asymptomatic uk biobank subjects, *Circulation* 148 (2023) 1932–1944.
- [22] A. B. Halima, D. Kobaa, M. B. Halima, S. Ayachi, M. Belkhiria, H. Addala, Assessment of premature ventricular beats in athletes, in: *Annales de Cardiologie et d'Angéiologie*, volume 68, Elsevier, 2019, pp. 175–180.
- [23] F. D. Marco, L. D. Biasi, A. A. Citarella, M. Tucci, G. Tortora, Identification of morphological patterns for the detection of premature ventricular contractions, in: E. Banissi, A. Ursyn, M. W. M. Bannatyne, J. M. Pires, N. Datia, K. Nazemi, B. Kovalerchuk, R. Andonie, M. Nakayama, F. Sciarrone, W. Huang, Q. V. Nguyen, M. S. Mabakane, A. Rusu, M. Temperini, U. Cvek, M. Trutschl, H. Müller, H. Siirtola, W. L. Woo, R. Francese, V. Rossano, T. D. Mascio, F. Bouali, G. Venturini, S. Kernbach, D. Malandrino, R. Zaccagnino, J. J. Zhang, X. Yang, V. Geroimenko (Eds.), *26th International Conference Information Visualisation, IV 2022*, Vienna, Austria, July 19–22, 2022, IEEE, 2022, pp. 393–398. URL: <https://doi.org/10.1109/IV56949.2022.00071>. doi:10.1109/IV56949.2022.00071.
- [24] A. H. Kashou, P. A. Noseworthy, T. J. Beckman, N. S. Anavekar, M. W. Cullen, K. B. Angstman, B. J. Sandefur, B. P. Shapiro, B. W. Wiley, A. M. Kates, et al., Ecg interpretation proficiency of healthcare professionals, *Current problems in cardiology* (2023) 101924.
- [25] A. T. Jebb, V. Ng, L. Tay, A review of key likert scale development advances: 1995–2019, *Frontiers in psychology* 12 (2021) 637547.
- [26] P. J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of computational and applied mathematics* 20 (1987) 53–65.
- [27] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, *Communications in Statistics-theory and Methods* 3 (1974) 1–27.
- [28] D. L. Davies, D. W. Bouldin, A cluster separation measure, *IEEE transactions on pattern analysis and machine intelligence* (1979) 224–227.