X-Heart: A Human-Centered Framework for Explainable PVC Detection and Clinical Feedback⋆

Anna Maria De Roberto[1,†], Domenico Rossi[1,*,†], Alessia Auriemma Citarella[1,†], Fabiola De Marco[1,†], Luigi Di Biasi[1,†] and Genoveffa Tortora[1,†]

[1]*University of Salerno, Via Giovanni Paolo II, 132, Fisciano Italy*

Abstract

The integration of artificial intelligence (AI) in cardiology has shown remarkable potential for improving diagnostic accuracy. One of the biggest challenges in bringing AI tools into clinical practice is ensuring they are truly usable and user-friendly by medical professionals. This paper presents X-Heart, a human-centered framework for explainable detection of premature ventricular contractions that combines automated classification with transparent decision-making and clinician feedback. The system embeds a human-in-the-loop approach, inviting clinicians to validate predictions, rate confidence, and provide rationales through an intuitive interface designed to minimize cognitive overhead. By aligning AI outputs with clinical workflows and prioritizing interpretability with Grad-CAM-based visual explanations, X-Heart addresses key challenges in medical AI: transparency, clinician engagement, and mitigation of automation bias. This work underscores the importance of explainable AI and human-computer collaboration in advancing cardiac diagnostics, with implications for real-world deployment in healthcare settings.

Keywords

Human-in-the-loop, Premature Ventricular Contractions, eXplainable Artificial Intelligence

1. Introduction

Advancements in Artificial Intelligence (AI) technology are rapidly revolutionizing medicine field [1]. AI has the potential to significantly improve healthcare by optimizing resource allocation, reducing operational costs, and enhancing diagnostic accuracy. In particular, advanced techniques such as machine learning (ML) and deep learning (DL) enable early and precise detection of pathological conditions by identifying complex patterns in large-scale medical data [2, 3]. ML algorithms can analyze structured datasets to support clinical decision-making, while DL models, especially convolutional and recurrent neural networks (RNN), excel in processing unstructured data such as medical images and time-series signals like electrocardiograms (ECGs). These AI-driven tools not only assist in identifying early signs of disease but also contribute to developing personalized treatment plans, ultimately leading to more efficient and effective patient care. Despite the potential benefits and advantages that such computer systems can bring to clinical evaluation, healthcare personnel still have a certain mistrust, probably due to the inability to understand all the mechanisms underlying AI's exceptional performance. The aim of eXplainable Artificial Intelligence (XAI) is precisely to develop computer systems that can clearly explain their decision-making processes, using techniques such as feature importance analysis, decision trees, and visualization tools like heat maps and saliency maps [4]. A medical field of great interest for the use of AI and XAI is cardiological field and, in particular, in electrocardiographic evaluation of premature ventricular contractions (PVCs). 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 sinoatrial node that represent "cardiac pacemaker" and from that

EMPATH-IA: Workshop on EMpowering PAtients THrough AI, Multimedia, and Explainable HCI: Innovations in Personalized Healthcare, in conjunction with the 16th Biannual Conference of the Italian SIGCHI Chapter (CHITALY 2025), 6–10 October 2025, Salerno, Italy

*Corresponding author.

†These authors contributed equally.

aderoberto@unisa.it (A. M. D. Roberto); dorossi@unisa.it (D. Rossi); aauriemmacitarella@unisa.it (A. A. Citarella); fdemarco@unisa.it (F. D. Marco); ldibiasi@unisa.it (L. D. Biasi); ldibiasi@unisa.it (G. Tortora)

situs it spreads along cardiac atria [5]. Subsequently, this impulse travels down through the conduction pathways and causes ventricular depolarization and contraction essential for pump out blood. PVC is a heart arrythmia which belong to the group of ectopic arrythmias. In ectopic arrythmias ectopic beats originate outside sinoatrial node determining an anticipation of cardiac contraction (electrocardiographic changes) and frequently symptoms like palpitations, a feeling of skipped or extra heartbeats, dizziness or shorts of breaths [6].
DL has shown remarkable potential in classifying electrocardiograms and detecting various cardiac pathologies, with models such as Convolutional Neural Networks (CNNs), RNNs, and Transformers achieving high diagnostic accuracy [7, 8, 9, 10]. However, despite their performance, many existing DL frameworks lack explainable AI mechanisms, making it difficult for clinicians to trust and interpret their decisions. The absence of robust XAI in ECG classification raises concerns about transparency, accountability, and bias, particularly when diagnosing critical conditions like arrhythmias or myocardial infarctions. To bridge this gap, in this preliminary study, we propose a human-in-the-loop (HITL) approach, integrating clinician feedback with AI-driven analysis to improve diagnostic accuracy while maintaining interpretability and supporting in clinical decision-making. By incorporating principles from human-computer interaction (HCI), our framework X-Heart ensures that AI outputs are clinically actionable and promoting usability in real-world settings. For this purpose, we utilized the MIT-BIH Arrhythmia Database, training a CNN on selected ECG segments for PVC classification. Our methodology not only prioritizes model performance but also emphasizes XAI-driven transparency, enabling cardiologists to validate AI-generated insights and refine diagnostic precision through iterative collaboration.
The paper is structured as follows: Section 2 reviews existing AI approaches for ECG analysis and identifies gaps in explainability. Section 3 details our proposed HITL framework while section 4 analyzes clinical implications. Finally, Section 5 summarizes key contributions and future research directions.

2. Related Works

In recent years, the use of explainable frameworks and tools in biomedical signal analysis has become increasingly important, particularly in the context of automated cardiovascular disease detection. While high-performing models such as DL offer impressive accuracy, their "black-box" nature limits clinical trust and adoption. This section presents the proposed framework and the tools employed for early detection of cardiovascular abnormalities from ECG signals. [11] proposed CardioView, an explainable AI framework that incorporates the GRAD-CAM (Gradient-weighted Class Activation Mapping) technique to make the classification of PVCs more transparent and interpretable. CardioView achieves impressive performance metrics, including 96.21% accuracy, 98.09% recall, 94.74% precision, and an AUC of 99.28%, highlighting its robust classification capabilities. A key feature of CardioView is its ability to visualize which parts of the ECG waveform are most relevant to the model's decisions. This enables both clinicians and users to understand how the system differentiates between PVC and non-PVC signals, as well as among PVC subtypes [11]. [12] proposed a lightweight DL model combining CNN and Long Short-Term Memory (LSTM) layers for classifying eight types of cardiac arrhythmias and normal rhythms. ECG signals were preprocessed through resampling and baseline wander removal before being input into an 11-layer neural network. Using ECG data from the MIT-BIH arrhythmia and long-term AF databases, the model achieved a high mean diagnostic accuracy of 98.24%, outperforming many existing methods. To enhance transparency, SHapley Additive exPlanations (SHAP) were applied, enabling clinicians to understand which ECG features influenced the model's decisions. [13] presented a novel approach that extracts interpretable features using the Gini Index (GI) applied to the Choi-Williams time-frequency distribution (TFD) of QRS complexes. This marks the first use of GI in conjunction with nonlinear TFD for ECG analysis. Features were computed from one-minute segments over 30-minute ECG recordings, enabling both short- and long-term characterization. Evaluation on the MIT-BIH Arrhythmia and Fantasia datasets showed strong performance across eight ML models, with the top classifier achieving 100% sensitivity and over 94% accuracy. The emphasis on interpretability and low false negative rates supports clinical usability and integration into smart devices for continuous CVD monitoring, both online and offline [14]. [15] proposed an explainable heart disease risk prediction framework using binary classification, with a focus on model interpretability. The system is trained using Random Forest and XGBoost, with XGBoost selected for its superior performance (85–87% accuracy). To ensure transparency, the SHAP algorithm is integrated, offering clear visual insights into feature contributions. Interactive visualizations, including Plotly and 3D scatter plots, help users and clinicians understand how variables such as cholesterol, blood pressure, and chest pain types influence predictions, with techniques like waterfall plots enhancing interpretability. The model is deployed through a Streamlit application, providing a user-friendly interface for real-time risk assessment [16].

Although numerous AI models and diagnostic tools have been proposed in the literature for cardiovascular analysis, relatively few have specifically focused on the identification and classification of PVCs. Moreover, even fewer frameworks incorporate a HITL approach, which is essential for enhancing clinical reliability and interpretability. The absence of mechanisms that allow cardiologists to review, correct, or guide model predictions limits both transparency and acceptance in real-world settings. Integrating HITL within AI pipelines could not only improve diagnostic performance through expert reinforcement but also facilitate the development of trustworthy, explainable systems aligned with medical practice.

3. The X-Heart System Overview

The X-Heart framework (see Figure 1) is designed to support clinicians in the interpretation of PVCs by combining automatic classification, visual explanation, and interactive feedback. The framework operates by generating an initial diagnostic suggestion, followed by an explainable and interactive interface that enables the participation of clinicians. The diagnostic suggestion is generated from an AI algorithm, potentially based on ML techniques and previously trained on annotated ECG datasets. The framework is composed of three main steps:

- *ECG prediction*: Users can upload an ECG image and receive a preliminary classification result.
- *Interactive explainable visualization*: A visual heatmap highlights the ECG regions that most influenced the algorithm decision, improving interpretability.
- *Clinician feedback survey*: The interface invites clinicians to confirm or reject the prediction, rate their confidence, and optionally provide a brief rationale for their decision.

The X-Heart design follows well-established HCI principles, specifically transparency, interpretability, and minimal cognitive overhead, ensuring that the user remains in control of the diagnostic process while being supported, not replaced, by automation.

3.1. ECG prediction

Once a single-lead ECG signal is uploaded to the framework, the system automatically launches a processing pipeline and generates a concise prediction that can be human-interpretable. This initial output serves as a preliminary prompt in the decision-making process of clinicians, acting more like a triage recommendation than a conclusive diagnosis.

As illustrated in Figure 2, the prediction panel comprises two key components: *a binary classification label* providing a high-level summary of the system's interpretation and *an optional confidence indicator*, displayed as a percentage or as a visual bar, communicating the internal accuracy of the system intuitively.

Crucially, the system's output is framed as a suggestion, not a directive. Visual neutrality in the interface (e.g., restrained color schemes, absence of prescriptive language) helps mitigate automation bias, inviting users to assess the prediction critically rather than accept it by default. This preliminary output facilitates rapid clinical orientation and supports the decision of whether to proceed with further analysis through the explainability module. It functions as an entry point for a more comprehensive diagnostic process, presenting the AI result as a working hypothesis to be examined and refined.

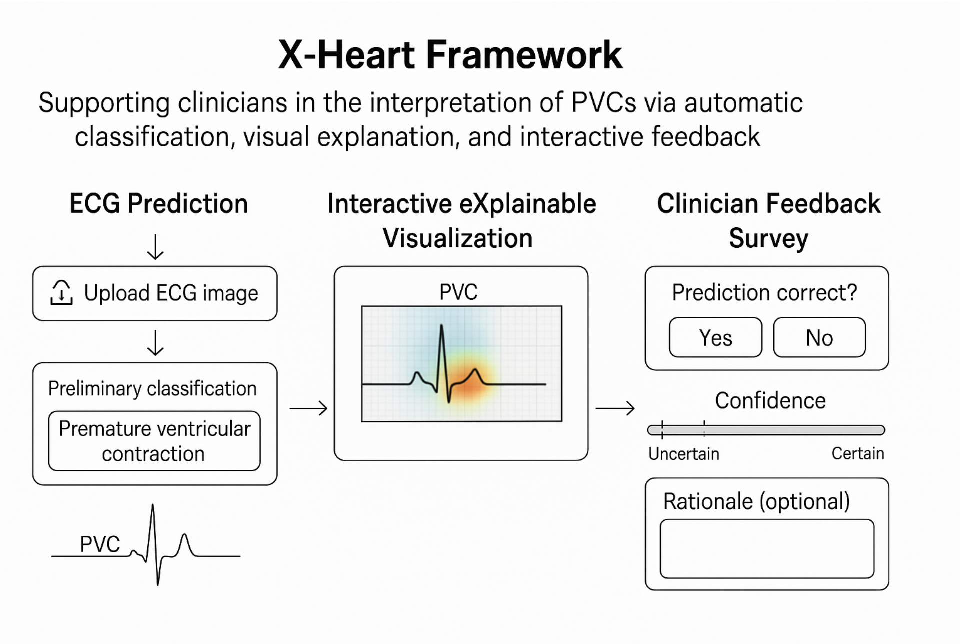

Figure 1: The X-Heart framework

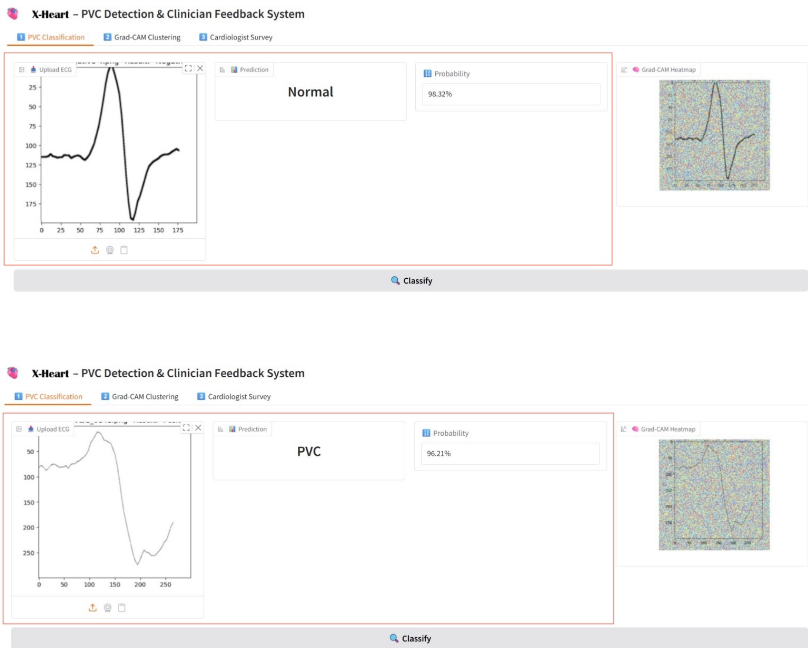

Figure 2: Prediction Panel Interface: The image above, from left to right, shows the loaded trace, with the prediction in the centre, which in this case is “normal”, and on the far right is the probability as a percentage, related to the prediction provided in the output. The second image shows the same example of the interface, but with a trace representing “PVC”.

3.2. Explainable Interface

Once the system’s initial prediction is displayed, clinicians can explore the internal reasoning of the model through a Grad-CAM-based visualization. This component is designed to bridge the cognitive gap between *what* the prediction of the model and *why* the decision made, promoting transparency and interpretability in clinical AI applications. The explanation is provided through a gradient-based class activation map (Grad-CAM) [17], which overlays a heatmap on the ECG waveform. This heatmap is color-coded (e.g., blue to red) to represent the relative importance of different temporal segments in influencing the output of the model. Regions with stronger activation reflect where the model focused its attention during decision-making.

This visualization offers insight into the internal logic of the model by highlighting waveform features, such as premature QRS complexes or abnormal repolarization patterns, that may have driven the classification. It also allows clinicians to compare the focus of the model with their own diagnostic reasoning, allowing a critical evaluation of AI's attention. In addition, it supports educational engagement, particularly for less experienced users, by illustrating how the model prioritizes parts of the ECG trace in ways that may confirm or challenge conventional clinical heuristics. Figure 3 illustrates two Grad-CAM overlays: one corresponding to a normal heartbeat, and the other to a PVC. The activation map in the PVC trace reveals a significant deviation in the QRS complex, aligning with the typical pathological patterns, whereas the normal beat shows more diffuse or baseline-aligned activation.

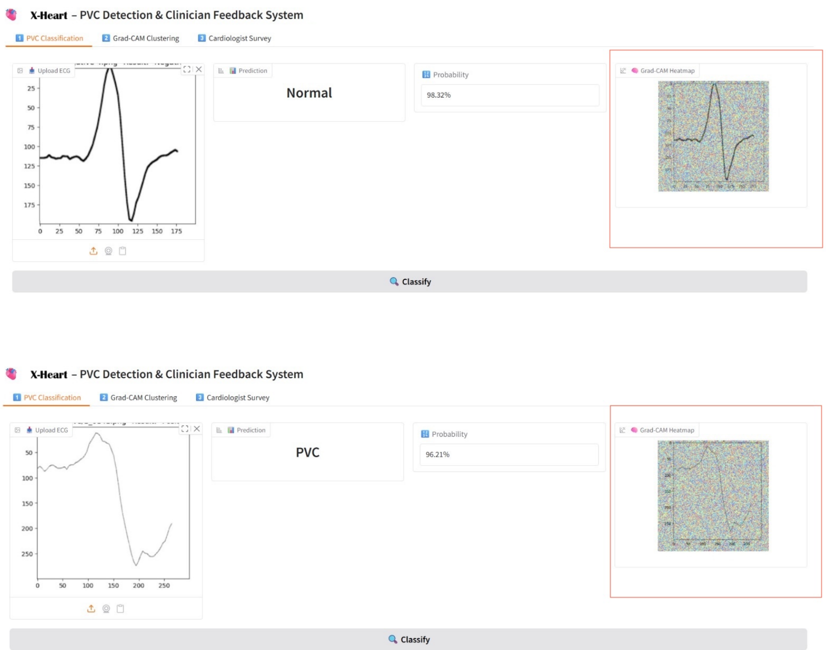

Figure 3: Grad-CAM Explanations for ECG Traces: In this image, the red box highlights the part of XAI showing the GradCam heatmap associated with the input trace.

Instead of relying on abstract or numerical indicators of "feature importance", the system provides spatial explanations that align more naturally with human visual reasoning. This design choice supports what Norman described as "external cognition", using visual representations to offload cognitive effort and promote perceptual reasoning [18].

The Grad-CAM visualization can also expose uncertainty or disagreement between the model and clinician. For instance, if the highlighted region does not correspond with the expected premature complex—or appears diffuse and unfocused—the clinician may question the model's confidence and choose to override the output.

3.3. Feedback Panel

The final stage of the X-Heart workflow invites the clinician to actively respond to the system's prediction and visual explanation. This step is not a mere formality but a core element of the HITL design philosophy: it transforms the user from a passive observer into a contributing agent in the decision-making process.

As shown in the Figure 4 the feedback panel is intentionally structured, yet flexible, allowing clinicians to express their evaluation through three complementary components.

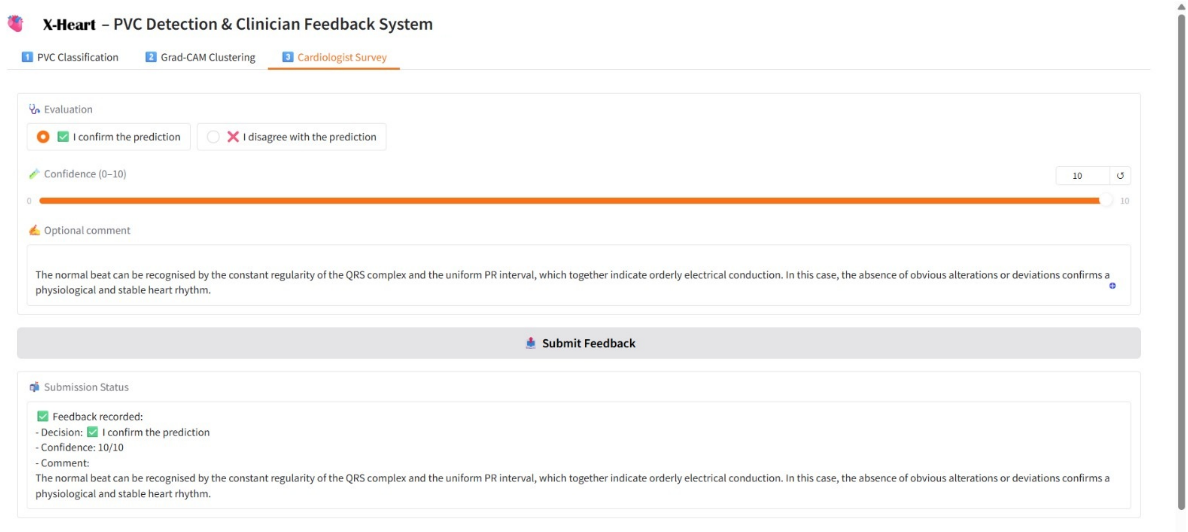

Figure 4: Clinician Feedback Panel: This image shows the feedback panel, which can be used by clinicians to evaluate the prediction made by the framework. It also provides the opportunity to give a confidence level and add a comment to the evaluation. When you click on 'Submit feedback', the framework automatically converts the clinician's choices into text.

3.3.1. Binary Diagnostic Judgment

The clinicians is first prompted to indicate if they agree or disagree with the classification of the system. This binary input serves as a clear validation signal and provides the foundation for measuring the concordance between human and AI decisions. Importantly, this interaction is structured in a non-coercive manner; the system neither implies nor promotes agreement, allowing the clinician to remain autonomous.

3.3.2. Confidence Scale

To capture the nuance behind the binary decision, the panel includes a confidence slider ranging from 0 (no confidence) to 10 (absolute certainty). This numerical input allows users to express medical uncertainty, which provides insight into how strongly they support their diagnostic choice. Over time, aggregated confidence data can help identify ambiguous ECG cases, highlight systematic discrepancies, or even flag regions of clinical disagreement.

3.4. Free-text Rationale

Finally, clinicians are optionally invited to provide an open comment explaining the reasoning behind their decision. This field enables the externalization of tacit knowledge, which is often difficult to capture through structured fields. Comments may include references to specific waveform anomalies, comparisons to known PVC patterns, signal quality concerns, or even clinical impressions based on history of the patient (if available).

4. Discussion

The X-Heart system represents a significant advancement in clinical decision support by addressing three critical challenges in medical AI: interpretability, clinician engagement, and diagnostic workflow integration. The system's three-stage workflow (prediction, explanation, feedback) mirrors the natural diagnostic reasoning process of clinicians, making it more likely to be adopted in practice. The preliminary classification serves as an effective triage mechanism, particularly valuable in high-volume clinical settings. Our neutral interface design successfully mitigates automation bias, a common problem in clinical decision support systems where users may over-rely on algorithmic outputs. The heatmap visualization provides immediate value by highlighting ECG features that may require closer inspection, potentially reducing interpretation time for complex cases.

In addition, the feedback mechanism offers particular clinical benefits by creating a structured process for documenting diagnostic disagreements between clinicians and AI, by capturing valuable clinical reasoning through free-text rationales and generating data that can identify systematic patterns in AI errors or clinician uncertainties. Specifically, our Grad-CAM implementation provides several advantages over traditional feature importance methods:

- the spatial heatmap aligns with existing visual interpretation patterns of the clinicians;
- it enables direct comparison between machine attention and human diagnostic heuristics;
- the visualization can reveal model limitations when activation patterns appear diffuse or misaligned with clinically significant features.

The system successfully implements the principles of external cognition of Norman by transforming abstract model decisions into visual representations that reduce cognitive load. This is particularly valuable for training scenarios, where the heatmaps can help less experienced clinicians develop their pattern recognition skills. By keeping the interface simple and the explanations visual, we've created a system that puts medical expertise first, with technology serving as a transparent aid rather than a barrier. The true test of this design comes when busy clinicians can use the system effectively without special training or technical support - and that's exactly the experience we've worked to create. After all, the value of AI in medicine isn't in its complexity, but in its ability to make complex information more accessible to those who care for patients.

5. Conclusion

This work presents X-Heart, a human-centered interface for PVC detection that combines machine-generated explanations with structured clinician feedback. Rather than focusing on classification performance, the framework emphasizes transparency, traceability, and collaboration. Through its three core modules—automated prediction, visual explanation, and interactive feedback, X-Heart suggests how AI tools can become more interpretable and clinically actionable. Although real-world validation remains essential, the system shows the potential of interactive and explainable AI to promote trust, clinician involvement, and improved diagnostic accuracy in cardiology.

Future directions include expanding the system for multi-class arrhythmia detection, investigating alternative explanation methods, and embedding the platform into clinical workflows for observational studies.

Declaration on Generative AI

The authors affirm that no generative artificial intelligence tools were used in the preparation of this work.

References

[1] A. G. Fraser, E. Biasin, B. Bijnens, N. Bruining, E. G. Caiani, K. Cobbaert, R. H. Davies, S. H. Gilbert, L. Hovestadt, E. Kamenjasevic, et al., Artificial intelligence in medical device software and high-risk medical devices–a review of definitions, expert recommendations and regulatory initiatives, Expert Review of Medical Devices 20 (2023) 467–491.

[2] S. Das, S. P. Nayak, B. Sahoo, S. C. Nayak, Machine learning in healthcare analytics: a state-of-the-art review, Archives of Computational Methods in Engineering 31 (2024) 3923–3962.

[3] A. Rahman, T. Debnath, D. Kundu, M. S. I. Khan, A. A. Aishi, S. Sazzad, M. Sayduzzaman, S. S. Band, Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities, AIMS Public Health 11 (2024) 58.

[4] Q. Sun, A. Akman, B. W. Schuller, Explainable artificial intelligence for medical applications: A review, ACM Transactions on Computing for Healthcare 6 (2025) 1–31.

[5] 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.

[6] I. S. Abdalla, R. J. Prineas, J. D. Neaton, D. R. Jacobs Jr, R. S. Crow, Relation between ventricular premature complexes and sudden cardiac death in apparently healthy men, The American journal of cardiology 60 (1987) 1036–1042.

[7] F. De Marco, F. Ferrucci, M. Risi, G. Tortora, Classification of qrs complexes to detect premature ventricular contraction using machine learning techniques, Plos one 17 (2022) e0268555.

[8] F. De Marco, L. Di Biasi, A. Auriemma Citarella, G. Tortora, Improving pvc detection in ecg signals: A recurrent neural network approach, in: Italian Workshop on Artificial Life and Evolutionary Computation, Springer, 2023, pp. 256–267.

[9] D. B. Olawade, N. Aderinto, G. Olatunji, E. Kokori, A. C. David-Olawade, M. Hadi, Advancements and applications of artificial intelligence in cardiology: Current trends and future prospects, Journal of Medicine, Surgery, and Public Health (2024) 100109.

[10] K. Niroda, C. Drudi, J. Byers, J. Johnson, G. Cozzi, L. A. Celi, H. Khraishah, Artificial intelligence in cardiology: insights from a multidisciplinary perspective, Journal of the Society for Cardiovascular Angiography & Interventions 4 (2025) 102612.

[11] G. Arienzo, A. A. Citarella, F. De Marco, A. M. De Roberto, L. Di Biasi, R. Francese, G. Tortora, Cardioview: a framework for detection premature ventricular contractions with explainable artificial intelligence, in: INI-DH 2024: Workshop on Innovative Interfaces in Digital Healthcare, in conjunction with International Conference on Advanced Visual Interfaces, 2024, pp. 3–7.

[12] N. Alamatsaz, L. Tabatabaei, M. Yazdchi, H. Payan, N. Alamatsaz, F. Nasimi, A lightweight hybrid cnn-lstm explainable model for ecg-based arrhythmia detection, Biomedical Signal Processing and Control 90 (2024) 105884. URL: https://www.sciencedirect.com/science/article/pii/S1746809423013174. doi:https://doi.org/10.1016/j.bspc.2023.105884.

[13] M. Aashiq, S. J. Hashim, F. Zaman Rokhani, M. Hanafi, A. Faeq Hussein, Tfdginixml: A novel explainable machine learning framework for early detection of cardiac abnormalities based on nonlinear time-frequency distribution gini index features, IEEE Access 13 (2025) 86666–86685. doi:10.1109/ACCESS.2025.3566094.

[14] M. Aashiq, S. J. Hashim, F. Z. Rokhani, M. Hanafi, A. F. Hussein, Tfdginixml: A novel explainable machine learning framework for early detection of cardiac abnormalities based on nonlinear time-frequency distribution gini index features, IEEE Access (2025).

[15] A. Sethi, S. Dharmavaram, S. K. Somasundaram, Explainable artificial intelligence (xai) approach to heart disease prediction, in: 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), 2024, pp. 1–6. doi:10.1109/AIIoT58432.2024.10574635.

[16] V. Aarthi, et al., Explainable ai for cardiovascular health: A shap-based framework, in: 2025 International Conference on Pervasive Computational Technologies (ICPCT), IEEE, 2025, pp. 353–358.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, International journal of computer vision 128 (2020) 336–359.

[18] D. A. Norman, Some observations on mental models. mental models, Lawrence Erlbaum 99 (1983) 129.