CureNet: Improving Explainability of AI Diagnosis Using Custom Large Language Models

Subhash Khambampati[*], Sushanth Dondapati [*], Tejo Vardhan Kattamuri[*], Rahul Krishnan Pathinarupothi[†]
[*]School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, India
[*]amenu2aie21026@am.students.amrita.edu
[†]Center for Wireless Networks & Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri, India
[†]rahulkrishnan@am.amrita.edu

***Abstract*—Cardiac abnormalities are a leading global cause of fatalities, necessitating precise diagnostic decisions, particularly in the context of cardiac health. Explainable Artificial Intelligence (XAI) has emerged as a valuable tool to interpret and clarify opaque algorithms used in Machine Learning (ML) and Deep Learning (DL). In our research, we emphasize the need for DL tools that are both precise and explainable, especially in domains where human expertise is crucial. We introduce a novel clinician-in-the-loop, prompt-based XAI tool that integrates multiple DL models with different XAI techniques, offering both accuracy and transparency. Our DL model, trained on a comprehensive ECG dataset, achieves high precision. We also employ XAI techniques, including Integrated Gradients, Layer-wise Relevance Propagation (LRP), and DeepLift, to generate heatmaps that highlight the model's decision-making regions within ECG signals. These visual outputs enhance the interpretability of the diagnostic process. Additionally, we present a chatbot that uses a pre-trained language model and an innovative image retrieval feature, offering an intelligent conversational agent for medical inquiries.**

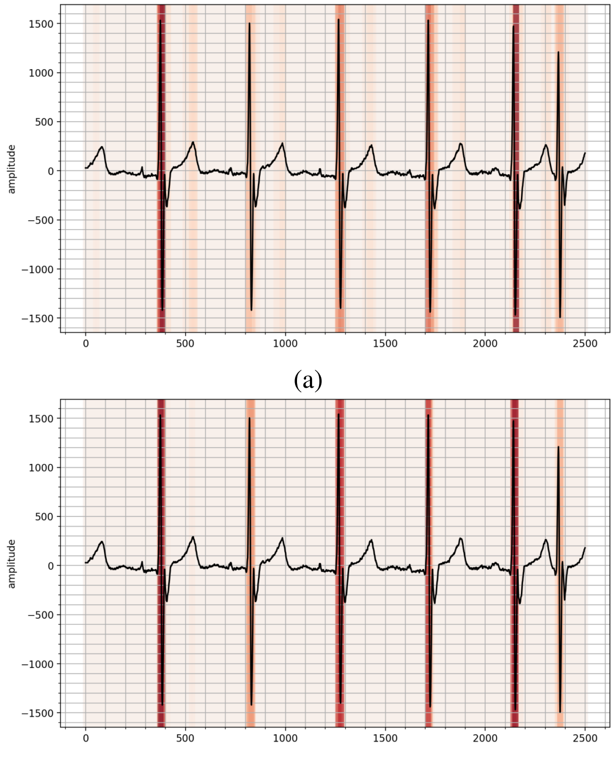
***Index Terms*—Cardiac abnormalities, XAI, Integrated Gradients, Layer-wise Relevance Propogation, DeepLift, Language Model**

I. Introduction

Cardiac abnormalities currently rank as the leading global cause of fatalities. In the field of automated AI enhanced medical diagnosis, especially within the context of cardiac health, precision of diagnostic decisions is along with its interpretability. The techniques of eXplainable Artificial Intelligence (XAI) is employed to interpret and infer from often-opaque algorithms used in Machine Learning (ML) and Deep Learning (DL).

The trade-off between precision and explaining ability presents a dilemma in AI, with extensive impacts in fields of healthcare and engineering. High-precision models are of limited use if their decision-making processes are opaque. Our work seeks to navigate this trade-off, offering solutions that provide both accuracy and transparency.

XAI tools have been applied to images, texts, and signals. However, their applications in bio-signals have presented a new set of challenges. These XAI tools offer only a preliminary insight on the regions of interest in signals, which may not provide clinicians with enough information on how a specific region is selected by the model for disease inference. Additionally, each of these XAI models has been observed to produce varying focus regions, where the intensity of colors represents the regions within the signal to which the model directs its attention to predict heart abnormality (see Figure 1). Therefore, to achieve better clinical acceptance of machine learning models, these XAI outputs should ideally be amenable to be more interactive with a clinician in the loop system.



(a)

(b)

Fig. 1. (a) Integrated Gradient and (b) Deep Lift heat maps; XAI techniques that highlight the prediction-concentrated regions of an ECG signal.

We propose a novel clinician-in-the-loop, prompt-based XAI tool that integrates multiple DL models with different XAI techniques, coupled with a fine-tuned LLM system.

We trained our DL model using widely available ECG datasets, we employed the Falcon 7B pre-trained model for conversational LLM, along with a fine-tuned LLM specifically for heart disease diagnosis, using medical files and textbooks as training data. We then integrated it with an automated prompt-based query system to retrieve relevant images through medical image scraping.

Instead of converting the image into a vector residing in a latent space that requires training with LLM, our methodology relies on image captions that precisely represent the clinical description of the image. By comparing the topic in the prompt to these image captions, we demonstrate a computationally simpler method for retrieving images, although it may not achieve the accuracy provided by the 50 Billion+ parameter models.

II. Related Works

Within the domain of cardiac abnormality prediction, several studies have leveraged machine learning models. Jia et al., 2020 utilized the PhysioNet/Computing in Cardiology Challenge 2020 dataset for ECG classification. Data preprocessing involved resampling to 500Hz, truncating to 10 seconds, and no signal filtering. The main model was a 34-layer ResNet with skip connections and Squeeze-and-Excitation layers. Demographic data was integrated, and a threshold was applied to predict ECG classes. Baseline models included ResNet50 and LightGBM classifiers. The models showed great results, the AUC scores of the four models are as follows Rsenet50: 0.938, Rsenet50+lgb: 0.951, Resnet34: 0.959, and SE-Resnet34: 0.967. This research's approach and models offer valuable insights for future ECG classification studies [1]. There are more different kinds of paper where it showed similar or greater results [3]–[5].

Nils et al., addressed the challenge of detecting and interpreting myocardial infarction using convolutional neural networks. In this context, the authors introduced a range of attribution techniques, with a specific reference to Gradient*Input and integrated gradients. Notably, the experimental results presented in the paper consistently demonstrated the accurate and unambiguous identification of ST-segment elevation. This is particularly significant as ST-segment elevation is a critical hallmark on the electrocardiogram (ECG) for diagnosing myocardial infarction, and serves as a diagnostic criterion for ST-segment elevation myocardial infarction [2].

Shrikumar et al., described a method called DeepLIFT that interprets neural network behavior. It attributes output differences to input differences from a reference input. This method uses contribution scores, multipliers, a chain rule, and reference inputs to analyze model decisions. DeepLIFT offers advantages over gradient-based methods, handling situations where gradients are zero and avoiding discontinuities due to bias terms. It also introduces the revealcancel rule for improved shapely value approximations [6].

The work by Bach et al., introduced Layer-wise Relevance Propagation (LRP) as an approach to clarify which pixels in an image contribute to a classification decision [7]. Initially applied to neural networks, LRP was later extended to include other models such as bag-of-word models by Csurka et al., and Sande et al., [8], [9]. Subsequently, Bach et al., further extended its application to Fisher vectors [10].

Binder et al., suggest an expansion of LRP to encompass neural networks featuring nonlinearities that deviate from the typical neural network structure, including local renormalization layers, which are challenging for standard LRP techniques [11]. The proposed method relies on applying a first-order (or higher) Taylor expansion. The focus is on a classification scenario with real-valued outputs, where a classifier, denoted as f, maps an input space X to real numbers, and f(x) greater than 0 signifies the presence of a class.

III. CureNet Architecture

The architecture used for this model is shown in Figure 2. In this architecture, ECG data is used to train a neural network model, which is then saved as a pre-trained model. When new test data is uploaded, this pre-trained model is utilized to predict the diseases or abnormalities in the patient.

Subsequently, the signal that the model predicts is subjected to three distinct eXplainable Artificial Intelligence (XAI) techniques, namely Integrated Gradients, LRP, and Deep Lift. These XAI techniques are applied to provide insight into the model's decision-making process.

The outcome of these techniques is the generation of heatmaps, which visually represent and highlight the regions in the ECG signal that are most relevant to the prediction made by the model. These heatmaps serve as a valuable output, offering a clear and interpretable understanding of the model's assessment and contributing to the interpretability of the diagnostic process.

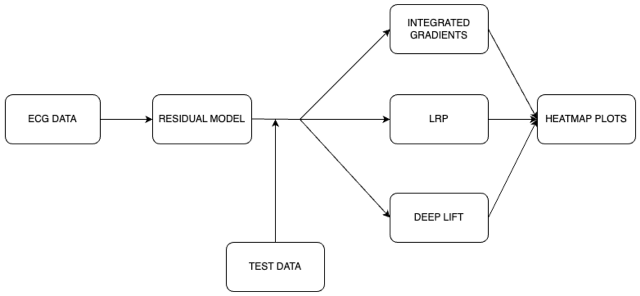


Fig. 2. Flowchart illustrating the XAI model for ECG analysis: Training the residual model with ECG data, testing with patient ECG signals, and employing explainable models (Integrated Gradients, LRP, and DeepLift) to generate prediction-concentrated heatmaps.

The following is a description of Figure 3. The PDF Loader component loads the contents of the PDFs comprising various files. The character text splitter comes into action, by breaking down the pdf content into individual characters. The vector representations of these characters are stored in the vector store, and subsequently, the Falcon 7B model transforms these vectors back into text. The final step involves the Interface, which displays the reconstructed text to the user. Furthermore, if the user makes an inquiry regarding the images pertaining to heart diseases, the model promptly responds by accessing the relevant image repository and providing appropriate information.

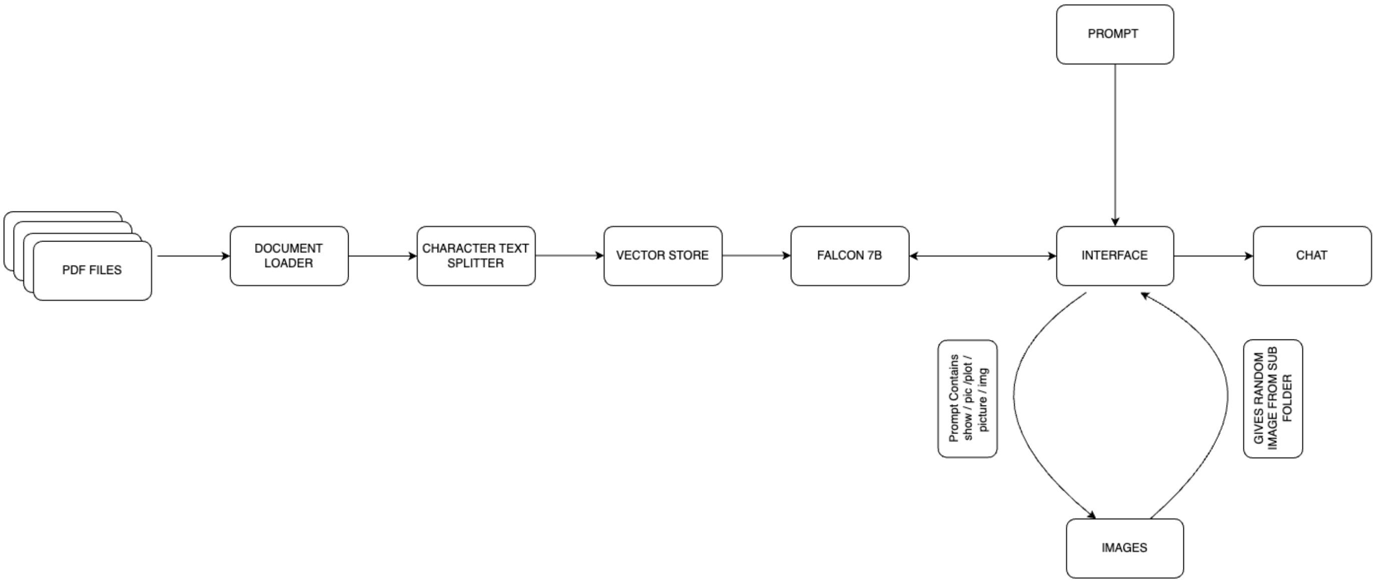


Fig. 3. Flowchart depicting the architecture of the medical data chatbot: PDF files loaded via document loader, character splitting with a text splitter, storage in a vector database, training using the 'Falcon 7B' pre-trained LLM model, and interaction with the user or clinician through a user interface, providing contextually relevant textual or visual responses based on user prompts.

IV. Deep Learning Model for Cardiac Abnormality Detection

We developed a specialized Convolutional Neural Networks (CNN) - ConvNet model for the classification of Electrocardiogram (ECG) signals, with a primary focus on diagnosing heart-related conditions. To enhance training stability and speed up convergence, we also integrated Batch Normalization layers after each convolutional layer, and ReLU activation functions are used throughout the network to maintain strong information flow during training. In the final convolutional layer, we employed a Global Average Pooling to condense spatial information, making the model more computationally lightweight. For multi-label classification tasks, the output layer utilizes a softmax activation function, and binary cross-entropy loss was used as a suitable choice for ECG signal classification.

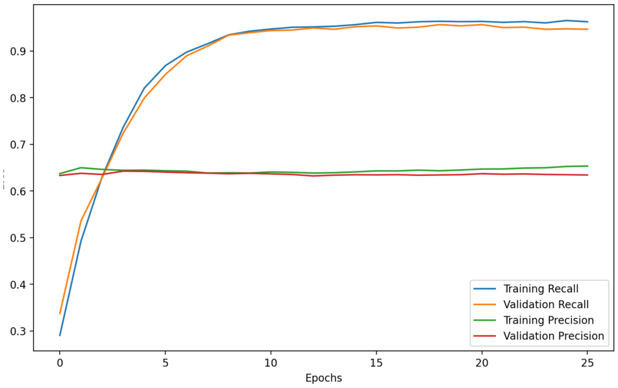


Fig. 4. The precision and recall of the trained model.

To ensure efficient model training, we optimize the network with Stochastic Gradient Descent (SGD) using a learning rate of 0.001. Our model is trained over 100 epochs with a batch size of 32, and we carefully considered class weights to address data imbalances. The model is trained on 12 lead ECG signal data of patients from different hospitals.

The dataset for this research comprises ECG (Electrocardiography) recordings sourced from various databases, including:

1) **CPSC Database and CPSC-Extra Database:** Public and unused data from the China Physiological Signal Challenge in 2018 (CPSC2018), with over 10,000 recordings [13].
2) **INCART Database:** A dataset from St Petersburg INCART 12-lead Arrhythmia Database, consisting of 74 annotated recordings [16].
3) **PTB and PTB-XL Database:** The Physikalisch Technische Bundesanstalt (PTB) provides two public databases, PTB Diagnostic ECG Database (516 records) and PTB-XL (21,837 clinical 12-lead ECGs) [14], [15].
4) **Georgia 12-lead ECG Challenge (G12EC) Database:** A database representing the Southeastern United States, with 10,344 12-lead ECGs [16].
5) **Undisclosed Database:** An undisclosed American database containing 10,000 ECGs, solely used for test data.

The ConvNet model achieved a precision score of 0.67 and a recall score of 0.43 on the test set. Figure 4 represents the same metrics on the training and validation sets with respect to epochs.

V. eXplainable AI for Bio-signals

Explainability is a cornerstone of our research, and in this section, we delve into the eXplainable AI (XAI) techniques applied to interpret our DL models. LRP, Integrated

Gradients, and Deep Lift are explored in detail, providing insights into how these methods generate heat maps that elucidate the decision-making process of our models. The comparative analysis of these XAI techniques sheds light on their respective strengths and limitations in the context of ECG signal analysis.The below explanations and formulae are referred from Żyliński and Ksenia [4] [12].

A. *Integrated Gradients:*

Integrated Gradients (IG) is a method for attributing the prediction of a DL model to its input features. It provides a clear and interpretable way to understand the contribution of each feature to the model's output as shown in Figure 5. The integrated gradients for a particular input feature x_i is calculated as follows:

$$IG_i(f) = (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial f(\mathbf{x}' + \alpha \times (\mathbf{x} - \mathbf{x}'))}{\partial x_i} \, d\alpha$$

where f is the model's prediction function, $\mathbf{x}$ is the input, and $\mathbf{x}'$ is the baseline or reference input.

Integrated Gradients computes the integral of the model's gradient with respect to the input along a straight path from a baseline input to the actual input. This integral represents the accumulated change in the model's output concerning the input feature.

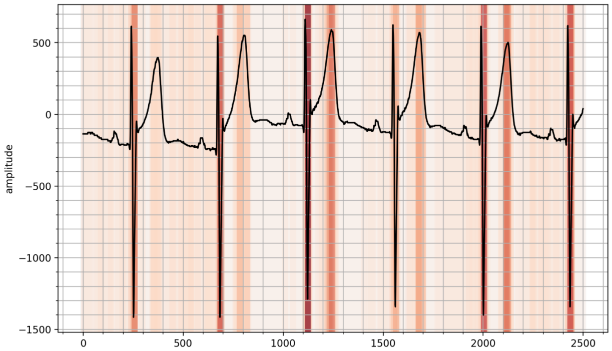


Fig. 5. Integrated Gradient heat map, an XAI technique that highlights the prediction-concentrated regions of an ECG signal.

B. *Layer-wise Relevance Propagation (LRP)*

Layer-wise Relevance Propagation (LRP) is a powerful technique in the realm of interpretability for deep neural networks. Its goal is to attribute relevance or importance scores to each input feature, helping to understand which features contribute the most to a model's predictions as shown in Figure 6. LRP enables a more transparent and interpretable view of complex models, making it a valuable tool in various applications, from image recognition to natural language processing.

LRP Relevance Score for a Neuron: Let $R_i^{(L)}$ be the relevance score for the i-th neuron in layer L, and $z_i^{(L)}$ represent the pre-activation of that neuron. The relevance score can be calculated as follows:

$$R_i^{(L)} = \frac{z_i^{(L)}}{\sum_j z_j^{(L)}} \times R^{(L+1)}$$

This formula distributes the relevance of the neuron in layer $L+1$ proportionally to the contribution of the neuron in layer L to its output.

LRP Relevance Score for an Input Feature: To compute the relevance of an input feature x_i, given the relevance scores $R_i^{(L)}$ for neurons in the last hidden layer, the formula is:

$$R_i = \sum_j \frac{x_i w_{ij}}{\sum_k x_k w_{kj}} \times R_j^{(L)}$$

Here, w_{ij} is the weight connecting the i-th input feature to the j-th neuron in the last hidden layer.

LRP for Neurons: LRP works by distributing the relevance of the output across the neurons in the network. The relevance of a neuron is proportional to its contribution to the output. The ratio $\frac{z_i^{(L)}}{\sum_j z_j^{(L)}}$ normalizes the importance of each neuron in layer L, and $R^{(L+1)}$ is the relevance received from the layer above.

LRP for Input Features: The relevance of an input feature is computed by considering the contribution of that feature to the activation of neurons in the last hidden layer. The term $\frac{x_i w_{ij}}{\sum_k x_k w_{kj}}$ represents the relevance of the input feature x_i in activating the j-th neuron. This is then multiplied by the relevance of the neuron $R_j^{(L)}$, distributing the importance back to the input features.

LRP's strength lies in its ability to provide a detailed and fine-grained understanding of the model's decision process, allowing practitioners to identify the most influential features and enhance the interpretability of deep neural networks.

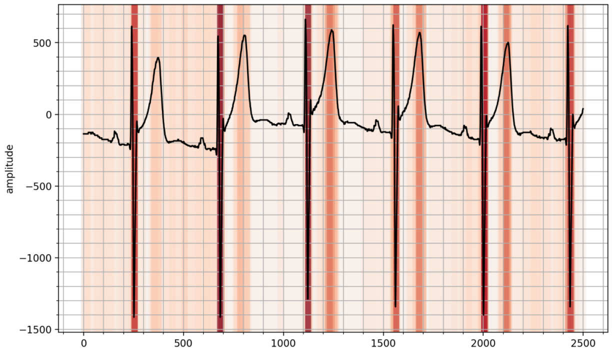


Fig. 6. LRP heat map, an XAI technique that highlights the prediction-concentrated regions of an ECG signal.

C. *DeepLift*

The DeepLIFT method flows through the neural network in the same way as LRP. It provides an interpretable way to understand the prediction-concentration regions of the ECG signals as shown in Figure 7. Each neuron i is assigned an attribution, which is the relative effect of activation of a given unit at the input x of the network compared to activation at some basic input, or baseline $\overline{x}$. The control values $\overline{z}_j$ for all hidden units are determined by passing through the network directly, using $\overline{x}$ the baseline as input and recording the activation of each neuron. The baseline is often selected to be zero. The propagation of relevance for this method is:

$$r_i{}^{(L)} = \begin{cases} S_i(x) - S_i(\bar{x}), & \text{if } i \text{ is the target neuron} \\ 0, & \text{else} \end{cases}$$

$$r_i{}^{(l)} = \sum_j \frac{Z_{ji} - \bar{Z_{ji}}}{\sum_i Z'_{ji} - \sum_i \bar{Z'_{ji}}}$$

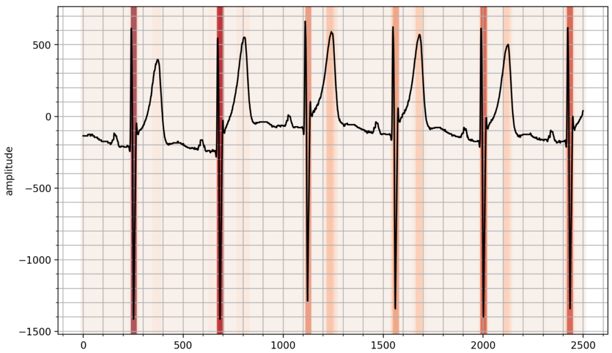


Fig. 7. Deep Lift heat map, an XAI technique that highlights the prediction-concentrated regions of an ECG signal.

VI. ChatBot

This section discusses the development of a conversational chatbot that is based on the Falcon 7B pre-trained language model from Hugging Face. The model is trained on a wide range of literature which includes topics such as cardiac diseases, Deep Learning, and XAI. The chatbot engages users in informative discussions across these domains and thus serves as an intelligent conversational agent. An innovative inclusion to its functionality involves an image retrieval feature; the method involves web scrapping images from the various trusted websites relevant to cardiac abnormalities. These images are organized systematically into folders, each labeled with the corresponding heart disease.

When requested by the user for an image, the chatbot displays the relevant image retrieved from the specific folder along with the textual details. This combination of a pre-trained language model and targeted image retrieval provides a comprehensive understanding of medical topics. This implementation smoothly combines the chatbot's ability to address user inquiries with visual assistance, thereby enhancing explainability.

VII. System Analysis

Working of our prototype model is represented in Figure 8.

In Figure 8(a), we begin by selecting a model capable of providing insights into various diseases based on a in input ECG signal. This model generates a bar chart, offering a visual representation of its predictions regarding the likelihood of different diseases (Prediction probability). Furthermore, it informs us about the disease it considers most probable and the degree of certainty it has in that prediction.

In Figure 8(b), we introduce the flexibility of interpreting the model's responses. Picture a graph illustrating heart rate, where different colors are used. Here, the intensity of the color red signifies the model's focus on specific regions of the graph. This feature aids us in comprehending the model's thought process and enables us to enhance its performance.

For healthcare practitioners, this colorful graph serves as a valuable tool. It directs their attention to the segments of the heart rate graph where the model detects potential issues, thus assisting them in making well-informed decisions about patient care.

This vibrant graph not only facilitates our understanding of machine learning models designed for predicting heart-related issues but also plays a crucial role in their improvement. It contributes to the development of more accurate predictive models and enhances the quality of care provided to patients experiencing heart health concerns.

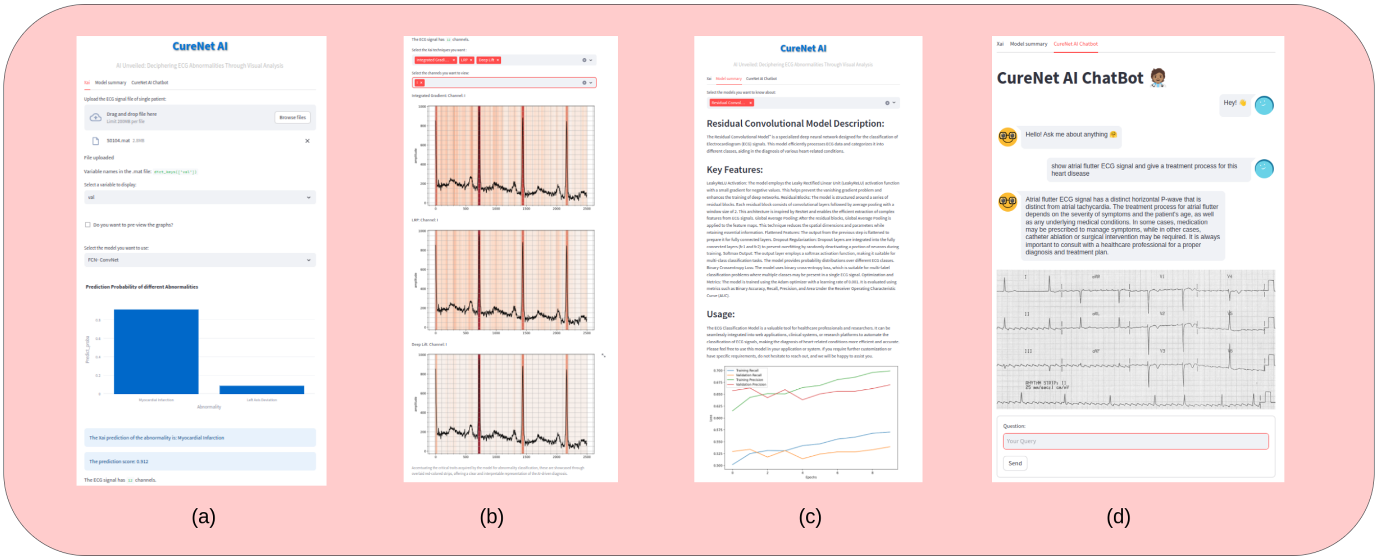
In Figure 8(c), we can know the summary of the model. This summary includes information about the model's design. This knowledge is crucial for ensuring the model's reliability and understanding its limitations.

In Figure 8(d), we delve into how the chatbot is used. Imagine a scenario where a doctor is using this chatbot. The doctor has a question about ”atrial flutter”, which is a heart condition. The chatbot is designed to answer medical questions, so it provides information about atrial flutter. It might explain what atrial flutter is, its symptoms, potential causes, and treatment options. Essentially, it acts as a virtual assistant that can provide medical information quickly, which can be very helpful for healthcare professionals seeking information on various health topics.

VIII. Conclusion

In this study, we have outlined the importance of developing precise and interpretable tools for cardiac health diagnostics. The trade-off between precision and explainability in AI poses significant challenges, particularly in domains where human expertise is indispensable. Our approach addresses this challenge by introducing a clinician-in-the-loop, prompt-based XAI tool that combines multiple DL models with various XAI techniques. This innovative approach offers both accuracy and transparency in diagnostic decisions. Our DL model, trained on a diverse ECG dataset, demonstrates high precision in cardiac abnormality prediction. We further enhance the interpretability of our models by employing XAI techniques such as Integrated Gradients, LRP, and DeepLift, which generate heatmaps highlighting the model's decision-making regions within ECG signals.

Additionally, we introduce a chatbot that leverages a pre-trained language model to provide comprehensive responses to medical inquiries. The chatbot's integration with an image retrieval feature enhances the user experience by providing visual assistance in addition to textual responses. Our research takes a significant step toward bridging the gap between precision and explainability in AI-driven cardiac health diagnostics. Our combination of DL models, XAI techniques, and a chatbot



CureNet AI

AI Unveiled: Deciphering ECG Abnormalities Through Visual Analysis

Upload the ECG signal file of single patient:

Drag and drop file here
Limit 200MB per file

S0104.mat 2.8MB

File uploaded

Variable names in the .mat file: `dict_keys(['val'])`

Select a variable to display: val

□ Do you want to pre-view the graphs?

Select the model you want to use: FCN- ConvNet

Prediction Probability of different Abnormalities

The Xai prediction of the abnormality is: Myocardial Infarction

The prediction score: 0.912

The ECG signal has 12 channels.

The ECG signal has 12 channels.

Select the Xai techniques you want : Integrated Gradi... × LRP × Deep Lift ×

Select the channels you want to view: I ×

Integrated Gradient: Channel: I

LRP: Channel: I

Deep Lift: Channel: I

Accentuating the critical traits acquired by the model for abnormality classification, these are showcased through overlaid red-colored strips, offering a clear and interpretable representation of the AI-driven diagnosis.

CureNet AI

AI Unveiled: Deciphering ECG Abnormalities Through Visual Analysis

Select the models you want to know about: Residual Convol... ×

Residual Convolutional Model Description:

The Residual Convolutional Model" is a specialized deep neural network designed for the classification of Electrocardiogram (ECG) signals. This model efficiently processes ECG data and categorizes it into different classes, aiding in the diagnosis of various heart-related conditions.

Key Features:

LeakyReLU Activation: The model employs the Leaky Rectified Linear Unit (LeakyReLU) activation function with a small gradient for negative values. This helps prevent the vanishing gradient problem and enhances the training of deep networks. Residual Blocks: The model is structured around a series of residual blocks. Each residual block consists of convolutional layers followed by average pooling with a window size of 2. This architecture is inspired by ResNet and enables the efficient extraction of complex features from ECG signals. Global Average Pooling: After the residual blocks, Global Average Pooling is applied to the feature maps. This technique reduces the spatial dimensions and parameters while retaining essential information. Flattened Features: The output from the previous step is flattened to prepare it for fully connected layers. Dropout Regularization: Dropout layers are integrated into the fully connected layers (fc1 and fc2) to prevent overfitting by randomly deactivating a portion of neurons during training. Softmax Output: The output layer employs a softmax activation function, making it suitable for multi-class classification tasks. The model provides probability distributions over different ECG classes. Binary Crossentropy Loss: The model uses binary cross-entropy loss, which is suitable for multi-label classification problems where multiple classes may be present in a single ECG signal. Optimization and Metrics: The model is trained using the Adam optimizer with a learning rate of 0.001. It is evaluated using metrics such as Binary Accuracy, Recall, Precision, and Area Under the Receiver Operating Characteristic Curve (AUC).

Usage:

The ECG Classification Model is a valuable tool for healthcare professionals and researchers. It can be seamlessly integrated into web applications, clinical systems, or research platforms to automate the classification of ECG signals, making the diagnosis of heart-related conditions more efficient and accurate. Please feel free to use this model in your application or system. If you require further customization or have specific requirements, do not hesitate to reach out, and we will be happy to assist you.

CureNet AI ChatBot

Hey!

Hello! Ask me about anything

show atrial flutter ECG signal and give a treatment process for this heart disease

Atrial flutter ECG signal has a distinct horizontal P-wave that is distinct from atrial tachycardia. The treatment process for atrial flutter depends on the severity of symptoms and the patient's age, as well as any underlying medical conditions. In some cases, medication may be prescribed to manage symptoms, while in other cases, catheter ablation or surgical intervention may be required. It is always important to consult with a healthcare professional for a proper diagnosis and treatment plan.

RHYTHM STRIP: II
25 mm/sec;1 cm/mV

Question: Your Query

Fig. 8. (a): ECG input, DL model selection, and output (refer Figure 2 for architectural diagram). (b): XAI heat-maps for ECG channels (refer Figure 2 for architectural diagram). (c): DL model details and accuracy plots. (d): Chatbot's response to user prompt. (refer Figure 3 for architectural diagram).

offers a holistic solution for improving patient care in the realm of cardiac health. We are preparing for a much larger training of CureNet on large medical datasets, and anticipate that our approach will find applications in other domains where the interpretability of AI models is paramount. The code and data used in this study are available on GitHub at the following link: [17].

Acknowledgment

We express our gratitude to Amrita University for providing computational resources and other resources required for this project. We are also deeply indebted to the constant support given by our Chancellor, Sri. Mata Amritanandamayi Devi, to bring together a highly interdisciplinary team of researchers and physicians to help build societal useful solutions.

References

[1] Jia, Wenxiao, Xiao Xu, Xian Xu, Yuyao Sun and Xiaoshuang Liu. "Arrhythmia Detection and Classification of 12-lead ECGs Using a Deep Neural Network." 2020 Computing in Cardiology Conference (CinC) (2020): n. pag.

[2] Nils Strodthoff, Claas Strodthoff, "Detecting and interpreting myocardial infarction using fully convolutional neural networks", 2019.

[3] S. Yang, H. Xiang, Q. Kong and C. Wang, "Multi-label Classification of Electrocardiogram With Modified Residual Networks," 2020 Computing in Cardiology, Rimini, Italy, 2020, pp. 1-4, doi: 10.22489/CinC.2020.007.

[4] M. Żyliński and G. Cybulski, "Selected Features for Classification of 12-lead ECGs," 2020 Computing in Cardiology, Rimini, Italy, 2020, pp. 1-4, doi: 10.22489/CinC.2020.061.

[5] M. A. Reyna et al., "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," 2020 Computing in Cardiology, Rimini, Italy, 2020, pp. 1-4, doi: 10.22489/CinC.2020.236.

[6] Shrikumar, Avanti, Peyton Greenside and Anshul Kundaje. "Learning Important Features Through Propagating Activation Differences." International Conference on Machine Learning (2017).

[7] Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W (2015) On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLoS ONE 10(7): e0130140. https://doi.org/10.1371/journal.pone.0130140August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[8] Csurka, Gabriella, Christopher R. Dance, Lixin Fan, Jutta Katharina Willamowski and Cédric Bray. "Visual categorization with bags of keypoints." European Conference on Computer Vision (2002).

[9] van de Sande KE, Gevers T, Snoek CG. Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell. 2010 Sep;32(9):1582-96. doi: 10.1109/TPAMI.2009.154. PMID: 20634554.

[10] Bach, Sebastian, Alexander Binder, Grégoire Montavon, Klaus-Robert Müller and Wojciech Samek. "Analyzing Classifiers: Fisher Vectors and Deep Neural Networks." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 2912-2920.

[11] Binder, Alexander, Grégoire Montavon, Sebastian Lapuschkin, Klaus-Robert Müller and Wojciech Samek. "Layer-Wise Relevance Propagation for Neural Networks with Local Renormalization Layers." ArXiv abs/1604.00825 (2016): n. pag.

[12] Ksenia Shkileva, Nikolai Zolotykh, Explainable Artificial Intelligence Techniques in Medical Signal Processing, Procedia Computer Science, Volume 212, 2022, Pages 474-484, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.11.031.

[13] Liu, Feifei Liu, Chengyu Zhao, Lina Zhang, Xiangyu Wu, Xiaoling Xu, Xiaoyan Liu, Yulin Ma, Caiyun Wei, Shoushui He, Zhiqiang Li, Jianqing Ng, Eddie. (2018). An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. Journal of Medical Imaging and Health Informatics. 8. 1368-1373. 10.1166/jmihi.2018.2442.

[14] Bousseljot R, Kreiseler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1 (1995) S 317

[15] Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T. (2020), PTB-XL: A Large Publicly Available ECG Dataset. Scientific Data. https://doi.org/10.1038/s41597-020-0495-6

[16] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P.C., Mark, R., Mietus, J.E., Moody, G.B., Peng, C.K. and Stanley, H.E., 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[17] https://github.com/TejoVK/ECG_XAi