MEETI: A Multimodal ECG Dataset from MIMIC-IV-ECG with Signals, Images, Features and Interpretations

Deyun Zhang[1,#], Xiang Lan[2,#], Shijia Geng[1], Qinghao Zhao[3], Sumei Fan[4], Mengling Feng[2,*], and Shenda Hong[5,6,*]

[1]HeartVoice Medical Technology, Hefei, China
[2]Saw Swee Hock School of Public Health and Institute of Data Science, National University of Singapore, Singapore
[3]Department of Cardiology, Peking University People's Hospital, Beijing, China
[4]College of Integrative Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei, China
[5]National Institute of Health Data Science, Peking University, Beijing, China
[6]Institute for Artificial Intelligence, Peking University, Beijing, China.
[#]These authors contributed equally
[*]Correspondence: hongshenda@pku.edu.cn, ephfm@nus.edu.sg

Abstract

Electrocardiogram (ECG) plays a foundational role in modern cardiovascular care, enabling non-invasive diagnosis of arrhythmias, myocardial ischemia, and conduction disorders. While machine learning has achieved expert-level performance in ECG interpretation, the development of clinically deployable multimodal AI systems remains constrained, primarily due to the lack of publicly available datasets that simultaneously incorporate raw signals, diagnostic images, and interpretation text. Most existing ECG datasets provide only single-modality data (e.g., raw waveform signals) or, at most, dual modalities (e.g., image-text pairs), making it difficult to build models that can understand and integrate diverse ECG information in real-world settings. To address this gap, we introduce MEETI (MIMIC-IV-Ext ECG-Text-Image), the first large-scale ECG dataset that synchronizes raw waveform data, high-resolution plotted images, and detailed textual interpretations generated by large language models (LLMs). In addition, MEETI includes beat-level quantitative ECG parameters extracted from each lead, offering structured parameters that support fine-grained analysis and model interpretability. MEETI is built upon MIMIC-IV-ECG, one of the largest open-access collections of 12-lead clinical ECG recordings, comprising over 800,000 ten-second recordings paired with expert reports. MEETI expands this foundation with three key additions: high-resolution ECG images, beat-level quantitative ECG parameters, and detailed textual interpretations. Each MEETI record is aligned across four components: (1) the raw ECG waveform, (2) the corresponding plotted image, (3) extracted feature parameters, and (4) detailed interpretation text. This alignment is achieved using consistent, unique identifiers. This unified structure supports transformer-based multimodal learning and supports fine-grained, interpretable reasoning about cardiac health. By bridging the gap between traditional signal analysis, image-based interpretation, and language-driven understanding, MEETI established a robust foundation for the next generation of explainable, multimodal cardiovascular AI. It offers the research community a comprehensive benchmark for developing and evaluating ECG-based AI systems.

Background & Summary

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide[1], contributing to over 17 million deaths each year. The electrocardiogram (ECG) remains the primary non-invasive

modality to assess cardiac electrophysiology[2]. It provides temporal waveform information that is critical for diagnosing arrhythmias, myocardial ischemia, and conduction abnormalities[3,4]. However, manually interpreting large volumes of ECG recordings is time-consuming and prone to inter-observer variability. Recently, machine learning and artificial intelligence (AI) approaches have emerged as powerful tools for ECG analysis[5–8]. They can achieve expert-level performance in waveform analysis, arrhythmia detection, and risk stratification. These advances underscore the urgent need for comprehensive, multimodal ECG datasets. Such datasets are essential to ensure reproducible research and to accelerate clinical translation. Over the past few years, multimodal large language models (LLMs) have demonstrated remarkable performance in interpreting complex medical imaging data[9]. They can seamlessly integrate radiographic, histopathological, and clinical text inputs to support diagnostic and prognostic workflows[10,11]. These advanced architectures use transformer-based encoders to fuse pixel-level features with semantic context. As a result, tasks such as lesion detection, disease classification, and report generation can be performed with near-human precision[12]. Despite these successes, current LLMs remain incapable of directly processing and understanding electrocardiographic waveforms. Cardiac signals have unique temporal dynamics and amplitude variations that differ fundamentally from spatial image patterns. Bridging this gap requires specialized datasets and model designs that jointly represent ECG signals alongside complementary textual and visual information.

Most existing ECG datasets provide only single-modality data (e.g., raw waveform signals) or, at most, dual modalities such as image–text pairs, as commonly seen in clinical reports. However, contemporary AI research increasingly requires multimodal representations to model the complex nature of ECG data, which inherently combine temporal waveforms, visual patterns, and diagnostic text[13–15]. This mismatch between available data formats and modeling needs has hindered progress in developing explainable and generalizable ECG-based AI systems. Models trained on digitized time series cannot easily leverage the many image-based clinical interpretations available online[16,17]. Traditional analysis pipelines rely on handcrafted parameter extraction (e.g., QRS complex duration, ST segment deviations) to support diagnostic decisions[18]. In contrast, deep learning models act as opaque black boxes without explicit feature semantics[19]. These two limitations highlight that ECG data naturally combine visual, temporal, and textual information. As a result, we need a unified multimodal approach. Training multimodal language models to understand ECG data requires genuine ECG multimodal examples—paired waveform images, raw signals, and textual annotations[20,21]. However, existing public ECG datasets only include signal–text pairs and leave out image data[22,23]. This gap prevents a systematic study of how to combine visual and temporal cardiac representations in transformer-based models. The lack of integrated datasets is therefore a bottleneck for ECG multimodal research. It motivates the creation of resources that encode images, signals, and semantics together for end-to-end model training.

To address these limitations, we introduce MEETI (MIMIC-IV-Ext ECG-Text-Image, Figure 1), a multimodal ECG dataset that enables LLMs to process and interpret ECG data more effectively. MEETI is built on the MIMIC-IV-ECG dataset[14,24], one of the largest and most representative public 12-lead clinical ECG collections available. It covers a wide range of cardiac conditions and signal qualities that reflect real-world clinical settings. MEETI is the first multimodal ECG dataset to provide synchronized access to four key components: raw ECG signals, plotted waveform images, extracted quantitative parameters, and detailed interpretation texts. By aligning these modalities, MEETI establishes a unified framework that integrates signal, visual, and language representations. This framework standardizes input formats for training LLMs and promotes explainable AI by enabling clinically interpretable reasoning, rather than replying on traditional 'black-box' classification approaches. The integration of visual, temporal, and semantic modalities facilitates cross-modal alignment and supports the development of models capable of performing ECG analysis from multiple perspectives. Moreover, MEETI encourages a paradigm shift from
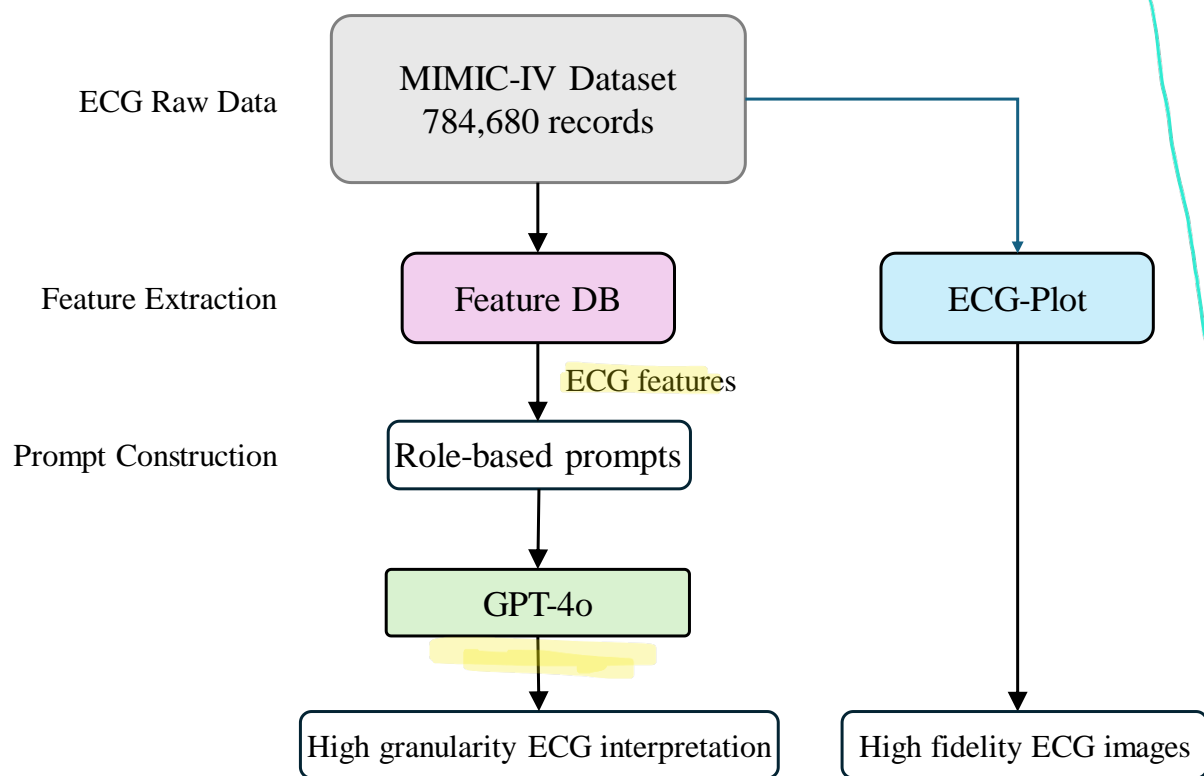
2

Figure 1: Schematic overview of the components that constitute the MEETI dataset.

conventional end-to-end diagnostic pipelines toward inference-based reasoning. Ultimately, this resource enhances the ability of multimodal LLMs to generate meaningful, context-aware interpretations of cardiac health.

Methods

ECG Signals

As part of routine clinical care at the Beth Israel Deaconess Medical Center (BIDMC), diagnostic 12-lead ECGs were collected using standard ECG machines and subsequently archived in the MIMIC-IV-ECG[14]. This dataset comprises approximately 800,000 ten-second, 12-lead ECG recordings sampled at 500 Hz, matched to nearly 160,000 unique patients included in the broader MIMIC-IV Clinical Database. In addition to the raw waveform data, machine-generated measurements and cardiologist-interpreted reports are also available for a substantial proportion of ECGs. All records have been de-identified in accordance with the HIPAA Safe Harbor requirements to ensure patient privacy.

ECG Parameter Extraction

To extract a comprehensive set of quantitative descriptors from the MIMIC-IV-ECG waveforms, we integrate the open source FeatureDB toolkit (https://github.com/PKUDigitalHealth/FeatureDB) into our pre-processing pipeline[18,25]. FeatureDB employs a multi-stage workflow beginning with

adaptive peak detection to localize the P wave, QRS complex, and T wave across each lead. Following initial peak localization, discrete wavelet transforms are applied at multiple scales to refine the onset and offset boundaries of these waveform components, thereby ensuring precise measurement of key temporal intervals (PR interval, QRS duration, QT interval, and QTc). We applied this pipeline to all de-identified 12 lead ECG records after verifying signal integrity (i.e., complete lead set, consistent sampling rate, and absence of major artifacts). Extracted parameters were stored in a structured relational database, with each record linked to its raw WFDB waveform files and corresponding expert reports via unique study identifiers. This systematic application of FeatureDB not only facilitates scalable, reproducible extraction of clinically meaningful ECG metrics but also lays the groundwork for downstream multimodal integration with image and text modalities, ultimately supporting the development of interpretable AI models in cardiac electrophysiology.

ECG Plot

To generate high-fidelity ECG images, we rendered each preprocessed 12 lead waveform into a standardized graphical format using the open-source ecg plot Python library (`https://github.com/PKUDigitalHealth/ecg_plot`)[26]. For each record, we configured the plotting parameters to emulate clinical paper recordings: a paper speed of 25 mm/s and an amplitude scaling of 10 mm/mV. Under these settings, each large grid block corresponds to 0.5 mV in voltage and 0.2 s in time, thereby ensuring precise spatial encoding of waveform morphology. Leads were arranged in the conventional 3×4, 6×2, or 12×1 (I, II, III; aVR, aVL, aVF; V1-V6) with appropriate inter-panel spacing and grid visibility. Images were exported at 300 dpi resolution to facilitate downstream vision-based model training and to maintain pixel-level detail of low-amplitude features. Each image file was named using the study identifier, then stored alongside the corresponding waveform and text report for seamless multimodal alignment. This procedural standardization yields a uniform image corpus suitable for transformer-based and convolutional model development in ECG interpretation.

ECG Interpretation Generation

Human-written ECG reports are typically brief and often lack the detail needed for comprehensive interpretation. To produce high-granularity ECG interpretations that explicitly link diagnoses to measurable parameters such as QRS durations or PR intervals, we develop role-based prompts that, combined with extracted ECG parameters, allow GPT-4o[27] to effectively utilize its latent medical knowledge for generating clinically detailed and parameter-grounded interpretations. For each ECG record, the extracted parameters are structured into sequential data. For example, an ECG signal containing ten visible heartbeats allows us to construct a QRS duration sequence as $[QRS_1, QRS_2, ..., QRS_{10}]$, where each element precisely represents the QRS duration for a individual heartbeat. These sequences offer fine-grained descriptions of ECG physiology, enabling assessment of cardiac function at the individual heartbeat level. The role-based prompts then guide GPT-4o[27] with cardiology-specific instructions to analyze various aspects of the ECG parameters, conditioned on the ground-truth report. For example, the prompt may instruct GPT-4o[27] to examine P wave amplitude and duration in Lead II for signs of atrial enlargement. These instructions also reflect the logical structure of clinical diagnostic workflows. Through this approach, we are able to generate high-granularity ECG interpretations at scale without relying on manual annotations.

4

Data Records

Data Released as Part of This Dataset

MEETI can be accessed via the following link: `https://github.com/PKUDigitalHealth/MIMIC-IV-ECG-Ex` MEETI consists of four complementary components: raw ECG signals and report, plotted ECG images, per-beat ECG parameters, and LLM-based textual interpretations. Each component is linked by a unique record identifier so that users can cross-reference signals, images, parameters, and interpretations. Our goal is to lower the barrier to entry for multimodal ECG research. By making all four data types available in one cohesive resource, we hope to accelerate developments in ECG-based diagnostics, prognostics, and educational tools. Researchers can seamlessly integrate waveform analysis, image-based deep learning, feature engineering, and natural language processing within a single dataset. MEETI thus represents a comprehensive, publicly accessible resource for advancing AI in cardiology.

- **Raw ECG data**: The raw ECG signals and accompanying text reports are included directly from MIMIC-IV-ECG. These original signals cover a wide range of patient demographics, clinical conditions, and recording settings. Researchers can access the same high-resolution waveform data that clinicians used at Beth Israel Deaconess Medical Center. By depositing MEETI on PhysioNet, we ensure long-term accessibility and enable others to reproduce our results without additional data collection efforts.

- **Plotted ECG image**: MEETI provides a set of plotted ECG images. Each image arranges twelve leads in a 12×1 grid, replicating the standard clinical layout. We generated these plots using the matplotlib library. The plotting scripts apply consistent scaling and labeling conventions to preserve the clinical context. Clinicians and machine learning practitioners alike can use these images for visual analysis, teaching, or training convolutional neural networks. Since plotted images are derived directly from the same high-fidelity signals, they faithfully represent the original ECG morphology.

- **ECG parameter**: MEETI also includes quantitative ECG parameters for every beat in each lead. These parameters were automatically extracted using the FeatureDB toolkit. For each lead, FeatureDB identifies fiducial points (P-wave, QRS complex, T-wave) and computes intervals (PR, QRS duration, QT/QTc) as well as amplitudes (R- and S-wave heights in leads V1 and V5, for example). By capturing beat-by-beat variability, these detailed metrics support both statistical analyses and deep learning approaches. Researchers interested in heart rate variability, conduction delays, or repolarization abnormalities will find this dataset especially useful. Providing per-beat parameters from all twelve leads means that MEETI can facilitate studies of subtle waveform changes that might be missed when relying on summary statistics alone.

- **Interpretation of the LLM**: MEETI offers natural language interpretations generated by an LLM. We used GPT-4o[27] to produce these interpretations based on the combined input of the original textual reports and the extracted ECG parameters. For each recording, the LLM output highlights significant findings, such as rhythm abnormalities or evidence of ischemia. By comparing LLM-generated text with clinician-authored reports, users can evaluate the model's diagnostic suggestions. These interpretations serve as a benchmark for future work on AI-driven ECG reporting, and they demonstrate how multimodal inputs (e.g., text plus structured parameters) can inform more accurate, context-aware explanations.

Each subject's data is organized into a dedicated folder named by their unique subject ID, which corresponds directly to the IDs used in the MIMIC-IV-ECG database. Within each folder, plotted

ECG images are saved as individual PNG files, with filenames that reflect lead arrangement and recording timestamps. Beat-by-beat parameters—such as PR interval, QRS duration, QT/QTc values, and fiducial point amplitudes—are stored in the MAT file. This file includes detailed metrics for every beat across all twelve leads, making it easy to perform quantitative analyses without further preprocessing. The LLM-generated textual interpretations from GPT-4o[27] are also provided in the MAT file. Those interpretations link directly to the corresponding ECG parameter sets, allowing users to compare AI-driven commentary with numerical results. By grouping all PNG and MAT files under a single subject-specific directory (Table 1), MEETI ensures that researchers can quickly locate ECG images, per-beat ECG parameters, and LLM-based textual interpretations for each individual (Table 2). This streamlined structure supports seamless multimodal analysis.

Descriptive Statistics

Except for a small number of samples that could not be processed by the algorithm, MEETI cover the full MIMIC-IV-ECG dataset, up to 784,680 records from 160,597 patients.

We used several tables and figures to present the ECG parameters included in MEETI. Table 3 shows the median values of all ECG parameters for each complete ECG recording. These statistics provide a clear overview of the central tendencies across subjects and recordings. Researchers interested in overall waveform characteristics can refer to Table 3, while those focused on lead-specific details will find Table 3 especially useful. All of these parameters are stored in the .mat files for easy access. The clear linkage between the tables, figures, and .mat files ensures that datasets are both transparent and reproducible (Table 3, Figure 2).

The ECGs are grouped into subdirectories based on subject_id. Each record path follows the pattern: files/pNNNN/pXXXXXXXX/sZZZZZZZZ/ZZZZZZZZ, where:

NNNN is the first four characters of the subject_id, XXXXXXXX is the subject_id, ZZZZZZZZ is the file_name

An example of the file structure is as follows:

```
MEETI
|---p1000
    |---p10000032
        |---s40689238
            |---40689238.mat
 |......
 |---p1571
    |---p15718729
        |---s40000369
            |---40000369.mat
            |---40000369.png
        |---s40461152
            |---40461152.mat
            |---s44659160
        |---44659160.mat
            |---s48122836
            |---48122836.mat
```

Above we find two subjects p10000032 (under the p1000 group level directory) and p15718729 (under the p1571 group level directory). For subject p10000032 we find one study: s40689238. For p15718729 we find four studies: s40000369, s40461152, s44659160, s48122836. The study identifiers are completely random, and their order has no implications for the chronological order of the actual studies. Each study has a like-named .png and .mat file (Table 1, Table 2).

Table 1: Description of key values of the provided datasets.

	key value	Description
.png file	—	ECG image generated from ECG signal data
.mat file	id	MIMIC-IV-ECG identifier
	LLM_Interpretation	ECG interpretation generated by LLM based on input prompts
	report	ECG report from MIMIC-IV-ECG
	featuredb_lead_X	Beat parameters were extracted from each lead using FeatureDB. 'X' denotes the lead name, which can be one of the following: I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, or V6

The record_list.csv file contains the file name and path for each record. It also provides the corresponding subject_id and file_name. The subject_id can be used to link a subject from MEETI to the other modules in the MIMIC-IV and MIMIC-IV-ECG Database.

Below is a basic script for reading the downloaded records from this project:

```python
from PIL import Image
from scipy.io import loadmat

loaddata = loadmat('./p1571/p15718729/s40000369/40000369.mat')
print(loaddata.keys())
print(loaddata['id'])
print(loaddata['LLM_Interpretation'])
print(loaddata['report'])

img = Image.open('./p1571/p15718729/s40000369/40000369.png')
print(img.size)
```

Technical Validation

The technical validation of MEETI includes two main components. First, we present the distribution and descriptive statistics of the extracted ECG parameters. For each ECG parameter, such as PR interval, QRS duration, QT interval, and T-wave axis, we calculate the mean and standard deviation. Distributions are visualized across all twelve leads to show consistency and variability within the dataset. This step helps ensure that the extracted parameters fall within expected clinical ranges and reflect real-world physiological diversity. Second, we demonstrate the richness of MEETI by showcasing examples of the dataset's four core components: ECG image, ECG report, beat-level parameters, and LLM interpretations. We highlight how these elements are linked through shared identifiers and are consistently formatted for ease of use. Together, they provide a comprehensive view of each cardiac event, from waveform morphology to diagnostic explanation. This combination supports multiple downstream tasks, such as signal analysis,

7

Table 2: Descriptions of ECG parameters were extracted from each lead using FeatureDB.

ECG parameter name	Description
HR	The heart rate was calculated based on the average of RR1 and RR2, along with the sampling rate
RR1	The RR interval between the current heartbeat and the previous one
RR2	The RR interval between the current heartbeat and the next one
P_amplitude	The amplitude of the P wave, measured from the isoelectric baseline to its peak
P_duration	The duration of the P wave, from onset to its end
PR_interval	The interval from the start of the P wave to the start of the QRS complex
QRS_amplitude	The peak amplitude of the QRS complex
QRS_duration	The duration of the QRS complex, from its onset to its end
T_amplitude	The amplitude of the T wave was measured from baseline to peak
T_duration	The duration of the T wave, from onset to its end
ST_duration	The length of the ST segment, from the end of the QRS complex to the beginning of the T wave
ST_form	The morphological pattern of the ST segment (e.g., horizontal, upsloping, downsloping)
QT_interval	The interval from the start of the QRS complex to the end of the T wave
QTc	The QT interval was corrected for heart rate

image classification, and natural language modeling. By validating both the statistical integrity and multimodal structure of the data, we aim to show that MEETI is a reliable and versatile resource for training and evaluating ECG-based machine learning models.

ECG parameters: distribution of extracted parameters. To validate the quality of the extracted ECG parameters, we analyzed their statistical distributions across the entire MEETI dataset. For each parameter—such as PR interval, QRS duration, QT interval, QTc, heart rate, etc.—we calculated the mean, standard deviation, median, interquartile range, etc. These statistics were computed at both the recording level. We also report the missing data ratio, which was effectively zero across all records, indicating high data completeness. Most parameters showed distributions consistent with known physiological ranges. For example, the median PR interval was approximately 140 ms, and the QRS duration clustered around 90 ms, as expected in standard adult ECGs.

These results are presented in Table 3, providing an overview of the central tendency and variability across the entire dataset. The clear, clinically plausible patterns observed across

metrics confirm that the extracted parameters are reliable and suitable for downstream modeling tasks such as classification, clustering, or trend analysis in machine learning applications.

Table 3: Statistical analysis of electrocardiogram parameters was conducted by extracting data from MIMIC-IV-ECG using FeatureDB.
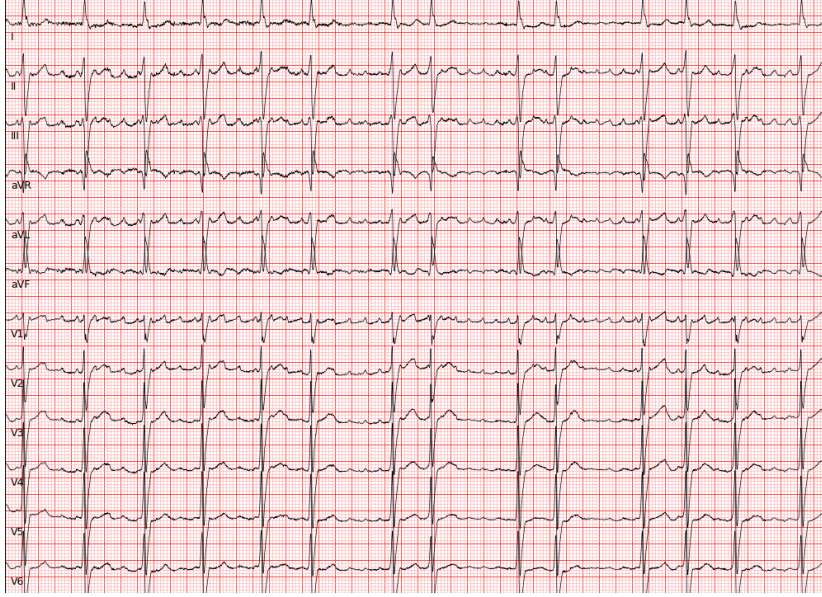
	Mean	Std	95%CI	Median	IQR	Skewness	Kurtosis	Missing ratio
HR	78.19	19.77	78.17-78.20	74.50	24.00	1.13	1.78	0.00
RR1	403.44	94.38	403.38-403.50	400.00	128.50	0.38	1.17	0.00
RR2	403.70	94.41	403.64-403.76	400.5	128.00	0.41	1.72	0.00
P_amplitude	0.30	0.43	0.30-0.30	0.10	0.31	2.95	12.33	0.00
P_duration	94.01	9.42	94.01-94.02	94.00	12.00	-0.18	-0.16	0.00
PR_interval	135.07	60.76	135.03-135.11	140.00	61.00	-0.58	0.83	0.00
QRS_amplitude	0.59	0.62	0.59-0.59	0.49	0.96	1.21	2.32	0.00
QRS_duration	84.65	36.48	84.63-84.68	91.00	70.00	0.06	-1.20	0.00
T_amplitude	0.34	0.42	0.34-0.34	0.20	0.32	3.54	20.94	0.00
T_duration	145.15	18.45	145.14-145.16	144.00	26.00	-0.03	0.74	0.00
ST_duration	133.10	71.08	133.06-133.15	127.00	75.00	0.98	4.33	0.00
ST_form	0.00	1.00	0.00-0.00	0.00	1 .00	-0.58	-1.00	0.00
QT_interval	363.98	85.51	363.92-364.03	371.00	94 .00	-0.09	2.04	0.00
QTc	409.30	91.74	409.24-409.35	418.00	94.5 0	-0.27	0.76	0.00

Multimodal components: Synchronized ECG signals, waveform images, beat-level quantitative parameters, and interpretation texts generated by LLMs. We confirmed that each MEETI record reliably links four data modalities around a single visual workspace. In Figure 2, the twelve-lead ECG image appears on the left. To its right, we display the corresponding per-beat parameters and the clinician's original text report side by side. This arrangement allows users to see numerical metrics—such as PR interval, QRS duration, and wave amplitudes—aligned directly with expert observations. Beneath the entire panel, we place the GPT-4o generated interpretation. This parameter-based summary references specific parameter values and mirrors the clinical report's structure, for example: "The rhythm suggests atrial fibrillation, characterized by the irregular RR intervals observed throughout the ECG. This arrhythmia is further supported by the absence of distinct P waves, vals observed throughout the ECG" or "Nonspecific ST-T wave changes are present, particularly noticeable in the lateral leads (V5-V6), which display a subtle upslope in the ST segment, although without a clear pattern of ischemia or infarction."

Usage Notes

This module gives MEETI users an additional, potentially valuable source of information for research involving the ECG LLM. The dataset does have some limitations. Because of image size constraints, it includes about 10,000 ECG images. Researchers who need more ECG image data can use MIMIC-IV-ECG as a foundation to generate additional samples from the GitHub repository (`https://github.com/PKUDigitalHealth/ecg_plot`). Our recent work GEM[28] demonstrates how to leverage such multimodal ECG data to train MLLMs for ECG understanding.

ECG

ECG Parameters

Name	Parameter	Unit
HR	82	bpm
RR1	351	——
RR2	351	——
P_amplitude	0.06	mV
P_duration	94	ms
PR_interval	102	ms
QRS_amplitude	2.26	mV
QRS_duration	122	ms
T_amplitude	0.13	mV
T_duration	166	ms
ST_duration	128	ms
ST_form	Upsloping	——
QT_interval	404	ms
QTc	477	ms

ECG Report

Atrial fibrillation, Marked left axis deviation, Intraventricular conduction defect, Lateral ST-T changes are nonspecific, Repolarization changes may be partly due to rhythm, Abnormal ECG.

LLM Interpretation

In analyzing the provided ECG, several significant findings merit attention. Initially, the rhythm suggests atrial fibrillation, characterized by the irregular RR interrit attention. Initially, the rhythm suggests atrial fibrillation, characterized by the irregular RR intervals observed throughout the ECG. This arrhythmia is further supported by the absence of distinct P waves,vals observed throughout the ECG. This arrhythmia is further supported by the absence of distinct P waves, especially evident in Leads II and III. Furthermore, there is clear evidence of a marked left axis deviation, indicated by the negative QRS complexes in Lead III, which typically suggests a leftward shift in the heart's electrical axis. The QRS duration is generally prolonged in Leads I and V1-V6, suggesting an intraventricular conduction defect. In Lead I, the QRS duration consistently exceeds 120 ms, which aligns with the criteria for conduction abnormalities. Additionally, nonspecific ST-T wave changes are present, particularly noticeable in the lateral leads (V5-V6), which display a subtle upslope in the ST segment, although without a clear pattern of ischemia or infarction. The QTc intervals are prolonged in several leads, such as Lead V4, indicating potential repolarization abnormalities that may be partly related to the irregular rhythm of atrial fibrillation. Synthesizing these findings, the primary clinical diagnoses are atrial fibrillation, a marked left axis deviation, and an intraventricular conduction defect, with lateral ST-T changes that are nonspecific in nature. These diagnoses are substantiated by the irregular rhythm, atypical QRS complex morphology, and ST-T changes observed in the leads as explained. These findings together result in an abnormal ECG indicative of these specific cardiac conditions.

Figure 2: Multimodal ECG data.

RESOURCE AVAILABILITY

The code of our method and evaluations is publicly available at `https://github.com/PKUDigitalHealth/ecg_plot` and `https://github.com/PKUDigitalHealth/FeatureDB`. All the datasets used in our work are publicly available: MIMIC-IV-ECG (`https://physionet.org/content/mimic-iv-ecg/1.0/`) and MEETI (`https://github.com/PKUDigitalHealth/MIMIC-IV-ECG-Ext-Text-Image`).

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Creation and maintenance of the original database: D.Z., X.L., M.F. and S.H.; Conception of the release process: D.Z., X.L. and S.H.; Data harmonization: D.Z., X.L., S.G. and Q.Z.; Providing conversion routines: D.Z.; Manuscript preparation: D.Z., X.L., S.G., S.F. and S.H.; Supervision of the project: M.F. and S.H.; Critical comments and revision of manuscript: all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

1. Martin, S.S., Aday, A.W., Allen, N.B., Almarzooq, Z.I., Anderson, C.A., Arora, P., Avery, C.L., Baker-Smith, C.M., Bansal, N., Beaton, A.Z. et al. (2025). 2025 heart disease and stroke statistics: a report of us and global data from the american heart association. Circulation *151*, e41–e660.

2. Tan, J., Fan, H., Luo, J., Zhou, Y., Wang, N., Wang, X., Liu, G., Liu, C., and Wang, Z. (2025). A pediatric ecg database with disease diagnosis covering 11643 children. Scientific Data *12*, 867.

3. Yagi, R., Mori, Y., Goto, S., Iwami, T., and Inoue, K. (2024). Routine electrocardiogram screening and cardiovascular disease events in adults. JAMA Internal Medicine *184*, 1035–1044.

4. Zheng, J., Fu, G., Anderson, K., Chu, H., and Rakovski, C. (2020). A 12-lead ecg database to identify origins of idiopathic ventricular arrhythmia containing 334 patients. Scientific data *7*, 98.

5. Hong, S., Zhou, Y., Shang, J., Xiao, C., and Sun, J. (2020). Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Computers in biology and medicine *122*, 103801.

6. Lan, X., Ng, D., Hong, S., and Feng, M. (2022). Intra-inter subject self-supervised learning for multivariate cardiac signals. In Proceedings of the AAAI Conference on Artificial Intelligence vol. 36. pp. 4532–4540.

7. Lan, X., Yan, H., Hong, S., and Feng, M. (2024). Towards enhancing time series contrastive learning: A dynamic bad pair mining approach. In The Twelfth International Conference on Learning Representations. URL: `https://openreview.net/forum?id=K2c04ulKXn`.

8. Li, J., Aguirre, A.D., Junior, V.M., Jin, J., Liu, C., Zhong, L., Sun, C., Clifford, G., Brandon Westover, M., and Hong, S. (2025). An electrocardiogram foundation model built on over 10 million recordings. NEJM AI *2*, AIoa2401033.

9. Liu, F., Zhu, T., Wu, X., Yang, B., You, C., Wang, C., Lu, L., Liu, Z., Zheng, Y., Sun, X. et al. (2023). A medical multimodal large language model for future pandemics. NPJ Digital Medicine *6*, 226.

10. Sandmann, S., Hegselmann, S., Fujarski, M., Bickmann, L., Wild, B., Eils, R., and Varghese, J. (2025). Benchmark evaluation of deepseek large language models in clinical decision-making. Nature Medicine pp. 1–1.

11. Zhou, S., Xu, Z., Zhang, M., Xu, C., Guo, Y., Zhan, Z., Fang, Y., Ding, S., Wang, J., Xu, K. et al. (2025). Large language models for disease diagnosis: A scoping review. npj Artificial Intelligence *1*, 9.

12. Hong, Q., Liu, S., Wu, L., Lu, Q., Yang, P., Chen, D., Rao, G., Liu, X., Ye, H., Zhuang, P. et al. (2025). Evaluating the performance of large language & visual-language models in cervical cytology screening. NPJ Precision Oncology *9*, 153.

13. Wagner, P., Strodthoff, N., Bousseljot, R.D., Kreiseler, D., Lunze, F.I., Samek, W., and Schaeffter, T. (2020). Ptb-xl, a large publicly available electrocardiography dataset. Scientific data *7*, 1–15.

14. Gow, B., Pollard, T., Nathanson, L.A., Johnson, A., Moody, B., Fernandes, C., Greenbaum, N., Waks, J.W., Eslami, P., Carbonati, T. et al. (2023). Mimic-iv-ecg: Diagnostic electrocardiogram matched subset. Type: dataset *6*, 13–14.

15. Kansal, A., Chen, E., Jin, B.T., Rajpurkar, P., and Kim, D.A. (2025). Mc-med, multimodal clinical monitoring in the emergency department. Scientific Data *12*, 1094.

16. Al-Zaiti, S.S., Martin-Gill, C., Zègre-Hemsey, J.K., Bouzid, Z., Faramand, Z., Alrawashdeh, M.O., Gregg, R.E., Helman, S., Riek, N.T., Kraevsky-Phillips, K. et al. (2023). Machine learning for ecg diagnosis and risk stratification of occlusion myocardial infarction. Nature Medicine *29*, 1804–1813.

17. Chen, J., Huang, S., Zhang, Y., Chang, Q., Zhang, Y., Li, D., Qiu, J., Hu, L., Peng, X., Du, Y. et al. (2024). Congenital heart disease detection by pediatric electrocardiogram based deep learning integrated with human concepts. Nature Communications *15*, 976.

18. Fan, S., Zhang, D., Wang, Y., Geng, S., Lu, K., Sang, M., Xu, W., Wang, H., Zhao, Q., Cheng, C. et al. (2025). Detecting long qt syndrome and first-degree atrioventricular block using single-lead ai-ecg: A multi-center real-world study. arXiv preprint arXiv:2502.17499.

19. ŞAHiN, E., Arslan, N.N., and Özdemir, D. (2025). Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning. Neural Computing and Applications *37*, 859–965.

20. Li, J., Liu, C., Cheng, S., Arcucci, R., and Hong, S. (2024). Frozen language model helps ecg zero-shot learning. In Medical Imaging with Deep Learning. PMLR pp. 402–415.

21. Lai, Y., Chen, J., Zhao, Q., Zhang, D., Wang, Y., Geng, S., Li, H., and Hong, S. (2025). Diffusets: 12-lead ecg generation conditioned on clinical text reports and patient-specific information. Patterns.

22. Tsutsui, K., Brimer, S.B., Ben-Moshe, N., Sellal, J.M., Oster, J., Mori, H., Ikeda, Y., Arai, T., Nakano, S., Kato, R. et al. (2025). Shdb-af: a japanese holter ecg database of atrial fibrillation. Scientific Data *12*, 454.

23. Zheng, J., Zhang, J., Danioko, S., Yao, H., Guo, H., and Rakovski, C. (2020). A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific data *7*, 48.

24. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L.A., and Mark, R. (2020). Mimic-iv. PhysioNet. Available online at: https://physionet. org/content/mimiciv/1.0/(accessed August 23, 2021) pp. 49–55.

25. Hong, S., Fu, Z., Zhou, R., Yu, J., Li, Y., Wang, K., and Cheng, G. (2020). Cardiolearn: a cloud deep learning service for cardiac disease detection from electrocardiogram. In Companion Proceedings of the Web Conference 2020. pp. 148–152.

26. Zhang, D., Geng, S., Zhou, Y., Xu, W., Wei, G., Wang, K., Yu, J., Zhu, Q., Li, Y., Zhao, Y. et al. (2023). Artificial intelligence system for detection and screening of cardiac abnormalities using electrocardiogram images. arXiv preprint arXiv:2302.10301.

27. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. et al. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

13

28. Lan, X., Wu, F., He, K., Zhao, Q., Hong, S., and Feng, M. (2025). Gem: Empowering mllm for grounded ecg understanding with time series and images. arXiv preprint arXiv:2503.06073.

407

408