



CardioAssist: A Review of Deep Learning, Machine Learning and Explainable AI Approaches for Integrated Cardiology Decision Support

Sayali Bhanage¹, Simrat Singh², Aanchal More³, Prof. Jaspreet Kaur⁴

¹Department of Computer Engineering Smt. Indira Gandhi College of Engineering Ghansoli, India 2023ce05f@sigce.edu.in

²Department of Computer Engineering Smt. Indira Gandhi College of Engineering Ghansoli, India 2023ce06f@sigce.edu.in

³Department of Computer Engineering Smt. Indira Gandhi College of Engineering Ghansoli, India 2023ce35f@sigce.edu.in

⁴Department of Computer Engineering Smt. Indira Gandhi College of Engineering Ghansoli, India

ABSTRACT—

Cardiovascular disease is the leading cause of death worldwide, and doctors who see a cardiac patient usually have to do three separate things: read an ECG, weigh a set of clinical risk factors, and check whether any prescribed drug could prolong the QT interval. Each task has its own body of machine-learning research, but very little of it asks how the three might sit in one tool that a clinician can actually use. This paper reviews the literature behind each task: deep learning for 12-lead ECG classification, explainability methods for ECG models, machine-learning risk estimation from clinical parameters, and QTc-based medication safety screening. We then describe how these pieces are planned to fit together in CardioAssist, a doctor-facing web platform that is still under development. The planned design pairs a ResNet50 ECG classifier with Grad-CAM heat maps, a Random Forest risk model built on the UCI Heart Disease data, and a QTc module that combines published clinical rules with a learned component. No performance results for the system are reported here. Instead we point out the gaps that motivate it: single-task tools, thin evidence on image-based ECG analysis, explanations that are rarely checked by clinicians, small tabular datasets that invite optimistic scores, and scarce external validation. We close with a staged evaluation plan.

Index Terms—ECG classification, ResNet50, Grad-CAM, explainable AI, Random Forest, cardiovascular risk, QTc, clinical decision support, literature review

1. INTRODUCTION

A. Motivation

Cardiovascular disease (CVD) causes more deaths than any other group of conditions. The Global Burden of Disease study put the toll at roughly 18.6 million deaths in 2019 [1]. India carries a large share of this burden, and studies of Indian populations suggest that CVD tends to appear earlier in life than in many Western cohorts [2]. Where specialist care is unevenly spread, a practical question follows: how can a busy physician, who may not be a cardiologist, get a dependable second look at a patient's heart?

In day-to-day practice that second look rarely comes from a single source. A typical cardiac work-up involves reading a 12-lead ECG, combining history, blood pressure, cholesterol and exercise findings into some sense of overall risk, and checking whether a prescribed drug might lengthen the QT interval and invite a dangerous arrhythmia. Each of these steps has drawn its own stream of machine-learning work. What has drawn far less attention is how they could live together in one tool that a doctor can open during a consultation.

Clinical decision support systems (CDSS) are the usual name for such tools. A well-known systematic review of controlled trials found that these systems were more likely to change practice when they worked automatically inside the clinician's workflow and offered concrete recommendations rather than bare assessments [3]. More recent overviews add that the benefits come with risks, among them alert fatigue and over-reliance on the machine [4]. Any system we propose has to be read against those lessons.

B. Project Context

This review grew out of our work on CardioAssist, a doctor-facing web platform planned around three tasks: classification of 12-lead ECG images, cardiovascular risk estimation from clinical parameters, and QTc-based medication safety screening. The platform is not finished, so this is not a results paper. What follows is a literature review that explains why we chose the techniques we did, where the existing evidence is thin, and what we intend to test once the prototype is stable. The guiding principle throughout is that the system supports a clinician's judgement and never replaces it.

C. Contributions

The novelty we claim is modest. It lies in the combination of tasks and in the attention given to validation, not in any single algorithm. Specifically, the paper offers:

1. A task-by-task review of ECG classification, explainability, tabular risk modelling and qtc screening, with comparison tables of representative work;
2. A five-dimension taxonomy that places each study by input, task, model family, explainability and level of integration (Table I);
3. A discussion of why resnet50 with Grad-CAM, a Random Forest, and a rule-plus-ML qtc module are reasonable choices for an early-stage system, together with their known weaknesses;
4. A list of research gaps, each tied to the part of cardioassist meant to address it (Table IX); and
5. A staged evaluation and validation plan (Sections VIII and XI).

D. Relation to Earlier Reviews

Several good reviews already cover parts of this ground. Hong *et al.* survey deep learning on ECG data [5], Siontis *et al.* discuss AI-enhanced electrocardiography from a clinical management viewpoint [6], Tjoa and Guan survey explainable AI in medicine [7], and Sutton *et al.* cover clinical decision support in general [4]. Each of these looks at one slice. This paper is narrower in depth but different in aim: it follows the three tasks an integrated cardiology tool would have to cover, and asks, for each, what the evidence supports and what a small team building such a tool should be careful about.

E. Organisation

Section II describes how the literature was collected, and Section III introduces the taxonomy. Sections IV to VII review ECG classification, explainability, risk assessment and QTc safety. Section VIII sets out the planned CardioAssist framework. Sections IX to XI cover gaps, challenges and future work, and Section XII concludes.

II. REVIEW METHODOLOGY

This is a narrative review, not a formal systematic one, and we want to be clear about that. We did not follow the PRISMA protocol or count screened records, and we do not claim exhaustive coverage. The aim was narrower: to gather the studies and resources that bear directly on the design choices of a small cardiology support system, and to read them with attention to what was actually validated.

We searched IEEE Xplore, PubMed, Google Scholar, arXiv and the ACM Digital Library with combinations of terms such as “12-lead ECG deep learning”, “ECG image classification”, “Grad-CAM ECG”, “heart disease prediction random forest”, “QTc prolongation drug risk” and “clinical decision support cardiology”. Deep-learning work from about 2015 onward was the main focus, but older papers were kept where they define a method, a dataset or a clinical threshold that newer work still relies on. We preferred peer-reviewed articles, work built on public datasets, and studies that described their evaluation clearly enough to judge. For the QTc material we leaned on guideline and consensus documents, because the safety rules come from cardiology practice rather than from machine learning. Fig. 1 summarises the process.

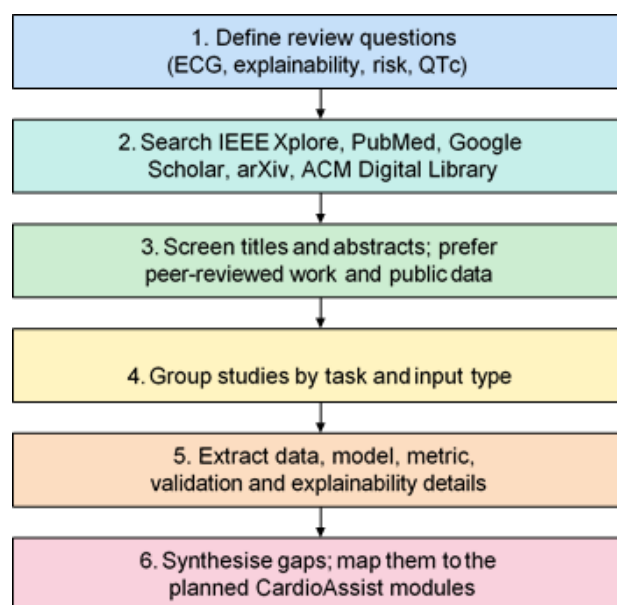


Fig. 1. Review process followed in this paper.

III. TAXONOMY OF CARDIAC AI SUPPORT SYSTEMS

To keep the comparison orderly we classify each study along five dimensions (Table I). They are not fully independent, but together they show where a given system sits and, more usefully, where few people seem to be working. Most published ECG papers occupy the corner marked ECG signal / classification / CNN / single task. The system we plan sits elsewhere: ECG image input, a post-hoc explanation, and an integrated platform that also handles risk and drug safety.

TABLE I

Taxonomy Used in This Review

Dimension	Categories
Input modality	ECG signal; ECG image or scan; tabular clinical data; medication and QT data
Task	Multi-class or multi-label classification; risk prediction; rule-based safety screening
Model family	Classical ML (SVM, Random Forest); CNN; recurrent or Transformer; hybrid rule + ML
Explainability	None; post-hoc saliency; model-agnostic (LIME, SHAP); intrinsic
Integration level	Single task; multi-task model; integrated platform

Fig. 2 shows, in broad strokes, how computer reading of the ECG has moved over the decades. The steps overlap in time and older methods are still used, but the direction is clear: less manual feature design, more reliance on data. That shift is the reason the dataset questions in Sections IV and IX matter as much as the model questions.

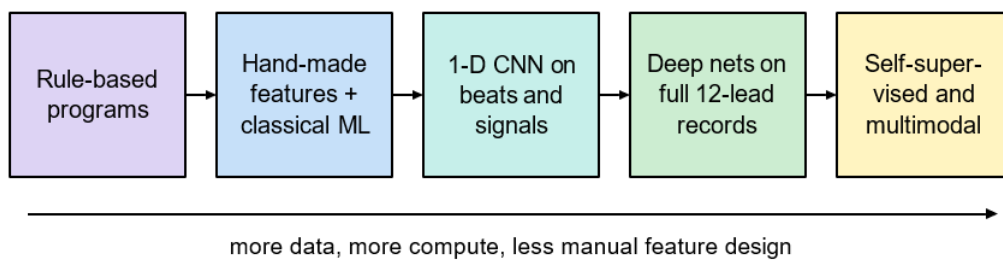


Fig. 2. Broad progression of computer-based ECG interpretation.

4. ECG-BASED CARDIAC CONDITION CLASSIFICATION

A. From Rules to Learned Features

Computer-assisted ECG reading began with rule-based programs built into recording devices. These pipelines typically filter the signal, locate the QRS complex with detectors such as the Pan–Tompkins algorithm [8], measure intervals and amplitudes, and then apply threshold rules or a conventional classifier such as an SVM. They behave reasonably on clean recordings but have a built-in weakness: every stage depends on the previous one, so a missed R-peak or a wandering baseline quietly spoils what follows. Hand-picked features also tend to encode only what the designer already knew to look for.

B. Beat-Level and Signal-Level CNNs

Convolutional networks removed much of the manual feature step. Kiranyaz *et al.* trained compact patient-specific 1-D CNNs for heartbeat classification on the MIT-BIH arrhythmia data [9], and Acharya *et al.* applied a deeper network to the same database [10]. Both showed that features learned from raw samples can compete with hand-built ones. The evaluation setting matters, though: when beats from the same patient land in both training and test sets, scores look better than they will in a clinic. We return to this patient-leakage problem in Section VI because it affects tabular data too.

C. Large-Scale 12-Lead and Single-Lead Studies

Three studies shaped how the field thinks about what is possible. Hannun *et al.* trained a 34-layer convolutional network on 91,232 single-lead ambulatory recordings from 53,549 patients and reported that it exceeded the average of a panel of cardiologists across twelve rhythm classes [11]; the group had described the approach earlier in a preprint [12]. Ribeiro *et al.* trained a residual network on more than two million 12-lead records from a Brazilian telehealth network and reported strong results for six abnormalities on a held-out set [13]. Attia *et al.* showed that a network could detect signs of atrial fibrillation from ECGs recorded during normal sinus rhythm, with an area under the curve of 0.87 [14]. These papers explain the current excitement around AI-enabled ECG analysis [6], [15]. They also

share a feature worth noticing: all of them work on digital signals and depend on data volumes that a student project cannot match. Reviews of the area point to the same recurring concerns: limited dataset size, uneven labelling and weak generalisation across sites [5].

D. PTB-XL and Public Benchmarks

For researchers without private hospital data, PTB-XL has become the reference resource. It holds 21,837 ten-second 12-lead records from 18,885 patients, annotated with standardised ECG statements that roll up into five superclasses: normal (NORM), myocardial infarction (MI), ST/T change (STTC), conduction disturbance (CD) and hypertrophy (HYP) [16]. The data are distributed through PhysioNet [17]. Strodthoff *et al.* used PTB-XL to benchmark a range of deep architectures, including ResNet-style and Inception-style 1-D CNNs and recurrent models, and found ResNet-type models among the strongest across tasks [18]. Later work looked at self-supervised pre-training to cope with limited labels [19]. Other large 12-lead collections, such as the 2020 PhysioNet/Computing in Cardiology Challenge data [20] and the Chapman–Shaoxing database [21], offer extra material for external testing. The five classes planned for CardioAssist are exactly the PTB-XL superclasses, which gives a clear label scheme and a route to larger-scale training later.

TABLE II

Representative Deep Learning Studies on ECG Data

Study	Input	Model	Data / target
Kiranyaz <i>et al.</i> [9]	Beat segments	Patient-specific 1-D CNN	MIT-BIH; heartbeat classes
Acharya <i>et al.</i> [10]	Beat segments	Deep 1-D CNN	MIT-BIH; heartbeat classes
Hannun <i>et al.</i> [11]	Single-lead ECG	34-layer CNN	91,232 records; 12 rhythm classes
Ribeiro <i>et al.</i> [13]	12-lead signal	Residual 1-D CNN	>2 million records; 6 abnormalities
Attia <i>et al.</i> [14]	12-lead signal	CNN	AF in sinus rhythm; AUC 0.87
Strodthoff <i>et al.</i> [18]	12-lead signal	ResNet, Inception, LSTM (benchmark)	PTB-XL; diagnostic, form, rhythm
Mehari and Strodthoff [19]	12-lead signal	Self-supervised pre-training	PTB-XL and others; few labels

Table III collects the datasets that appear in this review and says what role each plays in our plans.

TABLE III

Datasets Referred to in This Review

Dataset	Content	Size	Role here
PTB-XL [16]	12-lead ECG signals, 10 s	21,837 records; 18,885 patients	Label scheme; larger-scale training later
Chapman–Shaoxing [21]	12-lead ECG, arrhythmia	>10,000 patients	Possible external test set
PhysioNet/CinC 2020 [20]	12-lead ECG, several sources	Pooled challenge data	Possible external test set
MIT-BIH [22]	Two-channel ECG, annotated beats	48 half-hour excerpts; 47 subjects	Background (beat-level work)
Khan <i>et al.</i> [23]	ECG images	Image collection	Image-based reference
UCI Heart Disease [24]	Tabular clinical attributes	303 records (Cleveland)	Risk module

E. Image-Based ECG Classification

Signal-based models need the raw samples, but in many hospitals and clinics the ECG exists only as a printout, a scan, a photograph or a PDF. That is the practical case for image-based classification. A trace rendered as an image can be given to any standard vision network, and transfer learning from natural-image models is available off the shelf [25], [26]. Khan *et al.* released a public set of ECG images from cardiac and COVID-19 patients [23], one of the few openly available image collections we found. The price of working with images is real: amplitude calibration is lost or distorted, leads can overlap or be cropped, grid lines and handwriting add clutter, and fine timing detail is limited by resolution. We did not find a widely adopted benchmark for 12-lead ECG image classification comparable to PTB-XL for signals, so accuracies reported by different image-based papers are hard to compare. We come back to this in Section IX.

F. Why ResNet50

Very deep plain networks become harder to optimise as layers are added. He *et al.* addressed this by letting each block learn a residual function and adding the block's input back through an identity shortcut [27]:

$$y = F(x, \{W\}) + x \quad (1)$$

Fig. 3 shows the bottleneck block used in ResNet50. A 1×1 convolution reduces the channel count, a 3×3 convolution does the spatial work, and a second 1×1 convolution restores the width. Sixteen of these blocks plus a stem and a classifier give the 50 weight layers, roughly 25 million parameters, and a final feature map of 7×7×2048 for a 224×224 input

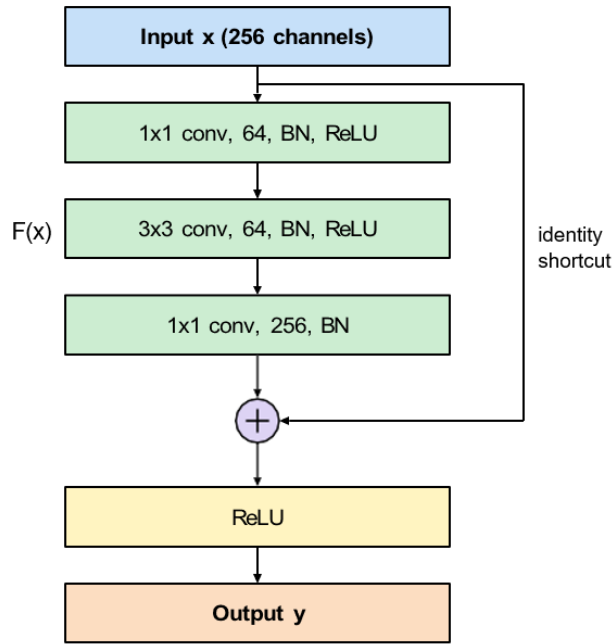


Fig. 3. Bottleneck residual block used in ResNet50 (BN: batch normalisation).

We prefer ResNet50 here for four practical reasons. First, ImageNet-pretrained weights exist in every major framework, which matters when the labelled ECG set is small. Second, it is among the most thoroughly studied architectures, so its failure modes are better understood than those of newer models. Third, its last convolutional stage gives Grad-CAM a clean layer to work with. Fourth, it is small enough to run on a modest server. We do not claim it is the most accurate option; vision transformers and newer convolutional families may do better, and 1-D networks on raw signals probably do better still when signals are available. For an image-based, explainable, doctor-facing prototype, ResNet50 is a defensible and low-risk baseline.

A caveat on pretraining. Pretraining is not a free lunch. Natural photographs and ECG traces look very little alike, and Raghu et al. found for several medical imaging tasks that ImageNet pretraining gave modest gains at best, with much smaller networks performing comparably [28]. Yosinski et al. had earlier shown that transferability falls as the source and target tasks drift apart [26]. We therefore plan to compare fine-tuning the whole network, fine-tuning only the later stages, and training from scratch, and to report which works best on our data rather than assume that pretraining helps.

G. Single-Label versus Multi-Label Output

A real ECG often shows more than one finding. In PTB-XL a record can carry several diagnostic statements, and therefore belong to more than one superclass [16]. A five-way softmax output, which forces a single winner, is a simplification of that reality. We start with it because it is easy to explain and to evaluate, but multi-label output with independent sigmoid units and per-class thresholds is the more faithful design and is listed among the next steps in Section XI.

5. EXPLAINABLE AI FOR ECG MODELS

A. Why Explanation Matters

A class label alone gives a doctor little to argue with. If the model says “MI”, the natural reply is “where do you see it?”. Explainability methods try to answer that question. Their use in medicine has been surveyed by Tjoa and Guan [7], while Rudin has argued forcefully that high-stakes decisions are better served by inherently interpretable models than by post-hoc explanations of black boxes [29]. Both points are relevant to our design, and Table IV lists the main families of methods.

TABLE IV

Common Explanation Methods and Their Limits

Method	How it works	Main limitation
Saliency map [30]	Gradient of score w.r.t. input pixels	Noisy; sensitive to training details
CAM [31]	Weights final feature maps via global pooling layer	Needs a specific architecture
Grad-CAM [32]	Gradient-weighted sum of last conv. maps	Coarse; may not match clinical features

Method	How it works	Main limitation
Grad-CAM++ [33]	Higher-order gradient weighting	Same coarse resolution; more compute
LIME [34]	Local surrogate model on perturbed inputs	Unstable; slow for images
SHAP [35]	Shapley-value feature attribution	Costly; approximations for images

B. Grad-CAM

Grad-CAM [32] works on a trained network without changing it.

For a target class c it takes the gradient of the class score S_c with respect to each feature map A of the last convolutional layer, averages it over spatial positions to get a weight, and forms a ReLU-filtered weighted sum:

$$\alpha_c = (1/Z) \sum_i \sum_j \partial S_c / \partial A(i,j) \quad (2)$$

$$L_c = \text{ReLU}(\sum_k \alpha_{k,c} A_k) \quad (3)$$

Here Z is the number of spatial positions. The result is a coarse map at feature-map resolution, which is upsampled and overlaid on the input, as in Fig. 4. Only positive contributions survive the ReLU, so the map shows what supports the chosen class rather than what argues against it.

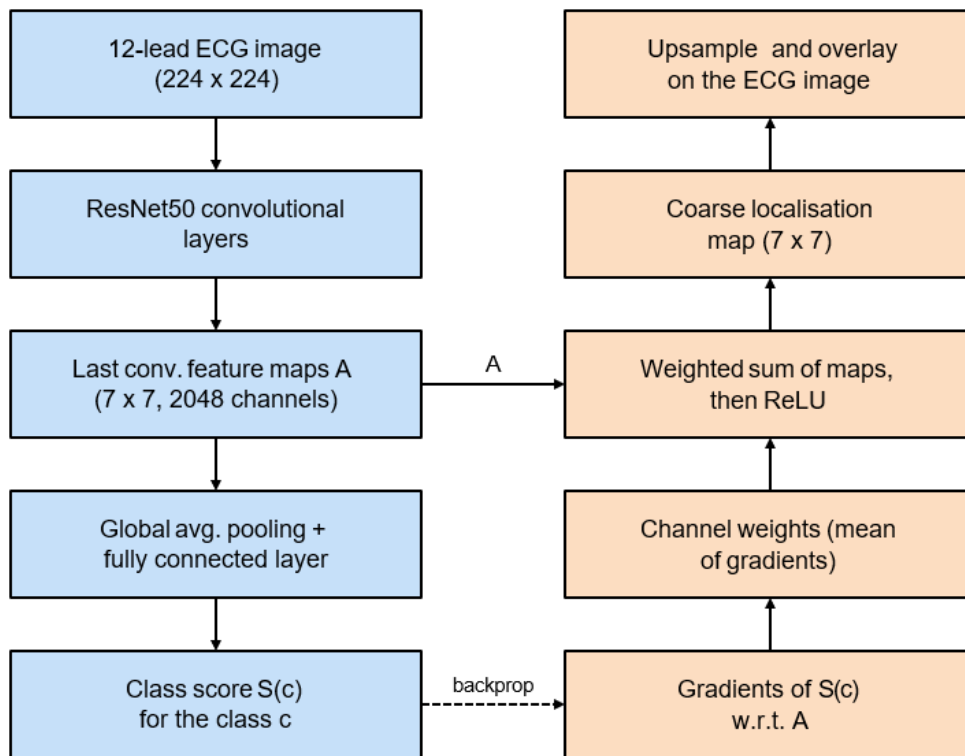


Fig. 4. Grad-CAM pipeline for an ECG image classifier (forward pass on the left, explanation branch on the right).

C. Limits for ECG Images

The practical limits are worth stating. With a 224×224 input the final ResNet50 feature map is only 7×7, so each cell covers a 32×32 pixel patch. A heat map at that resolution can point to a region of the tracing but cannot isolate a single ST segment. Adebayo *et al.* showed that some saliency methods produce maps that barely change when the network's weights are randomised, so a plausible-looking picture does not prove that the model used the highlighted evidence [36]; plain gradients and Grad-CAM held up better than some alternatives in that study, which is encouraging but not a guarantee for ECG models. A heat map that lands on the right lead also says only that the model looked somewhere sensible. It does not explain why the finding is MI rather than STTC.

Few of the ECG papers we read tested their explanations with cardiologists; quality was usually judged by eye. Without that step, “explainable” can end up meaning “accompanied by a coloured picture”. A modest remedy, which we plan to adopt, is to ask clinicians to mark the regions they consider diagnostic on a sample of images and measure how well the model's map overlaps with them.

D. Explaining the Other Modules

The risk and QTc modules need explanations too, although of a simpler kind. For the Random Forest, a ranked list of the factors that mattered for this particular patient is more useful than a global importance chart. For the QTc module the explanation is the rule itself: which threshold was crossed, which drug category applied, and which patient factors added to the risk. Rule-based reasons are one place where interpretability by design, in Rudin's sense [29], is easy to achieve.

6. CARDIOVASCULAR RISK ASSESSMENT

A. Conventional Risk Scores

Clinicians have long used regression-based scores such as the Framingham general cardiovascular risk profile [37] and the pooled cohort equations from the 2013 ACC/AHA guideline [38]. They are transparent and built on large cohorts, but they were derived for primary prevention in particular populations, use a small fixed set of inputs, and assume simple additive effects. Applying them to a different population usually calls for recalibration.

B. The UCI Heart Disease Data

Machine-learning papers on heart disease rely overwhelmingly on the UCI Heart Disease collection [24]. It originates from four clinical sites (Cleveland, Hungary, Switzerland and the Long

Beach VA centre) and was described by Detrano et al. [39]. The Cleveland subset, with 303 records and 14 commonly used attributes, is what most papers actually use. Table V lists the attributes. Merged versions of about a thousand rows circulate in public repositories (CardioAssist plans to use a version of roughly 1,026 records), and these are reported to contain repeated records. Duplicates that fall on both sides of a train–test split inflate accuracy, so the data must be checked before any score is trusted.

TABLE V

Attributes in the UCI Heart Disease Data (Cleveland Format)

Attribute	Type	Meaning
age	Numeric	Age in years
sex	Binary	Male or female
cp	Categorical	Chest pain type (four values)
trestbps	Numeric	Resting blood pressure (mm Hg)
chol	Numeric	Serum cholesterol (mg/dl)
fbs	Binary	Fasting blood sugar above 120 mg/dl
restecg	Categorical	Resting ECG result
thalach	Numeric	Maximum heart rate achieved
exang	Binary	Exercise-induced angina
oldpeak	Numeric	ST depression, exercise vs. rest
slope	Categorical	Slope of peak-exercise ST segment
ca	Numeric	Major vessels coloured by fluoroscopy (0–3)
thal	Categorical	Thalassaemia test result

C. Random Forest

Breiman's Random Forest builds many decision trees on bootstrap resamples of the training data and, at each split, considers only a random subset of the features [40]. The predicted class is the majority vote of the B trees (Fig. 5):

$$\hat{y}(x) = \text{majority vote } \{ h(x) : b = 1, \dots, B \} \quad (4)$$

Each tree chooses splits that reduce an impurity measure such as

the Gini index, $G = 1 - \sum_k p_k^2$, where p_k is the share of class k in a node.

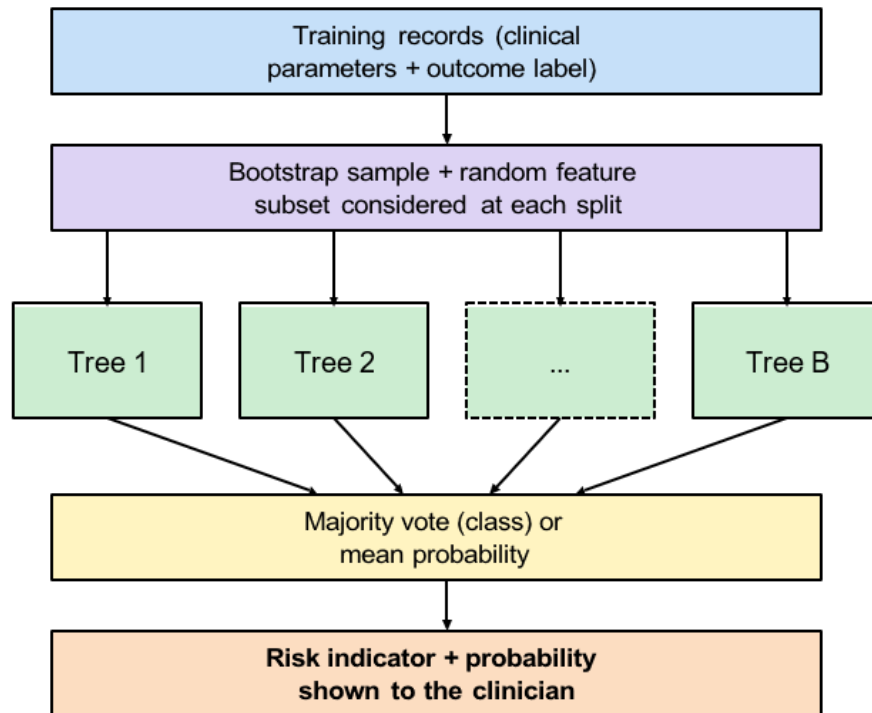


Fig. 5. Random Forest for clinical risk classification.

Several properties suit this problem. The data are small and of mixed type (numeric, binary, categorical), which tree ensembles handle with little preprocessing; they need modest tuning; they give out-of-bag error estimates without a separate validation set; and they offer feature-importance scores a doctor can read. Grinsztajn et al. found that tree-based models still tended to beat deep networks on medium-sized tabular data [41], which supports this choice over a neural network. One caveat: impurity-based importance can favour variables with many distinct values, so permutation importance or SHAP [35] is a safer way to report which factors drove a prediction.

D. Reported Results and Their Reliability

Published accuracies on the UCI data vary widely. Mohan et al. combined a Random Forest with a linear model and reported an accuracy of 88.7 percent [42]; Ali et al. proposed an ensemble deep-learning system with feature fusion [43]. Some papers on this dataset claim accuracies above 95 percent, and we read such claims cautiously: with a few hundred patients a single train-test split has wide confidence intervals, and results are rarely checked on another hospital's data. Calibration deserves attention as well, since a model can rank patients well and still output probabilities that are too confident [44]. In CardioAssist the output will be shown as a risk indicator next to its main contributing factors, never as a diagnosis. Our minimum standards are stratified cross-validation, duplicate removal before splitting, recall and AUC reported alongside accuracy, and published calibration curves.

E. B. Overlap with the ECG Module

Some UCI attributes, such as the resting ECG result, ST depression and the slope of the ST segment, describe the same physiology that the ECG module reads from the tracing. In a combined system this raises a design question: should the ECG model's output feed the risk model, or should the two stay separate? The UCI data contain no ECG images, so a fused model cannot be trained on them, and a hand-built combination would have no ground truth to be checked against. For that reason CardioAssist keeps the two modules independent for now and presents their outputs side by side. Learning from both together is listed as future work and would need a dataset in which ECGs and clinical variables belong to the same patients.

F. Choosing an Operating Point

A risk model outputs a probability, and someone has to decide where to draw the line. A low threshold catches more patients at risk but sends more people for further tests; a high threshold does the opposite. Because the clinician still checks the result, we lean toward sensitivity, but the right balance depends on what happens to a flagged patient. We will therefore show the probability itself and report performance at several thresholds instead of at a single one.

7. QTc AND MEDICATION SAFETY

A. Background

Many drugs, from antiarrhythmics to some antibiotics, antipsychotics and antiemetics, block cardiac potassium channels and lengthen ventricular repolarisation, which shows up as a longer QT interval. A markedly prolonged QT can precede torsades de pointes, a polymorphic ventricular tachycardia

that may degenerate into ventricular fibrillation [45]. Because QT shortens as heart rate rises, it is normally corrected for heart rate to give QTc. Table VI lists three common formulas, where RR is the interval between successive R-waves in seconds.

TABLE VI

Common Heart-Rate Correction Formulas for the QT Interval

Formula	Expression	Remark
Bazett [46]	$QTc = QT / \sqrt{RR}$	Built into most ECG machines; over-corrects at fast heart rates [47]
Fridericia [48]	$QTc = QT / RR^{1/3}$	Less dependent on heart rate in comparisons on normal ECGs [47]
Framingham [49]	$QTc = QT + 154(1 - RR)$	Linear; QT in ms, RR in s

There is no single universal cut-off. Reference limits differ between sources, and values around 450 ms in men and 460 ms in women are commonly quoted as the start of the borderline range [50]. A QTc above 500 ms, or a rise of more than 60 ms from baseline after a new drug is started, is widely treated as a warning sign [45], [51]. The exact thresholds used in a deployed system should follow the guideline of the institution that uses it.

B. Drug Lists and Risk Scores

For drugs, the CredibleMeds list is widely used and sorts agents into known, possible and conditional risk of torsades de pointes [52]. Tisdale *et al.* developed and validated a points-based score for hospitalised patients that combines factors such as age, sex, loop diuretic use, low potassium, baseline QTc, heart failure and the number of QT-prolonging drugs [53]. These two resources are usually consulted separately from the ECG itself, which is exactly the fragmentation an integrated tool could reduce.

C. Measurement Uncertainty

Measuring QT is itself uncertain. The end of the T-wave is often hard to define, results differ between manual and automated methods, and the problem grows when T-waves are flat, biphasic or merged with U-waves [54]. This matters for CardioAssist because the QT value may come either from the doctor's entry or from the ECG image, and the system must state which one it used.

D. Rules, Machine Learning, or Both

Rule-based logic has two advantages here: it is auditable, and it encodes guideline thresholds directly. Its weakness is that it treats risk factors one at a time. A learned component can in principle weigh combinations, such as female sex, low potassium and two interacting drugs, but it needs outcome-labelled data, which is hard to obtain. A hybrid is therefore attractive: rules set the floor, and the ML component, once validated, may refine the level. In our planned design the rules stay visible to the user, and the ML component cannot lower a warning that the rules have raised (Fig. 6). We did not find published systems that connect QTc screening to automatic ECG analysis and cardiovascular risk estimation in one interface

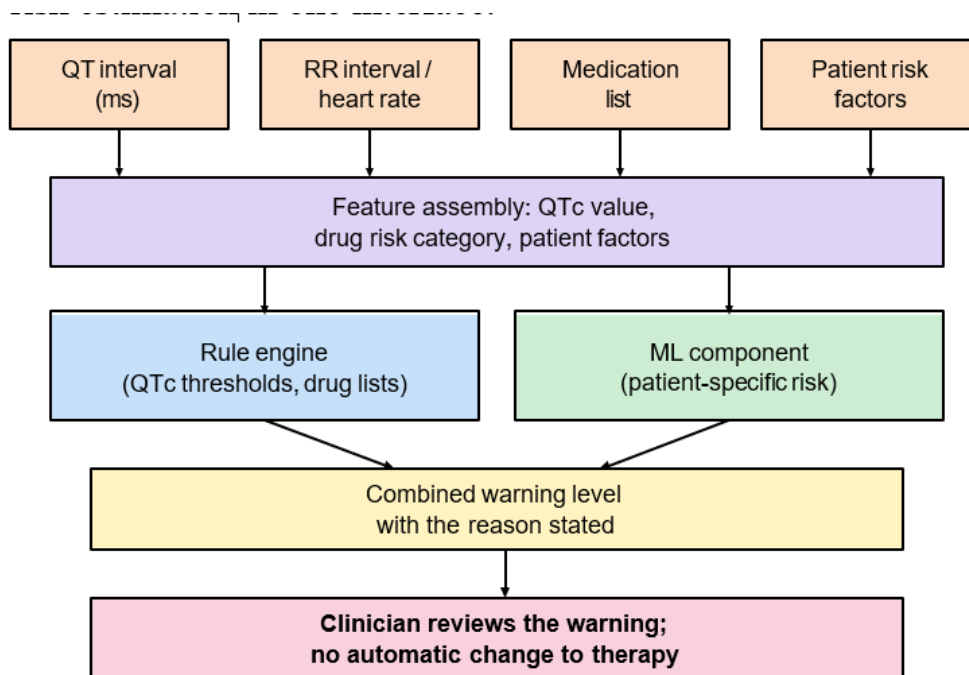


Fig. 6. Planned QTc and medication safety workflow.

E. Situations Where QTc Needs Extra Care

A single corrected value is not equally trustworthy in every patient. In atrial fibrillation the RR interval varies from beat to beat, so the choice of RR for the correction changes the answer. When the QRS complex is wide, as in bundle branch block or ventricular pacing, part of the apparent QT prolongation comes from slow depolarisation rather than repolarisation, and standardisation documents discuss adjusted measures for these cases [50]. Reference limits also differ by sex, and QTc is on average longer in women, which is one reason sex appears in drug-risk scores [53]. CardioAssist will flag these situations when the inputs allow, and will say that the number should be interpreted with caution rather than silently apply a single formula to everyone.

VIII. PLANNED CARDIOASSIST FRAMEWORK

A. Architecture

CardioAssist is planned as a web application with a doctor-oriented dashboard (Fig. 7). The front end uses HTML, CSS and JavaScript. The server side is written in Python, with PyTorch [55] for the ECG model and scikit-learn [56] for the risk model. The three modules run side by side and report to one results page, so the doctor does not have to move between tools. Table VII summarises the modules. All of them are at the planning and prototyping stage

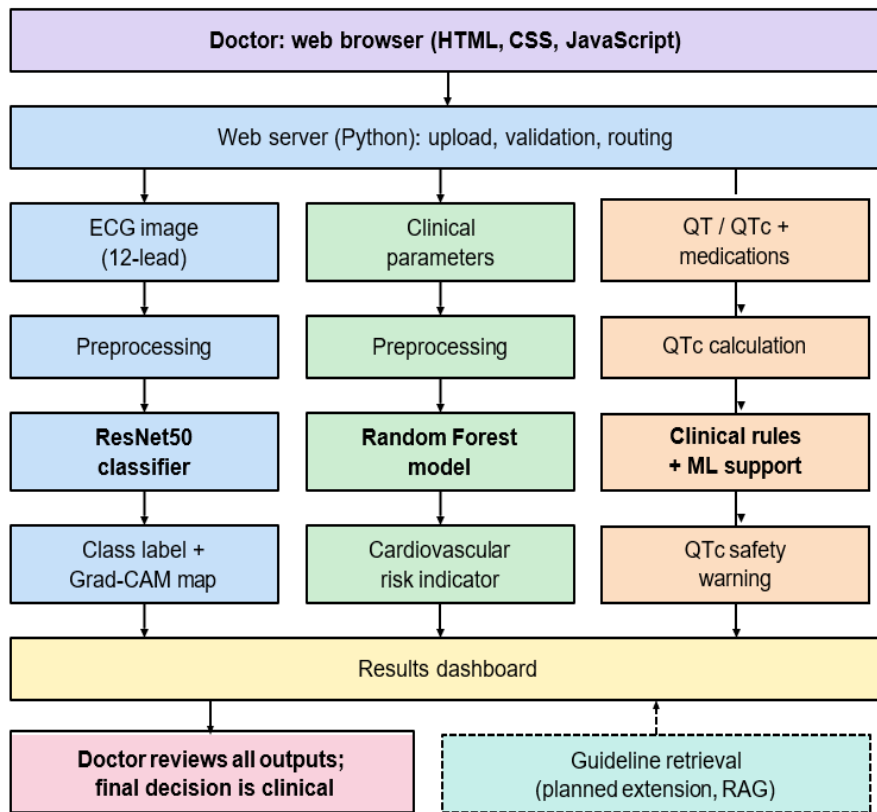


Fig. 7. Planned CardioAssist architecture. The dashed box marks a future extension.

TABLE VII

Planned Modules (Status: Under Development)

Module	Input	Method	Output
ECG	12-lead ECG image	ResNet50 + Grad-CAM	Normal / MI / STTC / CD / HYP; heat map
Risk	Clinical parameters	Random Forest	Risk indicator, key factors
QTc safety	QT/QTc, drugs, patient factors	Clinical rules + ML support	Warning level, reason

B. Planned Methods

ECG module. Images will be resized to 224×224, normalised with ImageNet statistics, and lightly augmented with small shifts, scaling and brightness changes [57]. Horizontal flips will be avoided because they change the meaning of the leads. The pretrained ResNet50 will be fine-tuned with a new five-way output layer, a class-weighted cross-entropy loss to handle imbalance, and the Adam optimiser [58].

Risk module. Features will be imputed and encoded, duplicate records removed, and the Random Forest tuned with stratified cross-validation. Feature importance will be reported with a permutation-based method.

QTc module. QTc will be computed with a configurable correction formula, with Fridericia and Bazett as the candidates. Thresholds will be taken from a stated guideline, drug flags from a licensed drug list, and the ML component will sit behind the rules as described in Section VII-D.

C. Evaluation Plan

Because this is a safety-relevant tool, the evaluation plan matters as much as the models. For the ECG module we will split by patient, report per-class and macro-averaged F1 and AUC, and use the PTB-XL patient-wise folds when that dataset is added. With precision $P = TP/(TP+FP)$ and recall $R = TP/(TP+FN)$,

use the PTB-XL patient-wise folds when that dataset is added. With precision $P = TP/(TP+FP)$ and recall $R = TP/(TP+FN)$,

$$F_1 = 2PR / (P + R) \quad (5)$$

and macro-F1 is the unweighted mean of the per-class values, so rare classes count as much as common ones. Calibration of the risk module will be checked with the Brier score,

$$BS = (1/N) \sum_i (p_i - o_i)^2 \quad (6)$$

where p_i is the predicted probability and o_i the observed outcome. We will also test the ECG module against image degradation (blur, rotation, scan noise), compare Grad-CAM maps with clinician-marked regions, and check the QTc module against hand-worked cases.

Pitfalls to avoid. Prediction-model papers in medicine often fall short on reporting rather than on modelling. The TRIPOD statement lists what a paper on a diagnostic or prognostic model should disclose, from data sources and missing-value handling to how performance was estimated [59], and we will use it as a checklist. The pitfalls we most want to avoid are four: splitting at the record level when patients have several records; tuning on the test set; reporting accuracy alone when classes are imbalanced; and presenting a single-split score without a confidence interval. For rendered ECG images there is a fifth trap, since the plotting style or scanner can act as a shortcut that tracks the label. We will therefore test on images made by a different rendering pipeline.

D. Interface Principles

The interface follows five principles drawn from the decision-support literature [3], [4]:

- Show the evidence beside the prediction: the heat map, the main contributing factors, and the rule that triggered a warning.
- Show uncertainty: probabilities with calibration notes instead of bare labels, and a clear message when an uploaded image is of low quality.
- Keep the doctor in control: every output is advisory, and nothing is written to a prescription or a record automatically.
- Keep alerts rare and specific, to limit alert fatigue.
- Log inputs, model versions and outputs so that any result can be audited later.

In a typical session the doctor uploads a patient's 12-lead ECG image and enters the clinical parameters and current medication list. The results page then shows the predicted ECG category with its heat map, a risk indicator with the factors that contributed most, and any QTc warning together with the reason it was raised. The doctor reads these alongside the patient and decides.

Table VIII summarises the alternatives we weighed for each module and the reason for the current choice.

TABLE VIII

Candidate Methods Considered for Each Module

Module	Candidates	Current choice and reason
ECG classifier	ResNet50; EfficientNet or ViT on images; 1-D CNN on raw signal	ResNet50: pretrained, well studied, suits Grad-CAM; a 1-D CNN is preferred once signals are available
Risk model	Logistic regression; Random Forest; gradient boosting; neural network	Random Forest: mixed data, little tuning, out-of-bag error; logistic regression kept as a baseline
QTc module	Rules only; ML only; hybrid	Hybrid with rules first: auditable, and ML cannot lower a rule-based warning
Explanation	Saliency; Grad-CAM; LIME or SHAP	Grad-CAM for the image model; permutation importance or SHAP for the risk model

E. Design Rationale

Each module uses a well-studied method whose failure modes are documented, which is a deliberate choice for a tool that clinicians may come to rely on. We would rather start from a transparent baseline and measure it honestly than start from a novel architecture whose weaknesses are unknown. What we expect to add is the integration, the explicit explanation check, and a validation path, not a new network design.

F. Intended Use and Clinical Role

CardioAssist is meant as a decision-support tool, not as an autonomous diagnostic system. Its outputs are an extra source of information for a qualified clinician, who stays responsible for the decision. We say this plainly because the line between support and replacement can blur in practice, especially when a system is fast and usually right. If the benefits we hope for do appear, they would be a quicker first look at an ECG, a second opinion on risk, a visual cue showing where the model looked, a reminder when a drug combination deserves attention, and less switching between separate tools. These are expectations to be tested, not findings. The system should not be used in patient care until the validation steps in Section XI have been completed and its regulatory position has been checked.

RESEARCH GAPS

Reading across the four areas, six gaps stand out. Table IX pairs each one with the evidence behind it and the response planned in CardioAssist.

TABLE IX

Research Gaps and Planned Responses

Gap	Evidence in the literature	Planned response
Single-task tools	Most studies address one task; support tools work best inside the clinical workflow [3], [4]	One interface for ECG, risk and QTc
Image-based ECG	Large results are signal-based [11], [13]; no standard image benchmark found	Use PTB-XL classes; test scan and rendering variation
Unchecked explanations	Saliency maps can mislead [36]; clinician checks are rare	Compare Grad-CAM with clinician-marked regions
Small tabular data	UCI data are small and merged versions repeat records [24], [39]	Deduplicate; cross-validate; report calibration
QTc apart from ECG	QT is hard to measure [54]; drug lists and scores used separately [52], [53]	Rule-first hybrid; state the QT source
External validation	Held-out data from other sites are often missing in imaging AI [60]	Staged external validation (Section XI)

X. CHALLENGES

Data and labels. PTB-XL labels come from annotation of reports and cardiologist review, with varying certainty, and ECG image datasets are smaller and less uniformly labelled. Class imbalance is common, with normal records far outnumbering some abnormal classes. Label noise puts a ceiling on what any model can show.

Domain shift. A model trained on one set of recorders, paper speeds and printing styles may stumble on another. None of the main public datasets discussed here (PTB-XL, UCI Heart Disease, the Chapman–Shaoxing data) was collected in India, so local validation will be necessary before any claim about Indian patients can be made.

Human factors. Decision support can fail even when the model is right. Clinicians may defer to a wrong suggestion, or learn to ignore frequent alerts [4]. Kelly et al. list further obstacles to clinical impact, from workflow fit to regulation [61], and Topol argues that the realistic goal is human and machine working together rather than machine alone [62]. This is why CardioAssist is designed so that the final decision stays with the clinician.

Regulation and privacy. Software that influences diagnosis is likely to be treated as medical device software in many jurisdictions and would need review before clinical use. ECG images carry patient identifiers, and processing them falls under India's Digital Personal Data Protection Act, 2023 [63], so storage, consent and access control need to be designed in from the start.

Subgroup performance. Normal ECG values vary with age and sex, and children, older adults and women are not equally represented in every dataset. A model with a good overall score can still do worse for one group. PTB-XL records age and sex, so subgroup results can at least be reported for the ECG module; for the other datasets the metadata are thinner, and we will say so where that limits what can be concluded.

Limitations. This review is narrative, so relevant studies may have been missed, and our reading of reported numbers depends on what the papers disclose. Several statements about CardioAssist describe intentions rather than measured behaviour. The planned system will inherit the weaknesses of its data: ECG image quality, the make-up of public datasets, and the small size of the UCI data. A machine-learning risk score cannot replace a full clinical evaluation, QTc interpretation depends on the measurement method and the patient, and nothing in this paper amounts to clinical validation.

XI. FUTURE DIRECTIONS

Table X sets out the staged path we intend to follow. The stages are ordered by dependency, not by calendar dates, since progress will depend on data access and clinical collaboration.

TABLE X

Staged Development and Validation Plan

Stage	Focus	Expected outcome
1	Consolidate the three modules on current data; error analysis	Internal baseline report
2	Add PTB-XL; patient-wise folds; render signals as images	Generalisation tests
3	Raw-signal models, beat-level analysis, multi-label output, arrhythmia detection	Comparison with image baseline
4	Guideline retrieval (RAG) with visible sources	Referenced support text
5	External data, clinician review, prospective study, safety and regulatory review	Evidence for a deployment decision

Fig. 8 shows the validation ladder behind Stage 5 in more detail. Each rung answers a different question, and a result that holds only on the lower rungs should not be presented as evidence of clinical usefulness.

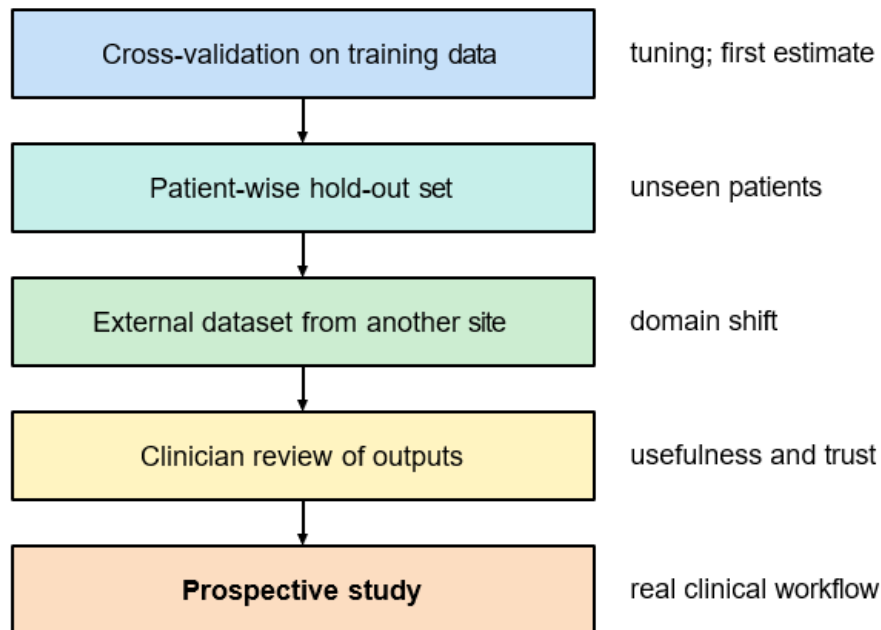


Fig. 8. Validation ladder planned for CardioAssist.

Guideline retrieval. Retrieval-augmented generation [64] could fetch passages from trusted sources, such as the ESC guidelines on ventricular arrhythmias [51], and show them beside a prediction. The risk is that a language model paraphrases a guideline wrongly. The planned safeguard is to restrict retrieval to a curated corpus, display the source passage and its reference, and refuse to answer outside it.

Richer ECG analysis and explainability. Moving from images to raw signals would remove the rendering step and allow finer timing analysis, and multi-label output would reflect that real ECGs often show more than one finding. On the explanation side, Grad-CAM++ [33] and model-agnostic methods can be compared against Grad-CAM, using the clinician-overlap measure described earlier.

XII. CONCLUSION

This paper reviewed the research behind the three tasks that CardioAssist aims to bring together: ECG classification with explanation, cardiovascular risk estimation, and QTc medication safety. The evidence for each task is strong enough to justify building on it, but it is uneven. Signal-based ECG deep learning is well developed while image-based work lacks a shared benchmark. Heart-disease prediction on small public data is easy to over-report. Grad-CAM gives a useful but coarse view that has rarely been checked against clinicians. QTc screening is well grounded clinically yet seldom connected to automated ECG reading.

TABLE XI

Summary of the Reviewed Evidence by Task

Task	Strongest evidence	Main weakness	Implication for CardioAssist
ECG classification	Large signal-based studies with cardiologist-level rhythm results [11], [13]	Image-based work lacks a shared benchmark	ResNet50 as a baseline; add PTB-XL and signal models later
Explainability	Grad-CAM is simple to attach to trained CNNs [32]	Coarse and rarely validated by clinicians [36]	Check maps against clinician-marked regions
Risk estimation	Tree ensembles do well on tabular data [41]	Small data, repeated records, little external testing	Deduplicate; report calibration and intervals
QTc safety	Established formulas, drug lists and risk scores [52], [53]	Measurement uncertainty; tools used separately	Rule-first hybrid with the QT source stated

CardioAssist is planned as a modest answer to those gaps: familiar methods, joined in one doctor-facing tool, with an evaluation plan that treats validation as part of the design. The system is not yet ready, and no claim about its accuracy is made here. The next steps are to consolidate the prototype, test it on larger and external data, and put its explanations in front of clinicians.

ACKNOWLEDGMENT

The authors thank the Department of Computer Engineering, Smt. Indira Gandhi College of Engineering, Ghansoli, for its support during this work.

ABBREVIATIONS

AUC, area under the ROC curve; CD, conduction disturbance; CDSS, clinical decision support system; CNN, convolutional neural network; CVD, cardiovascular disease; ECG, electrocardiogram; Grad-CAM, gradient-weighted class activation mapping; HYP, hypertrophy; MI, myocardial infarction; NORM, normal ECG; QTc, heart-rate-corrected QT interval; RAG, retrieval-augmented generation; STTC, ST/T change; TdP, torsades de pointes.

REFERENCES

1. G. A. Roth *et al.*, "Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study," *J. Am. Coll. Cardiol.*, vol. 76, no. 25, pp. 2982–3021, 2020.
2. D. Prabhakaran, P. Jeemon, and A. Roy, "Cardiovascular diseases in India: Current epidemiology and future directions," *Circulation*, vol. 133, no. 16, pp. 1605–1620, 2016.
3. K. Kawamoto, C. J. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success," *BMJ*, vol. 330, no. 7494, p. 765, 2005.
4. R. T. Sutton *et al.*, "An overview of clinical decision support systems: Benefits, risks, and strategies for success," *npj Digit. Med.*, vol. 3, Art. no. 17, 2020.
5. S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Comput. Biol. Med.*, vol. 122, Art. no. 103801, 2020.
6. K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," *Nat. Rev. Cardiol.*, vol. 18, pp. 465–478, 2021.
7. E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, 2021.
8. J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985.

9. S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, 2016.
10. U. R. Acharya *et al.*, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, 2017.
11. A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nat. Med.*, vol. 25, pp. 65–69, 2019.
12. P. Rajpurkar, A. Y. Hannun, M. Haghpahani, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," arXiv:1707.01836, 2017.
13. A. H. Ribeiro *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nat. Commun.*, vol. 11, Art. no. 1760, 2020.
14. Z. I. Attia *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction," *Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.
15. A. Mincholé and B. Rodriguez, "Artificial intelligence for the electrocardiogram," *Nat. Med.*, vol. 25, pp. 22–23, 2019.
16. P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, Art. no. 154, 2020.
17. A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
18. N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1519–1528, 2021.
19. T. Mehari and N. Strodthoff, "Self-supervised representation learning from 12-lead ECG data," *Comput. Biol. Med.*, vol. 141, Art. no. 105114, 2022.
20. E. A. Perez Alday *et al.*, "Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Meas.*, vol. 41, no. 12, Art. no. 124003, 2020.
21. J. Zheng *et al.*, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Sci. Data*, vol. 7, Art. no. 48, 2020.
22. G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
23. A. H. Khan, M. Hussain, and M. K. Malik, "ECG images dataset of cardiac and COVID-19 patients," *Data Brief*, vol. 34, Art. no. 106762, 2021.
24. A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart Disease," UCI Machine Learning Repository, 1988. [Online]. Available: <https://archive.ics.uci.edu/dataset/45/heart+disease>
25. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 248–255.
26. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3320–3328.
27. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
28. M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 3347–3357.
29. C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, pp. 206–215, 2019.
30. K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," arXiv:1312.6034, 2013.
31. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2921–2929.
32. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
33. A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2018, pp. 839–847.

34. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 1135–1144.
35. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2017, pp. 4765–4774.
36. J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2018, pp. 9505–9515.
37. R. B. D'Agostino Sr. et al., "General cardiovascular risk profile for use in primary care: The Framingham Heart Study," *Circulation*, vol. 117, no. 6, pp. 743–753, 2008.
38. D. C. Goff Jr. et al., "2013 ACC/AHA guideline on the assessment of cardiovascular risk," *Circulation*, vol. 129, no. 25, suppl. 2, pp. S49–S73, 2014.
39. R. Detrano et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," *Am. J. Cardiol.*, vol. 64, no. 5, pp. 304–310, 1989.
40. L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
41. L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in Proc. NeurIPS Datasets Benchmarks Track, 2022.
42. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.
43. . Ali et al., "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Inf. Fusion*, vol. 63, pp. 208–222, 2020.
44. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), PMLR 70, 2017, pp. 1321–1330.
45. D. M. Roden, "Drug-induced prolongation of the QT interval," *N. Engl. J. Med.*, vol. 350, no. 10, pp. 1013–1022, 2004.
46. H. C. Bazett, "An analysis of the time-relations of electrocardiograms," *Heart*, vol. 7, pp. 353–370, 1920.
47. S. Luo, K. Michler, P. Johnston, and P. W. Macfarlane, "A comparison of commonly used QT correction formulae: The effect of heart rate on the QTc of normal ECGs," *J. Electrocardiol.*, vol. 37, suppl., pp. 81–90, 2004.
48. L. S. Fridericia, "Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken," *Acta Med. Scand.*, vol. 53, pp. 469–486, 1920.
49. A. Sagie, M. G. Larson, R. J. Goldberg, J. R. Bengtson, and D. Levy, "An improved method for adjusting the QT interval for heart rate (the Framingham Heart Study)," *Am. J. Cardiol.*, vol. 70, no. 7, pp. 797–801, 1992.
50. P. M. Rautaharju et al., "AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: Part IV," *Circulation*, vol. 119, no. 10, pp. e241–e250, 2009.
51. S. G. Priori et al., "2015 ESC guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death," *Eur. Heart J.*, vol. 36, no. 41, pp. 2793–2867, 2015.
52. R. L. Woosley, C. W. Heise, and K. A. Romero, "QTdrugs list," CredibleMeds, AZCERT. [Online]. Available: <https://www.crediblemeds.org>
53. J. E. Tisdale et al., "Development and validation of a risk score to predict QT interval prolongation in hospitalized patients," *Circ. Cardiovasc. Qual. Outcomes*, vol. 6, no. 4, pp. 479–487, 2013.
54. P. G. Postema and A. A. M. Wilde, "The measurement of the QT interval," *Curr. Cardiol. Rev.*, vol. 10, no. 3, pp. 287–294, 2014.
55. A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2019, pp. 8024–8035.
56. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
57. C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, Art. no. 60, 2019.
58. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.
59. G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," *Ann. Intern. Med.*, vol. 162, no. 1, pp. 55–63, 2015.

-
60. X. Liu et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis," *Lancet Digit. Health*, vol. 1, no. 6, pp. e271–e297, 2019.
 61. C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Med.*, vol. 17, Art. no. 195, 2019.
 62. E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, pp. 44–56, 2019.
 63. Ministry of Law and Justice, Government of India, "The Digital Personal Data Protection Act, 2023," Act No. 22 of 2023.
 64. P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020, pp. 9459–9474.