**Heritage Journal of Health Sciences and Research**
Kapoor S. Heritage J Health Sci Res. 2026;1(2):38-42
**Volume 1, Issue 2, July–December 2026** | Open Access | Peer Reviewed
Received: 26-10-2026 | Accepted: 03-11-2026 | Published: 06-11-2026
**Research Article** | Heritage J Health Sci Res

# Artificial Intelligence-Based Early Detection of Cardiovascular Diseases Using Multimodal Clinical Data

**Dr. Simran Kapoor** [1*]
[1] Department of Respiratory Medicine, Horizon Medical College, Indore, India

## Abstract

Cardiovascular diseases (CVDs) caused about 20.5 million deaths in 2021, close to one-third of all deaths worldwide, and many of these deaths follow a long silent phase in which disease is present but undiagnosed. Artificial intelligence (AI) can detect this silent phase by reading patterns in routine clinical data that clinicians cannot see. This paper reviews published, peer-reviewed evidence on AI for early CVD detection across four data modalities: the electrocardiogram (ECG), cardiac imaging (echocardiography and retinal photographs), electronic health records (EHR), and their combination. Across the studies reviewed, a single 12-lead ECG read by a convolutional neural network identified left ventricular systolic dysfunction with an AUC of 0.93, and machine-learning risk models built on EHR and cohort data raised AUC from 0.724–0.728 (guideline scores) to 0.764–0.774. A pragmatic randomized trial showed that an AI-ECG alert raised new diagnoses of low ejection fraction by 32% in routine primary care. Multimodal fusion frameworks improved AUROC by 6–33% over single-source models. We propose a late-fusion architecture that combines ECG, imaging and EHR embeddings, and we discuss the barriers to deployment: data bias, explainability, external validation and regulation.

**Keywords:** artificial intelligence, deep learning, cardiovascular disease, early detection, multimodal fusion, electrocardiogram, electronic health records, echocardiography

## Introduction

Cardiovascular disease is the leading cause of death in the world. The World Heart Federation reports that CVDs accounted for 20.5 million deaths in 2021, about one-third of global deaths, and roughly 80% of these occurred in low- and middle-income countries[1]. India and the wider WHO South-East Asia Region carry a large share of this burden: an estimated 4.3 million people in the region died from CVDs in 2021[2].

Much of this mortality is preventable if disease is found early. Conditions such as asymptomatic left ventricular dysfunction (ALVD), paroxysmal atrial fibrillation (AF) and subclinical atherosclerosis often produce no symptoms for years. ALVD alone is present in 3–6% of the general population and is treatable when found[3]. Yet the tools that confirm these conditions—echocardiography, long-term rhythm monitoring, CT angiography—are costly and are not offered to people who feel well.

Current prevention relies on risk scores such as Framingham, SCORE and the ACC/AHA pooled cohort equations. These scores use fewer than ten variables and assume simple, mostly linear relationships. As a result they miss many people who go on to have events and over-treat others[4].

AI offers a different approach. Deep neural networks can learn features directly from raw signals and images, and gradient-boosted or ensemble models can capture non-linear interactions among hundreds of EHR variables. Each data source sees a different side of the heart: the ECG records electrical activity, echocardiography records structure and motion, the retina shows the state of small blood vessels, and the EHR records risk factors, laboratory values and history. Combining them—multimodal learning—should give a more complete picture than any single source.

This paper makes three contributions:

1. It summarises published evidence on AI for early CVD detection by data modality, using reported metrics from peer-reviewed studies (Tables 1–3).
2. It proposes a reference architecture for multimodal fusion of ECG, imaging and EHR data (Figure 1).
3. It compares AI models with guideline risk scores (Figure 2) and sets out the main barriers to clinical deployment.

## Related Work

Published AI models for early CVD detection fall into four groups by input data. Table 1 lists landmark studies in each group with the cohort size and headline result each reported.

**Table 1:** Landmark AI studies for cardiovascular detection and risk prediction (values as reported by the authors).

| Study | Modality | Task | Cohort | Model | Key result |
|---|---|---|---|---|---|
| Attia et al., 2019, *Nat Med* [3] | 12-lead ECG | Detect LV dysfunction (EF ≤ 35%) | 44,959 train; 52,870 test | CNN | AUC 0.93 |
| Attia et al., 2019, *Lancet* [5] | 12-lead ECG in sinus rhythm | Detect hidden atrial fibrillation | 180,922 patients; 649,931 ECGs | CNN | AUC 0.87 (0.90 with multiple ECGs) |
| Hannun et al., 2019, *Nat Med* [6] | Single-lead ambulatory ECG | Classify 12 rhythm classes | 91,232 ECGs; 53,549 patients | 34-layer DNN | AUC 0.97; F1 0.837 vs 0.780 for cardiologists |
| Ouyang et al., 2020, *Nature* [7] | Echocardiogram video | Estimate EF; detect HFrEF | 10,030 videos | 3D CNN (EchoNet-Dynamic) | EF MAE 4.1%; HFrEF AUC 0.97 |
| Poplin et al., 2018, *Nat Biomed Eng* [8] | Retinal fundus photo | Predict risk factors and MACE | 284,335 patients | Inception-v3 CNN | MACE AUC 0.70; age MAE 3.26 years |

| Study | Modality | Task | Cohort | Model | Key result |
|---|---|---|---|---|---|
| Weng et al., 2017, *PLoS One* [4] | Primary-care EHR | 10-year CVD risk | 378,256 patients | Neural network | AUC 0.764 vs 0.728 for ACC/AHA |
| Alaa et al., 2019, *PLoS One* [9] | UK Biobank, 473 variables | 5-year CVD risk | 423,604 participants | AutoPrognosis ensemble | AUC 0.774 vs 0.724 for Framingham |
| Soenksen et al., 2022, *npj Digit Med* [10] | Tabular + time-series + text + X-ray | 12 tasks incl. cardiomegaly | 34,537 samples; 6,485 patients | HAIM fusion (14,324 models) | AUROC +6–33% over single source |

**ECG-based models.** The ECG is cheap, quick and recorded in almost every clinic, which makes it the strongest candidate for population screening. Attia et al. trained a convolutional network on paired ECG–echocardiogram data and found that patients flagged by the model but with a normal echo had four times the risk of developing future LV dysfunction[3]. This suggests the network detects disease before it becomes visible on imaging. The same group showed that an ECG recorded in normal sinus rhythm still carries a signature of past or future AF[5]. Hannun et al. showed that a deep network reading single-lead patch ECGs matched or exceeded a panel of cardiologists across 12 rhythm classes[6].

**Imaging-based models.** EchoNet-Dynamic reads full echocardiogram videos rather than single frames and estimates EF beat by beat; it kept an AUC of 0.96 for HFrEF on an external hospital dataset[7]. Poplin et al. showed that retinal photographs, which do not image the heart at all, still predict blood pressure, smoking status and five-year major adverse cardiac events[8].

**EHR-based risk models.** Weng et al. compared four machine-learning methods with the ACC/AHA equations on UK primary-care records; the best neural network correctly identified 355 more patients who later developed CVD[4]. Alaa et al. found that adding more variables ("information gain") helped more than using more complex models ("modeling gain")[9].

**Multimodal models.** Fusion work is newer. The HAIM framework of Soenksen et al. trained models on every combination of four modalities from MIMIC-IV and MIMIC-CXR and found that each added modality raised performance, with diminishing returns[10].

## Methodology: A Multimodal Fusion Framework

We propose a framework that encodes each data modality separately, fuses the learned embeddings, and outputs both a diagnosis probability and a risk score. Figure 1 shows the design; Table 2 lists the inputs.

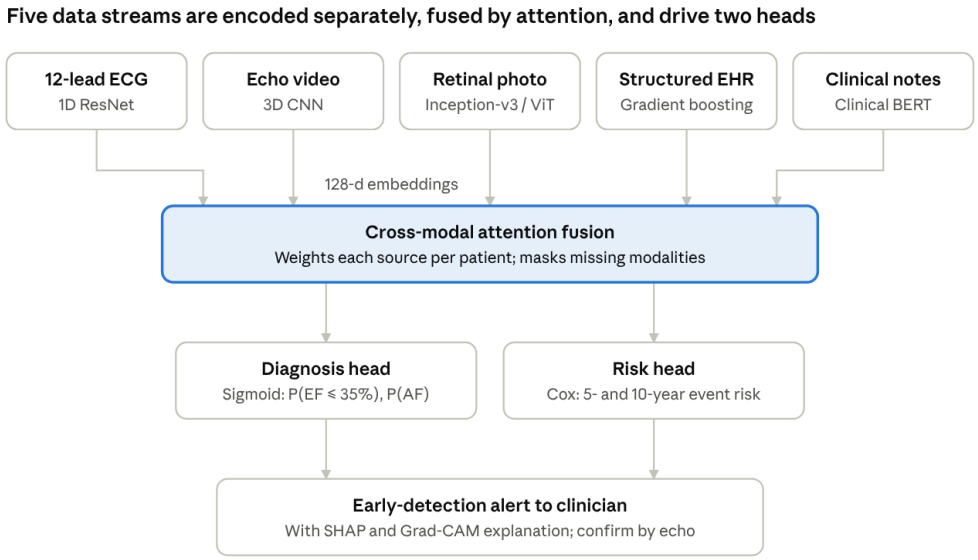



**Fig 1:** Proposed multimodal fusion framework (five inputs, one fusion layer, two output heads). Each modality is encoded to a 128-dimensional embedding; an attention layer fuses them and feeds a diagnosis head and a risk head, whose output reaches the clinician as an explained alert.

## Data sources

**Table 2:** Input modalities, typical sources and encoders used in the framework.

| Modality | What it captures | Example public dataset | Preprocessing | Encoder |
|---|---|---|---|---|
| 12-lead ECG (10 s, 500 Hz) | Electrical activity, conduction, repolarisation | PTB-XL (21,799 ECGs) | Band-pass filter 0.5–40 Hz, baseline removal, z-score per lead | 1D ResNet |
| Echocardiogram video | Chamber size, wall motion, ejection fraction | EchoNet-Dynamic (10,030 videos) | Apical-4-chamber view, 112×112 px, 32-frame clips | 3D CNN (R(2+1)D) |
| Retinal fundus photo | Microvascular health, vessel calibre | UK Biobank retinal subset | Crop, resize to 587×587, colour normalise | Inception-v3 / ViT |
| Structured EHR | Age, sex, BP, lipids, HbA1c, BMI, smoking, medications | MIMIC-IV, UK Biobank | Unit harmonisation, median imputation + missing flags | Gradient boosting or MLP |
| Clinical notes | Symptoms, family history | MIMIC-IV-Note | De-identification, tokenisation | Clinical BERT |

## Preprocessing

Records are linked by patient ID and time-aligned so that every input precedes the prediction date; this avoids label leakage. Continuous EHR variables are standardised, and each variable that can be missing gets a binary "missing" flag, because missingness itself often carries clinical meaning[4]. Signals with poor quality (flat leads, motion artefact) are removed using a signal quality index.

## Model architecture

Each encoder maps its input to a 128-dimensional embedding. We use **intermediate (embedding-level) fusion**: the embeddings are concatenated and passed through a cross-modal attention layer, which learns how much weight to give each source for each patient. If a modality is missing, its embedding is masked and attention is redistributed over the rest, so the model still works for patients who have only an ECG and basic EHR data. Two heads sit on top: a sigmoid classifier for the presence of disease (e.g., EF ≤ 35%, AF) and a Cox-style survival head for 5- and 10-year event risk.

The overall loss is a weighted sum:

$$L = \lambda_1 L_{BCE}(y, \hat{y}) + \lambda_2 L_{Cox}(t, \delta, \hat{r})$$

where y is the diagnostic label, t the time to event, δ the event indicator and $\hat{r}$ the predicted risk.

## Training and evaluation

Data are split by patient (70/10/20 for train, validation and test) so that no individual appears in more than one set. Class imbalance is handled with focal loss and stratified sampling. We report AUROC, sensitivity, specificity, F1 score and, for risk models, the C-index and calibration slope. External validation on a hospital system not used in training is required before any claim of generalisability. SHAP values and Grad-CAM saliency maps are produced to explain which variables and which ECG segments drove each prediction.

## Results and Discussion

The evidence shows that AI detects hidden cardiac disease from routine tests with AUCs of 0.87–0.97, and that it beats guideline risk scores by 0.036–0.050 AUC points when given richer data.

### Diagnostic accuracy of single-modality models

**Table 3:** Reported diagnostic performance of ECG and echocardiography AI models on held-out test sets.

| Model and task | AUC | Sensitivity (%) | Specificity (%) | Accuracy (%) | Other |
|---|---|---|---|---|---|
| AI-ECG, LV dysfunction (EF ≤ 35%) [3] | 0.93 | 86.3 | 85.7 | 85.7 | F1 49.5%; NPV 98.7% |
| AI-ECG, AF from one sinus-rhythm ECG [5] | 0.87 | 79.0 | 79.5 | 79.4 | F1 39.2% |
| AI-ECG, AF from all ECGs in first month [5] | 0.90 | 82.3 | 83.4 | 83.3 | F1 45.4% |
| DNN, 12 rhythm classes, single lead [6] | 0.97 | — | — | — | F1 0.837 (cardiologists 0.780) |
| EchoNet-Dynamic, HFrEF, internal test [7] | 0.97 | — | — | — | EF MAE 4.1%; Dice 0.92 |
| EchoNet-Dynamic, HFrEF, external hospital [7] | 0.96 | — | — | — | EF MAE 6.0% |

Two points stand out. First, the high negative predictive value of the LV-dysfunction model (98.7%) makes it useful as a rule-out test: a negative AI-ECG can safely spare a patient an echocardiogram. Second, the low F1 scores (39–50%) show that positive predictive value is modest when disease is rare. A positive AI result is therefore a reason for a confirmatory test, not a diagnosis. Performance also drops when a model moves to a new hospital—EchoNet's EF error rose from 4.1% to 6.0%—which is why external validation matters.

### AI risk models versus guideline scores

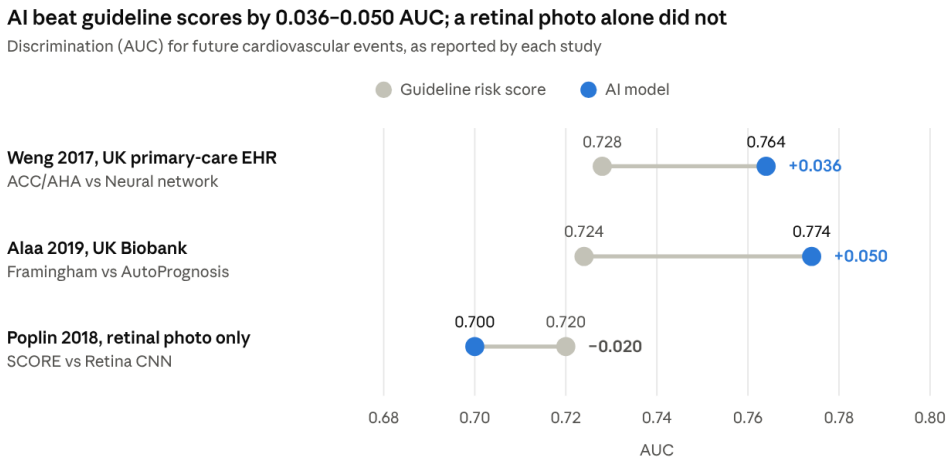



**Fig 2:** AUC of guideline risk scores versus AI models in three published cohorts. The two models given hundreds of EHR or cohort variables beat the guideline score; the model given only a retinal photograph came within 0.02 of SCORE. Sources: Weng et al. 2017 [4], Alaa et al. 2019 [9], Poplin et al. 2018 [8].

Gains are consistent but moderate. In the UK Biobank study, a Cox model using all 473 variables already reached AUC 0.758, close to the AutoPrognosis ensemble at 0.774[9]. This supports the view that richer data matters more than more complex algorithms. The retinal model is a useful counter-example: from one photograph it nearly matched SCORE, which needs a blood test, but did not exceed it[8]. Each single source has a ceiling; combining sources is the way past it.

## Evidence of clinical benefit

Accuracy on a test set does not prove that patients benefit. The EAGLE trial tested this directly by cluster-randomising 120 primary-care teams in 45 clinics and hospitals; 22,641 adults had an ECG during the eight-month trial, and 6% screened positive[11].

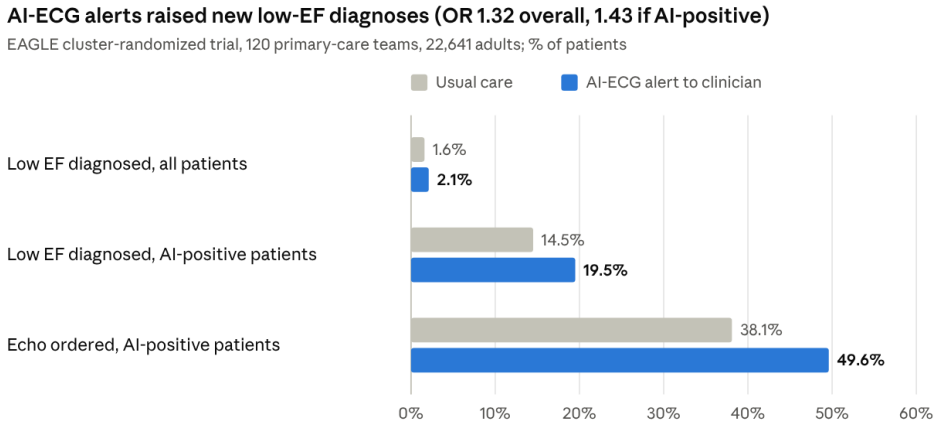



**Fig 3:** Outcomes in the EAGLE trial. Access to the AI-ECG result raised low-EF diagnosis from 1.6% to 2.1% overall (OR 1.32, P = 0.007) and from 14.5% to 19.5% among AI-positive patients (OR 1.43, P = 0.01), mainly by prompting more echocardiograms. Source: Yao et al., Nat Med 2021 [11].

In absolute terms, the AI alert produced about five new diagnoses of low EF per 1,000 patients screened[11]. The trial also exposed a workflow gap: even after a positive AI result, clinicians ordered an echocardiogram only about half the time. Deployment therefore needs workflow design, not just a good model.

## Value of multimodal fusion

In the HAIM study, multimodal models beat single-source models on every task, with AUROC gains of 6–22% for chest pathology (10% for cardiomegaly), 8–20% for length of stay and 11–33% for 48-hour mortality[10]. Each added modality improved results, but with diminishing returns. For CVD, this suggests a practical tiering: start with ECG plus basic EHR data, which are almost always available, and add imaging when it exists.

## Challenges and Future Directions

**Data bias and generalisability.** Most landmark models were trained in single health systems in the US or UK, such as Mayo Clinic, Stanford and UK Biobank. Their performance in South Asian, African or rural populations is largely untested, and ECG morphology, risk-factor profiles and disease prevalence all differ across populations. Multi-centre datasets from low- and middle-income countries, where most CVD deaths occur, are the most urgent need.

**Missing and messy data.** Real patients rarely have every modality. A fusion model must degrade gracefully when imaging or notes are absent, as the masked-attention design described under Model architecture is meant to do.

**Explainability and trust.** Clinicians are reluctant to act on a score they cannot interpret. Saliency maps and SHAP values help, but they show correlation, not causation, and need clinical review.

**Low positive predictive value.** In low-prevalence screening, most positive alerts will be false. Programmes must budget for confirmatory tests and decide thresholds based on local capacity.

**Privacy and regulation.** Linking ECG, imaging and EHR data raises the risk of re-identification. Federated learning, in which models are trained across hospitals without moving raw data, is a promising answer. Regulators also need evidence of safety after deployment, including monitoring for performance drift.

**Future work.** Priorities include prospective multimodal trials with hard outcomes (events and deaths, not just diagnoses), smartphone and wearable ECG inputs for community screening, and foundation models pre-trained on millions of unlabelled ECGs and images that can be fine-tuned for local populations.

## Conclusion

AI can find cardiovascular disease years before symptoms by reading routine tests that are already collected. A single ECG analysed by a neural network detects LV dysfunction with an AUC of 0.93, machine-learning risk models beat guideline scores by up to 0.050 AUC points, and a randomized trial has shown that AI-ECG alerts raise real-world diagnosis rates by 32%. Each modality has a ceiling, and multimodal fusion of ECG, imaging and EHR data is the clearest route past it. The next step is not a better algorithm but better evidence: diverse training data, external validation, and prospective trials that measure whether patients live longer. If those are delivered, multimodal AI could make early cardiac screening affordable at population scale, including in the low-resource settings that carry most of the global burden.

## References


1. World Heart Federation. World Heart Report 2023: Confronting the World's Number One Killer. Geneva; 2023.
2. World Health Organization, South-East Asia Region. Management of cardiovascular disease.
3. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med. 2019;25:70–74.
4. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. 2017;12(4):e0174944.
5. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. Lancet. 2019;394:861–867.
6. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25:65–69.
7. Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020;580:252–256.
8. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. 2018;2:158–164.
9. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. PLoS One. 2019;14(5):e0213653.
10. Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. npj Digit Med. 2022;5:149.
11. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence–enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med. 2021;27:815–819.