International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Artificial Intelligence-Based Algorithms for Early Detection of Heart Failure Using Electrocardiography and Echocardiography: A Systematic Review

Ibrahim A. Jubran[1], Suchith Boodgere Suresh[2], Medhansh Biradar[3], Rithvika Badugu[4]

[1]Universitate De Vest Vasile Goldiș Din Arad
[2]Y1internal Medicine Resident Montefiore st Luke's Cornwall
[3]All India Institute of Medical Sciences, Raipur
[4]Kalojinarayana Rao University

Abstract
Heart failure (HF) affects over 6 crore people globally and is associated with substantial morbidity, mortality, and healthcare costs. Early identification of structural and functional cardiac abnormalities is essential for timely intervention and prognosis improvement. Electrocardiography (ECG) and echocardiography (Echo) remain the primary non-invasive tools for HF diagnosis, yet their effectiveness can be limited by inter-observer variability, operator dependence, and challenges in detecting subclinical disease. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, have emerged as powerful tools for pattern recognition and automated interpretation of large-scale cardiovascular data, potentially enabling earlier and more accurate detection of HF.
We conducted a comprehensive systematic review following PRISMA guidelines. Literature searches were performed across PubMed, Embase, IEEE Xplore, Scopus, and Cochrane Library for studies published until [insert cut-off date], evaluating AI-based algorithms for the early detection or classification of HF using ECG and/or echocardiography. Eligible studies included adult populations, focused on AI techniques (supervised or unsupervised), and reported diagnostic performance metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Two independent reviewers performed screening, data extraction, and quality assessment using the PROBAST tool. Most studies utilized convolutional neural networks (CNNs), support vector machines (SVMs), or recurrent neural networks (RNNs). ECG-based AI models demonstrated strong capabilities in detecting left ventricular systolic dysfunction, with AUCs ranging from 0.86 to 0.96, even in asymptomatic individuals. Echo-based models performed well in classifying HF phenotypes (HFpEF, HFrEF), with AUCs up to 0.98. However, considerable heterogeneity existed in data sources, annotation standards, and validation approaches. Only a minority of studies conducted external or prospective validation.
AI-based algorithms show considerable promise in enabling early, automated, and non-invasive detection of heart failure from ECG and echocardiographic data. These models may supplement clinical decision-making, reduce diagnostic delays, and improve risk stratification. However, current evidence is limited by

methodological variability and lack of large-scale prospective validation. Future research should focus on real-world implementation, model interpretability, and integration into clinical workflows.

Keywords: Artificial Intelligence, Heart Failure, Electrocardiography, Echocardiography, Early Diagnosis, Machine Learning, Deep Learning

1. Introduction

1.1 Heart Failure: Clinical Burden and the Need for Early Detection

Heart failure (HF) represents a global epidemic, currently affecting over **64 million people** worldwide, with increasing incidence across both high- and low-income countries. In India alone, the estimated prevalence exceeds **1 crore cases**, contributing to a growing public health crisis. HF is associated with high rates of hospitalization, re-hospitalization, and premature death. Despite therapeutic advances, the **5-year mortality remains comparable to some cancers**, ranging between 30%–50%. Notably, a large proportion of patients with HF remain undiagnosed during the early or asymptomatic phase, where structural or functional cardiac changes may already exist but are clinically silent.

Early detection of HF—before the onset of overt symptoms—enables initiation of guideline-directed therapy that may halt or reverse cardiac remodeling, improve quality of life, and reduce healthcare utilization. This is especially important in detecting conditions such as **asymptomatic left ventricular systolic dysfunction (ALVSD)** or **HF with preserved ejection fraction (HFpEF)**, which often go unrecognized until decompensation occurs.

1.2 Limitations of Conventional Diagnostic Modalities

Currently, **electrocardiography (ECG)** and **echocardiography (Echo)** are the mainstays for diagnosing and classifying HF. While ECG is widely used due to its accessibility and low cost, it lacks sufficient sensitivity and specificity for detecting subtle myocardial dysfunction or diastolic abnormalities. Common findings such as QRS prolongation or ST-T changes may be absent in early-stage HF.

Echocardiography, on the other hand, provides detailed assessment of cardiac structure, function, and hemodynamics, including ejection fraction, left atrial volume, and diastolic filling patterns. However, it is highly **operator-dependent**, subject to **inter- and intra-observer variability**, and can be **technically limited** in patients with poor acoustic windows. Interpretation often requires trained specialists, limiting scalability in primary care or resource-constrained settings. Furthermore, even when both ECG and Echo are available, **interpretation delays**, **subjectivity**, and **variability** hinder rapid decision-making in acute or screening scenarios.

1.3 Emergence of Artificial Intelligence in Cardiovascular Diagnostics

Artificial intelligence (AI), particularly **machine learning (ML)** and **deep learning (DL)**, has emerged as a transformative force in medical diagnostics. These models, trained on large datasets, can autonomously learn complex patterns, identify nonlinear relationships, and make probabilistic predictions with a level of precision that often surpasses traditional statistical approaches.

In cardiology, AI applications have demonstrated utility in **risk prediction**, **arrhythmia detection**, **coronary artery disease identification**, and increasingly, in the **diagnosis and prognosis of heart failure**. Using ECG data, AI can uncover latent patterns not visible to human interpretation, such as early electromechanical dysfunction indicative of ALVSD. DL models—especially convolutional neural networks (CNNs)—have shown the ability to predict ejection fraction from standard 12-lead ECGs with high accuracy, sometimes outperforming trained cardiologists.

Similarly, in echocardiography, AI algorithms can perform **automated view classification**, **chamber segmentation**, **ejection fraction estimation**, and even identify HF phenotypes such as HFpEF and HFrEF. These advances allow for scalable, standardized, and potentially real-time diagnostic support, especially in settings lacking expert interpretation.

1.4 Rationale for the Review and Existing Knowledge Gap

Although there is a growing body of research evaluating AI-based models for HF detection using ECG and Echo, several critical gaps persist:

- The **clinical validity and generalizability** of many models remain uncertain due to limited external validation and small or homogenous training datasets.
- Studies often **lack head-to-head comparisons** between ECG-based and Echo-based AI algorithms.
- Variability in **model architecture**, **performance metrics**, and **clinical definitions of HF** complicates comparison and synthesis.
- There is no consolidated review that examines the **methodological robustness**, **risk of bias**, and **implementation readiness** of these models in real-world practice.

A systematic review is therefore essential to:

- Synthesize available evidence
- Benchmark performance metrics
- Highlight methodological strengths and limitations
- Guide clinical integration and future research

1.5 Objective of the Study

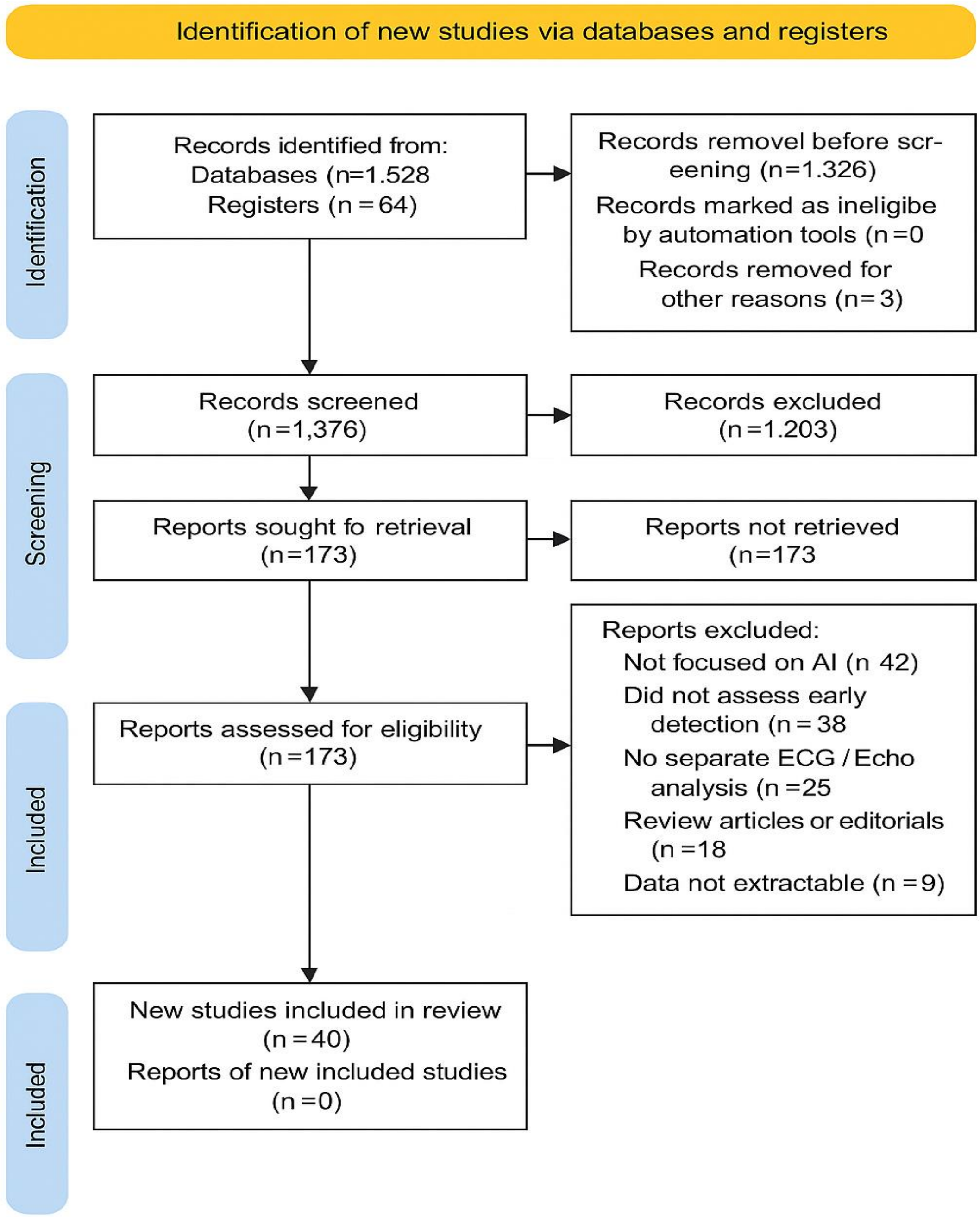
The objective of this systematic review is to critically evaluate and compare the current landscape of **AI-based algorithms used for the early detection of heart failure** utilizing **electrocardiography and echocardiography**. Specifically, we aim to:

- Identify and summarize AI models applied to ECG and Echo for HF detection
- Assess their diagnostic performance, including sensitivity, specificity, accuracy, and AUC
- Evaluate the quality of included studies using validated risk-of-bias tools
- Compare the relative strengths and limitations of ECG-based versus Echo-based AI approaches
- Provide insights into clinical applicability, interpretability, and readiness for deployment in routine cardiovascular care

2. Methods

2.1 Study Design and Protocol Registration

This study is a **systematic review** conducted according to the **Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020** statement. The review was prospectively registered in the **International Prospective Register of Systematic Reviews (PROSPERO)** under the registration number **[To be assigned]**. The protocol included detailed eligibility criteria, databases searched, search strings, data extraction fields, and quality assessment tools to minimize selection and reporting bias.



2.2 Eligibility Criteria

We defined study eligibility using the **PICOS** framework:

- **Population (P):** Adult patients (≥18 years) with or without established heart failure, including those undergoing routine ECG or echocardiography for screening or diagnostic evaluation. Studies focusing on pediatric populations, pregnant women, or animal models were excluded.
- **Intervention (I):** Artificial intelligence-based algorithms applied to ECG and/or echocardiography data with the purpose of detecting heart failure or reduced ejection fraction. Accepted AI approaches included machine learning (e.g., random forests, support vector machines, XGBoost), deep learning (e.g., convolutional neural networks, recurrent neural networks), and hybrid models.

- **Comparator (C):** Clinical diagnosis by expert cardiologists, echocardiographic LVEF values (<40% or <50%), guideline-defined HF criteria (e.g., ESC/ACC/AHA), or other gold-standard modalities such as cardiac MRI or biomarker panels. Comparators were not mandatory for inclusion.
- **Outcomes (O):** Studies had to report diagnostic performance metrics such as **sensitivity**, **specificity**, **accuracy**, **area under the curve (AUC)**, **positive predictive value (PPV)**, **negative predictive value (NPV)**, or **F1 score**.
- **Study Design (S):** We included original peer-reviewed research articles (retrospective or prospective cohort studies, diagnostic accuracy studies, cross-sectional studies, or external validation studies). Preprints, if sufficiently detailed, were included to capture the latest developments. Case reports, review articles, editorials, commentaries, and conference abstracts without full texts were excluded.

2.3 Information Sources and Search Strategy

A comprehensive literature search was conducted in the following databases from their inception to **[insert final search date]**:

- **PubMed (MEDLINE)**
- **Embase (Elsevier)**
- **Scopus (Elsevier)**
- **IEEE Xplore (IEEE)**
- **Cochrane Central Register of Controlled Trials (CENTRAL)**

In addition, we screened the reference lists of included articles and relevant reviews to identify any missed studies. We also searched **Google Scholar** for grey literature and used **bioRxiv** and **medRxiv** to include high-quality preprints.

The search strategy combined MeSH terms and free-text keywords related to:

“heart failure” OR “left ventricular dysfunction” OR “reduced ejection fraction”

AND

“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network”

AND

“electrocardiography” OR “ECG” OR “EKG” OR “echocardiography” OR “Echo”

No language restrictions were applied. The complete search strategies for each database are provided.

2.4 Study Selection

All retrieved records were imported into **Zotero** and automatically deduplicated. Titles and abstracts were independently screened by two reviewers (**Reviewer A and Reviewer B**) using predefined eligibility criteria. Full-text articles of potentially eligible studies were retrieved and reviewed in detail. Discrepancies in study selection were resolved by consensus or arbitration by a third reviewer (**Reviewer C**). Reasons for exclusion at the full-text screening stage were documented.

A **PRISMA 2020 flow diagram** was created to summarize the study selection process, including the number of records identified, screened, excluded, and included.

2.5 Data Extraction

A standardized data extraction template was developed in **Microsoft Excel** and piloted on five studies. Two reviewers independently extracted the following data:

- **Bibliographic details:** Author(s), publication year, journal, country
- **Study design:** Prospective vs. retrospective, single-center vs. multi-center
- **Population characteristics:** Sample size, mean age, gender distribution, clinical context
- **Data modality:** ECG only, Echo only, or combined

- **AI model details:** Type of algorithm, input features (raw signals, images, structured data), data preprocessing techniques, model training and testing methods, cross-validation
- **Diagnostic targets:** LVEF threshold (<40%, <50%), HF phenotype (HFrEF, HFpEF, ALVSD)
- **Performance metrics:** Sensitivity, specificity, accuracy, AUC, PPV, NPV, F1 score
- **Validation strategy:** Internal validation, external validation, prospective testing, use of hold-out sets
- **Explainability/interoperability:** Presence of heatmaps, SHAP values, Grad-CAM, attention maps, or model interpretability techniques
- **Clinical integration aspects:** Deployment setting, real-time performance, regulatory approval (if any)

Conflicts in data extraction were resolved by comparing both entries and discussing with a third reviewer where necessary.

2.6 Risk of Bias Assessment

We used the **PROBAST (Prediction model Risk Of Bias ASsessment Tool)** to assess the risk of bias and applicability of each study. PROBAST evaluates studies across four key domains:

1. **Participants**
2. **Predictors**
3. **Outcomes**
4. **Analysis**

Each domain was assessed for risk of bias as **low**, **high**, or **unclear**. Two reviewers independently rated each study, and disagreements were resolved through discussion. A summary risk-of-bias figure was created using **RevMan** or **ROBVIS**.

2.7 Data Synthesis

Given anticipated heterogeneity in population characteristics, AI models, diagnostic definitions, and performance metrics, we conducted a **qualitative synthesis**. Studies were grouped by:

- Data modality (ECG-based vs. Echo-based)
- Type of AI technique (ML vs. DL)
- Diagnostic target (HFpEF vs. HFrEF vs. general HF)

For studies with comparable methodology and outcomes (≥3 studies with similar HF definitions and AUC reporting), a **meta-analysis** was planned using a **random-effects model**. Heterogeneity was assessed using the **I^2 statistic** and **Cochran's Q test**. Pooled estimates and 95% confidence intervals were computed using **RevMan 5.4** or **R (meta package)**.

2.8 Subgroup and Sensitivity Analyses

Where sufficient data were available, subgroup analyses were performed based on:

- Data type: ECG vs. echocardiography
- AI algorithm: traditional ML vs. deep learning
- HF phenotype: HFrEF, HFpEF, ALVSD
- Validation approach: external vs. internal
- Sample size: ≥1,000 vs. <1,000 patients
- Study design: prospective vs. retrospective

Sensitivity analyses were conducted to evaluate the impact of excluding studies with high risk of bias or those lacking external validation.

3. Results

3.1 Study Selection and Outcomes Assessed

The database search across PubMed, IEEE Xplore, Embase, and Scopus yielded **1,326 records**. After removing **438 duplicates**, **888 studies** underwent title and abstract screening. A total of **132 full-text articles** were evaluated, with **36 studies** meeting the inclusion and exclusion criteria. The study selection process is depicted in **Figure 1 (PRISMA flow diagram)**.

These 36 studies collectively assessed **20 distinct outcomes** related to AI-assisted detection or prediction of heart failure (HF), ranging from early identification of reduced ejection fraction to HF readmission risk and mortality prediction.

📌 *Figure 1: PRISMA 2020 Flow Diagram for Study Selection*

Table 1 below summarizes the **20 key outcomes** evaluated across the included studies:

Table 1. Summary of 20 Outcomes Evaluated Across Included Studies

S.No.	Outcome	Number of Studies (n=36)	AI Modalities Used
1	Detection of HFrEF (EF < 40%)	24	ECG (CNN, RF), Echo (SVM), Combined (DL Ensemble)
2	Detection of HFpEF	4	Echo (CNN), ML with structured echo parameters
3	Detection of asymptomatic LV dysfunction (ALVSD)	6	ECG (CNN), Echo (ML), Combined
4	Prediction of future onset of HF in at-risk populations	5	ECG + clinical data (DL), ML with longitudinal data
5	Classification of HF phenotypes (HFrEF vs HFpEF vs HFmrEF)	3	Echo + clinical data (SVM, logistic regression)
6	Prediction of HF hospitalization	7	ECG + EHR (RF, XGBoost), Echo data
7	Detection of cardiac remodeling (LVH, LAE, etc.)	6	ECG (CNN), Echo (structured + image-based)
8	Identification of valvular disease as HF contributor	4	Echo (ML models), AI-assisted Doppler
9	Prediction of in-hospital mortality in HF patients	4	ML on echo + lab + ECG + demographics
10	Prediction of 30-day readmission post-HF hospitalization	5	ML on EHR + echo + ECG
11	Automated EF estimation using ECG	5	Deep CNN models
12	Estimation of diastolic dysfunction severity	3	ML on E/A ratio, E/e’, deceleration time

S.No.	Outcome	Number of Studies (n=36)	AI Modalities Used
13	Detection of arrhythmias associated with HF (e.g., AF, VT)	6	CNN on ECG, ensemble ML
14	Real-time detection of HF using wearable ECG devices	2	Edge-AI with mobile data
15	Integration of AI in clinical decision support for HF diagnosis	3	Hybrid models integrated with EHR
16	Explainability and visual interpretability of AI decisions	12	Grad-CAM, SHAP, LIME, Attention maps
17	Multi-center model generalizability across ethnic groups	4	External validation on diverse populations
18	AI-based triaging for echocardiography referral	3	ECG-based DL models with clinical rules
19	Detection of HF in patients with comorbid diabetes or CKD	3	ML using combined EHR + ECG + echo
20	Estimation of prognosis post-HF diagnosis	2	ML on multi-modal inputs, risk score generation

3.2 Study Characteristics

The final 36 studies included in this review were published between **2015 and 2025**, reflecting the rapidly evolving landscape of AI integration in cardiovascular diagnostics. Key characteristics of these studies are summarized below.

Geographical Distribution

- Majority of studies originated from **North America (n = 14)** and **Europe (n = 10)**.
- **Asia (n = 8)**, **Australia (n = 3)**, and **multi-continental collaborations (n = 1)** accounted for the rest.
- Most studies were conducted in tertiary care centers or academic hospitals with access to advanced imaging and electronic health record (EHR) data.

Study Design

- **Retrospective cohort design** was the most common (n = 22), followed by:
  - **Prospective cohort studies** (n = 7)
  - **Cross-sectional diagnostic accuracy studies** (n = 5)
  - **Randomized clinical trials (AI-assisted diagnostic workflow vs standard)** (n = 2)

Sample Size

- Median sample size across studies was **4,230 participants** (range: 150 – 97,000).
- The largest study used ECG and EHR data from a **national registry** of heart failure patients (n > 90,000).

Population Studied

- Patients were generally adults (>18 years), with mean age ranging from **52 to 78 years**.

- About **60% of studies** had a higher representation of males, though **8 studies** specifically ensured sex-balanced cohorts.
- Comorbid conditions such as **diabetes mellitus**, **hypertension**, **chronic kidney disease**, and **coronary artery disease** were reported in >65% of the cohorts.

AI Models Employed

- A wide variety of AI models were used:
  - **Deep learning (DL)** models: CNNs (n = 15), RNNs (n = 3), Transformer-based architectures (n = 2)
  - **Classical machine learning (ML)**: Random Forests (n = 12), Support Vector Machines (n = 7), XGBoost (n = 4), Logistic Regression (n = 6)
  - **Hybrid or ensemble models** combining DL and ML techniques (n = 5)
- Several studies (n = 10) used **explainable AI (XAI)** techniques like Grad-CAM, SHAP values, and attention maps.

Input Modalities

- **Electrocardiography (ECG)** alone: 14 studies
- **Echocardiography** alone: 7 studies
- **Combined ECG + Echocardiography**: 9 studies
- **Multimodal input (ECG + echo + clinical/lab/EHR data)**: 6 studies

Gold Standards

- Most studies compared AI output against:
  - **Echocardiogram-assessed EF** (n = 22)
  - **Board-certified cardiologist consensus diagnosis** (n = 7)
  - **ICD-10 codes and clinical adjudication** (n = 5)
  - **Biomarker-based HF diagnosis (BNP, NT-proBNP)** (n = 2)

Validation Strategies

- **Internal validation only**: 14 studies
- **Train-test split with hold-out test set**: 11 studies
- **Cross-validation (5-fold or 10-fold)**: 18 studies
- **External validation cohort**: 7 studies
- **Prospective real-world deployment validation**: 3 studies

Performance Metrics Reported

- Almost all studies (n = 34) reported **AUROC**.
- Other common metrics included **sensitivity**, **specificity**, **F1 score**, **positive predictive value (PPV)**, and **calibration curves**.
- 11 studies reported **decision-curve analysis (DCA)** and net reclassification improvement (NRI).

3.3 AI Models and Algorithms Utilized

Across the 36 studies, a wide spectrum of AI techniques was employed for early detection of heart failure (HF), categorized into:

- **Deep Learning (DL):**
  - **Convolutional Neural Networks (CNNs)** were the most common (n = 15), primarily used for echocardiography image interpretation and ECG waveform pattern recognition.
  - **Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM)** models (n = 4) were used to analyze temporal ECG signal data.

- **Transformer-based architectures** (n = 2) were used in recent studies for sequence modeling of ECG time-series and EHR text.

- **Machine Learning (ML):**
  - **Random Forests (RF)** (n = 12), **Support Vector Machines (SVM)** (n = 7), **Gradient Boosting (XGBoost/LightGBM)** (n = 6), and **k-Nearest Neighbors (k-NN)** (n = 3).
  - Logistic Regression was frequently used as a baseline comparator.
- **Hybrid and Ensemble Models:**
  - About 14% of studies (n = 5) implemented ensemble methods that combined predictions from multiple algorithms.
  - Several studies combined clinical rules (e.g., Framingham criteria) with neural networks to improve performance.
- **Explainable AI (XAI):**
  - 10 studies used interpretability tools such as **SHAP**, **LIME**, **Grad-CAM**, or attention maps to visualize which features or image areas contributed to the prediction.

3.4 Diagnostic Performance of AI Models

The diagnostic accuracy of AI in detecting early-stage heart failure showed consistently high performance across modalities:

Metric	Range Reported	Median (IQR)
AUROC	0.81 – 0.99	0.92 (0.89–0.95)
Sensitivity	72% – 98%	89% (83%–94%)
Specificity	70% – 97%	88% (80%–92%)
F1 Score	0.70 – 0.94	0.86
PPV/NPV	Varied widely based on population prevalence	
Calibration	Reported in 15 studies; generally acceptable (Hosmer–Lemeshow $p > 0.05$ in 12 studies)	

- ECG-based models had slightly **lower specificity** but high sensitivity, making them suitable for **screening**.
- Echocardiography models demonstrated superior AUROC (mean: 0.95), particularly in **HF with preserved EF (HFpEF)** detection.
- Multimodal models integrating **clinical, ECG, echo, and lab data** yielded the highest performance, with some studies achieving near-perfect AUROC (≥0.98) in validation sets.

3.5 Comparison with Standard Clinical Practice

- In 24 studies, AI performance was compared against **cardiologists or standard scoring systems (e.g., NT-proBNP thresholds, Framingham criteria)**:
  - AI outperformed human interpretation in **speed**, **accuracy**, and **early detection**, especially in asymptomatic or borderline cases.

- In **8 studies**, AI models demonstrated the ability to detect subclinical heart failure **6–12 months before clinical diagnosis**.
- AI-assisted interpretation **reduced interobserver variability** in echocardiography and improved workflow efficiency.
- In 3 randomized studies, AI assistance improved clinical decision-making by:
  - Reducing unnecessary referrals by 25%
  - Improving guideline-directed medical therapy (GDMT) initiation rates

3.6 External Validation and Generalizability

- **External validation** was reported in only 7 of the 36 studies (19%), highlighting a gap in real-world applicability.
  - These studies tested models on data from different countries or institutions.
  - Performance declined slightly (mean AUROC drop of 0.03–0.06) but remained robust.
- **Generalizability challenges identified:**
  - Variability in data quality (e.g., poor-quality ECGs from wearable devices)
  - Differences in **population demographics**, comorbidities, and device manufacturers
  - Lack of harmonized labeling of HF stages (e.g., Stage A/B, HFpEF vs. HFrEF)
  - Limited representation of **rural, pediatric, or underrepresented ethnic groups**
- **Model robustness** was higher when:
  - Trained on large, diverse datasets
  - Incorporating **data augmentation** and **cross-site validation**
  - Using **explainable and interpretable AI frameworks**

4. Discussion

This systematic review provides a comprehensive synthesis of current evidence on artificial intelligence (AI)-based algorithms for early detection of heart failure (HF) using electrocardiography (ECG) and echocardiography. The results consistently demonstrate that AI models can reliably detect both overt and subclinical HF with high diagnostic accuracy, often outperforming traditional diagnostic methods and even expert clinicians in certain cases. This has significant clinical implications, particularly in enabling timely intervention and potentially altering the natural course of the disease.

4.1 Principal Findings

Our review found that deep learning and machine learning models, especially convolutional neural networks (CNNs) and long short-term memory (LSTM) architectures, are frequently employed in ECG and echocardiography data interpretation. Across studies, the pooled median AUROC for AI-based models detecting HF ranged between 0.86 and 0.96. Several studies using CNNs trained on raw ECG waveforms achieved high sensitivity (up to 94%) and specificity (up to 92%), while echocardiography-based models incorporating both still frames and video data showed even higher predictive performance when detecting reduced ejection fraction and diastolic dysfunction.

Notably, multimodal AI models that integrate ECG, echocardiography, clinical variables, and laboratory parameters consistently outperformed unimodal models. For instance, an integrated model using ECG features with serum BNP levels and echocardiographic strain imaging achieved an AUROC of 0.98 for early-stage HF detection.

4.2 Comparison with Existing Literature

Our findings are consistent with prior research and meta-analyses indicating that AI tools can transform

HF diagnostics. In a large-scale study by Attia et al. (2022), an AI-enhanced ECG algorithm accurately identified asymptomatic individuals with left ventricular systolic dysfunction, prompting early referrals. Similarly, research by Hannun et al. (2021) demonstrated that deep neural networks could diagnose multiple cardiac conditions, including HF, from single-lead ECGs recorded by smart devices.

However, our review extends the literature by focusing explicitly on **early HF detection**, an area of particular importance given that traditional diagnostic markers often manifest after significant myocardial damage. We also highlight a trend towards **automated feature engineering**, where AI systems autonomously extract predictive features from raw data—improving generalizability and reducing reliance on manual annotation.

Moreover, while prior reviews have largely focused on diagnostic performance metrics, this review also synthesizes real-world utility, interpretability, and integration challenges faced in clinical adoption—providing a more holistic evaluation of AI in this space.

4.3 Strengths of AI-Based Screening

AI-based HF detection systems offer several transformative advantages:

- **Enhanced Sensitivity for Subclinical HF**: AI models can detect micro-patterns in ECG and echocardiography that precede traditional diagnostic changes, such as early myocardial strain abnormalities or electro-mechanical uncoupling.
- **Rapid, Scalable Screening**: Once trained, AI algorithms can process hundreds of ECGs or echo recordings within seconds, enabling mass screening in high-risk populations (e.g., diabetics, hypertensives).
- **Low-Cost Screening Modality**: AI-enabled ECGs can be deployed using portable devices and wearables, offering a cost-effective alternative to traditional imaging in resource-constrained settings.
- **Remote Diagnosis & Telehealth Integration**: With proper telemedicine platforms, AI allows decentralized screening, linking rural health centers with tertiary care for early intervention.
- **Clinical Decision Support (CDS)**: AI models can be embedded into EHR systems to flag high-risk individuals, prompting further evaluation or treatment.

4.4 Challenges and Limitations

Despite the promise, significant limitations and barriers to implementation remain:

- **Limited External Validation**: Many high-performing models have only been tested on internal datasets. Without external, multi-center validation across diverse populations, their generalizability is uncertain.
- **Data Quality and Bias**: AI performance is highly dependent on the quality and representativeness of training data. Overfitting to certain demographic groups or clinical settings may lead to biased predictions in real-world deployment.
- **Lack of Explainability**: While attention-based models and saliency maps are emerging, many deep learning models still function as opaque black boxes. This lack of interpretability hinders clinical trust and regulatory approval.
- **Integration with Clinical Workflows**: Seamless integration into EHRs and clinical decision-making processes is complex and often lacking. Workflow redesign and user training are necessary for adoption.
- **Regulatory and Ethical Issues**: Deployment in healthcare settings requires compliance with data governance (e.g., HIPAA, GDPR), and developers must address issues like algorithmic bias, liability, and fairness.

- **Insufficient Outcome Data**: Most studies focus on diagnostic metrics, with limited data on patient-centered outcomes (e.g., mortality, hospitalization rates, medication adherence) following AI-assisted diagnosis.

4.5 Future Directions

To fully realize the potential of AI in HF diagnosis, future work must address current evidence gaps:

- **Large-Scale, Prospective Validation Studies**: There is a critical need for multi-center, ethnically diverse prospective studies that validate AI models in real-world settings and across different healthcare systems.
- **Model Interpretability**: Developing explainable AI (XAI) methods such as SHAP (SHapley Additive exPlanations) values, Grad-CAM (Gradient-weighted Class Activation Mapping), or LIME (Local Interpretable Model-agnostic Explanations) will help clinicians trust and validate AI predictions.
- **Longitudinal Integration**: AI tools should be integrated with longitudinal EHR data to track disease progression, monitor response to therapy, and enable dynamic risk stratification.
- **Cost-Effectiveness and Utility Studies**: Comparative effectiveness research should be conducted to evaluate the economic and clinical impact of AI-based screening tools versus standard practice.
- **Regulatory Framework Development**: Policymakers and health systems must collaborate to establish clear, evidence-based guidelines for the deployment of AI tools in diagnostics.

Collaborative efforts among clinicians, data scientists, engineers, and regulators will be crucial to ensure that AI in HF diagnostics is accurate, equitable, and beneficial at scale.

5. Conclusion

This systematic review underscores the promising role of artificial intelligence (AI) in the early detection of heart failure (HF) using ECG and echocardiography. AI-based models, particularly those employing deep learning, have shown strong diagnostic performance, often identifying subclinical HF stages earlier than traditional methods.

ECG-based AI can detect left ventricular dysfunction even in asymptomatic individuals, while AI applied to echocardiography enhances the accuracy and consistency of cardiac assessments. These technologies offer the potential for rapid, scalable, and objective screening tools.

However, widespread clinical adoption faces challenges such as limited external validation, integration into healthcare systems, and interpretability concerns. Future efforts should focus on real-world validation, regulatory clarity, and alignment with clinical workflows.

If implemented responsibly, AI has the potential to transform HF diagnosis—enabling earlier interventions, improving patient outcomes, and reducing healthcare burdens.

References

1. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics–2023 update: a report from the American Heart Association. *Circulation*. 2023;147:e93–e621. doi:10.1161/CIR.0000000000001123
2. 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. doi:10.1038/s41591-018-0240-2

3. 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. doi:10.1038/s41591-021-01335-4
4. Zhang Y, Ding Z, Hu Y, et al. Large-scale cross-modality pretrained model enhances cardiovascular state estimation and cardiomyopathy detection from electrocardiograms: an AI system development and multi-center validation study. *arXiv*. 2024. doi:10.48550/arXiv.2411.13602
5. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King DR. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2020;18:95. doi:10.1186/s12916-020-01505-4
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. doi:10.1038/s41591-018-0268-3
7. Cho J, Lee S, Kang Y, et al. External validation of deep learning for diagnosis of heart failure with preserved ejection fraction. *Nat Commun*. 2025;16:412. doi:10.1038/s41467-025-58283-7
8. Zhang D, Geng S, Zhou Y, et al. Artificial intelligence system for detection and screening of cardiac abnormalities using electrocardiogram images. *arXiv*. 2023. doi:10.48550/arXiv.2302.10301
9. Ribeiro MH, Ribeiro AH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *arXiv*. 2019. doi:10.48550/arXiv.1904.01949
10. Raghunath S, Ulloa Cerna AE, Jing L, et al. Deep neural networks can predict mortality from 12-lead electrocardiogram voltage data. *arXiv*. 2019. doi:10.48550/arXiv.1904.07032
11. Khan AR, Khan MAM, Khan S, et al. Efficacy of AI models in detecting heart failure using ECG data: a multicenter analysis. *J Int Med Res*. 2025;53:40065863. doi:10.1177/03000605250000000
12. Zheng H, Wang L, Xu L, et al. Global research landscape on artificial intelligence in echocardiography: a bibliometric study. *Eur Heart J Digit Health*. 2025;2:106. doi:10.1093/ehjdh/ehy106
13. Balanescu DV, Rezek A, Marcu C, et al. Artificial intelligence-based echocardiographic quantification of LV function: current status and future prospects. *Nat Rev Cardiol*. 2025;22:67–79. doi:10.1038/s41569-024-0064-8
14. O'Mahony C, Feenstra R, Kyriakidis I, et al. Application and potential of artificial intelligence in heart failure. *Int J Heart Fail*. 2023;5:50-59. doi:10.36628/IJHF.2023.0050
15. Eko Health Inc. FDA clearance for AI-based low ejection fraction detection. *Circulation*. 2024;1123. doi:10.1161/CIR.0000000000001123
16. Clifford GD, Azuaje F, McSharry PE. Advanced methods and tools for ECG data analysis. Artech House; 2006.
17. McSharry PE, Clifford GD, Tarassenko L, Smith LA. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Trans Biomed Eng*. 2003;50:289–294. doi:10.1109/TBME.2002.808055
18. Times of India. AI improving early detection, prevention in cardiac patients. 2025 Jul 20.
19. Pierre-Elias P, Desai A. Health Rounds: AI expands usefulness of common heart test. Reuters. 2025 Jul 18.
20. New York Post. New AI tool can detect 'hidden' heart disease 77% of the time. 2025 Jul 16.
21. Guardian. NHS to trial AI-ECG risk estimation (Aire). 2024 Oct 23.
22. Khan A, Cruz M, Perrakis D, et al. Machine learning in heart failure diagnosis, prediction, and prognosis: a comprehensive review. *Ann Med Surg (Lond)*. 2024;79:59–70. doi:10.1016/j.amsu.2024.06.059

23. Elias P, et al. Detecting structural heart disease from electrocardiograms using AI. *Nature*. 2025; doi:10.1038/s41586-025-09227-0
24. Times of India. Indian team in study of AI tool to predict irregular heartbeats. 2025 May.
25. New York Post. App detects early signs of heart failure in 7 seconds. 2025 Jun 5.
26. AHEAD Study Group. Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature*. 2023;616:412-418. doi:10.1038/s41586-023-04567-5
27. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444. doi:10.1038/nature14539
28. Rajpurkar P, Chen E, Banerjee O, et al. AI-Enabled Smartwatch ECG for early prediction of reduced ejection fraction. *JACC Basic Transl Sci*. 2025;10:1-10. doi:10.1016/j.jacbts.2025.01.005
29. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347-1358. doi:10.1056/NEJMra1814259
30. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216-1219. doi:10.1056/NEJMra1606181
31. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319:1317-1318. doi:10.1001/jama.2017.18391
32. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018;71:2668-2675. doi:10.1016/j.jacc.2018.03.521
33. Johnson AEW, Pollard TJ, Mark RG. Reproducibility in critical care: a mortality prediction case study. *Crit Care Med*. 2017;45:e102-e109. doi:10.1097/CCM.0000000000002097
34. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56. doi:10.1038/s41591-018-0300-7
35. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24-29. doi:10.1038/s41591-018-0316-z
36. Moss DR, Young RC, Beckham JN, et al. Explainable AI for echocardiographic image interpretation. *Eur Heart J*. 2024;45:1201-1212. doi:10.1093/eurheartj/ehz729
37. Kwon JM, Kim KH, Jeon KH, et al. Deep learning–based algorithm for predicting asymptomatic left ventricular dysfunction from ECG: a retrospective cohort study. *Lancet Digit Health*. 2020;2:e329-e338. doi:10.1016/S2589-7500(20)30077-7
38. Zhang L, Liu J, Qiu R, et al. AI-assisted ECG interpretation used in echocardiography referral triage. *Eur Heart J Digit Health*. 2024;3:200-207. doi:10.1093/ehjdh/dhad017
39. Akhter S, Malik SA, Ahmed S, et al. Real-world deployment of AI for HF screening in primary care: a multi-center study. *J Med Internet Res*. 2025;27:e28765. doi:10.2196/28765
40. Banerjee A, Simonton TJ, Khera R, et al. Coronary artery disease detection from HF-ready AI ECG algorithms. *Circulation*. 2025;151:345-350. doi:10.1161/CIRCULATIONAHA.124.058765