Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence Driven Software Systems for Cardiovascular Disease Detection Using Physiological Signals: A Systematic Review

MD Tasnim Alam

mdtasnim.alam@uq.edu.au

University of Queensland

Abdullah Al-Mamun Bulbul

University of Queensland

AKM Azad

Imam Mohammad ibn Saud Islamic University

Asaduzzaman Khan

University of Queensland

Mohammad Ali Moni

University of Queensland

Research Article

Keywords: Cardiovascular disease, machine learning, deep learning, ECG, PPG, PCG, mobile health

Posted Date: February 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8737888/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Artificial Intelligence Review on November 14th, 2026. See the published version at https://doi.org/10.1007/s10462-026-11756-7.

Artificial Intelligence Driven Software Systems for Cardiovascular Disease Detection Using Physiological Signals: A Systematic Review

MD Tasnim Alam[1*], Abdullah Al-Mamun Bulbul[1], AKM Azad[2], Asaduzzaman Khan[1], Mohammad Ali Moni[1]

[1*]School of Health & Rehabilitation Sciences, The University of Queensland, Brisbane, QLD 4072, Australia.
[2]Department of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 13318, Saudi Arabia.

*Corresponding author(s). E-mail(s): mdtasnim.alam@uq.edu.au;
Contributing authors: a.bulbul@uq.edu.au; akmazad@imamu.edu.sa;
a.khan2@uq.edu.au; m.moni@uq.edu.au;

Abstract

Cardiovascular disease (CVD) exacts an extraordinary global toll, with annual mortality exceeding 17.9 million lives and representing nearly one-third of all deaths worldwide. Reducing diagnostic delays, particularly in resource-limited settings, requires innovative, scalable, and cost-effective solutions. Artificial intelligence (AI) integrated with mobile and wearable technologies offers a promising pathway, leveraging physiological signals such as electrocardiogram (ECG), photoplethysmogram (PPG), and phonocardiogram (PCG) for real-time cardiovascular assessment. Despite rapid advances, the current evidence base remains fragmented, impeding identification of validated approaches and actionable pathways toward clinical translation. A systematic review is therefore essential to consolidate existing evidence and establish the current technological landscape. A PRISMA-guided search across five electronic databases-Scopus, Web of Science, PubMed, Embase, and IEEE Xplore-yielded 30 eligible studies. Most systems utilised ECG (66.6%) and deep learning models such as long short-term memory (LSTM) and convolutional neural networks (CNNs). Although several studies (n = 10) reported accuracies approaching or exceeding 98%, only a few (n = 7) demonstrated external validation, and fewer still (n = 3) reported large-scale clinical deployment. While AI-enabled CVD systems are becoming

technically sophisticated, limitations in data diversity, interpretability, regulatory compliance, and clinical integration hinder real-world translation. This review establishes a structured taxonomy of AI-enabled software and mobile-based CVD diagnostic systems, revealing substantial technical progress alongside persistent barriers. A three-directional framework is proposed, addressing generalisability through multi-signal models, interpretability via LLM-driven clinical narratives, and accessibility through edge-cloud deployment, to guide future research toward clinically deployable systems.

Keywords: Cardiovascular disease, machine learning, deep learning, ECG, PPG, PCG, mobile health

1 Introduction

Cardiovascular disease (CVD) constitutes the foremost cause of global mortality, claiming approximately 17.9 million lives each year-a burden disproportionately borne by low- and middle-income countries where access to timely diagnosis remains critically limited [1]. Early detection and continuous monitoring are crucial for improving clinical outcomes, reducing hospital admissions, and enabling timely interventions. Yet access to diagnostic tools such as ECGs, echocardiograms, and specialist consultations remains limited outside major urban centres. In response, recent advances in mobile health (mHealth), wearable sensing, and artificial intelligence (AI) have opened new frontiers in digital cardiology.

At the heart of this transformation is the integration of physiological signals-including electrocardiogram (ECG), photoplethysmogram (PPG), and phonocardiogram (PCG)-with machine learning (ML) and deep learning (DL) techniques, particularly convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and hybrid architectures, which can achieve performance on par with human experts in classifying ECG abnormalities [2, 3]. These intelligent software systems automate the detection, classification, and monitoring of cardiac conditions such as arrhythmia, atrial fibrillation (AF), myocardial infarction (MI), and heart failure, and are being embedded into mobile applications, smartwatches, web-based dashboards, and cloud infrastructures, offering unprecedented opportunities for real-time, decentralised, and user-centred cardiac care [4–6].

Despite the growing volume of literature, the field remains fragmented across several dimensions that impede evidence synthesis and informed decision-making. Methodological heterogeneity is pervasive: studies employ varying preprocessing pipelines, model architectures, validation protocols, and performance metrics, making cross-study comparison difficult [7, 8]. Most systems rely on legacy benchmark datasets such as MIT-BIH, which lack demographic diversity and real-world signal variability, raising questions about generalisability [9]. The absence of standardised explainability mechanisms renders many DL models opaque to clinicians, hindering trust and adoption. External and prospective validation remains sparse, with few multi-centre

trials reported [10]. Regulatory and deployment readiness are rarely addressed, creating uncertainty about translation pathways [11, 12]. Health equity concerns-including accessibility for older adults, individuals with limited digital literacy, and populations in resource-limited settings-receive minimal attention [13]. These challenges collectively highlight the need for a systematic synthesis to consolidate fragmented evidence and identify priorities for future research.

Given this fragmentation, a systematic review is essential to consolidate the dispersed evidence base into a coherent overview, identify which approaches demonstrate robust validation versus benchmark-only performance, map the current technological landscape across signal modalities, platforms, and clinical applications, and chart evidence-informed directions for future research and clinical translation. While individual studies offer valuable technical contributions, no comprehensive synthesis currently exists that evaluates AI-driven CVD detection systems across these dimensions. This review addresses that gap, serving researchers designing new systems, clinicians evaluating adoption decisions, and policymakers developing regulatory frameworks.

This systematic review aims to comprehensively evaluate the current landscape of AI-driven software systems and mobile health applications for CVD detection using physiological signals. The following research questions guide this review:

RQ1. What software systems or mobile health applications currently exist for the detection, monitoring, or screening of CVD using physiological signals (ECG/EKG, PPG, or PCG) and machine/deep learning techniques?

RQ2. What machine learning and deep learning models are most commonly used in these systems, and how do they perform in detecting or classifying cardiovascular disease from ECG, PPG, or PCG signals?

RQ3. How are these software systems or mobile applications architected in terms of signal acquisition, preprocessing, ML/DL integration, data infrastructure, and output interpretation?

By synthesising findings from both technical and clinical perspectives, this review offers a structured overview of the field and highlights essential directions for future research.

2 Methods

This systematic review has been conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [14]. A structured review protocol was developed prior to the search process, encompassing a review approach, research questions, definition of inclusion criteria, data extraction fields, and quality assessment indicators.

2.1 Review Approach

To frame the scope and guide inclusion, we adopted the PICOC model [15], as outlined below:

Population (P): Individuals at risk of, or diagnosed with, cardiovascular disease (CVD), or general populations undergoing screening or monitoring.

Intervention (I): Software-based systems or mobile applications (e.g., apps, dashboards, chatbots) that utilise ML or DL applied to physiological signals such as ECG, PPG, or PCG.

Comparison (C): Not applicable. The focus is descriptive; performance comparisons between different systems/models are noted where available.

Outcome (O): Functional software systems or mobile apps designed for CVD detection, prediction, or monitoring; reported performance metrics and system architectures.

Context (C): Research published in English in peer-reviewed journals with accessible full text under the University of Queensland license.

2.2 Search Strategy and Information Sources

A comprehensive literature search was conducted from January 2013 to October 2024 across five databases: Scopus (n = 1,626), Web of Science (n = 792), PubMed (n = 746), Embase (n = 580), and IEEE Xplore (n = 68), yielding a total of 3,812 records. Boolean search terms combined CVD-related conditions with signal types and software modalities:

```
("Cardi*" OR "Heart failure" OR "Arrhythmia*" OR "Fibrillation")
AND ("ECG" OR "PPG" OR "PCG") AND ("Mobile App*" OR "Dashboard" OR
"Chatbot" OR ("Android" OR "IOS") AND "Apps")
```

Where supported, filters excluded review articles and meta-analyses.

2.3 Inclusion and Exclusion Criteria

After these questions have been set, the following inclusion and exclusion criteria are defined.

2.3.1 Inclusion criteria

The inclusion criteria (IC) for this systematic review are as follows:

***IC1*:** The solutions presented in the articles are applied to detect, monitor, and/or identify CVDs or cardiac-related issues utilising ECG/PPG/PCG and software systems like APP(S), Dashboard or Chatbot AND

***IC2*:** The papers are available to download or view in full under the license from the University of Queensland AND

***IC3*:** The papers are in a Journal AND

***IC4*:** The papers are written in English AND

***IC5*:** The papers provide a practical solution

2.3.2 Exclusion criteria

The exclusion criteria (EC) for this systematic review are listed below:

***EC1*:** The solutions outlined in the articles are not applied to detect, monitor, and/or identify CVDs or cardiac-related issues utilising ECG/PPG/PCG and software systems like APP(S), Dashboard or Chatbot OR

***EC2*:** The papers are not available to download or view in full under the license from the University of Queensland OR

***EC3*:** The papers are Book, Preprint, review articles or Conference OR
***EC4*:** The papers are not written in English OR
***EC5*:** The papers do not provide a cardiovascular disease analysis solution

2.4 Screening and PRISMA Flow

Fig. 1 presents the systematic review workflow following the PRISMA approach, tracing the process from initialisation to final inclusion of studies. It begins with the initialisation phase, where protocol sheets are defined, search strings are identified and refined, and queries are formulated in consultation with co-authors and supervisors. Following this, the identification phase involved selecting a database and conducting a systematic search across multiple sources, including Scopus, Web of Science, PubMed, Embase, and IEEE, which resulted in 3812 articles. After duplicate removal via Covidence, 1972 articles remained.

The screening phase was conducted in Covidence using a two-reviewer independent process. At the first level, two reviewers (M.T.A. and A.A.B.) independently screened titles and abstracts, excluding 1,443 irrelevant articles and retaining 529 for full-text assessment. At the second level, the same two reviewers independently evaluated the full texts, with disagreements resolved through discussion. This process ensured methodological rigour and minimised selection bias, ultimately yielding 30 studies eligible for data extraction. The included studies were then subjected to bibliometric analysis. The pie chart highlights the distribution of publications across journals, showing that research is spread across multiple venues, with a few journals contributing more prominently. The bar chart shows the temporal distribution of publications, indicating a clear increase in outputs over recent years, particularly from 2019 onwards, which demonstrates accelerating interest and activity in this research domain.

2.5 Data Extraction and Synthesis

For each included study, data were extracted on signal modality (ECG, PPG, PCG, HR), type of software platform (mobile, cloud, web), ML/DL algorithm used (e.g., SVM, CNN, CatBoost), application type (screening, monitoring), deployment status (prototype, deployed app), evaluation metrics (accuracy, sensitivity, AUC), and validation methods. The findings were synthesised thematically and structured according to the three predefined research questions. Each included study was examined along the following dimensions:

- *System Description*: Type of application (e.g., mobile app, dashboard, cloud-based service), software architecture, deployment environment (on device, cloud, hybrid), user interface modalities, and stated clinical or operational objectives.
- *Signal and Data Handling*: Type(s) of physiological signals used (e.g., ECG, PPG, PCG), signal acquisition methods (sensor type, sampling rate, duration), and preprocessing steps (e.g., denoising, normalisation, segmentation).
- *Machine Learning Integration*: Type of ML/DL models implemented (e.g., CNN, LSTM, ensemble models), training methodologies (cross-validation, hold-out), data sources and labelling strategies, performance metrics (accuracy, sensitivity,

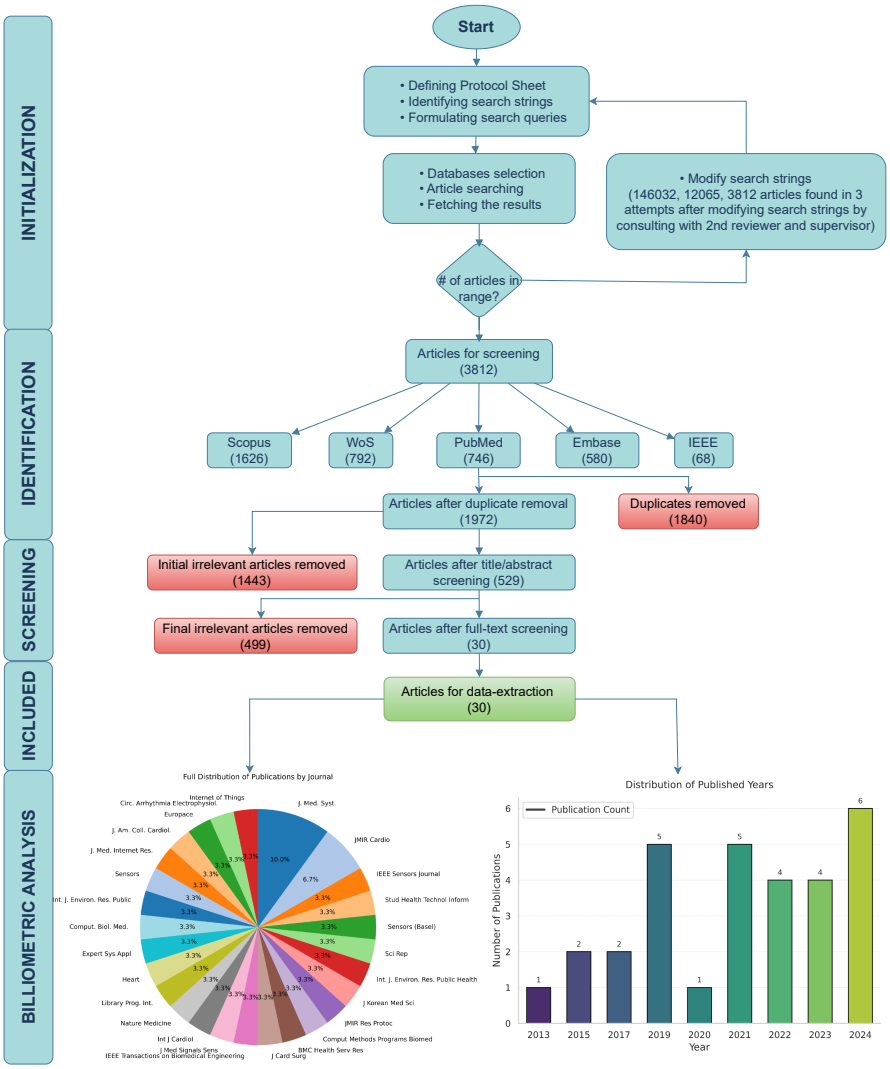


Fig. 1: PRISMA 2020 flow diagram summarising the screening process, followed by the synthesis of articles via bibliographic analysis.

specificity, F1-score, AUC), and model interpretability (e.g., feature attribution, explainable AI components).

- *Validation and Evaluation*: Experimental setup (e.g., simulated, real-world deployment, retrospective analysis), validation strategies (internal vs external), sample size, population demographics (if available), and outcome measurement standards.

- *Stated Limitations and Future Directions*: Author-reported constraints in scalability, generalisability, model robustness, and data quality; recommendations for future system enhancements or clinical integration.

3 Results

This section presents findings from 30 peer-reviewed studies on AI-enabled CVD detection systems. Table 1 provides a comprehensive overview of all included studies, while subsequent subsections address the three research questions through synthesis of signal modalities, model architectures, and system infrastructure.

Table 1: Summary Characteristics of Included Studies (N=30)

Study (Year)	Signal	AI Model	Condition	Acc%	F1%	AUC	Dataset	N	Architecture	Deploy
El-Sofany (2024)	ECG	XGBoost	CVD Risk	97.6	92.7	.98	Cleveland+Fram.	1,190	Mobile App	Proto
Tiwari (2023)	ECG	CNN-LSTM	Arrhythmia	98.0	99.0[a]	NR	MIT-BIH	47 rec	Mobile App	Proto
Han (2023)	PPG	Custom	AF	NR	NR	NR	Pulsewatch	34	Wearable+Cloud	Pilot
Shao (2020)	ECG	CatBoost	AF	99.6	92.0	NR	MITBIH-AFDB	7,270	Wearable+App	Proto
Alshraideh (2015)	ECG	DT (J48)	MI	98.3	NR	NR	UCI Arrhythmia	410	Web-based	Proto
Smieszek (2021)	ECG+PPG	NR	HF/AF	NR	NR	NR	NR	NR	Wearable+Cloud	Pilot
Ma (2024)	PCG	CNN-SSL	VHD	98.7	NR	NR	Custom PCG	NR	Edge Device	Proto
Fu (2021)	ECG	CNN-RNN	Multi-CVD	99.3	66.9	.99	Hospital ECGs	15,437	Mobile App	Proto
Lee (2013)	PPG	Statistical	AF	93.7[b]	NR	NR	MIT-BIH+Custom	50	Mobile App	Proto
Verbrugge (2019)	PPG	RNN+RT	AF	NR	NR	NR	Clinical valid.	223	Mobile App	Clinical
Wong (2024)	ECG	CNN+UNet	Multi-CVD	NR	93.4	NR	Custom clinical	8,368	Mobile App	Proto
Melillo (2015)	ECG+Act	RF/SVM	CVD Risk	84.0	NR	.82	Custom clinical	139	Mobile+Cloud	Proto
Yashudas (2024)	Multi	BiGRU-Att	CVD Risk	99.9	98.0	.999	Fram.+Statlog	4,238	IoT Platform	Proto
Klier (2023)	ECG	FDA-ML	AF	NR	NR	NR	Clinical valid.	200	Mobile App	Clinical
Xiong (2022)	ECG	CRN	AF	93.8	96.4	NR	Custom+PhysioNet	2,917	Wearable+App	Proto
Desai (2022)	NR	Ensemble	CHD	86.0	NR	.96	Cleveland HD	303	Cloud-based	Proto
Lei (2021)	HR	BiLSTM	AF	98.5	NR	NR	AFDB+LTAFDB	71 rec	Mobile App	Proto
Kim (2024)	ECG	ResNet50	AF	70.5	71.9	.79	Custom+External	1,266	Mobile App	Proto
Sanamdikar (2024)	ECG	DNN	Multi-CVD	NR	NR	NR	NR	NR	Cloud-based	Proto
Hermans (2022)	PPG	Custom	AF	NR	NR	NR	Clinical study	45	Mobile App	Pilot
Golzar (2017)	ECG	Rule-based	Arrhythmia	91.6	NR	NR	Hospital valid.	20	Mobile App	Proto
Guo (2019)	PPG	PPG-Algo	AF	99.0[c]	NR	NR	HUAWEI Heart	187,912	Wearable+Cloud	Clinical
Islam (2023)	Multi	Stacking	CVD Risk	87.5	91.0	.96	Custom IoT	1,000[c]	IoT Platform	Proto
Moreno-A (2017)	Multi	NR	CVD Screen	NR	NR	NR	Custom clinical	16	Mobile App	Pilot
Attia (2022)	ECG	CNN	LVSD	NR	NR	.89	Mayo Clinic	701	Wearable+Cloud	Clinical
Atilgan (2021)	Multi	NR	Post-op	NR	NR	NR	Clinical study	50	Telemedicine	Pilot
Wasserlauf (2019)	ECG+PPG	CNN	AF	NR	NR	NR	Clinical (ICM)	24	Wearable+Cloud	Proto
Kilimo (2019)	ECG	Wavelet	Arrhythmia	NR	NR	NR	Protocol only	-	Mobile App	Proto
Lucki (2021)	ECG	NR	Post-stroke	NR	NR	NR	Clinical study	32	Mobile App	Pilot
Zhang (2019)	PPG	Boosting	AF	NR	NR	NR	Clinical valid.	361	Wearable+App	Proto

Abbreviations: Multi=Multimodal; Act=Actigraphy; VHD=Valvular Heart Disease; CHD=Coronary Heart Disease; LVSD=Left Ventricular Systolic Dysfunction; N=Number of participants/samples; rec=recordings; Deploy=Deployment (Proto=Prototype, Pilot=Pilot study, Clinical=Clinical deployment); NR=Not Reported; Fram=Framingham; valid.=validation. [a]Best class (Normal beat). [b]Average across methods. [c]Approximate value.

3.1 Landscape of Software Systems and Applications for CVD Detection

The technological landscape for CVD detection has diversified rapidly, spanning consumer-grade health apps to clinical-grade diagnostic tools. Fig. 2 provides a consolidated overview across four dimensions: signal types, clinical contexts, primary applications, and system components.

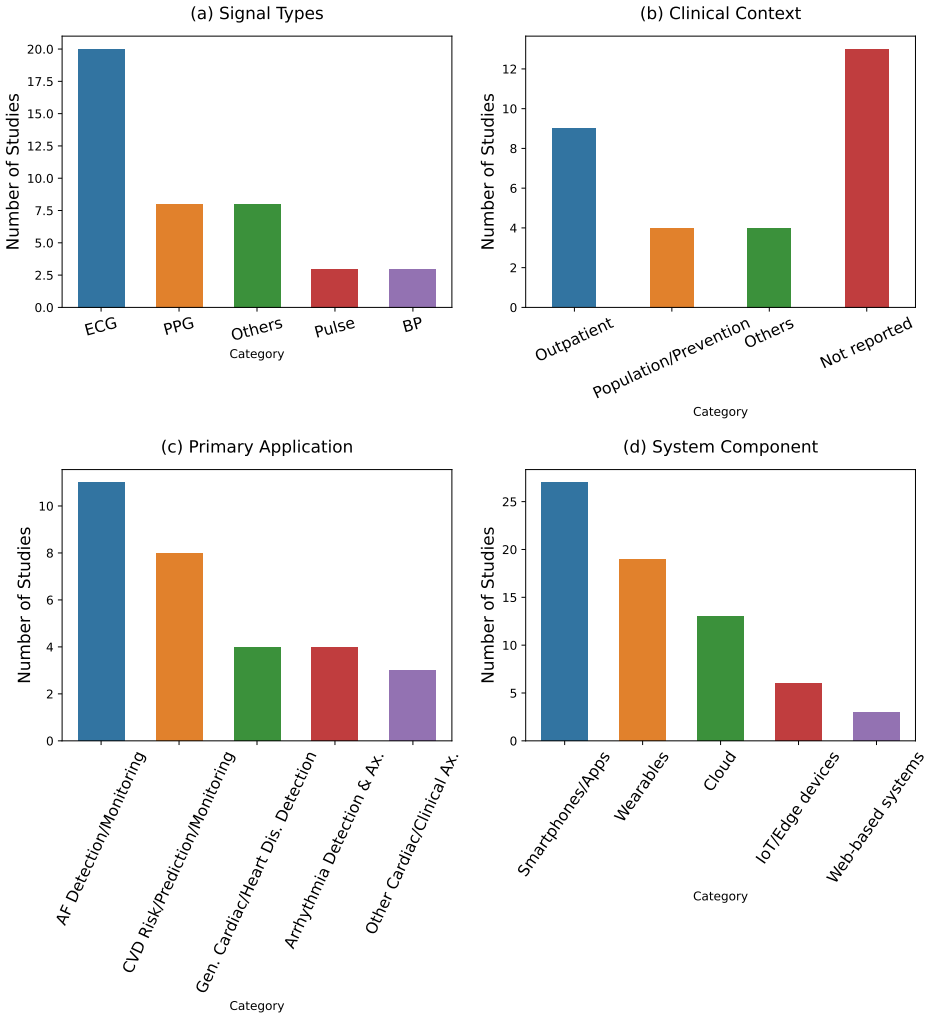


Fig. 2: Distribution of studies by (a) signal types, (b) clinical context, (c) primary application, and (d) system components. ECG dominates signal modalities (66.6%), while 43% of studies omit clinical context reporting.

3.1.1 Signal Types and Input Modalities

Terminology Note: This review distinguishes between *multimodal systems*-combining a primary signal with supplementary features (e.g., ECG + SpO_2 + blood pressure)- and *multi-signal fusion*-principled integration of complementary physiological signals (ECG + PPG + PCG) through dedicated architectures.

As shown in Fig. 2a, ECG signals dominated (n=20, 66.6%), reflecting their clinical validation as the gold standard for diagnosing arrhythmias and conduction abnormalities [16, 17]. PPG signals appeared in 8 studies (26.7%), typically via smartwatch

optical sensors for AF screening [18]. Only one study (3.3%) utilised PCG, reflecting limited standardisation in acoustic signal acquisition.

Seven studies employed multimodal designs integrating physiological signals with supplementary features such as SpO_2, blood pressure, or activity data [19–25]. However, true multi-signal fusion-integrating ECG (electrical), PPG (haemodynamic), and PCG (mechanical) through principled architectures-remains unexplored (0/30 systems), representing a critical gap for comprehensive cardiovascular AI.

3.1.2 Clinical Context and Target Conditions

Clinical contextualisation remains inadequate: 43% of systems failed to disclose care settings, creating ambiguity about intended use-cases (Fig. 2b). Among those specifying settings, 9 studies targeted outpatient environments, 4 addressed population screening, and 4 covered other contexts. Hospital or high-risk clinical validation was notably absent.

AF detection dominated clinical applications (n=11, 36.7%), followed by CVD risk prediction (n=9, 30%) and general cardiac assessment (n=7, 23.3%) (Fig. 2c). Notable systems include the Pulsewatch smartwatch platform processing over 33 million seconds of PPG data [18] and Shao *et al.*'s CatBoost classifier achieving F1=0.92 on 7,000+ ECG samples [17].

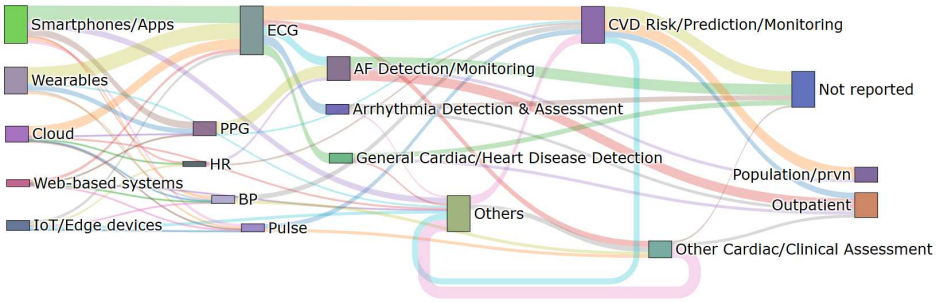


Fig. 3: Flow diagram linking system types, signal inputs, application domains, and deployment contexts, illustrating heterogeneity in design and weak alignment between technical innovation and clinical practice. The Sankey visualisation traces the cascade from platform architecture through signal modality to clinical application, revealing that mobile app-based systems predominantly utilise ECG signals for AF detection, while wearable platforms show greater diversity in signal integration.

Fig. 3 traces the cascade from system type through signal modality to clinical application, revealing weak alignment between technical innovation and deployment contexts. Klier *et al.* [26] validated mobile ECG tools and demonstrated that performance was highly context-sensitive, reinforcing the importance of deployment-specific validation.

3.1.3 System Deployment Modalities

Smartphone/app-based systems predominated (n=27, 90%), typically enabling real-time signal analysis via external sensors (Fig. 2d). Representative examples include El-Sofany *et al.*'s XGBoost-based mobile application (97.57% accuracy) [16] and Tiwari *et al.*'s ConvNet-LSTM hybrid (98% accuracy) [27]. Wearables (n=19) and cloud-based systems (n=13) were also prevalent, with hybrid edge-cloud architectures emerging to balance computational constraints with analytical capability.

Only three studies reported clinical deployment at scale: the HUAWEI Heart Study screening 187,912 participants [28], the Mayo Clinic-Apple Watch evaluation [10], and the Pulsewatch system [18]. Notably, no study implemented chatbot or conversational interfaces for cardiovascular screening-a complete absence of patient-facing conversational AI in this domain.

3.2 Model Architectures and Performance Metrics

Fig. 4 presents a consolidated view of model architectures, input features, validation methods, and reported metrics across the 30 studies.

3.2.1 Architectural Patterns

CNN-based architectures represented 27% of implementations, leveraging hierarchical feature extraction from raw waveforms without handcrafted preprocessing [17]. Recurrent models (LSTM, GRU) accounted for 20%, addressing temporal dependencies in arrhythmia patterns [29]. Ensemble and hybrid architectures (27% combined) reflect an emergent design philosophy fusing DL's representational power with ML's interpretability-methods such as XGBoost and CatBoost provide transparent decision boundaries suitable for clinical adoption [27]. Legacy ML classifiers retain relevance in resource-constrained settings where computational efficiency and explainability are paramount.

Fig. 5 compares performance across model categories. CNN-based models achieved the highest overall balance, while RNN/LSTM architectures demonstrated superior AUC, reflecting their strength in temporal pattern recognition. Ensemble methods maintained robust sensitivity, and traditional ML approaches remained competitive for interpretable deployments.

3.2.2 Performance Metrics and Validation Deficiencies

Performance metric reporting was inconsistent across studies (Table 1). Among the 30 systems reviewed, accuracy was reported in 15 studies (50%), with values ranging from 70.5% to 99.9%. Sensitivity was reported in 18 studies (60%), specificity in 13 studies (43%), F1-score in 10 studies (33%), and AUC in 10 studies (33%). PPV and NPV were rarely disclosed. Notably, 9 studies failed to report any quantitative performance metrics, and 11 omitted validation protocol descriptions entirely [25, 26].

Validation rigour represents the weakest dimension across studies. As shown in Fig. 4d, k-fold cross-validation and train-test splits were equally adopted (n=6 each),

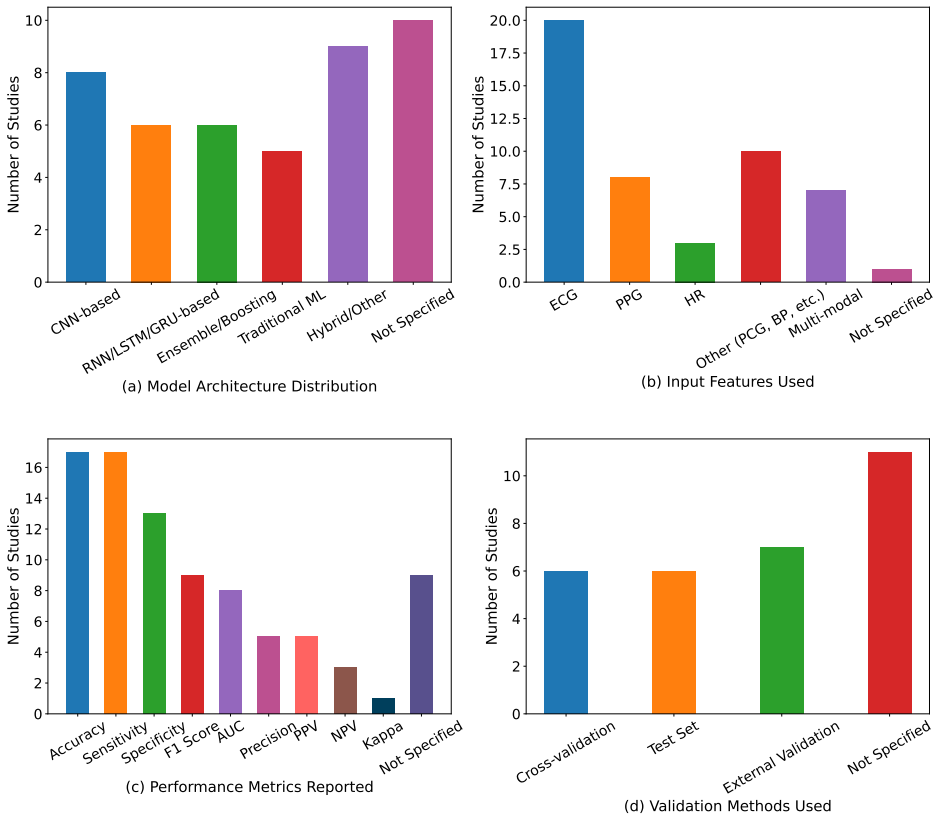


Fig. 4: Dashboard of model characteristics: (a) architecture distribution showing CNN prevalence (27%), (b) input feature reliance on ECG, (c) inconsistent performance metric reporting, and (d) validation method distribution with limited external validation (23%).

but only 7 studies (23%) pursued external validation. Fig. 6 demonstrates that external validation studies reported broader metrics, suggesting more rigorous evaluation standards.

Fig. 7 provides an integrated visualisation linking model types, input features, performance metrics, and validation approaches. The figure reveals that CNN-ECG pipelines dominate the landscape, while multimodal hybrids are more frequently associated with external validation and richer metric reporting. This pattern suggests that methodological rigour correlates with system complexity, though the limited use of hospital-acquired or real-world data undermines overall translational readiness.

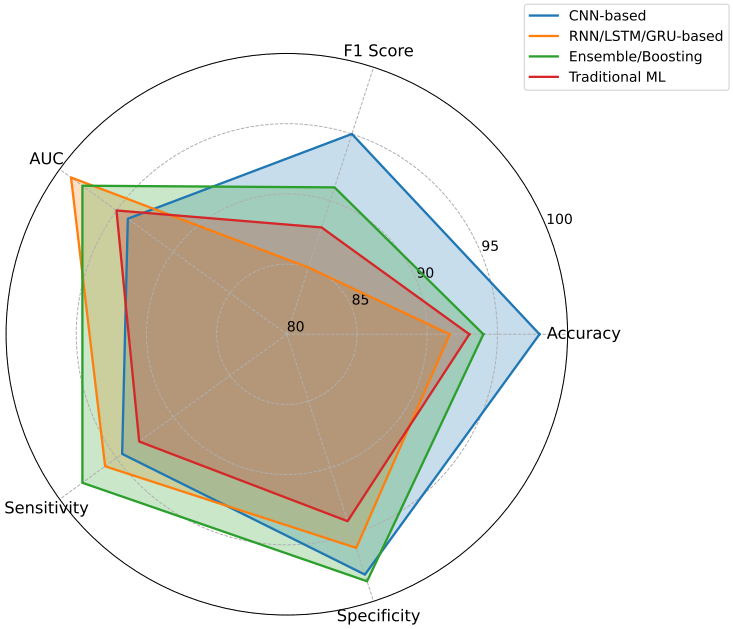


Fig. 5: Radar chart comparing average performance metrics across model types. CNN-based models achieve the highest balance (≈98% accuracy, 98% specificity), while RNN/LSTM architectures show superior AUC (≈0.99).

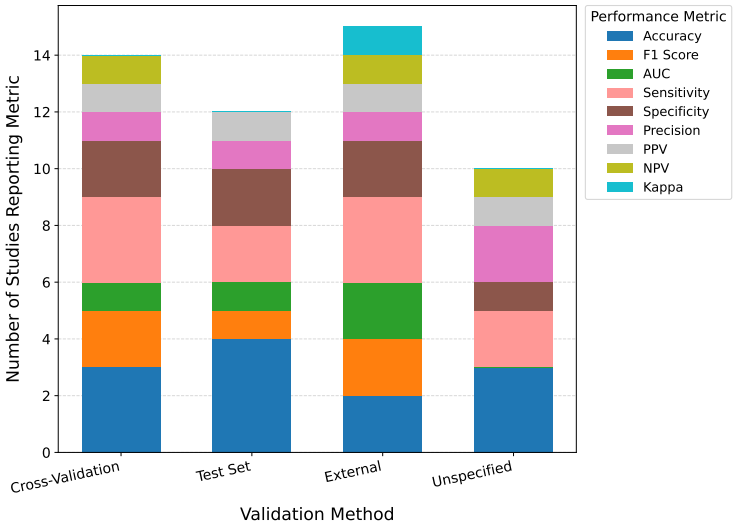


Fig. 6: Performance metrics stratified by validation method. External validation studies reported broader metric sets (AUC, F1, PPV), while internal validations relied predominantly on accuracy alone.

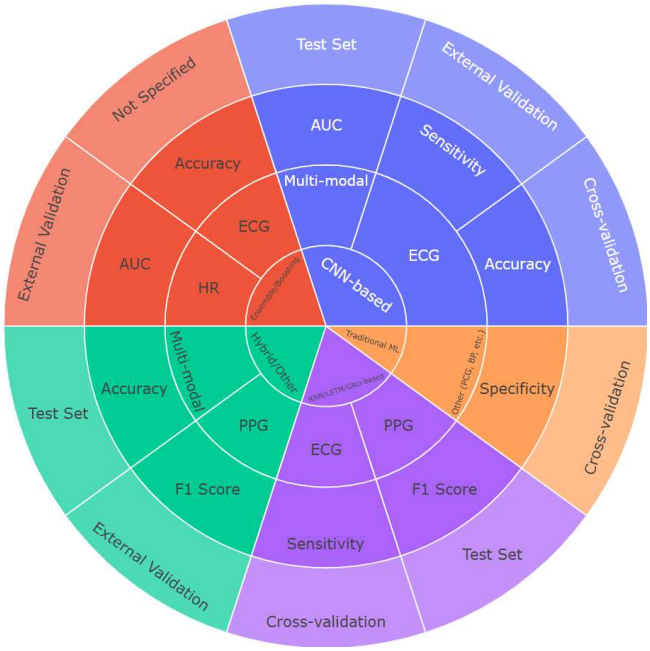


Fig. 7: Integrated view linking model architectures, input features, performance metrics, and validation methods. The sunburst visualisation reveals that CNN-ECG pipelines dominate, while multimodal systems are more frequently associated with external validation and comprehensive metric reporting.

3.3 System Architecture and Database Infrastructure

Modern CVD detection systems comprise five interoperable layers: signal acquisition, preprocessing, model execution, user interface, and data stewardship. Fig. 8 illustrates this conceptual pipeline.

At the acquisition layer, biosignals are sourced from dedicated ECG devices (AliveCor, Kardia, MD-Link), consumer wearables (Apple Watch, Fitbit), and custom IoT sensors [30]. Signal heterogeneity necessitates adaptable preprocessing: ECG signals undergo band-pass filtering (0.5-40 Hz), QRS detection, and wavelet decomposition [22, 31], while PPG signals require motion artifact removal and beat segmentation.

The model execution layer spans embedded inference engines (CoreML, TensorFlow Lite) to cloud-hosted ensembles. Hybrid edge-cloud architectures, exemplified by Pulsewatch [32], optimise latency and scalability while enabling continuous learning. Foundation models such as the knowledge-enhanced ECG diagnosis system [33] show promise for reducing retraining burden through pre-trained signal representations.

User interfaces vary from patient-facing apps with risk scores to clinician dashboards with ECG overlays. Interpretability modules using SHAP, Grad-CAM, or class activation maps appear in some systems [29], though formal explainable AI integration remains limited despite regulatory demands.

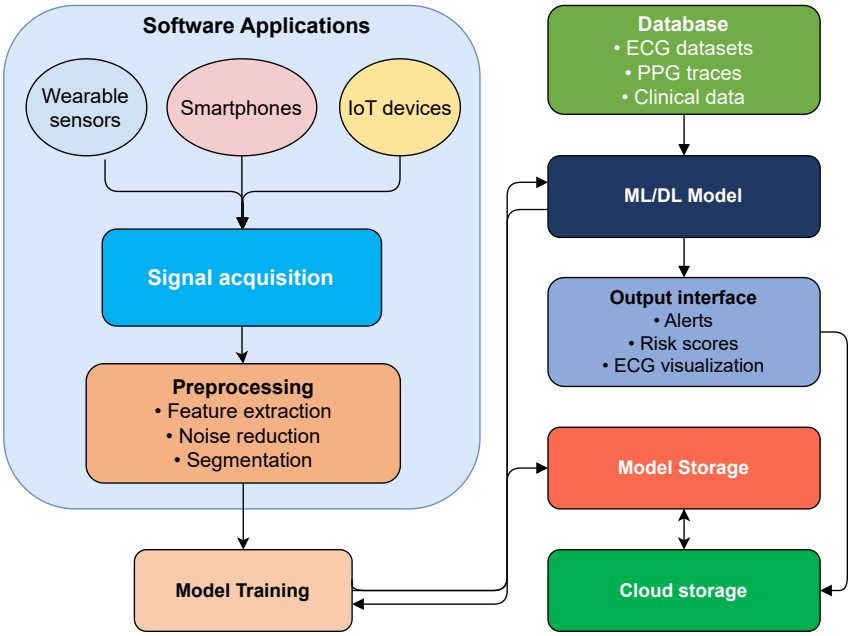


Fig. 8: Conceptual architecture of ML/DL-based CVD systems showing layered workflows from signal acquisition through preprocessing, model inference, and output delivery, with supporting database and cloud infrastructure.

Dataset utilisation was uneven (Fig. 9): benchmark public datasets dominated (n=10), followed by custom real-time datasets (n=8), while hospital repositories were rarely used (n=2). The MIT-BIH Arrhythmia Database appeared in only 3 studies (10%) [17, 27], and custom clinical datasets in just 2 studies (6.7%). This over-reliance on legacy benchmarks raises concerns about generalisability to real-world signal noise, demographic variation, and sensor heterogeneity.

Database infrastructure varied by deployment context: public datasets utilise WFDB standards with SQL interfaces; private platforms combine InfluxDB with Firebase for real-time logging; and hospital repositories employ PostgreSQL with HL7/FHIR compliance. Despite architectural sophistication, few studies reported security implementations (AES-256 encryption, role-based access) critical for GDPR, HIPAA, and MDR compliance.

4 Discussion

This systematic review synthesised evidence from 30 studies to evaluate AI-driven CVD detection systems using physiological signals. The analysis reveals a field characterised by substantial technical achievement alongside persistent barriers to clinical

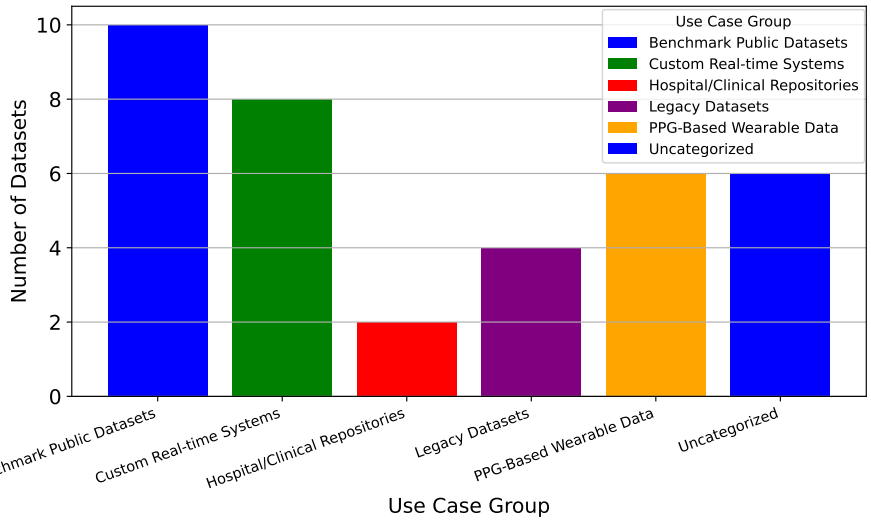


Fig. 9: Dataset utilisation across studies showing dependence on public benchmarks (n=10) and underuse of hospital repositories (n=2). This pattern raises concerns about generalisability to real-world signal variability.

translation. This discussion summarises key findings and their real-world implications, then presents a three-directional framework-derived directly from the identified gaps-that structures actionable pathways toward clinically viable systems.

Four principal findings emerge from this synthesis, each with distinct implications for clinical practice and future development. The first two findings directly inform the first two framework directions, while findings three and four collectively inform the third direction addressing deployment and patient engagement.

Unimodal Signal Reliance: The dominance of single-modality ECG systems (66.6%) reflects both clinical convention and technical convenience, yet imposes fundamental constraints on diagnostic scope. ECG captures only electrical cardiac activity, missing mechanical dimensions (valve function, heart sounds) and haemodynamic dynamics (blood flow, peripheral perfusion). This limitation manifests clinically as reduced robustness in ambulatory settings with motion artefacts and constrained generalisability across diverse acquisition conditions-limitations explicitly acknowledged by landmark studies [2, 3, 6]. The severe underrepresentation of PCG (3.3%) and complete absence of multi-signal fusion frameworks (0/30 systems) represents a critical gap, as multi-signal integration would provide complementary diagnostic information and signal redundancy when individual modalities are corrupted.

Interpretability Deficit: Despite deep learning dominance, no system integrates large language models with cardiovascular AI for clinical narrative generation. Existing explainable AI techniques (SHAP, Grad-CAM) are applied retrospectively rather than embedded within real-time decision support, and remain fundamentally misaligned with clinical reasoning-saliency maps are not linked to established clinical features such as ST changes or QT prolongation. The core problem: clinicians need narratives,

not saliency maps. While Med-PaLM demonstrated that LLMs can encode substantial clinical knowledge (67.6% MedQA accuracy) [34], no existing system leverages this capability to translate AI probability outputs into guideline-aligned diagnostic language.

Unrealistic Deployment Assumptions: Current systems assume constant cloud connectivity and unlimited battery-conditions incompatible with real-world wearable and mobile deployment. Only 23% of studies conducted external validation, with edge-cloud architectures appearing conceptually but rarely operationalised. Real-world deployment benchmarking is systematically absent: no studies reported latency testing, battery constraints, network variability handling, or signal quality degradation under practical conditions. Furthermore, regulatory alignment with IEC 62304, ISO 14971, or FDA SaMD criteria is virtually absent, meaning promising prototypes cannot progress to clinical certification without substantial redesign [11, 12].

Patient Engagement Void: Zero of 30 systems (0/30) implemented patient chatbots or conversational interfaces. Current approaches treat patients as passive alert recipients without education (patients do not understand what AF detection means), personalised feedback (generic warnings only), symptom tracking, or bidirectional clinician-patient communication. This entirely unidirectional approach fails to leverage conversational AI for sustained cardiovascular care engagement and health literacy [20, 35].

These four findings reveal interconnected barriers that cannot be addressed in isolation. We therefore propose a three-directional framework Fig. 10 that synthesises these insights into an integrated pathway toward clinically viable AI-CVD systems. Each direction directly addresses specific gaps identified in this review: Gap 1 informs Direction 1, Gap 2 informs Direction 2, and Gaps 3 and 4 collectively inform Direction 3.

The first direction focuses on generalisable multi-signal AI models to address unimodal signal limitations. A generalised multi-signal framework incorporating ECG, PPG, and PCG would capture three distinct dimensions of cardiac function: electrical conduction (arrhythmias, ischaemia), haemodynamic activity (blood volume changes, heart rate variability, peripheral perfusion), and mechanical activity (heart sounds S1/S2, valvular disorders, structural abnormalities). Technical approaches include transformer-based fusion architectures for cross-modal feature alignment [21], neural networks with uncertainty quantification for calibrated predictions [36], and attention mechanisms enabling interpretable weighting of signal contributions. The multi-signal approach provides robustness through redundancy-if one signal is corrupted by artefact, others maintain diagnostic capability-while enabling comprehensive cardiovascular assessment that single-modality systems fundamentally cannot achieve.

The second direction addresses the interpretability deficit through LLM-driven clinical narratives. Large language model-based systems can bridge the gap between low-level AI explanations and high-level clinical narratives, transforming opaque probability outputs into structured reports aligned with physiological reasoning. Rather than displaying "AF probability: 0.87," systems would generate interpretable clinical narratives: "Irregular RR intervals with absent P-waves detected in leads II and V1, consistent with atrial fibrillation. Heart rate variability analysis shows reduced

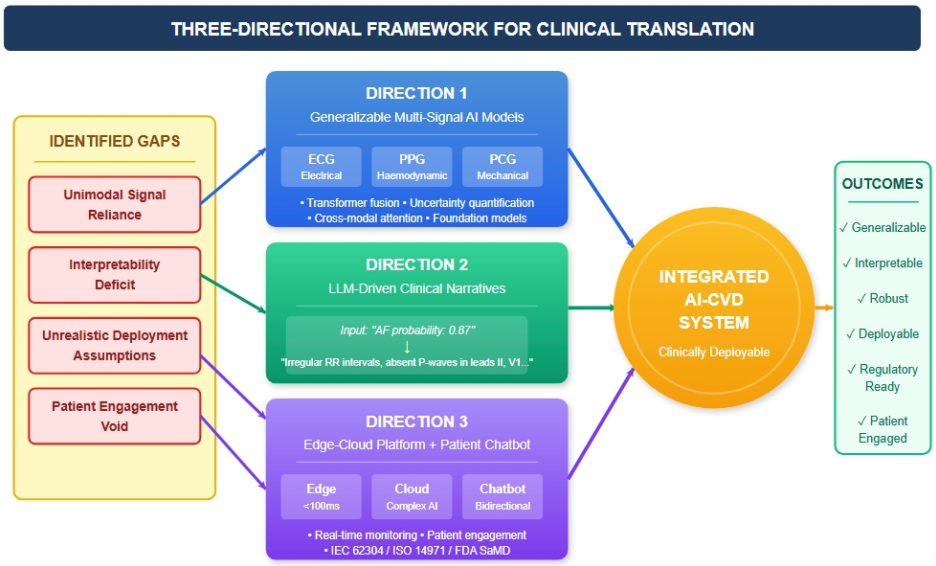


Fig. 10: Three-directional framework synthesising review findings into an integrated pathway toward clinically viable AI-CVD systems, addressing generalisability (multi-signal AI), interpretability (LLM-driven narratives), and accessibility (edge-cloud platforms with patient chatbots).

SDNN indicating autonomic dysfunction." This approach extends existing attention and SHAP mechanisms [16] into guideline-aligned diagnostic language that integrates with clinical workflows, facilitating clinician-in-the-loop validation. The narrative format directly addresses regulatory requirements for AI explainability while building clinician trust through alignment with established diagnostic reasoning patterns.

The third direction proposes an edge-cloud integrated platform with patient chatbot to address both deployment constraints and the patient engagement void. An edge-cloud hybrid architecture would optimise task delegation through context-aware orchestration: lightweight preprocessing and initial screening on edge devices (targeting ¡100ms latency with TensorFlow Lite) with complex inference delegated to cloud resources when connectivity permits [24, 29]. This architecture must be empirically tested under real-world constraints including battery limitations, network variability, and signal quality degradation. Crucially, this platform should integrate a patient-facing conversational interface enabling symptom reporting, health education, intelligent triage, and bidirectional clinician-patient communication-transforming episodic alerts into sustained, interactive cardiovascular care. A clinician dashboard for professional oversight, result review, and clinical workflow integration completes the ecosystem. Regulatory compliance (IEC 62304, ISO 14971, FDA SaMD) must be embedded from the design phase through systematic version control, audit trails, and risk assessment practices.

Several limitations of this review should be acknowledged. The search was limited to English-language peer-reviewed journal articles, potentially excluding relevant conference proceedings, preprints, and non-English publications; however, this ensured consistent quality standards. Heterogeneity in reported metrics and validation approaches precluded quantitative meta-analysis; the qualitative synthesis nonetheless enables pattern identification not apparent in individual studies. Rapid technological advancement means newer systems may not be captured. The focus on ECG, PPG, and PCG excludes other modalities (cardiac MRI, echocardiography) contributing to comprehensive CVD assessment. Despite these limitations, the systematic methodology provides a robust foundation for understanding current capabilities and future directions.

This systematic review makes four principal contributions to the field. First, it establishes a comprehensive taxonomy of AI-enabled CVD detection systems, mapping the landscape across platforms, signals, architectures, and applications. Second, it moves beyond reporting claimed performance to critically evaluate validation rigour, quantifying unimodal signal reliance (66.6%), external validation gaps (23%), and patient engagement absence (0/30 chatbots). Third, it identifies four specific, quantified research gaps with direct implications for system development. Fourth, it synthesises findings into an evidence-informed three-directional framework providing a structured roadmap from current fragmentation toward integrated, clinically deployable systems.

5 Conclusion

This systematic review of 30 AI-driven CVD detection systems reveals a field characterised by technical sophistication alongside critical translational barriers: unimodal signal reliance, interpretability deficits, unrealistic deployment assumptions, and absent patient engagement mechanisms. The proposed three-directional framework-multi-signal AI models, LLM-driven clinical narratives, and edge-cloud platforms with patient chatbots-provides an integrated pathway from current fragmentation toward clinically deployable systems. Realising AI's potential in cardiovascular diagnostics requires progression beyond accuracy-centric benchmarks toward generalisability, interpretability, and regulatory readiness.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: Not applicable for this systematic review study.

Data availability: All data extracted from the included studies are presented in the tables and figures within this manuscript.

Author contributions: M.T.A. conceptualised the study, designed the methodology, conducted the systematic search, performed data extraction and analysis, prepared figures and drafted the manuscript. A.A.B. participated in the screening

process, conducted an independent full-text assessment, and contributed to data validation. A.K.M.A. provided feedback and contributed to manuscript revision. A.K. provided critical feedback on the methodology and contributed to manuscript revision. M.A.M. supervised the project, provided guidance throughout the research process, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

References

[1] World Health Organization: Cardiovascular Diseases (CVDs). WHO Fact Sheets (2023)

[2] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory ECGs using a deep neural network. Nat. Med. **25**, 65–69 (2019) https://doi.org/10.1038/s41591-018-0268-3

[3] Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P.S., Andersson, C.R., Macfarlane, P.W., Meira Jr., W., Schön, T.B., Ribeiro, A.L.P.: Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat. Commun. **11**, 1760 (2020) https://doi.org/10.1038/s41467-020-15432-4

[4] Topol, E.J.: High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. **25**, 44–56 (2019) https://doi.org/10.1038/s41591-018-0300-7

[5] Meskó, B., Drobni, Z., Bényei, E., Gergely, B., Győrffy, Z.: Digital health is a cultural transformation of traditional healthcare. mHealth **3**, 38 (2017) https://doi.org/10.21037/mhealth.2017.08.07

[6] Attia, Z.I., Noseworthy, P.A., Lopez-Jimenez, F., Asirvatham, S.J., Deshmukh, A.J., Gersh, B.J., Carter, R.E., Yao, X., Rabinstein, A.A., Erickson, B.J., Kapa, S., Friedman, P.A.: An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. Lancet **394**(10201), 861–867 (2019) https://doi.org/10.1016/S0140-6736(19)31721-0

[7] Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., Tsaneva-Atanasova, K.: Artificial intelligence, bias and clinical safety. BMJ Qual. Saf. **28**(3), 231–237 (2019) https://doi.org/10.1136/bmjqs-2018-008370

[8] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. **1**, 206–215 (2019) https://doi.org/10.1038/s42256-019-0048-x

[9] Rajpurkar, P., Lungren, M.P.: AI in healthcare: The black box issue. JAMA **320**(11), 1105–1106 (2018) https://doi.org/10.1001/jama.2018.13396

[10] Attia, Z.I., Harmon, D.M., Dugan, J., Manka, L., Lopez-Jimenez, F., Lerman, A., Siontis, K.C., Friedman, P.A., Noseworthy, P.A.: Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. Nature Medicine **28**, 2497–2503 (2022) https://doi.org/10.1038/s41591-022-02053-1

[11] Alshraideh, H., Otoom, M., Al-Araida, A., Bawaneh, H., Bravo, J.: A web based cardiovascular disease detection system. J. Med. Syst. **39**(10), 122 (2015) https://doi.org/10.1007/s10916-015-0290-7

[12] Hermans, A.N.L., Gawalko, M., Dohmen, L., Velden, R.M.J., Betz, K., Duncker, D., Verhaert, D.V.M., Heidbuchel, H., Svennberg, E., Linz, D., Hindricks, G., Crijns, H.J.G.M., Pison, L., Hendriks, J.M.L.: Mobile app-based symptom-rhythm correlation assessment in patients with persistent atrial fibrillation. Int. J. Cardiol. **367**, 29–37 (2022) https://doi.org/10.1016/j.ijcard.2022.08.021

[13] Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used for healthcare management. Science **366**(6464), 447–453 (2019) https://doi.org/10.1126/science.aax2342

[14] Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int. J. Surg. **8**(5), 336–341 (2010) https://doi.org/10.1016/j.ijsu.2010.02.007

[15] Petticrew, M., Roberts, H.: Systematic Reviews in the Social Sciences: A Practical Guide vol. 11. Blackwell Publishing, Oxford, UK (2006). https://doi.org/10.1002/9780470754887

[16] El-Sofany, H., Bouallegue, B., El-Latif, Y.M.A.: A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. Scientific Reports **14**(1), 23277 (2024) https://doi.org/10.1038/s41598-024-74656-2

[17] Shao, M., Zhou, Z., Bin, G., Bai, Y., Wu, S.: A wearable electrocardiogram telemonitoring system for atrial fibrillation detection. Sensors **20**(3), 606 (2020) https://doi.org/10.3390/s20030606

[18] Han, D., Ding, E.Y., Cho, C., Jung, H., Dickson, E.L., Mohagheghian, F., Peitzsch, A.G., DiMezza, D., Tran, K.V., McManus, D.D., Chon, K.H.: A smartwatch system for continuous monitoring of atrial fibrillation in older adults after stroke or transient ischemic attack: Application design study. JMIR Cardio **7**, 41691 (2023) https://doi.org/10.2196/41691

[19] Smieszek, M., Kindermann, A., Amr, A., Meder, B., Dieterich, C.: An Apple Watch dashboard for HiGHmed heart insufficiency patients. In: Stud. Health Technol. Inform., vol. 283, pp. 146–155 (2021). https://doi.org/10.3233/SHTI210553

[20] Melillo, P., Orrico, A., Scala, P., Crispino, F., Pecchia, L.: Cloud-based smart health monitoring system for automatic cardiovascular and fall risk assessment in hypertensive patients. J. Med. Syst. **39**(10), 109 (2015) https://doi.org/10.1007/s10916-015-0294-3

[21] Yashudas, A., Gupta, D., Prashant, G.C., Dua, A., AlQahtani, D., Reddy, A.S.K.: DEEP-CARDIO: Recommendation system for cardiovascular disease prediction using IoT network. IEEE Sensors J. **24**(9), 14539–14547 (2024) https://doi.org/10.1109/JSEN.2024.3373429

[22] Islam, M.N., Mustafina, J., Mahmud, T., Khan, N.I., *et al.*: Predictis: an IoT and machine learning-based system to predict risk level of cardio-vascular diseases. BMC Health Serv. Res. **23**(1), 171 (2023) https://doi.org/10.1186/s12913-023-09104-4

[23] Moreno-Alsasua, L., García-Zapirain, B., Méndez-Zorrilla, A., Ruiz-Zubero, I., Amayra-Caro, I.: Primary prevention of asymptomatic cardiovascular disease using physiological sensors connected to an iOS app. J. Med. Syst. **41**(12), 189 (2017) https://doi.org/10.1007/s10916-017-0840-2

[24] Atilgan, K., Onur Hanedan, M., Yilik, L., Beyaztas, S., Ozkan, M., Cevirme, D., Yeniterzi, M., Kestelli, M., Yucel, O.: Remote patient monitoring after cardiac surgery: The utility of a novel telemedicine system. J. Card. Surg. **36**(11), 4226–4234 (2021) https://doi.org/10.1111/jocs.15962

[25] Wasserlauf, J., You, C., Patel, R., Valys, A., Albert, D., Passman, R.: Smartwatch performance for the detection and quantification of atrial fibrillation. Circ. Arrhythmia Electrophysiol. **12**(6), 006834 (2019) https://doi.org/10.1161/CIRCEP.118.006834

[26] Klier, K., Koch, L., Graf, L., Schinköthe, T., Schmidt, A.: Diagnostic accuracy of single-lead electrocardiograms using the Kardia Mobile app and the Apple Watch 4: Validation study. JMIR Cardio **7**, 50701 (2023) https://doi.org/10.2196/50701

[27] Tiwari, S., Jain, A., Sapra, V., Koundal, D., Alenezi, F., Polat, K., Alhudhaif, A., Nour, M.: A smart decision support system to diagnose arrhythmia using ensembled ConvNet and ConvNet-LSTM model. Expert Syst. Appl. **213**, 118933 (2023) https://doi.org/10.1016/j.eswa.2022.118933

[28] Guo, Y., Wang, H., Zhang, H., Liu, T., Liang, Z., Xia, Y., Yan, L., Xing, Y., Shi, H., Li, S., Liu, Y., Liu, F., Feng, M., Chen, Y., Lip, G.Y.H., Lane, D.A., Wang, L., Eckstein, J., Thomas, G.N.: Mobile photoplethysmographic technology to detect atrial fibrillation. J. Am. Coll. Cardiol. **74**(19), 2365–2375 (2019) https://doi.org/10.1016/j.jacc.2019.08.019

[29] Kim, J., Lee, S.J., Ko, B., Lee, M., Lee, Y.S., Lee, K.H.: Identification of atrial fibrillation with single-lead mobile ECG during normal sinus rhythm using deep

learning. J. Korean Med. Sci. **39**(5), 56 (2024) https://doi.org/10.3346/jkms.2024.39.e56

[30] Kilimo, P.J., Otieno, F., Wanjala, P., Wafula, E.M.: Testing MD-Link, a low-cost mobile electrocardiography monitoring device, in patients with irregular heartbeat: Protocol for a cross-sectional study. JMIR Res. Protoc. **8**(1), 10863 (2019) https://doi.org/10.2196/resprot.8762

[31] Lee, J., Reyes, B.A., McManus, D.D., Maitas, O., Chon, K.H.: Atrial fibrillation detection using an iPhone 4S. IEEE Trans. Biomed. Eng. **60**(1), 203–206 (2013) https://doi.org/10.1109/TBME.2012.2208112

[32] Ma, S., Chen, J., Ho, J.W.K.: An edge-device-compatible algorithm for valvular heart diseases screening using phonocardiogram signals with a lightweight convolutional neural network and self-supervised learning. Comput. Methods Programs Biomed. **244**, 107906 (2024) https://doi.org/10.1016/j.cmpb.2023.107906

[33] Tian, Y., Li, Z., Jin, Y., Wang, M., Wei, X., Zhao, L., Liu, Y., Liu, J., Liu, C.: Foundation model of ECG diagnosis: Diagnostics and explanations of any form and rhythm on ECG. Cell Reports Medicine **5**(12), 101875 (2024) https://doi.org/10.1016/j.xcrm.2024.101875

[34] Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamber, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Arcas, B., Webster, D., Corrado, G.S., Matias, Y., Chou, K., Gottweis, J., Tomasev, N., Liu, Y., Rajkomar, A., Barral, J., Semturs, C., Karthikesalingam, A., Natarajan, V.: Large language models encode clinical knowledge. Nature **620**(7972), 172–180 (2023) https://doi.org/10.1038/s41586-023-06291-2

[35] Desai, F., Chowdhury, D., Kaur, R., Peeters, M., Arya, R.C., Wander, G.S., Gill, S.S., Buyya, R.: HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing. Internet of Things **17**, 100485 (2022) https://doi.org/10.1016/j.iot.2021.100485

[36] Lei, N., Kareem, M., Moon, S.K., Ciaccio, E.J., Acharya, U.R., Faust, O.: Hybrid decision support to monitor atrial fibrillation for stroke prevention. Int. J. Environ. Res. Public Health **18**(2), 813 (2021) https://doi.org/10.3390/ijerph18020813