REVIEW | OPEN ACCESS

Medical Science

To Cite:
Długosz KJ, Witkowska A, Piotrowska J, Łubińska-Kowalska A, Domańska A, Witkowska AT, Zapalska BA, Żak-Gontarz A, Minda A, Janikowska J, Wendland M. Current trends in AI-ECG interpretation studies. *Medical Science* 2025; 29: e99ms3628
doi: https://doi.org/10.54905/disssi.v29i160.e99ms3628

Current trends in AI-ECG interpretation studies

Krzysztof Julian Długosz[1*], Adrianna Witkowska[2], Julia Piotrowska[2], Aleksandra Łubińska-Kowalska[2], Adrianna Domańska[2], Antonina Teresa Witkowska[3], Barbara Anna Zapalska[4], Agata Żak-Gontarz[5], Aleksandra Minda[6], Justyna Janikowska[7], Monika Wendland[8]

Authors' Affiliation:
[1]Clinical University Hospital in Olsztyn, Aleja Warszawska 30, 11-041 Olsztyn, Poland
[2]Medical Department, University of Warmia and Mazury, Aleja Warszawska 30,
11-082 Olsztyn, Poland
[3]Praga Hospital of the Transfiguration of the Lord Aleja Solidarności 67, 03-401 Warsaw, Poland
[4]Independent Public Central Clinical Hospital of University Clinical Center of Medical University of Warsaw, Banacha 1A, 02-097 Warsaw, Poland,
[5]Stefan Cardinal Wyszyński Provincial Specialist Hospital SPZOZ in Lublin, Aleja Kraśnicka 100, 20-718 Lublin, Poland
[6]The Infant Jesus Clinical Hospital, Lindleya 4, 02-005 Warsaw, Poland
[7]National Medical Institute of the Ministry of the Interior and Administration in Warsaw, Wołoska 137, 02-507 Warsaw, Poland
[8]Mazovian "Bródnowski" Hospital in Warsaw, Kondratowicza 8, 03-242 Warsaw, Poland

[*]Corresponding author
Krzysztof Julian Długosz - Clinical University Hospital in Olsztyn, Aleja Warszawska 30, 11-041 Olsztyn, Poland
Poland; Email: krzysztofjdlugosz@gmail.com

ORCID List:

Krzysztof Julian Długosz	0009-0000-8134-6115
Adrianna Witkowska	0009-0008-7314-8045
Julia Piotrowska	0009-0006-3261-018X
Aleksandra Łubińska- Kowalska	0009-0007-2699-5965
Adrianna Domańska	0009- 0002-2720-2641
Antonina Teresa Witkowska	0009-0007-9954-1462
Barbara Anna Zapalska	0009- 0004- 6417-877X
Agata Żak- Gontarz	0009-0003-6533-9048
Aleksandra Minda	0009-0004-8862-712X
Justyna Janikowska	0009-0001-8277-0855
Monika Wendland	0009-0009-6894-8846

Peer-Review History
Received: 01 April 2025
Reviewed & Revised: 07/April/2025 to 18/June/2025
Accepted: 25 June 2025
Published: 30 June 2025

Peer-review Method
External peer-review was done through double-blind method.

Medical Science
pISSN 2321–7359; eISSN 2321–7367

© The Author(s) 2025. Open Access. This article is licensed under a Creative Commons Attribution License 4.0 (CC BY 4.0)., which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of this license, visit
http://creativecommons.org/licenses/by/4.0/.

DISCOVERY
SCIENTIFIC SOCIETY

ABSTRACT

It is not an easy task to be continually updated on what are the latest trends and searched areas in the world of medicine. Our review comes here to help by summarizing some recent advancements and main points of interest regarding ECG (electrocardiography) and AI (artificial intelligence) deployment in its interpretation. In our work we want to present the results of our research, expose popular themes that ejected among studies and introduce the reader to selected ML (machine learning) terminology used in the construction of AI algorithms. Moreover, we would like to discuss what might have been the recurring limitations of the reviewed works and speculate about which enhancements may benefit further papers. From 108 open-access articles published between 01-12-2024 and 02-03-2025 we found 41 original research papers. We categorized them into five major categories: AI-ECG in risk stratification, quality of data and preprocessing in AI-ECG, edge devices and telemedicine, AI-ECG algorithms general utilization and unclassified. The modern world is changing quickly, and new technologies like deep learning models will ultimately gain its significance in the art of medicine. Our task as physicians and scientists is to be aware of the recent technological achievements, try to familiarize with them and implement in our work, if it's beneficiary. We hope that our review will be an inspiration for researcher to explore this promising area of modern science.

Keywords: AI, ECG, ECG interpretation, risk stratification, edge devices, quality of data, preprocessing, review

1. INTRODUCTION

AI (artificial intelligence) is a trending topic, especially in the scientific area of interest. AI tools such as CNN (convulational neural network), ViT (vision transformer), RNN (recurrent neural network), LSTM (Long Short-Term Memory), Random Forest and others are involved in an increasing amount of scientific research (Kim & Lee, 2024). This does not exclude medicine. In pharmacology, especially in the drug combination prediction models, AI tools pose a groundbreaking solution. Combining omics (i.e. biological data) and advanced prediction algorithm, AI improves the pace and effectiveness of multi-drug therapy discovery (Chen et al., 2025). Moreover, AI capabilities are increasingly used to analyze and classify complex medical data such as ECG recordings, as reflected by the number of scientific papers focusing on the use of AI in ECG. Recent ESC Guidelines regarding atrial fibrillation management devoted some attention to the use of big data and AI, primarily emphasizing the need to develop this line of research. (Van Gelder et al., 2024). This emphasizes the perspective nature of the issue we raised in the following review and is a signpost to the path of future medicine.

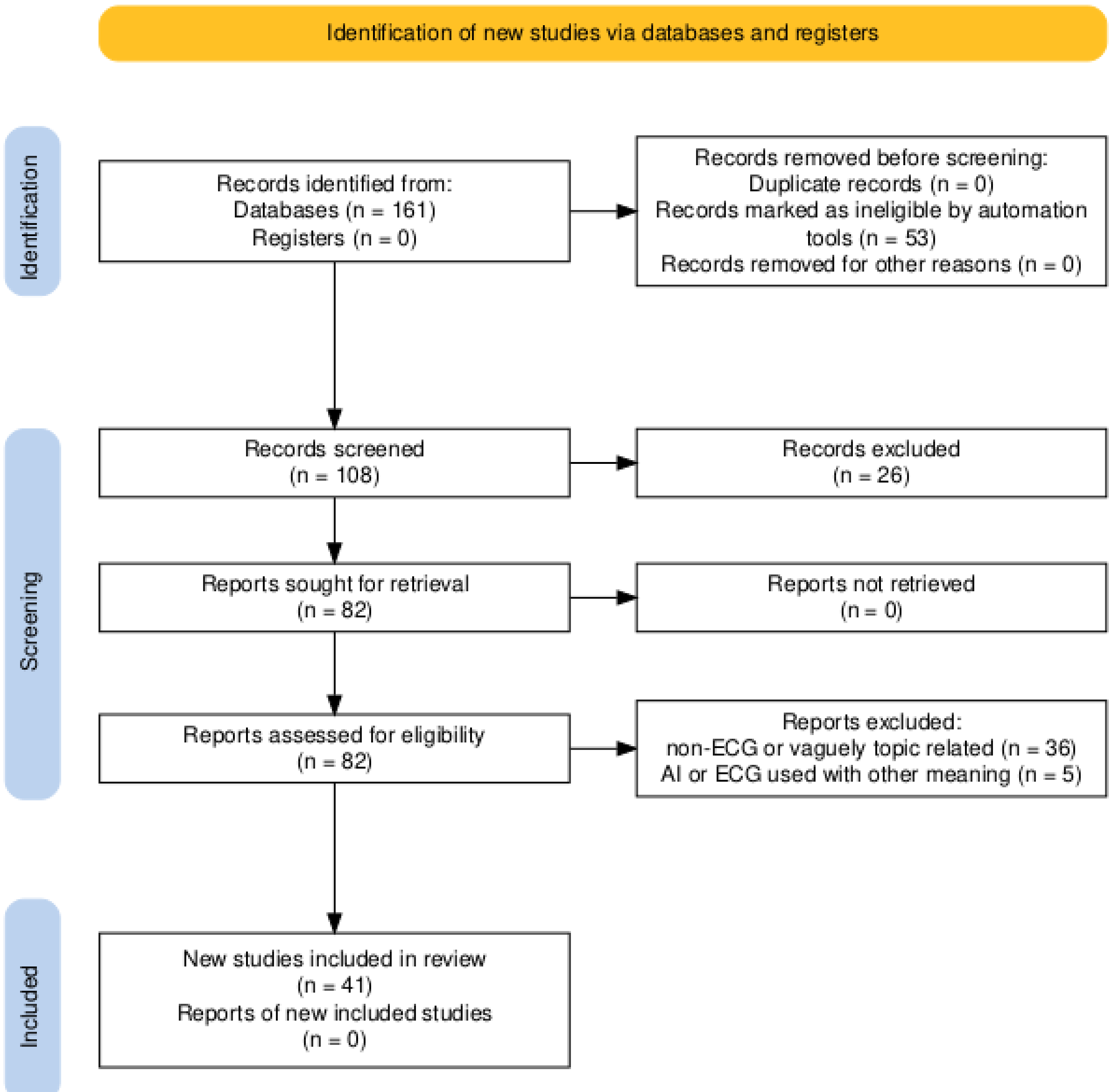

Figure 1. PRISMA flow chart of selected studies

2. METHODOLOGY

The data for this review article were collected using the PUBMED browser. In the search, we used two keyword abbreviations used in the article title: “AI” AND “ECG”. Publication date included articles published between 01-12-2024 and 02-03-2025. We found 108 open-access results with free full text from which we excluded: reviews (n=21), preprints (n=3), clinical trials (n=2) and, due to eligibility criteria for full text, non-ECG or vaguely topic-related studies (n=36), as well as papers, where AI or ECG were used with other meaning (n=5). That way, we ended up with 41 original papers presenting different approaches to AI utilization in ECG analysis and its usage in clinical practice. In our reviewing process, we divided the results into five major categories: AI-ECG in risk stratification (n=25), quality of data and preprocessing in AI-ECG (n=5), edge devices and telemedicine (n=5), AI-ECG algorithms general utilization (n=3) and unclassified (n=3). Methodology of our research is summarized in the provided PRISMA flow chart (Figure 1).

3. RESULTS AND DISCUSSION

3.1 AI-ECG in risk stratification

The most prevalent utilization of artificial intelligence algorithms addressed various risk assessing models in which AI-ECG played a role as a sole risk assessing tool or with a combination of clinical data. We found 23 papers focused on the topic. Most of them touched the problem of atrial fibrillation (7 research studies).

Jabbour et al., (2024) created an AI-ECG model using ResNet 50 (Residual neural network), a complex deep learning architecture operating on visual data combined with an XAI (explainable artificial intelligence) of TensorFlow 2.0 Gradient Type (which highlighted sections of ECG chart which contributed the most to ResNet 50 analysis outcome) to calculate risk of developing atrial fibrillation (AF) in the future among 145 323 patients with no history of AF in the past, have not undergone any cardiac surgery within 30 days of assessment and presented with sinus rhythm ECG. AI-ECG analysis where compared to four different clinical risk scores (of which ultimately selected one was CHARGE-AF score) and polygenic score (PGS). According to the study AI-ECG model outperformed both clinical risk score and PGS scores. It reached an AUC-ROC (area under the receiver operating characteristic curve) of 0.78 compared to accordingly 0.62 and 0.59 AUC-ROC scores for best clinical risk score and PGS. Surprisingly, a combination of AI-ECG, clinical score and PGS did not ameliorated the AUC-ROC (=0.77). Although, it improved "goodness of fit", compromising be-tween explainability and statistic values scores.

An ECG analysis model was created by Jin et al., (2025) on 318 321 patients with normal sinus rhythm (NSR). It served for detecting presymptomatic changes that may predispose to PAF (paroxysmal AF). It assessed 18 clinical parameters and a 12-lead ECG. The authors combined 50-layer 1D-CNN (as one of a default architecture for visual data analysis), residual blocks (to minimize vanishing gradient problem), batch normalization (to reduce internal covariate shift problem), rectified linear activation (ReLU – for enhancing complex non-direct reasoning), stride 2 in CNN (for downsampling), max pooling (for overfitting reduction), concatenation (for combining feature vectors) softmax (to convert probability into decision) and some other AI tools (Table 1). Thanks to layer-wise relevance propagation (LRP), a tool for XAI, the most significant relationship was established between NSTTA (non-specific ST-T segment changes) and PAF. The authors concluded, that despite the limitations to extrapolate the use of the model resulting from the methodology of the study itself, perhaps such a model or a similar one may be used in the future for proactive monitoring of patients with presymptomatic PAF according to the algorithm - it may allow earlier diagnosis of the disease and increase the chances of moderation and possibly accelerate the implementation of anticoagulant treatment based on CHA2DS2-VA score, reducing the risk of ischemic cerebral events.

Table 1. Basic AI and ML terminology grouped by their function.

ECG data	Preprocessing	AI/ML Models	Enhancements	Explainability (XAI)
.dat, .mat	Wavelet	CNN	Batch Normalization	SHAP
images	Pan-Tompkins	ResNet	ResNet	LIME
.hea	Graphs	RF	Max Pooling	Grad-CAM
.csv	STNF	Transformers	Dropout	Int. Gradients
.xml		ViT		
		RNN, LSTM		

ECG- electrocardiogram, STNF- short-time Fourier transform, CNN – convolutional neural network, RF – Random Forest, ViT- vision transformer, RNN – recurrent neural network, LSTM – Long Short-Term Memory, ResNet residual neural network, XAI – explainable AI, SHAP - Shapley Additive Explanations, LiME - Local Interpretable Model-agnostic Explanations), Grad-CAM - Gradient-weighted Class Activation Mapping

Another study regarding AF was published by Khan et al., (2025) 911 patients with implantable cardiomonitor (ICM) and symptomatic heart failure were searched for incidence of AF. Input data consisted of two-dimensional (2D) ECG charts and clinical metadata, which were analyzed by a 6-layer CNN. The researchers in their work assumed that the risk of AF calculated by the algorithm at over 90% is a true positive result, indicating very high confidence in the AI-ECG model.

Yao et al., (2024) published a study, where the researchers involved four features to base their prognostic tool (in a population of AVIC study) for AF early detection, which were three non-AI related criteria of: clinical risk scores like CHARGE-AF, polygenic risk score, and NT-proBNP and FGF-23 proteins level measured, and an AI-ECG charts analyzed by 1D-CNN algorithm which assessed AF risk using regularized logistic regression model. Authors stated that combining these four factors in risk assessment improved the accuracy and recall in the prognostic model. They concluded that it presents a promising idea for further studies in the area of primary prophylaxis and personalized medicine.

An interesting approach was presented by Lin et al., (2025) who developed a knowledge-embedded (i.e. based on scientific consensuses) multimodal (i.e. using various methodology) pseudosiamese (i.e. combining two complementary neural networks in one) model for detecting AF. Raw ECG signals (from Physionet databases: MIT - BIH AF dataset, CinC 2017 and Chapman dataset) ready for interpretation by 1D-CNN were used as input da-ta. It was analyzed by 1D-CNN ResNet_Wang and transformed into GAF (Gramian Angular Field) images by the GASF (Gramian Angular Summation Fields) algorithm interpreted by 2D-CNN, which was a 2D-CNN ResNet 18. Despite the apparent secondary nature of such a solution, it combined both models' advantages: the one-dimensional (temporal) model focuses more on the interpretation of data in the context of temporal information, ergo, the rhythm itself, while the two-dimensional (spatial) model allows for more focus on the interpretation of the image, morphology and interpoint relations. However, the results were not similar between the analyzed databases, which underscores the significance of preprocessing and data standardization.

Tao et al., (2024) in their retrospective study, utilized AI to try to identify left atrium low-voltage areas (LA-LVA) in 1133 patients who underwent radiofrequency ablation and compared sole CNN-ECG model, clinical data risk scores (DR-FLASH and APPLE) and a combination of both (in a CNN-RF multimodal model). The authors found the combined analysis to reach the highest degree of accuracy. Gradient-weighted Class Activation Map-ping (Grad-CAM) served as an XAI for highlighting regions of ECG which were essential for model interpretation outcome.

An original approach demonstrated that She et al., (2025) tried to construct an AI-ECG model for assessing the risk of reoccurrence of AF among postablation patients based on their ECG charts performed after ablation. They addressed the problem of how to gain clinicians trust toward AI analysis outcomes. After theoretical consultations with medical professionals, the authors chose a model where manually extracted features from 503 postablation patients' ECG charts were presented to a Cox regression AI model with SHAP (Shapley Additive Explanations) XAI.

Although the model did not reach high AUC-ROC and sensitivity, the discussion contained fascinating section engaging potential users of such software. Cardiologists, who provided extensive feedback, constituted the basis for refining the model in future studies. They underscored an urging need for AI cause and effect reasoning, which would gain more confidence in the use of AI analyses in clinical work by practicing physicians. The paper poses an insight into professionals' perspective and expectations toward further AI studies in medicine and its implementation in daily practice.

Some other authors decided to employ AI-ECG architectures for assessing cardiovascular risks. In a study conducted by Oikonomou et al., (2025), 1550 patients with a diagnosis of breast cancer or non-Hodgkin lymphoma, previously without cardiomyopathy, before initiating anticancer therapy with anthracyclines or trastuzumab, were classified into risk groups of developing cardiac dysfunction in the course of the treatment. 12-lead ECG were performed before initiating treatment, converted to 300 by 300 pixel images. Then analyzed by EfficientNet B3 CNN model with an XAI (Gradient-weighted Class Activation method). The algorithm assessed risk of early chemotherapy-related myocardial dysfunction with high accuracy.

In another study regarding cardio-oncology and AI-ECG, Ayoub et al., (2024) used Tensor-Flow to find a suitable AI to evaluate risk of MACE (major adverse cardiovascular events) and myocarditis. They analyzed data from 2258 cancer patients before receiving immune checkpoint inhibitor therapy. They decided to implement a 5-blocked 1D-CNN with a ReLU activation and a SHAP XAI. The results concluded that more valuable for risk outcome were 23 descriptive data attached to the analysis than AI-ECG model.

Another popular topic among studies conducted in recent time was estimating biological age or gender by evaluating ECG by various AI algorithms (Liu et al., 2025; Hempel et al., 2025; Singh et al., 2025; Sau et al., 2025; Adel et al., 2024). Using different AI tools, the researchers investigated whether the AI-ECG-based biological age or gender discrepancy adds new predictive value that clinicians can incorporate into existing clinical risk scores. Unfortunately, among the five studies listed, only two used XAI (variational autoencoder with CNN and Integrated Gradients with a post hoc attribution method).

Apart from AF, Cardiooncology and Demographics, AI in ECG interpretation was also employed in different topics, some being less popular, but clinically significant. Latest papers have also addressed such issues as cardiac amyloidosis, Brugada syndrome, pulmonary hypertension, progression of heart failure, hypertension, cardiac wall motion abnormalities, cardiovascular risk in athletes and overall mortality (Amadio et al., 2025; Randazzo et al., 2025; Kishikawa et al., 2025; Kim et al., 2025; Sau et al., 2025; Rogers et al., 2025; Nechita et al., 2025; Sau et al., 2024). Populations in these studies varied from 210 to 189,539 patients. The most prevalent AI architecture used was CNN models. However, some of researchers decided to experiment with the AI used in their studies and implemented solutions such as Vit (vision transformer) or a random forest classifier and a complex multimodal AI-ECG (1D-CNN + transformer + triplet loss). Two of them utilized XAI (variational autoencoder and SHAP). In all of the works, AI-ECG analysis played as a promising statistic biomarker, which can ameliorate the prognostic algorithms alone or in combination of clinical data.

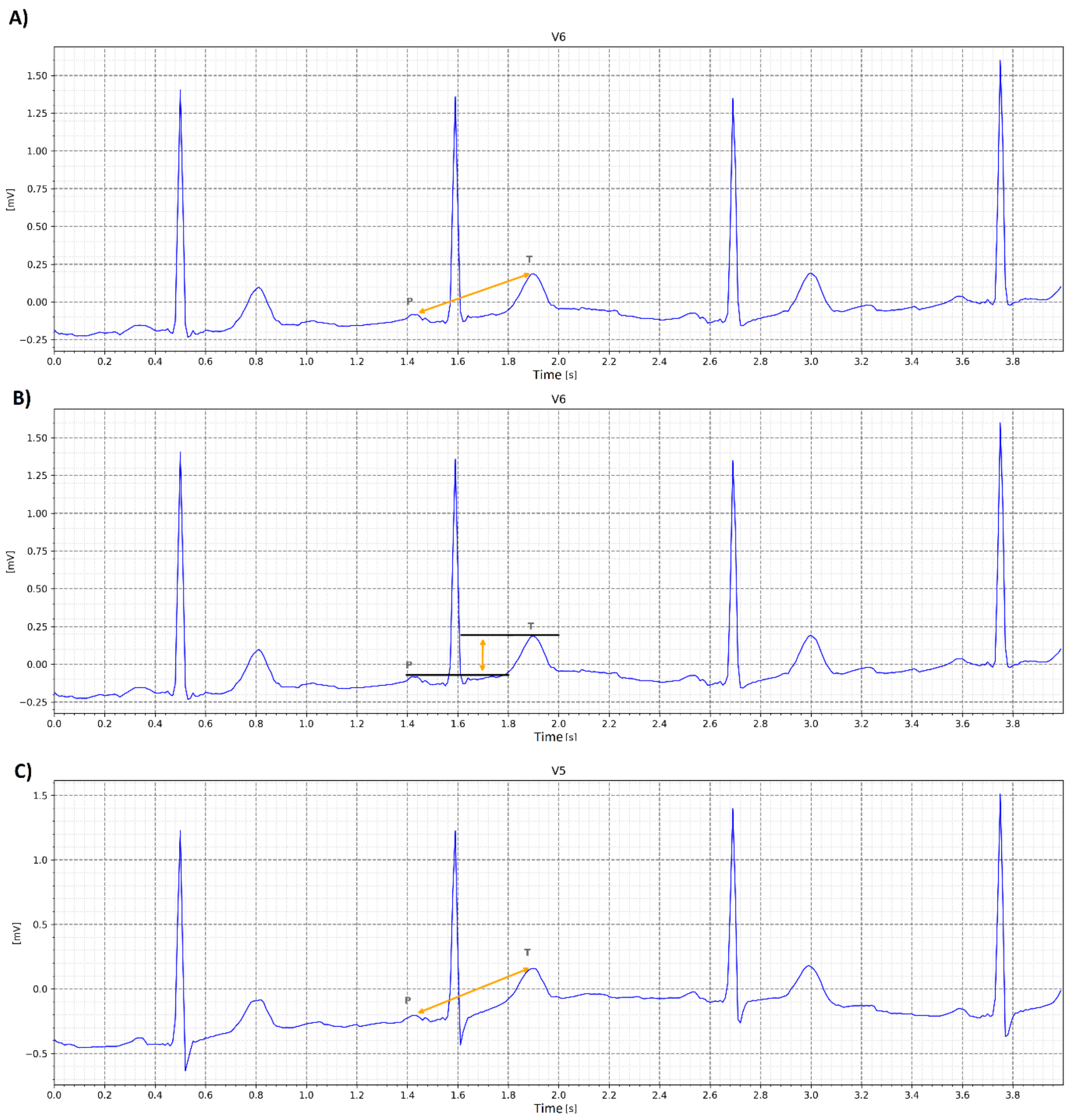

Figure. 2 – Top 3 features in Yeh. et al, study for CAD risk prediction in both sexes (Yeh et al., 2024); A) P peak to T peak slope in lead V6; B) P peak to T peak amplitude difference in lead V6; C) P peak to T peak slope in lead V5

A different problem occurred in another study. Butler et al. compared two AI-ECG models' accuracy on data from around 50 000 patients for fatal coronary heart disease risk prediction (Butler et al., 2024). Surprisingly, their work concluded that a ResNet 1D-CNN model reached a higher AUC when analyzing only lead I ECG, as opposed to a 12-lead ECG. This conclusion gives a fascinating insight into utilization of AI into biometrical analysis. Moreover, it underscores the recurring problem in deep learning of data noise overload.

In a study group from Taiwan, Yeh et al., (2024) investigated how well an AI-ECG tree-based model of XGBoost (eXtreme Gradient Boosting) with SHAP XAI performs in coronary artery disease (CAD) risk assessment. Scientists preprocessed approximately 5000 12-lead ECGs in raw .xml signal format. After removing noise, drift, and marking fiducial points (i.e., P, Q, R, S, T), the researchers extracted the data into a numerical form (using wavelet transformation), which could be analyzed by XGBoost. AI-enhanced ECG assessed CAD risk with 61-66% accuracy and 79-83% sensitivity. The authors speculated that AI-ECG may develop as an alternative for myocardial perfusion scynthygraphy with thalium 201. The paper presented instances where the AI-ECG model predicted a positive coronary angiography (CAG) result preceded by a negative scynthygraph. Interestingly, the algorithm identified the most significant features out of 561 extracted ECG elements for AI-ECG CAD risk stratification (Figure 2). Moreover, such characteristics have never occurred in ECG traditional interpretation guidelines before.

3.2 Preprocessing tools

An interesting section in the results of our research presents works regarding preprocessing and data quality. Chen et al., (2024) presented an unconventional, yet very innovative approach. The researchers ameliorated CNN performance of ECG interpretation by incorporating clinical data into ECG record via ECG coloring technology. Imposing demographic information onto patients' ECG chart resulted in a substantial rise in the AUC result. It detect-ed AF in ECG by 7,6% better compared to the AI-ECG model using only the original signal value.

A study group from Korea focused on the very fundamental aspect of the ECG data collection. They proposed a preprocessing tool for graph structure-based data augmentation, which reflects the concept of three-dimensional data of the heart's electric potential (Kim & Lee, 2024). The researchers implemented their idea into ResNet and DenseNet models of ECG interpretation. As a result, it improved the F1 score by 1,44%.

Kwon et al., (2025) investigated yet another new method of presenting data to the AI deep learning models. By preprocessing ECG signals with polar transformations of short-time Fourier transform (STNF) spectrograms the authors could present preprocessed 30-second ECG charts acquired from open-source PhysioNet online database to the CNN (MobileNet, ResNet, DenseNet) prediction model of AF. The algorithm reached comparable to previous works results, but provided longer ECG chart duration analysis. Thanks to the above, such developed models may be a better choice regarding the nature of diagnosing supraventricular arrhythmia.

A promising insight introduced researchers from Israel and Japan, who developed an ECG analysis model that allows for the interpretation of even poor quality ECG images (taken with a smartphone, with shadows, and with a random background) - which is supposed to improve its usability in clinical practice (Gliner et al., 2025).

A completely different approach presented a study group of Galanty et al., (2024) who, using the BEAMRADS protocol ("Bias Evaluation And Monitoring for Transparent And Reliable Medical Datasets"), assessed the quality of databases available on platforms such as PhysioNet (ECG) and Grand Challenge (MRI images and Color Fundus Photography). The authors emphasized that the published databases contained specific errors (in varying percentages) that could affect the incorrect training (bias) of AI algorithms on these datasets. The most common errors found are listed in Table 2. The authors underscored the urgent necessity for more strict quality checks by data collecting organizations.

Table 2. Most common errors recurring in the open-source medical databases such as PhysioNet and Grand Challenge

Identified databases' common quality issues
1) Annotations problem – rarely reported percentages of misdiagnosis in supplied annotations
2) Shortage of transparent inclusion criteria of patients
3) Lack of demographic metadata of the cohorts – limits extrapolation to diverse populations
4) Insufficient data regarding the collection process and preprocessing tools used
5) No descriptions of dataset limitations – scarce data on underrepresented patients
6) Inadequate or inconsistent implementation of existing data collection standards such as BIAS or datasheets for datasets

3.3. Algorithms competing with humans

An extensive study by Johnson et al., (2025) compared the performance of AI to 167 ECG technicians in screening 10-14 day ECG Holter. The study group consisted of 14,596 patients with the most common monitoring indications of palpitations, syncope, dizziness, and paroxysmal AF suspicion. The aim of the screening was to detect critically important ECGs for physician review. The authors designed the model (DeepRythmAI for Autonomous Analysis of Rhythm Investigation, ie, "DRAI MARTINI") in cooperation with Medicalgorithmics®, a Polish software company. AI was less likely to miss diseased ECGs but more likely to report pathology in healthy participants compared to the performance of qualified ECG technicians. F1 was similar in both groups. The researchers suggest that their tool can safely replace technicians, which may herald a spectacular breakthrough in the field of AI-ECG study.

Other authors decided to assess the ECG interpreting capabilities of AI-LLM (Large Language Model) like popular ChatGPT and BioMed GPT-LM-7B (Çamkıran et al., 2025; Yang et al., 2025). ChatGPT-based algorithms achieved much worse results (accuracy ranging between 57-62%) when interpreting 107 ECG charts. In comparison, two cardiologist professionals reached an accuracy of 92,52%. Also BioMed GPT-LM-7B with ResNet-18 CNN preprocessing model, reached an unsatisfactory F1 (0.5). It may be attributed to a prevalent and urgent problem in AI-LLM models, known as "hallucination" (i.e., creating made-up facts and answers), which could be a considerable limitation in deploying such solutions in further studies.

3.4. Digital wearable health technologies and telemedicine

Among reviewed studies, a significant portion of works regarding edge devices (ED) and edge intelligence technology drew our attention. Most of them were sponsored, albeit they presented a different perspective from the rest of the articles. To gain a comprehensive understanding of recent study trends, it is also beneficial to take a look at this area of study. A substantial portion of researchers noticed an increasing need to implement more complex AI solutions into portable devices assessing ECG charts as they emerged recently as a possible first-line screening tool in detecting arrhythmia (especially early AF detection). These trends created a need for better hardware in the limited physical space of an edge device. To solve this problem, An et al., (2024) studied a concept of Knowledge Distillation, i.e. adjusting edge devices AI to big model (as a teacher-student relation). They tried to compress algorithms to detect 11 different arrhythmia in an ECG chart. In this study, the authors trained a device imitating an ED (technical specifications: STM32F429 microcontroller characterized by ARM Cortex-M4 core, 2 MB of flash, a maximum clock frequency of 180 MHz and 256 kB RAM) and reached 96% accuracy, whereas "the teacher" AI model (big ResNet, LSTM and SENet models) reached 97% accuracy. Second work implemented a solution of connecting edge device of Kardia Mobile to a commercially developed end-to-end Willem AI to ameliorate AF detection, reaching above 96% accuracy (Guio et al., 2025).

Other studies regarding ED and AI-ECG, utilized various devices for miscellaneous tasks. Disrud et al., (2025) developed a 12-lead ECG monitoring device as a telehealth diagnostic tool in cooperation with patients, but met with a lot of technical limitations. Blok et al., (2025) studied ED smart band with photoplethysmography technology's ability to detect early AF. Smiley and Finkelstein (2025) using a sample of only 27 patients, developed an AI model to assess the current level of physical exertion. It used a single-lead ECG measured from a device worn on the chest and a pulse oximeter on the wrist. Finnish Kubios Heart Rate Variability software (HRV) and LSTM model processed the raw ECG data to reach F1 of 91,7% when compared to patients' Borg scale score. Another approach was presented an Italian study group. The researchers investigated the possibility of a non-invasive screening for the risk of type 1 diabetes based on single-lead ECG charts assembled with a wearable device (n=27 patients) (Gragnaniello et al., 2025). Microprocessor converted the signal into spectrograms for analysis by 1D-CNN. The model reached F1 score of 0.9.

3.5. Miscellaneous studies

Three studies could not be categorized into other subgroups, so we decided to enclose them separately. Maleki Lonbar et al., (2024) developed a model for biometric recognition of a person based on the ECG signal. For noise reduction they used Pan-Tompkin's algorithm and Wigner-Ville distribution to convert signals into image. An advanced CNN model of GoogLeNet analyzed the data and reached almost 100% sensitivity and specificity. Choudhury et al., (2025) studied single-lead ECG data correlation with sleep apnea episodes using three databases with several-hour ECG recordings from 113 patients. After data preprocessing, the CWT (continuous wavelet transform) model created standardized scalograms with compressed visual information from multi-hour ECG recordings and subjected them to CNN analysis (GoogLeNet). The LIME model (Local Interpretable Model-agnostic Explanations) served as an XAI. Some correlations between apnea episodes and ECG events like slowing of the rhythm, QT prolongation, cases of ST depression and cases of provoking an AF episode were found. It provided a fascinating insight into pathophysiology in the study.

Huang et al., (2025) presented another interesting work. The AI-ECG model of Cat-Boost, analyzed ECG data after Philips DXL algorithm preprocessing. It predicted Left Ventricle Hypertrophy better than popular ECG criteria like Sokolov-Lyon, Cornell product, Peguero-Lo Presti or Framingham criterion, achieving AUC of 0,795 in the validation sample (n=8403). The algorithm highlighted some

ECG features that were previously rarely reported in traditional ECG criteria. An example is the peak-to-peak QRS complex amplitude in aVF lead. These findings constitute an interesting point for further studies.

Although some authors, especially when it comes to the models calculating the probability of the disease based on the ECG, tend to overestimate the significance of their analysis in their article's discussion section, it is worth mentioning that the features that their algorithms extracted from the raw ECG signals or visual data are a fascinating field to base future studies upon.

In our opinion many publications focused more on exhibiting sole statistical correlation rather than giving a pathophysiological explanation on why their AI models made the stratification decisions that they did. An interesting tool that was used to mitigate this effect in some of the studies were LIME (Local Interpretable Model-agnostic Explanations) and SHAP. They are a software methods (ie. an XAI) created to help understand how complex AI models make their decisions. They underscored which features are the most important in the analysis. It is a promising step towards making current AI deployment studies more transparent and scientific, thereby earning more trust among clinicians and academicians in such solutions. A fascinating and deep insight into this matter is made in one Japanese study about CNN work, whose task was to rank risk in postablation AF patients, where a significant section is dedicated to covering clinicians' feedback on how they assess the results of the study, do they trust the algorithm and how prone are they to implement such an augmentation of risk stratification clinical assessment in their future work.

A big part of the participants expressed that although results seem to be promising, they lack the conclusive evidence-based argument why clinicians should give credence to these tools as they support their decisions with vague logic reasoning. Another noteworthy study addressed the issue of maintaining quality in databases used to train AI tools in medicine. In our opinion too little credit is being given to how the data are collected. An urging issue is to create direct guidelines on how to construct such databases. Open-source websites that currently dominate this area should consider revising their submission policies to help mitigate this problem.

4. CONCLUSIONS

The modern world is changing rapidly, and new technologies, such as deep learning models, will ultimately gain significance in the field of medicine. Our task as physicians and scientists is to be aware of recent technological achievements, familiarize ourselves with them, and implement them in our work if beneficial. We hope that our review will be an inspiration for researchers to explore this promising area of modern science.

Author's Contributions

Krzysztof Julian Długosz: Conceptualization, coordinating, writing – main draft and finishing

Julia Piotrowska, Antonina Teresa Witkowska, Barbara Anna Zapalska, Aleksandra Łubińska- Kowalska, Adrianna Domańska, Agata Żak- Gontarz, Aleksandra Minda, Justyna Janikowska, Monika Wendland: writing – data selection, rough preparation

Adrianna Witkowska: writing – review, editing

All authors contributed in the preparation of the final manuscript.

Acknowledgments

Figure captions

Figure 1. Visualization created using PRISMA 2020 Flow Diagram online open-source application [https://estech.shinyapps.io/prisma_flowdiagram/] (Haddaway et al., 2021).

Figure 2. Visualization created using PTB-XL, publicly available electrocardiography dataset, available via PhysioNet and processed with Python (Goldberger et al., 2000; Wagner et al., 2020; Wagner et al., 2022).

Informed consent

Not applicable.

Ethical approval

Not applicable.

Funding
This study has not received any external funding.

Conflict of interest
The authors declare that there is no conflict of interest.

Data and materials availability
All data sets collected during this study are available upon reasonable request from the corresponding author.

REFERENCES

1. Adel FW, Sang P, Walsh C, Rothschild E, Marjanovic M, Majeed M, Fernandes R, Thibault H, Ahmed A, Zahir A, Patel M, Aggarwal S. Artificial intelligence evaluation of electrocardiographic characteristics and interval changes in transgender patients on gender-affirming hormone therapy. Eur Heart J Digit Health. 2024;6(1):55–62. doi:10.1093/ehjdh/ztae076
2. Amadio JM, Grogan M, Muchtar E, Uy-Evanado A, Rao R, Dispenzieri A, Gertz MA, Groarke JD, Kristen AV, Nativi-Nicolau J, Rapezzi C. Predictors of mortality by an artificial intelligence enhanced electrocardiogram model for cardiac amyloidosis. ESC Heart Fail. 2025;12(1):677–682. doi:10.1002/ehf2.15061
3. An X, Shi S, Wang Q, Yu Y, Liu Q. Research on a Lightweight Arrhythmia Classification Model Based on Knowledge Distillation for Wearable Single-Lead ECG Monitoring Systems. Sensors. 2024;24(24):7896. doi:10.3390/s24247896
4. Ayoub C, Appari L, Pereyra M, Farina JM, Chao CJ, Scalia IG, Mahmoud AK, Abbas MT, Baba NA, Jeong J, Lester SJ, Patel BN, Arsanjani R, Banerjee I. Multimodal Fusion Artifi-cial Intelligence Model to Predict Risk for MACE and Myocarditis in Cancer Patients Re-ceiving Immune Checkpoint Inhibitor Therapy. JACC Adv. 2024;4(1):101435. doi:10.1016/j.jacadv.2024.101435
5. Blok S, Gielen W, Piek MA, de Groot JR, Schalij MJ, Doevendans PAJM. Performance of a medical smartband with photoplethysmography technology and artificial intelligence algo-rithm to detect atrial fibrillation. mHealth. 2025;11:10–15. doi:10.21037/mhealth-24-10
6. Butler L, Ivanov A, Celik T, Ando K, Wang F, Healy L, Krummen DE. Time-Dependent ECG-AI Prediction of Fatal Coronary Heart Disease: A Retrospective Study. J Cardiovasc Dev Dis. 2024;11(12):395. doi:10.3390/jcdd11120395
7. Çamkıran V, Tunç H, Achmar B, Ürker TS, Kutlu İ, Torun A. Artificial intelligence (ChatGPT) ready to evaluate ECG in real life? Not yet! Digit Health. 2025;11:20552076251325279. doi:10.1177/20552076251325279
8. Chen J, Lin A, Jiang A, Qi C, Liu Z, Cheng Q, Yuan S, Luo P. Computational frameworks transform antagonism to synergy in optimizing combination therapies. NPJ Digit Med. 2025;8(1):44. doi:10.1038/s41746-025-01435-2
9. Chen WW, Liu CM, Tseng CC, Lee CC, Lin YT, Huang CT, Tsao CW, Chiang CH. Identify-ing the presence of atrial fibrillation during sinus rhythm using a dual-input mixed neural network with ECG coloring technology. BMC Med Res Methodol. 2024;24(1):318. doi:10.1186/s12874-024-02421-0
10. Choudhury M, Tanvir M, Yousuf MA, Islam N, Uddin MZ. Explainable AI-driven scalogram analysis and optimized transfer learning for sleep apnea detection with single-lead electro-cardiograms. Comput Biol Med. 2025;187:109769. doi:10.1016/j.compbiomed.2025.109769
11. Disrud LW, Swain WH, Davison H, Lewis GF, Smith SA, Maguire P. A Pilot Study of the Home-Based 12-Lead Electrocardiogram in Clinical Practice. Mayo Clin Proc Innov Qual Outcomes. 2025;9(2):100598. doi:10.1016/j.mayocpiqo.2025.100598
12. Galanty M, Luitse D, Noteboom SH, Juric A, van der Schaaf M, Wisselink W, van Dijk M. Assessing the documentation of publicly available medical image and signal datasets and their impact on bias using the BEAMRAD tool. Sci Rep. 2024;14(1):31846. doi:10.1038/s41598-024-83218-5
13. Gliner V, Levy I, Tsutsui K, Neufeld S, Preukschas AG, Blazeck J. Clinically meaningful in-terpretability of an AI model for ECG classification. NPJ Digit Med. 2025;8(1):1–15. doi:10.1038/s41746-025-01467-8
14. Gragnaniello M, Marrazzo VR, Borghese A, Maresca L, Breglio G, Riccio M. Edge-AI Ena-bled Wearable Device for Non-Invasive Type 1 Diabetes Detection Using ECG Signals. Bio-engineering. 2025;12(1):4. doi:10.3390/bioengineering12010004
15. Guio FD, Rienstra M, Lillo-Castellano JM, Kirchhof P, Lip GYH, McMurray JJ. Enhanced detection of atrial fibrillation in single-lead electrocardiograms using a Cloud-based artifi-cial intelligence platform. Heart Rhythm. 2025;0(0)\:S1–S10. doi:10.1016/j.hrthm.2024.12.048
16. Hempel P, Ribeiro AH, Vollmer M, Bender T, Dörr M, Krefting D, Spicher N, Dorr F, Frantz S, Zink F, Bittner FL,

Katzenberger T, Neumann JT, Thiele H, Böhm M, Werdan K, Westen-feld R, Dörge H, Rollmann M, Heider TM. Explainable AI associates ECG aging effects with increased cardiovascular risk in a longitudinal population study. NPJ Digit Med. 2025;8(1):25:1–11. doi:10.1038/s41746-024-01428-7

17. Huang JT, Tseng CH, Huang WM, Chen YJ, Lin YS, Chang WC. Comparison of machine learning and conventional criteria in detecting left ventricular hypertrophy and prognosis with electrocardiography. Eur Heart J Digit Health. 2025;6(2):252–260. doi:10.1093/ehjdh/ztaf003
18. Jabbour G, Nolin-Lapalme A, Tastet O, Corbin D, Jorda P, Sowa A, Delfrate J, Busseuil D, Hussin JG, Dube M-P, Tardif J-C, Rivard L, Macle L, Cardin-Tourigny J, Khairy P, Avram R, Tadros R. Prediction of incident atrial fibrillation using deep learning, clinical models, and polygenic scores. Eur Heart J. 2024;45(46):4920-4934. doi:10.1093/eurheartj/ehae595
19. Jin Y, Ko B, Chang W, Choi KH, Lee KH. Explainable paroxysmal atrial fibrillation diagno-sis using an artificial intelligence-enabled electrocardiogram. Korean J Intern Med. 2025;40(2):251-261. doi:10.3904/kjim.2024.130
20. Johnson LS, Zadrozniak P, Jasina G, Wang J, Smith GC, Parikh V, Raisaro JL, Hoppmann P. Artificial intelligence for direct-to-physician reporting of ambulatory electrocardiography. Nat Med. 2025;31(3):925–931. doi:10.1038/s41591-025-03516-x
21. Khan MS, Zile MR, Kahwash R, Sarkar S, Van Dorn B, Koehler J, Franco N, Gerritse B, Butler J. Incidence of atrial fibrillation in patients with an insertable cardiac monitor and symptomatic heart failure. ESC Heart Failure. 2025;12(3):1693-1702. doi:10.1002/ehf2.15180
22. Kim DY, Lee SW, Lee DH, Park JH, Choi Y, Kim JH, Lee HS, Hwang J. Electrocardiog-raphy-based artificial intelligence predicts the upcoming future of heart failure with mildly reduced ejection fraction. Front Cardiovasc Med. 2025;12:1418914. doi:10.3389/fcvm.2025.1418914
23. Kim KG, Lee BT. Graph structure based data augmentation method. Biomed Eng Lett. 2024;15(2):283–289. doi:10.1007/s13534-024-00446-4
24. Kishikawa R, Kodera S, Setoguchi N, Minamikawa T, Tanaka R, Yanagawa Y, Inoue S, Oka K, Hiramatsu O, Takahashi H, Ikeda S. An ensemble learning model for detection of pulmonary hypertension using electrocardiogram, chest X ray, and brain natriuretic peptide. Eur Heart J Digit Health. 2025;6(2):209–217. doi:10.1093/ehjdh/ztae097
25. Kwon D, Kang H, Lee D, Kim YC, Cho YH, Park SY, Shin S. Deep learning-based predic-tion of atrial fibrillation from polar transformed time-frequency electrocardiogram. PLOS One. 2025;20(3)\:e0317630. doi:10.1371/journal.pone.0317630
26. Lin C, Lu H, Sang P, Pan C. A knowledge embedded multimodal pseudo-siamese model for atrial fibrillation detection. Sci Rep. 2025;15(1):3133. doi:10.1038/s41598-025-87115-3
27. Liu CM, Kuo MJ, Kuo CY, Phan TG, Muller-Edenborn B, Johnson AEW, Cushman M, Wagner P, Lessner SM, Jaffe AS, Sanchez O, Weng W, Hill NR, Attia ZI, Froelicher VF, Ti-son GH, Wilson FP, Tait N, Akinbami LJ, Basu S. Reclassification of the conventional risk assessment for aging-related diseases by electrocardiogram-enabled biological age. npj Ag-ing. 2025;11(1):1–7. doi:10.1038/s41514-025-00198-0
28. Maleki Lonbar S, Beigi A, Bagheri N, Peris-Lopez P, Camara C. Deep learning based bio-metric authentication system using a high temporal/frequency resolution transform. Front Digit Health. 2024;6:1463713. doi:10.3389/fdgth.2024.1463713
29. Nechita LC, Tutunaru D, Nechita A, Benea DI, Sandulescu M, Ion S, Ursaru A, Pantea M, Dan GA, Piros S. A Resting ECG Screening Protocol Improved with Artificial Intelligence for the Early Detection of Cardiovascular Risk in Athletes. Diagnostics. 2025;15(4):477. doi:10.3390/diagnostics15040477
30. Oikonomou EK, Sangha V, Dhingra LS, Aminorroaya A, Coppi A, Krumholz HM, Baldas-sarre LA, Khera R. Artificial Intelligence–Enhanced Risk Stratification of Cancer Therapeu-tics–Related Cardiac Dysfunction Using Electrocardiographic Images. Circ Cardiovasc Qual Outcomes. 2025;18(1):e011504. doi:10.1161/CIRCOUTCOMES.124.011504
31. Randazzo V, Caligari S, Pasero E, Lanza GA, Pieroni N, Prota C, Pezzuto B, Crea F. A Vi-sion Transformer Model for the Prediction of Fatal Arrhythmic Events in Patients with Bru-gada Syndrome. Sensors. 2025;25(3):824. doi:10.3390/s25030824
32. Rogers AJ, Bhatia NK, Bandyopadhyay S, Grosjean J, Zemedikun DT, Prats-Galino A, Petri E, Kowey PR, Guldenpfennig LJ, Tafti D, Poon S. Identification of cardiac wall motion ab-normalities in diverse populations by deep learning of the electrocardiogram. NPJ Digit Med. 2025;8(1):1–10. doi:10.1038/s41746-024-01407-y
33. Sau A, Barker J, Pastika L, Zeidabadinejad B, Sieliwonczyk E, Patlatzoglou K, Melville D, Huber S, Roy S. Artificial Intelligence–Enhanced Electrocardiography for Prediction of In-cident Hypertension. JAMA Cardiol. 2025;10(3):214–223. doi:10.1001/jamacardio.2024.4796
34. Sau A, Sieliwonczyk E, Patlatzoglou K, Zeidabadinejad B, Barker J, Pastika L, Khurana RK, Betancur-balanzo M, Egiziano MR, Wang Y, Panchal B, Jha R, Mehta H. Artificial intelli-gence-enhanced electro-cardiography for the identification of a sex-related cardiovascular risk continuum:

a retrospective cohort study. Lancet Digit Health. 2025;7(3)\:e184–e194. doi:10.1016/j.landig.2024.12.003

35. Sau A, Zeidabadinejad B, Patlatzoglou K, Barker J, Pastika L, Sieliwonczyk E, Roy S, Mel-ville D. A comparison of artificial intelligence-enhanced electrocardiography approaches for the prediction of time to mortality using electrocardiogram images. Eur Heart J Digit Health. 2024;6(2):180–189. doi:10.1093/ehjdh/ztae090
36. She WJ, Siriaraya P, Iwakoshi H, Kuwahara N, Senoo K. An Explainable AI Application (AF'fective) to Support Monitoring of Patients With Atrial Fibrillation After Catheter Abla-tion: Qualitative Focus Group, Design Session, and Interview Study. JMIR Human Factors. 2025;12(1):e65923. doi:10.2196/65923
37. Singh M, Friedman PA, Gulati R, Masoudi FA, Thomas LE, Reza N, O'Connor CM, Peter-son ED. Role of Biological Age in the Determination of Long-Term Cause-Specific Death Following Percutaneous Coronary Interventions. J Am Heart Assoc. 2025;14(5)\:e036876. doi:10.1161/JAHA.124.036876
38. Smiley A, Finkelstein J. Dynamic Prediction of Physical Exertion: Leveraging AI Models and Wearable Sensor Data During Cycling Exercise. Diagnostics. 2025;15(1):52. doi:10.3390/diagnostics15010052
39. Tao Y, Zhang D, Pang N, Geng S, Tan C, Tian Y, Hong S, Liu XP. Multi-modal artificial in-telligence algorithm for the prediction of left atrial low-voltage areas in atrial fibrillation patient based on sinus rhythm electrocardiogram and clinical characteristics: a retrospective, multicentre study. Eur Heart J Digit Health. 2024;6(2):200-208. doi:10.1093/ehjdh/ztae095
40. Van Gelder IC, Rienstra M, Bunting KV, Casado Arroyo R, Caso V, Crijns HJGM, De Potter TJR, Dwight J, Guasti L, Hanke T, Jaarsma T, Lettino M, Løchen M L, Lum-bers RT, Maesen B, Mølgaard I, Rosano GMC, Sanders P, Schnabel RB, Suwalski P, Svennberg E, Tamargo J, Tica O, Traykov V, Tzeis S, Kotecha D. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): Developed by the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC), with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Endorsed by the European Stroke Organisation (ESO). Eur Heart J. 2024;45(36):3314-3414. doi:10.1093/eurheartj/ehae176
41. Yang K, Hong M, Zhang J, Li W, Chen X, Zhou Y. ECG-LM: Understanding Electrocardio-gram with a Large Language Model. Health Data Sci. 2025;5:0221. doi:10.34133/hds.0221
42. Yao Y, Zhang MJ, Wang W, Zuang Z, He R, Ji Y, Knutson KA, Norby FL, Alonso A, Soli-man EZ,Tang W, Pankow JS, Pan W, Chen LY. Multimodal data integration to predict atrial fibrillation. Eur Heart J Digit Health. 2024;6(1):126-136. doi:10.1093/ehjdh/ztae081
43. Yeh CH, Tsai TH, Chen CH, Wu YW, Wang CL, Liao KM, Huang CC, Chen CY. Artificial intelligence-enhanced electrocardiography improves the detection of coronary artery dis-ease. Comput Struct Biotechnol J. 2024;27:278–286. doi:10.1016/j.csbj.2024.12.032
44. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA 2020: An updated framework for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
45. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and Physio-Net. Circulation. 2000;101(23)\:e215–e220. doi:10.1161/01.CIR.101.23.e215
46. Wagner P, Strodthoff N, Bousseljot RD, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Sci Data. 2020;7(1):154. doi:10.1038/s41597-020-0495-6
47. Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). Published online 2022 April 9, 2025. doi:10.13026/kfzx-aw45