



Artificial intelligence enabled ECG screening for left ventricular systolic dysfunction: a systematic review

Laura Vindeløv Bjerken¹ · Søren Nicolaj Rønborg² · Magnus Thorsten Jensen^{3,4,5} · Silas Nyboe Ørting⁶ · Olav Wendelboe Nielsen^{1,2}

Accepted: 18 October 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Screening for left ventricular systolic dysfunction (LVSD), defined as reduced left ventricular ejection fraction (LVEF), deserves renewed interest as the medical treatment for the prevention and progression of heart failure improves. We aimed to review the updated literature to outline the potential and caveats of using artificial intelligence-enabled electrocardiography (AIeECG) as an opportunistic screening tool for LVSD.

We searched PubMed and Cochrane for variations of the terms “ECG,” “Heart Failure,” “systolic dysfunction,” and “Artificial Intelligence” from January 2010 to April 2022 and selected studies that reported the diagnostic accuracy and confounders of using AIeECG to detect LVSD.

Out of 40 articles, we identified 15 relevant studies; eleven retrospective cohorts, three prospective cohorts, and one case series. Although various LVEF thresholds were used, AIeECG detected LVSD with a median AUC of 0.90 (IQR from 0.85 to 0.95), a sensitivity of 83.3% (IQR from 73 to 86.9%) and a specificity of 87% (IQR from 84.5 to 90.9%). AIeECG algorithms succeeded across a wide range of sex, age, and comorbidity and seemed especially useful in non-cardiology settings and when combined with natriuretic peptide testing. Furthermore, a false-positive AIeECG indicated a future development of LVSD. No studies investigated the effect on treatment or patient outcomes.

This systematic review corroborates the arrival of a new generic biomarker, AIeECG, to improve the detection of LVSD. AIeECG, in addition to natriuretic peptides and echocardiograms, will improve screening for LVSD, but prospective randomized implementation trials with added therapy are needed to show cost-effectiveness and clinical significance.

Keywords Artificial intelligence · Electrocardiogram · Left ventricular systolic dysfunction · Reduced left ventricular ejection fraction · Screening

Introduction

The first randomized controlled trial recently demonstrated how an AI-enabled electrocardiogram (AIeECG) could increase the number of patients detected with reduced left ventricular ejection fraction (LVEF) in a broad clinical setting [1]. Heart Failure (HF) occurs in about 37.7 million people worldwide, and a similar number of people have undetected or asymptomatic left ventricular systolic dysfunction [2]. Treating systolic HF with reduced LVEF is strongly recommended, even in some patients with LVEF below 50% [3, 4]. The reference method for detecting reduced LVEF, echocardiography, is time-consuming, expert-dependent, and costly [5]. Therefore, other more widely applicable methods are needed to enable detection, treatment and prevent the onset of HF, which is why screening deserves a renewed interest.

✉ Laura Vindeløv Bjerken
laura.bjerken@gmail.com

¹ Department of Clinical Medicine, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark

² Department of Cardiology, Copenhagen University Hospital Bispebjerg, Copenhagen, Denmark

³ Department of Cardiology, Copenhagen University Hospital Amager and Hvidovre, Hvidovre, Denmark

⁴ Present Address: Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

⁵ William Harvey Research Institute, Queen Mary University Hospital, London, UK

⁶ Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Screening community patients with brain natriuretic peptide (BNP) followed by clinical intervention reduces subsequent HF events compared with the standard of care [6]. We have previously shown that natriuretic peptides, history of hypertension, myocardial infarction and the ECG are useful biomarkers to detect reduced LVEF in a community cohort [7, 8].

The readily available and cheap ECG could be a perfect tool to identify patients who should have an echocardiogram examination. But the primary concern that ECG-reading skills are inadequate among general practitioners [9] must be addressed. Therefore, an AI algorithm that can make general physicians experts in reading ECGs could offer a perfect screening option in general practice and non-cardiology departments [10].

However, it is unknown whether AIeECG is a generally applicable generic biomarker, or whether it only works in specific populations and for dedicated research groups. To our knowledge, no one has made a comprehensive review of the studies that used AIeECG to screen for reduced LVEF or left ventricular systolic dysfunction (LVSD).

Our aim is to review the existing literature with AIeECG to detect LVSD and explore the following sub-questions: What are the similarities and differences between the research groups' algorithms for accuracy in detecting LVSD in relation to LVEF thresholds, study populations, and is AIeECG better than natriuretic peptides.

Method

A systematic search on Pubmed and Cochrane focused on studies using AIeECG in a clinical context, but not a technical. The literature review and reporting adhered to the Prisma guidelines [11].

Pubmed search

The search contained variations of three main aspects regarding the use of AIeECG to screen for LVSD: "ECG," "Heart Failure" and "Artificial Intelligence" (Supplementary Appendix 1). The search consisted of both a MeSH term for each of the main aspects as well as variations of that main aspect. The MeSH term searches and variations hereof were combined using "AND" and the search was restricted to the last 10 years (Supplementary Appendix 2).

Cochrane search

A supplementary search was also performed in Cochrane using the search words: machine learning, deep learning, neural network, and artificial intelligence. This search did not contribute any new articles.

Study selection

Study selection was performed by two separate reviewers (LB and SR), and in case of agreement articles were included. If the two reviewers disagreed, a third-party reviewer (OW) was consulted, and each article was discussed until consensus was reached. The risk of bias was reduced by excluding studies with a poor definition of the outcome variable (LVEF threshold or HF event).

A total of 147 articles were found. 107 articles were discarded based on title and abstract because they did not address ECG, LVSD, or artificial intelligence. The remaining 40 articles were assessed for full-text eligibility (Fig. 1) and a further 25 articles were excluded for the following reasons: no focus on LVSD screening ($n=12$), a technical rather than clinical scope ($n=1$), not original study but a meta-analysis or review ($n=4$), or not an original study but a comment, protocol or editorial ($n=5$). Exclusion due to poor quality related to studies that did not focus on using AIeECG ($n=3$) (Fig. 1). The remaining 15 articles were included in the present investigation. Certainty in evidence was assessed by noting whether algorithms had been tested externally or in prospective randomized trials.

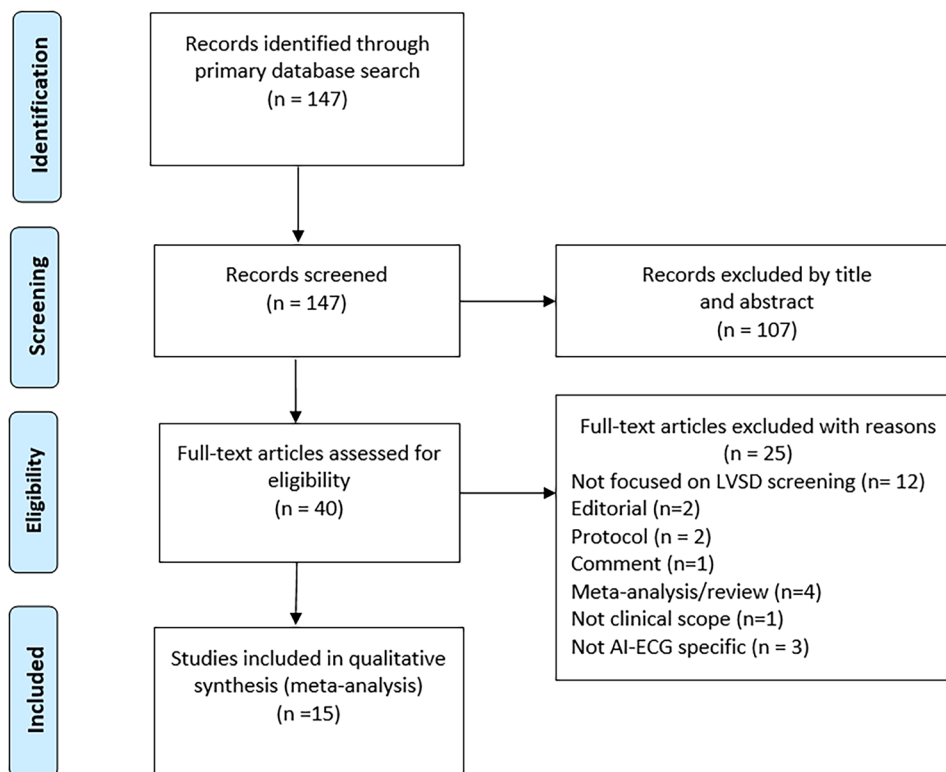
Data collection from articles

Numerical values in this review are reported as stated in the original articles, but we estimated data from figures or Forest plots when numbers were missing. We categorized studies after the first author. We report diagnostic accuracy data from the external validation rather than from training and internal validation. A formal statistical testing was inappropriate because of too few studies, therefore we report the median values and interquartile range for AUC, sensitivity, and specificity across all algorithms instead.

Results

Overview

We identified fifteen studies (Fig. 1 and Table 1) investigating algorithms from seven different research groups. Attia et al. [12] from the Mayo Clinic created an algorithm that was tested in eight studies [12–17], one group from Seoul (the republic of Korea) developed and tested their algorithms in two studies [18, 19]. Five research groups (Sbrollini et al.) developed in-house algorithms and tested them in separate studies [20–24]. A detailed overview of each study population and outcome can be found in "Supplementary results" (Supplementary Appendix 3), where details are provided for AUC, type of outcome, model adjustments and comparison to other test (e.g., BNP/NT-proBNP).

Fig. 1 PRISMA flowchart of study selection

A retrospective design was used in eleven studies [12–14, 17–24], one was a case series [15] and three were prospective cohort studies [16, 25, 26]. Apart from those 15 studies that fulfilled eligibility criteria for development and testing of algorithms, we identified one randomized controlled trial [1] and one cost-effectiveness study [27]. No studies examined the effects of using AI-enabled screening in relation to treatment changes, patient outcomes or quality of life.

The diversity of study cohorts reflected a broad clinical spectrum, spanning from unselected digital ECG databases [12, 15, 16, 20, 22–24], emergency department patients [13], cardiac intensive care unit patients [17], unspecified hospitalized patients [18, 19, 21], general populations [14, 25] and patients diagnosed with Chagas disease [26].

The reference standard for AIeECG in fourteen of fifteen studies was an echocardiogram performed within a prespecified period of 1 week to 1 year apart from the examined ECG. The studies used different LVEF cutoffs to define their outcome LVSD (e.g., EF < 35%, EF < 40% and EF < 50%) (Table 1). One study used the outcome, “altered clinical status,” instead of a reduced LVEF, and that study examined serial ECGs instead of a single baseline ECG [20].

The algorithms appear to perform better in the hospital populations and when the outcome (LVSD) was defined by a lower EF cut-off (e.g., EF ≤ 35/40% instead of EF < 50%) (Table 1).

Some of the included studies investigated whether the algorithms performed better or worse in selected

subgroups (age, sex, and comorbidities). Algorithm performance seems to be lower in populations with comorbidity (acute coronary syndrome, PCI intervention performed, diabetes mellitus, renal disease, hypertension, hypercholesterolemia, and previous myocardial infarction), with the exception of obesity which did not affect performance. Age and sex did not affect the algorithms performance significantly. Overall, algorithm performance was strong and stable across all investigated subgroups with AUCs > 0.765 (Table 2).

Furthermore, five studies compared their AIeECG with BNP or NT-proBNP (Table 1 and Supplementary Appendix 3). The few studies that made this comparison found that AIeECG performed better with a numerical higher AUC [12, 22]. Two studies also combined the two methods and found that this combination outperformed both methods individually [13, 26].

The external validation of all fifteen algorithms resulted in a median AUC of 0.90 (IQR from 0.85 to 0.95), a sensitivity 83.3% (IQR from 73–86.9%) and a specificity 87% (IQR from 84.5 to 90.9%) (Table 1).

Discussion

We identified 7 different AIeECG algorithms (Table 1) in this first comprehensive review of studies using AIeECG algorithms to screen for LVSD. Despite

Table 1 Included studies in review

Study and algorithm	Population, design, and method			Outcome		External diagnostic value			Comparison with other tests
	Population and ECG recording	Study design	AI model	Definitions	Prevalence and no. patients	AUC (95% CI)	Sensitivity	Specificity	
Attia algorithm									
Attia et al. [12]	97,829 patients from the Mayo Clinic “digital data vault”. 35,970 for training, 8989 for internal validation and 52,870 to test the AI algorithm. All had ECG and TTE taken within 2 weeks of each other	Retrospective cohort	Convolutional neural network (CNN)	EF \leq 35%	7.8% of 52,870	0.932	86.3%	85.7%	BNP—AUC 0.60
Attia et al. [16]	6008 adult patients from the “Mayo Clinic ECG Laboratory” who had a routine ECG and subsequent echocardiogram taken within a year	Prospective cohort	CNN ad modum Attia et al. [12]	EF \leq 35% Echo within a month EF \leq 35% Echo within a year	7.0% of 3874 5.4% of 6008	0.918 (0.902–0.934) 0.911 (0.897–0.925)	82.5% 81.5%	86.8% 86.3%	AIeECG + NT-proBNP (> 125 pg/mL) reduced the number of false positives with 5 without increasing the number of false negatives
Attia et al. [15]	27 patients with Covid-19 from the mayo clinic electronic medical record who had an ECG and echocardiogram taken within 2 weeks	Case-series	CNN ad modum Attia et al. [12]	EF \leq 40%	11.1% of 27	0.950	Not specified	Not specified	Not specified
Attia et al. [14]	4277 participants (35–69 years) from the “Know your heart study” who had an ECG and TTE taken within an undefined period	Retrospective cohort	CNN ad modum Attia et al. [12]	EF \leq 35%	0.6% of 4277	0.82 (0.75–0.90)	26.9%	97.4%	
Jentzer et al. [17]	5680 patients from a cardiac intensive care unit who had ECG and TTE taken within 7 days	Retrospective cohort	CNN ad modum Attia et al. [12]	EF \leq 40%	33.6% of 5680	0.83 (0.82–0.84)	73%	78%	

Table 1 (continued)

Study and algorithm	Population, design, and method			Outcome	External diagnostic value			Comparison with other tests	
	Population and ECG recording	Study design	AI model		Definitions	Prevalence and no. patients	AUC (95% CI)		Sensitivity
Adedinswo et al. [13]	1606 emergency department patients (>18 years) with dyspnea who had an ECG and echocardiogram taken within 30 days	Retrospective cohort	CNN ad modum Attia et al. [12]	EF ≤ 35% EF < 50%	10.2% of 1606 11.6% of 1606	0.89 (0.86–0.91) 0.85 (0.83–0.88)	74% 63%	87% 91%	NT-proBNP (>800 pg/mL) AUC 0.80 (0.76–0.94) AIeECG + NT-proBNP AUC 0.91
Kashou et al. [25]	2,041 randomly selected people (>45 years) from Olmsted County, Minnesota, who underwent ECG and echocardiogram on the same day	Prospective cohort	CNN ad modum Attia et al. [12]	EF ≤ 40% Preclinical all patients Preclinical high risk	2.0% of 2,041 1.1% of 1,996 1.5% of 1,348	0.971 0.967 0.968	90% 86.4% 90%	92% 92.5% 90.8%	
Brito et al. [26]	1304 patients (>19 years) from the “SaMITrop study” diagnosed with Chagas disease who had an EKG, echocardiogram and NT-proBNP measurement performed	Prospective cohort	CNN ad modum Attia et al. [12]	EF ≤ 40%	7.1% of 1,304	0.839 (0.793–0.885)	73%	83%	AIeECG + NT-proBNP AUC 0.874 (0.831–0.918)
Katsushika algorithm									
Katsushika et al. [21]	18,919 patients (>18 years) from The University of Tokyo Hospital who had an ECG and echocardiogram taken within 28 days between Jan 2015 and Dec 2019. 12,108 for training, 3027 for internal validation and 3784 for testing	Retrospective cohort	CNN – inspired by Attia et al. [12]	EF ≤ 40%	10.2% of 3,784	0.945 (0.936–0.954)	85.8%	90%	Compared to accuracy of cardiologists predicting LVSD from ECGs Accuracy without model 78% and with model 88%

Table 1 (continued)

Study and algorithm	Population, design, and method		Outcome		External diagnostic value		Comparison with other tests	
	Population and ECG recording	Study design	AI model	Definitions	Prevalence and no. patients	AUC (95% CI)		Sensitivity
Sun algorithm								
Sun et al. [22]	26,792 Adult patients from "the digital database of Wuxi People's Hospital" who had ECG and TTE taken within 14 days. 21,732 for training, 2530 for internal validation and 2530 for testing	Retrospective cohort	Convolutional neural network	EF \leq 50%	4.7% of 26,792 2530	0.713	69.2%	70.5%
								BNP—AUC 0.60 NT-proBNP—AUC 0.70
Kwon algorithm								
Kwon et al. [19]	22,765 patients (> 18 years) from 2 different hospitals in Seoul who had ECG and echocardiogram taken within 4 weeks. 13,486 for training, 3378 for internal validation and 5901 to test the algorithm	Retrospective cohort	Deep neural network based on preselected ECG features	EF \leq 40% EF < 50%	6.1% of 22,765 5901 6.8% of 22,765 5901	0.889 (0.887–0.891) 0.850 (0.848–0.852)	Not specified	Not specified
Cho et al. [18]	24,211 patients (> 18 years) from two different hospitals in Seoul who had ECG and echocardiogram taken within 30 days. 17,127 for training, 2908 for internal validation and 4176 to test the algorithm	Retrospective cohort	Convolutional neural network combined with short time fourier transform	EF \leq 40%	6.8% of 24,211 4176	1.0961	91.5%	91.1%

Table 1 (continued)

Study and algorithm	Population, design, and method			Outcome	External diagnostic value			Comparison with other tests
	Population and ECG recording	Study design	AI model		Definitions	Prevalence and no. patients	AUC (95% CI)	
Sbrollini algorithm								
Sbrollini et al. [20]	A total of 129 patients divided into 47 cases from a HF database and 81 controls. 54 patients for training, 10 for internal validation and 65 to test the algorithm	Retrospective cohort	Deep neural network based on preselected ECG features from serial ECGs	Altered clinical status	58% of 129 65	0.84 (0.73–0.95)	Not specified	Not specified
Chen algorithm								
Chen et al. [24]	58,431 patients from “the electronic medical records of Tri-Service General Hospital of Taipei, Taiwan” who had ECG and TTE taken within 7 days. 30,531 for development, 7271 for internal validation and 20,629 to test the algorithm and for follow-up	Retrospective cohort	Convolutional neural network	EF ≤ 35% EF < 50%	3.6% of 20,629 1.5% of 20,629	0.947 0.885	86.9% 72.1%	89.6% 88%
Vaid algorithm								
Vaid et al. [23]	147,636 Patients from 5 different New York City hospitals within “the EPIC electronic healthcare record system” with ECG and TTE taken within 7 days	Retrospective cohort	Convolutional neural network	EF ≤ 40% EF 41–50% EF > 50% EF ≤ 35	25.9% of 597 14.9% of 597 59.3% of 597 23.07% of not specified	0.94 (0.94–0.95) 0.73 (0.72–0.74) 0.87 (0.87–0.88) 0.95 (0.95–0.96)	87% 78% 84% 88%	85% 57% 81% 87%

Table 2 AUC in subgroups

Outcome	Study	Clinical subgroup	AUC	Reported significant difference between groups
EF ≤ 35%	Attia et al. [12]	Comorbidity	0.93 (CI not reported)	Yes, <i>p</i> -value not reported
		+ Comorbidity % Comorbidity	0.98 (CI not reported)	
EF ≤ 40%	Jentzer et al. [17]	+ Acute coronar syndrome	0.800 (0.780–0.820)	Yes, <i>p</i> -value not reported
		% Acute coronar syndrome	0.860 (0.840–0.880)	
EF ≤ 40%	Jentzer et al. [17]	+ PCI during hospitalization	0.810 (0.780–0.830)	Yes, <i>p</i> -value not reported
		% PCI during hospitalization	0.840 (0.830–0.850)	
EF ≤ 40%	Cho et al. [18]	BMI < 30	0.962 (0.955–0.967)	No, <i>p</i> = 0.902
		BMI ≥ 30	0.963 (0.938–0.981)	
Sex				
EF ≤ 40%	Jentzer et al. [17]	Male	0.840 (0.830–0.860)	Yes, <i>p</i> -value not reported
		Female	0.790 (0.770–0.810)	
EF ≤ 35%	Adedinsewo et al. [13]	Male	0.869 (0.830–0.907)	No, <i>p</i> -value not reported
		Female	0.904 (0.863–0.944)	
EF ≤ 50%	Adedinsewo et al. [13]	Male	0.834 (0.800–0.869)	No, <i>p</i> -value not reported
		Female	0.875 (0.834–0.915)	
EF ≤ 40%	Cho et al. [18]	Male	0.960 (0.951–0.967)	No, <i>p</i> = 0.700
		Female	0.964 (0.9550.972)	
EF ≤ 50%	Attia et al. [14]	Male	0.852 (0.762–0.942)	Not reported
		Female	0.761 (0.620–0.901)	
EF ≤ 35%	Attia et al. [12] *	Male	87.5% / 85.5%	No, <i>p</i> -value not reported
		Female	82% / 90%	
EF ≤ 40%	Kashou et al. [25]	Male	0.962 (CI not reported)	Not reported
		Female	0.980 (CI not reported)	
Age				
EF ≤ 40%	Jentzer et al. [17]	< 70 years	0.850 (0.830–0.860)	Yes, <i>p</i> -value not reported
		≥ 70 years	0.800 (0.790–0.820)	
EF ≤ 40%	Cho et al. [18]	< 65 years	0.974 (0.968–0.980)	Yes, <i>p</i> < 0.001
		≥ 65 years	0.932 (0.917–0.945)	
EF ≤ 35%	Adedinsewo et al. [13]	18–59 years	0.909 (0.864–0.955)	No, <i>p</i> -value not reported
		59–69 years	0.875 (0.819–0.932)	
		69–89 years	0.871 (0.827–0.916)	
EF ≤ 50%	Adedinsewo et al. [13]	18–59 years	0.893 (0.854–0.931)	No, <i>p</i> -value not reported
		59–69 years	0.842 (0.787–0.896)	
		69–89 years	0.832 (0.792–0.873)	
EF ≤ 50%	Attia et al. [14]	30–55 years	0.905 (0.829–0.981)	Not reported
		55–75 years	0.765 (0.652–0.877)	
EF ≤ 35%	Attia et al. [12] *	< 60 years	77%/91.5%	No, <i>p</i> -value not reported
		≥ 60 years	89%/82.5%	

*Sensitivity/Specificity as AUC was not reported

investigating various study populations and using different LVEF thresholds to define LVSD, the majority of studies obtained a high diagnostic accuracy on external validation. Overall, AIeECG seems to be a robust and potentially universal tool to screen for LVSD, which could improve when combined with clinical characteristics such as gender and comorbidities as well as NTpro-BNP.

AIeECG screening for LVSD

Overall, AIeECG had a high diagnostic value when screening for LVSD and resulted in median AUC of 0.90 (IQR from 0.85 to 0.95). The median sensitivity of 83.3% ensures a low number of false negative screen failures while a high specificity of 87% leads to a low number of false positive screen subjects if used for screening.

The studies examined different populations with different prevalences and outcome definitions (Table 1), making direct comparison difficult. Most consistently, there appears to be higher AUC, sensitivity, and specificity in the studies examining hospital populations. A higher AUC may be explained by populations with a few severe cases of LVSD and a large number of patients admitted with non-cardiac conditions.

Risk of bias and certainty of evidence

Retrospective designs were dominant as these are obviously easier to complete. Furthermore, most algorithms were tested in selected populations which is a drawback since results cannot directly translate to future value in clinical practice with consecutive patients. Accordingly, the prevalence of LVSD in the tested populations was higher than in the general population. Selection bias occurs when including patients who already had an ECG and echocardiogram performed, as it selects higher-risk patients who already had an indication for an echocardiogram. This could result in a better performance than may be observed in a prospective study.

Truly external validation was only applied to one algorithm by Attia et al. [12] as this was the only algorithm to be tested in other studies with separate populations [13–17, 25, 26]. More than half of the algorithms were not validated in a truly external dataset as the same dataset was used for both development and validation (the same dataset was split into a training, internal validation, and external validation group) [20–22, 24]. The remaining studies used different datasets for training/validation and external validation, respectively, but the majority of algorithms were not truly externally validated [18, 19, 23].

Despite all the differences between the populations used for the development of algorithms, a consistent high AUC was shown by several groups, not just by one dedicated group in a selected population. Therefore, AIeECG has the potential to become a widespread and useful technique in clinical practice in the future.

Clinical characteristics

Six of the included studies [12–14, 17, 18, 25] performed sub-analyses according to comorbidity, age, and sex (Table 2). AI diagnostic performance seemed lower in populations with comorbidity (Table 2), but data lacked power for a formal statistical analysis. Populations without comorbidities were associated with a higher AUC in two out of three studies [12, 17]. Overall gender did not affect the algorithms, although a single study found that the algorithm was significantly better in men [17]. The algorithms seemed to perform better in the younger

populations, maybe due to less comorbidity, but this was only reported significant in two of the five studies [17, 18]. It is a major limitation that the impact of ethnicity on the performance of the algorithms was not investigated, because several studies have shown that different ethnicities display different ECG characteristics. Therefore, ethnicity specific ECG reference ranges/cut-offs are pertinent to investigate [28–30].

It is a strength regarding screening that the performance of AIeECG was not strongly associated with gender and age as the overall performance was strong in these subgroups. Only few studies investigated this, therefore it needs to be further investigated in future studies.

Potential for improvement

In clinical practice, more useful information beyond the ECG is available for the clinician and for an AI algorithm. Therefore, the diagnostic value of AIeECG combined with demographics and clinical information should be further tested.

Very few studies compared their AIeECG algorithm with the performance of BNP or NT-proBNP measurements as a screening method for HF [12, 13, 16, 22, 26]. Two studies found that their AIeECG algorithm outperformed natriuretic peptide measurement when identifying reduced LVEF [12, 22]. The addition of NT-proBNP to the AIeECG marginally improved detection of LVSD [13, 26] and resulted in a higher specificity and fewer false-positive screen cases, without increasing the number of false-negative screenings [16].

We cannot yet conclude that AIeECG outperforms BNP and NT-proBNP measurement as a screening method, but it seems that AIeECG may be more stable across age and gender than BNP [12, 18, 31], and a combination might therefore be the optimal way of screening.

Detection vs. prediction

Besides detecting LVSD with a high diagnostic accuracy some of algorithms also predicts future LVSD or HF events.

In three studies [12, 23, 25], patients with a *false* positive AIeECG at the time of screening had a significantly increased risk of subsequently developing LVSD compared to patients with a *true* negative AIeECG screening. Another study found that patients with a *false* positive AIeECG were more susceptible to major adverse cardiovascular events (HR 1.5) compared to patients with a *true* negative AIeECG [24].

In one of the studies, the algorithm was able to predict newly emerging HF pathology, as well as aggravating cardiac pathology by detecting subtle changes in a newly recorded ECG compared to a previously recorded ECG [20].

These findings support the notion that AI algorithms detect subtle or subclinical ECG changes that are associated with risk of future LVSD. This notion was corroborated in a study screening for HF with preserved ejection fraction (HFpEF) where a significant proportion of *false* positives remarkably developed HFpEF during follow-up [32]. The AIeECG has the ability to predict a wide range of pathologies, even simultaneously, such as HFpEF, right ventricular dysfunction and more [20, 23, 24, 32, 33]. Hence, the gain of AIeECG is most likely greater than we demonstrated in this review.

Clinical implications of AIeECG

With the emerging AIeECG technology, the idea of screening for LVSD is worth revisiting, especially in combination with BNP/NTproBNP and basic clinical information. The health expenditures for patients diagnosed with HF are expected to rise in the coming years [34]. Lifesaving treatment to prevent the development of HF, hospitalization and death are evident if early identification is possible [12].

Clinical implication in regard to early diagnosis of low EF has been investigated in the recently published EAGLE trial [1] which substantiates that the concept of screening in the primary sector is viable. The study found that the use of an AIeECG algorithm increased the diagnosis of low EF in the overall cohort (1.6% in the control arm versus 2.1% in the intervention arm), suggesting a modest but significant gain from using AIeECG. Importantly, the use of AIeECG was not associated with an overall increased use of echocardiography, but instead an increased use of echocardiography on more relevant patients.

Notably, AIeECG had the highest value for primary care physicians and the lowest value when applied in hospital wards. These findings suggest that AIeECG may be most useful for clinicians who are not routinely interpreting ECGs, and less useful in settings where echocardiograms are performed routinely. A clear distinction does not exist as Katsushika et al. [21] showed how AIeECG proved to help even cardiologist with > 7 years' experience.

ECG screening for LVSD is potentially feasible but cost-effectiveness and clinical implication are yet to be fully investigated [35]. So far, only two studies have investigated cost-effectiveness of screening for LVSD [25, 27]. Under most clinical scenarios, screening was cost-effective with a cost of < \$50,000 per QALY [27]. It was estimated that numbers needed to screen to identify one case of LVSD corresponds to 90.7 AIeECGs' and 8.8 echocardiograms when screening the total population. But it could be reduced to 67.4 AIeECGs' and 5.6 echocardiograms when screening a "high risk" population [25]. Cost-effectiveness increased with higher disease prevalence and better sensitivity of the AIeECG method. Thus, cost-effectiveness

can possibly be improved if screening is applied to subjects with preexisting cardiovascular risk factors, abnormal natriuretic peptide, diabetes mellitus, hypertension, and ischemic heart disease [27].

Future use of AIeECG

AIeECG algorithms will most likely be implemented in ECG machines in a few years. A new study has even shown that it can be built into a stethoscope to detect low EF during cardiac auscultation [36], the possibilities are plenty. One attribute of AIeECG is the high specificity which may guide physicians to refer patients with abnormal AIeECG findings and increase the likelihood of identifying the patients at the highest risk.

In the future, multiple groups can and will be able to produce effective algorithms, but standardization is required to compare effectiveness of algorithms. One solution could be external validation of algorithms in a multinational common dataset of paired ECG's and outcome measures.

We anticipate that a breakthrough for these algorithms will occur when they are combined with other risk markers, possibly natriuretic peptides, and tested prospectively including management actions based on the screening results. Success may be defined when a 15–20% reduction of HF hospitalization or all-cause mortality is demonstrated in comparison with standard of care. But less ambitious goals such as demonstrating a reduction in the cost per identified patient with LVSD are also valid as it leads to more rational use of resources.

Strengths and limitations

The strengths of this literature study are the simple, well-defined research questions and the clinically focused literature search, consisting of both Mesh terms and free-text words. The study followed PRISMA guidelines for reporting results, but some limitations of this work must be acknowledged when considering the findings. We searched Pubmed and Cochrane databases and only included peer-reviewed studies of high quality to focus on clinical aspects rather than technical differences between the algorithms. We focused solely on LVSD and a 12-lead ECG which could have led to exclusion of otherwise relevant studies that examined AIeECG based on one or three lead electrocardiograms. Furthermore, although it is a limitation that we did not use a formal "risk-of-bias assessment tool," we aimed to minimize the risk of bias by using strict study selection criteria.

Due to few studies and lack of statistical power, we summarized data instead of making formal statistical tests. Many more studies and algorithms will without doubt evolve over the next years, allowing for more accurate estimates of

accuracy. We suspect such result will most likely lie within the range reported here. Still, the main objective of our review was to examine whether AIeECG works generically and to point at strengths and possibilities for improvement.

Conclusion

This systematic review corroborates the arrival of a new generic biomarker, AIeECG, to screen for LVSD. Overall, the algorithms identified LVSD with a high diagnostic value and predict LVSD in addition to natriuretic peptides and echocardiograms. AIeECG has the potential to increase general physicians' proficiency to use ECG and choose to refer the right patients for a diagnostic echocardiogram. Further randomized implementation studies are needed, especially in the primary sector, to show cost-effectiveness and clinical significance, preferably in combination with other biomarkers such as the natriuretic peptides.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10741-022-10283-1>.

Author contribution Olav Wendelboe Nielsen and Laura Vindeløv Bjerken had the idea for the article and performed the literature search. Study selection was performed by Laura Vindeløv Bjerken and Søren Nicolaj Rønborg and was discussed with Olav Wendelboe Nielsen for appropriateness. The first draft of the manuscript was written by Laura Vindeløv Bjerken with input from Olav Wendelboe Nielsen and Søren Nicolaj Rønborg. The manuscript was critically revised by all authors.

Data availability Not applicable.

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

References

1. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM et al (2021) Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med* 27(5):815–819
2. Atherton JJ (2010) Screening for left ventricular systolic dysfunction: is imaging a solution? *JACC Cardiovasc Imaging* 3(4):421–428
3. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M et al (2021) ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 00:1–128
4. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM et al (2022) AHA/ACC/HFSA Guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 145(18):e895–e1032
5. Jahmunah V, Oh SL, Wei JKE, Ciaccio EJ, Chua K, San TR et al (2019) Computer-aided diagnosis of congestive heart failure using ECG signals - a review. *Phys Med* 62:95–104
6. Ledwidge M, Gallagher J, Conlon C, Tallon E, O'Connell E, Dawkins I et al (2013) Natriuretic peptide-based screening and collaborative care for heart failure - the STOP-HF randomized trial. *JAMA* 310(1):66–74
7. Nielsen OW, Cowburn PJ, Sajadieh A, Morton JJ, Dargie H, McDonagh T (2007) Value of BNP to estimate cardiac risk in patients on cardioactive treatment in primary care. *Eur J Heart Fail* 9(12):1178–1185
8. Nielsen OW, McDonagh TA, Robb SD, Dargie HJ (2003) Retrospective analysis of the cost-effectiveness of using plasma brain natriuretic peptide in screening for left ventricular systolic dysfunction in the general population. *J Am Coll Cardiol* 41(1):113–120
9. Goudie BM, Jarvis RI, Donnan PT, Sullivan FM, Pringle SD, Jeyaseelan S et al (2007) Screening for left ventricular systolic dysfunction using GP-reported ECGs. *Br J Gen Pract* 57:191–195
10. Reichlin T, Abacherli R, Twerenbold R, Kuhne M, Schaer B, Muller C et al (2016) Advanced ECG in 2016: is there more than just a tracing? *Swiss Med Wkly* 146:w14303
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71
12. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G et al (2019) Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 25(1):70–74
13. Adedinsewo D, Carter RE, Attia Z, Johnson P, Kashou AH, Dugan JL et al (2020) Artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea. *Circ Arrhythm Electrophysiol* 13(8):e008437
14. Attia IZ, Tseng AS, Benavente ED, Medina-Inojosa JR, Clark TG, Malyutina S et al (2021) External validation of a deep learning electrocardiogram algorithm to detect ventricular dysfunction. *Int J Cardiol* 329:130–135
15. Attia ZI, Kapa S, Noseworthy PA, Lopez-Jimenez F, Friedman PA (2020) Artificial intelligence ECG to detect left ventricular dysfunction in COVID-19: a case series. *Mayo Clin Proc* 95(11):2464–2466
16. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA et al (2019) Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol* 30(5):668–674
17. Jentzer JC, Kashou AH, Attia ZI, Lopez-Jimenez F, Kapa S, Friedman PA et al (2020) Left ventricular systolic dysfunction identification using artificial intelligence-augmented electrocardiogram in cardiac intensive care unit patients. *Int J Cardiol*
18. Cho J, Lee B, Kwon JM, Lee Y, Park H, Oh BH et al (2021) Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography. *ASAIO J* 67(3):314
19. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM et al (2019) Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J* 49(7):629–639
20. Sbröllini A, De Jongh MC, Ter Haar CC, Treskes RW, Man S, Burattini L et al (2019) Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: a deep-learning approach. *Biomed Eng Online* 18(1):15
21. Katsushika S, Kodera S, Nakamoto M, Ninomiya K, Inoue S, Sawano S et al (2021) The effectiveness of a deep learning model to detect left ventricular systolic dysfunction from electrocardiograms. *Int Heart J* 62(6):1332–1341
22. Sun JY, Qiu Y, Guo HC, Hua Y, Shao B, Qiao YC et al (2021) A method to screen left ventricular dysfunction through ECG based

- on convolutional neural network. *J Cardiovasc Electrophysiol* 32(4):1095–1102
23. Vaid A, Johnson KW, Badgeley MA, Somani SS, Bicak M, Landi I et al (2022) Using deep-learning algorithms to simultaneously identify right and left ventricular dysfunction from the electrocardiogram. *JACC Cardiovasc Imaging* 15(3):395–410
 24. Chen HY, Lin CS, Fang WH, Lou YS, Cheng CC, Lee CC et al (2022) Artificial intelligence-enabled electrocardiography predicts left ventricular dysfunction and future cardiovascular outcomes: a retrospective analysis. *J Pers Med* 12(3)
 25. Kashou AH, Medina-Inojosa JR, Noseworthy PA, Rodeheffer RJ, Lopez-Jimenez F, Attia IZ et al (2021) Artificial intelligence-augmented electrocardiogram detection of left ventricular systolic dysfunction in the general population. *Mayo Clin Proc* 96(10):2576–2586
 26. Brito BOF, Attia ZI, Martins LNA, Perel P, Nunes MCP, Sabino EC et al (2021) Left ventricular systolic dysfunction predicted by artificial intelligence using the electrocardiogram in Chagas disease patients-The SaMi-Trop cohort. *PLoS Negl Trop Dis* 15(12):e0009974
 27. Tseng AS, Thao V, Borah BJ, Attia IZ, Medina Inojosa J, Kapa S et al (2021) Cost effectiveness of an electrocardiographic deep learning algorithm to detect asymptomatic left ventricular dysfunction. *Mayo Clin Proc* 96(7):1835–1844
 28. Burden A (2016) QRS duration and ethnicity: implications for heart failure therapy. *Heart* 102(18):1427–1428
 29. Okin PM, Wright JT, Nieminen MS, Jern S, Taylor AL, Phillips R et al (2022) Ethnic differences in electrocardiographic criteria for left ventricular hypertrophy: the LIFE study. *Am J Hypertens* 15(8):663–71
 30. Santhanakrishnan R, Wang N, Larson MG, Magnani JW, Vasan RS, Wang TJ et al (2016) Racial differences in electrocardiographic characteristics and prognostic significance in whites versus Asians. *J Am Heart Assoc* 5(3):e002956
 31. Anderson KP (2019) Artificial intelligence-augmented ECG assessment: the promise and the challenge. *J Cardiovasc Electrophysiol* 30(5):675–678
 32. Robert MAvdB (2021) Predicting heart failure with preserved ejection fraction: revisiting an old friend with new knowledge. *Eur Heart J Dig Health* 2(1):104–105
 33. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA (2021) Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*
 34. Ziaean B, Fonarow GC (2016) Epidemiology and aetiology of heart failure. *Nat Rev Cardiol* 13(6):368–378
 35. Khunti K, Squire I, Abrams KR, Sutton AJ (2004) Accuracy of a 12-lead electrocardiogram in screening patients with suspected heart failure for open access echocardiography: a systematic review and meta-analysis. *Eur J Heart Fail* 6(5):571–576
 36. Attia Z, Dugan J, Rideout A, Maidens JN, Venkatraman S, Guo L et al (2022) Automated detection of low ejection fraction from a one-lead electrocardiogram: application of an AI algorithm to an ECG-enabled digital stethoscope. *Eur Heart J Dig Health* ztac030

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.