

# Deep learning analysis of resting electrocardiograms for the detection of myocardial dysfunction, hypertrophy, and ischaemia: a systematic review

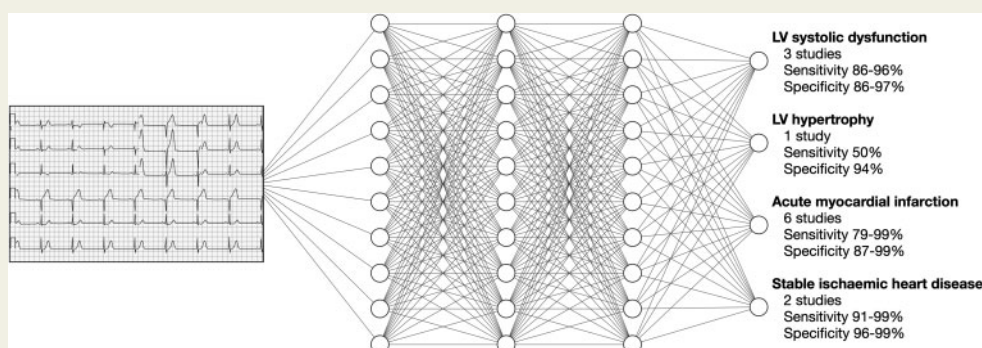
Ghalib Al Hinai<sup>1</sup>, Samer Jammoul<sup>1</sup>, Zara Vajihi<sup>2</sup>, and Jonathan Afilalo<sup>1,3\*</sup>

<sup>1</sup>Division of Cardiology, Jewish General Hospital, McGill University, 3755 Cote Ste Catherine Rd, E-222, Montreal, QC H3T 1E2, Canada; <sup>2</sup>Department of Emergency Medicine, Jewish General Hospital, McGill University, 3755 Cote Ste Catherine Rd, H-126, Montreal, QC H3T 1E2, Canada; and <sup>3</sup>Centre for Clinical Epidemiology, Jewish General Hospital, 3755 Cote Ste Catherine Rd, H-411, Montreal, QC H3T 1E2, Canada

Received 30 January 2021; revised 14 April 2021 online publish-ahead-of-print 7 August 2021

The aim of this review was to assess the evidence for deep learning (DL) analysis of resting electrocardiograms (ECGs) to predict structural cardiac pathologies such as left ventricular (LV) systolic dysfunction, myocardial hypertrophy, and ischaemic heart disease. A systematic literature search was conducted to identify published original articles on end-to-end DL analysis of resting ECG signals for the detection of structural cardiac pathologies. Studies were excluded if the ECG was acquired by ambulatory, stress, intracardiac, or implantable devices, and if the pathology of interest was arrhythmic in nature. After duplicate reviewers screened search results, 12 articles met the inclusion criteria and were included. Three articles used DL to detect LV systolic dysfunction, achieving an area under the curve (AUC) of 0.89–0.93 and an accuracy of 98%. One study used DL to detect LV hypertrophy, achieving an AUC of 0.87 and an accuracy of 87%. Six articles used DL to detect acute myocardial infarction, achieving an AUC of 0.88–1.00 and an accuracy of 83–99.9%. Two articles used DL to detect stable ischaemic heart disease, achieving an accuracy of 95–99.9%. Deep learning models, particularly those that used convolutional neural networks, outperformed rules-based models and other machine learning models. Deep learning is a promising technique to analyse resting ECG signals for the detection of structural cardiac pathologies, which has clinical applicability for more effective screening of asymptomatic populations and expedited diagnostic work-up of symptomatic patients at risk for cardiovascular disease.

## Graphical Abstract



## Keywords

Electrocardiogram • Deep learning • Artificial intelligence • Heart failure • Myocardial infarction • Coronary artery disease • Left ventricular hypertrophy

\* Corresponding author. Tel: (514) 340-7540, Fax: (514) 340-7534, Email: [jonathan.afilalo@mcgill.ca](mailto:jonathan.afilalo@mcgill.ca)

© The Author(s) 2021. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Introduction

The electrocardiogram (ECG) is one of the most commonly used diagnostic tools in clinical medicine, providing a broad range of information vital in the diagnosis and management of cardiovascular disease.<sup>1</sup> The utility of the ECG extends broadly beyond acute hospital care to outpatient primary care, home care, preoperative screening, athletic screening, telemedicine, and self-monitoring.

Computer-assisted interpretation of the ECG has become integral in clinical workflows since its introduction over 50 years ago, serving as an adjunct to physician interpretation.<sup>2,3</sup> Traditional models are dependent on computer recognition and measurement of pre-defined ECG features (waves, segments, and intervals) and rules-based classification of their normality or abnormality. These classification rules are programmed by humans based on known criteria for various pathologies, such that the computer algorithm ‘sees’ what the expert human would see, but faster and more consistently without the influence of fatigue and other human factors. However, the performance of traditional models for computer-assisted ECG interpretation remains suboptimal,<sup>4,5</sup> which has been attributed to the low accuracy of (archaic) classification rules and their lack of robustness in the face of imperfect tracings.

To address this, artificial intelligence models have been applied to ECG analysis with varying success. Earlier models used machine learning algorithms such as support vector machines and random forests to predict the likelihood of specific cardiac pathologies irrespective of pre-defined classification rules; notwithstanding, training these models still required the analyst to laboriously define and extract (‘engineer’) the features of interest from the ECG tracing. More recent models employed deep learning (DL) algorithms such as convolutional neural networks to perform the feature engineering step or obviate the need for this step altogether, and ultimately improve efficiency and predictive accuracy. Deep learning is a form of representation-based learning that consists of an input layer for the raw ECG signals, multiple hidden layers for the signal analysis, and an output layer for the final prediction of cardiac pathology (Figure 1).<sup>6</sup> Thus, the DL algorithm may ‘see’ informative features that the expert human may not visually appreciate or be trained to look for.

Much of the published research on DL-based analysis of ECGs has focused on the detection of atrial arrhythmias from ambulatory ECG devices and wearables, with less emphasis on the detection of

structural cardiac pathologies from the resting ECG. Structural cardiac pathologies such as heart failure (HF), hypertensive heart disease, and ischaemic heart disease are the pre-eminent causes of cardiovascular mortality and morbidity globally.<sup>7</sup> Therefore, our goal was to conduct a systematic review to address this gap and ascertain whether DL models could be used to detect left ventricular (LV) systolic dysfunction, hypertrophy, and acute or chronic forms ischaemic heart disease.

## Methods

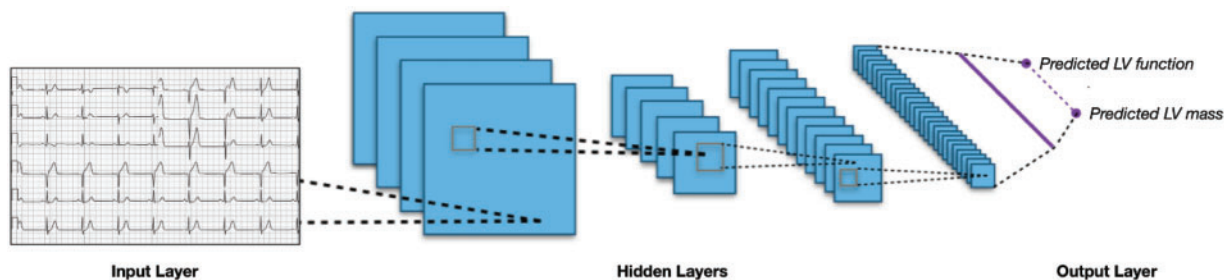
A systematic review was conducted to identify and aggregate published original studies that reported on DL-based analyses of resting ECGs (DL ECG) for the assessment of structural cardiac pathologies. The manuscript was prepared in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines.<sup>8</sup>

### Data sources and search strategy

PubMed MEDLINE was systematically searched from inception to 19 December 2019 and then focused on articles from 1 January 2009 to 19 December 2019 to correspond with the contemporary *big bang* of DL.<sup>9</sup> The following search terms were used: (‘Artificial Intelligence’[Mesh] OR ‘artificial intelligence’ OR ‘machine learning’ OR ‘deep learning’) AND (‘Electrocardiography’[MeSH] or electrocardiogra\* or ‘ECG’ or ‘EKG’). Search results were imported and screened using Rayyan (<https://rayyan.qcri.org/>)—a web-based software platform that allows duplicate reviewers to independently screen abstracts and full-text manuscripts and flag those that meet inclusion and exclusion criteria in a blinded fashion. In addition to our search results, references from retrieved studies were hand searched. When necessary, study investigators were contacted for clarification or to provide missing data. The complete list of search results may be made available upon request.

### Study selection

Two independent reviewers screened search results for articles that met the following inclusion criteria: (i) human adults; (ii)  $\geq 18$  years of age; (iii) underwent a resting surface ECG; (iv) end-to-end DL model used for analysis. Studies in which a DL algorithm was used for feature extraction and a non-DL algorithm was used for classification, or vice versa, were not considered to be end-to-end DL and therefore were not included. Exclusion criteria were: (i) underwent an ambulatory, stress, intracardiac, implantable, or bedside cardiac monitor ECG; (ii) non-original research



**Figure 1** Deep neural network. Sample architecture of a deep convolutional neural network composed of a first input layer for receiving the electrocardiogram signals, four hidden convolutional layers with multiple kernels for analysing the electrocardiogram signal features, and a dense output layer for generating the predicted left ventricular function and mass.

articles (reviews, editorials, and opinions); and (iii) non-English language articles. The reviewers were blinded to each other's selections for inclusion and exclusion, and disagreements were resolved by consensus with the senior author.

## Data extraction

Included studies were categorized according to whether the output of interest was prediction of LV dysfunction, hypertrophy, acute myocardial infarction, or stable ischaemic heart disease. For each study, the following

parameters were extracted: author, journal, year of publication, number of patients, proportion of females, mean age, duration of ECG recording, number of ECG leads, and algorithms used. The following statistical metrics of model performance were extracted and presented for the test set of ECGs (i.e. ECGs other than those analysed as part of the training set): sensitivity, specificity, accuracy, and area under the receiver operating characteristics curve (AUC).

## Results

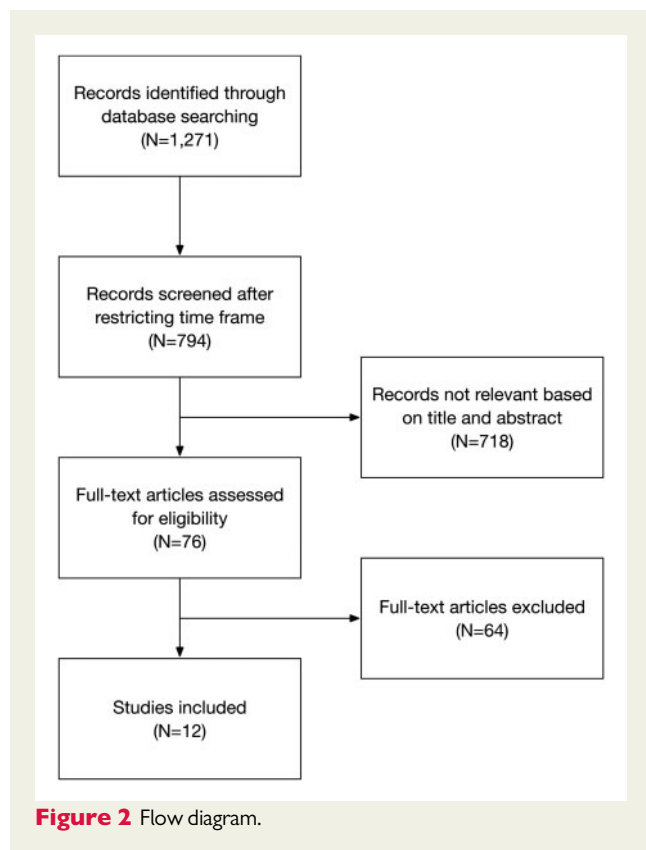
Our literature search returned 794 unique articles. After screening, 76 articles were deemed to be potentially eligible based on their titles and abstracts. After full-text review, 12 articles fulfilled the selection criteria and were included in this systematic review. The flow diagram for study selection is shown in [Figure 2](#) and study characteristics are shown in [Table 1](#).

### Studies of left ventricular function and morphology

Three studies used DL ECG analysis to detect LV systolic dysfunction, achieving an AUC of 0.89–0.93 and an accuracy of 98% ([Table 2](#)). Left ventricular systolic dysfunction was defined by echocardiography as an ejection fraction of <35% or <40% (depending on the study), with the ECG and echocardiogram having been done within 2 to 4 weeks of each other with no major changes in clinical status. Two of these studies compared different models and found that the predictive accuracy of neural network DL models was superior to other non-DL models and superior to expert interpretation by board-certified cardiologists. One study used DL ECG to detect left ventricular hypertrophy (LVH), achieving an AUC of 0.87 and an accuracy of 87%. No published study used DL ECG to detect LV chamber dilation.

### Studies of ischaemic heart disease

Six studies used DL ECG analysis to detect acute myocardial infarction and two to detect stable ischaemic heart disease, achieving an



**Figure 2** Flow diagram.

**Table 1** Characteristics of included studies

	Patients, N	Female, %	Age, mean (SD)	ECG, leads	ECG, seconds
Attia et al. <sup>10</sup>	52 870	43.0	61.8 (16.5)	12	10
Kwon et al. <sup>11</sup>	22 765	36.21	64.3 (14.2)	12	●
Li et al. <sup>12</sup>	573	●	●	1	2
Kwon et al. <sup>13</sup>	21 286	50.6	59.5	12	8
Liu et al. <sup>14</sup>	200	25.5	51.9	12 + 3	>30
Han and Shi <sup>15</sup>	165	25.5	51.9	12 + 3	>30
Acharya et al. <sup>16</sup>	200	25.5	51.9	12 + 3	>30
Liu et al. <sup>17</sup>	200	25.5	51.9	12 + 3	>30
Liu et al. <sup>18</sup>	290	●	●	4	>30
Tan et al. <sup>19</sup>	47	55.0	58	1	5
Acharya et al. <sup>20</sup>	47	55.3	58	1	5
Goto et al. <sup>21</sup>	243	●	●	12	10

ECG, electrocardiogram; SD, standard deviation; ●, not reported/available.

**Table 2** Left ventricular function and hypertrophy—performance of various deep learning models

	Model	AUC	Accuracy	Sensitivity	Specificity
LV systolic function					
Attia <i>et al.</i> <sup>10</sup>	CNN	0.932	●	86.0	86.0
Kwon <i>et al.</i> <sup>11</sup>	DNN	0.889	●	●	●
	RF	0.853	●	●	●
	LR	0.847	●	●	●
	CNN-RNN	●	97.6	96.3	97.4
Li <i>et al.</i> <sup>12</sup>	MLP	●	93.3	85.7	84.4
	RF	●	82.1	83.4	81.7
	CART	●	72.3	76.6	78.8
	SVM	●	66.0	73.3	61.2
LV hypertrophy					
Kwon <i>et al.</i> <sup>13</sup>	Combination NN	0.868	86.6	49.6	93.6
	RF	0.831	85.2	40.3	85.2
	LR	0.81	84.6	36.4	84.6
	ECG machine interpretation	0.679	85.1	34.5	93.6
	Expert interpretation	●	85.5	28.4	95.1

AUC, area under the curve; CART, classification and regression tree; CNN, convolutional neural network; DNN, deep neural network; ECG, electrocardiogram; LR, logistic regression; LV, left ventricular; MLP, multi-layer perceptron; NN, neural network; RF, random forest; RNN, recurrent neural network; SVM, support vector machine; ●, not reported/available.

AUC of 0.88–1.00 and accuracy of 83–99.9% (Table 3). Electrocardiograms were extracted from the Physikalisch-Technische Bundesanstalt (PTB) Database<sup>22</sup> and the St Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database.<sup>22</sup> Ischaemic heart disease was confirmed by coronary angiography and cardiac biomarkers alongside clinical history. The predictive accuracy of DL models was very high to detect both acute or stable forms of ischaemic heart disease, and furthermore, two studies demonstrated an accuracy of 99.81%<sup>14</sup> and 99.72%<sup>15</sup> to localize the territory of infarction.

## Discussion

This systematic review has highlighted the evidence on the accuracy and discrimination of DL models for the detection of structural cardiac pathologies. Our review has shown that DL models achieved a high degree of sensitivity and specificity for the detection of LV systolic dysfunction and acute or stable forms of ischaemic heart disease. A single study showed that DL achieved a high degree of specificity albeit low sensitivity to detect LVH. Our review has also shown that DL models appeared to outperform other common machine learning models such as support vector machines, random forests, and logistic regression. Thus, initiatives to implement machine learning into ECG analysis systems should consider DL as a favorable approach.

There are relevant use-cases for implementing DL ECG analysis alongside clinician interpretation in cardiovascular medicine; ranging from screening electrocardiography performed in routine primary care, preoperative evaluation, or competitive sports, to diagnostic electrocardiography performed in inpatient, outpatient, and pre-hospital settings. The use of DL ECG analysis may help clinicians and paramedical personnel uncover new cardiac pathologies that would

not otherwise have been suspected or that would have been diagnosed hours, days, or even weeks later after a specialist's assessment or an echocardiogram. Expedited diagnosis of cardiovascular disease translates to earlier initiation of treatment and better outcomes, while missed diagnosis translates to the opposite scenario. The use of DL ECG analysis goes beyond the clinical setting and may be employed in wearable and implantable devices allowing for continuous monitoring of health. This can improve the quality of care by allowing people, in particular older persons, to continue living independently at home while providing a means for the early identification of structural and electrical cardiac abnormalities.<sup>23,24</sup>

In patients presenting to the emergency department with signs and symptoms of HF, the odds of hospital mortality increased by 2.1% for every 4-h delay in diagnosis and 'door-to-furosemide time'.<sup>25</sup> In patients with Stage B HF, defined as structural heart disease without current or prior symptoms, 26% developed symptomatic Stage C HF and 40% died during an average follow-up of 5 years in the Framingham Study.<sup>26</sup> The risk of death or incident HF was reduced by 39% when asymptomatic patients were treated with the angiotensin-converting enzyme inhibitor enalapril in the Studies of Left Ventricular Dysfunction (SOLVD) trial.<sup>27</sup> HF affects 26 million people worldwide and 3.5 million new people every year,<sup>28</sup> however, effective population-based screening is still lacking.<sup>29,30</sup> Given that the number of persons living with Stage B HF is four times greater than Stages C and D combined, the potential benefits of screening to detect and treat asymptomatic LV systolic dysfunction, with tools such as DL ECG analysis, are considerable.<sup>31</sup>

With the rising global burden of hypertension and hypertensive heart disease, the population-level benefits of detecting LVH with DL ECG analysis are even greater. LVH is a haemodynamic manifestation of hypertensive end-organ damage and a risk factor for incident HF, stroke, and cardiovascular mortality.<sup>32</sup> In numerous studies, these

**Table 3** Ischaemic heart disease—performance of various deep learning models

	Model	Accuracy	AUC	Sensitivity	Specificity	
Acute MI						
	Liu et al. <sup>14</sup>	MFB-CNN	99.95	0.9998	99.97	99.90
	Han et al. <sup>15</sup>	ML-ResNet	99.92	1	99.98	99.77
	Acharya et al. <sup>16</sup>	CNN	93.5	●	93.71	92.83
	Liu et al. <sup>17</sup>	MFB-CBRNN	99.9	●	99.97	99.54
	Liu et al. <sup>18</sup>	CNN	96.0	●	95.40	97.37
	Goto et al. <sup>21</sup>	CNN-BLSTM	83	0.88	79	87
Stable IHD						
	Tan et al. <sup>19</sup>	Stacked CNN-LSTM	99.9	●	99.84	99.85
	Acharya et al. <sup>20</sup>	Deep CNN	95.1	●	91.13	95.88

AUC, area under the curve; CBRNN, convolutional bidirectional recurrent neural network; CNN, convolutional neural network; IHD, ischaemic heart disease; (B)LSTM, (bidirectional) long short-term memory; MFB, multi-feature branch; MI, acute myocardial infarction; ML-ResNet, multi-lead residual neural network; ●, not reported/available.

risks were reduced by early initiation of antihypertensive drugs (notably with angiotensin-converting enzyme inhibitors and angiotensin receptor blockers), weight loss, or dietary sodium reduction. Progressive LVH leads to diastolic dysfunction and, in turn, HF with preserved ejection fraction—a growing epidemic.<sup>33</sup> In addition to detecting LVH from the resting ECG signals, machine learning models have been developed to detect diastolic dysfunction as defined by Doppler echocardiography<sup>34–36</sup>; while these studies are of great interest, they did not strictly use end-to-end DL models and hence were not included in this systematic review. One of these studies used a hybrid model with a DL algorithm for feature engineering of the 12-lead ECG and a non-DL machine learning algorithm for the classification of structural cardiac pathologies.<sup>37</sup> This hybrid model yielded good discrimination for the detection of LVH (AUC 0.87), diastolic dysfunction (AUC 0.84), and hypertrophic cardiomyopathy (AUC 0.91).

In certain cases, LVH may be a phenotypic manifestation of hypertrophic cardiomyopathy, which is the most common cause of sudden cardiac death in competitive athletes<sup>38</sup>—a tragic outcome that can often be prevented by pre-emptive detection. Unfortunately, traditional ECG criteria are only 7–35% sensitive for mild LVH and 10–50% sensitive for moderate-to-severe LVH,<sup>39</sup> and echocardiography is not logistically feasible for all those at-risk. Adoption of DL ECG analysis as a screening modality for hypertrophic cardiomyopathy (AUC 0.91) or for LVH in general (AUC 0.87) could be justified as it performs similarly to other common screening modalities such as cervical cytology for cervical cancer (AUC 0.7), mammography for breast cancer (AUC 0.85), and prostate-specific antigen for prostate cancer (0.92).<sup>10</sup>

Traditional ECG criteria are imperfect for the diagnosis of acute or chronic presentation of ischaemic heart disease, underperforming as a gatekeeper to stress cardiac imaging and invasive cardiac catheterization. Factors associated with false positive or negative ECG interpretations include pre-existing conduction disturbances, early repolarization patterns, pacemakers, lateral infarcts, and less experienced reading clinicians. One study showed that the inter-rater reliability for the diagnosis of acute myocardial infarction was particularly poor when minimal clinical information was provided to the reading

clinicians, an increasingly common scenario in the era of telemedicine and pre-hospital activations.<sup>40</sup> The sensitivity and specificity for the pre-hospital diagnosis of acute myocardial infarction were shown to be 69% and 99% with rules-based computerized interpretation,<sup>41</sup> as compared with an average of 95% and 96% with DL-based interpretation in this review. The clinical implications should not be understated as ischaemic heart disease is a leading cause of death worldwide.<sup>42</sup> The ECG is an inexpensive and non-invasive test that can be coupled with DL models to assist clinicians in diagnosis; of special interest in developing countries where access to more costly testing is limited.

In this review, the highest reported accuracies were achieved with DL models that combined convolutional and other neural networks, effectively learning different types of functions in a single network model. One of the main strengths of end-to-end DL models is their ability to learn the discriminating features from complex and heterogeneous types of inputs (such as ECG signals or radiographic images) automatically without necessarily requiring the analyst to define, extract, and process the features of interest, also known as feature engineering. Non-DL machine learning models, on the other hand, require pre-processing and feature engineering, which is a multi-step process that has the potential to miss potentially informative features. There are numerous such features present on the resting ECG that reflect the structural and metabolic changes associated with LV dysfunction,<sup>10</sup> but many of these are subtle and not discernible to the human eye. Even after building a DL model, the nature of the informative features remains 'hidden', rendering it difficult for the clinician to understand or apply them in a non-computer-assisted interpretation.

Whereas high predictive accuracy is one of the main advantages of DL, low interpretability is one of its main disadvantages. Deep learning is sometimes referred to as a 'black box', wherein the user cannot comprehend precisely what the DL model is seeing (in terms of features) and how it is reaching a particular prediction. Conversely, traditional statistical modelling approaches tend to have lower predictive accuracy but higher interpretability, sometimes referred to as algorithmic transparency, wherein the user can gain insights into the specific features and their relative contributions to the final prediction. There are emerging DL techniques to enable (to some extent) the

**Table 4** Deep learning model architectures and data partitioning

	DL model	Input ECG format	Train and internal validation sets	External validation set
Attia et al. <sup>10</sup>	CNN: 6 layers of Conv + BN + MxP	12-lead ECG with 5000 data points in each lead as a 2D matrix of size 12 × 5000	50% of entire data: 80% for train and 20% for internal validation	50% of entire data: for test
Kwon et al. <sup>11</sup>	DNN: 5 hidden layers, 45 nodes, and dropout layers	10 ECG and demographic features as a 1D array	Data from Hospital 1: 80% for train and 20% for internal validation	Data from Hospital 2: for test
Li et al. <sup>12</sup>	CNN: 8 layers of Conv + MxP + FC RNN: 4 layers of LSTM + FC	Each ECG lead as 1D array	10-fold cross-validation	●
Kwon et al. <sup>13</sup>	ENN: CNN (6 layers of Conv + MxP) + DNN (5 layers and 56 nodes)	12-lead ECG with 4000 data points in each lead as a 2D matrix of size 12 × 4000 for CNN Demographic data and features from CNN as a 1D array for DNN	Data from Hospital A: 80% for train and 20% for internal validation	Data from Hospital B: for test
Liu et al. <sup>14</sup>	MFB-CNN: 7 layers of Conv + MP + FC	Each ECG lead as a 1D array, results combined in final FC layer	Five-fold cross-validation	●
Han et al. <sup>15</sup>	ML-ResNet: 12 feature branch of residual blocks + GAP + DropOut + Flatten + FC	Each ECG lead as a 1D array	Intra-patient scheme: five-fold cross-validation	Inter-patient scheme: 4740 controls and 10 721 patients for train, 2205 controls and 6491 patients for test
Acharya et al. <sup>16</sup>	CNN: 11 layers of Conv + MxP + FC	Lead II as a 1D array	90% of entire data: 70% for train and 30% for internal validation 10-fold cross-validation	10% of entire data: for test
Liu et al. <sup>17</sup>	MFB-CBRNN: 10 layers of Conv + BN + MP + GAP + LRM + BLSTM + FC	Each ECG lead as a 1D array, results combined in final FC layer	Five-fold cross-validation for class-based and subject-based experiments	●
Liu et al. <sup>18</sup>	CNN: 7 layers of Conv + LAP	4 selected ECG leads as a 2D matrix	Five-fold cross-validation	●
Tan et al. <sup>19</sup>	CNN-LSTM: 8 layers of Conv + MxP + LSTM + FC	Lead II as a 2D matrix of size 211 × 24	Approach 1: 10% of randomly selected data for train Approach 2: first 37.5% of controls and 43% of patients for train	Approach 1: 90% of randomly selected data for test Approach 2: first 62.5% of controls and 57% of patients for test
Acharya et al. <sup>20</sup>	Deep CNN: 11 layers of Conv + MxP + FC	Lead II as a 1D array	10-fold cross-validation	●
Goto et al. <sup>21</sup>	CNN-LSTM: 7 layers of Conv + BLSTM + Dense	12-lead ECG with 10 000 data points in each lead as a 2D matrix of size 12 × 10 000	249 urgent and 300 non-urgent revascularizations for train	113 urgent and 130 non-urgent revascularizations for test

*Types of Algorithms:* CNN, convolutional neural network; DNN, deep neural network; RNN, recurrent neural network; LSTM: Long-Short Term Memory, a type of RNN with backward feedback loops; BLSTM: Bilateral Long-Short Term Memory, a type of RNN with backward and forward feedback loops; MFB-CBRNN: Multiple Feature Branch Convolutional Bidirectional Recurrent Neural Network; ENN: Ensemble Neural Network, more than one type of neural network used on the same data.

*Types of Layers:* BN: Batch-Normalization layer, used to normalize the batches of input ECG data; Conv: Convolutional layer, used to extract the various features from the input ECG data; MP: Mean-Pooling layer, used to downsample the mean values from sections of the feature map; MxP: Max-Pooling layer, used to down sample the maximum values from sections of the feature map; GAP: Global Average Pooling layer, used to downsample the mean values from the full feature map; LAP: Lead Asymmetric Pooling layer used to downsample a multiscale feature map; FC: Fully Connected layer, used to compile values extracted by previous layers to classify the output(s); LRM: Lead Random Mask layer, used to randomly drop feature branches in the training phase.

interpretation of model predictions.<sup>43–45</sup> While a detailed discussion of these techniques is beyond the scope of this review, one interesting study in this field generated images of synthetic ECG signals corresponding to the condition of interest (hyperkalaemia) in order to visually illustrate the model's predictive features (widened QRS complexes, peaked T waves, etc.).<sup>44</sup>

It is worth noting that the pre-processing steps taken for the preparation of ECG signals before feeding them into DL models were most frequently upsampling and downsampling, denoising, regularization, normalization, and segmentation. The reviewed papers mostly employed z-score transformation,<sup>12,16,17,19</sup> Pan-Tompkins QRS-wave detection algorithm,<sup>14,16,17</sup> fuzzy information granulation,<sup>18</sup> or discrete wavelet transform<sup>20</sup> for the aforementioned pre-processing steps. The modelling steps summarized in [Table 4](#) were equally variable in terms of number and type of layers. Some papers<sup>15,16</sup> employed state-of-the-art architectures for feature extraction; others employed methods such as grid search<sup>13</sup> to determine the optimal number of layers and nodes for their respective models to maximize accuracy according to their dataset and testing method. These methodological differences show that the approach for model selection and optimization remains an open topic in the field of DL ECG signal analysis.

As opposed to the other fields of DL such as computer vision where there are multiple large-scale annotated image datasets (e.g. ImageNet, Open Images), there are relatively few publicly available annotated ECG datasets. Even then, annotated labels typically span a narrow range of cardiovascular changes and diagnoses. Accordingly in our review, five studies used local hospital-based ECG datasets to train and test their proposed model,<sup>10–13,21</sup> five studies used the publicly available PTB ECG dataset,<sup>14–18</sup> and two studies used a combination of ECG datasets from PhysioNet, Fantasia, and St-Petersburg Institute of Cardiology Technics.<sup>19,20</sup> Expanding the volume and depth of publicly available annotated ECG datasets would appear to be a priority to equip researchers with the source data needed to catalyse further research efforts, ultimately leading to improvements in predictive accuracy and reliability.

There are limitations that merit discussion. First, a number of studies particularly in the field of ischaemic heart disease used databases that consisted of ECGs from a singular hospital system or narrow patient population. External validation in geographically diverse multi-centre populations with multi-vendor ECG systems would be crucial for generalizability. Second, few studies provided direct head-to-head comparisons against traditional rules-based computer programs or expert interpretations. Extrapolating from historical studies suggests that DL would likely outperform them, since expert cardiologists achieved a pooled 75% accuracy for detecting ECG pathologies,<sup>46</sup> and rules-based computer programs achieved 57% sensitivity for detecting LV hypertrophy and 59–77% for detecting myocardial infarction.<sup>47,48</sup> Third, there was study-to-study variability in technical ECG acquisition in terms of the number of leads and the duration of recording, and it is unclear to what extent these parameters may or may not influence the performance of the DL models. From a clinical standpoint, the 12-lead 10-second resting ECG is of specific interest given that this is the current standard of care in most centres. Finally, implementation of the DL models was not a focal point of the reviewed studies. Further research is needed to

determine the effect of these DL models on clinical decision-making and ultimately patient outcomes.

## Conclusions

When applied to the analysis of resting ECG signals, DL models achieve a high degree of accuracy and (inherent) reliability in detecting LV systolic dysfunction, LVH, and acute or chronic forms of ischaemic heart disease. Deep learning models appear to outperform traditional computerized interpretations and non-DL machine learning models. Gains in predictive performance could translate to earlier diagnosis of symptomatic cardiovascular pathologies and pre-emptive detection of asymptomatic ones. Enhanced screening with DL ECG has the potential to shift the emphasis towards the prevention of cardiovascular disease and its complications by early detection of at-risk groups. While current screening and diagnostic pathways rely on resource-intensive imaging tests and biomarkers, implementation of DL to a widely available tool like the ECG could help provide an accessible front-line option to assist clinicians in caring for their patients.

## Acknowledgements

The authors would like to thank Dr Eli Segal for his diligent review of this manuscript.

## Funding

Dr. Afilalo is supported by the Canadian Institutes of Health Research and the Fonds de recherche en santé du Québec.

**Conflict of interest:** none declared.

## References

- Schlant RC, Adolph RJ, DiMarco JP, Dreifus LS, Dunn MI, Fisch C, Garson A Jr, Haywood LJ, Levine HJ, Murray JA. Guidelines for electrocardiography. A report of the American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Committee on Electrocardiography). *Circulation* 1992;**85**:1221–1228.
- Sansone M, Fusco R, Pepino A, Sansone C. Electrocardiogram pattern recognition and analysis based on artificial neural networks and support vector machines: a review. *J Healthc Eng* 2013;**4**:465–504.
- Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;**25**:65–69.
- Guglin ME, Thatai D. Common errors in computer electrocardiogram interpretation. *Int J Cardiol*. 2006;**106**:232–237.
- Poon K, Okin PM, Kligfield P. Diagnostic performance of a computer-based ECG rhythm algorithm. *J Electrocardiol* 2005;**38**:235–238.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–444.
- Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, Barengo NC, Beaton AZ, Benjamin EJ, Benziger CP, Bonny A, Brauer M, Brodmann M, Cahill TJ, Carapetis J, Catapano AL, Chugh SS, Cooper LT, Coresh J, Criqui M, DeCleene N, Eagle KA, Emmons-Bell S, Feigin VL, Fernández-Solá J, Fowkes G, Gakidou E, Grundy SM, He FJ, Howard G, Hu F, Inker L, Karthikeyan G, Kassebaum N, Koroshetz W, Lavie C, Lloyd-Jones D, Lu HS, Mirijello A, Temesgen AM, Mokdad A, Moran AE, Muntner P, Narula J, Neal B, Ntsekhe M, Moraes de Oliveira G, Otto C, Owolabi M, Pratt M, Rajagopalan S, Reitsma M, Ribeiro ALP, Rigotti N, Rodgers A, Sable C, Shakil S, Sliwa-Hahnle K, Stark B, Sundström J, Timpel P, Tleyjeh IM, Valgimigli M, Vos T, Whelton PK, Yacoub M, Zuhlke L, Murray C, Fuster V; GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global burden of cardiovascular diseases and risk factors, 1990–2019. *J Am Coll Cardiol* 2020;**76**:2982–3021.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;**339**:b2535.

9. Wikipedia contributors. Deep learning. In Wikipedia, The Free Encyclopedia. [https://en.wikipedia.org/w/index.php?title=Deep\\_learning&oldid=1026572762](https://en.wikipedia.org/w/index.php?title=Deep_learning&oldid=1026572762) (12 June 2021).
10. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;**25**:70–74.
11. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM, Song PS, Park J, Choi RK, Oh BH. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J* 2019;**49**:629–639.
12. Li D, Li X, Zhao J, Bai X. Automatic staging model of heart failure based on deep learning. *Biomed Signal Process Control* 2019;**52**:77–83.
13. Kwon JM, Jeon KH, Kim HM, Kim MJ, Lim SM, Kim KH, Song PS, Park J, Choi RK, Oh BH. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace* 2020;**22**:412–419.
14. Liu W, Huang Q, Chang S, Wang H, He J. Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram. *Biomed Signal Process Control* 2018;**45**:22–32.
15. Han C, Shi L. ML-ResNet: a novel network to detect and locate myocardial infarction using 12 leads ECG. *Comput Methods Programs Biomed* 2020;**185**:105138.
16. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf Sci* 2017;**415–416**:190–198.
17. Liu W, Wang F, Huang Q, Chang S, Wang H, He J. MFB-CBRNN: a hybrid network for MI detection using 12-lead ECGs. *IEEE J Biomed Health Inform* 2020;**24**:503–514.
18. Liu W, Zhang M, Zhang Y, Liao Y, Huang Q, Chang S, Wang H, He J. Real-time multilead convolutional neural network for myocardial infarction detection. *IEEE J Biomed Health Inform* 2018;**22**:1434–1444.
19. Tan JH, Hagiwara Y, Pang W, Lim I, Oh SL, Adam M, Tan RS, Chen M, Acharya UR. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput Biol Med* 2018;**94**:19–26.
20. Acharya UR, Fujita H, Lih OS, Adam M, Tan JH, Chua CK. Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowledge-Based Syst* 2017;**132**:62–71.
21. Goto S, Kimura M, Katsumata Y, Goto S, Kamatani T, Ichihara G, Ko S, Sasaki J, Fukuda K, Sano M. Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients. *PLoS One* 2019;**14**:e0210103.
22. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000;**101**:E215–E220.
23. Luz EJD, Schwartz WR, Cámara-Chávez G, Menotti D. ECG-based heartbeat classification for arrhythmia detection: a survey. *Comput Methods Programs Biomed* 2016;**127**:144–164.
24. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep learning for health informatics. *IEEE J Biomed Health Inform* 2017;**21**:4–21.
25. Matsue Y, Damman K, Voors AA, Kagiama N, Yamaguchi T, Kuroda S, Okumura T, Kida K, Mizuno A, Oishi S, Inuzuka Y, Akiyama E, Matsukawa R, Kato K, Suzuki S, Naruke T, Yoshioka K, Miyoshi T, Baba Y, Yamamoto M, Murai K, Mizutani K, Yoshida K, Kitai T. Time-to-furosemide treatment and mortality in patients hospitalized with acute heart failure. *J Am Coll Cardiol* 2017;**69**:3042–3051.
26. Wang TJ, Evans JC, Benjamin EJ, Levy D, LeRoy EC, Vasan RS. Natural history of asymptomatic left ventricular systolic dysfunction in the community. *Circulation* 2003;**108**:977–982.
27. Yusuf S, Pitt B, Davis CE, Hood WB Jr, Cohn JN. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N Engl J Med* 1992;**327**:685–691.
28. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, Nodari S, Lam CSP, Sato N, Shah AN, Gheorghide M. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol* 2014;**63**:1123–1133.
29. McDonagh TA, McDonald K, Maisel AS. Screening for asymptomatic left ventricular dysfunction using B-type natriuretic Peptide. *Congest Heart Fail* 2008;**14**(4 Suppl. 1):5–8.
30. Redfield MM, Rodeheffer RJ, Jacobsen SJ, Mahoney DW, Bailey KR, Burnett JC Jr. Plasma brain natriuretic peptide to detect preclinical ventricular systolic or diastolic dysfunction: a community-based study. *Circulation* 2004;**109**:3176–3181.
31. Goldberg LR, Jessup M. Stage B heart failure: management of asymptomatic left ventricular systolic dysfunction. *Circulation* 2006;**113**:2851–2860.
32. Oparil S, Acelajado MC, Bakris GL, Berlowitz DR, Cifková R, Dominiczak AF, Grassi G, Jordan J, Poulter NR, Rodgers A, Whelton PK. Hypertension. *Nat Rev Dis Primers* 2018;**4**:18014.
33. Yoon S, Eom GH. Heart failure with preserved ejection fraction: present status and future directions. *Exp Mol Med* 2019;**51**:1–9.
34. Kagiama N, Piccirilli M, Yamamala N, Farjo PD, Casacang-Verzosa G, Tarhuni WM, Nezarat N, Budoff MJ, Narula J, Sengupta PP. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. *J Am Coll Cardiol* 2020;**76**:930–941.
35. Sabovic F, Cauwenberghs N, Kouznetsov D, Haddad F, Alonso-Betanzos A, Vens C, Kuznetsova T. Applying machine learning to detect early stages of cardiac remodelling and dysfunction. *Eur Heart J Cardiovasc Imaging* 2020.
36. Sengupta PP, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. *J Am Coll Cardiol* 2018;**71**:1650–1660.
37. Tison GH, Zhang J, Dellling FN, Deo RC. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ Cardiovasc Qual Outcomes* 2019;**12**:e005289.
38. Emery MS, Kovacs RJ. Sudden cardiac death in athletes. *JACC Heart Fail* 2018;**6**:30–40.
39. Ang D, Lang C. The prognostic value of the ECG in hypertension: where are we now? *J Hum Hypertens* 2008;**22**:460–467.
40. McCabe JM, Armstrong EJ, Ku I, Kulkarni A, Hoffmayer KS, Bhavne PD, Waldo SW, Hsue P, Stein JC, Marcus GM, Kinlay S, Ganz P. Physician accuracy in interpreting potential st-segment elevation myocardial infarction electrocardiograms. *J Am Heart Assoc* 2013;**2**:e000268.
41. de Champlain F, Boothroyd LJ, Vadeboncoeur A, Huynh T, Nguyen V, Eisenberg MJ, Joseph L, Boivin JF, Segal E. Computerized interpretation of the prehospital electrocardiogram: predictive value for ST segment elevation myocardial infarction and impact on on-scene time. *CJEM* 2014;**16**:94–105.
42. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, Ahmed M, Aksut B, Alam T, Alam K, Alla F, Alvis-Guzman N, Amrock S, Ansari H, Ärnlöv J, Asayesh H, Atey TM, Avila-Burgos L, Awasthi A, Banerjee A, Barac A, Barnighausen T, Barregard L, Bedi N, Ketema EB, Bennett D, Berhe G, Bhutta Z, Bitew S, Carapetis J, Carrero JJ, Malta DC, Castañeda-Orjuela CA, Castillo-Rivas J, Catalá-López F, Choi JY, Christensen H, Cirillo M, Cooper L Jr, Criqui M, Cundiff D, Damasceno A, Dandona L, Dandona R, Davletov K, Dharmaratne S, Dorairaj P, Dubey M, Ehrenkranz R, Zaki MES, Faraon EJA, Esteghamati A, Farid T, Farvid M, Feigin V, Ding EL, Fowkes G, Gebrehiwot T, Gillum R, Gold A, Gona P, Gupta R, Habtewold TD, Hafezi-Nejad N, Hailu T, Hailu GB, Hankey G, Hassen HY, Abate KH, Havmoeller R, Hay SI, Horino M, Hotez PJ, Jacobsen K, James S, Javanbakht M, Jeemon P, John D, Jonas J, Kalkonde Y, Karimkhandi C, Kasaian A, Khader Y, Khan A, Khang YH, Khera S, Khoja AT, Khubchandani J, Kim D, Kolte D, Kosen S, Krohn KJ, Kumar GA, Kwan GF, Lal DK, Larsson A, Linn S, Lopez A, Lotufo PA, El Razek HMA, Malekzadeh R, Mazidi M, Meier T, Meles KG, Mensah G, Meretoja A, Mezgebe H, Miller T, Mirrakhimov E, Mohammed S, Moran AE, Musa KI, Narula J, Neal B, Ngallesoni F, Nguyen G, Obermeyer CM, Owolabi M, Patton G, Pedro J, Qato D, Qorbani M, Rahimi K, Rai RK, Rawaf S, Ribeiro A, Safiri S, Salomon JA, Santos I, Milicevic MS, Sartorius B, Schutte A, Sepanlou S, Shaikh MA, Shin MJ, Shishehbor M, Shore H, Silva DAS, Sobngwi E, Stranges S, Swaminathan S, Tabarés-Seisdedos R, Atanfu NT, Tesfay F, Thakur JS, Thrift A, Topor-Madry R, Truelsen T, Tyrovolas S, Ukwaja KN, Uthman O, Vasankari T, Vlassov V, Vollset SE, Wakayo T, Watkins D, Weintraub R, Werdecker A, Westerman R, Wiysonge CS, Wolfe C, Workicho A, Xu G, Yano Y, Yip P, Yonemoto N, Younis M, Yu C, Vos T, Naghavi M, Murray C. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol* 2017;**70**:1–25.
43. Erhan D, Bengio Y, Courville AC, Vincent P. *Visualizing Higher-Layer Features of a Deep Network*. 2009. Technical Report 1341, Département d'Informatique et Recherche Opérationnelle.
44. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, Ackerman MJ, Noseworthy PA. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol* 2019;**4**:428–436.
45. Olah C, Mordvintsev A, Schubert L. Feature visualization. *Distill* 2017;**2**.
46. Cook DA, Oh SY, Pusic MV. Accuracy of physicians' electrocardiogram interpretations: a systematic review and meta-analysis. *JAMA Intern Med* 2020;**180**:1–11.
47. Schläpfer J, Wellens HJ. Computer-interpreted electrocardiograms: benefits and limitations. *J Am Coll Cardiol* 2017;**70**:1183–1192.
48. Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Gehring J, Graham I, van Herpen G, Machado H, Macfarlane PW, Michaelis J, Mouloupoulos SD, Rubel P, Zywiec C. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991;**325**:1767–1773.