

---

Review

# Machine Learning in Left Ventricular Hypertrophy Detection: Systematic Review and Meta-Analysis

---

Yilin Li<sup>1</sup>, MD; Ke Zhao<sup>2</sup>, MM; Jing Wu<sup>1</sup>, BS

<sup>1</sup>Department of Geriatrics, The Third People's Hospital of Chengdu, Chengdu, Sichuan, China

<sup>2</sup>Department of Critical Care Medicine, Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan, China

**Corresponding Author:**

Jing Wu, BS

Department of Geriatrics

The Third People's Hospital of Chengdu

82 Qinglong Street, Qingyang District, Chengdu, Sichuan Province, China, 610031

Chengdu, Sichuan

China

Phone: 86 15881707332

Email: [wjing0103@163.com](mailto:wjing0103@163.com)

---

## Abstract

**Background:** In recent years, researchers have investigated machine learning (ML)-based approaches for the detection of left ventricular hypertrophy (LVH). However, the accuracy of ML in detecting LVH varies across different modeling variables and models. Systematic evidence is lacking in understanding how different ML approaches affect LVH detection accuracy.

**Objective:** The aim of this study is to systematically assess the diagnostic accuracy of these ML approaches to inform the development of artificial intelligence tools.

**Methods:** PubMed, Embase, Cochrane Library, and Web of Science were comprehensively searched up to November 12, 2025. The Prediction Model Risk of Bias Assessment Tool was used to evaluate the risk of bias. Subgroup analyses were performed based on ML model types and modeling variables (electrocardiogram [ECG], clinical features, and echocardiography). Only diagnostic 2×2 tables from validation sets were pooled for meta-analysis, with all statistical analyses performed using Stata.

**Results:** A total of 25 studies were included in the analysis. The performance of ML models varied with input data types and algorithms. A meta-analysis showed that ECG-based models, in comparison, exhibited a sensitivity of 0.76 (95% CI 0.66-0.84) and a specificity of 0.84 (95% CI 0.78-0.89). Echocardiography-based models had a sensitivity ranging from 0.71 to 0.94 and a specificity ranging from 0.67 to 0.96. The models based on clinical features had a sensitivity of 0.78 (95% CI 0.69-0.85) and a specificity of 0.71 (95% CI 0.65-0.76). A subgroup analysis of the ECG-based models revealed that the deep learning model produced a sensitivity of 0.71 (95% CI 0.60-0.80) and a specificity of 0.79 (95% CI 0.65-0.88).

**Conclusions:** ML demonstrates reasonably high accuracy in detecting LVH. However, these conclusions are derived from limited evidence. Meanwhile, the extreme heterogeneity reported in the meta-analysis requires more critical interpretation. Current conclusions regarding model accuracy should be interpreted with caution. Therefore, future research should focus on constructing high-performance ML models based on imaging data for LVH diagnosis.

*J Med Internet Res* 2026;28:e76637; doi: [10.2196/76637](https://doi.org/10.2196/76637)

**Keywords:** artificial intelligence; AI; machine learning; left ventricular hypertrophy; meta-analysis; deep learning; electrocardiogram; ECG; echocardiography; cardiovascular risk

---

## Introduction

Left ventricular hypertrophy (LVH) represents an increase in left ventricular mass driven by various cardiovascular risk factors [1], and it confers a 5-10 times greater risk of developing cardiovascular disease [2,3]. Accumulating

evidence demonstrates that LVH is an independent predictor of cardiovascular risk [4-6]. However, the pathological and physiological alterations induced by LVH occur without overt symptoms or clinical signs [1]. Therefore, the early detection of LVH through sensitive screening methods is imperative,

as it holds significant positive implications for patient health outcomes.

Currently, echocardiography (ECHO) and electrocardiogram (ECG) are the conventional diagnostic methods for detecting LVH [7,8]. While ECHO is a relatively accurate approach [9], it is costly and requires specialized equipment and trained operators, which poses challenges for early diagnosis [9,10]. Studies also show high rates of disagreement in ECHO readings among cardiologists, sometimes as much as 42.1% [11]. Cardiologists may mistakenly read ECHOs, particularly when tired, possibly resulting in misdiagnoses and misinterpretations [12,13]. Machine learning (ML) has garnered increasing attention in the medical field over the past decade, particularly in cardiovascular medicine [14-16]. ML is important since it can discern critical features from complex datasets [17].

Given its capacity to process intricate, high-volume health data, ML has ascended as an indispensable tool for developing predictive models in medicine [17,18]. In this context, researchers have begun constructing ML-based approaches for diagnosing LVH.

Nonetheless, there are various models and modeling variables for constructing ML models [17,19,20]. Multiple

ML models have been proposed for detecting LVH [1,10,21,22], yet their comparative diagnostic accuracy remains debated. Robust evidence assessing both the accuracy and real-world implementation of ML models for LVH identification remains insufficient. Therefore, a systematic review and meta-analysis was conducted in this study to evaluate the performance of ML in detecting LVH and furnish empirical support for artificial intelligence (AI) development in this field.

## Methods

### Study Registration

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was prospectively registered in PROSPERO (registration ID: CRD42024617183) [23].

### Eligibility Criteria

Detailed inclusion and exclusion criteria were formulated for this review. To improve the visualization effect, these standards were presented in tabular form (Textbox 1).

#### Textbox 1. Inclusion and exclusion criteria.

##### Inclusion criteria

- Population: individuals suspected of having left ventricular hypertrophy
- Intervention: machine learning (ML) models for left ventricular hypertrophy identification
- Comparison: no diagnostic tool was used as a comparator
- Outcome: quantitative diagnostic performance metrics, including C-statistic, sensitivity, specificity, accuracy, recall, precision, confusion matrices, diagnostic 2x2 tables,  $F_1$ -score, or calibration curves
- Study design: case-control, cohort, or cross-sectional studies published in English

##### Exclusion criteria

- Intervention: studies that only examined differential factors without developing complete ML models
- Outcomes: studies lacking the following metrics for ML performance: C-statistic, sensitivity, specificity, accuracy, recall, precision, confusion matrices, diagnostic 2x2 tables,  $F_1$ -score, or calibration curves
- Study design: meta-analyses, reviews, guidelines, expert opinions, and unpublished conference abstracts

### Data Sources and Search Strategy

PubMed, Embase, Cochrane Library, and Web of Science were systematically searched up to November 12, 2025, using both Medical Subject Headings (MeSH) terms and free-text words. Search terms included combinations of “artificial intelligence,” “deep learning,” “machine learning,” “random forest,” “support vector machine,” and “left ventricular hypertrophy.” No filters were applied to constrain studies by publication year or geographic region. The details on the search strategy are available in [Multimedia Appendix 1](#).

### Study Selection and Data Extraction

All identified records were uploaded to EndNote, and duplicates were filtered out. The titles and abstracts of the remaining studies were screened for potentially relevant studies. Full texts of the selected studies were obtained and evaluated for eligibility. An electronic extraction sheet was prepared in advance to record author information, publication year, country of the author, study design, patient

sources, generation method of validation set, model types, and modeling variables. The entire screening process and data extraction were performed independently by 2 investigators, followed by cross-verification. Discrepancies were addressed through discussion with a third researcher.

### Risk of Bias in Studies

The Prediction Model Risk of Bias Assessment Tool (PROBAST) was leveraged to appraise the risk of bias (RoB) in the included original studies [24]. The evaluation covers 4 domains: participants, predictor variables, outcomes, and statistical analysis, which collectively determine the overall RoB and applicability. Each domain contains specific signaling questions (2, 3, 6, and 9 questions, respectively), with 3 possible responses: “yes/probably yes” (low RoB), “no or probably no” (high RoB), and “no information” (unclear). Two researchers independently conducted the PROBAST assessments, followed by cross-checking. Any dissents were addressed through discussion with a third researcher.

### Synthesis Methods

Some studies had multiple validation sets. When both an internal and an external validation set were available, only the results from the external set were retained. If an external validation set is unavailable, the results from the internal validation set with optimal performance were selected.

A meta-analysis was conducted on the validation set, leveraging diagnostic 2x2 tables. For studies without 2x2 tables, sensitivity, specificity, positive predictive value, and accuracy, combined with case numbers, were used for estimation (formulas 1-5). The bivariate mixed-effects model was used to pool diagnostic metrics (sensitivity, specificity, positive likelihood ratio [PLR], negative likelihood ratio [NLR], diagnostic odds ratio [DOR], and summary receiver operating characteristic [SROC]). Deek funnel plot was used for assessing publication bias [25], and Fagan nomogram was adopted for evaluating clinical utility [26]. Subgroup analyses were carried out based on modeling variables and model types. A significance threshold of  $P < .05$  was established.

$$TP = Sensitivity \times Events \quad (1)$$

$$TN = Specificity \times (sample\ size - Events) \quad (2)$$

$$FN = Events - TP \quad (3)$$

$$FP = samplesize - Events - TN \quad (4)$$

OR

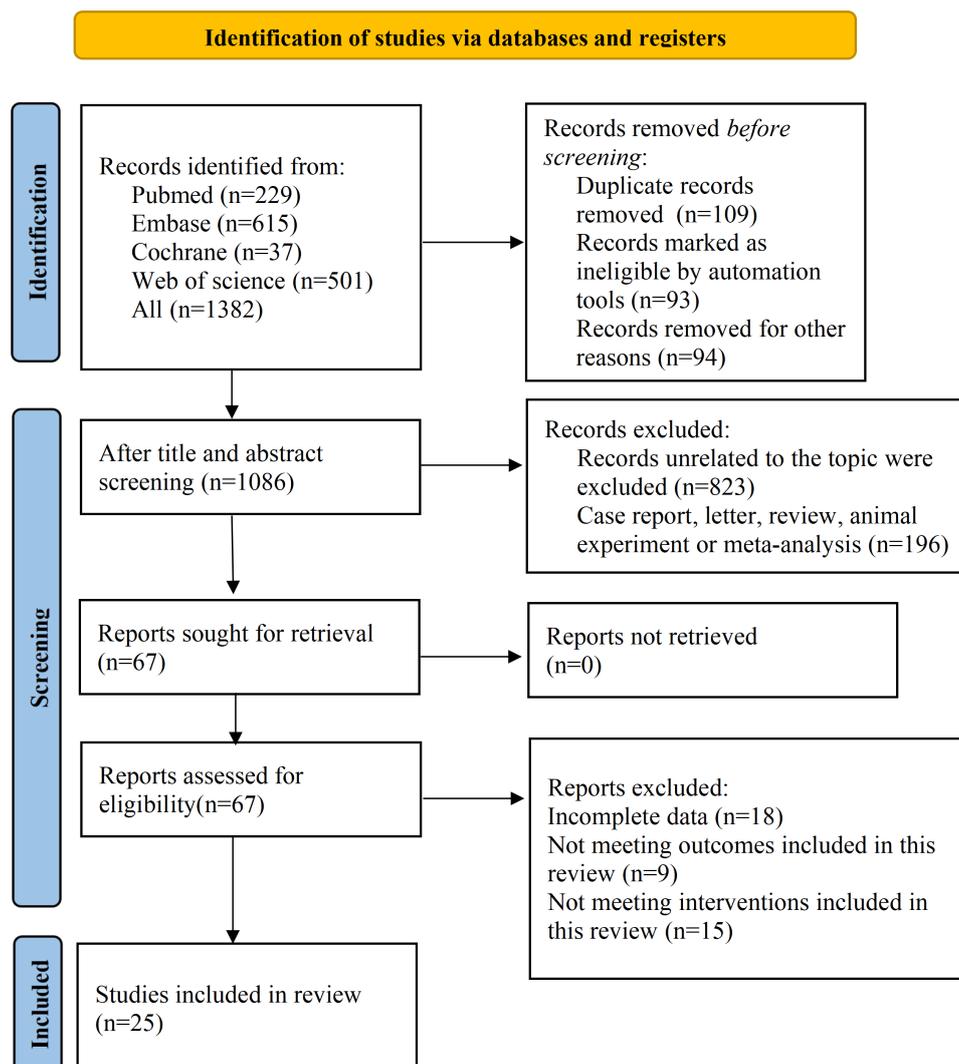
$$FP = TP/precision - TP \quad (5)$$

## Results

### Study Selection

A total of 1382 records were searched. After removing duplicates, 1086 records remained. Then, 1019 records were excluded through title and abstract screening, leaving 67 records for full-text review. Of those, 42 reports were removed due to incomplete data, incompatible outcomes, and incompatible interventions. Ultimately, 25 studies were selected for analysis (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for literature screening.



### Study Characteristics

Of the 25 case-control studies included (2018-2025) [1,10, 21,22,27-47], 19 used single-center designs, and 6 adopted multicenter approaches (including 4 studies based on databases that leveraged the UK Biobank resource). These studies were conducted across diverse regions: 10 studies in China, 6 in the United States, 2 in Japan, 2 in Korea, 1 in Mexico, 1 in the United Kingdom, 1 in Greece, 1

in Italy, and 1 in the United Arab Emirates. ECG-based models were included in 16 studies, clinical feature-based models were included in 5 studies, and ECHO-based models were included in 4 studies. Six studies performed external validation, including 5 ECG-based models and 1 clinical feature-based model. The main features of these studies are summarized in Table 1.

**Table 1.** Basic features of the included studies.

Author and publication year	Country of author	Study design	Patient source	Validation set generation method	Model type	Modeling variable	Reference standard
Ryu et al [22] (2023)	Korea	Case-control study	Single center	RS <sup>a</sup>	DL <sup>b</sup>	12-lead ECG <sup>c</sup>	ECHO <sup>d</sup>
Farhad et al [28] (2023)	United Arab Emirates	Case-control study	LVH <sup>e</sup> database	RS 10 folds	DL	ECHO	Clinical consensus
Zhao et al [32] (2022)	China	Case-control study	Single center	Folds RS	DL	12-lead ECG	Clinical consensus
Liu et al [33] (2022)	China	Case-control study	Multicenter	Folds EV <sup>f</sup>	DL	12-lead ECG	Clinical consensus
De la Garza Salazar et al [45] (2021)	Mexico	Case-control study	Single center	RS EV	DT <sup>g</sup>	12-lead ECG	Clinical consensus
Kwon et al [40] (2020)	Korea	Case-control study	Multicenter	RS EV	DL	12-lead ECG	Clinical consensus
Ghorbani et al [39] (2020)	United States	Case-control study	Single center	RS	DL	ECHO	Clinical consensus
Sparapani et al [43] (2019)	United States	Case-control study	The MESA <sup>h</sup> database	RS	DT	12-lead ECG	CMR <sup>i</sup>
Madani et al [44] (2018)	United States	Case-control study	Single center	RS	DL	ECHO	Clinical consensus
Yuan et al [29] (2023)	China	Case-control study	Single center	RS	LR <sup>j</sup>	Clinical features	Clinical consensus
Dwivedi et al [30] (2023)	United States	Case-control study	Single center	RS	RF <sup>k</sup>	12-lead ECG	ECHO
Yu et al [35] (2022)	China	Case-control study	Single center	RS	DL	ECHO	Clinical consensus
Wu et al [34] (2022)	China	Case-control study	Single center	EV	LR	Clinical features	Clinical consensus
Kokubo et al [36] (2022)	Japan	Case-control study	Single center	RS	DL	12-lead ECG	Clinical consensus
Ye et al [37] (2021)	China	Case-control study	Single center	10 folds <sup>l</sup>	LASSO <sup>m</sup> regression	Clinical features	Clinical consensus
Angelaki et al [38] (2021)	Greece	Case-control study	Single center	RS	RF	12-lead ECG	Clinical consensus
Lin and Liu [41] (2020)	United States	Case-control study	Single center	RS 4 folds	SVM <sup>o</sup>	12-lead ECG	ECHO
Tison et al [42] (2019)	United States	Case-control study	Single center	RS	GBM <sup>n</sup>	12-lead ECG	ECHO
Liu et al [31] (2023)	China	Case-control study	Single center	RS	ANN <sup>p</sup>	12-lead ECG	Clinical consensus
Zhang et al [1] (2024)	China	Case-control study	Single center	RS	LR	Clinical features	Clinical consensus
Wan et al [27] (2024)	China	Case-control study	Single center	RS	LR	Clinical features	Clinical consensus

Author and publication year	Country of author	Study design	Patient source	Validation set generation method	Model type	Modeling variable	Reference standard
Naderi et al [21] (2024)	United Kingdom	Case-control study	UK Biobank Database+ Single center	RS EV	SVM	12-lead ECG	CMR
Cai et al [10] (2024)	Japan	Case-control study	Single center	RS	DL	12-lead ECG	Clinical consensus
Huang et al [46] (2025)	China	Case-control study	Single center	RS	CatBoost	12-lead ECG	Clinical consensus
Taconné et al [47] (2025)	Italy	Case-control study	PTB-XL ECG database+ Georgia 12-lead ECG Challenge Database	10 folds EV	SVM	12-lead ECG	Clinical consensus

<sup>a</sup>RS: random sampling.

<sup>b</sup>DL: deep learning.

<sup>c</sup>ECG: electrocardiography.

<sup>d</sup>ECHO: echocardiography.

<sup>e</sup>LVH: left ventricular hypertrophy.

<sup>f</sup>EV: external validation.

<sup>g</sup>DT: decision tree.

<sup>h</sup>MESA: Multi-Ethnic Study of Atherosclerosis.

<sup>i</sup>CMR: cardiac magnetic resonance.

<sup>j</sup>LR: logistic regression.

<sup>k</sup>RF: random forest.

<sup>l</sup>10 folds: 10-fold cross-validation.

<sup>m</sup>LASSO: least absolute shrinkage and selection operator.

<sup>n</sup>GBM: gradient boosted machine.

<sup>o</sup>SVM: support vector machine.

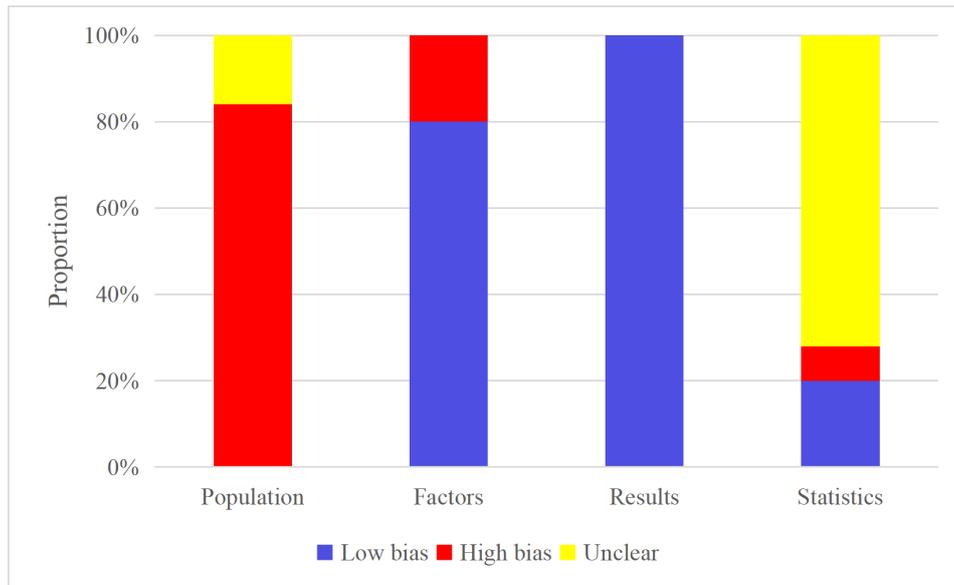
<sup>p</sup>ANN: artificial neural network.

## RoB in Studies

A high RoB was identified in 21 case-control studies (single- and multi-center) due to potential selection bias. An unclear RoB was assigned to 4 studies based on public databases. The 21 case-control studies were rated as having a high RoB due to selection bias. Due to inherent selection biases, such as the “healthy volunteer” effect commonly found in biobanks, the 4 studies based on public databases were rated as having an unclear RoB. Five case-control studies utilized clinical features for modeling and were rated as having a high RoB due to case-control influence during data collection. The remaining studies, which used image-based ML models, were assessed as having a low RoB in terms of predictors. All studies used valid outcome definitions, and the avoidance of predictor variables in model construction ensured a low RoB. For the imaging-based validation set, the accurate calculation of events per variable (EPV)

proved challenging, leading to an uncertain RoB. Developers typically rely on EPV, especially EPV of 10 higher, to determine the minimum sample size required and the maximum number of candidate predictions that can be tested [48]. However, when building models based on images, the basic unit of validation data is the image, not the traditional “patient case.” Therefore, we cannot calculate the sample size that meets the EPV requirements based on the number of patient cases, making it difficult to directly apply this standard in such studies. In contrast, studies using clinical features for modeling consistently met the EPV>10 criterion and thus were classified as low RoB. In 2 studies, data were segmented according to images, and patients’ images were shared between the training set and the validation set, which may lead to data leakage, and thus were classified as high RoB (Figure 2).

**Figure 2.** Risk of bias of the included machine learning models.



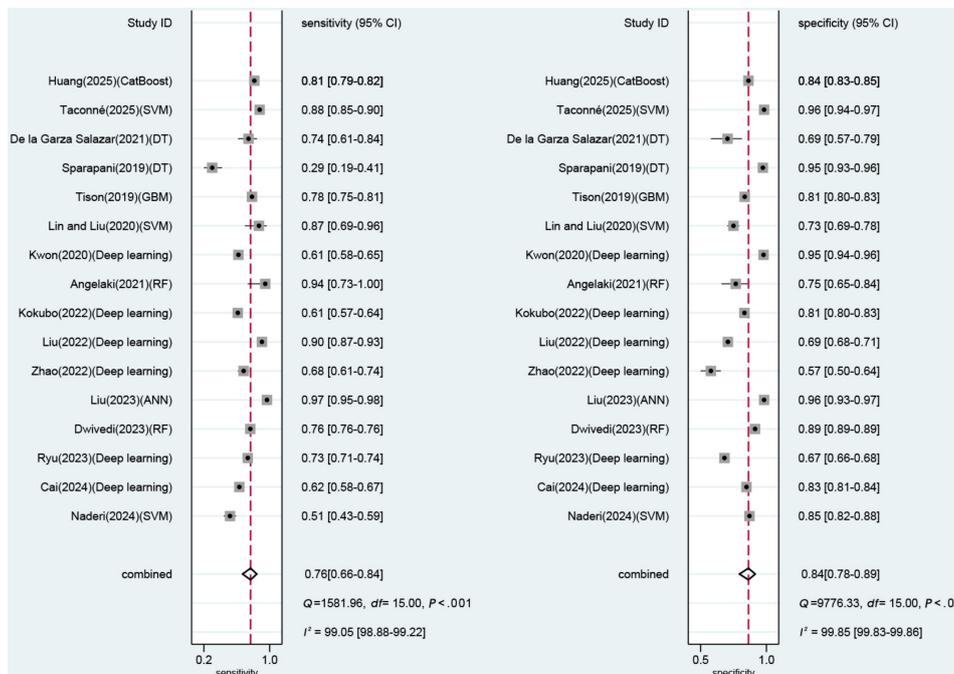
## Meta-Analysis

### ECG-Based ML Model for LVH Detection

In the validation set, the performance of ECG-based ML models was evaluated in 16 diagnostic 2x2 tables. The meta-analysis results demonstrated the following estimates: sensitivity 0.76 (95% CI 0.66-0.84), specificity 0.84 (95% CI 0.78-0.89), PLR 4.8 (95% CI 3.3-7.0), NLR 0.28 (95% CI 0.19-0.41), DOR 17 (95% CI 9-33), and SROC 0.88

(95% CI 0.75-0.95; Figure 3 and Figure S1 in Multimedia Appendix 2). The Deek funnel plot showed no significant publication bias among the included studies ( $P=.43$ ; Figure S2 in Multimedia Appendix 2). Given a pretest high-risk probability of 10%, if the model result indicated LVH, the posterior probability of LVH was 35%; if the model result indicated non-LVH, the posterior probability of LVH was 3% (Figure S3 in Multimedia Appendix 2).

**Figure 3.** Forest plot of sensitivity and specificity for electrocardiogram-based models [10,21,22,30-33,36,38,40-43,45-47]. ANN: artificial neural network; DT: decision tree; GBM: gradient boosting machine; RF: random forest; SVM: support vector machine.



Subgroup analyses were performed across various ML approaches. The deep learning (DL) model produced a sensitivity of 0.71 (95% CI 0.60-0.80) and a specificity of 0.79 (95% CI 0.65-0.88).

Subgroup analyses were performed across various validation approaches. The results showed that internal validation produced a sensitivity of 0.77 (95% CI 0.64-0.86) and a specificity of 0.83 (95% CI 0.75-0.89), while external

validation produced a sensitivity of 0.76 (95% CI 0.60-0.87) and a specificity of 0.87 (95% CI 0.74-0.94; [Table 2](#)).

**Table 2.** Subgroup analyses of electrocardiography (ECG)-based machine learning (ML) models across various validation approaches.

Methods	Machine learning approaches, n	Sen <sup>a</sup> (95% CI)	Spe <sup>b</sup> (95% CI)	PLR <sup>c</sup> (95% CI)	NLR <sup>d</sup> (95% CI)	DOR <sup>e</sup> (95% CI)	SROC <sup>f</sup> (95% CI)
IV <sup>g</sup>	11	0.77 (0.64-0.86)	0.83 (0.75-0.89)	4.5 (2.9-6.8)	0.28 (0.17-0.46)	16 (7-35)	0.87 (0.74-0.94)
EV <sup>h</sup>	5	0.76 (0.60-0.87)	0.87 (0.74-0.94)	5.8 (2.8-12.2)	0.28 (0.16-0.48)	21 (7-60)	0.89 (0.52-0.98)
Overall	16	0.76 (0.66-0.84)	0.84 (0.78-0.89)	4.8 (3.3-7.0)	0.28 (0.19-0.41)	17 (9-33)	0.88 (0.75-0.95)

<sup>a</sup>Sen: sensitivity.

<sup>b</sup>Spe: specificity.

<sup>c</sup>PLR: positive likelihood ratio.

<sup>d</sup>NLR: negative likelihood ratio.

<sup>e</sup>DOR: diagnostic odds ratio.

<sup>f</sup>SROC: summary receiver operating characteristic.

<sup>g</sup>IV: internal validation.

<sup>h</sup>EV: external validation.

As for internal validation, the meta-analysis results demonstrated an SROC of 0.87 (95% CI 0.74-0.94; [Figure S4 in Multimedia Appendix 2](#)). The Deek funnel plot showed no significant publication bias among the included studies ( $P=.30$ ; [Figure S5 in Multimedia Appendix 2](#)). Given a pretest high-risk probability of 10%, if the model result indicated LVH, the posterior probability of LVH was 33%; if the model result indicated non-LVH, the posterior probability of LVH was 3% ([Figure S6 in Multimedia Appendix 2](#)).

As for external validation, the meta-analysis results showed an SROC of 0.89 (95% CI 0.52-0.98; [Figure S7 in Multimedia Appendix 2](#)). The Deek funnel plot showed no significant publication bias across the included studies ( $P=.37$ ; [Figure S8 in Multimedia Appendix 2](#)). Given a pretest high-risk probability of 10%, if the model result indicated LVH, the posterior probability of LVH was 39%; if the model result indicated non-LVH, the posterior probability of LVH was 3% ([Figure S9 in Multimedia Appendix 2](#)).

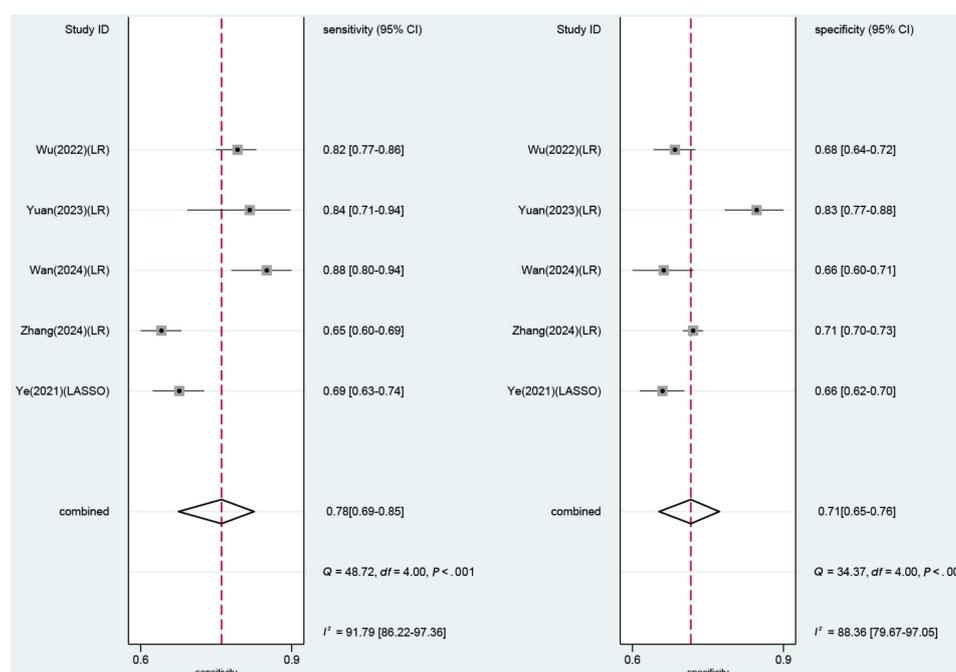
Dwivedi et al [30] used image-level splitting, which involved sharing patient images between training and validation. This method posed a high RoB due to data leakage. Accordingly, a sensitivity analysis was conducted on the ECG-based ML model group to exclude this study. The

meta-analysis results demonstrated the following estimates: sensitivity 0.76 (95% CI 0.66-0.85), specificity 0.84 (95% CI 0.77-0.89), PLR 4.7 (95% CI 3.2-7.0), NLR 0.28 (95% CI 0.19-0.42), DOR 17 (95% CI 8-33), and SROC 0.88 (95% CI 0.74-0.94).

### Clinical Feature-Based ML Model for LVH Detection

In the validation set, there were 5 diagnostic 2×2 tables for evaluating the performance of ML models based on clinical features. The meta-analysis results demonstrated the following estimates: sensitivity 0.78 (95% CI 0.69-0.85), specificity 0.71 (95% CI 0.65-0.76), PLR 2.7 (95% CI 2.1-3.4), NLR 0.31 (95% CI 0.21-0.46), DOR 9 (95% CI 5-16), and SROC 0.79 (95% CI 0.62-0.90; [Figure 4 and Figure S10 in Multimedia Appendix 2](#)). The Deek funnel plot showed no significant publication bias among the included studies ( $P=.09$ ; [Figure S11 in Multimedia Appendix 2](#)). Given a pretest high-risk probability of 10%, if the model result indicated LVH, the posterior probability of LVH was 23%; if the model result indicated non-LVH, the posterior probability of LVH was 3% ([Figure S12 in Multimedia Appendix 2](#)).

**Figure 4.** Forest plot of sensitivity and specificity for clinical feature-based models [1,27,29,34,37]. LASSO: least absolute shrinkage and selection operator; LR: logistic regression.



## ECHO-Based ML Model for LVH Detection

Farhad et al [28] used image-level splitting, which involved sharing patient images between training and validation. This method posed a high RoB due to data leakage. Accordingly, a sensitivity analysis was conducted in the ECHO-based ML model group to exclude this study. Ghorbani et al [39] developed a DL model based on ECHO. In their validation set of 373 cases (including 142 LVH cases), the model had a pooled sensitivity of 0.71 and a specificity of 0.67. Madani et al [44] also developed a DL model using echocardiographic data. Their validation set comprised 200 cases (100 LVH cases), with the model demonstrating a pooled sensitivity of 0.78 and a specificity of 0.96. Similarly, Yu et al [35] implemented a DL approach for ECHO. Their validation set included 150 cases (97 LVH cases), and the model yielded a pooled sensitivity of 0.94 and a specificity of 0.92.

## Discussion

### Principal Findings

This study examined the efficacy of 3 types of ML models for LVH detection. In the validation cohort, ECG-based models had a sensitivity of 0.76 (95% CI 0.66-0.84) and a specificity of 0.84 (95% CI 0.78-0.89). The clinical feature-based models had a sensitivity of 0.78 (95% CI 0.69-0.85) and a specificity of 0.71 (95% CI 0.65-0.76). The ECHO-based models had a sensitivity ranging from 0.71 to 0.94 and a specificity ranging from 0.67 to 0.96. These findings confirm the significant efficacy of ML in LVH detection.

### Comparison to Prior Work

ML is increasingly accurate in LVH detection, which enables the identification and treatment of earlier subclinical cardiovascular diseases [49]. In clinical practice, ECG is

now widely used as a screening method for LVH due to its affordability and high accessibility, and is recommended as part of standard clinical evaluation for patients with hypertension in the Chinese Hypertension Clinical Practice Guidelines [50]. ECG-based LVH detection has been a major research priority in hypertensive cardiac target organ damage studies [1]. The clinical diagnosis of LVH via 12-lead ECG relies principally on the quantitative assessment of QRS complex (Q-wave, R-wave, S-wave) voltage parameters [51,52]. Despite recent innovations in diagnostic criteria, the overall sensitivity remains suboptimal, particularly in Chinese and other East Asian populations [1,53,54]. The low sensitivity of ECG in LVH detection limits its clinical application [32]. Consistent with existing research, our results showed that ECG-based models produced a sensitivity of 0.76 (95% CI 0.66-0.84) and a specificity of 0.84 (95% CI 0.78-0.89). Although limited by its sensitivity, studies have confirmed that ECG-detected LVH is associated with adverse cardiovascular outcomes [32,55]. Given its widespread availability and low cost, ECG remains an ideal initial assessment tool for LVH screening in patients with hypertension [56]. Therefore, some researchers recommend using ECG primarily for the risk stratification and prognosis prediction of LVH [43,57].

It has also been noted that different ML methods show significant differences in their ability to identify positive or outcome events. With its powerful automatic feature extraction capabilities, DL models can directly identify complex and subtle patterns from imaging data, integrate multiple imaging variables, and improve diagnostic accuracy [58,59]. Its automated segmentation reduces the subjective bias of traditional manual segmentation in radiomics and can integrate segmentation with diagnosis, offering significant advantages [60,61]. Therefore, it is imperative to prioritize the development of automated methods based on DL. Decision tree models are highly interpretable models that

generate rules by recursively splitting variables in a dataset [62]. Decision tree models add a penalty term proportional to the absolute value of the feature coefficient for feature selection and model building [63]. Multiple studies have confirmed that RF models typically exhibit higher predictive performance than other ML methods because they can model nonlinear relationships and are insensitive to overfitting [64]. RF models demonstrate significant advantages in handling imbalanced datasets [65]. RF models excel at handling nonlinear and missing data, are conducive to high-dimensional data analysis, and demonstrate strong predictive performance [66]. Support vector machine (SVM) models have high predictive accuracy [67]. For structured and well-defined data, SVM models can skillfully capture the association between biomarkers and diseases through mathematical and statistical principles, thereby achieving high diagnostic accuracy [68,69]. SVM models are known for their ability to find the optimal decision boundary in high-dimensional feature spaces, resulting in effective classification based on extracted features [70,71].

ML models that incorporated clinical features demonstrated a sensitivity of 0.78 (95% CI 0.69-0.85) and a specificity of 0.71 (95% CI 0.65-0.76) in the validation set. Unlike traditional ECG criteria, the clinical features used to construct these models encompassed demographic characteristics (eg, sex, age) and clinical factors (eg, systolic blood pressure, duration of hypertension, and abdominal obesity). The possible reasons for using these factors are as follows.

Sex constitutes a significant determinant of LVH [1,37]. Substantial evidence has established the female sex as an independent predictor. In previous predictive models, women scored 70, while men scored 0, possibly due to the older study population, where most women had reached menopause [1]. Women exhibit higher salt sensitivity than men due to female hormonal influences, which amplifies adverse renin-angiotensin-aldosterone system effects [72,73]. Regardless of race or menopausal status, women with salt-sensitive hypertension face a higher LVH risk because this sex-specific mechanism promotes aldosterone-driven endothelial dysfunction [72,73]. This association is further exacerbated by diminished cardiovascular protection due to decreased estrogen levels in postmenopausal women [37].

Age plays a pivotal role in forecasting LVH in young patients with salt-sensitive hypertension. The evidence reveals a dual pattern: (1) patients 40 years or older demonstrate a significantly higher risk than younger populations, consistent with age-related left ventricular mass increase [1,37,74], while (2) patients with early-onset hypertension exhibit more severe cardiac damage and elevated cardiovascular mortality [74,75]. Such observations suggest that the interaction between age and the timing of hypertension requires special attention.

The development of LVH is closely associated with systolic blood pressure level and duration of hypertension. Prospective studies demonstrate that each 19 mm Hg increase in systolic blood pressure elevates the LVH risk by 49% [6], while effective systolic blood pressure control markedly

reduces the disease risk in old patients [76]. These findings provide crucial evidence for clinical treatment decisions. Hypertension duration, as a cumulative exposure metric, has been consistently validated in multiple studies to be significantly associated with LVH. Shortening the course of hypertension can effectively decrease the risk of myocardial remodeling, which shows the importance of early intervention in delaying LVH progression [77].

Obesity constitutes an independent risk factor for LVH [78,79]. Waist circumference, the principal diagnostic metric for abdominal obesity, is linked with both cardiovascular and metabolic disorders [80]. The pathophysiological linkage between abdominal obesity and LVH can be attributed to the robust correlation between waist circumference and excessive visceral adipose tissue accumulation [80,81]. As a metabolically active depot, visceral adipose tissue exhibits substantially higher energy demands compared to subcutaneous fat [81]. This augmented metabolic requirement exacerbates cardiac workload in patients with hypertension, ultimately resulting in LVH [78,81]. Studies have identified waist circumference as a superior predictor of cardiovascular risk and mortality compared to BMI [80,82]. Notably, elevated systolic blood pressure and prolonged hypertension duration may synergistically accelerate concentric cardiac remodeling, while abdominal obesity plus aging compounds metabolic risk.

The complex interactions among clinical features were quantified and integrated into ML models. While previous studies have confirmed the underlying mechanism of individual factors, models combining features from diverse dimensions provide more complete predictive insights. Specifically, sex-specific risk mechanisms dynamically interact with age-related cardiovascular changes, while blood pressure's temporal cumulative effects and the spatial distribution of metabolic factors form a 3D LVH prediction network. This finding provides a theoretical foundation for developing clinical feature-based ML tools for precise prediction.

ECHO serves as the primary imaging tool for LVH identification. However, its clinical application faces multiple technical challenges, including a relatively low signal-to-noise ratio, inconsistent image quality, and suboptimal reproducibility [28,35]. These limitations are particularly pronounced in patients with obesity due to enhanced chest wall and subcutaneous tissue thickness, which significantly impede image acquisition [28,35]. ML technology offers a novel approach to overcoming these technical barriers. ML models incorporating ECHO demonstrated a sensitivity range of 0.71 to 0.94 and a specificity range of 0.67 to 0.96.

Existing hospital ECHO systems are integrated with this technology to enable automated image analysis and real-time LVH diagnosis. Without replacing existing manual verification procedures, it markedly increases diagnostic efficiency and proves particularly beneficial as a clinical decision aid for less experienced physicians [28]. In contemporary practice, LVH detection chiefly relies on ECHO, where ML proves to be powerful in image processing stages [35,83].

Conventional approaches that depend on manual segmentation and feature extraction are prone to information loss, whereas DL models incorporate image processing, texture extraction, and feature selection throughout the modeling process, which better preserves original image information [35,84,85]. All ML models based on ECHO in this study were DL models. Beyond improved diagnostic efficiency, ML models can automatically identify high-risk cases and enable cardiologists to prioritize urgent interventions, thus shortening care pathways and improving outcomes [49]. Additionally, this technique can enhance diagnostic capabilities at primary care levels, optimize the allocation of cardiovascular imaging resources, reduce unnecessary tests, and be significantly cost-effective in population screening and subclinical disease management [28].

Despite the superior image-processing capabilities of DL models, their clinical integration with ultrasound systems must overcome real-world challenges, including real-time performance, device compatibility, and patient data privacy. Furthermore, the stability of ML models across demographics needs further validation in multicenter research. Future research should prioritize enlarging and diversifying ECHO image databases and validating the actual performance of ML in the detection of LVH in prospective trials.

Meta-analysis of ECG-based ML models for detecting LVH revealed substantial heterogeneity. This restricts the statistical reliability of pooled performance metrics, preventing their interpretation as definitive generalization conclusions. Potential sources of such heterogeneity are manifold. Methodologically, substantial variation in ECG signal preprocessing, feature extraction, and used algorithms introduces fundamental differences. Clinically, diverse underlying disease profiles and severity among studied populations affect model consistency. Regional demographic variations further increase result dispersion. Consequently, pooled estimates from our research require cautious interpretation. Their value lies more in demonstrating the broad performance range and limited confirmatory nature of current evidence than in providing a stable performance expectation. Future efforts should prioritize methodological standardization and prospective multicenter validation to enhance evidence quality in this domain.

When the PROBAST was used to assess model quality, we found that the results raised some concerns, particularly in terms of its strict evaluation of the study population. We believe that some of its criteria may be overly stringent. Although the PROBAST is suitable for evaluating multivariate models for diagnosis and prediction, it is primarily based on retrospective case-control studies, with few prospective studies included [86]. This design preference often leads to a high risk of bias in assessment results, which poses a significant challenge to diagnostic model research. In the studies we included, most involved diagnostic models, and the case selection process itself might introduce bias. The PROBAST uses the EPV principle for sample size estimation, but image-based ML models face challenges in applying EPV rules for bias risk assessment. In future studies, we may explore prior

sample size estimation during the research design phase, leveraging ML-based expected performance metrics, such as area under the curve and its 95% CI. Moreover, learning curve analysis serves as a powerful empirical tool. By plotting how a model's performance on both the training and validation sets changes as the sample size increases, we can intuitively determine whether the model's performance has reached a plateau, thereby establishing the optimal number of cases. In the field of DL, there exist several widely accepted rules. For instance, in complex image classification tasks, it is generally recommended that each category requires at least several thousand or even tens of thousands of training images to achieve good generalization performance. Although these guidelines are somewhat rough, they provide a practical benchmark for assessing sample sizes in large-scale imaging studies. In addition, the PROBAST requires the verification of the consistency between model statistical weights and actual reported results, which is particularly difficult for models with low interpretability (such as SVM), as they typically do not provide the weight information required in the original validation set [24]. These issues indicate that while the PROBAST is widely used for bias risk assessment in ML research, the observed limitations arise more fundamentally from the inherent biases of the retrospective case-control design—which is applied by most of the current ML diagnostic studies—than from an inherent excessive stringency in the PROBAST criteria. Advancing this field requires constructing and validating models through prospective, multicenter, double-blinded cohort studies, thereby improving evidence quality at its source. Specifically, future work should emphasize prospective sample size estimation and learning curve analysis during development, enhance model interpretability, conduct independent external validation, and ultimately assess the real-world impact of a model on clinical workflows and patient outcomes via pragmatic effectiveness studies. Additionally, we recommend using QUADAS-AI (Quality Assessment of Diagnostic Accuracy Studies Using AI) to assess the risk of bias in included original studies for diagnostic ML tasks [87]. QUADAS-AI focuses on data sources, sample size, eligibility criteria, the rationale for splitting training, validation, test sets, imaging protocols, and preprocessing methods—rather than simply labeling all case-control studies as “high risk.” In the studies we included, a substantial number of studies clearly described the source, size, and eligibility criteria, the principles of splitting into training, validation, and test sets, as well as the imaging protocols and preprocessing methods, thereby mitigating the high risk of bias associated with the study design.

This study has high statistical heterogeneity. Even when we conducted subgroup analyses using various ML approaches and validation methods to further elucidate the sources of heterogeneity, our explanatory power remains limited. This may be due to underlying heterogeneity present during model training. We consider that the following ML processes may introduce potential heterogeneity. First, regarding image preprocessing, the included studies exhibited certain variations in their approaches. These differences may introduce significant heterogeneity. For different images,

there is considerable variation in image parameters, yet there is a lack of effective discussion on how this impacts modeling. The images in the original studies exhibit significant variations across imaging protocols. Although we performed subgroup analyses based on different models, even within the same model, differences in model iterations and parameter adjustment rules persist, potentially contributing to heterogeneity.

### **Strengths and Limitations of the Study**

Our systematic review is the first to quantitatively examine ML performance in LVH detection. While informative, several limitations should be considered when interpreting these results. First, the systematic search, while rigorous, identified solely small case-control studies, thereby narrowing the spectrum of studies for subgroup comparisons. Second, the limited number of included studies precluded meaningful comparisons between different ML methodologies. Third, while most studies completed external validation, some relied solely on random sampling for internal validation. Owing to the limited number of publications, a systematic comparison between internal and external validation approaches is infeasible. Fourth, the RoB in the included studies is high, which is inevitable given the stringent criteria for evaluating ML-based research. This observation corresponds with methodological appraisals indicating that 87% of ML models for medical prediction carry a high RoB [88]. Future studies should optimize design, implementation, reporting, and validation approaches to facilitate clinical translation of ML-based prediction models. Fifth, some studies divide data according to images, and patients' images are shared between the training set and validation sets, which may result in data leakage. Sixth, there are different reference standards for different models. This discrepancy in reference standards indicates that any comparison of diagnostic accuracy between different model types is inherently confounded. The presented pooled estimates for each subgroup should be interpreted independently and not compared directly to each other. Seventh, the findings of this study may have been influenced to some extent by the fact that 3 papers used the 2x2 formula for estimation. The results should be interpreted with caution. Eighth, several included studies used "clinical consensus" as the reference standard for diagnosing LVH. While this consensus typically adhered to established guidelines (defining LVH as left ventricular mass index  $>115$  g/m<sup>2</sup> in men and  $>95$  g/m<sup>2</sup> in women [89]), original studies generally did not explicitly describe whether the consensus process was entirely independent of the outputs of evaluated ML models. Theoretically, "incorporation bias" could be introduced if the predictions of a model were known or incorporated during the adjudication of the reference standard, potentially leading to an overestimation of its diagnostic accuracy. This constitutes a key limitation for evidence interpretation in this systematic review and underscores the necessity for future original research to explicitly report reference standard independence in methodological descriptions.

### **Future Directions**

ML is widely used in LVH research, but its clinical application still faces multiple challenges. Many models are developed based on limited samples and lack standardized processes for key steps, such as image selection, segmentation, and feature extraction, resulting in high reliance of models on the experience of developers [62,90]. This raises concerns about the robustness of the models. Significant differences in image quality across different institutions further weaken the universality of the models. Existing models often focus on accuracy and neglect population risk stratification and regional differences in medical resources, limiting their clinical utility. Considering interpretability and a lack of trust among physicians, it remains difficult to integrate ML into clinical practice [91,92]. Although some models perform well in retrospective studies, their effectiveness and applicability in real-world clinical settings still need to be verified through rigorous prospective trials. Before clinical application, it is necessary to fully address the issues of patient data privacy protection, lack of trust among doctors due to insufficient model interpretability, and ethical considerations associated with new systems [93-95]. Additionally, we must be wary of overreliance on technological means at the expense of humanistic care. The need for systematic management and individualized treatment should not be overlooked. Technology should serve, rather than replace, patient-centered medical practices.

To promote the safe, effective, and responsible application of ML in LVH clinical practice, it is necessary to standardize the model development process, strengthen research on model generalizability and interpretability, conduct rigorous prospective validation, and focus on risk stratification and resource differences. Furthermore, it is necessary to establish a comprehensive ethical and regulatory framework covering data privacy, patient rights, algorithm transparency, and clinical workflow integration. Clear implementation guidelines are needed to ensure that technological development truly benefits patients and improves the quality of medical care.

### **Conclusions**

This study suggests that ML demonstrates diagnostic potential in detecting LVH. However, the overall certainty of the evidence is low, primarily constrained by the limited number of included studies and substantial heterogeneity. Consequently, current conclusions about model performance should be cautiously interpreted. Future investigations should adopt prospective designs and implement standardized data acquisition and model validation protocols to develop and evaluate ML models with enhanced robustness and clinical interpretability. This will advance this technology and make it a reliable diagnostic tool.

---

**Funding**

This research was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

---

**Data Availability**

All data generated or analyzed during this study are included in this published article and its multimedia appendices.

---

**Authors' Contributions**

Conceptualization: JW, KZ, YL

Formal analysis and investigation: JW, KZ, YL

Methodology: JW, KZ, YL

Supervision: JW, KZ, YL

Writing – original draft preparation: JW, KZ, YL

Writing – review and editing: JW, KZ, YL

All authors commented on previous versions of the manuscript, read and approved the final manuscript, and contributed to the study conception and design.

---

**Conflicts of Interest**

None declared.

---

**Multimedia Appendix 1**

Literature search strategy.

[\[DOCX File \(Microsoft Word File\), 20 KB-Multimedia Appendix 1\]](#)

---

**Multimedia Appendix 2**

Summary receiver operating characteristic curves, funnel plots, and nomograms for electrocardiogram (overall and validation analyses) and clinical features.

[\[DOCX File \(Microsoft Word File\), 1411 KB-Multimedia Appendix 2\]](#)

---

**Checklist 1**

PRISMA flowchart.

[\[PDF File \(Adobe File\), 100 KB-Checklist 1\]](#)

---

**References**

1. Zhang X, He C, Lu S, et al. Construction and validation of a nomogram to predict left ventricular hypertrophy in low-risk patients with hypertension. *J Clin Hypertens*. Mar 2024;26(3):274-285. [doi: [10.1111/jch.14780](https://doi.org/10.1111/jch.14780)]
2. Đorđević DB, Koračević GP, Đorđević AD, Lović DB. Hypertension and left ventricular hypertrophy. *J Hypertens*. Sep 1, 2024;42(9):1505-1515. [doi: [10.1097/HJH.0000000000003774](https://doi.org/10.1097/HJH.0000000000003774)] [Medline: [38747417](https://pubmed.ncbi.nlm.nih.gov/38747417/)]
3. Han Y, Li Y, Wu Z, et al. Progress in diagnosis and treatment of hypertension combined with left ventricular hypertrophy. *Ann Med*. Dec 2024;56(1):2405080. [doi: [10.1080/07853890.2024.2405080](https://doi.org/10.1080/07853890.2024.2405080)] [Medline: [39301864](https://pubmed.ncbi.nlm.nih.gov/39301864/)]
4. Miller RJH, Mikami Y, Heydari B, et al. Sex-specific relationships between patterns of ventricular remodelling and clinical outcomes. *Eur Heart J Cardiovasc Imaging*. Sep 1, 2020;21(9):983-990. [doi: [10.1093/ehjci/jeaa164](https://doi.org/10.1093/ehjci/jeaa164)] [Medline: [32594163](https://pubmed.ncbi.nlm.nih.gov/32594163/)]
5. Lewis AA, Ayers CR, Selvin E, et al. Racial differences in malignant left ventricular hypertrophy and incidence of heart failure: a multicohort study. *Circulation*. Mar 24, 2020;141(12):957-967. [doi: [10.1161/CIRCULATIONAHA.119.043628](https://doi.org/10.1161/CIRCULATIONAHA.119.043628)] [Medline: [31931608](https://pubmed.ncbi.nlm.nih.gov/31931608/)]
6. Cao X, Broughton ST, Waits GS, Nguyen T, Li Y, Soliman EZ. Interrelations between hypertension and electrocardiographic left ventricular hypertrophy and their associations with cardiovascular mortality. *Am J Cardiol*. Jan 15, 2019;123(2):274-283. [doi: [10.1016/j.amjcard.2018.10.006](https://doi.org/10.1016/j.amjcard.2018.10.006)] [Medline: [30390988](https://pubmed.ncbi.nlm.nih.gov/30390988/)]
7. Siranart N, Deepan N, Techasatian W, et al. Diagnostic accuracy of artificial intelligence in detecting left ventricular hypertrophy by electrocardiograph: a systematic review and meta-analysis. *Sci Rep*. Jul 10, 2024;14(1):15882. [doi: [10.1038/s41598-024-66247-y](https://doi.org/10.1038/s41598-024-66247-y)] [Medline: [38987652](https://pubmed.ncbi.nlm.nih.gov/38987652/)]
8. Pedersen LR, Kristensen AMD, Petersen SS, et al. Prognostic implications of left ventricular hypertrophy diagnosed on electrocardiogram vs echocardiography. *J Clin Hypertens (Greenwich)*. Sep 2020;22(9):1647-1658. [doi: [10.1111/jch.13991](https://doi.org/10.1111/jch.13991)] [Medline: [32813895](https://pubmed.ncbi.nlm.nih.gov/32813895/)]
9. Woythaler JN, Singer SL, Kwan OL, et al. Accuracy of echocardiography versus electrocardiography in detecting left ventricular hypertrophy: comparison with postmortem mass measurements. *J Am Coll Cardiol*. Aug 1983;2(2):305-311. [doi: [10.1016/s0735-1097\(83\)80167-3](https://doi.org/10.1016/s0735-1097(83)80167-3)] [Medline: [6223063](https://pubmed.ncbi.nlm.nih.gov/6223063/)]

10. Cai C, Imai T, Hasumi E, Fujii K. One-shot screening: utilization of a two-dimensional convolutional neural network for automatic detection of left ventricular hypertrophy using electrocardiograms. *Comput Methods Programs Biomed.* Apr 2024;247:108097. [doi: [10.1016/j.cmpb.2024.108097](https://doi.org/10.1016/j.cmpb.2024.108097)] [Medline: [38428250](https://pubmed.ncbi.nlm.nih.gov/38428250/)]
11. Spahillari A, McCormick I, Yang JX, Quinn GR, Manning WJ. On-call transthoracic echocardiographic interpretation by first year cardiology fellows: comparison with attending cardiologists. *BMC Med Educ.* Jun 14, 2019;19(1):213. [doi: [10.1186/s12909-019-1634-7](https://doi.org/10.1186/s12909-019-1634-7)] [Medline: [31200721](https://pubmed.ncbi.nlm.nih.gov/31200721/)]
12. Rao S, Ferris TG, Hidrue MK, et al. Physician burnout, engagement and career satisfaction in a large academic medical practice. *Clin Med Res.* Mar 2020;18(1):3-10. [doi: [10.3121/cmr.2019.1516](https://doi.org/10.3121/cmr.2019.1516)] [Medline: [31959669](https://pubmed.ncbi.nlm.nih.gov/31959669/)]
13. Quinn GR, Ranum D, Song E, et al. Missed diagnosis of cardiovascular disease in outpatient general medicine: insights from malpractice claims data. *Jt Comm J Qual Patient Saf.* Oct 2017;43(10):508-516. [doi: [10.1016/j.jcjq.2017.05.001](https://doi.org/10.1016/j.jcjq.2017.05.001)] [Medline: [28942775](https://pubmed.ncbi.nlm.nih.gov/28942775/)]
14. Wang Z, Gu Y, Huang L, et al. Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data. *Cardiovasc Diabetol.* Sep 28, 2024;23(1):351. [doi: [10.1186/s12933-024-02439-0](https://doi.org/10.1186/s12933-024-02439-0)] [Medline: [39342281](https://pubmed.ncbi.nlm.nih.gov/39342281/)]
15. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biol.* Feb 2025;79:103470. [doi: [10.1016/j.redox.2024.103470](https://doi.org/10.1016/j.redox.2024.103470)] [Medline: [39700695](https://pubmed.ncbi.nlm.nih.gov/39700695/)]
16. Layton AT. AI, machine learning, and ChatGPT in hypertension. *Hypertension.* Apr 2024;81(4):709-716. [doi: [10.1161/HYPERTENSIONAHA.124.19468](https://doi.org/10.1161/HYPERTENSIONAHA.124.19468)] [Medline: [38380541](https://pubmed.ncbi.nlm.nih.gov/38380541/)]
17. Ying Y, Ju R, Wang J, et al. Accuracy of machine learning in diagnosing microsatellite instability in gastric cancer: a systematic review and meta-analysis. *Int J Med Inform.* Jan 2025;193:105685. [doi: [10.1016/j.ijmedinf.2024.105685](https://doi.org/10.1016/j.ijmedinf.2024.105685)] [Medline: [39515046](https://pubmed.ncbi.nlm.nih.gov/39515046/)]
18. Beam AL, Kohane IS. Big data and machine learning in health care. *J Am Med Assoc.* Apr 3, 2018;319(13):1317-1318. [doi: [10.1001/jama.2017.18391](https://doi.org/10.1001/jama.2017.18391)] [Medline: [29532063](https://pubmed.ncbi.nlm.nih.gov/29532063/)]
19. Zhao K, Zhu Y, Chen X, et al. Machine learning in hypertrophic cardiomyopathy: nonlinear model from clinical and CMR features predicting cardiovascular events. *JACC Cardiovasc Imaging.* Aug 2024;17(8):880-893. [doi: [10.1016/j.jcmg.2024.04.013](https://doi.org/10.1016/j.jcmg.2024.04.013)] [Medline: [39001729](https://pubmed.ncbi.nlm.nih.gov/39001729/)]
20. You J, Guo Y, Kang JJ, et al. Development of machine learning-based models to predict 10-year risk of cardiovascular disease: a prospective cohort study. *Stroke Vasc Neurol.* Dec 29, 2023;8(6):475-485. [doi: [10.1136/svn-2023-002332](https://doi.org/10.1136/svn-2023-002332)] [Medline: [37105576](https://pubmed.ncbi.nlm.nih.gov/37105576/)]
21. Naderi H, Ramírez J, van Duijvenboden S, et al. Diagnostic and prognostic value of ECG-predicted hypertension-mediated left ventricular hypertrophy using machine learning. medRxiv. Preprint posted online on Apr 23, 2024. [doi: [10.1101/2024.04.22.24306204](https://doi.org/10.1101/2024.04.22.24306204)]
22. Ryu JS, Lee S, Chu Y, Ahn MS, Park YJ, Yang S. CoAt-mixer: self-attention deep learning framework for left ventricular hypertrophy using electrocardiography. *PLoS ONE.* 2023;18(6):e0286916. [doi: [10.1371/journal.pone.0286916](https://doi.org/10.1371/journal.pone.0286916)] [Medline: [37289800](https://pubmed.ncbi.nlm.nih.gov/37289800/)]
23. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
24. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med.* Jan 1, 2019;170(1):W1-W33. [doi: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377)] [Medline: [30596876](https://pubmed.ncbi.nlm.nih.gov/30596876/)]
25. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* Sep 2005;58(9):882-893. [doi: [10.1016/j.jclinepi.2005.01.016](https://doi.org/10.1016/j.jclinepi.2005.01.016)] [Medline: [16085191](https://pubmed.ncbi.nlm.nih.gov/16085191/)]
26. Fagan TJ. Letter: nomogram for Bayes's theorem. *N Engl J Med.* Jul 31, 1975;293(5):257. [doi: [10.1056/NEJM197507312930513](https://doi.org/10.1056/NEJM197507312930513)] [Medline: [1143310](https://pubmed.ncbi.nlm.nih.gov/1143310/)]
27. Wan J, Wang P, Liu S, Wang X, Zhou P, Yang J. Risk factors and a predictive model for left ventricular hypertrophy in young adults with salt-sensitive hypertension. *J Clin Hypertens (Greenwich).* Aug 2024;26(8):933-944. [doi: [10.1111/jch.14863](https://doi.org/10.1111/jch.14863)] [Medline: [38940286](https://pubmed.ncbi.nlm.nih.gov/38940286/)]
28. Farhad M, Masud MM, Beg A, Ahmad A, Ahmed LA, Memon S. A data-efficient zero-shot and few-shot Siamese approach for automated diagnosis of left ventricular hypertrophy. *Comput Biol Med.* Sep 2023;163:107129. [doi: [10.1016/j.combiomed.2023.107129](https://doi.org/10.1016/j.combiomed.2023.107129)] [Medline: [37343469](https://pubmed.ncbi.nlm.nih.gov/37343469/)]
29. Yuan R, Chen J, Zhang S, Zhang X, Yu J. Establishment of an individual-specific nomogram for predicting the risk of left ventricular hypertrophy in Chinese postmenopausal hypertensive women. *Medicina (Kaunas).* Mar 17, 2023;59(3):598. [doi: [10.3390/medicina59030598](https://doi.org/10.3390/medicina59030598)] [Medline: [36984599](https://pubmed.ncbi.nlm.nih.gov/36984599/)]

30. Dwivedi T, Xue J, Treiman D, Dubey A, Albert D. Machine learning models of 6-lead ECGs for the interpretation of left ventricular hypertrophy (LVH). *J Electrocardiol*. 2023;77:62-67. [doi: [10.1016/j.jelectrocard.2022.12.001](https://doi.org/10.1016/j.jelectrocard.2022.12.001)] [Medline: [36641988](https://pubmed.ncbi.nlm.nih.gov/36641988/)]
31. Liu CW, Wu FH, Hu YL, et al. Left ventricular hypertrophy detection using electrocardiographic signal. *Sci Rep*. Feb 13, 2023;13(1):2556. [doi: [10.1038/s41598-023-28325-5](https://doi.org/10.1038/s41598-023-28325-5)] [Medline: [36781924](https://pubmed.ncbi.nlm.nih.gov/36781924/)]
32. Zhao X, Huang G, Wu L, et al. Deep learning assessment of left ventricular hypertrophy based on electrocardiogram. *Front Cardiovasc Med*. 2022;9:952089. [doi: [10.3389/fcvm.2022.952089](https://doi.org/10.3389/fcvm.2022.952089)] [Medline: [36035939](https://pubmed.ncbi.nlm.nih.gov/36035939/)]
33. Liu CM, Hsieh ME, Hu YF, et al. Artificial intelligence-enabled model for early detection of left ventricular hypertrophy and mortality prediction in young to middle-aged adults. *Circ Cardiovasc Qual Outcomes*. Aug 2022;15(8):e008360. [doi: [10.1161/CIRCOUTCOMES.121.008360](https://doi.org/10.1161/CIRCOUTCOMES.121.008360)] [Medline: [35959675](https://pubmed.ncbi.nlm.nih.gov/35959675/)]
34. Wu Z, Shi M, Wang L, Yao Y. Identification of major risk factors and non-linear effects to the development of left ventricular hypertrophy in chronic kidney disease by constructing and validation of nomograms. *Front Med (Lausanne)*. 2022;9:914800. [doi: [10.3389/fmed.2022.914800](https://doi.org/10.3389/fmed.2022.914800)] [Medline: [35911424](https://pubmed.ncbi.nlm.nih.gov/35911424/)]
35. Yu X, Yao X, Wu B, et al. Using deep learning method to identify left ventricular hypertrophy on echocardiography. *Int J Cardiovasc Imaging*. Apr 2022;38(4):759-769. [doi: [10.1007/s10554-021-02461-3](https://doi.org/10.1007/s10554-021-02461-3)] [Medline: [34757566](https://pubmed.ncbi.nlm.nih.gov/34757566/)]
36. Kokubo T, Kodera S, Sawano S, et al. Automatic detection of left ventricular dilatation and hypertrophy from electrocardiograms using deep learning. *Int Heart J*. Sep 30, 2022;63(5):939-947. [doi: [10.1536/ihj.22-132](https://doi.org/10.1536/ihj.22-132)] [Medline: [36104234](https://pubmed.ncbi.nlm.nih.gov/36104234/)]
37. Ye C, Wang T, Gong J, et al. Development of a nomogram for screening the risk of left ventricular hypertrophy in Chinese hypertensive patients. *J Clin Hyperten*. Jun 2021;23(6):1176-1185. [doi: [10.1111/jch.14240](https://doi.org/10.1111/jch.14240)]
38. Angelaki E, Marketou ME, Barmparis GD, et al. Detection of abnormal left ventricular geometry in patients without cardiovascular disease through machine learning: an ECG-based approach. *J Clin Hypertens (Greenwich)*. May 2021;23(5):935-945. [doi: [10.1111/jch.14200](https://doi.org/10.1111/jch.14200)] [Medline: [33507615](https://pubmed.ncbi.nlm.nih.gov/33507615/)]
39. Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med*. 2020;3(10):10. [doi: [10.1038/s41746-019-0216-8](https://doi.org/10.1038/s41746-019-0216-8)] [Medline: [31993508](https://pubmed.ncbi.nlm.nih.gov/31993508/)]
40. Kwon JM, Jeon KH, Kim HM, et al. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace*. Mar 1, 2020;22(3):412-419. [doi: [10.1093/europace/euz324](https://doi.org/10.1093/europace/euz324)] [Medline: [31800031](https://pubmed.ncbi.nlm.nih.gov/31800031/)]
41. Lin GM, Liu K. An electrocardiographic system with anthropometrics via machine learning to screen left ventricular hypertrophy among young adults. *IEEE J Transl Eng Health Med*. 2020;8:1800111. [doi: [10.1109/JTEHM.2020.2990073](https://doi.org/10.1109/JTEHM.2020.2990073)] [Medline: [32419990](https://pubmed.ncbi.nlm.nih.gov/32419990/)]
42. Tison GH, Zhang J, Delling FN, Deo RC. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ Cardiovasc Qual Outcomes*. Sep 2019;12(9):e005289. [doi: [10.1161/CIRCOUTCOMES.118.005289](https://doi.org/10.1161/CIRCOUTCOMES.118.005289)] [Medline: [31525078](https://pubmed.ncbi.nlm.nih.gov/31525078/)]
43. Sparapani R, Dabbouseh NM, Gutterman D, et al. Detection of left ventricular hypertrophy using Bayesian additive regression trees: the MESA. *J Am Heart Assoc*. Mar 5, 2019;8(5):e009959. [doi: [10.1161/JAHA.118.009959](https://doi.org/10.1161/JAHA.118.009959)] [Medline: [30827132](https://pubmed.ncbi.nlm.nih.gov/30827132/)]
44. Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit Med*. 2018;1(1):59. [doi: [10.1038/s41746-018-0065-x](https://doi.org/10.1038/s41746-018-0065-x)] [Medline: [31304338](https://pubmed.ncbi.nlm.nih.gov/31304338/)]
45. De la Garza Salazar F, Romero Ibarguengoitia ME, Azpiri López JR, González Cantú A. Optimizing ECG to detect echocardiographic left ventricular hypertrophy with computer-based ECG data and machine learning. *PLoS ONE*. 2021;16(11):e0260661. [doi: [10.1371/journal.pone.0260661](https://doi.org/10.1371/journal.pone.0260661)] [Medline: [34847202](https://pubmed.ncbi.nlm.nih.gov/34847202/)]
46. Huang JT, Tseng CH, Huang WM, et al. Comparison of machine learning and conventional criteria in detecting left ventricular hypertrophy and prognosis with electrocardiography. *Eur Heart J Digit Health*. Mar 2025;6(2):252-260. [doi: [10.1093/ehjdh/ztaf003](https://doi.org/10.1093/ehjdh/ztaf003)] [Medline: [40110220](https://pubmed.ncbi.nlm.nih.gov/40110220/)]
47. Taconne M, Corino VDA, Mainardi L. An ECG-based model for left ventricular hypertrophy detection: a machine learning approach. *IEEE Open J Eng Med Biol*. 2025;6:219-226. [doi: [10.1109/OJEMB.2024.3509379](https://doi.org/10.1109/OJEMB.2024.3509379)] [Medline: [39698126](https://pubmed.ncbi.nlm.nih.gov/39698126/)]
48. van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res*. Aug 2019;28(8):2455-2474. [doi: [10.1177/0962280218784726](https://doi.org/10.1177/0962280218784726)] [Medline: [29966490](https://pubmed.ncbi.nlm.nih.gov/29966490/)]
49. Duffy G, Cheng PP, Yuan N, et al. High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. *JAMA Cardiol*. Apr 1, 2022;7(4):386-395. [doi: [10.1001/jamacardio.2021.6059](https://doi.org/10.1001/jamacardio.2021.6059)] [Medline: [35195663](https://pubmed.ncbi.nlm.nih.gov/35195663/)]

50. Liu LS, Writing Group of 2010 Chinese Guidelines for the Management of Hypertension. 2010 Chinese guidelines for the management of hypertension. *Zhonghua Xin Xue Guan Bing Za Zhi*. Jul 2011;39(7):579-615. [Medline: [22088239](#)]
51. Casale PN, Devereux RB, Kligfield P, et al. Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria. *J Am Coll Cardiol*. Sep 1985;6(3):572-580. [doi: [10.1016/s0735-1097\(85\)80115-7](#)] [Medline: [3161926](#)]
52. Sokolow M, Lyon TP. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *Am Heart J*. Feb 1949;37(2):161-186. [doi: [10.1016/0002-8703\(49\)90562-1](#)] [Medline: [18107386](#)]
53. Xia Y, Li X, Zhang H, et al. Diagnostic capability and influence factors for a new electrocardiogram criterion in the diagnosis of left ventricular hypertrophy in a Chinese population. *Cardiology*. 2020;145(5):294-302. [doi: [10.1159/000505421](#)] [Medline: [32289773](#)]
54. Wang D, Xu JZ, Zhang W, et al. Performance of electrocardiographic criteria for echocardiographically diagnosed left ventricular hypertrophy in Chinese hypertensive patients. *Am J Hypertens*. Sep 10, 2020;33(9):831-836. [doi: [10.1093/ajh/hpaa083](#)] [Medline: [32484222](#)]
55. Bang CN, Devereux RB, Okin PM. Regression of electrocardiographic left ventricular hypertrophy or strain is associated with lower incidence of cardiovascular morbidity and mortality in hypertensive patients independent of blood pressure reduction—a LIFE review. *J Electrocardiol*. 2014;47(5):630-635. [doi: [10.1016/j.jelectrocard.2014.07.003](#)] [Medline: [25052475](#)]
56. Hancock EW, Deal BJ, Mirvis DM, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part V: electrocardiogram changes associated with cardiac chamber hypertrophy: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. *Circulation*. Mar 17, 2009;119(10):e251-261. [doi: [10.1161/CIRCULATIONAHA.108.191097](#)] [Medline: [19228820](#)]
57. Rautaharju PM, Soliman EZ. Electrocardiographic left ventricular hypertrophy and the risk of adverse cardiovascular events: a critical appraisal. *J Electrocardiol*. 2014;47(5):649-654. [doi: [10.1016/j.jelectrocard.2014.06.002](#)] [Medline: [25012077](#)]
58. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun (Lond)*. Apr 2020;40(4):154-166. [doi: [10.1002/cac2.12012](#)] [Medline: [32277744](#)]
59. Jiang X, Li J, Kan Y, et al. MRI based radiomics approach with deep learning for prediction of vessel invasion in early-stage cervical cancer. *IEEE/ACM Trans Comput Biol Bioinform*. 2021;18(3):995-1002. [doi: [10.1109/TCBB.2019.2963867](#)] [Medline: [31905143](#)]
60. She L, Li Y, Wang H, et al. Imaging-based AI for predicting lymphovascular space invasion in cervical cancer: systematic review and meta-analysis. *J Med Internet Res*. Jun 16, 2025;27:e71091. [doi: [10.2196/71091](#)] [Medline: [40523236](#)]
61. Zuo H, Huang B, He J, Fang L, Huang M. Machine learning approaches in high myopia: systematic review and meta-analysis. *J Med Internet Res*. Jan 3, 2025;27:e57644. [doi: [10.2196/57644](#)] [Medline: [39753217](#)]
62. Zhang H, Zou P, Luo P, Jiang X. Machine learning for the early prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage: systematic review and meta-analysis. *J Med Internet Res*. Jan 20, 2025;27:e54121. [doi: [10.2196/54121](#)]
63. Zhu J, Yang F, Wang Y, et al. Accuracy of machine learning in discriminating Kawasaki disease and other febrile illnesses: systematic review and meta-analysis. *J Med Internet Res*. Nov 18, 2024;26:e57641. [doi: [10.2196/57641](#)] [Medline: [39556821](#)]
64. Becker T, Rousseau AJ, Geubbelmans M, Burzykowski T, Valkenburg D. Decision trees and random forests. *Am J Orthod Dentofacial Orthop*. Dec 2023;164(6):894-897. [doi: [10.1016/j.ajodo.2023.09.011](#)] [Medline: [38008491](#)]
65. Liu L, Li Z, Hu Y, et al. Predictive performance of machine learning for suicide in adolescents: systematic review and meta-analysis. *J Med Internet Res*. Jun 16, 2025;27:e73052. [doi: [10.2196/73052](#)] [Medline: [40522723](#)]
66. Yang J, Zeng S, Cui S, Zheng J, Wang H. Predictive modeling of acute respiratory distress syndrome using machine learning: systematic review and meta-analysis. *J Med Internet Res*. May 13, 2025;27:e66615. [doi: [10.2196/66615](#)] [Medline: [40359510](#)]
67. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2019;9(4):e1312. [doi: [10.1002/widm.1312](#)] [Medline: [32089788](#)]
68. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. May 2019;20(5):e262-e273. [doi: [10.1016/S1470-2045\(19\)30149-4](#)] [Medline: [31044724](#)]

69. Parvatikar PP, Patil S, Khaparkhantkar K, et al. Artificial intelligence: machine learning approach for screening large database and drug discovery. *Antiviral Res.* Dec 2023;220:105740. [doi: [10.1016/j.antiviral.2023.105740](https://doi.org/10.1016/j.antiviral.2023.105740)] [Medline: [37935248](https://pubmed.ncbi.nlm.nih.gov/37935248/)]
70. Tsai CA, Chang YJ. Efficient selection of Gaussian kernel SVM parameters for imbalanced data. *Genes (Basel).* Feb 25, 2023;14(3):583. [doi: [10.3390/genes14030583](https://doi.org/10.3390/genes14030583)] [Medline: [36980852](https://pubmed.ncbi.nlm.nih.gov/36980852/)]
71. Park JS, Park SY, Moon JW, Kim K, Suh DI. Artificial intelligence models for pediatric lung sound analysis: systematic review and meta-analysis. *J Med Internet Res.* Apr 18, 2025;27:e66491. [doi: [10.2196/66491](https://doi.org/10.2196/66491)] [Medline: [40249944](https://pubmed.ncbi.nlm.nih.gov/40249944/)]
72. Liao YY, Gao K, Fu BW, et al. Risk factors for electrocardiographic left ventricular hypertrophy in a young Chinese general population: the Hanzhong adolescent cohort study. *BMC Cardiovasc Disord.* Mar 31, 2021;21(1):159. [doi: [10.1186/s12872-021-01966-y](https://doi.org/10.1186/s12872-021-01966-y)] [Medline: [33789587](https://pubmed.ncbi.nlm.nih.gov/33789587/)]
73. Tiyerili V, Müller CFH, Fung S, Panek D, Nickenig G, Becher UM. Estrogen improves vascular function via peroxisome-proliferator-activated-receptor- $\gamma$ . *J Mol Cell Cardiol.* Aug 2012;53(2):268-276. [doi: [10.1016/j.yjmcc.2012.05.008](https://doi.org/10.1016/j.yjmcc.2012.05.008)] [Medline: [22634137](https://pubmed.ncbi.nlm.nih.gov/22634137/)]
74. Vasani RS, Song RJ, Xanthakis V, et al. Hypertension-mediated organ damage: prevalence, correlates, and prognosis in the community. *Hypertension.* Mar 2022;79(3):505-515. [doi: [10.1161/HYPERTENSIONAHA.121.18502](https://doi.org/10.1161/HYPERTENSIONAHA.121.18502)] [Medline: [35138872](https://pubmed.ncbi.nlm.nih.gov/35138872/)]
75. Wang C, Yuan Y, Zheng M, et al. Association of age of onset of hypertension with cardiovascular diseases and mortality. *J Am Coll Cardiol.* Jun 16, 2020;75(23):2921-2930. [doi: [10.1016/j.jacc.2020.04.038](https://doi.org/10.1016/j.jacc.2020.04.038)] [Medline: [32527401](https://pubmed.ncbi.nlm.nih.gov/32527401/)]
76. Zhou B, Li C, Shou J, Zhang Y, Wen C, Zeng C. The cumulative blood pressure load and target organ damage in patients with essential hypertension. *J Clin Hypertens (Greenwich).* Jun 2020;22(6):981-990. [doi: [10.1111/jch.13875](https://doi.org/10.1111/jch.13875)] [Medline: [32427414](https://pubmed.ncbi.nlm.nih.gov/32427414/)]
77. Carlsson AC, Ruge T, Sundström J, et al. Association between circulating endostatin, hypertension duration, and hypertensive target-organ damage. *Hypertension.* Dec 2013;62(6):1146-1151. [doi: [10.1161/HYPERTENSIONAHA.113.02250](https://doi.org/10.1161/HYPERTENSIONAHA.113.02250)] [Medline: [24082055](https://pubmed.ncbi.nlm.nih.gov/24082055/)]
78. Zhang X, Li G, Zhang D, Sun Y. Influence of hypertension and global or abdominal obesity on left ventricular hypertrophy: a cross-sectional study. *J Clin Hypertens (Greenwich).* Dec 2024;26(12):1342-1350. [doi: [10.1111/jch.14716](https://doi.org/10.1111/jch.14716)] [Medline: [37794691](https://pubmed.ncbi.nlm.nih.gov/37794691/)]
79. Tsujimoto T, Kajio H. Abdominal obesity is associated with an increased risk of all-cause mortality in patients with HFpEF. *J Am Coll Cardiol.* Dec 5, 2017;70(22):2739-2749. [doi: [10.1016/j.jacc.2017.09.1111](https://doi.org/10.1016/j.jacc.2017.09.1111)] [Medline: [29191321](https://pubmed.ncbi.nlm.nih.gov/29191321/)]
80. Choi D, Choi S, Son JS, Oh SW, Park SM. Impact of discrepancies in general and abdominal obesity on major adverse cardiac events. *J Am Heart Assoc.* Sep 17, 2019;8(18):e013471. [doi: [10.1161/JAHA.119.013471](https://doi.org/10.1161/JAHA.119.013471)] [Medline: [31480883](https://pubmed.ncbi.nlm.nih.gov/31480883/)]
81. Ross R, Neeland IJ, Yamashita S, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nat Rev Endocrinol.* Mar 2020;16(3):177-189. [doi: [10.1038/s41574-019-0310-7](https://doi.org/10.1038/s41574-019-0310-7)] [Medline: [32020062](https://pubmed.ncbi.nlm.nih.gov/32020062/)]
82. Wang Y, Howard AG, Adair LS, Wang H, Avery CL, Gordon-Larsen P. Waist circumference change is associated with blood pressure change independent of BMI change. *Obesity (Silver Spring).* Jan 2020;28(1):146-153. [doi: [10.1002/oby.22638](https://doi.org/10.1002/oby.22638)] [Medline: [31755247](https://pubmed.ncbi.nlm.nih.gov/31755247/)]
83. Wu D, Ono R, Wang S, Kobayashi Y, Sugimoto K, Liu H. Pulse wave signal-driven machine learning for identifying left ventricular enlargement in heart failure patients. *Biomed Eng Online.* Jun 22, 2024;23(1):60. [doi: [10.1186/s12938-024-01257-5](https://doi.org/10.1186/s12938-024-01257-5)] [Medline: [38909231](https://pubmed.ncbi.nlm.nih.gov/38909231/)]
84. Lau ES, Di Achille P, Koppurapu K, et al. Deep learning-enabled assessment of left heart structure and function predicts cardiovascular outcomes. *J Am Coll Cardiol.* Nov 14, 2023;82(20):1936-1948. [doi: [10.1016/j.jacc.2023.09.800](https://doi.org/10.1016/j.jacc.2023.09.800)] [Medline: [37940231](https://pubmed.ncbi.nlm.nih.gov/37940231/)]
85. Wehbe RM, Katsaggelos AK, Hammond KJ, et al. Deep learning for cardiovascular imaging: a review. *JAMA Cardiol.* Nov 1, 2023;8(11):1089-1098. [doi: [10.1001/jamacardio.2023.3142](https://doi.org/10.1001/jamacardio.2023.3142)] [Medline: [37728933](https://pubmed.ncbi.nlm.nih.gov/37728933/)]
86. Maleki F, Ovens K, Gupta R, Reinhold C, Spatz A, Forghani R. Generalizability of machine learning models: quantitative evaluation of three methodological pitfalls. *Radiol Artif Intell.* Jan 2023;5(1):e220028. [doi: [10.1148/ryai.220028](https://doi.org/10.1148/ryai.220028)] [Medline: [36721408](https://pubmed.ncbi.nlm.nih.gov/36721408/)]
87. Sounderajah V, Ashrafian H, Rose S, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat Med.* Oct 2021;27(10):1663-1665. [doi: [10.1038/s41591-021-01517-0](https://doi.org/10.1038/s41591-021-01517-0)] [Medline: [34635854](https://pubmed.ncbi.nlm.nih.gov/34635854/)]
88. Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ.* Oct 20, 2021;375:n2281. [doi: [10.1136/bmj.n2281](https://doi.org/10.1136/bmj.n2281)] [Medline: [34670780](https://pubmed.ncbi.nlm.nih.gov/34670780/)]

89. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. Jan 2015;28(1):1-39. [doi: [10.1016/j.echo.2014.10.003](https://doi.org/10.1016/j.echo.2014.10.003)] [Medline: [25559473](https://pubmed.ncbi.nlm.nih.gov/25559473/)]
90. Wang B, Jiang B, Liu D, Zhu R. Early predictive accuracy of machine learning for hemorrhagic transformation in acute ischemic stroke: systematic review and meta-analysis. *J Med Internet Res*. May 23, 2025;27:e71654. [doi: [10.2196/71654](https://doi.org/10.2196/71654)] [Medline: [40408765](https://pubmed.ncbi.nlm.nih.gov/40408765/)]
91. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. Oct 29, 2019;17(1):195. [doi: [10.1186/s12916-019-1426-2](https://doi.org/10.1186/s12916-019-1426-2)] [Medline: [31665002](https://pubmed.ncbi.nlm.nih.gov/31665002/)]
92. Alkhanbouli R, Matar Abdulla Almadhaani H, Alhosani F, Simsekler MCE. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. *BMC Med Inform Decis Mak*. Mar 4, 2025;25(1):110. [doi: [10.1186/s12911-025-02944-6](https://doi.org/10.1186/s12911-025-02944-6)] [Medline: [40038704](https://pubmed.ncbi.nlm.nih.gov/40038704/)]
93. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey. *J Med Internet Res*. Aug 26, 2021;23(8):e26162. [doi: [10.2196/26162](https://doi.org/10.2196/26162)] [Medline: [34236994](https://pubmed.ncbi.nlm.nih.gov/34236994/)]
94. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics*. Sep 15, 2021;22(1):122. [doi: [10.1186/s12910-021-00687-3](https://doi.org/10.1186/s12910-021-00687-3)] [Medline: [34525993](https://pubmed.ncbi.nlm.nih.gov/34525993/)]
95. Larson DB, Magnus DC, Lungren MP, Shah NH, Langlotz CP. Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology*. Jun 2020;295(3):675-682. [doi: [10.1148/radiol.2020192536](https://doi.org/10.1148/radiol.2020192536)] [Medline: [32208097](https://pubmed.ncbi.nlm.nih.gov/32208097/)]

## Abbreviations

**AI:** artificial intelligence  
**DL:** deep learning  
**DOR:** diagnostic odds ratio  
**ECG:** electrocardiogram  
**ECHO:** echocardiography  
**EPV:** events per variable  
**LVH:** left ventricular hypertrophy  
**MeSH:** Medical Subject Headings  
**ML:** machine learning  
**NLR:** negative likelihood ratio  
**PLR:** positive likelihood ratio  
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
**PROBAST:** Prediction Model Risk of Bias Assessment Tool  
**QUADAS-AI:** Quality Assessment of Diagnostic Accuracy Studies Using AI  
**RoB:** risk of bias  
**SROC:** summary receiver operating characteristic  
**SVM:** support vector machine

*Edited by Andrew Coristine; peer-reviewed by Nirajan Acharya, Nupur Shah; submitted 28.Apr.2025; final revised version received 28.Jan.2026; accepted 30.Jan.2026; published 27.Feb.2026*

### *Please cite as:*

Li Y, Zhao K, Wu J  
*Machine Learning in Left Ventricular Hypertrophy Detection: Systematic Review and Meta-Analysis*  
*J Med Internet Res* 2026;28:e76637  
URL: <https://www.jmir.org/2026/1/e76637>  
doi: [10.2196/76637](https://doi.org/10.2196/76637)

© Yilin Li, Ke Zhao, Jing Wu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 27.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org>, as well as this copyright and license information must be included.