Machine learning to classify left ventricular hypertrophy using ECG feature extraction by variational autoencoder

Short title: ML models for LVH

Amulya Gupta, MBBS,[1]* Christopher J. Harvey,[1]* Ashley DeBauge, MD,[2] Sumaiya Shomaji,[3] PhD, Zijun Yao, PhD,[3] Amit Noheria, MBBS, SM[1]

[1]Department of Cardiovascular Medicine, The University of Kansas Medical Center, Kansas City, Kansas

[2]Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri, USA

Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, Kansas

*Contributed equally to the manuscript

Word count: Abstract- 250 words, Main text- 3031 words; Tables: 3; Figures: 5; References: 35

Corresponding Author: Amit Noheria, MBBS, SM
3901 Rainbow Blvd.
Kansas City, KS 66160
Phone: (913)588-9600
Email: noheriaa@gmail.com

Disclosures: None

ABSTRACT

Background: Traditional ECG criteria for left ventricular hypertrophy (LVH) have low diagnostic yield. Machine learning (ML) can improve ECG classification.

Methods: ECG summary features (rate, intervals, axis), R-wave, S-wave and overall-QRS amplitudes, and QRS/QRST voltage-time integrals (VTIs) were extracted from 12-lead, vectorcardiographic X-Y-Z-lead, and root-mean-square (3D) representative-beat ECGs. Latent features were extracted by variational autoencoder from X-Y-Z and 3D representative-beat ECGs. Logistic regression, random forest, light gradient boosted machine (LGBM), residual network (ResNet) and multilayer perceptron network (MLP) models using ECG features and sex, and a convolutional neural network (CNN) using ECG signals, were trained to predict LVH (left ventricular mass indexed in women >95 g/m², men >115 g/m²) on 225,333 adult ECG-echocardiogram (within 45 days) pairs. AUROCs for LVH classification were obtained in a separate test set for individual ECG variables, traditional criteria and ML models.

Results: In the test set (n=25,263), AUROC for LVH classification was higher for ML models using ECG features (LGBM 0.790, MLP 0.789, ResNet 0.788) as compared to the best individual variable ($VTI_{QRS\text{-}3D}$ 0.677), the best traditional criterion (Cornell voltage-duration product 0.647) and CNN using ECG signal (0.767). Among patients without LVH who had a follow-up echocardiogram >1 (closest to 5) years later, LGBM false positives, compared to true negatives, had a 2.63 (95% CI 2.01, 3.45)-fold higher risk for developing LVH ($p<0.0001$).

Conclusions: ML models are superior to traditional ECG criteria to classify—and predict future—LVH. Models trained on extracted ECG features, including variational autoencoder latent variables, outperformed CNN directly trained on ECG signal.

Keywords: Left ventricular hypertrophy, LVH, machine learning, deep learning, artificial intelligence, electrocardiogram, ECG, variational autoencoder.

Abbreviations:

ECG, electrocardiogram

LVH, left ventricular hypertrophy

ML, machine learning

AI, artificial intelligence

MLP, multilayered perceptron

LGBM, light gradient-boosting machine

AUROC, area under the receiver operator characteristic curve

VAE, variational Autoencoder

LVMi, left ventricular mass indexed

INTRODUCTION

Left ventricular hypertrophy (LVH) refers to increased left ventricular mass, characterized by an increase in left ventricular wall thickness and/or enlargement of the left ventricular cavity. This is often secondary to pathological or physiological stressors such as chronic hypertension, valvular heart disease, athletic training, or genetic conditions. LVH is associated with over a two-fold increase in cardiovascular morbidity and all-cause mortality (1). Early detection and initiation of pharmacological treatment, along with lifestyle modifications, have been associated with improved outcomes (2).

Transthoracic echocardiography is the standard-of-care for the diagnosis of LVH. However, despite its non-invasive nature and widespread utilization, universal screening for LVH using echocardiography even in high-risk groups, such as those with hypertension, is not cost-effective (3,4).

Electrocardiography (ECG) is an affordable, widely accessible, and frequently used diagnostic tool for cardiovascular screening. Often considered an extension of the cardiovascular physical examination, it is estimated that over 100-300 million ECGs are performed annually in the United States (5). Several criteria for 12-lead ECG diagnosis of LVH have been published over many decades, mainly based on the magnitude of QRS voltages in various—especially precordial—leads. However, these criteria have poor sensitivities in detecting LVH, making them unsuitable for standalone ECG screening (6-8). In a 2023 consensus statement, the International Society of Electrocardiology and the International Society for Holter Monitoring and Noninvasive Electrocardiology highlighted the need for a paradigm shift in ECG-based LVH diagnosis (9). The statement emphasized the limitations of traditional ECG criteria and discussed the potential of artificial intelligence (AI)-driven approaches for LVH detection.

Machine learning (ML) can reduce reliance on human interpretation and yet increase the diagnostic accuracy of ECG (10,11). Several ECG-based ML models have been developed for detecting LVH, with varying sensitivities and specificities (12). Many of these studies use convolutional neural network (CNN) deep learning architecture to train models using ECG signals often with fewer than 10,000 training ECGs. Given that each 12-lead 10-second ECG signal at 500 Hz consists of 60,000 data points, using such a high-dimensionality input for ML training with a limited number of samples can result in overfitting and reduced generalizability (13-15). On the other hand, non-neural network ML architectures—such as logistic regression, random forest, gradient boosted machine—are not suited to use high-dimensional ECG signal data as input and are usually limited to using extracted ECG features with potential loss of diagnostic information (15).

To mitigate these limitations—while preserving the advantages of deep learning—we developed a variational autoencoder (VAE) that can encode 0.75-sec-representative-beat from either X-Y-Z-lead or root-mean-squared ECG into 30 variables (15-17). These VAE latent encodings retain the ECG morphological information and can reconstruct back the ECG signal with high fidelity. In this study, we aimed to train and test different ML models using extracted ECG features including the latent encodings or the ECG signal to classify LVH from the representative-beat ECG.

METHODS

Patient selection and data retrieval: An automated retrospective retrieval of records was performed from our clinical database at the University of Kansas Medical Center between May 2010 and Jan 2022 to search for ECG and echocardiogram performed on the same patient within

45 days of each other. Echocardiograms-ECG pairs with echocardiographic left ventricular mass index (LVMi) >95 g/m^2 for females and >115 g/m^2 for males were labelled as 'LVH' while rest of the pairs were assigned to the 'no LVH' group (15). The study was conducted under an approval from the Institutional Review Board.

Data extraction: ECGs were acquired with Philips 12-lead ECG machines. The 12-lead ECG 10-second and 1200-ms-representative-beat signals along with standard features like heart rate, PR interval, etc. were exported to a research SQL data server. Echocardiograms were standard clinical studies performed for clinical indications both as outpatient and inpatient evaluations. Individual echocardiogram numeric variables including diastolic measurements of left ventricular internal diameter (LVIDd), interventricular septum (IVSd) and posterior wall (PWd) from 2D parasternal long-axis view were extracted using a backend query in HERON (Healthcare Enterprise Repository for Ontological Narration), a search discovery tool that facilitates searches on various hospital electronic data sources (18,19). The query results were recombined using medical record number, encounter number and study date to generate back the list of variables belonging to each echocardiogram study. Left ventricular mass was calculated using the American Society of Echocardiography recommended formula: $0.8 \times 1.04[(\text{LVIDd} + \text{IVSd} + \text{PWd})^3$ and indexed to body surface area (20).

ECG processing: The details of ECG processing performed using Python are provided in prior publications (15,21,22). In summary, vectorcardiographic X-Y-Z-lead ECGs were constructed from 12-lead ECGs using Kors' matrix (23). Using these orthogonal X, Y, Z leads, the root-mean-square (RMS or 3D) ECG was constructed. Voltage-time integrals (VTIs) were obtained by the integration of the instantaneous voltage over the duration of QRS (VTI_{QRS}) or QRS-T (VTI_{QRST}).

Traditional Criteria and Univariable Models: Based on review of literature, we selected 5 widely used ECG-based LVH diagnostic criteria for comparison, i.e. Peguero-Lo Presti criteria (max S + S_{v4}), Cornell voltage (R_{avL} + S_{v3}), Cornell voltage-duration product (VDP), Sokolow-Lyon criteria (S_{V1} + max $R_{(V5\ or\ V6)}$), and Gubner-Ungerleider critera (R_I + S_{III}). We also selected 3 ECG variables for comparison namely QRS duration, $amplitude_{QRS\text{-}3D}$, and $VTI_{QRS\text{-}3D}$ (21,22,24). The latter 2 were calculated off the QRS from the RMS/3D ECG.

Variational Autoencoder: We trained a variational autoencoder (VAE) on 1.18 million unlabeled ECG signals to encode a 0.75-sec segment centered on the representative beat ECG signal into 60 variables (30 variables for X, Y, Z leads and 30 for RMS of these leads). The VAE has a dual neural network architecture with the encoder taking the ECG input and outputting 30 latent variables, and the decoder inputting the 30 latent variables and outputting the ECG signal. The network is rewarded in training to encode the signal such as to learn accurate reconstruction of the original signal from the latent variables alone. Our VAEs are able to reconstruct the original signal back from the latent variables with high fidelity (16,17,25). The X-Y-Z-lead and RMS/3D representative-beat ECGs included in this study were processed using these 2 VAEs to generate latent encodings or variables.

ECG Features: The following features were available for ML model training:

- Summary features like heart rate, PR interval, QRS duration, corrected QT interval (26), frontal plane QRS axis, etc.
- From 16 leads—each of 12-leads, 3 X-Y-Z-leads and 1 RMS ECG—we obtained QRS amplitudes, VTI_{QRS}, VTI_{QRST}, R-wave amplitudes, S-wave amplitudes.
- 30 latent variables each from VAEs trained to reconstruct the X-Y-Z-lead and RMS representative-lead ECGs.

- Sex

Model Training and Testing: Approximately 10% of the medical record numbers in the dataset were withheld as the testing set, and remainder used for model training (**Figure 1**). We trained the following ML architectures on the training set – logistic regression, random forests, light gradient boosted machine (LGBM), residual neural network (ResNet), multilayered perceptron (MLP) and CNN. The CNN was trained on the representative-beat X-Y-Z-lead ECG signal, and the other 5 ML models trained on the extracted ECG features (as above) plus sex. Sex was provided to the models as the definition of LVH is sex specific. The results are reported from the performance of the trained models in the holdout test set. We also report the models' performance in 4 subgroups based on intraventricular conduction – QRS duration <120 ms, typical right bundle branch block (RBBB, QRS duration ≥120), typical left bundle branch block (LBBB, QRS duration ≥120 ms), and interventricular conduction delay (IVCD, QRS duration ≥ 120 ms but not meeting either RBBB or LBBB criteria). American Heart Association-American College of Cardiology Foundation-Heart Rhythm Society criteria for bundle branch blocks were used (27).

Statistical analysis: Continuous variables are reported as mean ± standard deviation, and categorical variables as percentages. Comparisons were made using Student's t-test for continuous variables and χ^2-test for categorical variables. Statistical analysis was conducted in Python version 3.12.7 and 2-tailed p-value of less than 0.05 was considered statistically significant.

RESULTS

Patient characteristics: A total of 250,596 ECG-echocardiogram pairs were included, with 149,612 (59.7%) pairs belonging to females. The mean age of the overall population of ECG-echocardiogram samples was 63.8 ± 15.3 years. In the training sets, 40,839 (28.2%) of the female samples and 23,309 (24.3%) male samples had LVH on echocardiography. The testing set consisted of 25,263 ECG-echocardiogram pairs. In the testing set, 4470 (27.8%) female samples and 2672 (24.6%) male samples had LVH. The detailed distributions of the ECG and echocardiographic variables in the testing set are shown in **Table 1** and for the training set in **Supplementary Table 1**. The testing samples were divided into 4 subgroups i.e. narrow QRS <120 ms (*n*= 215,228), typical RBBB (*n*=24,800), typical LBBB (*n*=13,893), and IVCD (*n*=13,714).

LVH classification models: The testing set performance of the 3 univariable models, 5 traditional criteria and the 6 ML models is summarized in **Table 2** and **Supplementary Table 2A-D**.

Univariable models: Amongst the linear univariable models, $VTI_{QRS\text{-}3D}$ was the best predictor of LVH in the overall population, with an AUROC 0.677. Further, $VTI_{QRS\text{-}3D}$ performed the best in all subgroups except in typical LBBB (narrow QRS 0.659, RBBB 0.674, LBBB 0.585, IVCD 0.578). In typical LBBB, $amplitude_{QRS\text{-}3D}$ performed the best, with an AUROC 0.590.

Traditional criteria: Overall, the performance of traditional ECG criteria for predicting LVH was poor, with AUROCs ranging from 0.507 to 0.647. Cornell VDP was the best performing criteria overall and in narrow QRS subgroup (overall 0.647; narrow QRS 0.643). In

other subgroups, Peguero-Lo Presti criteria performed the best (RBBB 0.598, LBBB 0.572, IVCD 0.578). In general, these criteria performed better in females as compared to males.

ML Models: All ML models outperformed the traditional criteria and univariate models. LGBM (AUROC 0.790), MLP (0.789) and ResNet (0.788), which were trained on ECG features including VAE latent encodings and sex, were the best performing models in the overall population. The CNN model, which was trained on the raw ECG signal alone, demonstrated an AUROC 0.767. The ROC curves, separately for females and males, for the top 4 ML models vis-à-vis the best univariable and best traditional criteria are plotted in **Figure 2**.

When evaluated in the 4 ECG subgroups by intraventricular conduction, models with highest AUROCs were LGBM in narrow QRS (0.785), MLP in RBBB (0.778) and LBBB (0.698) and ResNet in IVCD (0.720). The ROC curves of the best model each amongst univariable, traditional criteria and ML for each of the 4 subgroups separately for females and males is shown in **Figure 3 and 4**.

Linear analysis of LGBM prediction probabilities: LVMi was plotted against the prediction probabilities output generated by LGBM model for females and males as shown in **Figure 5**. A strong linear trend between prediction probabilities and LVMi can be noted for both females and males (respectively R^2 0.851 and 0.833, or correlation coefficient ρ 0.922 and 0.913).

Longitudinal analysis of LVH negatives: Among false positives and true negatives produced by the LGBM model in the testing set, we searched for the ECG-echocardiogram pairs where a follow-up echocardiogram >1 year and closest to 5 years later was available for further analysis. We used a 2x2 table to compare the development of LVH in 161 false-positive as compared to the 1,019 true-negative samples. On mean follow-up of 3.9 ± 1.8 years, 54/161 (33.5%) patients in false-positive group, and 130/1019 (12.8%) patients in true-negative group

developed LVH. The risk ratio for development of LVH was 2.63 (95% CI 2.01, 3.45) in false-positives compared to true-negatives from the LGBM model (**Table 3**).

DISCUSSION

To the best of our knowledge, this is the largest evaluation of ECG criteria and ML models for predicting LVH till date. We have applied the innovative framework of using DL-based latent space ECG encodings for building ML models, which allows simpler models to make accurate predictions without overfitting.

Salient findings: First, traditional ECG-based criteria demonstrate suboptimal performance in diagnosing LVH, with the Cornell VDP showing the highest accuracy among them (AUROC 0.647). Second, univariable models including QRS duration, $\text{amplitude}_{\text{QRS-3D}}$, and $\text{VTI}_{\text{QRS-3D}}$ were at par or better than traditional criteria for the diagnosis of LVH , with $\text{VTI}_{\text{QRS-3D}}$ achieving the best overall results (AUROC 0.677). Third, ML models outperform both traditional and univariable models, with LGBM models demonstrating the highest performance in our study (overall AUROC 0.790). Last, the performance of traditional, univariable, and ML models vary across sex and QRS morphologies. Further, the LGBM model trained on ECG latent encodings and features successfully captured the underlying trend of LVMi, showing strong correlation and predicting future development of LVH.

Univariable models: Previous studies have demonstrated the utility of linear univariable predictors of LVH, such as QRS duration and QRS-VTIs (22,31). In our analysis, we evaluated QRS duration, $\text{amplitude}_{\text{QRS-3D}}$, and $\text{VTI}_{\text{QRS-3D}}$ for predicting LVH across various subgroups. Our findings indicate that these measures generally outperform traditional LVH criteria. Among them, $\text{VTI}_{\text{QRS-3D}}$ emerged as the best overall criteria, except in the typical LBBB subgroup, where

amplitude$_{QRS\text{-}3D}$ was superior. Similar to Cornell VDP, VTI$_{QRS\text{-}3D}$ incorporates both QRS voltage and duration. Since VTI$_{QRS\text{-}3D}$ is calculated from the reconstructed 3D-orthogonal leads, ostensibly, it captures the QRS complex more comprehensively as compared to Cornell VDP, which uses information from a pair of 2-D leads (V3 and aVL).

Traditional ECG criteria: As demonstrated in previous studies, our analysis reaffirmed the poor discrimination of LVH offered by standard electrocardiographic criteria using a large dataset (28,29). Unlike other voltage-based rules, Cornell VDP, which emerged as the best overall criterion, accounts for both QRS voltage and duration in its calculation. Both of these parameters are affected in LVH (30). In the subset of ECGs with conduction abnormalities (RBBB, LBBB, and IVCD), Peguero-Lo Presti criteria performed better than Cornell VDP. Although the difference in performance was marginal, if this trend is real, it could be explained by obfuscation of LVH-related changes in QRS duration due to QRS prolongation inherent to conduction delays. However, this cannot be verified in our study. Notably, compared to the combined population, individual criteria generally performed better in females and males separately. This underscores the importance of using different cut-off values for females and males, recognizing the sex-based differences in ECGs and definition of LVH. (28,29).

ML models: We tested several ML architectures for LVH prediction, including simple models (LR), tree-based models (RF, LGBM), and neural networks (ResNet, MLP, and CNN). The LGBM model demonstrated the best overall performance (AUROC 0.790), with AUROCs comparable to those of the MLP (0.789) and ResNet (0.788) models. The performance of all the models was worse in the subgroups with conduction abnormalities. MLP was the best performing model in typical RBBB and LBBB subgroups (0.778 and 0.698) while ResNet

performed the best in the IVCD subgroup (0.720). Nevertheless, it is important to note that the differences in the performance these models were only marginal.

We further evaluated the interpretability and physiological relevance of the LGBM model. First, we plotted the prediction probabilities from this model against LVMi, which showed a strong linear positive correlation, suggesting that the model captures meaningful physiological patterns rather than artificial class boundaries. Second, we analyzed the false positives produced by this model for future development of LVH, finding that the false positives were more than 2.5 times as likely to develop LVH in the future compared to true negatives. This indicates that the model captures underlying ECG abnormalities even before patients meet the criteria for overt LVH diagnosis.

Previous literature*:* In a recently published study from China, Zhu et al. used a large dataset comprising of over 90,000 ECGs to create deep learning multilabel classifier algorithms. They achieved AUROCs ranging from 0.78-0.92 using their 12-lead model, and showed that a reduced 4-lead model using lead I, aVR, V1 and V5 had equivalent performance (32). In a Taiwanese study, Liu et al. developed a deep learning model for predicting LVH using approximately 23,000 training samples (33). They achieved high AUROCs ranging from 0.83-0.89 across different testing sets. However, the definition of LVH used in this study was different, using LV mass >186 g for females and >258 g for males. In a South Korean study, Kwon et al. developed an ensemble deep neural network + CNN model using approximately 36,000 training samples, combining information from ECG signal, ECG features, and patient demographics (34). While using higher cut-off values for LVMi (109 g/m^2 females and 132 g/m^2 males), their model achieved AUROCs ranging from 0.87-0.88 in testing sets.

In a study from Massachusetts General Hospital, Haimovich et al. create ML models for predicting LVH in specific disease populations like cardiac amyloidosis, hypertrophic cardiomyopathy, aortic stenosis, and others using a total of 34,258 training samples (35). Similar to our approach, they used a pretrained deep learning model to produce latent encodings and trained a simpler classifier for LVH classification although they used full 10-second ECG signal instead of representative beat ECG. Their model achieved AUROCs ranging from 0.69 to 0.96 in various subgroups. Khurshid et al. used data from the UK Biobank to create a CNN model trained on 32,000 samples and achieved AUROCs ranging from 0.62 to 0.65 in predicting LVH. Owing to heterogeneity in study populations, data structures, and labels for LVH, it is difficult to evaluate the performance of models across studies. Nonetheless, the AUROCs attained by ML models in our study are comparable to previous work.

Limitations: Our work is best understood in the context of its limitations. Both training and testing sets for the models were from a single center, and these models might have sub-optimal performance when generalized to other datasets. Further, since the median beat ECGs were derived from a proprietary system, additional steps may be required in processing ECGs from other systems. Additionally, to calculate ECG parameters for traditional criteria and univariate models, automated feature extraction was done, which might not be as accurate as expert-created labels.

CONCLUSIONS

Traditional voltage-based criteria for ECG diagnosis have poor diagnostic performance. Simple univariable models, especially $VTI_{QRS\text{-}3D}$, perform better than the traditional criteria. ML techniques can significantly enhance the accuracy of ECG-based diagnosis of LVH over both

traditional voltage-based criteria and univariable models. Dimensionality reduction of ECG using variational autoencoder can facilitate utilization of non-deep learning ML architectures, which may otherwise struggle with high dimensionality of ECG data. Further external testing and testing is needed for clinical utilization of these ML models.

ACKNOWLEDGEMENT

Research reported in this publication was supported by the KUMC Research Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the KUMC Research Institute.

This work was supported by a CTSA grant from NCATS awarded to the University of Kansas for Frontiers: University of Kansas Clinical and Translational Science Institute (# UL1TR002366) The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or NCATS.

TABLE LEGENDS

Table 1. Patient characteristics of the testing set.

Table 2. Model performance for LVH prediction in the entire testing set. Area under receiver-operating characteristic curve (AUROC) and sensitivity at specificity fixed at 0.75 are provided.

Table 3. Comparison between presence of LVH on subsequent echocardiogram (>1 year and closest to 5 years after index echocardiogram) in false positives versus true negatives of LVH LGBM model in testing set

FIGURE LEGENDS

Figure 1. Data pipeline for model training and testing

Figure 2. ROC curves from the <u>entire testing set</u> for males (*left panel*) and females (*right panel).*

Figure 3. ROC curves for subgroups of testing set in females, narrow QRS (*top left*), typical right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*), intraventricular conduction delay (IVCD, *bottom right*).

Figure 4. ROC curves for subgroups of testing set in males, narrow QRS (*top left*), typical right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*), intraventricular conduction delay (IVCD, *bottom right*)

Figure 5. Scatterplots of echocardiographic left ventricular mass indexed (LVMi) plotted against prediction probabilities from the LGBM model for females (*left panel*) and males (*right panel*).

REFERENCES

1. Vakili BA, Okin PM, Devereux RB. Prognostic implications of left ventricular hypertrophy. Am Heart J 2001;141:334-41.
2. Sayin BY, Oto A. Left Ventricular Hypertrophy: Etiology-Based Therapeutic Options. Cardiol Ther 2022;11:203-230.
3. Cuspidi C, Meani S, Valerio C, Fusi V, Sala C, Zanchetti A. Left ventricular hypertrophy and cardiovascular risk stratification: impact and cost-effectiveness of echocardiography in recently diagnosed essential hypertensives. Journal of Hypertension 2006;24.
4. Whelton PK, Carey RM, Aronow WS et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018;71:e13-e115.
5. Tison GH, Zhang J, Delling FN, Deo RC. Automated and Interpretable Patient ECG Profiles for Disease Detection, Tracking, and Discovery. Circulation: Cardiovascular Quality and Outcomes 2019;12:e005289.
6. Ricciardi D, Vetta G, Nenna A et al. Current diagnostic ECG criteria for left ventricular hypertrophy: is it time to change paradigm in the analysis of data? Journal of Cardiovascular Medicine 2020;21.
7. Leese PJ, Viera AJ, Hinderliter AL, Stearns SC. Cost-Effectiveness of Electrocardiography vs. Electrocardiography Plus Limited Echocardiography to Diagnose LVH in Young, Newly Identified, Hypertensives. American Journal of Hypertension 2010;23:592-598.
8. 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 2009;119:e251-61.
9. Bacharova L, Chevalier P, Gorenek B et al. ISE/ISHNE Expert Consensus Statement on ECG Diagnosis of Left Ventricular Hypertrophy: The Change of the Paradigm. The joint paper of the International Society of Electrocardiology and the International Society for Holter Monitoring and Noninvasive Electrocardiology. Journal of Electrocardiology 2023;81:85-93.
10. Ose B, Sattar Z, Gupta A, Toquica C, Harvey C, Noheria A. Artificial Intelligence Interpretation of the Electrocardiogram: A State-of-the-Art Review. Curr Cardiol Rep 2024;26:561-580.
11. Ranka S, Reddy M, Noheria A. Artificial intelligence in cardiovascular medicine. Curr Opin Cardiol 2021;36:26-35.
12. 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. Scientific Reports 2024;14:15882.
13. Ying X. An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series 2019;1168:022022.

14. Kligfield P, Gettes LS, Bailey JJ et al. Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology 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. J Am Coll Cardiol 2007;49:1109-27.
15. Harvey CJ, Shomaji S, Yao Z, Noheria A. Comparison of Autoencoder Encodings for ECG Representation in Downstream Prediction Tasks. arXiv preprint 2024:2410.02937.
16. Harvey C, Noheria A. DEEP LEARNING ENCODED ECG - AVOIDING OVERFITTING IN ECG MACHINE LEARNING. Journal of the American College of Cardiology 2024;83:172-172.
17. Harvey C, Noheria A. REDUCING DATA DIMENSIONALITY OF ECG SIGNAL USING DEEP LEARNING. Journal of the American College of Cardiology 2024;83:26-26.
18. Murphy SN, Weber G, Mendis M et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010;17:124-30.
19. Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. AMIA Annu Symp Proc 2011;2011:1454-63.
20. 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 2015;28:1-39 e14.
21. Fairbank T, DeBauge A, Harvey CJ et al. Electrocardiographic Z-axis QRS-T voltage-time-integral in patients with typical right bundle branch block - Correlation with echocardiographic right ventricular size and function. J Electrocardiol 2024;82:73-79.
22. DeBauge A, Fairbank T, Harvey CJ et al. Electrocardiographic prediction of left ventricular hypertrophy in women and men with left bundle branch block - Comparison of QRS duration, amplitude and voltage-time-integral. J Electrocardiol 2023;80:34-39.
23. Kors JA, van Herpen G, Sittig AC, van Bemmel JH. Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: diagnostic comparison of different methods. Eur Heart J 1990;11:1083-92.
24. DeBauge A, Harvey CJ, Gupta A et al. Evaluation of electrocardiographic criteria for predicting left ventricular hypertrophy and dilation in presence of left bundle branch block. Journal of Electrocardiology 2024;87:153787.
25. Harvey CJ, Shomaji S, Yao Z, Noheria A. Comparison of Autoencoder Encodings for ECG Representation in Downstream Prediction Tasks: arXiv.
26. Fridericia LS. Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken. Acta Medica Scandinavica 1920;53:469-486.
27. Surawicz B, Childers R, Deal BJ et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part III: intraventricular conduction disturbances: 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 2009;119:e235-40.

28. Fragola PV, Autore C, Ruscitti G, Picelli A, Cannata D. Electrocardiographic diagnosis of left ventricular hypertrophy in the presence of left bundle branch block: a wasted effort. Int J Cardiol 1990;28:215-21.

29. Haskell RJ, Ginzton LE, Laks MM. Electrocardiographic diagnosis of left ventricular hypertrophy in the presence of left bundle branch block. J Electrocardiol 1987;20:227-32.

30. Molloy TJ, Okin PM, Devereux RB, Kligfield P. Electrocardiographic detection of left ventricular hypertrophy by the simple QRS voltage-duration product. J Am Coll Cardiol 1992;20:1180-6.

31. Okin PM, Roman MJ, Devereux RB, Kligfield P. Time-Voltage Area of the QRS for the Identification of Left Ventricular Hypertrophy. Hypertension 1996;27:251-258.

32. Zhu H, Jiang Y, Cheng C et al. Four-Channel ECG as a Single Source for Early Diagnosis of Cardiac Hypertrophy and Dilation — A Deep Learning Approach. NEJM AI 2024;1:AIoa2300297.

33. Liu C-M, Hsieh M-E, Hu Y-F et al. Artificial Intelligence–Enabled Model for Early Detection of Left Ventricular Hypertrophy and Mortality Prediction in Young to Middle-Aged Adults. Circulation: Cardiovascular Quality and Outcomes 2022;15:e008360.

34. Kwon J-M, Jeon K-H, Kim HM et al. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. EP Europace 2020;22:412-419.

35. Haimovich JS, Diamant N, Khurshid S et al. Artificial intelligence–enabled classification of hypertrophic heart diseases using electrocardiograms. Cardiovascular Digital Health Journal 2023;4:48-59.

Table 1. Patient characteristics of the testing set.

	Females			Males		
	No LVH	**LVH**	**p**	**No LVH**	**LVH**	**p**
n (%)	10782 (71.6%)	4268 (28.4%)		7466 (73.1%)	2747 (26.9%)	
Age (years), mean±S.D.	61.6 ± 16.6	67.4 ± 16.1	< 0.0001	63.2 ± 14.9	67.0 ± 13.6	< 0.0001
QRS duration (ms), mean±S.D.	96.1 ± 23.3	108.6 ± 27.8	< 0.0001	104.8 ± 25.0	123.6 ± 33.2	< 0.0001
Frontal plane QRS axis (°), mean±S.D.	28.5 ± 53.1	20.3 ± 66.3	< 0.0001	24.2 ± 58.7	24.9 ± 82.3	< 0.0001
$Amplitude_{QRS-3D}$ (µV), mean±S.D.	938.7 ± 405.5	1128.9 ± 525.5	< 0.0001	975.1 ± 416.6	1147.5 ± 525.1	< 0.0001
VTI_{QRS-3D} (nVs), mean±S.D.	30507.8 ± 16684.7	42976.0 ± 24444.6	< 0.0001	34575.5 ± 17143.0	49621.2 ± 25866.2	< 0.0001
QTc (ms), mean±S.D.	431.6 ± 40.1	450.0 ± 45.5	< 0.0001	433.5 ± 43.2	458.9 ± 49.5	< 0.0001
LVEF (%), mean±S.D.	57.7 ± 10.3	50.9 ± 15.6	< 0.0001	54.4 ± 12.4	43.7 ± 17.5	< 0.0001
LV mass index (g/m^2), mean±S.D.	70.2 ± 14.2	122.6 ± 224.6	< 0.0001	83.2 ± 17.7	150.2 ± 192.3	< 0.0001
ECG-defined subgroups			<0.0001			<0.0001
Narrow QRS[a]	9279 (86.1%)	3091 (72.7%)		5836 (78.5%)	1556 (56.8%)	
Typical RBBB[b]	664 (6.2%)	439 (10.3%)		902 (12.1%)	530 (19.4%)	
Typical LBBB[b]	434 (4.0%)	375 (8.8%)		253 (3.4%)	296 (10.8%)	
IVCD[b]	404 (3.7%)	346 (8.1%)		448 (6.0%)	356 (13.0%)	

RBBB, Right Bundle Branch Block; LBBB, Left Bundle Branch Block; IVCD, Intraventricular Conduction Delay

[a]QRS duration <120 ms, [b]QRS duration ≥120 ms

Table 2. Model performance for LVH prediction in the entire testing set. Area under receiver-operating characteristic curve (AUROC) and sensitivity at specificity fixed at 0.75 are provided.

Testing set	**Combined (n=25,263)**		**Females (n=15,050)**		**Males (n=10,213)**	
	AUROC	**Sensitivity***	**AUROC**	**Sensitivity***	**AUROC**	**Sensitivity***
Univariable models[a]						
▪ QRS duration	0.648	0.462	0.651	0.452	0.667	0.46
▪ $\text{Amplitude}_{\text{QRS-3D}}$	0.601	0.389	0.602	0.395	0.603	0.388
▪ $\text{VTI}_{\text{QRS-3D}}$	**0.677**	**0.515**	**0.671**	**0.504**	**0.699**	**0.543**
Traditional criteria						
▪ Peguero-Lo Presti voltage (max S + S_{V4})	0.626	0.384	0.634	0.392	0.623	0.377
▪ Cornell voltage (R_{avL} + S_{V3})	0.617	0.388	0.631	0.413	0.603	0.352
▪ Cornell VDP	**0.647**	**0.419**	**0.657**	**0.425**	**0.644**	**0.401**
▪ Sokolow-Lyon voltage (S_{V1} + max R (V5 or V6))	0.507	0.283	0.514	0.31	0.502	0.264
▪ Gubner-Ungerleider voltage (R_I + S_{III})	0.541	0.317	0.560	0.334	0.511	0.295
ML models						
▪ Logistic regression[b]	0.764	0.647	0.759	0.634	0.771	0.656
▪ Random forest[b]	0.772	0.654	0.766	0.646	0.783	0.679
▪ Light gradient boosted machine[b]	**0.790**	**0.687**	**0.787**	**0.677**	**0.796**	**0.703**
▪ Residual network[b]	0.788	0.684	0.783	0.674	0.796	0.693
▪ Multilayered perceptron network[b]	0.789	0.686	0.784	0.676	0.796	0.702
▪ Convolutional neural network[c]	0.767	0.638	0.773	0.646	0.769	0.644

* At specificity 0.75

[a] Logistic regressions

[b] Input of 117 ECG statistics like QRS duration, heart rate etc. and 60 variational autoencoder latent variables from ECG representative beat and sex

[c] Input of representative-beat ECG signal (X, Y, Z leads)

Table 3. Comparison between presence of LVH on subsequent echocardiogram (>1 year and closest to 5 years after index echocardiogram) in false positives versus true negatives of LVH LGBM model in testing set

	Follow-up echocardiogram		**Total**	**Risk**	**Risk Ratio**	**p**
	No LVH	**LVH**				
True negatives	889	130	1,019	0.13	Ref.	-
False positives	107	54	161	0.34	2.63 (2.01-3.45)	<0.0001
Total	996	184	1,180	-	-	-

Figure 1. Data pipeline for model training and testing

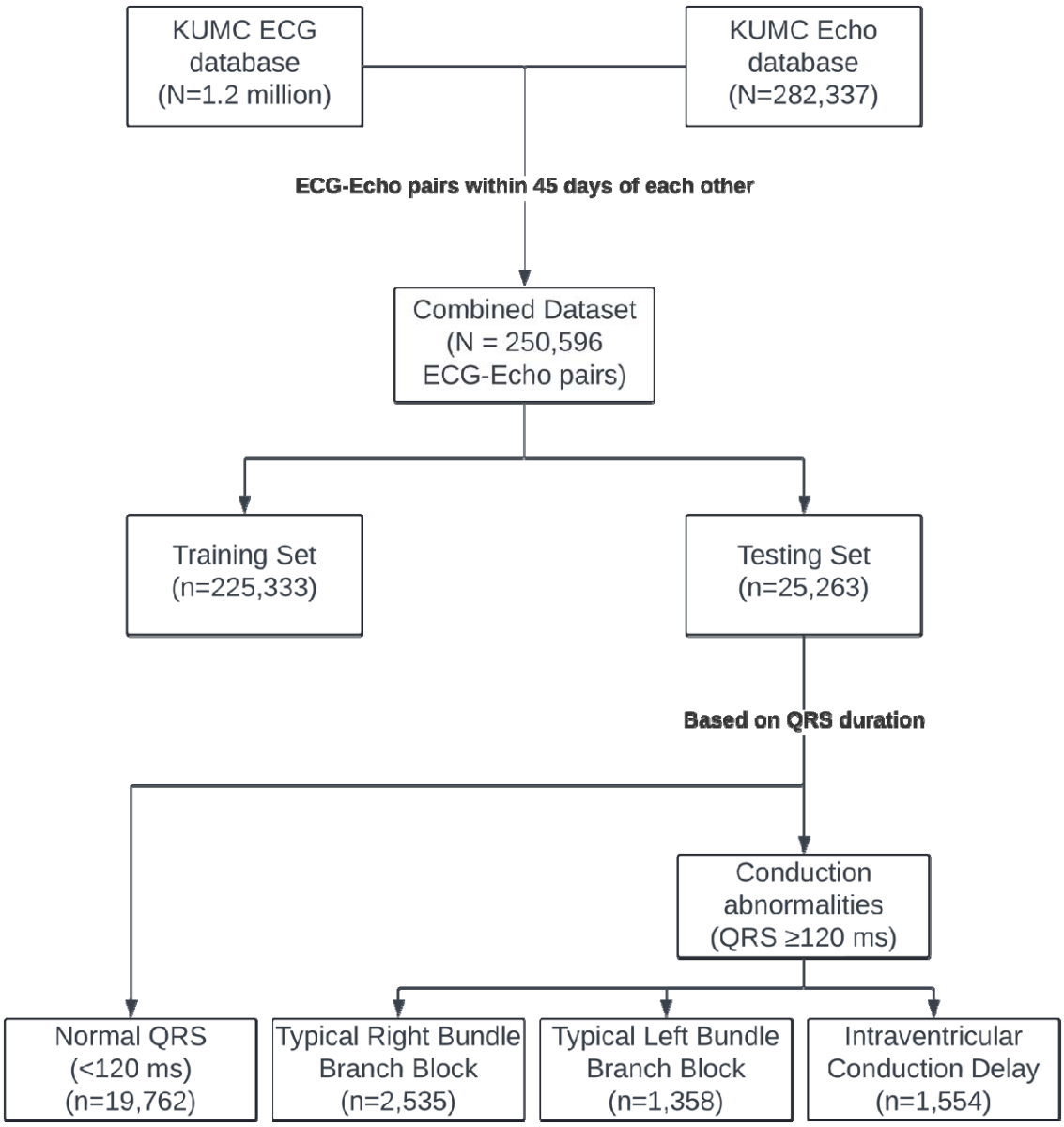

Figure 2. ROC curves from the <u>entire testing set</u> for males (*left panel*) and females (*right panel).*

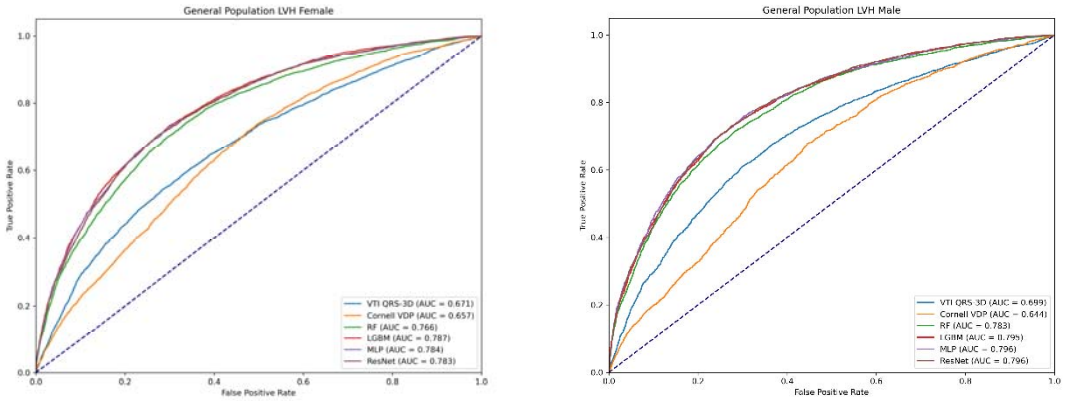

Figure 3. ROC curves for subgroups of females in testing set, narrow QRS (*top left*), typical right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*), intraventricular conduction delay (IVCD, *bottom right*).

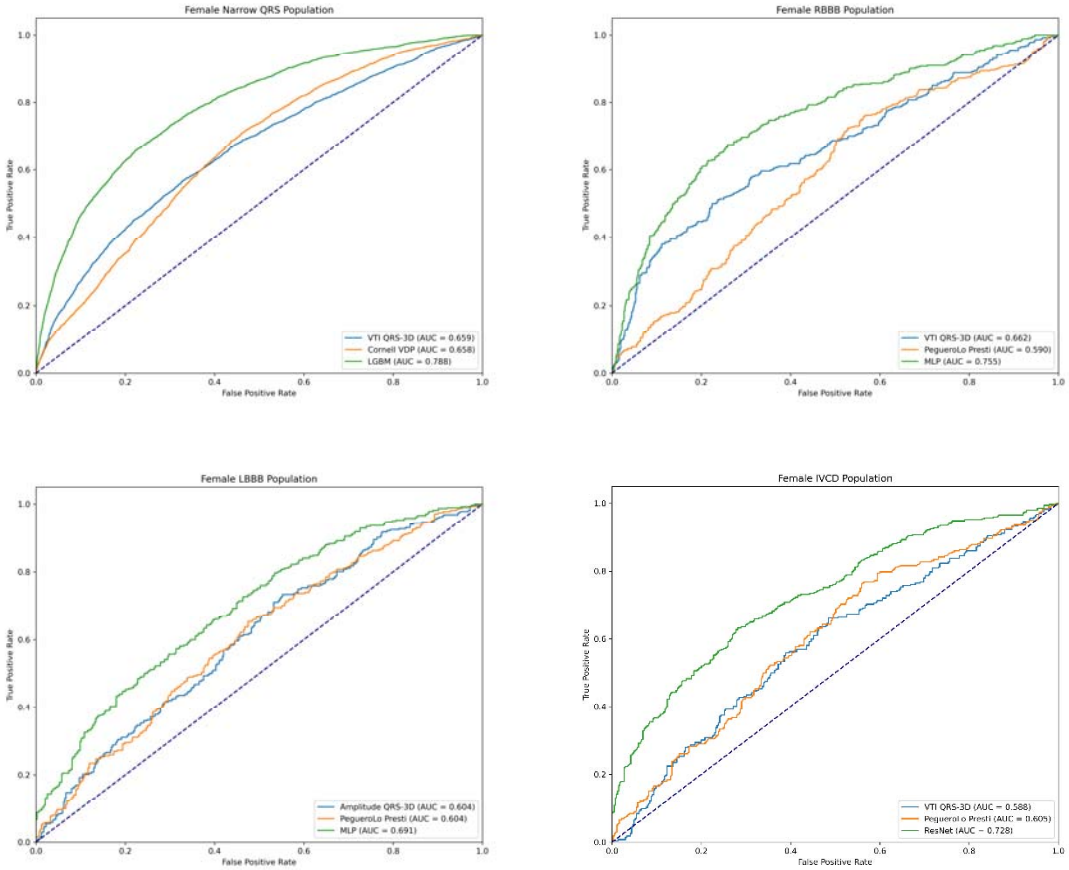

Figure 4. ROC curves for subgroups of males in testing set, narrow QRS (*top left*), typical right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*), intraventricular conduction delay (IVCD, *bottom right*)

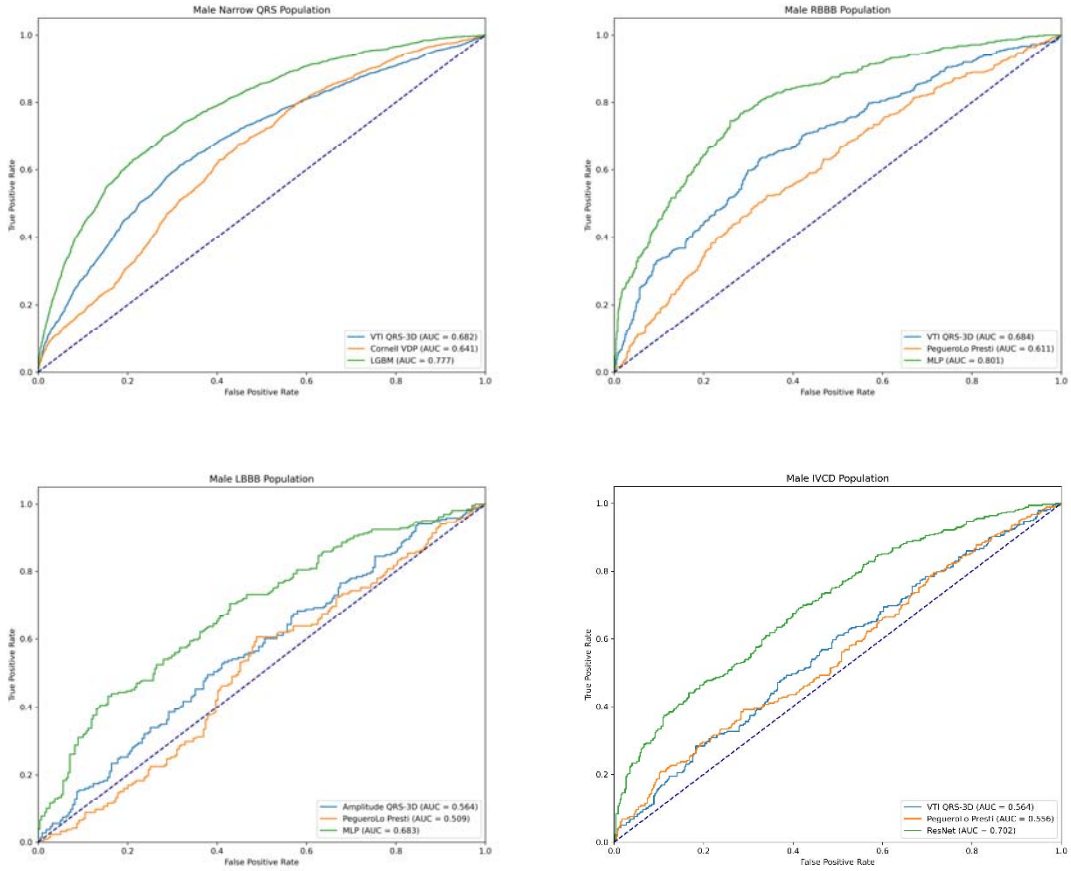
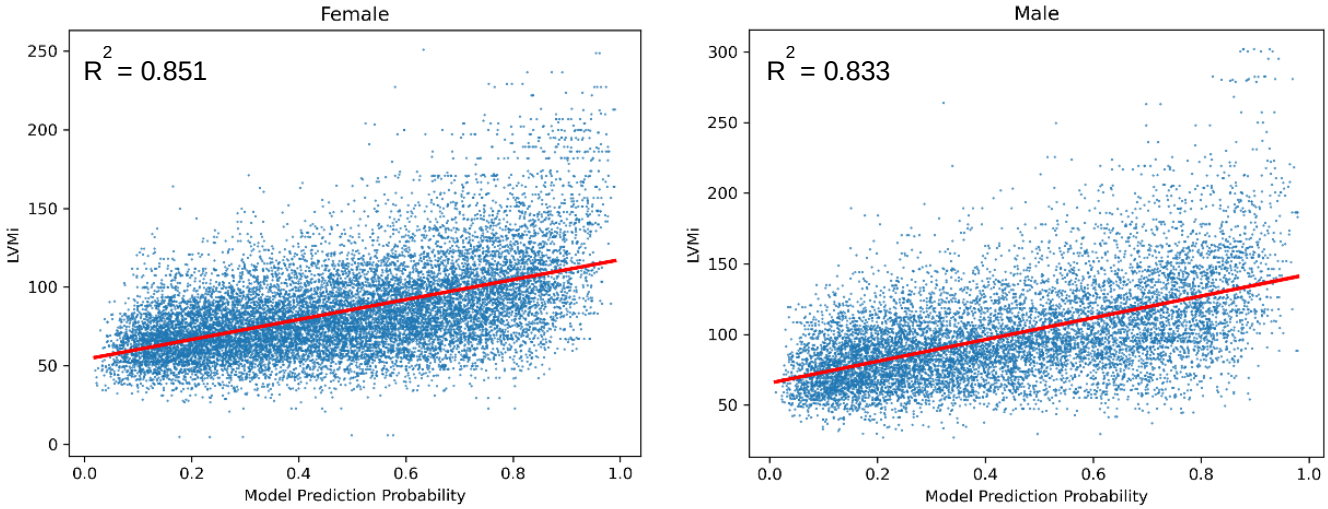
Figure 5. Scatterplots of echocardiographic left ventricular mass indexed (LVMi) plotted against prediction probabilities from the LGBM model for females (*left panel*) and males (*right panel*).