1 **Machine learning to classify left ventricular hypertrophy using ECG feature**

2 **extraction by variational autoencoder**

3 Short title: ML models for LVH

4

5 Amulya Gupta, MBBS,[1]* Christopher J. Harvey,[1]* Ashley DeBauge, MD,[2] Sumaiya Shomaji,[3]

6 PhD, Zijun Yao, PhD,[3] Amit Noheria, MBBS, SM[1]

7 [1]Department of Cardiovascular Medicine, The University of Kansas Medical Center, Kansas

8 City, Kansas

9 [2]Department of Internal Medicine, Washington University School of Medicine, St. Louis,

10 Missouri, USA

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

12 Lawrence, Kansas

13 *Contributed equally to the manuscript

14

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

16 Corresponding Author: Amit Noheria, MBBS, SM

17 3901 Rainbow Blvd.

18 Kansas City, KS 66160

19 Phone: (913)588-9600

20 Email: noheriaa@gmail.com

21

22 Disclosures: None

23

24   **ABSTRACT**

25   **Background**: Traditional ECG criteria for left ventricular hypertrophy (LVH) have low

26   diagnostic yield. Machine learning (ML) can improve ECG classification.

27   **Methods**: ECG summary features (rate, intervals, axis), R-wave, S-wave and overall-

28   QRS amplitudes, and QRS/QRST voltage-time integrals (VTIs) were extracted from 12-lead,

29   vectorcardiographic X-Y-Z-lead, and root-mean-square (3D) representative-beat ECGs. Latent

30   features were extracted by variational autoencoder from X-Y-Z and 3D representative-beat

31   ECGs. Logistic regression, random forest, light gradient boosted machine (LGBM), residual

32   network (ResNet) and multilayer perceptron network (MLP) models using ECG features and sex,

33   and a convolutional neural network (CNN) using ECG signals, were trained to predict LVH (left

34   ventricular mass indexed in women >95□g/m², men >115□g/m²) on 225,333 adult ECG-

35   echocardiogram (within 45 days) pairs. AUROCs for LVH classification were obtained in a

36   separate test set for individual ECG variables, traditional criteria and ML models.

37   **Results**: In the test set (n=25,263), AUROC for LVH classification was higher for ML

38   models using ECG features (LGBM 0.790, MLP 0.789, ResNet 0.788) as compared to the best

39   individual variable (VTI$_{QRS-3D}$ 0.677), the best traditional criterion (Cornell voltage-duration

40   product 0.647) and CNN using ECG signal (0.767). Among patients without LVH who had a

41   follow-up echocardiogram >1 (closest to 5) years later, LGBM false positives, compared to true

42   negatives, had a 2.63 (95% CI 2.01, 3.45)-fold higher risk for developing LVH (p<0.0001).

43   **Conclusions**: ML models are superior to traditional ECG criteria to classify—and predict

44   future—LVH. Models trained on extracted ECG features, including variational autoencoder

45   latent variables, outperformed CNN directly trained on ECG signal.

46

47   **Keywords**: Left ventricular hypertrophy, LVH, machine learning, deep learning, artificial

48   intelligence, electrocardiogram, ECG, variational autoencoder.

49

50   **Abbreviations**:

51   ECG, electrocardiogram

52   LVH, left ventricular hypertrophy

53   ML, machine learning

54   AI, artificial intelligence

55   MLP, multilayered perceptron

56   LGBM, light gradient-boosting machine

57   AUROC, area under the receiver operator characteristic curve

58   VAE, variational Autoencoder

59   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.

83   Machine learning (ML) can reduce reliance on human interpretation and yet increase the

84   diagnostic accuracy of ECG (10,11). Several ECG-based ML models have been developed for

85   detecting LVH, with varying sensitivities and specificities (12). Many of these studies use

86   convolutional neural network (CNN) deep learning architecture to train models using ECG

87   signals often with fewer than 10,000 training ECGs. Given that each 12-lead 10-second ECG

88   signal at 500 Hz consists of 60,000 data points, using such a high-dimensionality input for ML

89   training with a limited number of samples can result in overfitting and reduced generalizability

90   (13-15). On the other hand, non-neural network ML architectures—such as logistic regression,

91   random forest, gradient boosted machine—are not suited to use high-dimensional ECG signal

92   data as input and are usually limited to using extracted ECG features with potential loss of

93   diagnostic information (15).

94   To mitigate these limitations—while preserving the advantages of deep learning—we

95   developed a variational autoencoder (VAE) that can encode 0.75-sec-representative-beat from

96   either X-Y-Z-lead or root-mean-squared ECG into 30 variables (15-17). These VAE latent

97   encodings retain the ECG morphological information and can reconstruct back the ECG signal

98   with high fidelity. In this study, we aimed to train and test different ML models using extracted

99   ECG features including the latent encodings or the ECG signal to classify LVH from the

100   representative-beat ECG.

101

102   **METHODS**

103   Patient selection and data retrieval: An automated retrospective retrieval of records was

104   performed from our clinical database at the University of Kansas Medical Center between May

105   2010 and Jan 2022 to search for ECG and echocardiogram performed on the same patient within

106   45 days of each other. Echocardiograms-ECG pairs with echocardiographic left ventricular mass

107   index (LVMi) >95 g/m^2 for females and >115 g/m^2 for males were labelled as 'LVH' while rest

108   of the pairs were assigned to the 'no LVH' group (15). The study was conducted under an

109   approval from the Institutional Review Board.

110   Data extraction: ECGs were acquired with Philips 12-lead ECG machines. The 12-lead

111   ECG 10-second and 1200-ms-representative-beat signals along with standard features like heart

112   rate, PR interval, etc. were exported to a research SQL data server. Echocardiograms were

113   standard clinical studies performed for clinical indications both as outpatient and inpatient

114   evaluations. Individual echocardiogram numeric variables including diastolic measurements of

115   left ventricular internal diameter (LVIDd), interventricular septum (IVSd) and posterior wall

116   (PWd) from 2D parasternal long-axis view were extracted using a backend query in HERON

117   (Healthcare Enterprise Repository for Ontological Narration), a search discovery tool that

118   facilitates searches on various hospital electronic data sources (18,19). The query results were

119   recombined using medical record number, encounter number and study date to generate back the

120   list of variables belonging to each echocardiogram study. Left ventricular mass was calculated

121   using the American Society of Echocardiography recommended formula: $0.8 \times 1.04[(\text{LVIDd} +$

122    $\text{IVSd} + \text{PWd})^3$ and indexed to body surface area (20).

123   ECG processing: The details of ECG processing performed using Python are provided in

124   prior publications (15,21,22). In summary, vectorcardiographic X-Y-Z-lead ECGs were

125   constructed from 12-lead ECGs using Kors' matrix (23). Using these orthogonal X, Y, Z leads,

126   the root-mean-square (RMS or 3D) ECG was constructed. Voltage-time integrals (VTIs) were

127   obtained by the integration of the instantaneous voltage over the duration of QRS (VTI$_{QRS}$) or

128   QRS-T (VTI$_{QRST}$).

129   <u>Traditional Criteria and Univariable Models</u>: Based on review of literature, we selected 5

130   widely used ECG-based LVH diagnostic criteria for comparison, i.e. Peguero-Lo Presti criteria

131   (max S + S_{v4}), Cornell voltage (R_{avL} + S_{v3}), Cornell voltage-duration product (VDP), Sokolow-

132   Lyon criteria (S_{V1} + max R $_{(V5\ or\ V6)}$), and Gubner-Ungerleider critera (R_I + S_{III}). We also selected

133   3 ECG variables for comparison namely QRS duration, amplitude$_{QRS-3D}$, and VTI$_{QRS-3D}$

134   (21,22,24). The latter 2 were calculated off the QRS from the RMS/3D ECG.

135   <u>Variational Autoencoder</u>: We trained a variational autoencoder (VAE) on 1.18 million

136   unlabeled ECG signals to encode a 0.75-sec segment centered on the representative beat ECG

137   signal into 60 variables (30 variables for X, Y, Z leads and 30 for RMS of these leads). The VAE

138   has a dual neural network architecture with the encoder taking the ECG input and outputting 30

139   latent variables, and the decoder inputting the 30 latent variables and outputting the ECG signal.

140   The network is rewarded in training to encode the signal such as to learn accurate reconstruction

141   of the original signal from the latent variables alone. Our VAEs are able to reconstruct the

142   original signal back from the latent variables with high fidelity (16,17,25). The X-Y-Z-lead and

143   RMS/3D representative-beat ECGs included in this study were processed using these 2 VAEs to

144   generate latent encodings or variables.

145   <u>ECG Features</u>: The following features were available for ML model training:

146   •   Summary features like heart rate, PR interval, QRS duration, corrected QT interval (26),

147   frontal plane QRS axis, etc.

148   •   From 16 leads—each of 12-leads, 3 X-Y-Z-leads and 1 RMS ECG—we obtained QRS

149   amplitudes, VTI$_{QRS}$, VTI$_{QRST}$, R-wave amplitudes, S-wave amplitudes.

150   •   30 latent variables each from VAEs trained to reconstruct the X-Y-Z-lead and RMS

151   representative-lead ECGs.

152   • Sex

153   Model Training and Testing: Approximately 10% of the medical record numbers in the

154   dataset were withheld as the testing set, and remainder used for model training (**Figure 1**). We

155   trained the following ML architectures on the training set – logistic regression, random forests,

156   light gradient boosted machine (LGBM), residual neural network (ResNet), multilayered

157   perceptron (MLP) and CNN. The CNN was trained on the representative-beat X-Y-Z-lead ECG

158   signal, and the other 5 ML models trained on the extracted ECG features (as above) plus sex. Sex

159   was provided to the models as the definition of LVH is sex specific. The results are reported

160   from the performance of the trained models in the holdout test set. We also report the models'

161   performance in 4 subgroups based on intraventricular conduction – QRS duration <120 ms,

162   typical right bundle branch block (RBBB, QRS duration ≥120), typical left bundle branch block

163   (LBBB, QRS duration ≥120 ms), and interventricular conduction delay (IVCD, QRS duration ≥

164   120 ms but not meeting either RBBB or LBBB criteria). American Heart Association-American

165   College of Cardiology Foundation-Heart Rhythm Society criteria for bundle branch blocks were

166   used (27).

167   Statistical analysis: Continuous variables are reported as mean ± standard deviation, and

168   categorical variables as percentages. Comparisons were made using Student's t-test for

169   continuous variables and χ^2-test for categorical variables. Statistical analysis was conducted in

170   Python version 3.12.7 and 2-tailed p-value of less than 0.05 was considered statistically

171   significant.

172

173

174

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-3D} was the best predictor of LVH in the overall population, with an AUROC 0.677. Further, VTI_{QRS-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, $amplitude_{QRS-3D}$ 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

197   other subgroups, Peguero-Lo Presti criteria performed the best (RBBB 0.598, LBBB 0.572,

198   IVCD 0.578). In general, these criteria performed better in females as compared to males.

199   ML Models: All ML models outperformed the traditional criteria and univariate models.

200   LGBM (AUROC 0.790), MLP (0.789) and ResNet (0.788), which were trained on ECG features

201   including VAE latent encodings and sex, were the best performing models in the overall

202   population. The CNN model, which was trained on the raw ECG signal alone, demonstrated an

203   AUROC 0.767. The ROC curves, separately for females and males, for the top 4 ML models vis-

204   à-vis the best univariable and best traditional criteria are plotted in **Figure 2**.

205   When evaluated in the 4 ECG subgroups by intraventricular conduction, models with

206   highest AUROCs were LGBM in narrow QRS (0.785), MLP in RBBB (0.778) and LBBB

207   (0.698) and ResNet in IVCD (0.720). The ROC curves of the best model each amongst

208   univariable, traditional criteria and ML for each of the 4 subgroups separately for females and

209   males is shown in **Figure 3 and 4**.

210   Linear analysis of LGBM prediction probabilities: LVMi was plotted against the

211   prediction probabilities output generated by LGBM model for females and males as shown in

212   **Figure 5**. A strong linear trend between prediction probabilities and LVMi can be noted for both

213   females and males (respectively R^2 0.851 and 0.833, or correlation coefficient ρ 0.922 and 0.913).

214   Longitudinal analysis of LVH negatives: Among false positives and true negatives

215   produced by the LGBM model in the testing set, we searched for the ECG-echocardiogram pairs

216   where a follow-up echocardiogram >1 year and closest to 5 years later was available for further

217   analysis. We used a 2x2 table to compare the development of LVH in 161 false-positive as

218   compared to the 1,019 true-negative samples. On mean follow-up of 3.9 ± 1.8 years, 54/161

219   (33.5%) patients in false-positive group, and 130/1019 (12.8%) patients in true-negative group

220   developed LVH. The risk ratio for development of LVH was 2.63 (95% CI 2.01, 3.45) in false-

221   positives compared to true-negatives from the LGBM model (**Table 3**).

222

223   **DISCUSSION**

224   To the best of our knowledge, this is the largest evaluation of ECG criteria and ML models for

225   predicting LVH till date. We have applied the innovative framework of using DL-based latent

226   space ECG encodings for building ML models, which allows simpler models to make accurate

227   predictions without overfitting.

228   *Salient findings*: First, traditional ECG-based criteria demonstrate suboptimal

229   performance in diagnosing LVH, with the Cornell VDP showing the highest accuracy among

230   them (AUROC 0.647). Second, univariable models including QRS duration, $\text{amplitude}_{\text{QRS-3D}}$,

231   and $\text{VTI}_{\text{QRS-3D}}$ were at par or better than traditional criteria for the diagnosis of LVH , with

232    $\text{VTI}_{\text{QRS-3D}}$ achieving the best overall results (AUROC 0.677). Third, ML models outperform both

233   traditional and univariable models, with LGBM models demonstrating the highest performance

234   in our study (overall AUROC 0.790). Last, the performance of traditional, univariable, and ML

235   models vary across sex and QRS morphologies. Further, the LGBM model trained on ECG latent

236   encodings and features successfully captured the underlying trend of LVMi, showing strong

237   correlation and predicting future development of LVH.

238   *Univariable models*: Previous studies have demonstrated the utility of linear univariable

239   predictors of LVH, such as QRS duration and QRS-VTIs (22,31). In our analysis, we evaluated

240   QRS duration, $\text{amplitude}_{\text{QRS-3D}}$, and $\text{VTI}_{\text{QRS-3D}}$ for predicting LVH across various subgroups. Our

241   findings indicate that these measures generally outperform traditional LVH criteria. Among them,

242    $\text{VTI}_{\text{QRS-3D}}$ emerged as the best overall criteria, except in the typical LBBB subgroup, where

243   amplitude$_{QRS-3D}$ was superior. Similar to Cornell VDP, VTI$_{QRS-3D}$ incorporates both QRS voltage

244   and duration. Since VTI$_{QRS-3D}$ is calculated from the reconstructed 3D-orthogonal leads,

245   ostensibly, it captures the QRS complex more comprehensively as compared to Cornell VDP,

246   which uses information from a pair of 2-D leads (V3 and aVL).

247   ***Traditional ECG criteria***: As demonstrated in previous studies, our analysis reaffirmed

248   the poor discrimination of LVH offered by standard electrocardiographic criteria using a large

249   dataset (28,29). Unlike other voltage-based rules, Cornell VDP, which emerged as the best

250   overall criterion, accounts for both QRS voltage and duration in its calculation. Both of these

251   parameters are affected in LVH (30). In the subset of ECGs with conduction abnormalities

252   (RBBB, LBBB, and IVCD), Peguero-Lo Presti criteria performed better than Cornell VDP.

253   Although the difference in performance was marginal, if this trend is real, it could be explained

254   by obfuscation of LVH-related changes in QRS duration due to QRS prolongation inherent to

255   conduction delays. However, this cannot be verified in our study. Notably, compared to the

256   combined population, individual criteria generally performed better in females and males

257   separately. This underscores the importance of using different cut-off values for females and

258   males, recognizing the sex-based differences in ECGs and definition of LVH. (28,29).

259   ***ML models***: We tested several ML architectures for LVH prediction, including simple

260   models (LR), tree-based models (RF, LGBM), and neural networks (ResNet, MLP, and CNN).

261   The LGBM model demonstrated the best overall performance (AUROC 0.790), with AUROCs

262   comparable to those of the MLP (0.789) and ResNet (0.788) models. The performance of all the

263   models was worse in the subgroups with conduction abnormalities. MLP was the best

264   performing model in typical RBBB and LBBB subgroups (0.778 and 0.698) while ResNet

265   performed the best in the IVCD subgroup (0.720). Nevertheless, it is important to note that the

266   differences in the performance these models were only marginal.

267   We further evaluated the interpretability and physiological relevance of the LGBM model.

268   First, we plotted the prediction probabilities from this model against LVMi, which showed a

269   strong linear positive correlation, suggesting that the model captures meaningful physiological

270   patterns rather than artificial class boundaries. Second, we analyzed the false positives produced

271   by this model for future development of LVH, finding that the false positives were more than 2.5

272   times as likely to develop LVH in the future compared to true negatives. This indicates that the

273   model captures underlying ECG abnormalities even before patients meet the criteria for overt

274   LVH diagnosis.

275   *Previous literature:* In a recently published study from China, Zhu et al. used a large

276   dataset comprising of over 90,000 ECGs to create deep learning multilabel classifier algorithms.

277   They achieved AUROCs ranging from 0.78-0.92 using their 12-lead model, and showed that a

278   reduced 4-lead model using lead I, aVR, V1 and V5 had equivalent performance (32). In a

279   Taiwanese study, Liu et al. developed a deep learning model for predicting LVH using

280   approximately 23,000 training samples (33). They achieved high AUROCs ranging from 0.83-

281   0.89 across different testing sets. However, the definition of LVH used in this study was different,

282   using LV mass >186 g for females and >258 g for males. In a South Korean study, Kwon et al.

283   developed an ensemble deep neural network + CNN model using approximately 36,000 training

284   samples, combining information from ECG signal, ECG features, and patient demographics (34).

285   While using higher cut-off values for LVMi (109 g/m^2 females and 132 g/m^2 males), their model

286   achieved AUROCs ranging from 0.87-0.88 in testing sets.

287   In a study from Massachusetts General Hospital, Haimovich et al. create ML models for

288   predicting LVH in specific disease populations like cardiac amyloidosis, hypertrophic

289   cardiomyopathy, aortic stenosis, and others using a total of 34,258 training samples (35). Similar

290   to our approach, they used a pretrained deep learning model to produce latent encodings and

291   trained a simpler classifier for LVH classification although they used full 10-second ECG signal

292   instead of representative beat ECG. Their model achieved AUROCs ranging from 0.69 to 0.96 in

293   various subgroups. Khurshid et al. used data from the UK Biobank to create a CNN model

294   trained on 32,000 samples and achieved AUROCs ranging from 0.62 to 0.65 in predicting LVH.

295   Owing to heterogeneity in study populations, data structures, and labels for LVH, it is difficult to

296   evaluate the performance of models across studies. Nonetheless, the AUROCs attained by ML

297   models in our study are comparable to previous work.

298   *Limitations*: Our work is best understood in the context of its limitations. Both training

299   and testing sets for the models were from a single center, and these models might have sub-

300   optimal performance when generalized to other datasets. Further, since the median beat ECGs

301   were derived from a proprietary system, additional steps may be required in processing ECGs

302   from other systems. Additionally, to calculate ECG parameters for traditional criteria and

303   univariate models, automated feature extraction was done, which might not be as accurate as

304   expert-created labels.

305

306   **CONCLUSIONS**

307 Traditional voltage-based criteria for ECG diagnosis have poor diagnostic performance. Simple

308 univariable models, especially VTI_{QRS-3D}, perform better than the traditional criteria. ML

309 techniques can significantly enhance the accuracy of ECG-based diagnosis of LVH over both

310   traditional voltage-based criteria and univariable models. Dimensionality reduction of ECG

311   using variational autoencoder can facilitate utilization of non-deep learning ML architectures,

312   which may otherwise struggle with high dimensionality of ECG data. Further external testing

313   and testing is needed for clinical utilization of these ML models.

314

315

316   **ACKNOWLEDGEMENT**

324   **TABLE LEGENDS**

325   **Table 1.** Patient characteristics of the testing set.

326   **Table 2.** Model performance for LVH prediction in the entire testing set. Area under receiver-

327   operating characteristic curve (AUROC) and sensitivity at specificity fixed at 0.75 are provided.

328

329   **Table 3.** Comparison between presence of LVH on subsequent echocardiogram (>1 year and

330   closest to 5 years after index echocardiogram) in false positives versus true negatives of LVH

331   LGBM model in testing set

332   **FIGURE LEGENDS**

333   **Figure 1.** Data pipeline for model training and testing

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

335   **Figure 3.** ROC curves for subgroups of testing set in females, narrow QRS (*top left*), typical

336   right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom*

337   *left*), intraventricular conduction delay (IVCD, *bottom right*).

338   **Figure 4**. ROC curves for subgroups of testing set in males, narrow QRS (*top left*), typical right

339   bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*),

340   intraventricular conduction delay (IVCD, *bottom right*)

341   **Figure 5.** Scatterplots of echocardiographic left ventricular mass indexed (LVMi) plotted against

342   prediction probabilities from the LGBM model for females (*left panel*) and males (*right panel*).

343

REFERENCES

344
345
346 1. Vakili BA, Okin PM, Devereux RB. Prognostic implications of left ventricular
347   hypertrophy. Am Heart J 2001;141:334-41.
348 2. Sayin BY, Oto A. Left Ventricular Hypertrophy: Etiology-Based Therapeutic Options.
349   Cardiol Ther 2022;11:203-230.
350 3. Cuspidi C, Meani S, Valerio C, Fusi V, Sala C, Zanchetti A. Left ventricular hypertrophy
351   and cardiovascular risk stratification: impact and cost-effectiveness of echocardiography
352   in recently diagnosed essential hypertensives. Journal of Hypertension 2006;24.
353 4. Whelton PK, Carey RM, Aronow WS et al. 2017
354   ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the
355   Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A
356   Report of the American College of Cardiology/American Heart Association Task Force
357   on Clinical Practice Guidelines. Hypertension 2018;71:e13-e115.
358 5. Tison GH, Zhang J, Delling FN, Deo RC. Automated and Interpretable Patient ECG
359   Profiles for Disease Detection, Tracking, and Discovery. Circulation: Cardiovascular
360   Quality and Outcomes 2019;12:e005289.
361 6. Ricciardi D, Vetta G, Nenna A et al. Current diagnostic ECG criteria for left ventricular
362   hypertrophy: is it time to change paradigm in the analysis of data? Journal of
363   Cardiovascular Medicine 2020;21.
364 7. Leese PJ, Viera AJ, Hinderliter AL, Stearns SC. Cost-Effectiveness of
365   Electrocardiography vs. Electrocardiography Plus Limited Echocardiography to Diagnose
366   LVH in Young, Newly Identified, Hypertensives. American Journal of Hypertension
367   2010;23:592-598.
368 8. Hancock EW, Deal BJ, Mirvis DM et al. AHA/ACCF/HRS recommendations for the
369   standardization and interpretation of the electrocardiogram: part V: electrocardiogram
370   changes associated with cardiac chamber hypertrophy: a scientific statement from the
371   American Heart Association Electrocardiography and Arrhythmias Committee, Council
372   on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart
373   Rhythm Society: endorsed by the International Society for Computerized
374   Electrocardiology. Circulation 2009;119:e251-61.
375 9. Bacharova L, Chevalier P, Gorenek B et al. ISE/ISHNE Expert Consensus Statement on
376   ECG Diagnosis of Left Ventricular Hypertrophy: The Change of the Paradigm. The joint
377   paper of the International Society of Electrocardiology and the International Society for
378   Holter Monitoring and Noninvasive Electrocardiology. Journal of Electrocardiology
379   2023;81:85-93.
380 10. Ose B, Sattar Z, Gupta A, Toquica C, Harvey C, Noheria A. Artificial Intelligence
381   Interpretation of the Electrocardiogram: A State-of-the-Art Review. Curr Cardiol Rep
382   2024;26:561-580.
383 11. Ranka S, Reddy M, Noheria A. Artificial intelligence in cardiovascular medicine. Curr
384   Opin Cardiol 2021;36:26-35.
385 12. Siranart N, Deepan N, Techasatian W et al. Diagnostic accuracy of artificial intelligence
386   in detecting left ventricular hypertrophy by electrocardiograph: a systematic review and
387   meta-analysis. Scientific Reports 2024;14:15882.
388 13. Ying X. An Overview of Overfitting and its Solutions. Journal of Physics: Conference
389   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

435   the International Society for Computerized Electrocardiology. Circulation
436   2009;119:e235-40.
437   28.   Fragola PV, Autore C, Ruscitti G, Picelli A, Cannata D. Electrocardiographic diagnosis
438   of left ventricular hypertrophy in the presence of left bundle branch block: a wasted effort.
439   Int J Cardiol 1990;28:215-21.
440   29.   Haskell RJ, Ginzton LE, Laks MM. Electrocardiographic diagnosis of left ventricular
441   hypertrophy in the presence of left bundle branch block. J Electrocardiol 1987;20:227-32.
442   30.   Molloy TJ, Okin PM, Devereux RB, Kligfield P. Electrocardiographic detection of left
443   ventricular hypertrophy by the simple QRS voltage-duration product. J Am Coll Cardiol
444   1992;20:1180-6.
445   31.   Okin PM, Roman MJ, Devereux RB, Kligfield P. Time-Voltage Area of the QRS for the
446   Identification of Left Ventricular Hypertrophy. Hypertension 1996;27:251-258.
447   32.   Zhu H, Jiang Y, Cheng C et al. Four-Channel ECG as a Single Source for Early
448   Diagnosis of Cardiac Hypertrophy and Dilation — A Deep Learning Approach. NEJM
449   AI 2024;1:AIoa2300297.
450   33.   Liu C-M, Hsieh M-E, Hu Y-F et al. Artificial Intelligence–Enabled Model for Early
451   Detection of Left Ventricular Hypertrophy and Mortality Prediction in Young to Middle-
452   Aged Adults. Circulation: Cardiovascular Quality and Outcomes 2022;15:e008360.
453   34.   Kwon J-M, Jeon K-H, Kim HM et al. Comparing the performance of artificial
454   intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy
455   using electrocardiography. EP Europace 2020;22:412-419.
456   35.   Haimovich JS, Diamant N, Khurshid S et al. Artificial intelligence–enabled classification
457   of hypertrophic heart diseases using electrocardiograms. Cardiovascular Digital Health
458   Journal 2023;4:48-59.

459

460

461 **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%)	

462
463 RBBB, Right Bundle Branch Block; LBBB, Left Bundle Branch Block; IVCD, Intraventricular
464 Conduction Delay
465 [a]QRS duration <120 ms, [b]QRS duration ≥120 ms
466

467 **Table 2.** Model performance for LVH prediction in the entire testing set. Area under receiver-

468 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
▪ Amplitude$_{QRS-3D}$	0.601	0.389	0.602	0.395	0.603	0.388
▪ VTI$_{QRS-3D}$	**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

469 * At specificity 0.75
470 [a] Logistic regressions
471 [b] Input of 117 ECG statistics like QRS duration, heart rate etc. and 60 variational autoencoder latent
472 variables from ECG representative beat and sex
473 [c] Input of representative-beat ECG signal (X, Y, Z leads)
474

475 **Table 3.** Comparison between presence of LVH on subsequent echocardiogram (>1 year and

476 closest to 5 years after index echocardiogram) in false positives versus true negatives of LVH

477 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	-	-	-

478
479

480 **Figure 1.** Data pipeline for model training and testing
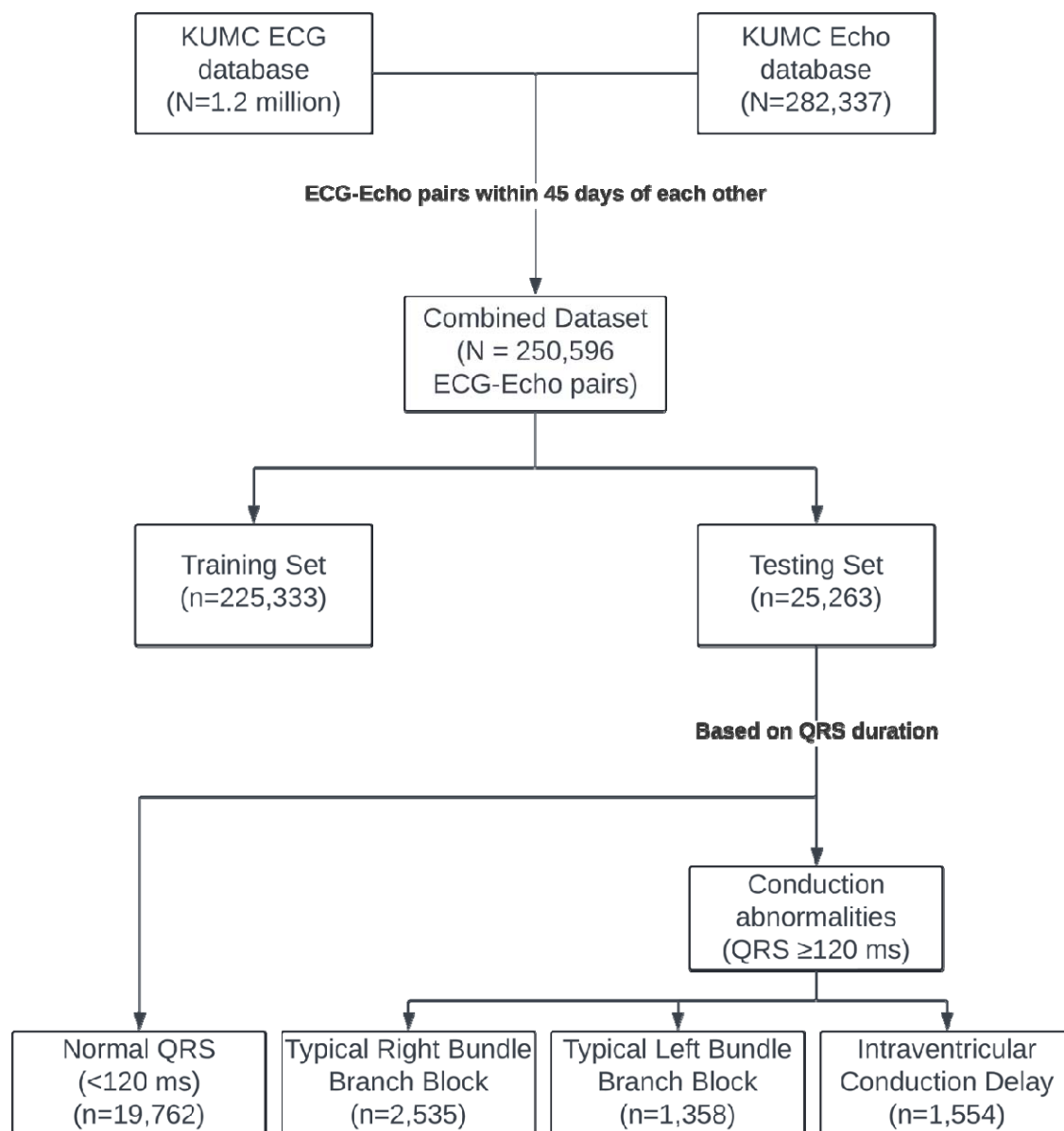
481
482

483   **Figure 2.** ROC curves from the <u>entire testing set</u> for males (*left panel*) and females (*right panel*).
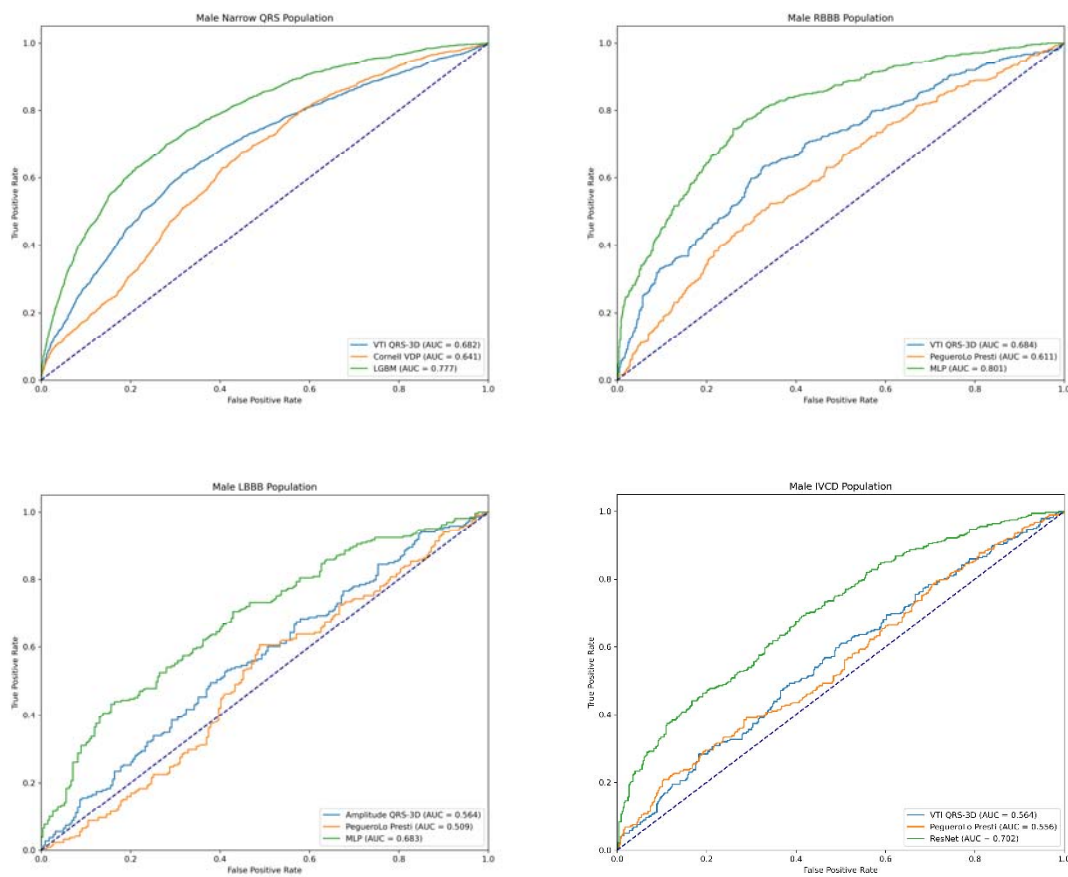
484

485
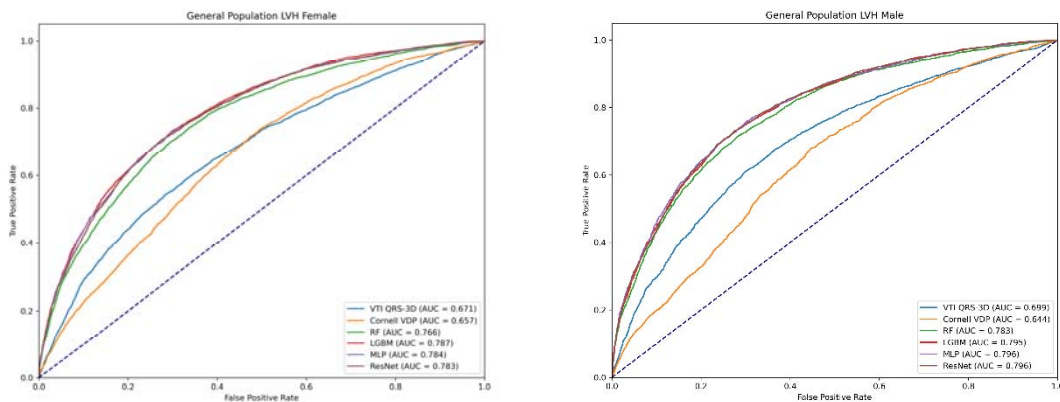
Figure 3. ROC curves for subgroups of females in testing set, narrow QRS (*top left*), typical right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*), intraventricular conduction delay (IVCD, *bottom right*).
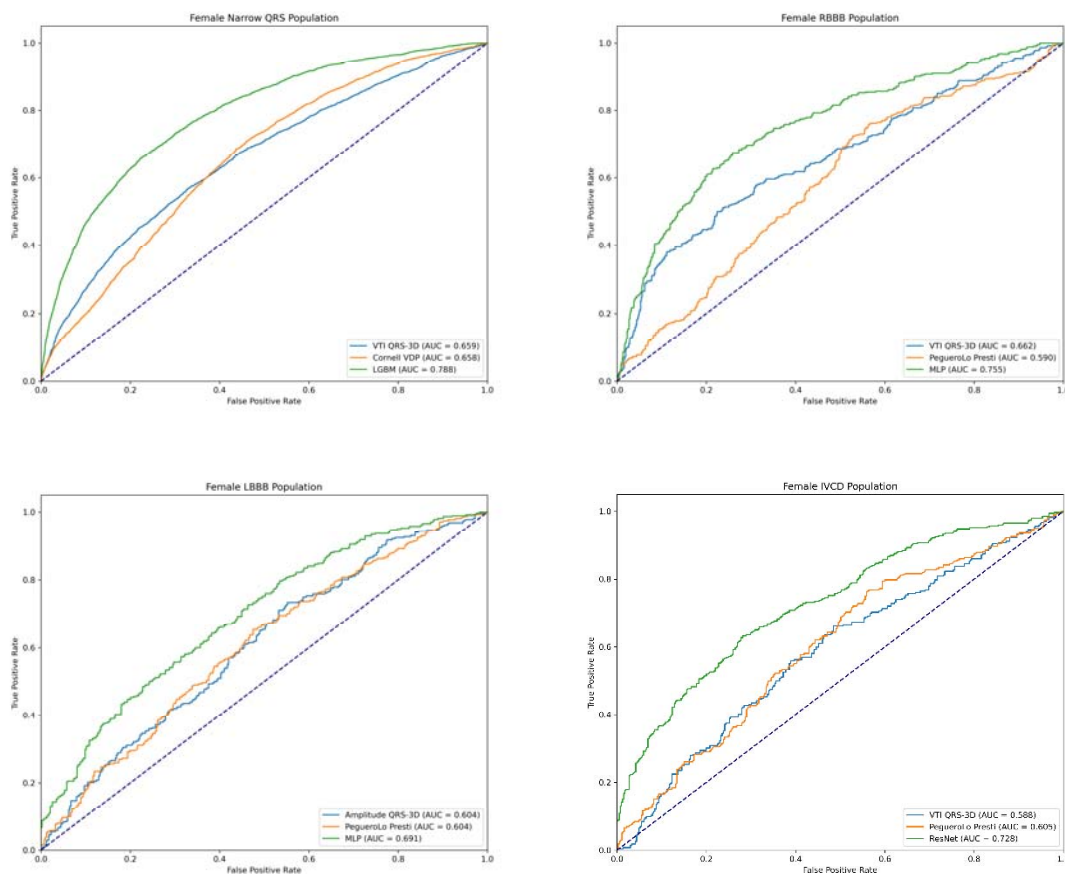
Figure 4. ROC curves for subgroups of males in testing set, narrow QRS (*top left*), typical right bundle branch block (RBBB, *top right*), typical left bundle branch block (LBBB, *bottom left*), intraventricular conduction delay (IVCD, *bottom right*)

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*).