ORIGINAL ARTICLE

Machine learning versus classical electrocardiographic criteria for echocardiographic left ventricular hypertrophy in a pre-participation cohort

Daniel YZ Lim[1], Gerald Sng[1], Wilbert HH Ho[1], Wang Hankun[1], Ching-Hui Sia[1–3], Joshua SW Lee[1], Xiayan Shen[1, 2], Benjamin YQ Tan[1, 4, 5], Edward CY Lee[1], Mayank Dalakoti[1, 2], Wang Kang Jie[1, 5], Clarence KW Kwan[1], Weien Chow[4], Ru San Tan[6], Carolyn SP Lam[6], Terrance SJ Chua[6], Tee Joo Yeo[1, 2], Daniel TT Chong[1, 6]

[1]Medical Classification Centre, Central Manpower Base, Singapore Armed Forces, Singapore
[2]Department of Cardiology, National University Heart Centre Singapore, Singapore
[3]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[4]HQ Medical Corps, Singapore Armed Forces, Singapore
[5]University Medicine Cluster, National University Health System, Singapore
[6]Department of Cardiology, National Heart Centre Singapore, Singapore

Correspondence to:
MBBS, MRCP, Daniel YZ Lim, Outram Rd, Singapore 169608,
phone: +65 97307915,
e-mail: daniel.lim@mohh.com.sg

Kardiol Pol. 2021; 79 (6): 654–661;
DOI: 10.33963/KP.15955

Received:
February 14, 2021

Revision accepted:
April 9, 2021

Published online:
April 16, 2021

ABSTRACT

Background: Classical electrocardiographic (ECG) criteria for left ventricular hypertrophy (LVH) are well studied in older populations and patients with hypertension. Their utility in young pre-participation cohorts is unclear.

Aims: We aimed to develop machine learning models for detection of echocardiogram-diagnosed LVH from ECG, and compare these models with classical criteria.

Methods: Between November 2009 and December 2014, pre-participation screening ECG and subsequent echocardiographic data was collected from 17 310 males aged 16 to 23, who reported for medical screening prior to military conscription. A final diagnosis of LVH was made during echocardiography, defined by a left ventricular mass index >115 g/m^2. The continuous and threshold forms of classical ECG criteria (Sokolow–Lyon, Romhilt–Estes, Modified Cornell, Cornell Product, and Cornell) were compared against machine learning models (Logistic Regression, GLMNet, Random Forests, Gradient Boosting Machines) using receiver-operating characteristics curve analysis. We also compared the important variables identified by machine learning models with the input variables of classical criteria.

Results: Prevalence of echocardiographic LVH in this population was 0.82% (143/17310). Classical ECG criteria had poor performance in predicting LVH. Machine learning methods achieved superior performance: Logistic Regression (area under the curve [AUC], 0.811; 95% confidence interval [CI], 0.738–0.884), GLMNet (AUC, 0.873; 95% CI, 0.817–0.929), Random Forest (AUC, 0.824; 95% CI, 0.749–0.898), Gradient Boosting Machines (AUC, 0.800; 95% CI, 0.738–0.862).

Conclusions: Machine learning methods are superior to classical ECG criteria in diagnosing echocardiographic LVH in the context of pre-participation screening.

Key words: biostatistics, electrocardiography, electronic medical records, myocardial disease

Kardiol Pol 2021; 79, 6: 654–661

INTRODUCTION

Left ventricular hypertrophy (LVH) is a clinically significant condition where there is an increased thickness of the left ventricular wall. It may be secondary to conditions such as athlete's heart, hypertension, valvular heart disease, and hypertrophic cardiomyopathy [1]. These conditions may need further evaluation, and may adversely impact fitness for participation in athletic endeavours and military training. Various electrocardiographic (ECG) criteria (Sokolow–Lyon [2], Cornell [3] etc.) have been proposed as markers of LVH, and are well studied in Western populations, particularly in hypertensive patients [4]. In our study, we examined ECG

WHAT'S NEW?

This large retrospective study examined the utility of machine learning algorithms in detecting echocardiographic left ventricular hypertrophy, when applied to screening electrocardiograms (ECG) of a young pre-participation cohort. Classical ECG criteria are not recommended in these individuals due to poor correlation with anatomic pathology, and no alternative algorithms currently exist. The machine learning algorithms applied showed good predictive power and performed better than classical criteria, whether clinical and anthropometric data were included as predictors or not. They also identified as important less recognized ECG parameters predictive of left ventricular hypertrophy, such as the mean QT interval, mean QRS duration, and R wave in lead I.

LVH criteria as predictors of LVH detected on transthoracic echocardiography, as defined by a left ventricular mass index (LVMI) >115 g/m^2.

Since 2008, it has been recognized that isolated electrocardiographic LVH may not apply to young athletic cohorts [5], and no specific recommendation exists for pre-participation cohorts. Referral for further cardiac investigations is no longer recommended by subsequent guidelines, such as the European Society of Cardiology (ESC) 2010 criteria [6], Seattle Criteria (2013) [7], Refined Criteria (2014) [8] and most recently the International Criteria for ECG interpretation in Athletes (2017) [9]. This was based on data showing no correlation between positive ECG LVH criteria and actual pathological LVH on cardiac imaging [10]. Nevertheless, detection of actual anatomic LVH remains a clinical outcome of interest. This is especially so in the context of fitness certification before participation in military training or sport, where patients with pathologic conditions (e.g. hypertrophic cardiomyopathy) should be excluded [11].

The main limitation of classical ECG criteria (e.g. Sokolow–Lyon, Cornell, Romhilt–Estes) is low sensitivity overall [4], and more so in younger populations [12]. This may stem from their statistical formulation, as most classical models assume each predictive factor is related in a linear fashion to LVH. For example, the Sokolow–Lyon criteria involve direct summation of S wave height in V_1, plus the larger of the R wave height in V_5 or V_6. This is an oversimplification of the overall information contained in an ECG. In addition, most current ECG criteria for LVH do not take into account demographic parameters, such as age, or anthropometric parameters, such as body mass index, body fat percentage, and presence of pectus excavatum.

Machine learning, or artificial intelligence, is an alternative approach that may improve the prediction of true LVH based on ECG parameters [13]. It can identify complex relationships between predictive parameters, and combine these in a non-linear fashion. This has the potential to improve prediction of clinical outcomes, and has been studied elsewhere for outcomes such as cardiovascular risk prediction [14]. There is limited literature on machine learning techniques to predict LVH from ECG parameters in adult cohorts [15, 16], but none in younger pre-participation cohorts.

Our objective in this study was to develop machine learning algorithms to predict LVH from resting ECG, as well as routine demographic, clinical, and anthropometric data in a cohort of young pre-participation individuals. All machine learning models were compared to classical ECG criteria for LVH. Machine learning models were first trained on ECG data alone, for a fair comparison with classical ECG criteria (which do not contain anthropometric data). They were also separately trained on the full set of predictors, including all ECG, demographic, clinical, and anthropometric data, to determine if this would improve the accuracy of predictions.

METHODS

Study population

The Singapore Armed Forces have conducted universal pre-participation screening before military enlistment for all young male Singaporeans [17]. Universal ECG screening was implemented since 2008, based on an Italian pre-participation cardiovascular screening system proposed by Corrado et al. [18], with referral for echocardiography if Sokolow–Lyon criteria for LVH were met. Echocardiography was performed within a year of referral to the national cardiology tertiary center, with all studies reported by a cardiology specialist. Our patient population comprised of 17 310 prospective male military recruits who had undergone transthoracic echocardiography from November 2009 to December 2014, as part of determination of their fitness to enlist into military service.

Approval for collection and use of data was granted by the Singapore Armed Forces Joint Medical Committee, and ethical approval was obtained from the local institutional review board.

Variables

For all individuals, demographic, anthropometric, and clinical parameters were collected, as well as a baseline resting 12 lead ECG. Resting ECGs were performed by trained personnel using the Philips Pagewriter TC70 ECG machine, which recorded at a sampling rate of 500 Hz. The electronic ECG readouts were analyzed using a proprietary Philips TM TraceMasterVue modular ECG analysis system

with automatic measurement of the ECG parameters. The full list of parameters collected are included in Supplementary material, *Table S1*.

The primary outcome was LVH, as assessed by LVMI on transthoracic echocardiography. Echocardiographic assessment was used as the determinant in view of operational considerations precluding general use of advanced imaging (such as cardiac magnetic resonance imaging [CMRI]) or histological diagnosis. A cutoff of LVMI >115 g/m^2 was used in this male population, based on the American Society of Echocardiography guidelines [19].

Prediction algorithms

For the machine learning algorithms, we employed a 70:30 train-test split (i.e. using 70% of the data to train the machine learning model, and the remaining 30% to assess accuracy). Continuous variables were scaled and normalized. Fivefold cross validation was used to tune the model parameters. We implemented some commonly used types of supervised machine-learning algorithms. We included regression based methods, namely conventional Logistic Regression and GLMNet (penalized logistic regression with the ElasticNet penalty) [20], as well as tree based methods, which were Random Forests [21] and Gradient Boosting Machines [22]. All machine learning models were trained first on ECG parameters only, and separately on the full set of predictive parameters (including demographic, clinical, and anthropometric data).

The machine learning models were compared with classical ECG criteria. We examined the commonly used Sokolow–Lyon, Romhilt–Estes, Modified Cornell, Cornell Product, and Cornell methods for assessing LVH on ECG, and calculated both their continuous and threshold forms. We summarize the different classical criteria being compared in Supplementary material, *Table S2*.

Statistical analysis

Summary statistics for anthropometric and ECG parameters were calculated, including counts and proportions for categorical data, as well as medians and interquartile ranges for continuous data.

The performance of the machine learning algorithms was assessed using the test cohort. We calculated receiver operating characteristic (ROC) curves for each algorithm, as well as the area under the curve (AUC). The 95% confidence interval (CI) for each AUC was determined using a bootstrap method. Likewise, we calculated the ROC and AUCs for the continuous versions of the various classical ECG methods. For machine learning models, variable importance was assessed using the final tuned coefficients for regression based models, and using the weighted average improvement in node impurity for tree based models. Analysis was performed using RStudio with R version 3.6.1, and using the packages caret, glmnet, randomForest, and pROC.

RESULTS

Study population characteristics

Our study included a total of 17 310 young men aged 16 to 23 years, of which 143 (0.82%) had LVH. The characteristics of the population stratified by LVH status are included in Table 1. Categorical variables are tabulated as frequencies with their respective percentages, and continuous variables are tabulated as means with standard deviation.

Performance of the various models

The predictive accuracy of the classical and machine learning models was assessed by the AUC. The values for the various models are tabulated in Table 2, with the bootstrapped 95% CIs. Other evaluation parameters are tabulated in Supplementary material, *Table S3*.

All machine learning models showed superior predictive accuracy compared to classical models, regardless of whether electrocardiographic (ECG) parameters alone were considered, or if all available predictive parameters were included. In particular, GLMNet, Random Forests, and Logistic Regression had excellent predictive accuracy with AUC surpassing 0.8. This can be seen in the numerical AUC values, as well as the ROC plots in Figure 1. The machine learning models which included ECG parameters only performed similarly to those with all predictive parameters included, as evidenced by the numerically similar AUC values as well as the overlapping 95% CIs.

Important variables

To interpret the machine learning models and compare them to classical criteria, the top ten most important variables for each machine learning model were ranked and listed in Table 3. Weight, height, body fat percentage, and systolic blood pressure were anthropometric parameters not used in the classical ECG criteria, but were deemed important to the machine learning algorithms. Mean QT interval, mean QRS duration and R wave in lead I were ECG parameters not used in the classical criteria, but were deemed important to the machine learning algorithms, both when ECG parameters alone were included, and when all predictive parameters were included.

DISCUSSION

General discussion

To the best of our knowledge, this is one of the few studies to employ machine learning to detect LVH. In our large, pre-participation cohort, we found that all 4 machine learning algorithms tested were superior to classical ECG criteria in identifying echocardiographic LVH. The high AUC values [23] of above 0.8 derived from the machine learning models are excellent and at a level which would generally be acceptable for clinical use. Other classical algorithms

Table 1. Population characteristics stratified by left ventricular hypertrophy (LVH) status

Variables	Overall (n = 17 310)	No LVH (n = 17 167)	Presence of LVH (n = 143)
Demographic and anthropometric parameters			
Age, years	18.0 (17.0–19.0)	18.0 (17.0–19.0)	18.0 (17.0–19.0)
Smoking	2921 (17)	2898 (17)	23 (16)
Urine dipstick blood (present)	33 (0)	33 (0)	0 (0)
Urine dipstick glucose (present)	28 (0)	28 (0)	0 (0)
Urine dipstick protein (present)	19 (0)	19 (0)	0 (0)
Pectus excavatum (present)	139 (1)	139 (1)	0 (0)
Hemoglobin, g/dl	15.6 (14.9–16.2)	15.6 (15.0–16.2)	15.3 (14.7–15.8)
Height, cm	172 (168–177)	172 (168–177)	172 (168–176)
Weight, kg	61 (54.5–69.5)	61 (54.5–69.4)	68 (60.4–75.35)
Body fat percentage, %	18.2 (14.5–23)	18.2 (14.5–23)	20.1 (17.3–24.3)
Systolic blood pressure, mm Hg	115 (105–125)	115 (105–125)	117 (108–127)
Diastolic blood pressure, mm Hg	66 (58–73)	66 (58–73)	64 (57–71)
Electrocardiogram parameters			
Mean PR interval, ms	144 (132–157)	144 (132–157)	148 (136.5–160.5)
Mean PR segment, ms	41 (31–51)	41 (31–51)	44 (33.5–54)
Mean QRS duration, ms	95 (89–102)	95 (89–102)	98 (92–105.5)
Mean QTc interval, ms	410 (395–425)	410 (395–425)	411 (394–427)
Mean QT interval, ms	364 (345–385)	364 (345–385)	388 (360.5–415.5)
Mean ventricular rate, bpm	43 (28–76)	43 (28–76)	54 (32–99.5)
QT interval dispersion, ms	76 (66–87)	76 (66–87)	68 (58–78)
R wave height in aVF, mV	1.4 (0.8–1.9)	1.4 (0.8–1.9)	1.6 (0.9–2.1)
R wave height in aVL, mV	0.2 (0.1–0.2)	0.2 (0.1–0.2)	0.2 (0.1–0.4)
R wave height in aVR, mV	0.1 (0.1–0.3)	0.1 (0.1–0.3)	0.1 (0–0.2)
R wave height in I, mV	0.5 (0.3–0.7)	0.5 (0.3–0.7)	0.7 (0.4–0.9)
R wave height in II, mV	1.5 (1–2.1)	1.5 (1–2.1)	1.9 (1.2–2.3)
R wave height in III, mV	1.2 (0.7–1.8)	1.2 (0.7–1.8)	1.3 (0.6–1.9)
R wave height in V1, mV	0.3 (0.2–0.5)	0.3 (0.2–0.5)	0.4 (0.2–0.5)
R wave height in V_2, mV	0.7 (0.5–1)	0.7 (0.5–1)	0.8 (0.6–1.1)
R wave height in V_3, mV	1 (0.7–1.4)	1 (0.7–1.4)	1.2 (0.9–1.6)
R wave height in V_4, mV	1.8 (1.3–2.4)	1.8 (1.3–2.4)	2.3 (1.6–2.9)
R wave height in V_5, mV	1.8 (1.3–2.3)	1.8 (1.3–2.2)	2.2 (1.6–2.6)
R wave height in V_6, mV	1.4 (1–1.8)	1.4 (1–1.8)	1.7 (1.3–2)
S wave depth in aVF, mV	0.1 (0–0.3)	0.1 (0–0.3)	0.1 (0–0.3)
S wave depth in aVL, mV	0.5 (0.2–0.8)	0.5 (0.2–0.8)	0.5 (0.2–0.9)
S wave depth in aVR, mV	0 (0–1.1)	0 (0–1.1)	0 (0–1.4)
S wave depth in I, mV	0.2 (0.1–0.4)	0.2 (0.1–0.4)	0.2 (0–0.3)
S wave depth in II, mV	0.2 (0–0.3)	0.2 (0–0.3)	0.1 (0–0.3)
S wave depth in III, mV	0.1 (0–0.2)	0.1 (0–0.2)	0.1 (0–0.3)
S wave depth in V_1, mV	1.1 (0.6–1.5)	1.1 (0.6–1.5)	1.3 (0.8–1.8)
S wave depth in V_2, mV	1.7 (1.1–2.3)	1.7 (1.1–2.3)	1.8 (1.4–2.7)
S wave depth in V_3, mV	1.5 (1.0–2.0)	1.5 (1.0–2.0)	1.6 (1.1–2.3)
S wave depth in V_4, mV	0.6 (0.3–1.0)	0.6 (0.3–1.0)	0.7 (0.4–1.1)
S wave depth in V_5, mV	0.3 (0.1–0.5)	0.3 (0.1–0.5)	0.3 (0.1–0.5)
S wave depth in V_6, mV	0.1 (0–0.3)	0.1 (0–0.3)	0.1 (0–0.3)
Ventricular activation time in aVL, ms	31 (20–62)	31 (20–62)	34 (22–60)
Ventricular activation time in V_5, ms	44 (37–49)	44 (37–49)	47 (39–51)
Ventricular activation time in V_6, ms	44 (38–48)	44 (38–48)	47 (42–52)

Values are expressed as counts (percentages) or median (interquartile range) as appropriate

based on ECG data were less discriminatory and had a lower predictive accuracy for LVH.

The machine learning models identified non-traditional ECG parameters that had predictive value for LVH. The approaches of the machine learning models were mathematically different, with Logistic Regression and GLMNet being regression type models, whereas Random Forest and Gradient Boosting Machines were decision tree-based models. We noted that some parameters that are less recognized and/or not included in all classical models, were instead consistently identified as important by the different machine learning modeling approaches. They are thus highly likely to represent true sources of predictive information about the underlying pathology, echocardiographic LVH. These included limb lead parameters (R wave in lead I) and parameters involving the duration of ventricular activity (such as mean QT interval and mean QRS duration). This suggests that electrical manifestations of LVH go beyond anteriorly directed depolarization forces in the precordial leads, but also involve abnormalities in depolarization and

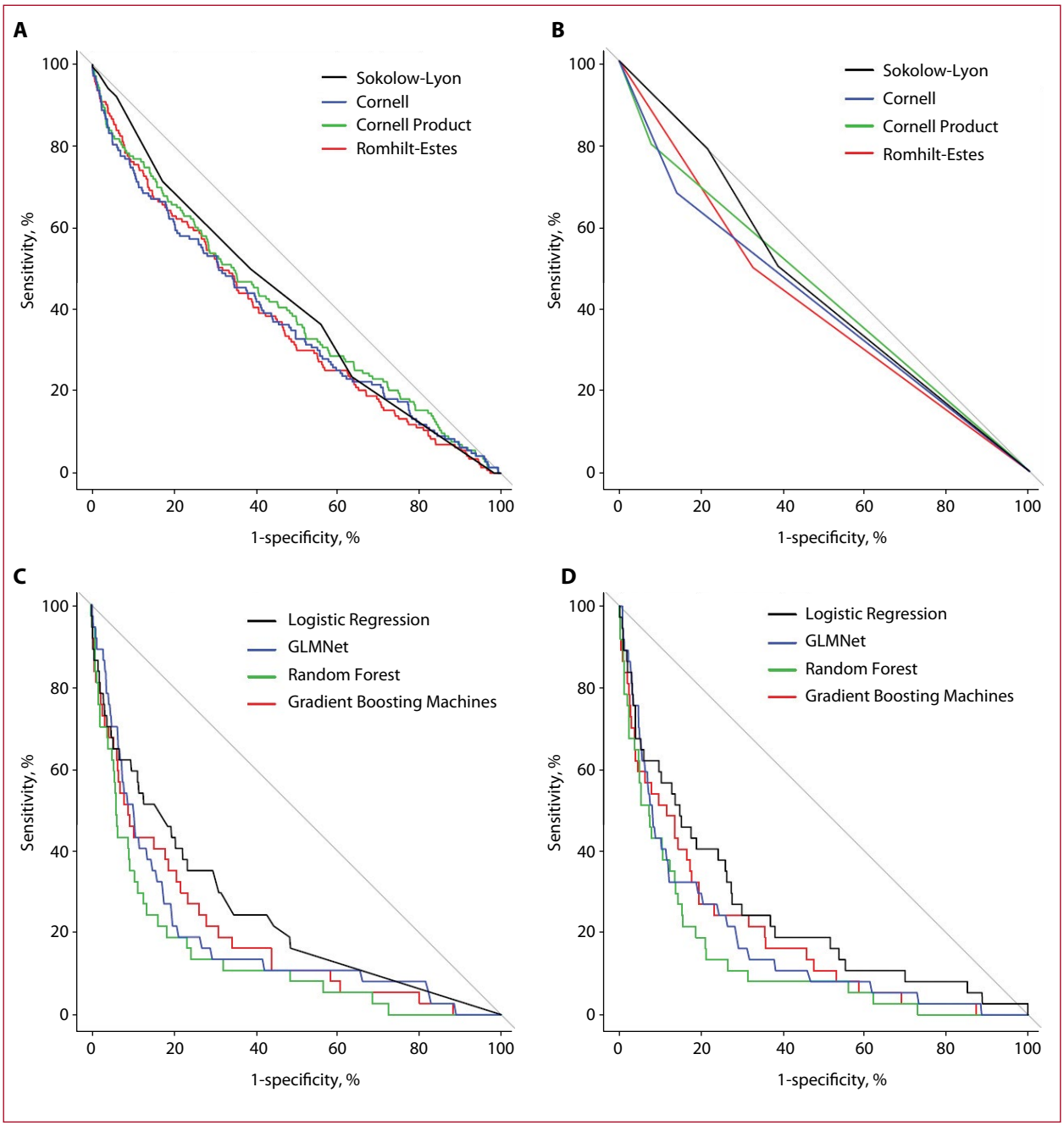

Figure 1. Composite plots of receiver operating characteristic curves for classical and machine learning models. **A.** Receiver operating characteristic plots for classical criteria (continuous). **B.** Receiver operating characteristic plots for classical criteria (thresholds). **C.** Receiver operating characteristic plots for machine learning models (with electrocardiogram parameters only). **D.** Receiver operating characteristic plots for machine learning models (with all predictive parameters)

repolarization. The left laterally directed depolarization forces may also be better represented by lead I, rather than the traditional precordial leads (such as V_5–V_6).

When trained on all available predictive parameters, the machine learning models were also able to integrate non-ECG data into their predictions, as evidenced by weight and body fat percentage being important variables identified. This suggests that the algorithms were able to learn adjustments for body habitus, an important factor that is not taken into account by classical models, but which prior studies have shown to improve diagnostic accuracy once included [24–26]. It is known that a person's habitus can affect the sensitivity and specificity of classical ECG criteria, and this is one factor that limits the accuracy of classical ECG criteria [27]. However, it is notable that the machine learning models which integrated non-ECG data had similar performance to those including ECG parameters only. This suggests that the machine learning algorithms are able to infer some of information about body habitus from the other ECG parameters, and that our models can still be applied even in cases where anthropometric information is not available.

Table 2. Area under receiver operating characteristic curve (AUC) of classical criteria and machine learning models

Predictive model		AUC (95% CI)
Classical criteria (continuous)	Sokolow–Lyon	0.629 (0.581–0.676)
	Cornell	0.599 (0.549–0.650)
	Cornell product	0.625 (0.575–0.675)
	Romhilt–Estes	0.582 (0.536–0.628)
Classical criteria (thresholds)	Sokolow–Lyon	0.589 (0.548–0.630)
	Cornell	0.562 (0.529–0.596)
	Cornell product	0.591 (0.552–0.629)
	Romhilt–Estes	0.544 (0.503–0.585)
Machine learning models (with electrocardiogram parameters only)	Logistic Regression	0.811 (0.738–0.884)
	GLMNet	0.873 (0.817–0.929)
	Random Forests	0.824 (0.749–0.898)
	Gradient Boosting Machines	0.800 (0.738–0.862)
Machine learning models (with all predictive parameters)	Logistic Regression	0.815 (0.745–0.885)
	GLMNet	0.864 (0.804–0.924)
	Random Forests	0.826 (0.756–0.897)
	Gradient Boosting Machines	0.793 (0.723–0.863)

Table 3. Table of important predictive variables for each machine learning model

Trained with electrocardiogram parameters only				Trained with all predictive parameters			
Logistic Regression	GLMNet	Random Forest	Gradient Boosting Machines	Logistic Regression	GLMNet	Random Forest	Gradient Boosting Machines
Mean QRS duration	R wave in I	Mean QT interval	Mean QT interval	Weight	Mean QT interval	Mean QT interval	R wave in I
S wave in V_4	Mean QT interval	R wave in I	R wave in V_4	Mean QRS duration	R wave in I	Weight	R wave in V_6
R wave in I	S wave in V_4	R wave in V_5	R wave in V1	Height	Weight	R wave in I	R wave in V_1
R wave in aVF	Mean QRS duration	R wave in aVL	S wave in aVR	S wave in V_4	R wave in V_4	Body fat percentage	Mean QT interval
Mean QT interval	R wave in V_4	R wave in V_4	R wave in V_6	Systolic blood pressure	S wave in V_4	R wave in V_4	R wave in V_4
R wave in III	S wave in V_3	Mean ventricular rate	R wave in I	R wave in aVF	Mean QRS duration	Mean ventricular rate	R wave in III
R wave in V_4	S wave in V_2	R wave in V_3	Mean QRS duration	R wave in III	S wave in V_3	R wave in V_3	S wave in aVR
S wave in I	R wave in II	Mean QRS duration	R wave in aVL	R wave in V_4	Systolic blood pressure	R wave in V_5	Mean ventricular rate
S wave in V_2	R wave in V_3	S wave in V_3	Mean ventricular rate	S wave in I	S wave in I	Mean QRS duration	S wave in V_3
S wave in V_5	S wave in aVR	R wave in II	R wave in V_3	S wave in V_6	S wave in V_2	R wave in aVL	R wave in V_3

A study by Kwon et al. [15] in an adult hospital cohort used a neural network based approach to improve on classical ECG criteria, with the best model (ensemble neural network) having an AUC of 0.868. Neural networks are a machine learning method which can accept unstructured data, such as an unprocessed ECG signal. They are able to use intensive computing methods to utilize information available in the signal that is not captured by conventional ECG parameters, but conversely, may be difficult to interpret (i.e. more "black-box") because the contributing information is not pre-defined as a parameter. In contrast, our study was done on a screening pre-participation cohort, and used machine learning methods which accepted only predefined parameters measured from the ECG, thus finding the most optimal ways that they could be mathematically combined. This approach also allowed identification of the most important conventional ECG parameters, for easier interpretability. Although we had a different approach, the best model in our study (GLMNet) achieved a comparable AUC of 0.873. We believe that both strategies are equally valid, and that combined strategies may be able to achieve further gains in predictive performance.

Strengths and limitations

The strength of our study is its large sample size, utilizing the data of 17 310 subjects. Furthermore, our dataset contained anthropometric variables, including body fat percentage, which are seldom collected systematically. However, we acknowledge that our study was performed in a pre-participation screening cohort, with a low incidence of actual LVH. The low number of true cases will limit gains to predictive accuracy, as subtle differences between true cases and controls will be harder to detect [28]. Further data collection with more pathological cases will help to increase the robustness of machine learning algorithms in this context.

Our study was performed exclusively in males, and there may be sex-specific differences in ECG-derived parameters or anthropometric findings. Further studies can be performed in populations of both sexes, but we believe that the techniques will be similarly applicable. We also recognize that CMRI derived LVH is the gold standard of diagnosis [29], whereas we used echocardiography to diagnose LVH. It is possible that classical criteria may perform better with CMRI as a diagnostic endpoint, as recent literature has showed that novel combinations of classical criteria can have good discriminatory power for CMRI derived LVH [30]. Nevertheless, this provides insight into the possibility of performing a similar analysis with CMRI data in the future.

One other potential limitation is that machine learning algorithms are by nature less interpretable than traditional predictive models [31]. Although it is possible to determine which variables are of importance to the algorithm, the underlying mathematical relationship to other variables is difficult to elucidate. This makes it challenging to check for problems such as overfitting [32]. Steps may be taken to guard against this, such as cross validation and separation of training and test sets [33, 34]. However, classical methods involving regression are just as susceptible to similar statistical problems, and are more likely to oversimplify the relationship between the variables. In general, all statistical models cannot determine the underlying pathophysiological reasons that make a variable predictive for pathology. For the novel predictive variables identified by our machine learning models, further studies will be needed to elucidate the underlying mechanism.

Future work

The gains in predictive accuracy found in our cohort using machine learning models should be validated in other large, external cohorts. It is possible that the novel parameters identified by our cohort will have applicability in older adults as well, and this may be explored by clinicians managing other cohorts where LVH prediction is clinically important, such as older adults, hypertensive patients and patients with cardiomyopathy.

Current consensus criteria recommend against the use of classical ECG criteria in determination of LVH from pre-participation ECGs [35], because their poor performance can lead to unnecessary investigation or exclusion from physical activities. However, no alternative criteria exists, and this may conversely lead to underdiagnosis of true LVH. More accurate machine learning models developed in this study suggest a renewed role for ECG detection of LVH, which can contribute to safer recommendations for pre-participation screening

CONCLUSIONS

Classical ECG criteria perform poorly in our young, pre-participation cohort with a low prevalence of LVH. Machine learning methods show superior predictive performance and accord high importance to less recognized predictors of LVH from ECG. Addition of anthropometric data did not improve performance of machine learning models. Further research is required to improve the predictive ability of machine learning models, to implement these models in clinical care, and to understand the underlying pathology of the novel ECG predictors identified.

Supplementary material

Supplementary material is available at https://journals.viamedica.pl/kardiologia_polska.

Article information

Conflict of interest: None declared.

How to cite: Daniel YZ Lim, Gerald Sng, Wilbert HH Ho, et al. Machine learning versus classical electrocardiographic criteria for echocardiographic left ventricular hypertrophy in a pre-participation cohort. Kardiol Pol. 2021; 79(6): 654–661, doi: 10.33963/KP.15955.

REFERENCES

1. Rubiś PP. Left ventricular hypertrophy: what lies beneath? Pol Arch Intern Med. 2019; 129(12): 945–948, doi: 10.20452/pamw.15118, indexed in Pubmed: 31868864.
2. Sokolow M, Lyon TP. Electrocardiographic patterns of ventricular hypertrophy as obtained by unipolar precordial and limb leads. Am J Med. 1947; 2(6): 656, doi: 10.1016/0002-9343(47)90055-7.
3. Casale PN, Devereux RB, Alonso DR, et al. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. Circulation. 1987; 75(3): 565–572, doi: 10.1161/01.cir.75.3.565, indexed in Pubmed: 2949887.
4. Pewsner D, Jüni P, Egger M, et al. Accuracy of electrocardiography in diagnosis of left ventricular hypertrophy in arterial hypertension: systematic review. BMJ. 2007; 335(7622): 711, doi: 10.1136/bmj.39276.636354.AE, indexed in Pubmed: 17726091.
5. Sohaib SM, Payne JR, Shukla R, et al. Electrocardiographic (ECG) criteria for determining left ventricular mass in young healthy men; data from the LARGE Heart study. J Cardiovasc Magn Reson. 2009; 11(1): 2, doi: 10.1186/1532-429X-11-2, indexed in Pubmed: 19149884.
6. Corrado D, Pelliccia A, Heidbuchel H, et al. Section of Sports Cardiology, European Association of Cardiovascular Prevention and Rehabilitation. Recommendations for interpretation of 12-lead electrocardiogram in the athlete. Eur Heart J. 2010; 31(2): 243–259, doi: 10.1093/eurheartj/ehp473, indexed in Pubmed: 19933514.
7. Drezner JA, Ackerman MJ, Anderson J, et al. Electrocardiographic interpretation in athletes: the ‚Seattle criteria'. Br J Sports Med. 2013; 47(3): 122–124, doi: 10.1136/bjsports-2012-092067, indexed in Pubmed: 23303758.
8. Sheikh N, Papadakis M, Ghani S, et al. Comparison of electrocardiographic criteria for the detection of cardiac abnormalities in elite black and white athletes. Circulation. 2014; 129(16): 1637–1649, doi: 10.1161/CIRCULATIONAHA.113.006179, indexed in Pubmed: 24619464.
9. Drezner JA, Sharma S, Baggish A, et al. International criteria for electrocardiographic interpretation in athletes: consensus statement. Br J Sports Med. 2017; 51(9): 704–731, doi: 10.1136/bjsports-2016-097331, indexed in Pubmed: 28258178.
10. Pelliccia A, Maron BJ, Culasso F, et al. Clinical significance of abnormal electrocardiographic patterns in trained athletes. Circulation. 2000; 102(3): 278–284, doi: 10.1161/01.cir.102.3.278, indexed in Pubmed: 10899089.

11. Maron BJ, Chaitman BR, Ackerman MJ, et al. Working Groups of the American Heart Association Committee on Exercise, Cardiac Rehabilitation, and Prevention, Councils on Clinical Cardiology and Cardiovascular Disease in the Young. Recommendations for physical activity and recreational sports participation for young patients with genetic cardiovascular diseases. Circulation. 2004; 109(22): 2807–2816, doi: 10.1161/01.CIR.0000128363.85581.E1, indexed in Pubmed: 15184297.
12. Sklyar E, Ginelli P, Barton A, et al. Validity of electrocardiographic criteria for increased left ventricular mass in young patients in the general population. World J Cardiol. 2017; 9(3): 248–254, doi: 10.4330/wjc.v9.i3.248, indexed in Pubmed: 28400921.
13. Sparapani R, Dabbouseh NM, Gutterman D, et al. Detection of left ventricular hypertrophy using bayesian additive regression trees: the MESA. J Am Heart Assoc. 2019; 8(5): e009959, doi: 10.1161/JAHA.118.009959, indexed in Pubmed: 30827132.
14. Quesada JA, Lopez-Pineda A, Gil-Guillén VF, et al. Machine learning to predict cardiovascular risk. Int J Clin Pract. 2019; 73(10): e13389, doi: 10.1111/ijcp.13389, indexed in Pubmed: 31264310.
15. 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. 2020; 22(3): 412–419, doi: 10.1093/europace/euz324, indexed in Pubmed: 31800031.
16. De la Garza-Salazar F, Romero-Ibarguengoitia ME, Rodriguez-Diaz EA, et al. Improvement of electrocardiographic diagnostic accuracy of left ventricular hypertrophy using a Machine Learning approach. PLoS One. 2020; 15(5): e0232657, doi: 10.1371/journal.pone.0232657, indexed in Pubmed: 32401764.
17. Ng CT, Ong HY, Cheok C, et al. Prevalence of electrocardiographic abnormalities in an unselected young male multi-ethnic South-East Asian population undergoing pre-participation cardiovascular screening: results of the Singapore Armed Forces Electrocardiogram and Echocardiogram screening protocol. Europace. 2012; 14(7): 1018–1024, doi: 10.1093/europace/eur424, indexed in Pubmed: 22308089.
18. Corrado D, Basso C, Schiavon M, et al. Pre-participation screening of young competitive athletes for prevention of sudden cardiac death. J Am Coll Cardiol. 2008; 52(24): 1981–1989, doi: 10.1016/j.jacc.2008.06.053, indexed in Pubmed: 19055989.
19. 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): 1–39.e14, doi: 10.1016/j.echo.2014.10.003, indexed in Pubmed: 25559473.
20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33(1): 1–22, indexed in Pubmed: 20808728.
21. Breiman L. Random forests. Mach Learn. 2001; 45: 5–32.
22. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Statist. 2001; 29(5): 1189–1232, doi: 10.1214/aos/1013203451.
23. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013; 4(2): 627–635, indexed in Pubmed: 24009950.
24. Norman JE, Levy D. Adjustment of ECG left ventricular hypertrophy criteria for body mass index and age improves classification accuracy: the effects of hypertension and obesity. J Electrocardiol. 1996; 29(Suppl 1): 241–247, doi: 10.1016/s0022-0736(96)80070-7.
25. Rider OJ, Ntusi N, Bull SC, et al. Improvements in ECG accuracy for diagnosis of left ventricular hypertrophy in obesity. Heart. 2016; 102(19): 1566–1572, doi: 10.1136/heartjnl-2015-309201, indexed in Pubmed: 27486142.
26. Okin PM, Roman MJ, Devereux RB, et al. ECG identification of left ventricular hypertrophy. Relationship of test performance to body habitus. J Electrocardiol. 1996; 29(Suppl 1): 256–261, doi: 10.1016/s0022-0736(96)80072-0, indexed in Pubmed: 9238409.
27. Jingi AM, Noubiap JJ, Kamdem P, et al. Determinants and improvement of electrocardiographic diagnosis of left ventricular hypertrophy in a black African population. PLoS One. 2014; 9(5): e96783, doi: 10.1371/journal.pone.0096783, indexed in Pubmed: 24810594.
28. Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell. 1991; 13(3): 252–264, doi: 10.1109/34.75512.
29. Grothues F, Smith GC, Moon JCC, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am J Cardiol. 2002; 90(1): 29–34, doi: 10.1016/s0002-9149(02)02381-0, indexed in Pubmed: 12088775.
30. Matusik PS, Bryll A, Matusik PT, et al. Electrocardiography and cardiac magnetic resonance imaging in the detection of left ventricular hypertrophy: the impact of indexing methods. Kardiol Pol. 2020; 78(9): 889–898, doi: 10.33963/KP.15464, indexed in Pubmed: 32598106.
31. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. BMC Med Inform Decis Mak. 2019; 19(1): 146, doi: 10.1186/s12911-019-0874-0, indexed in Pubmed: 31357998.
32. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010; 11: 2079–2107.
33. Kohavi R. Study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI'95. 1995; 2: 1137–1143.
34. Xu Y, Goodacre R. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. J Anal Test. 2018; 2(3): 249–262, doi: 10.1007/s41664-018-0068-2, indexed in Pubmed: 30842888.
35. Maron BJ, Thompson PD, Ackerman MJ, et al. American Heart Association Council on Nutrition, Physical Activity, and Metabolism. Recommendations and considerations related to preparticipation screening for cardiovascular abnormalities in competitive athletes: 2007 update: a scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism: endorsed by the American College of Cardiology Foundation. Circulation. 2007; 115(12): 1643–1655, doi: 10.1161/CIRCULATIONAHA.107.181423, indexed in Pubmed: 17353433.