ORIGINAL ARTICLE

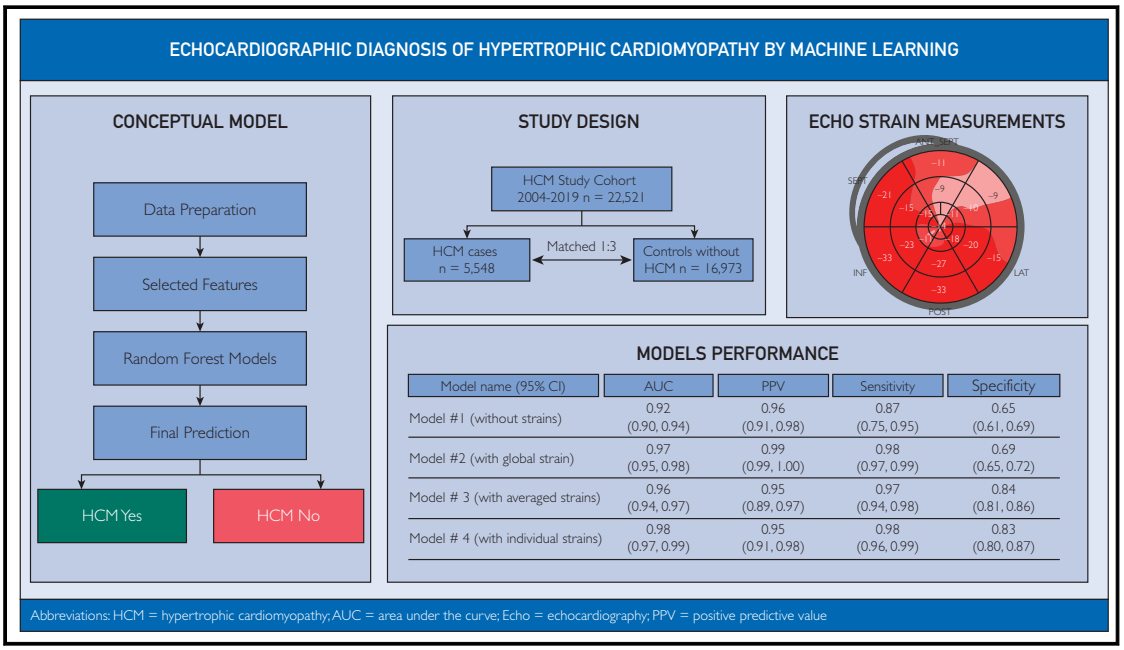
MAYO CLINIC PROCEEDINGS: DIGITAL HEALTH

MAYO CLINIC

Echocardiographic Diagnosis of Hypertrophic Cardiomyopathy by Machine Learning

Nasibeh Zanjirani Farahani, PhD; Mateo Alzate Aguirre, MD; Vanessa Karlinski Vizentin, MD; Moein Enayati, PhD; J. Martijn Bos, MD, PhD; Andredi Pumarejo Medina, MD; Kathryn F. Larson, MD; Kalyan S. Pasupathy, PhD; Christopher G. Scott, MS; April L. Zacher, MS; Eduard Schlechtinger, MS; Bruce K. Daniels, RDCS; Vinod C. Kaggal, MS; Jeffrey B. Geske, MD; Patricia A. Pellikka, MD; Jae K. Oh, MD; Steve R. Ommen, MD; Garvan C. Kane, MD; Michael J. Ackerman, MD, PhD; and Adelaide M. Arruda-Olson, MD, PhD

GRAPHICAL ABSTRACT



Abstract

Objective: To develop machine learning tools for automated hypertrophic cardiomyopathy (HCM) case recognition from echocardiographic metrics, aiming to identify HCM from standard echocardiographic data with high performance.

Patients and Methods: Four different random forest machine learning models were developed using a case-control cohort composed of 5548 patients with HCM and 16,973 controls without HCM, from January 1, 2004, to March 15, 2019. Each patient with HCM was matched to 3 controls by sex, age, and year of echocardiography. Ten-fold crossvalidation was used to train the models to identify HCM. Variables included in the models were demographic characteristics (age, sex, and body surface area) and 16 standard echocardiographic metrics.

Mayo Clin Proc Digital Health ■ December 2024;2(4):564-573 ■ https://doi.org/10.1016/j.mcpdig.2024.08.009
www.mcpdigitalhealth.org ■

Results: The models were differentiated by global, average, individual, or no strain measurements. Area under the receiver operating characteristic curves (area under the curve) ranged from 0.92 to 0.98 for the 4 separate models. Area under the curves of model 2 (using left ventricular global longitudinal strain; 0.97; 95% CI, 0.95-0.98), 3 (using averaged strain; 0.96; 95% CI, 0.94-0.97), and 4 (using 17 individual strains per patient; 0.98; 95% CI, 0.97-0.99) had comparable performance. By comparison, model 1 (no strain data; 0.92; 95% CI, 0.90-0.94) had an inferior area under the curve.

Conclusion: Machine learning tools that analyze echocardiographic metrics identified HCM cases with high performance. Detection of HCM cases improved when strain data was combined with standard echocardiographic metrics.

 ■ Mayo Clin Proc Digital Health 2024;2(4):564-573

Hypertrophic cardiomyopathy (HCM) is a heritable disorder characterized by cardiac hypertrophy without dilation not explained by other diseases.[1−3] Echocardiography is considered the gold standard imaging technology for HCM diagnosis.[2,3] Echocardiography requires considerable clinical skill for data acquisition, analysis, and interpretation. It has been previously proposed that machine learning (ML) may facilitate echocardiographic diagnoses.[4−6]

In adults, the diagnosis of HCM by echocardiography requires a maximal left ventricular end-diastolic wall thickness of ≥15 mm, in the absence of another cause of hypertrophy.[2,7] In addition to increased left ventricular wall thickness, other echocardiographic metrics obtained by Doppler imaging and myocardial strain imaging can increase diagnostic accuracy. These metrics are routinely obtained during comprehensive transthoracic echocardiography (TTE) of patients with HCM.

There have been previous attempts to use ML for the interpretation of echocardiographic images to enable automated diagnosis. A previous video-based echocardiographic study developed convolutional neural networks for interpretation of cardiac structure, function, and disease detection.[4] However, in that study, there was a high proportion of image segmentation failures, making these models unsuitable for use in clinical settings.[4] Another study used a video-based echocardiographic model aimed to discriminate 2-dimensional (2D) images of patients with increased thickness of cardiac walls due to cardiac amyloidosis, HCM, or end-stage renal disease.[5] This study also found limited performance with positive predictive value ranging from 74% to 77%.[5] Both the aforementioned studies used bidimensional echocardiographic image data only, which did not provide sufficient information to distinguish various conditions that may cause increased thickness of cardiac walls.

Myocardial strain measurements show distinct disease-specific patterns that discriminate different conditions that may increase the thickness of cardiac walls.[4−8] A previous echocardiographic measurement ML model included strain measurement variables and 2D variables.[9] However, a small number of patients with HCM (n=62) in that study limited the generalizability of observations. Accordingly, we developed and validated ML models to identify HCM cases using standard echocardiographic metrics retrieved from a large electronic health record (EHR)-based HCM cohort. We hypothesized that ML tools for the analysis of echocardiographic metrics can accurately identify HCM cases with high performance (graphical abstract).

From the Department of Cardiovascular Medicine (N.Z.F., M.A.A., V.K.V., J.M.B., A.P.M., K.F.L., B.K.D., J.B.G., P.A.P., J.K.O., S.R.O., G.C.K., M.J.A., A.M.A.-O.), Department of Quantitative Health Sciences (M.E., C.G.S.), Department of Molecular Pharmacology & Experimental Therapeutics, Windland Smith Rice Sudden Death Genomics Laboratory (J.M.B., M.J.A.), Department of Information Technology (A.L.Z., E.S., V.C.K.), Mayo Clinic, Rochester, MN; and Division of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL (K.S.P.).

PATIENTS AND METHODS

Cohort Selection

This case-control study was approved by the Mayo Institutional Review Board. Hypertrophic cardiomyopathy cases were matched to controls without HCM by age, sex, and date of echocardiography using a case-control ratio of 1:3 (Figure 1). All patients agreed to have their medical records used for research, and the institutional review board waived the need for informed consent.

In this study, the diagnosis of HCM was based on 2D echocardiography showing a maximal end-diastolic wall thickness of ≥15

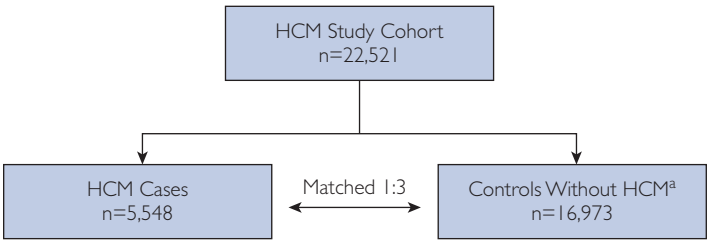


FIGURE 1. Study cohort. HCM, hypertrophic cardiomyopathy. [a]Cases and controls were matched by age, sex, and date of echocardiogram.

mm anywhere in the left ventricle, in the absence of other causes of hypertrophy.[8,9] The HCM case group included patients with HCM diagnosis who underwent clinically indicated TTE. Individuals with alternative causes for increased left ventricular mass that could mimic HCM, such as amyloidosis and hypertensive heart disease, were excluded from the HCM case group only but not from the control group of patients without HCM. This rigorous methodology ensures that the HCM case group represents a cohort of individuals with a diagnosis of HCM only.

The control group of patients without HCM included a broad spectrum of patients referred for clinically indicated echocardiograms for various reasons, including cardiac conditions that can mimic HCM such as amyloidosis and hypertensive heart disease. This approach enabled evaluation of model performance in distinguishing HCM from HCM mimics and from other indications for TTE in the routine clinical practice of an echocardiography laboratory. Additionally, the clinical decision support (CDS) tool that uses these ML models was designed to assist providers working in real-life echocardiographic clinical practices in which patients are referred for echocardiograms for a variety of reasons but are not limited to patients referred for assessment of HCM mimics. Another exclusion criterion for the case group and control group was age younger than 18 years.

For the control group, the most frequent indications for echocardiography were cardiac arrhythmias, valve disease, and heart failure (Table 1). For patients with more than 1 echocardiographic report, only the first echocardiography, both for cases and controls, was used for analysis.

TABLE 1. Indications for Transthoracic Echocardiography in the Control Group

Billing codes	Total (N=17,388), n (%)
Arrhythmias	8825 (50.8)
Valve diseases	6986 (40.2)
Hypertensive heart disease	6610 (38)
Heart failure	2235 (12.9)
Cardiac murmur	1424 (8.2)
Previous myocardial infarction	1324 (7.6)
Cardiomyopathies (not HCM)	1323 (7.6)
Amyloidosis	308 (1.8)
Acute coronary syndromes	249 (1.4)
Inherited myopathies	242 (1.4)
Sarcoidosis	164 (0.9)

HCM, hypertrophic cardiomyopathy.
Patients with more than 1 indication were included.

Data Preprocessing and Input Selection

The echocardiographic reports included 2 groups of variables: (1) patient demographic characteristics and (2) standard echocardiographic metrics relevant to HCM diagnosis (Table 2). Data from each patient's first echocardiographic report was retrieved from Mayo Clinic Unified Data Platform data warehouse electronically linked to the EHR.

All echocardiograms used guideline-directed methods for obtaining and annotating measurements. The interventricular septum diastolic thickness, posterior wall diastolic thickness, and left ventricular mass measurements were measured from bidimensional images.[10] If bidimensional images were not available, M-mode measurements were used for analysis. The left ventricular ejection fraction was calculated by the modified Quinones method, 3D volumes, or biplane Simpson method.[11] Supplemental Table 1 (available online at https://www.mcpdigitalhealth.org/) summarizes the definition of each considered echocardiographic variable, following the American Society of Echocardiography guidelines.[9,11] For reports containing more than 1 value of ejection fraction, the median values of all available measurements were used for

Mayo Clin Proc Digital Health ■ December 2024;2(4):564-573 ■ https://doi.org/10.1016/j.mcpdig.2024.08.009
www.mcpdigitalhealth.org

TABLE 2. Demographic and Echocardiographic Metrics of the Study Cohort

Variables[a]	HCM cases (N=5,548)	Controls (N=16,973)	*P*	Missing values in cases (%)	Missing values in controls (%)
Demographic characteristic					
Age (y)	58 (46-67)	58 (46-67)	.47	0	0
Male	3233 (56.7)	9685 (56.7)	.99	0	0
Body surface area (m^2)	2.0 (1.81-2.18)	1.98 (1.79-2.16)	<.001	<1	<1
Echocardiographic metrics					
Interventricular septum diastolic thickness (mm)	16 (13-20)	10 (9-11)	<.001	14	30
Thickest segment of the left ventricle (mm)	21 (18-23)	NA	NA	38	—
Left ventricular posterior wall diastolic thickness (mm)	12 (11-12)	10 (9-11)	<.001	14	31
Left ventricular mass (g/m^2)	276 (217.5-361)	175 (141-219)	<.001	15	33
Ejection fraction (%)	68 (63.5-72.5)	61 (56-65)	<.001	14	33
Mitral valve E-wave peak velocity (m/s)	0.8 (0.7-0.1)	0.8 (0.6-0.9)	<.001	8	23
Mitral valve A-wave peak velocity (m/s)	0.7 (0.5-0.9)	0.7 (0.5-0.8)	<.001	15	28
Mitral valve E:A ratio	1.13 (0.8-1.5)	1.14 (0.83-1.5)	.74	19	35
Mitral valve medial annulus e' velocity (m/s)	0.05 (0.04-0.06)	0.08 (0.06-0.1)	<.001	18	33
Mitral valve E/e' ratio (medial)	15.7 (11.7-21.7)	10 (7.5-12.9)	<.001	15	28
Mitral valve lateral annulus e' velocity (m/s)	0.07 (0.05-0.09)	0.1 (0.08-0.12)	<.001	25	39
Mitral valve E/e' ratio (lateral)	11.7 (8.6-16.7)	7.5 (5.7-10)	<.001	25	39
Left ventricular intracavitary peak velocity (m/s)	2.7 (2-3.9)	N/A	NA	35	—
Left ventricular outflow peak velocity (m/s)	3.5 (2.2-4.5)	N/A	NA	38	—
Left atrial volume index (mL/m^2)	42 (34-53)	30 (24-38)	<.001	41	52
Right ventricular systolic pressure (mm Hg)	33 (28-41)	30 (26-37)	<.001	19	23

[a]All numerical variables were summarized as median (Q1-Q3) or n (%).

E/e' ratio, ratio between peak early mitral inflow velocity and mitral early diastolic medial annular velocity; HCM, hypertrophy cardiomyopathy; NA, not available; Q1, first quartile; Q3, third quartile.

analysis. For data imputation, median values were used to address missing values as recommended by Acuna and Rodriguez.[12]

Model Development

We developed 4 supervised random forest models for the diagnosis of HCM, each including demographic characteristics and echocardiographic metrics. The input data for all models included 3 demographic variables (age, sex, and body surface area) and standard echocardiographic measurements relevant to HCM diagnosis, extracted from echocardiographic reports. Initially, a comprehensive list of echocardiographic metrics was reviewed by cardiologists with expertise in echocardiography. Metrics were categorized based on relevance to HCM diagnosis and considered for inclusion. After categorization, correlation analysis was performed with the exclusion of correlated variables. This process avoided redundancy and selected the most relevant metrics for HCM diagnosis. Demographic variables such as age, sex and body surface area were incorporated into the models to enhance predictive capability.

Throughout the model development process, demographic, standard echocardiographic metrics and strain measurements were added to the models as follows:

Model 1: This baseline model included only demographic and the standard echocardiographic metrics (Table 2) without strain variables.

Model 2: This model included left ventricular global longitudinal strain in addition to demographic and standard echocardiographic metrics. Global longitudinal strain provides a measure of myocardial deformation during the cardiac cycle.

Model 3: This model included averaged strains derived from speckle-tracking echocardiography in addition to demographic and standard echocardiographic metrics. The averaged strains were

obtained at the left ventricular base, mid, and apical segments. A total of 3 sets of averaged strains from each of these locations were included in model 3.

Model 4: This model included 17 individual strain values from each cardiac segment in addition to clinical and standard echocardiographic metrics. This granular approach enables a detailed characterization of myocardial mechanics. Supplemental Figure 1 (available online at https://www.mcpdigitalhealth.org/) shows strain images of an HCM patient in the study group from the echocardiographic review system where each numeric value of strain metrics is entered and stored.

Four ML models were trained using supervised random forest algorithms, selected for their high performance in classification tasks.[13,14] After data preparation and variable selection, the ML models classified patients as HCM cases or controls without HCM (Supplemental Figure 2, available online at https://www.mcpdigitalhealth.org/). The data set was divided into 10 subsets to perform 10-fold crossvalidation.[15] Within each fold, an 80:20 train:test split was applied. The trained models were evaluated on 20% of the data set in each fold. The models with the best performance across all folds were selected. For analysis, the following performed metrics were calculated: sensitivity, specificity, positive predictive value (PPV), negative predictive value, and area under the curve (AUC).

To enhance interpretability, we conducted a feature importance analysis to identify the most influential variables driving the classification decisions. This analysis aids in understanding the models underlying mechanisms and facilitates the interpretation of model predictions. Criteria used for classifying patients as HCM cases or controls without HCM are explicitly described to provide transparency and facilitate understanding of the model.

Statistical Analyses

Baseline demographic characteristics and echocardiographic variables were expressed as median (first quartile [Q1] and third quartile [Q3]). Differences between cases and controls were calculated using Kruskal-Wallis rank sum test for continuous data or a Pearson χ^2 test for categorical data. All tests were unpaired, and a 2-tailed P value of $<.05$ was considered statistically significant. All calculations and models were performed using R program version 4.1.2. The sensitivity, specificity, PPV, negative predictive value, and AUC of the receiver operating characteristics were calculated to evaluate model performance using the test set.

Missing values were addressed by imputing the median value of each variable separately for cases and controls. This approach was chosen to maintain consistency and accuracy within each data set while simulating real-world echocardiographic reporting scenarios. Although considering the exclusion of cases with missing values, we opted for imputation due to the limited number of cases affected and the potential introduction of bias through exclusion.[12]

RESULTS

Overall, the study cohort consisted of 5,548 HCM cases (56.7% males; median age at echocardiography, 58 years; interquartile range, 46-67 years) and 16,973 age-matched, sex-matched, and date of echocardiography-matched controls without HCM who underwent clinically indicated echocardiography from January 1, 2004, to March 15, 2019. Among the HCM cases, 4214 (76%) had mitral regurgitation, reported as mild in 2064 (37%), as moderate in 1407 (26%), and as severe in 743 (13%) cases. In total, 3863 (68%) patients reported evidence of systolic anterior motion of the mitral valve. Additionally, 3339 patients (60%) had obstructive HCM subtype with a resting left ventricular outflow tract peak velocity median value of 3.5 m/s (Q1, 2.2 m/s; Q3, 4.5 m/s).

As expected, compared with controls, patients with HCM had thicker left ventricular walls, higher left ventricular mass, and higher left ventricular ejection fraction (Table 2). The medial e' velocity was lower and the E/e' ratio was higher in HCM cases compared with those in controls without HCM, indicating that HCM cases were more likely to have elevated left ventricular filling pressure compared with controls. The left atrial volume index and the estimated right ventricular systolic

TABLE 3. Performance Metrics of ML Models to Identify HCM

Model name	AUC (95% CI)	PPV (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Model 1 (without strains)	0.92 (0.90-0.94)	0.96 (0.91-0.98)	0.87 (0.75-0.95)	0.65 (0.61-0.69)
Model 2 (with global strain)	0.97 (0.95-0.98)	0.99 (0.99-1.00)	0.98 (0.97-0.99)	0.69 (0.65-0.72)
Model 3 (with averaged strains)	0.96 (0.94-0.97)	0.95 (0.89-0.97)	0.97 (0.94-0.98)	0.84 (0.81-0.86)
Model 4 (with individual strains)	0.98 (0.97-0.99)	0.95 (0.91-0.98)	0.98 (0.96-0.99)	0.83 (0.80-0.87)

AUC, area under the curve; HCM, hypertrophic cardiomyopathy; ML, machine learning; PPV, positive predictive value.

pressure were also higher in HCM cases compared with controls (Table 2). The global left ventricular longitudinal strain was less negative for HCM cases (−15 [Q1, −17; Q3, −12]) compared with that for controls (−19 [Q1, −21; −17]; $P<.001$]. Table 2 shows percentage of missing values for each variable.

Model Performance

All models successfully distinguished HCM from controls with AUCs ranging from 0.92 to 0.98. The AUCs of model 2 (global strain; AUC, 0.97; 95% CI, 0.95-0.98), 3 (averaged strain; AUC, 0.96; 95% CI, 0.94-0.97), and 4 (individual strains; AUC, 0.98; 95% CI, 0.97-0.99) were comparable as evidenced by overlapping confidence intervals (Table 3).

Model 1 (with no strain data) had inferior AUC (0.92; 95% CI, 0.90-0.94) compared with model 2 (AUC, 0.97; 95% CI, 0.95-0.98), and model 4 (AUC, 0.98; 95% CI, 0.97-0.99). Model 1 had inferior sensitivity (0.87; 95% CI, 0.75-0.95) compared with model 2 (sensitivity, 0.98; 95% CI, 0.97-0.99), and model 4 (0.98; 95% CI, 0.96-0.99), and no significant difference to model 3 (0.97; 95% CI, 0.94-0.98). Models 3 and 4 had superior specificity (0.84; 95% CI, 0.81-0.86, and 0.83; 95% CI, 0.80-0.87, respectively) compared with models 1 (0.65; 95% CI, 0.61-0.69) and 2 (0.69; 95% CI, 0.65-0.72). All models had comparable PPV (Model 1: 0.96; 95% CI, 0.91-0.98; model 2: 0.99; 95% CI, 0.99-1.0; model 3: 0.95; 95% CI, 0.89-0.97; model 4: 0.95; 95% CI, 0.91-0.98).

The best predictors among the different models for identification of HCM are summarized in Supplemental Table 2 (available online at https://www.mcpdigitalhealth.org/). Although predictors varied depending on the model used, left ventricular outflow tract obstruction and interventricular septum diastolic thickness remained the best predictors for each model. All models that included left ventricular longitudinal strain as input (models 2-4) found these variables to be among the best predictors. Based on this, models 2-4 performed most accurately.

CDS Tool

These ML models generate input for the HCM-ML-based CDS tool. The ML-based CDS tool analyzes measurements obtained during routine echocardiograms, provides estimated probabilities of HCM, and real-time instructions to enhance diagnostic accuracy. This system automatically identifies missing metrics, promoting timely acquisition of necessary measurements before image interpretation completion and patient discharge.

Figure 2 illustrates the seamless integration of the ML-based CDS tool into the echocardiography laboratory workflows. Echocardiograms are ordered via the EHR system. Sonographers acquire cardiac ultrasound images and obtain measurements analyzed by ML models in real-time. If missing measurements are identified by the models, sonographers obtain these measurements which can be reanalyzed by the ML models automatically. The cardiologist releases the final echocardiographic report to the EHR.

The ML models have been installed in servers that interface with the echocardiographic review system. The server Operating System is Rocky Linux 8.7 (Green Obsidian), the platform is x86_64-pc-linux-gnu (64-bit) with R version 4.2.2 (2022-10-31). The processor is 10 core Intel Xeon W-1290 and Intel UHD Graphics P630. The memory is 64 GB (2 × 32 GB) DDR4-3200 UDIMM. Supplemental Figure 3 (available online at

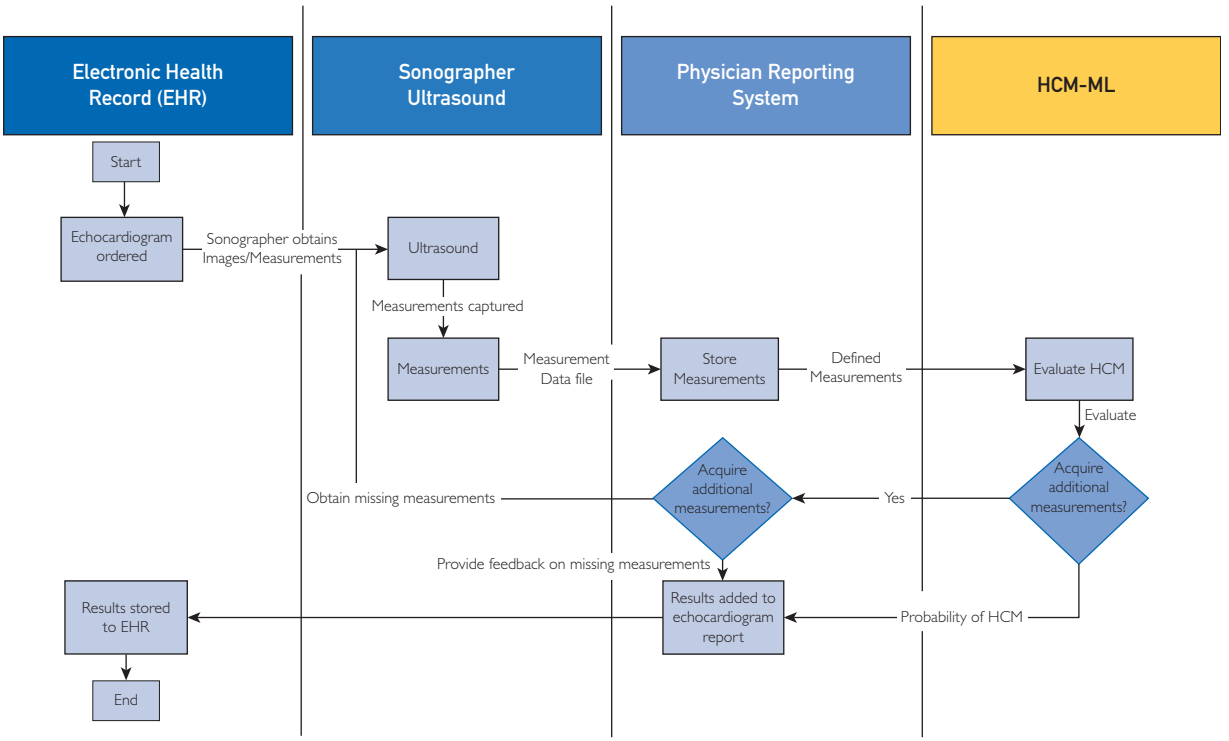


FIGURE 2. HCM-AI CDS dataflow diagram within the echocardiography laboratory clinical operations. The dataflow diagram shows the integration of the ML CDS tool within the clinical workflows of an echocardiography laboratory. This diagram delineates the processes and interactions involved in leveraging advanced computational methodologies to augment diagnostic accuracy while supporting provider workflows. CDS, clinical decision support; HCM, hypertrophic cardiomyopathy; ML, machine learning.

https://www.mcpdigitalhealth.org/) displays the architecture diagram of the HCM-CDS.

DISCUSSION

In this study, we developed, trained, and tested 4 random forest ML models to enable automated HCM case identification. These models discriminated HCM cases from controls with high performance as evident from high AUC, sensitivity, and PPV. The supervised ML models were developed to infer relationships between standard echocardiographic metrics and subsequently classify patients as HCM cases or controls. The performance of the outputs generated were evaluated through iterative training (10-fold crossvalidation) to reduce bias and variance.

For this study, supervised learning using labeled data sets were used.[16,17] Both labeled inputs and outputs were used for HCM classification, separating HCM cases from controls without HCM. The classifier used a random forest ML process, which consists of a combination of decision trees that produce a classification outcome.[11,13,14,18] Random forest can handle large data sets and works well with both categorical and continuous data.[13,14]

This study also confirms the importance of strain variables for accurate HCM identification. Models that included strain variables as input had superior performance compared with model 1, which did not include strain. Additionally, strain variables were among the best predictors for the models 2-4. Previous studies using traditional statistical analyses have also found left ventricular strain by echocardiographic aids in HCM diagnosis.[19–25]

The ML models found excellent performance in distinguishing HCM from controls without HCM, which included HCM mimics. Cases with limited image quality and diagnostically challenging situations were included in

Mayo Clin Proc Digital Health ■ December 2024;2(4):564-573 ■ https://doi.org/10.1016/j.mcpdig.2024.08.009
www.mcpdigitalhealth.org

the study sets. This approach enabled translation of these ML models to clinical practice assisting echocardiography providers in diagnosing HCM by echocardiogram. The inclusion of amyloidosis and hypertensive heart disease within the control group is a notable strength of our study. This design allowed for the assessment of model performance in differentiating HCM from both HCM mimics and other clinical indications for echocardiograms.

In our clinical practice, these ML models generate input for the HCM-ML-based clinical CDS tool. The HCM-ML CDS was designed to improve quality of echocardiographic reports, enhance report standardization, and facilitate the delivery of comprehensive and clinically relevant echocardiography reports to the EHR, thus promoting patient-centered clinical care. The HCM-ML CDS is portable to other systems.

This study focused on ML models that include demographic and echocardiographic metrics. Looking ahead, we propose to integrate electrocardiogram signals, clinical narratives, and image analysis with state-of-the-art ML techniques. Future studies should develop and validate ML for extraction of information from multiple sources to identify HCM among patients with more limited hypertrophy (13-14 mm) when present in family members of a patient with HCM or in conjunction with a positive genetic test identifying a pathogenic or likely pathogenic variant in a sarcomere gene. Data sources for these models would include clinical notes for extraction of a family history of HCM and genotype test reports in addition to echocardiographic metrics.[26,27] In the future, the ML models can be adapted to reflect updates of guideline criteria for HCM diagnosis.

Limitations

This study used real-world echocardiographic data and measurements, which were not always available from every echocardiographic report. To mitigate the effect of sparsity of echocardiographic data metrics, imputation was performed using the median value of each variable, following a previously established methodology.[12] In the study, the gold standard for comparison was the interpretation of clinically indicated echocardiograms by board-certified cardiologists. Future experiments can evaluate model performance with cardiologists blinded to referral diagnosis. In this study, the models were not tested against independent readers using the same echocardiographic parameters. Moreover, further studies are needed to assess the potential utility to differentiate obstructive and nonobstructive subtypes, particularly in cases where obstruction has not yet manifested. Data from all echocardiographic laboratories of Mayo Clinic enterprise were included in this study; portability of this model to other practices should be evaluated in the future. External validation studies using independent HCM data sets and cohorts of patients with HCM phenocopies including cardiac amyloidosis and hypertensive heart disease should also be conducted. Images of each echocardiographic were not reviewed owing to the large sample size. However, all images had been reviewed and interpreted by board-certified cardiologists trained in echocardiography during routine clinical practice.

CONCLUSION

Machine learning models for the analysis of standard echocardiographic metrics identified HCM cases with high performance. Detection of HCM cases was improved in models when echocardiographic strain measures were combined with other echocardiographic metrics. These tools may be useful for generating input to CDS to assist providers during real-time echocardiographic interpretation and reporting processes for patients with HCM. This study found the feasibility of automating echocardiographic diagnosis.

POTENTIAL COMPETING INTERESTS

Dr Pasupathy reports grants from Agency for Healthcare Research and Quality and National Institutes of Health - National Center for Advancing Translational Sciences; royalties from American College of Healthcare Executives and Mayo; consulting fees from Energesse and Guidepoint; honoraria from National Institutes of Health, Patient-Center Outcomes Research Institute, and National Science Foundation; and leadership roles in Intercultural Mutual Assistance Association. Dr Geske reports participation in the advisory board of Cytokinetics and is the site principle

investigator for industry-funded clinical trials for Bristol Myers Squibb and Cytokinetics. Dr Kane receives travel and registration honoraria from Mayo Clinic to attend scientific meetings. Dr Ackerman is consultant for Abbot, Boston Scientific, Bristol Myers Squibb, Daiichi Sankyo, Invitae, Thryv Therapeutics, and Medtronic and holds equity/Intelectual Property/royalties in AliveCor, Anumana, ARMGO Pharma, Pfizer, and UpToDate. Given her role as Editorial Board Member, Dr Arruda-Olson, had no involvement in the peer-review of this article and has no access to information regarding its peer-review. The other authors report no competing interests.

ETHICS STATEMENT

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and approved by the Institutional Review Board of Mayo Clinic. Data were anonymized before analysis to protect participant identity.

ACKNOWLEDGMENTS

The authors thank Tessa Flies and Kara M. Firzlaff for secretarial support. Drs Farahani and Aguirre contributed equally to this manuscript and should be considered joint first authors.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at https://www.mcpdigitalhealth.org/. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronyms: AUC, area under the curve; CDS, clinical decision support; EHR, electronic health record; HCM, hypertrophic cardiomyopathy; ML, machine learning; PPV, positive predictive value; Q1, first quartile; Q3, third quartile; TTE, transthoracic echocardiography

Grant Support: This work was supported by the Paul and Ruby Tsai Family Hypertrophic Cardiomyopathy Research Fund, Rochester, Minnesota, Mayo Clinic Windland Smith Rice Sudden Cardiac Death Genomics Laboratory, Rochester, Minnesota, the Mayo Clinic Center for Clinical and Translational Science, and Mayo Clinic K2R award.

Data Availability Statement: The data underlying this article cannot be shared publicly for the privacy of individuals that participated in the study. The data will be shared on reasonable request to the corresponding author.

Publication dates: Received for publication January 19, 2024; revisions received August 12, 2024; accepted for publication August 24, 2024.

Correspondence: Address to Adelaide M. Arruda-Olson, MD, PhD, Department of Cardiovascular Medicine, Mayo Clinic, Gonda Building 6, 200 First Street SW, Rochester, MN 55905 (@AMArrudaOlson).

REFERENCES

1. Maron BJ. Clinical course and management of hypertrophic cardiomyopathy. *N Engl J Med*. 2018;379(7):655-668. https://doi.org/10.1056/NEJMra1710575.
2. Ommen SR, Mital S, Burke MA, et al. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *Circulation*. 2020;142(25):e533-e557. https://doi.org/10.1161/CIR.0000000000000938.
3. Ommen SR, Mital S, Burke MA, et al. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *Circulation*. 2020;142(25):e558-e631. https://doi.org/10.1161/CIR.0000000000000937.
4. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138(16):1623-1635. https://doi.org/10.1161/CIRCULATIONAHA.118.034338.
5. Goto S, Mahara K, Beussink-Nelson L, et al. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *Nat Commun*. 2021;12(1):2726. https://doi.org/10.1038/s41467-021-22877-8.
6. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol*. 2016;68(21):2287-2295. https://doi.org/10.1016/j.jacc.2016.08.062.
7. Ommen SR, Mital S, Burke MA, et al. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *J Am Coll Cardiol*. 2020;76(25):e159-e240. https://doi.org/10.1016/j.jacc.2020.08.045.
8. Ommen SR, Ho CY, Asif IM, et al. 2024 HA/ACC/AMSSM/HRS/PACES/SCMR guideline for the management of hypertrophic cardiomyopathy: a report of the American Heart Association/American College of Cardiology Joint Committee on clinical practice guidelines. *Circulation*. 2024;149(23):e1239-e1311. https://doi.org/10.1161/CIR.0000000000001250.
9. Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr*. 2019;32(1):1-64. https://doi.org/10.1016/j.echo.2018.06.004.
10. Parato VM, Antoncecchi V, Sozzi F, et al. Echocardiographic diagnosis of the different phenotypes of hypertrophic cardiomyopathy. *Cardiovasc Ultrasound*. 2016;14(1):30. https://doi.org/10.1186/s12947-016-0072-5.
11. Quinones MA, Waggoner AD, Reduto LA, et al. A new, simplified and accurate method for determining ejection fraction with two-dimensional echocardiography. *Circulation*. 1981;64(4):744-753. https://doi.org/10.1161/01.cir.64.4.744.
12. Acuna E, Rodriguez C. The treatment of missing values and its effect on classifier accuracy. In: Banks D, McMorris FR, Arabie P, Gaul W, eds. *Classification, Clustering, and Data Mining Applications*. Springer; 2004:639-647.

13. GL Team. Random forest algorithm in machine learning: an overview. 2024: https://www.mygreatlerning.com/blog/random-forest-algorhithm/. Accessed September 27, 2024.
14. Corporate Finance Institute. Random forest. https://corporatefinanceinstitute.com/resources/data-science/random-forest/. Accessed September 27, 2024.
15. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Netw Open*. 2018;1(4):e181018. https://doi.org/10.1001/jamanetworkopen.2018.1018.
16. Delua Julianna IA. *IBM Blog.Supervised vs. unsupervised learning: what's the difference?*. IBM; 2021. https://wwwpoc.ibm.com/blog/supervised-vs-unsupervised-learning-what's-the-difference/. Accessed September 27, 2024.
17. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018;71(23):2668-2679. https://doi.org/10.1016/j.jacc.2018.03.521.
18. Uddin S, Ong S, Lu H. Machine learning in project analytics: a data-driven framework and case study. *Sci Rep*. 2022;12(1):15252. https://doi.org/10.1038/s41598-022-19728-x.
19. Kato TS, Noda A, Izawa H, et al. Discrimination of nonobstructive hypertrophic cardiomyopathy from hypertensive left ventricular hypertrophy on the basis of strain rate imaging by tissue Doppler ultrasonography. *Circulation*. 2004;110(25):3808-3814. https://doi.org/10.1161/01.CIR.0000150334.69355.00.
20. Yang H, Carasso S, Woo A, et al. Hypertrophy pattern and regional myocardial mechanics are related in septal and apical hypertrophic cardiomyopathy. *J Am Soc Echocardiogr*. 2010;23(10):1081-1089. https://doi.org/10.1016/j.echo.2010.06.006.
21. Haland TF, Almaas VM, Hasselberg NE, et al. Strain echocardiography is related to fibrosis and ventricular arrhythmias in hypertrophic cardiomyopathy. *Eur Heart J Cardiovasc Imaging*. 2016;17(6):613-621. https://doi.org/10.1093/ehjci/jew005.
22. Okada K, Kaga S, Mikami T, et al. Characteristic systolic waveform of left ventricular longitudinal strain rate in patients with hypertrophic cardiomyopathy. *Heart Vessels*. 2017;32(5):591-599. https://doi.org/10.1007/s00380-016-0906-y.
23. Tower-Rader A, Mohananey D, To A, Lever HM, Popovic ZB, Desai MY. Prognostic value of global longitudinal strain in hypertrophic cardiomyopathy: a systematic review of existing literature. *JACC Cardiovasc Imaging*. 2019;12(10):1930-1942. https://doi.org/10.1016/j.jcmg.2018.07.016.
24. Haland TF, Edvardsen T. The role of echocardiography in management of hypertrophic cardiomyopathy. *J Echocardiogr*. 2020;18(2):77-85. https://doi.org/10.1007/s12574-019-00454-9.
25. Geske JB, Bos JM, Gersh BJ, Ommen SR, Eidem BW, Ackerman MJ. Deformation patterns in genotyped patients with hypertrophic cardiomyopathy. *Eur Heart J Cardiovasc Imaging*. 2014;15(4):456-465. https://doi.org/10.1093/ehjci/jet234.
26. Moon S, Liu S, Scott CG, et al. Automated extraction of sudden cardiac death risk factors in hypertrophic cardiomyopathy patients by natural language processing. *Int J Med Inform*. 2019;128:32-38. https://doi.org/10.1016/j.ijmedinf.2019.05.008.
27. Rahit K, Avramovic V, Chong JX, Tarailo-Graovac M. GPAD: a natural language processing-based application to extract the gene-disease association discovery information from OMIM. *BMC Bioinformatics*. 2024;25(1):84. https://doi.org/10.1186/s12859-024-05693-x.