HCM-Echo-VAR-Ensemble: Deep Ensemble Fusion to Detect Hypertrophic Cardiomyopathy in Echocardiograms

Abdulsalam Almadani, *Graduate Student Member, IEEE*, Atifa Sarwar, Emmanuel Agu, Monica Ahluwalia, and Jacques Kpodonu

Abstract*—*Goal: **To detect Hypertrophic Cardiomyopathy (HCM) from multiple views of Echocardiogram (cardiac ultrasound) videos. *Methods:* we propose *HCM-Echo-VAR-Ensemble*, a novel framework that performs binary classification (HCM vs. no HCM) of echocardiogram videos directly using an ensemble of state-of-the-art deep VAR architectures models (SlowFast and I3D), and fuses their predictions using majority averaging ensembling. *Results: HCM-Echo-VAR-Ensemble* achieved state-of-the-art accuracy of 95.28%, an F1-Score of 95.20%, a specificity of 96.20%, a sensitivity of 93.97%, a PPV of 96.46%, an NPV of 94.17%, and an AUC of 98.42%, outperforming a comprehensive set of baselines including other ensembling approaches. *Conclusions:* Our proposed HCM-Echo-VAR-Ensemble framework demonstrates significant potential for improving the sensitivity and accuracy of HCM detection in clinical settings, particularly by ensembling the complementary strengths of the SlowFast and I3D deep VAR models. This approach can enhance diagnostic consistency and accuracy, enabling reliable HCM diagnoses even in low-resource environments.**

***Index Terms*—Cardiac assessment, computer vision, deep ensemble learning, deep learning, digital health, echocardiogram, hypertrophic cardiomyopathy (HCM), video analysis.**

***Impact Statement*— *HCM-Echo-VAR-Ensemble* can be integrated into Point of Care (PoC) tools that improve the consistency of cardiologists and enable non-experts to accurately diagnose HCM even in low-resource settings.**

I. Introduction

Motivation: Cardiovascular disease (CVD) critically impacts global health, accounting for over 19 million deaths annually worldwide [1]. In the United States, CVD causes more than 2 deaths for every 1,000 individuals [1]. Hypertrophic cardiomyopathy (HCM) is the most common genetic cardiac condition, affecting approximately one person in every 500 people [2]. However, HCM is a cardiomyopathy that is still underdiagnosed, as only 13% of HCM cases are clinically diagnosed. HCM affects the cardiac muscle, reducing contractile strength and relaxation, which can cause various complications such as Sudden Cardiac Death (SCD), abnormal heart rhythms, stroke, and Heart Failure (HF) [3]. HCM is a primary contributor to SCD in young athletes [4]. Notably, black patients with HCM have worse outcomes than white patients [5]. Longitudinal risk assessment can identify and diagnose family members who may have inherited the condition or are at risk of developing it early. Furthermore, such assessments can prevent complications such as SCD, abnormal rhythms, stroke, and heart failure [3] in patients with HCM.

Problem and Challenges: Detecting HCM in echocardiograms is challenging due to the condition's heterogeneous clinical presentations, complicating consistent diagnosis [6]. This variability necessitates advanced multi-view analysis, as single-view assessments may overlook critical details. HCM diagnoses also relies heavily on the cardiologist's expertise, introducing subjectivity and potential variability [7]. Moreover, similarities with other heart diseases, such as Fabry disease, hypertension, and cardiac amyloidosis, can lead to misdiagnosis or over-diagnosis [7]. Our study addresses these issues by employing ensemble deep VAR models to automate and improve HCM detection accuracy. We generated heatmaps to visualize predictive regions in multi-view echocardiograms. Our models can be the basis of a consistent and reliable tool for HCM detection, reducing dependence on specialized expertise, benefiting settings with limited access to cardiologists.

Prior work Recent studies have employed Artificial Intelligence (AI) methods, including machine learning and

Received 29 December 2023; revised 19 August 2024 and 21 October 2024; accepted 21 October 2024. Date of publication 25 October 2024; date of current version 13 December 2024. This work was supported by Academic & Research Computing Group at Worcester Polytechnic Institute. The review of this article was arranged by Editor Paolo Bonato. *(Corresponding author: Abdulsalam Almadani.)*

Abdulsalam Almadani is with the Data Science Program, Worcester Polytechnic Institute, Worcester, MA 01609 USA, and also with the College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh 13318, Saudi Arabia (e-mail: aalmadani@wpi.edu).

Atifa Sarwar and Emmanuel Agu are with the Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: asarwar@wpi.edu; emmanuel@cs.wpi.edu).

Monica Ahluwalia is with the Medical Director of Inherited Cardiac Diseases Program, Division of Cardiovascular Medicine, Boston Medical Center, Boston, MA 02118 USA, and also with the Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115 USA.

Jacques Kpodonu is with the Division of Cardiac Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115 USA.

Digital Object Identifier 10.1109/OJEMB.2024.3486541

TABLE I
COMPARISON OF PROPOSED APPROACH WITH EXISTING METHODS FOR CLASSIFYING HCM BASED ON ECHOCARDIOGRAMS

Author	Target Labels	Dataset Type	Sample Size	Proposed Approach	Targeted View	Accuracy (%)
Farahani et al. [12]	HCM/No/Possible	EHR data	11,562	Random forest	NA	70.51
Rosca et al. [13]	HCM/Normal	Clinical and Echo. Parameters	151	DL	NA	84
Zhang et al. [9]	HCM/Normal	Image	4,950	VGG-16	PLAX, A4C	93
Madani et al. [14]	HCM/Normal	Image	2,269	CNNs	A4C	92.3
Balaji et al. [8]	HCM/DCM/ Normal	Video	60	BPNN	PSAX	90.2
Ghorbani et al. [10]	HCM/Normal	Video	2,883	Inception-ResNet	A4C	75
Almadani et al. [11]	HCM/Normal	Video	1,553	SlowFast	A2C, A4C, PLAX, PSAX	93.13
HCM-Echo-VAR-Ensemble.	HCM/Normal	Video	1,553	SlowFast & I3D	A2C, A4C, PLAX, PSAX	**95.28**

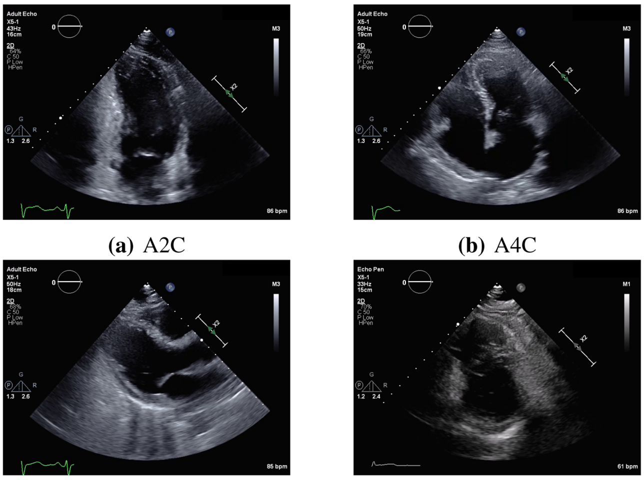
deep learning techniques such as BackPropagation Neural Network (BPNN) [8], VGG-16 [9], Inception-ResNet [10], and SlowFast [11] to detect HCM. These models, implemented using diverse datasets encompassing clinical, static, and dynamic echocardiographic data, have yielded accuracies of up to 93.13%. As detailed in Table I, the enhancements in our HCM-Echo-VAR-Ensemble method results in significant advancements in the accuracy and consistency of diagnosing HCM, positively impacting clinical outcomes. For a comprehensive review, refer to the supplementary material.

Our approach: In this paper, we propose *HCM-Echo-VAR-Ensemble*, a deep ensemble learning framework that utilizes Two-Stream Inflated 3D ConvNets (I3D) and SlowFast deep VAR models for analyzing echocardiogram videos. First, echocardiogram videos are segmented into frames, ordered based on their temporal position in the input video, and input to each model individually. SlowFast has two branches: a slow branch captures spatial features, and a fast branch that captures temporal features. Lateral connections integrate features from both paths to enhance spatio-temporal learning for detecting HCM. I3D builds upon image classification architectures by inflating pooling kernels and filters into three dimensions to enhance spatio-temporal classifiers. The learned features and predictions are combined in a majority averaging ensemble prediction to classify each video as HCM or normal. Additionally, *HCM-Echo-VAR-Ensemble* was pretrained on the EchoNet-Dynamic dataset [15], which includes 10 030 A4C echo videos, before fine-tuning on our HCM-Net dataset.

The novelty of our work: Our work introduces an innovative approach to echocardiogram video analysis for HCM prediction. While prior studies predominantly employed 2D/3D Convolutional Neural Networks (CNNs) to analyze still images, single echocardiogram views, or brief video clips with lower accuracy, our research takes a distinctive route. We employ ensemble techniques, combining multiple VAR models, specifically the I3D and SlowFast. This unique combination directly assesses multiple echocardiogram views such as A2C, A4C, PLAX, and PSAX used to capture echocardiogram videos. The overarching objective is to extract predictive spatio-temporal features and effectually detect HCM. The Grad-CAM technique [16] was used to interpret and reason about the results generated by *HCM-Echo-VAR-Ensemble*, providing insights into its inner workings and decision-making process. This study is the first of its kind to creatively ensemble and combine the capabilities of I3D and SlowFast VAR models for echocardiogram video analysis, particularly in the context of HCM prediction. Our innovative *HCM-Echo-VAR-Ensemble* framework not only achieves a remarkable level of accuracy but also shows an impressive sensitivity. These achievements are noteworthy as they achieve performance that human experts might find difficult to achieve consistently. By proposing a novel methodology that ensembles advanced deep VAR models, our research advances HCM detection.

Key contributions:

- *Novel Framework:* We present *HCM-Echo-VAR-Ensemble*, a deep ensemble learning framework that combines the strengths of I3D and SlowFast Deep VAR models with Majority Average Prediction to detect HCM from echocardiogram videos.
- *Rigorous Evaluation of HCM-Echo-VAR-Ensemble:* Extensive experiments were conducted to evaluate *HCM-Echo-VAR-Ensemble*, comparing it with state-of-the-art baselines, including individual VAR models such as R(2+1)D, TSM, TSN, I3D, and SlowFast, as well as various VAR ensembles fused using different ensembling methods including Majority Voting Prediction and Weight Prediction. In total, our rigorous evaluation explored over 90 combinations of VAR models and ensembling approaches. Ultimately, *HCM-Echo-VAR-Ensemble*'s approach of ensembling the outputs of SlowFast and I3D using Majority Averaging proved to be most effective. Using ensemble learning, *HCM-Echo-VAR-Ensemble* achieves a classification accuracy of 95.28%, an F1-score of 95.20%, specificity of 96.58%, sensitivity of 93.97%, PPV of 96.46%, NPV of 94.17%, and an AUC of 98.42%, thereby outperforming existing baselines.
- *Interpretability of HCM-Echo-VAR-Ensemble predictions:* We utilize Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight the most predictive regions of echocardiogram videos and provide insights into *HCM-Echo-VAR-Ensemble*'s decision-making process, enhances the transparency and interpretability of our model's results.
- *Clinical Applicability:* The framework's enhanced diagnostic consistency and accuracy, are crucial in clinical



(a) A2C **(b)** A4C

(c) PLAX **(d)** PSAX

Fig. 1. Echocardiogram views employed in HCM diagnoses.

settings, particularly in low-resource environments. It has the potential to be integrated into Point of Care (PoC) tools, aiding both cardiologists and non-experts in reliably diagnosing HCM.

The rest of the paper is organized as follows: Section II describes our dataset, baseline models, our proposed *HCM-Echo-VAR-Ensemble* framework, and evaluation metrics. Section III presents a comprehensive performance evaluation of *HCM-Echo-VAR-Ensemble* and outlines the results achieved. Section IV discusses our paper's findings and their implications, with Section V concluding the paper. The supplemental materials section reviews previous work and highlights their limitations, a detailed overview of all deep VAR architectures, and ensemble methods that are explored in this study.

II. Materials and Methods

A. Dataset

Our HCM echo dataset, *HCM-Net*, was collected through a collaborative effort of cardiac specialists from the HCM Program at Boston Medical Center (BMC) under the Institutional Review Board (IRB) approval number H-44101. The dataset includes 1,553 echocardiogram videos from both HCM patients and controls, recorded between 2016 and 2021. These videos capture diverse cardiac views, including Apical 2-Chamber (A2C), Apical 4-Chamber (A4C), Parasternal Long Axis (PLAX), and Parasternal Short Axis (PSAX), among others, as illustrated in Fig. 1. Diagnoses were confirmed by board-certified cardiologists through clinical history, echocardiography, and genetic testing.

The dataset was split into training (70%), validation (15%), and testing (15%) sets, maintaining a balanced prevalence of HCM across the partitions. Additional insights into the dataset's structure and characteristics can be found in the supplementary materials.

B. HCM-Echo-VAR-Ensemble: The Proposed Framework

Fig. 2 presents an overview of *HCM-Echo-VAR-Ensemble*, an ensembled improvement of *HCM-Dynamic-Echo*, previously proposed by Almadani et al. [11], which utilizes only the SlowFast model for HCM detection from echocardiography videos. *HCM-Echo-VAR-Ensemble* advances this approach by ensembling SlowFast with a second VAR architecture, and employing a majority averaging ensembling method to combine both models, thus improving HCM classification performance.

HCM-Echo-VAR-Ensemble innovatively combines the best of both models, leveraging the single-pathway architecture of I3D and the dual-pathway architecture of SlowFast to achieve more accurate classification. I3D provides efficient spatial modeling and parallel processing, while SlowFast captures both slow and fast motion patterns with its asymmetric design. By integrating their predictions through majority averaging as in (1), the ensemble overcomes the limitations of I3D in modeling complex long-term temporal dependencies.

$$EP(x) = \operatorname*{argmax}_{c} \left(\sum_{i=1}^{N} \mathrm{I}(M_i(x) = c) \right) \tag{1}$$

where argmax_c returns the class label with the highest sum of predictions, and $\mathrm{I}(M_i(x) = c)$ is an indicator function that returns 1 if $M_i(x)$ predicts class c for input x, and 0 otherwise. This method leverages the strengths of both models for superior accuracy in HCM detection. *HCM-Echo-VAR-Ensemble* is rigorously evaluated to demonstrate that the ensemble outperforms both of the individual I3D and SlowFast VAR models for the task of HCM diagnosis in echocardiograms, with an accuracy of 95.28% and a sensitivity of 93.97%, compared to 86.70% accuracy and 87.07% sensitivity for I3D, and 93.13% accuracy and 91.38% sensitivity for SlowFast.

C. Baseline Models

To rigorously evaluate HCM-Echo-VAR-Ensemble, the following state-of-the-art deep VAR models were trained as baselines for the task of detecting HCM from echocardiogram videos: R(2+1)D [17], The Temporal Shift Module (TSM) [18], Temporal Segment Networks (TSN) [19], SlowFast [20], and I3D [21].

For detailed information on the architecture and performance of these models, refer to the original papers. Additionally, the Supplementary Material includes descriptions of the VAR models and an explanation of the ensembling techniques used.

D. Evaluation Metrics

To evaluate the performance of *HCM-Echo-VAR-Ensemble*, we utilized a comprehensive set of machine learning metrics, including accuracy, sensitivity, specificity, and AUC-ROC, as well as clinically relevant measures such as Positive Predictive Value (PPV) and Negative Predictive Value (NPV). For more details on the definitions and formulas of these metrics, refer to the supplementary materials.

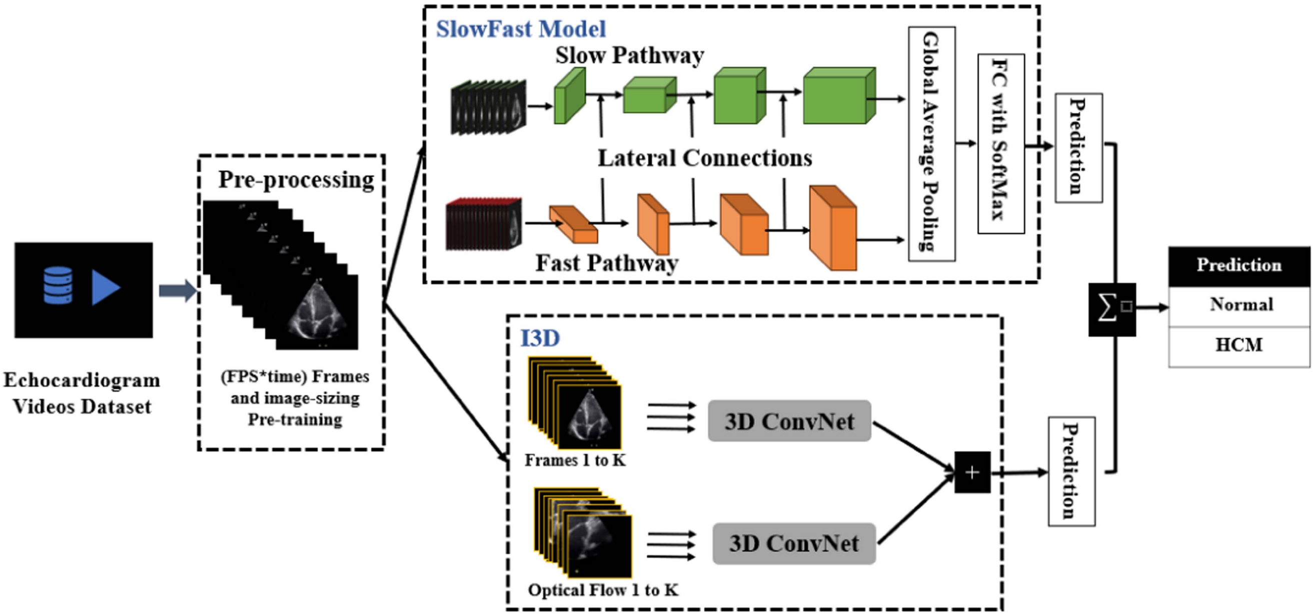


Fig. 2. HCM-Echo-VAR-Ensemble: Our proposed ensembling framework that uses deep VAR models for HCM classification. As a pre-processing step, all VAR models were pre-trained on a large EchoNet Dynamic dataset to address the data scarcity issue. Thereafter, all models were fine-tuned on our smaller HCM dataset to facilitate accurate classification of test echocardiograms as HCM vs. no HCM.

III. Results

In this study, a rigorous evaluation was conducted to assess the performance of the *HCM-Echo-VAR-Ensemble* model. The evaluation included a comparative analysis of various baseline deep VAR models, such as SlowFast, I3D, and others, as well as their ensemble combinations. Multiple ensemble techniques were explored to optimize the model's performance. The evaluation metrics encompassed a comprehensive set of measures, including accuracy, sensitivity, specificity, AUC-ROC, and confusion matrices. Additionally, Grad-CAM was employed to interpret and visualize the model's decision-making process, providing insights into the most predictive regions in the echocardiogram videos. The analysis demonstrated the superiority of the *HCM-Echo-VAR-Ensemble* model over individual models, achieving a classification accuracy of 95.28%, a sensitivity of 93.97%, and an AUC of 98.42%.

A. Ensembling Techniques vs. Baseline Models

To determine the best models for inclusion in *HCM-VAR-Echo-Ensemble*, Various deep VAR architectures were ensembled using different techniques. Evaluation on the test dataset showed that SlowFast was one of the most effective base networks to be inclcued in ensembles. As Fig. 3 shows that among the individual models, *HCM-Dynamic-Echo* had the highest accuracy of 93.13%, followed by I3D at 86.70%, TSN at 85.41%, and R(2+1)D and TSM at 77.68%. The performance of these models provided a basis against which to compare the improvements attributable to ensembling.

Table II shows detailed performance of various individual models and ensembles on various metrics with SlowFast using different ensemble techniques. The ensemble model *HCM-VAR-Echo-Ensemble* achieved the highest accuracy 95.28%, with an F1-Score of 95.20%, sensitivity of 93.97%, specificity of 96.58%, PPV of 96.46%, NPV of 94.17%, and AUC of 98.42%. These improvements enhance diagnostic consistency and reliability, ensuring more accurate and dependable HCM detection, which is crucial for good clinical outcomes. Refer to the supplementary materials for a visual comparison of the accuracies of individual models and ensembles.

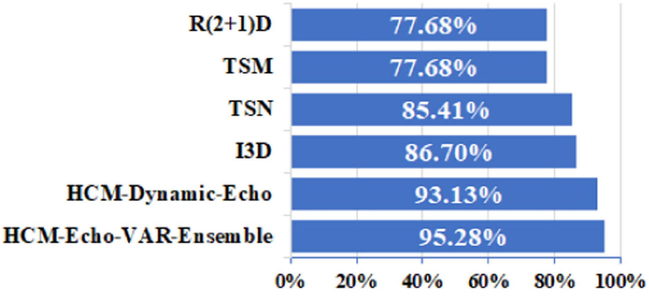


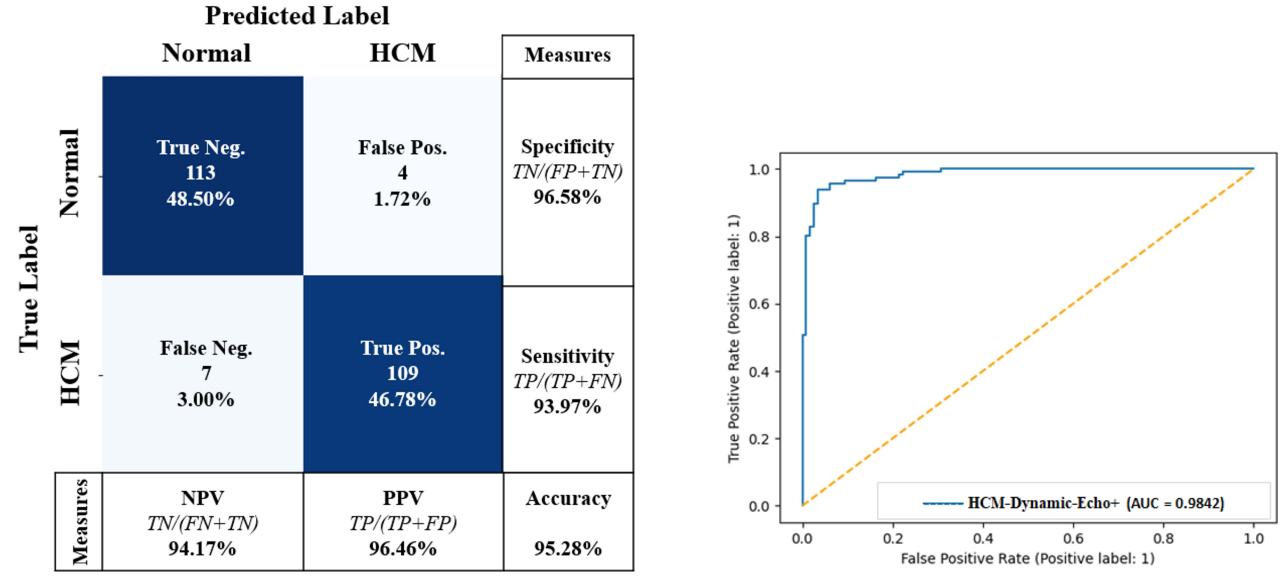
Fig. 3. Test accuracy comparison (%) for *HCM-Echo-VAR-Ensemble* and individual base models.

B. Evaluating the Performance of HCM-Echo-VAR-Ensemble

Fig. 4 illustrates the overall performance of the *HCM-Echo-VAR-Ensemble* model in accurately classifying HCM cases from echocardiogram videos. The figure combines both the confusion matrix and AUC-ROC curve to provide a complete overview of the model's reliability. Fig. 4(a) shows the confusion matrix along with key performance metrics, all derived from the model's performance on the test set. The confusion matrix

TABLE II
COMPARING PERFORMANCE METRICS FOR ENSEMBLING VARIOUS DEEP VAR MODELS USING DIFFERENT METHODS ALONG WITH INDIVIDUAL BASE MODELS

Method	Model Name	Accuracy	F1-Score	Specificity	Sensitivity	PPV	NPV	AUC
Majority Averaging Method	**HCM-Echo-VAR-Ensemble**	**95.28%**	**95.20%**	96.58%	**93.97%**	96.46%	**94.17%**	98.42%
	SlowFast & R(2+1)D	93.99%	93.81%	96.58%	91.38%	96.36%	91.87%	97.95%
	SlowFast & I3D & R(2+1)D	93.99%	93.91%	94.87%	93.10%	94.74%	93.28%	98.41%
	SlowFast & TSN	93.56%	93.39%	95.73%	91.38%	95.50%	91.80%	97.35%
	SlowFast & I3D & TSN	93.13%	92.98%	94.87%	91.38%	94.64%	91.74%	97.44%
Majority Voting Method	SlowFast & I3D & R(2+1)D	93.13%	92.98%	94.87%	91.38%	94.64%	91.74%	98.41%
	SlowFast & I3D & TSN	92.27%	92.04%	94.87%	89.66%	94.55%	90.24%	97.44%
	SlowFast & I3D & TSM	89.70%	89.47%	91.45%	87.93%	91.07%	88.43%	97.85%
	SlowFast & R(2+1)D & TSN	89.70%	89.19%	94.02%	85.34%	93.40%	86.61%	97.35%
	SlowFast & I3D	89.70%	88.68%	**98.29%**	81.03%	**97.92%**	83.94%	98.42%
Weighting predication Method	SlowFast & I3D	94.85%	94.74%	96.58%	93.10%	96.43%	93.39%	**98.45%**
	SlowFast & I3D & R(2+1)D	94.42%	94.32%	95.73%	93.10%	95.58%	93.33%	98.42%
	SlowFast & R(2+1)D	93.99%	93.81%	96.58%	91.38%	96.36%	91.87%	97.96%
	SlowFast & I3D & TSM	93.56%	93.45%	94.87%	92.24%	94.69%	92.50%	98.02%
	SlowFast & I3D & TSN	93.56%	93.45%	94.87%	92.24%	94.69%	92.50%	97.57%
Individual Base Model	HCM-Dynamic-Echo	93.13%	92.98%	94.87%	91.38%	94.64%	91.74%	93.13%
	I3D	86.70%	86.70%	86.32%	87.07%	86.32%	87.07%	86.70%
	TSN	85.41%	84.68%	89.74%	81.03%	88.68%	82.68%	85.39%
	TSM	77.68%	75.24%	87.18%	68.10%	84.04%	73.38%	77.64%
	R(2+1)D	77.68%	75.93%	84.62%	70.69%	82.00%	74.44%	77.65%



		Predicted Label		
		Normal	HCM	Measures
True Label	Normal	True Neg. 113 48.50%	False Pos. 4 1.72%	Specificity *TN/(FP+TN)* 96.58%
	HCM	False Neg. 7 3.00%	True Pos. 109 46.78%	Sensitivity *TP/(TP+FN)* 93.97%
	Measures	NPV *TN/(FN+TN)* 94.17%	PPV *TP/(TP+FP)* 96.46%	Accuracy 95.28%

(a) *HCM-Echo-VAR-Ensemble* confusion matrix and performance measures.

(b) AUC-ROC curve for the *HCM-Echo-VAR-Ensemble* framework.

Fig. 4. Evaluating the performance of HCM-Echo-VAR-Ensemble.

presents a side-by-side comparison of the actual labels versus the predicted labels assigned by *HCM-Echo-VAR-Ensemble*. These metrics are useful for evaluating the model's classification performance.

HCM-Echo-VAR-Ensemble demonstrated remarkable predictive capabilities by correctly classifying 113 normal videos and 109 HCM videos, yielding an impressive accuracy rate of 95.28%. The model only misclassified 4 normal videos as HCM and 7 HCM videos as normal, resulting in an error rate of 4.8%. Furthermore, *HCM-Echo-VAR-Ensemble* displayed an impressive sensitivity or True Positive Rate (TPR) of 93.9%. Additionally, *HCM-Echo-VAR-Ensemble* exhibited a high specificity or a True Negative Rate (TNR) of 96.5%. The model's precision was also noteworthy, with a PPV of 96.4%, indicating that 96.4% of the videos identified as HCM were indeed cases of HCM. Conversely, its NPV of 94.2% indicates that 94.2% of the videos it categorized as normal were indeed normal.

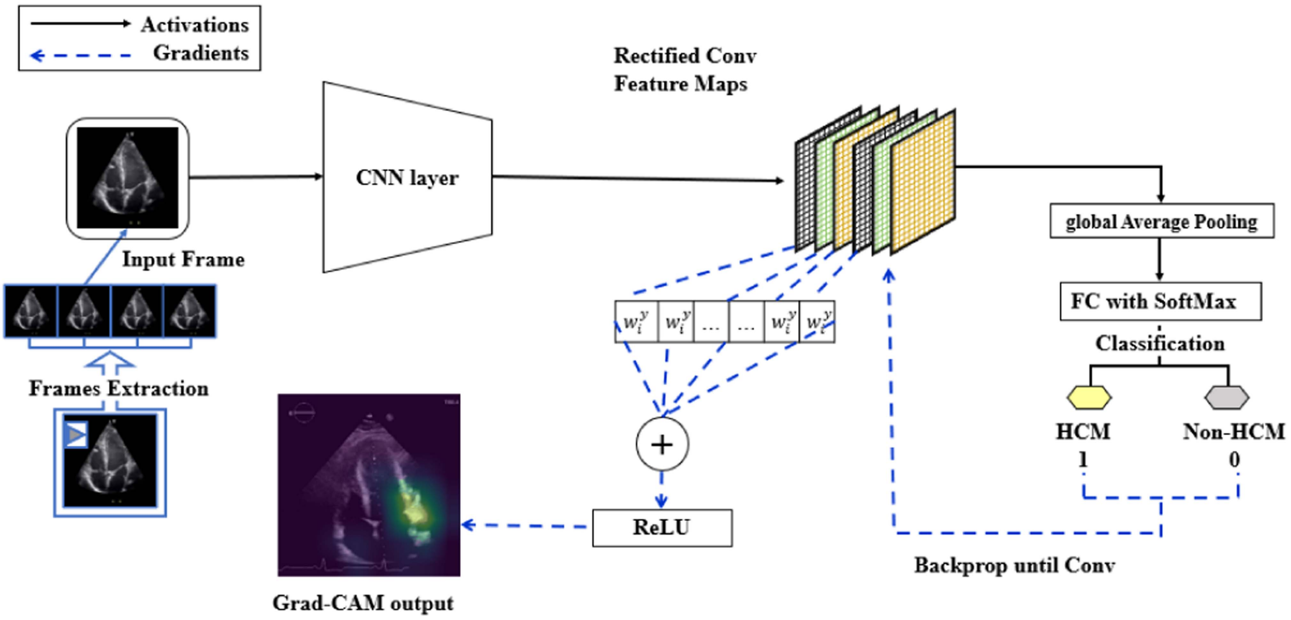


Fig. 5. A comprehensive visual representation of the key procedural steps in the *Grad-CAM* technique.

As shown in Fig. 4(b), *HCM-Echo-VAR-Ensemble* achieved an AUC of 98.42%, implying that the classifier has a very high ability to distinguish between positive and negative cases. This low error rate in predicting HCM labels confirms that *HCM-Echo-VAR-Ensemble* is an excellent binary classifier. These comprehensive analyses validate *HCM-Echo-VAR-Ensemble* as being effective and accurate in detecting HCM from echocardiogram videos.

C. Comparison of Accuracies for Various Ensembles Techniques

Various ensembling techniques were compared including Majority Averaging, Majority Voting, and Performance-Based Weighting techniques, achieving different performance levels compared to the individual models as shown in Table II. For example, the R(2+1)D and SlowFast ensemble using Majority Averaging had an accuracy of 93.99%, which was much better than the individual models. Moreover, the SlowFast and I3D ensemble with Majority Averaging, which is utilized in *HCM-Echo-VAR-Ensemble*, was the best performer with an accuracy of 95.28%, proving its effectiveness.

D. Evaluating Efficacy of Ensemble Techniques

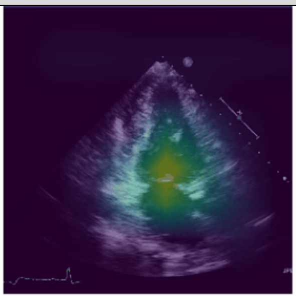
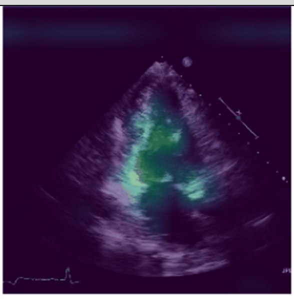
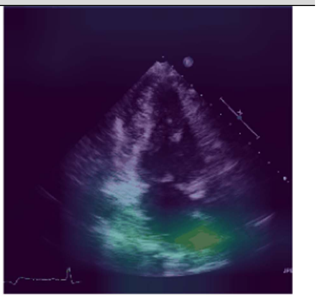
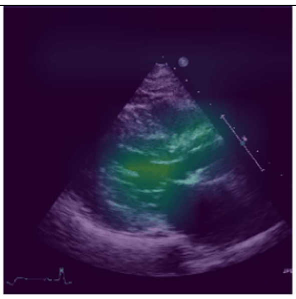
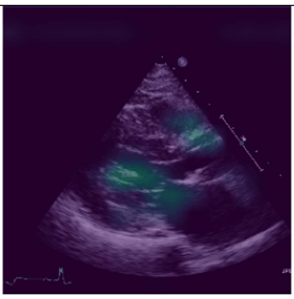
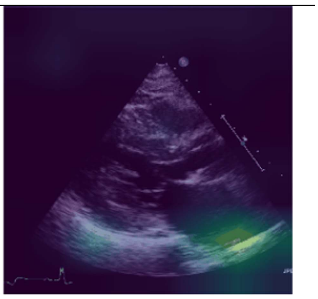
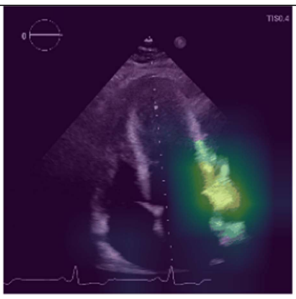
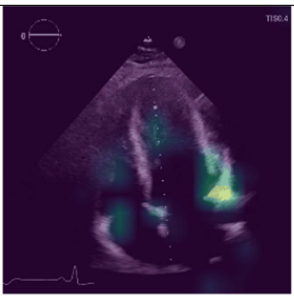
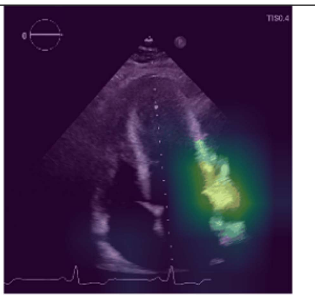
Ensembles have many advantages over individual models, especially in increasing predictive accuracy. Ensembling reduces the limitations of individual models by combining their strengths to achieve superior performance. The *HCM-Echo-VAR-Ensemble* achieved the best performance, demonstrating its utility. Its accuracy is higher than any individual model, demonstrating the power of model ensembling in achieving superior performance.

E. Explainability by Visualizing VAR Activation Maps With Grad-CAM

To facilitate interpretation of model outputs, the Gradient-weighted Class Activation Mapping (Grad-CAM) method was used to create heatmaps highlighting predictive echocardiogram regions that the deep learning models found to have the most predictive value. Grad-CAM presents insights on the inner workings of the Slowfast and I3D models, as well as the results of ensembling them. Features were extracted and the relative importance of each activation map for predicting a given class, were measured. Grad-CAM essentially generates a heatmap that visualizes the gradients of feature maps in the model's CNN layer for the task of HCM classification. Fig. 5 presents the key procedural steps of the Grad-CAM computation. For a comprehensive mathematical derivation of how Grad-CAM computes predictive regions, refer to the supplementary materials.

Table III illustrates the effectiveness of using Grad-CAM for interpretation. The Grad-CAM heatmap utilizes a yellowish color to highlight important regions that have a high influence on the model's final prediction. This enhanced visualization facilitates the identification of important factors that facilitate the decision-making process of model concerning the target class (y). It is instructive to note that the regions focused on by the HCM models varied based on the cardiac views under analysis. In particular, the main areas of focus include the left ventricular wall thickness (LVWT), the Right ventricle (RV), the left ventricle (LV), the left atrium (LA), and interventricular septum (IVS). From a medical perspective, human healthcare providers also tend to focus on these regions to gain insights necessary for diagnoses, and to support their final decisions.

TABLE III
INTERPRETATION OF SLOWFAST AND I3D MODEL PERFORMANCE DEPENDS ON NEARBY STRUCTURES AND FEATURES FROM DIFFERENT CARDIAC VIEWS

Structures and features	SlowFast Model		I3D Model
	Slow Pathway	Fast Pathway	
A2C			
PLAX			
A4C			

These representative cases illustrate regions of significant interest within the Slow and Fast pathways of the SlowFast model, along with insights from the I3D model's perspective. The ensemble of colored areas in each case depicts 95.5% confidence intervals. Based on the cardiac views analyzed, these regions, which include LV, RV, LVWT, LA, and IVS, are highly significant for healthcare practitioners because they facilitate judging how to model performance when making decisions about cardiac diagnoses.

IV. Discussion

We successfully achieved our primary objective, which was research and development of *HCM-Echo-VAR-Ensemble*, a novel Deep Learning Framework along with rigorous evaluation and interpretation of its results. This framework is able to classify echocardiogram videos into 'Normal' and 'HCM' categories with high accuracy, F1-Score, sensitivity, and various other classification metrics. The minimal divergence between training and validation accuracy trajectories of both *HCM-Echo-VAR-Ensemble* and base models demonstrated that overfitting was insignificant during the training process. Additional information on the training and validation accuracy curve trajectories can be found in the supplementary materials.

An ensemble of deep VAR models performs well for the task of HCM classification: This can be attributed partly to the previously established ability of VAR models for HCM prediction [10], [11], as well as the improved performance due to the fusion of multiple VAR models via ensembling, thereby combining the strengths of individual models. This finding is impressive given the complexity of interpreting dynamic echocardiograms, and the influence of a range of factors including the patient's demographics, health status, and echocardiogram settings (e.g. resolution, lighting, color Doppler settings, and background variations). These range of factors contribute significantly to an observed variability across different video clips belonging to the same class. Majority Averaging ensembling outperformed the Majority Voting, and Performance-Based Weighting baseline fusion strategies as well as individual VAR models.

Grad-Cam visualizations suggest that the inner workings of *HCM-Echo-VAR-Ensemble* are as expected We utilized Grad-CAM to generate visual insights into regions of importance based on cardiac views and HCM presence. This technique [16] sheds light on how these regions influence the model's classification predictions, particularly using the SlowFast and Inflated 3D deep VAR models. It detects regions associated with wall thickness, and the left ventricle regions appear as influential areas in *HCM-Echo-VAR-Ensemble*'s classification decisions. Moreover, it is imperative to note that *HCM-Echo-VAR-Ensemble*'s predictions are not intended to replace medical professionals but rather to serve as a valuable tool to expedite their diagnostic processes, improve their consistency, and enhance overall efficiency.

Study limitations: Our work has a few limitations. It explored only echocardiogram views such as A2C, A4C, PLAX, and PSAX for HCM diagnoses. However, prior work suggests that other views, namely, the apical 3-chamber [9], subcostal views [22] and speckle tracking [23], might also be useful along with Doppler measurements [23] and other clinical data, which may provide additional information or insights for diagnosis. Additionally, *HCM-Echo-VAR-Ensemble* uses an ensembling technique that combines the best performing combination of deep VAR models without including external data. Specifically, *HCM-Echo-VAR-Ensemble* utilizes an ensemble technique that prioritizes the selection of the most accurate model from each baseline without relying on Stacked Generalization or Bootstrap Aggregating methods, which we plan to explore in the future. Furthermore, The echocardiogram dataset has some missing participant demographic information, which is necessary for complete validation of *HCM-Echo-VAR-Ensemble* and to improve the generalizability of our results.

Future work: In future research, we intend to explore different frame segmentation methods such as extracting the Left Ventricle (LV) region [8], [10], [14] from each frame, as well as the relevant features that have an impact on HCM for echocardiogram interpretation. This will enable *HCM-Echo-VAR-Ensemble* to focus better on the area of interest and ignore irrelevant details. Additionally, we aim to explore a wider range of ensembling techniques and VAR models. Furthermore, we will research and develop novel techniques that enable models to first identify the specific cardiac view before then identifying the key features that indicate high HCM risk in each view. We believe these planned future directions have the potential to further advance HCM diagnosis, particularly ideas around incorporating view detection combined with leveraging the power of ensembling deep VAR techniques. Ultimately, these planned steps will facilitate the achievement of our goal of creating an interpretable AI model for diagnosing HCM and various heart conditions and associated complications early and accurately.

V. Conclusion

We proposed *HCM-Echo-VAR-Ensemble*, our solution for consistent and accurate HCM detection. This framework detects Hypertrophic Cardiomyopathy (HCM) in echocardiograms using an ensemble of deep Video Action Recognition (VAR) models. Our approach combines the strengths of I3D and SlowFast models, which capture both spatial and temporal features, and uses the majority averaging technique for fusing individual model predictions. In rigorous evaluation on a dataset of 1553 echocardiogram videos from different cardiac perspectives, *HCM-Echo-VAR-Ensemble* achieves state-of-the-art accuracy of 95.28% and an F1-score of 95.20%. This result demonstrates its potential as a useful tool for cardiologists and researchers, enabling accurate HCM diagnoses as well as the study of various cardiac conditions. We believe that our methodology will advance the field of echocardiography interpretation even beyond detecting HCM. By providing an improved method of cardiac assessment, *HCM-Echo-VAR-Ensemble* has the potential to improve clinical workflow for doctors, patient care and outcomes, and further understanding of cardiac pathologies.

Author Contributions

Abdulsalam Almadani contributed to the manuscript's conceptualization, design, methodology, data analysis, writing and editing. **Atifa Sarwar** provided thorough review and editing of the manuscript. **Emmanuel Agu** was responsible for data acquisition, conceptualization, supervision, resources, manuscript review, and project administration. **Monica Ahluwalia** and **Jacques Kpodonu** were responsible for data acquisition and review of the manuscript from a medical perspective.

Ethics and Conflict of Interest Statement

The dataset used in this research was collected with approval from the Institutional Review Board (IRB) of Boston Medical Center (IRB number H-44101, issued on 11/06/2023). Informed consent was obtained from all participants, and their data were anonymized to ensure confidentiality, in compliance with institutional and regulatory standards. The authors declare no conflicts of interest.

Acknowledgment

The authors would like to express our gratitude to all collaborators for their valuable contributions to our data collection and analyses. Additionally, the authors acknowledge the Academic & Research Computing group at Worcester Polytechnic Institute for their provision of computational resources, which supported the results presented in this paper.

References

[1] C. W. Tsao et al., "Heart disease and stroke statistics—2022 update: A report from the American Heart Association," *Circulation*, vol. 145, no. 8, pp. e153–e639, 2022.

[2] R. Wexler, T. Elton, A. Pleister, and D. Feldman, "Cardiomyopathy: An overview," *Amer. Fam. Physician*, vol. 79, no. 9, 2009, Art. no. 778.

[3] M. Glavaški, L. Velicki, and N. Vučinić, "Hypertrophic cardiomyopathy: Genetic foundations, outcomes, interconnections, and their modifiers," *Medicina*, vol. 59, no. 8, 2023, Art. no. 1424.

[4] C. Semsarian, J. Ingles, M. S. Maron, and B. J. Maron, "New perspectives on the prevalence of hypertrophic cardiomyopathy," *J. Amer. College Cardiol.*, vol. 65, no. 12, pp. 1249–1254, 2015.

[5] M. E. Arabadjian, Y. Gary, M. V. Sherrid, and V. V. Dickson, "Disease expression and outcomes in black and white adults with hypertrophic cardiomyopathy," *J. Amer. Heart Assoc.*, vol. 10, no. 17, 2021, Art. no. e019978.

[6] Z. Cheng, T. Fang, J. Huang, Y. Guo, M. Alam, and H. Qian, "Hypertrophic cardiomyopathy: From phenotype and pathogenesis to treatment," *Front. Cardiovasc. Med.*, vol. 8, 2021, Art. no. 722340.

[7] S. Ommen, R. Nishimura, and W. Edwards, "Fabry disease: A mimic for obstructive hypertrophic cardiomyopathy?," *Heart*, vol. 89, no. 8, pp. 929–930, 2003.

[8] G. Balaji, T. Subashini, and N. Chidambaram, "Detection and diagnosis of dilated cardiomyopathy and hypertrophic cardiomyopathy using image processing techniques," *Eng. Sci. Technol., Int. J.*, vol. 19, no. 4, pp. 1871–1880, 2016.

[9] J. Zhang et al., "Fully automated echocardiogram interpretation in clinical practice: Feasibility and diagnostic accuracy," *Circulation*, vol. 138, no. 16, pp. 1623–1635, 2018.

[10] A. Ghorbani et al., "Deep learning interpretation of echocardiograms," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–10, 2020.

[11] A. Almadani, E. Agu, A. Sarwar, M. Ahluwalia, and J. Kpodonu, "HCM-dynamic-echo: A framework for detecting hypertrophic cardiomyopathy (HCM) in echocardiograms," in *Proc. Int. Conf. Dig. Health*, 2023, pp. 217–226.

[12] N. Z. Farahani, D. S. B. Sundaram, M. Enayati, S. P. Arunachalam, K. Pasupathy, and A. M. Arruda-Olson, "Explanatory analysis of a machine learning model to identify hypertrophic cardiomyopathy patients from EHR using diagnostic codes," in *Proc. 2020 IEEE Int. Conf. Bio. Biomed.*, 2020, pp. 1932–1937.

[13] M. Rosca et al., "Machine learning model to predict the presence of paroxysmal atrial fibrillation in patients with hypertrophic cardiomyopathy," *Eur. Heart J.- Cardiovasc. Imag.*, vol. 24, no. Supplement_1, 2023, Art. no. jead119-344.

[14] A. Madani, J. R. Ong, A. Tibrewal, and M. R. Mofrad, "Deep echocardiography: Data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–11, 2018.

[15] D. Ouyang et al., "EchoNet-dynamic: A large new cardiac motion video data resource for medical machine learning," in *Proc. NeurIPS ML4H Workshop*, 2019, pp. 1–11.

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[17] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6450–6459.

[18] J. Lin, C. Gan, and S. Han, "TSM: Temporal shift module for efficient video understanding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 7082–7092.

[19] L. Wang et al., "Temporal segment networks: Towards good practices for deep action recognition," in *Proc. Eur. Conf. Comput. Vis.*. Springer, 2016, pp. 20–36.

[20] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "SlowFast networks for video recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6201–6210.

[21] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4724–4733.

[22] A. M. Anwar and F. J. TenCate, "Echocardiographic evaluation of hypertrophic cardiomyopathy: A review of up-to-date knowledge and practical tips," *Echocardiography*, vol. 38, no. 10, pp. 1795–1808, 2021.

[23] L. Mandeş, M. Roşca, D. Ciupercă, and B. A. Popescu, "The role of echocardiography for diagnosis and prognostic stratification in HCM," *J. Echocardiogr.*, vol. 18, pp. 137–148, 2020.