"Deep Learning Prediction of Ejection Fraction from Echocardiograms: Vector Embeddings for a Best Practice R3D Transformer"

Somin Mindy Lee
Dept. of Electrical & Computer Engineering
University of Toronto
Toronto, Ontario
somindy.lee@mail.utoronto.ca

Daniel Chung
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, United States
djechung@mit.edu

Vasu Kaker
Dept. of Mathematics
Massachusetts Institute of Technology
Cambridge, United States
vasuk@mit.edu

Yongyi Zhao
Dept. of Electrical Eng. & Computer Science
Massachusetts Institute of Technology
Cambridge, United States
yongyizh@mit.edu

Irbaz Riaz
Oncology
Mayo Clinic
Rochester, United States
riaz.irbaz@mayo.edu

Sudheesha Perera
T.H. Chan School of Public Health
Harvard University
Boston, United States
sperera@hsph.harvard.edu

Prabhu Sasankan
Internal Medicine
Beth Israel Deaconess Medical Center
Boston, United States
prabhu.sasankan.23@gmail.com

George Tang
College of Medicine
University of California, Irvine
Irvine, United States
georgehtang@gmail.com

Po-Chih Kuo
Dept. of Computer Science
National Tsing Hua University
Hsinchu City, Taiwan
kuopc@cs.nthu.edu.tw

Leo Anthony Celi
Laboratory for Computational Physiology
Massachusetts Institute of Technology
Cambridge, United States
lceli@mit.edu

***Abstract*—Ejection fraction (EF) estimation is critical to intensive care. Doctors commonly use estimated EF to create plans for patients in the ICU, and a low EF can indicate ventricular systolic dysfunction, which increases the risk of adverse events including heart failure. Heart failure affects approximately 6.2 million people in the US, appearing on 350,000 death certificates in 2018 and accounting for $30.7 billion of annual healthcare expenditures [1],[2]. There is growing interest in automatic EF estimation by models because human EF estimation has significantly higher variance [3]. Furthermore, the vector embeddings learned by such models would provide valuable mappings of echocardiograms within latent space. In this work, we repurpose an R3D transformer, a state-of-the-art deep learning model for Video Action Recognition, to classify whether patients have ventricular dysfunction or not (ejection fraction below or above 50%) using echocardiogram data. Our R3D model achieves a test AUC of 0.916 and a test accuracy of 87.5%, approaching the performance of previous comparable studies with a fraction of the training time. Emphasis is placed on sharing the vector embeddings learned by this model to advance the accessibility of deep learning tools in cardiac care. These vectors can be utilized for future echocardiogram-based clustering and classification problems with applications in intensive care, and their quality is evidenced by the strong results of the model that learned them.**

***Index Terms*—echocardiograms, ejection fraction, transformers**

I. Introduction

Heart failure is the chronic impairment of the heart's ability to effectively pump blood, markedly reducing patients' overall health and quality of life. However, a range of effective interventions, both medication-based and procedural, prolong life and reduce morbidity in cases where a heart failure exacerbation can be detected expediently [3]. Left ventricular dysfunction, caused by damage to or defectiveness of the left ventricle, raises the risk of heart failure [4]. Thus, from a clinical perspective, reliable and expedient diagnosis of ventricular dysfunction is critical to maintaining quality of care and preventing downstream sequelae of unaddressed ventricular dysfunction exacerbations.

The interpretation of imaging is a key component in the diagnosis of heart failure. In particular, the use of echocardiography, i.e. ultrasound of the heart, allows for the assessment of left ventricular ejection fraction (EF), a key metric in establishing the diagnosis and prognosis of ventricular dysfunction. EF represents the fraction of blood that exits the left ventricle during the systolic phase of the cardiac cycle.

Of particular interest are automated approaches within

echocardiography. Interpretation of echocardiography is typically performed manually by a trained clinician, but the quality of these insights are hindered by physician bias and lack of standardization. Additionally, lack of trained cardiologists in rural or low resource settings means that patients see delay of reading or lack of evaluation, leading to downstream negative health effects [2]. Therefore, a tool for automated EF prediction would significantly facilitate the early detection and treatment of ventricular dysfunction, constituting a preventative measure against heart failure [5].

A machine learning solution to this problem is challenging, however. Echocardiograms are stored as video files that suffer issues of high variability, low quality, and lack of standardization. Also challenging are the demands of videos on memory and storage [6]. Inferring cardiac function from echocardiography must also overcome the limited nature of the two-dimensional imaging modality and the complexity of cardiac physiology and kinesiology [7]. Nevertheless, others have demonstrated the efficacy of different approaches to cardiac pathologies, from video transforms to detect structural heart defects to utilizing CNN-LSTMs to predict cardiomyopathy and cardiac amyloidosis [8],[9].

In this work, we analyzed 10,030 echocardiograms from the EchoNet dataset first introduced by Ouyang et al. to predict low EF (defined as EF below 50%) utilizing a best-in-breed video transformer. EchoNet represents the largest labeled medical video dataset made available publicly to researchers and medical professionals. Our analysis corroborates the assessment of Ouyang et al. that transformer approaches effectively automate EF determination. Adding to this finding, we provide the vector embeddings for all 10,030 echocardiograms learned by our model to facilitate the development of future deep learning tools within this space. This step is necessary to democratize access to echocardiogram representations for clinical ML researchers.

Ultimately, accurate assessment of ventricular function is critical for patients with heart failure and associated conditions. Therefore, our models serve as a stepping stone towards developing robust clinical decision support tools that would aid in assessment of patients with suspected heart failure. Potential clinical applications include: (1) automated assessment of ventricular function in low resource settings where trained cardiologists are not available, and (2) expedited, actionable insight in high resource hospital settings where cardiology input is not immediately available.

II. Related Work

The EchoNet Dynamic dataset was first introduced by Ouyang et al. along with the performance of three 3D convolutional architectures for video classification used to assess EF as a numerical outcome (ie. percentage). The best of their models, an architecture based on decomposed R2+1D spatiotemporal convolutions, achieved an AUC of 0.97 when utilized for the same binary classification task of predicting whether EF was above or below 50% [10]. Others have also approached the regression problem, predicting continuous EF values, with encouraging results. Previous works by Asch et al. displayed R-squared of 0.95 on their proprietary database of roughly 50,000 echocardiograms [7]. The authors later followed this work with a Nature paper in 2020 in which they presented results from a model specifically trained to predict cases of low EF, this time displaying an AUC of 0.97 [10].

More recently in 2022, Almadani et al. utilized video action recognition (VAR) neural networks to perform binary classification of a far more complex set of echocardiograms. They reported an accuracy of 90.17% with inference time as low as 25.11 seconds using a Gate Shift Network with BNInception as its backbone [11]. Outside of binary classification of low EF, as well as numerical prediction, a number of associated models for determining other aspects of heart function based on echocardiograms have been explored in the literature. Notably, Zhang et al. (2018) were able to detect hypertrophic cardiomyopathy, cardiac amyloidosis, and pulmonary arterial hypertension with C statistics of 0.93, 0.87, and 0.85, respectively, as well as determine left ventricular mass, left ventricular diastolic volume, and left atrial volume within narrow range of machine-derived values [12].

General video transformers have also made insights beyond EF estimation from echocardiograms. For example, Jafaeezadeh et al. utilized a deep learning model with inception architecture as the backbone to 71% accuracy in the task of detecting mitral valve dysfunction, providing preliminary evidence that deep learning and transformers in echocardiographic videos can render quick, precise, and stable evaluations of various cardiac pathologies. [8]. Dai et. al. applied a novel Cyclical Self-Supervision (CSS) method for learning video-based LV segmentation, and showed that their method outperformed alternative semi-supervised methods to achieve mean absolute error (MAE) of 4.17, which is competitive with state-of-the-art supervised performance, using half the number of labels [13]. Apart from general video transformers and the convolutional neural networks presented above, CNN-LSTM models have also been applied in this context. Most recently, Hwang et al. achieved an accuracy of 92.8% on detection of LVH (i.e. hypertensive heart disease [HHD], hypertrophic cardiomyopathy [HCM], and light-chain cardiac amyloidosis [ALCA] [9].

III. Methods

A. R3D Transformer Model

The R3D model was among the best-performing approaches that Ouyang et al. (2019) used to tackle the echocardiogram classification problem [10]. Proposed by Tran et al. (2018) as an approach for video learning [14], R3D is especially apt for video data including echocardiograms because it uses 3D convolutional filters. These not only span the 2 spatial dimensions of a frame but also the temporal dimension from stacking them, allowing it to learn patterns not just spatially within frames but temporally between them. R3D in particular uses spatiotemporal kernels of shape $3 \times 3 \times 3$. 3D CNN architectures are typically shallow, but R3D combines the advantages of spatiotemporal kernels for video learning with

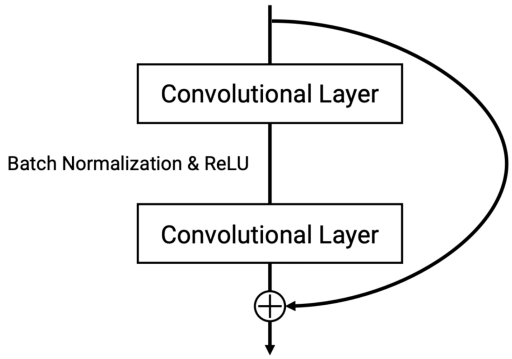

Fig. 1. Example of a shortcut ("skip") connection within a residual block.

those of deeper architectures for better overall performance. In particular, R3D builds on an 18-layer Resnet3D architecture. ResNet owes part of its performance to residual blocks that comprise its architecture: each residual block has a shortcut connection that allows the direct backpropagation of a gradient to earlier layers, which eases gradient calculations and cumulatively mitigates overfitting.

Finally, the R3D model we trained originally comes from PyTorch, and it was pre-trained on the Kinetics-400 dataset, which consists of clips of human actions [15]. Our modeling approach thus took advantage of transfer learning by initializing our R3D with weights already tuned to recognize particular actions from human video clips. In short, R3Ds use spatiotemporal kernels to capture both spatial and temporal patterns in video data, and they rely on the depth of the ResNet18 architecture. When initialized with the weights of pre-existing video transformers, R3D performance can also benefit from the "head start" of transfer learning. They are therefore powerful tools for video classification.

B. Training

Training was prohibitively expensive on local devices due to memory constraints. Initial mitigation strategies included downscaling all videos to a resolution of 56×56 instead of 112×112 and grayscaling all videos from 3 channels (RGB) to 1 (gray). Our final models, however, capitalized on cloud computing resources and were able to utilize the training data without downscaling or grayscaling. We also downsampled the videos by only using 1 of every 4 frames when training and testing our final model. Both the MIT Satori cluster and Google Cloud Platform (GCP) provided computational resources for model training. Model performance peaked at 4 epochs of training, declining from overfitting in subsequent epochs. Transfer learning likely contributed to such speed of convergence on the part of the model.

We used an Adam optimizer and achieved our best results based on validation sets using a learning rate of 1e-5 and a weight decay of 5e-4. To further manage our memory resources, we also used gradient accumulation, which involved accumulating gradients over successive mini-batches of size 20 before using them to update the model. This comes with computational costs, as there are more forward and backward passes involved with each parameter update. Nevertheless, memory was a tighter constraint, which motivated the final adoption of gradient accumulation in our modeling process. We also chose gradient accumulation as opposed to batching because the echocardiogram videos varied in frame count.

C. Classification Metrics

Our problem is a binary classification one. As such, we used binary cross entropy loss to guide the training process for our classification models. Accuracy, the ratio of true predictions to all predictions, was a necessary metric in order to compare the results of our model with those from related literature. It can become misleading, however, when the label distribution is imbalanced, which was the case for ventricular dysfunction. This motivated the use of more nuanced metrics. Precision (TP / (TP + FP)) describes the ability of a classifier to identify only relevant (positive) data points. Recall (TP / (TP + FN)) describes the ability of a classifier to identify any relevant data points to begin with. F1 score is the harmonic mean of both. Specificity (TN / (TN + FP)) conversely measures how well a classifier identifies negative data points. Given the importance of detecting ventricular dysfunction (positive class) early, we aimed primarily to maximize recall. In addition to AUC we recorded AUPRC, or the area under the precision-recall curve. This is because a small number of correct or incorrect predictions can result in large changes in the ROC curve for imbalanced data. This is not the case for the precision recall curve, making AUPRC a better holistic measure of discriminative ability for the type of imbalanced dataset at hand.

IV. Cohort

A. Cohort Selection

Our dataset consists of 10,030 echocardiograms. Ejection fraction (EF) served as our target variable, and we were interested in whether it was above or below 50%. This is because an EF below 50% warrants certain medical treatments that an EF above 50% does not. As with many health-related target variables, this caused a class imbalance program, as nearly 78% of the EchoNet echocardiograms had an EF that was greater than or equal to the 50% cutoff. Only 22% were below. The precise figures can be seen in Table 1.

The distribution of EF in our dataset was left-skewed with a mode around 60%. Specifically, the mean EF is 0.55, which is slightly lower, because there is a group of EFs between 20-50% that influence the average. Notably, there were virtually no records where EF exceeded 80% or fell below 10%. The standard deviation is 12.

We did not standardize frame length for each video, so the temporal dimension changes for each record in our dataset. Within the EchoNet dataset, echocardiograms have 173 frames on average with a standard deviation of 47 frames. This distribution appears normal. Video lengths could have been standardized, but we chose to preserve variations in frame

TABLE I
COHORT SPLIT BASED ON EJECTION FRACTION (EF)

Class	Description	Num Videos	Name
1	EF<50%	2,246	Unhealthy
0	EF≥50%	7,784	Healthy

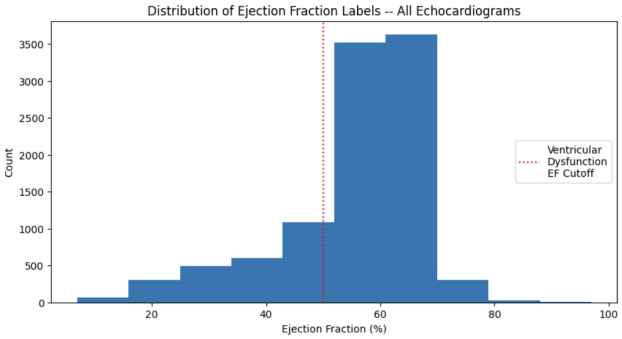

Fig. 2. Distribution of Ejection Fraction for All Echocardiograms

length for all videos in training, validation, and test sets. Keeping a variable frame length was important, as it reflects the medical reality that echocardiograms are measured in different lengths. Training a deep learning model on echocardiograms of frame length 170, for example, may not generalize well to echocardiograms of frame length 50 or 400. Non-standardized frame lengths thus functioned as a robustness feature.

Our dataset split resulted in 7,458 training videos, 1,284 validation videos, and 1,277 test videos. There was an underlying class imbalance wherein only 1,672 (22.4%) of the training videos were class 1 (EF<50%)

V. RESULTS

The performance of our models is on par with most state-of-the-art models, falling just 5 points short of the 0.97 AUC of the leading EchoNet Dynamic model [10]. This discrepancy arises from several sources. One is that each training echocardiogram is downsampled by a factor of 4, such that the processed training video contains one of every 4 frames from the original. This was done to work around limited

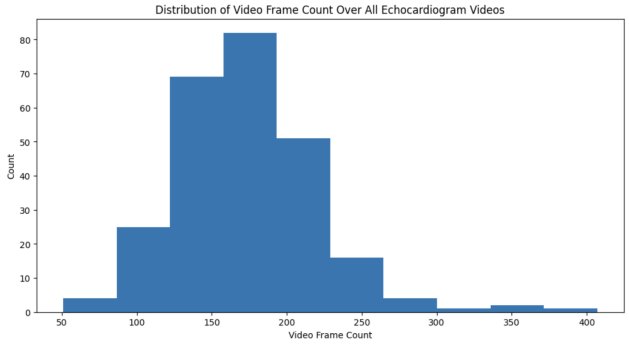

Fig. 3. Distribution of Frame Length for All Echocardiograms

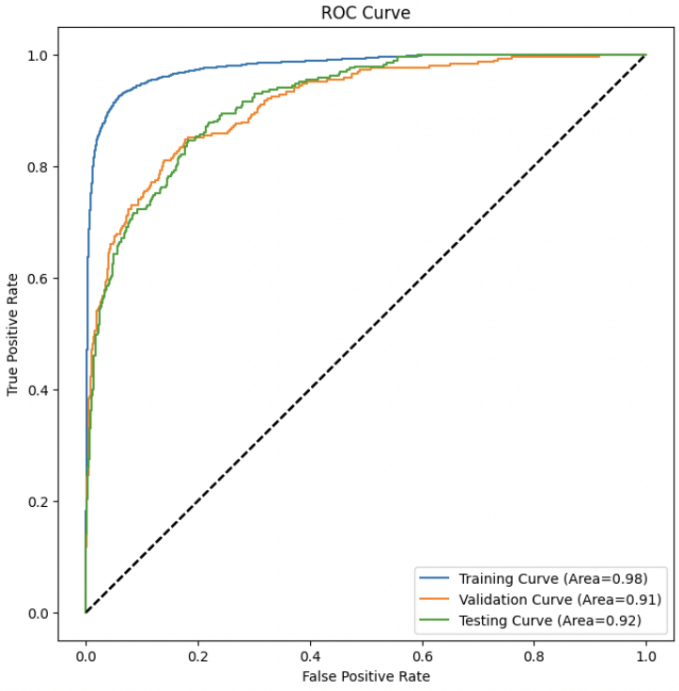

Fig. 4. AUC-ROC Curve for Train, Val, and Test Sets

TABLE II
R3D TRANSFORMER MODEL RESULTS

Model Metric	Data Split		
	Training	Validation	Testing
Accuracy	0.951	0.889	0.875
AUC	0.979	0.912	0.916
Precision	0.924	0.809	0.789
Recall	0.85	0.661	0.604
Specificity	0.98	0.955	0.954
F1 Score	0.885	0.728	0.684
AUPRC	0.949	0.811	0.798
BCE Loss	0.151	0.306	0.31

computational resources, but it also reduced the volume of the data that our R3D model was trained on, thus weakening performance. Furthermore, we did not augment the training data of our model as was done in the EchoNet Dynamic model, again for computational reasons. Data augmentations such as brightening and jittering the training videos have been shown to strengthen the performance of echocardiogram-based classifiers, as they increase the volume of and variation within the training data [11].

The AUPRC of our R3D was also strong. The AUPRC of a random classifier is theoretically 0.5, and for a perfect classifier it should be 1.0. The R3D test score was 0.798. Given how AUC can be misleading in contexts of class imbalance, such a high AUPRC value suggests it is genuinely the strength of the R3D classifier itself that explains the high AUC.

Model specificity was especially strong at 0.954 for the test set. This means the classifier makes negative predictions correctly almost every time. Test recall was lower at 0.604,

TABLE III
COMPARISON OF TEST ACCURACY AND AUC WITH EARLIER CLASSIFIER BY ALMADANI ET AL. (2022)

Model	Classification Approach	
Metric	*Our Approach*[a]	*Almadani et al.*[b]
Accuracy	87.5%	90.17%
AUC	0.916	0.847
EF Classification Threshold	50%	50%

[a]R3D transformer with ResNet18 architecture, transfer learning
[b]GSM with inception backbone, 32-frame echocardiograms

but this still means that the model predicts the majority of true positives (cases of ventricular dysfunction) correctly. Test precision, finally, was also high at 0.789. This means that a comfortable majority of positive predictions made by the R3D model were in fact true cases of ventricular dysfunction. Together, these metrics demonstrate the strong discriminative ability of this classifier to predict EF being above or below 50%. The vector embeddings of this model, therefore, are reflective of an effective learned representation of echocardiogram data.

VI. DISCUSSION

A. Comparison to State-of-the-Art Echocardiography Models

In 2022, Almadani et al. reported their utilization of video action recognition (VAR) neural networks to perform binary classification of echocardiograms [11]. A comparison of the AUC and accuracy of our R3D Transformer to that of the VAR model used by Almadani et al. is visible in Table III.

"Deep Video Action Recognition Models for Assessing Cardiac Function from Echocardiograms" achieved its highest accuracy of 90.17% with 32 frames on the GSM with BNInception. Its highest AUC of 0.847 was achieved with 16 frames on the GSM with InceptionV3. Our R3D model achieved a comparable accuracy of 87.5% and an even higher AUC of 0.916, demonstrating that it has at least the holistic discriminative ability of state-of-the-art ventricular dysfunction classifiers and has competitive accuracy. This is important to demonstrate, as it means that the vector embeddings learned by our model are of a quality consistent with the best classifiers in the current literature.

B. Additions for future development

Gradient weighted class activation mapping (Grad-CAM) is a method of finding which regions of an image play the biggest role in determining the prediction made by a neural network. Grad-CAM allows visualization of CNNs by generating a heatmap, produced by analyzing each feature map using gradients. Future work will use Grad-CAM to understand why the R3D model makes decisions, especially incorrect ones, as these problematic echocardiograms may be clustered together in latent space according to their embeddings.

C. Applications of Vector Embeddings

Beyond understanding the model's decision processes, the vector embeddings obtained from our R3D model open up a realm of possibilities of diverse applications in cardiac care. These embeddings encapsulate rich information about echocardiograms in a latent space enabling various downstream tasks and analyses. We describe a few of these applications here.

Echocardiogram-based Clustering: Clustering analyses can use the vector embeddings of echocardiograms to potentially reveal inherent patterns and subtypes within cardiac data that represent diverse presentations of ventricular dysfunction.

Automated Disease Classification: The learned vector embeddings can be used for training classifiers to automatically identify specific cardiac conditions beyond EF. This is because a lower EF can be caused by other conditions like coronary artery disease or systolic heart failure [16]. This extends the model's utility to a broader spectrum of cardiac pathologies.

Patient Risk Stratification: By incorporating vector embeddings, new risk prediction models could assist in stratifying patients based on their risk of developing ventricular dysfunction or related complications, enabling targeted interventions and personalized treatment plans.

Integration into Clinical Decision Support Systems: Vector embeddings can provide healthcare professionals with additional contextual information for interpreting echocardiograms. The collaborative approach in integrating these into either models and systems combines the strengths of deep learning with the expertise of clinicians.

These applications extend beyond model interpretability, contributing to advancements in automated diagnosis, risk prediction, and personalized treatment strategies.

D. Clinical Implications

On multiple metrics, the R3D transformer-based approach used in this study effectively predicts low EF. This means it can enhance shared decision-making which is a crucial element in patient-centric care, and it demonstrates that deep learning approaches can prove successful in clinical prediction tasks for ventricular dysfunction.

However, clinical decision making for a ventricularly dysfunctional patient is a complicated process and usually requires contextualization of EF. For example, implantable cardioverter-defibrillators (ICDs) are considered for primary prevention in patients at high risk for sudden cardiac death; an EF$\leq$35% is a clinically accepted predictor for elevated risk and strong benefit for ICD implant. may be considered in patients with clinical heart failure with an EF$\leq$35% for primary prevention of sudden death. Hence accurate and timely identification of ICD candidates is important in improving patient outcomes. For patients who might be at increased risk of ICD, insertion is detrimental for patient outcomes. While our R3D model demonstrated comparable performance in predicting EF$<$50% relative to other counterparts, its performance may plausibly be generalizable to predict EF$\leq$35% which could potentially facilitate the process of treatment selection in these patients. Likewise, the core architecture used in this study can be extrapolated to predict improvement of left ventricular EF following initiation of medical therapy to decide more rapidly

on primary device placement in those unlikely to show an EF increase to $\leq$35%.

Furthermore, while deep learning models may provide valuable insights and predictions, they are best utilized as decision support tools rather than standalone diagnostic tools. Contextualization of EF predictions by a cardiologist is crucial for comprehensive clinical decision-making in a heart failure patient. An EF prediction alone may not capture the entire clinical reality. Additional factors such as the patient's symptoms, prior response to medical therapy, and individualized treatment goals must be considered in real world clinical practice. Deep learning models can be integrated with the clinical expertise of cardiologists, by considering their predictions within the broader context of the patient's unique circumstances. Therefore these models have the potential to guide a more targeted approach to treatment selection addressing care gaps for high-risk individuals, and potentially reducing the need for costly and unnecessary procedures.

VII. Conclusion

In this paper, we adapt, train, and deploy an R3D deep video action recognition network for the objective of classifying whether or not a heart is healthy based on its EF. We achieve high AUC and accuracy scores with this model, on par with the performance of current best-practice models. Future work will aim to increase the training set size, robustify the model using data augmentation techniques, and interpret model decisions using Grad-CAM. At present, these strong results serve to legitimize the quality of the vector embeddings published in this work. These vector embeddings describe each echocardiogram in the EchoNet dataset in 400-dimensional latent space, constituting valuable lookup points for future work on echocardiogram cohort selection in a variety of ML applications within cardiology.

VIII. Code

The vector embeddings found by this paper can be accessed through the following GitHub repository: https://github.com/Team-Echo-MIT/r3d_v0_embeddings

IX. Acknowledgements

We extend a word of gratitude to our mentors, Prabhu Sasankhan, Po-Chih Kuo, George Tang, Jacques Kponopu, and Brigitte Kazzi for their continued guidance and support on clinical relevance, problem selection, model development, and training methodologies throughout the project.

References

[1] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. Circulation. 2020;141(9):e139-596.

[2] Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. Circulation. 2019;139(10):e56–528.

[3] Hunt SA, Abraham WT, Chin MH. 2009 Focused Update Incorporated Into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the International Society for Heart and Lung Transplantation. Circulation 2009;119:e391–479.

[4] Cleland JGF, Torabi A, Khan NKEpidemiology and management of heart failure and left ventricular systolic dysfunction in the aftermath of a myocardial infarctionHeart 2005;91:ii7-ii13.

[5] Ouyang, David, Bryan He, Amirata Ghorbani, Matthew P. Lungren, Euan A. Ashley, David H. Liang and James Y. Zou. "EchoNet-Dynamic: a Large New Cardiac Motion Video Data Resource for Medical Machine Learning." (2019).

[6] Xie, S., Sun, C., Huang, J., Tu, Z., Murphy, K. (2018). Rethinking Spatiotemporal Feature Learning: Speed-Accuracy Trade-offs in Video Classification. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11219. Springer, Cham. https://doi.org/10.1007/978-3-030-01267-0_19.

[7] Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, Romano N, Hong H, Mor-Avi V, Martin RP, Lang RM. Automated Echocardiographic Quantification of Left Ventricular Ejection Fraction Without Volume Measurements Using a Machine Learning Algorithm Mimicking a Human Expert. Circ Cardiovasc Imaging. 2019 Sep;12(9):e009303. doi: 10.1161/CIRCIMAGING.119.009303. Epub 2019 Sep 16. PMID: 31522550; PMCID: PMC7099856.

[8] Vafaeezadeh, M, Behnam, H, Hosseinsabet, A, Gifani, P. CarpNet: Transformer for mitral valve disease classification in echocardiographic videos. Int J Imaging Syst Technol. 2023; 1- 10. doi:10.1002/ima.22885.

[9] Hwang, IC., Choi, D., Choi, YJ. et al. Differential diagnosis of common etiologies of left ventricular hypertrophy using a hybrid CNN-LSTM model. Sci Rep 12, 20998 (2022). https://doi.org/10.1038/s41598-022-25467-w.

[10] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020 Apr;580(7802):252-256. doi: 10.1038/s41586-020-2145-8. Epub 2020 Mar 25. PMID: 32269341; PMCID: PMC8979576.

[11] A. Almadani, A. Shivdeo, E. Agu and J. Kpodonu, "Deep Video Action Recognition Models for Assessing Cardiac Function from Echocardiograms," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 5189-5199, doi: 10.1109/BigData55660.2022.10020947.

[12] Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018 Oct 16;138(16):1623-1635. doi: 10.1161/CIRCULATIONAHA.118.034338. PMID: 30354459; PMCID: PMC6200386.

[13] Dai, Angela, et al. "ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018. Crossref, https://doi.org/10.1109/cvpr.2018.00481.

[14] Tran, Du, et al. "A Closer Look at Spatiotemporal Convolutions for Action Recognition." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018. Crossref, https://doi.org/10.1109/cvpr.2018.00675.

[15] Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., ... & Zisserman, A. (2017). The kinetics human action video dataset. arXiv preprint arXiv:1705.06950.

[16] Vicent L, Álvarez-García J, Vazquez-Garcia R, González-Juanatey JR, Rivera M, Segovia J, Pascual-Figal D, Bover R, Worner F, Fernández-Avilés F, Ariza-Sole A, Martínez-Sellés M. Coronary Artery Disease and Prognosis of Heart Failure with Reduced Ejection Fraction. J Clin Med. 2023 Apr 21;12(8):3028. doi: 10.3390/jcm12083028. PMID: 37109365; PMCID: PMC10143946.