

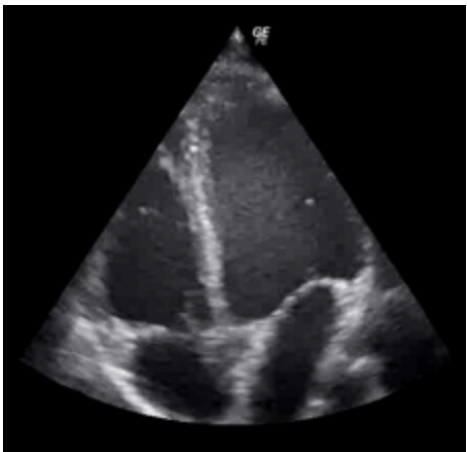


Abstract

Computer-assisted identification of end-diastolic (ED) and end-systolic (ES) heart phases within echocardiography (echo) scan enables a level of automation, a process previously only achievable by qualified medical staff members. This research focuses on the possible improvements using variations of neural network architectures, primarily focusing on the temporal aspect. A CNN and RNN model architecture usability applies to heart-phase detection, where expertly annotated ED and ES echo scans are the ground truth for model training. Results show that variates tested have the ability to outperform some existing solutions, falling behind architectures with more comprehensive spatial capturing characteristics. Additionally, performance analysis between variates suggests networks with bidirectional architectures increase computation time significantly, compared to their counterparts, while providing sub-optimal results.

Introduction

A key driver for this research is the investigation and potentially proposition of alternative neural network architectures to aid the accuracy and performance of self-driven heart assessments. Within the analysis stage of the self-driven heart exam, ML can assist with phase detection.



A standalone echocardiogram (echo) can help identify damage from a heart attack, heart failure, congenital heart disease, problems with the heart valves, and cardiomyopathy [1]. The previously mentioned automated heart assessment focuses on heart phase detection in an echo scan. The subject of this research is the find the time step of end-diastolic (ED) and end-systolic (ES) heart phases within an echo scan. The analysis of ED and ES heart phases could provide information on ejection fraction (EF). EF references the amount of blood pumped from the left ventricle per heart contraction [2] and is one of the various aspects that identifying ED and ES heart phases can help with heart assessment.

The aspect of no standardised dataset prompts the issue in the comparison of performance. While studies focus to provide metric for comparison (e.g., aaFD), differing data for each of the tested models trigger the issue of unfair testing as more than one variables differ. In the case of phase detection, various models are subject to comparison, yet they are tested on varying datasets.

Aims and Objectives

Core to this research is the analysis of various neural network structure for phase detection. Relevant studies have managed state-of-the-art results but using separate datasets, and different network structure, there is a difficult identifying the beneficial aspects for phase detection networks.

To combat the variability within the phase detection space, this research aims to:

1. Implement an artefact involving sequential phase detection.
2. Using transfer learning, utilise a common and readably available CNN for extracting spatial information.
3. Create several instances of the artefact (named variates).
4. Per variate, adjust the temporal network structure.
5. Using the same subsets, train, validate, and test each variate.
6. Find the average absolute frame difference to enable direct comparison between variates.

References

[1] - NHS, "Echocardiogram," 23 July 2018. [Online]. Available: <https://www.nhs.uk/conditions/echocardiogram/>. [Accessed 31 May 2021].

[2] - heart.org, "Ejection Fraction Heart Failure Measurement," 31 May 2017. [Online]. A available: <https://www.heart.org/en/health-topics/heart-failure/diagnosing-heart-failure/ejection-fraction-heart-failure-measurement>. [Accessed 31 May 2021].

[3] - B. Kong, Y. Zhan, M. Shin, T. Denny and S. Zhang, "Recognizing End-Diastole and End-Systole Frames via Deep Temporal Regression Network," Lecture Notes in Computer Science, vol. 9902, 2016.

[4] - F. T. Dezaki, N. Dhungel, A. H. Abdi, C. Luong, T. Tsang, J. Jue, K. Gin, D. Hawley, R. Rohling and P. Abolmaesumi, "Deep Residual Recurrent Neural Networks for Characterisation of Cardiac Cycle Phase from Echocardiograms," Lecture Notes in Computer Science, vol. 10553, 2017.

[5] - A. M. Fiorito, A. Østvik, E. Smistad, S. Leclerc, O. Bernard and L. Lovstakken, "Detection of Cardiac Events in Echocardiography Using 3D Convolutional Recurrent Neural Networks," 2018 IEEE International Ultrasonics Symposium (IUS), 2018.

[6] - American Heart Association, "Size, strength of heart's right side differs by age, gender, race/ethnicity," Science Daily, 9 June 2011. [Online]. Available: <https://www.sciencedaily.com/releases/2011/06/110606161024.htm>. [Accessed 31 May 2021].

Methodology

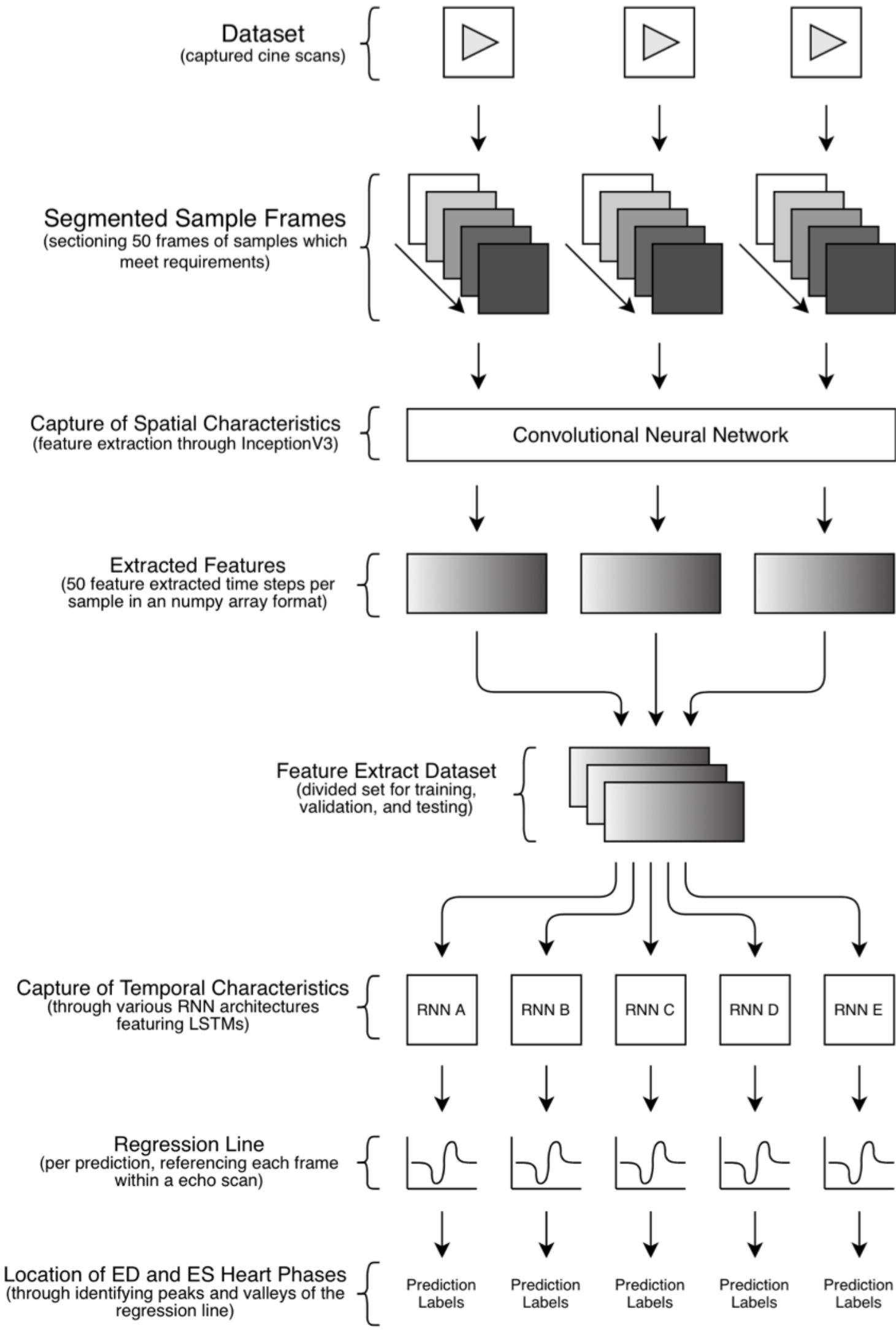
To ready the network for real-world usage, a subsection of the dataset will be used for training. The model will determine what features to pay attention to in the cine scan by knowing which are through available expert annotations.

Using transfer learning for capturing spatial characteristics, the CNN used in this research is the InceptionV3. RNNs handle the temporal network section which depends on the variate:

- A. 1x LSTM
- B. 2x LSTM
- C. Bidirectional LSTM + LSTM
- D. 2x Bidirectional LSTM
- E. 3x LSTM

The model formation consists of a regression-based approach which establishes a gradual relation between the ED and ES phases, recognising the structure of the data. The model benefits from gradual changes in heart phase labels, potentially recognising the dependencies before, during, or after a phase.

To calculate the error of the model during training and otherwise, Keras provides various loss functions. MSE and MAE is captured throughout the training and validation process, however the aaFD is the important metric which the custom testing process will calculate on the test dataset.



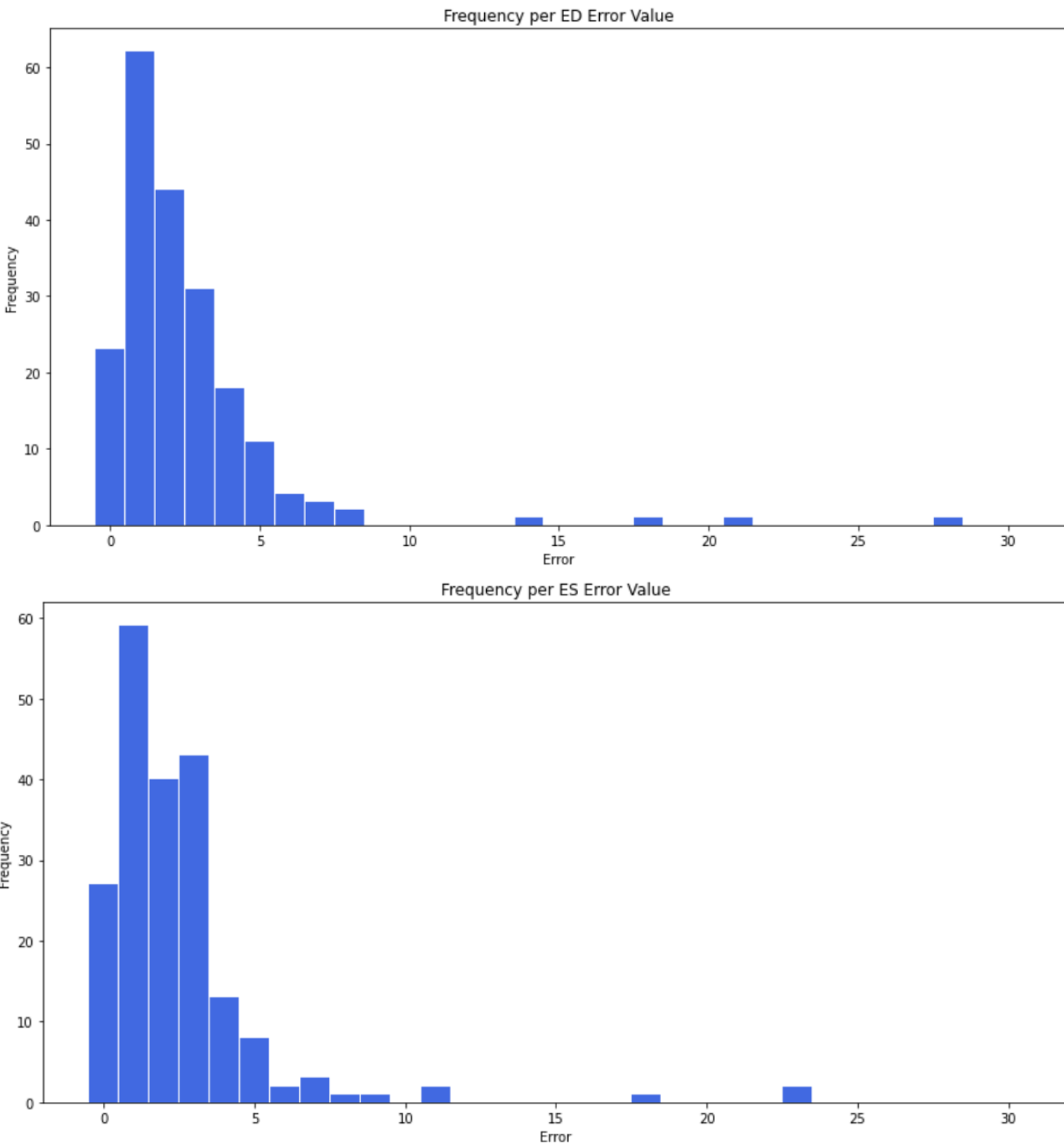
Results

Reviewing test results on a comprehensive test at the end of training (50 epochs) may not be adequate. Over-fitting is a critical problem plaguing machine learning models. The model begins to confine in the memorisation of the data instead of providing more general prediction with higher accuracy for unseen data. Instead, each variate is evaluated on their accuracy on the epoch that achieved the lowest validation error. Variate E performs the best from all variates in 22 epochs and represents the absolute best possible heart phase detection model version of all tested instances, as seen from table seen on the right hand side.

Variate	Error Value	Epoch
A	0.01736	11
B	0.01292	26
C	0.01289	20
D	0.01359	18
E	0.01270	22

Variate	Mean-Squared Error		Mean-Absolute Error (aaFD)	
	ED	ES	ED	ES
E (22 nd epoch)	16.05940	14.86138	2.53465	2.45544

LHS figures show aaFD ED and ES error frequency, and the information appears to be in line with 2.5 and 2.4 mae values for ED and ES, respectively. The frequency of ED and ES errors falls largely within 5, with very few entries for greater values. The model seems to be attempting to learn relevant patterns. Additionally, even though the large error values are infrequent, there is a hypothesis for their occurrence. Error values of 20 or higher could be predicting the phases heart beats before or after the EchoNet annotation. Regarding efficiency, there is a correlation evident surrounding the trainable model parameters, and the computation time. It is observed that variates with bidirectional LSTM increase computation time considerably.



Conclusions

Table located on the left contains accuracy values for various fixed-length phase detection implementations. Variate E at the 22nd epoch outperforms some ResNet and CNN phase detection implementations. However, the more advanced feature extraction on a three-dimensional CNN network achieves lower aaFD. When viewing the predicted results on a graph visible on the left, there is visible evidence that the model is learning and predicting different values depending on the input which results in an relative low average aaFD value 2.45. Despite promising results, not being able to validate the dataset causes uncertainty in its application in patients from different ages and ethnic backgrounds, who may exhibit slight differences in the structure of their heart, and more [6]. Citizens of these areas may not be properly represented in the datasets, a complication that could stem into abnormal predictions potentially causing false positives or true negatives in diagnosis. Most datasets mentioned in this study originates from university hospitals in extensively modernised societies. When considering implementation of an accurate model, there should be, at least, considerations of the origin of training dataset, and the intended destination. Preferably, all models should train on diverse enough datasets to account for those who immigrate from less represented countries.

Method	Mean-Absolute Error (aaFD)	
	ED	ES
CNN + LSTM + L_2 loss [3]	6.3	7.3
CNN + LSTM + L_2 loss + structured loss [3]	6.4	7.3
ResNet + LSTM + L_2 loss [4]	4.4	4.7
ResNet + LSTM + L_2 loss + structured loss [4]	3.7	4.1
InceptionV3 + LSTM + mse loss + Adam optimiser (variate E at 22 epoch)	2.5	2.4
3D CNN + LSTM + cross-entropy loss [5]	1.6	1.7

