

Unsupervised Discovery of Echocardiographic Views for Rheumatic Heart Disease Diagnosis

Yuri NJATHI¹, Lorna MUGAMBI², Liesl ZÜHLKE^{3,4,5}, Ciira wa MAINA⁶
^{1,2,6}*Centre for Data Science and Artificial Intelligence (DSAIL), Dedan Kimathi University of Technology, Private Bag-10143, Nyeri, Kenya*
Email: ¹yuri.njathi@interns.dkut.ac.ke, ²lorna.mugambi@dkut.ac.ke, ⁶ciira.maina@dkut.ac.ke
³*Division of Paediatric Cardiology, Department of Paediatrics and Child Health, Red Cross War Memorials Children's Hospital, Cape Town, South Africa*
⁴*Division of Cardiology, Department of Medicine, Groote Schuur Hospital, Cape Town, South Africa*
⁵*South African Medical Research Council, Francie Van Zyl Drive, Tygerberg Cape Town*
Email: ^{3,4,5}liesl.zuhlke@mrc.ac.za

Abstract: RHD is a cardiovascular disease that causes damage to the heart valves. If the damage is severe it is rectified using expensive valve replacement surgery. Early diagnosis of the disease allows for cost-friendly preventive measures. Specific views of the heart are required for proper assessment by heart specialists. Since routine screening is recommended for the rapid early identification of RHD, a large amount of patient data is generated. To handle this influx of data, trained AI is being used to automate view classification, unfortunately, the high cost of obtaining expert-labelled data in terms of time and money is prohibitive. Thus, we explore how the use of unsupervised AI methods can aid experts in the faster labelling of the data and what patterns PCA and agglomerative clustering identify in echo videos. We found that after appropriate preprocessing, these unsupervised methods can group videos with similar echocardiographic views. We also found that these methods were sensitive to the specific machines used to acquire the data and therefore care should be taken when applying them to data collected using different machines.

Keywords: Acute Rheumatic Fever, Echocardiograms, Principal Component Analysis, Rheumatic Heart Disease, Unsupervised Machine Learning

1. Introduction

Rheumatic Heart Disease (RHD) is a cardiovascular disease that affects the heart's valves, particularly the mitral and aortic valves. It is the only long-term consequence of acute rheumatic fever (ARF), which is caused by prolonged recurrent attacks of streptococcal infection. Curing RHD in its asymptomatic stage is more attractive than treatment in its symptomatic stage due to the availability of cost-friendly secondary prophylaxis in the form of penicillin injections every 3 to 4 weeks. It has been proven that this secondary prophylaxis prevents ARF thus preventing RHD for 5 to 10 years. Unfortunately, the majority of registered patients are symptomatic with advanced disease requiring heart valve surgery, a more expensive and complex treatment [1].

Guidelines by a group of experts from the World Heart Federation described diagnostic criteria for subclinical RHD [1]. Previously, listening for heart murmurs (auscultation) with a stethoscope was used to detect RHD. Echocardiography is more sensitive and specific at detecting early-stage RHD than auscultation. This feature of the echocardiographic criterion

allows for early detection of previously undiagnosed RHD. With the development of cheap and portable echocardiographic screening machines, the stage is set for technology to be used by the same professionals to increase early diagnosis rates.

Echocardiography has become routinely used in the diagnosis, management and follow-up of patients with any suspected or known heart conditions because it provides a lot of information about the heart. Multiple views of the heart are obtained to show different parts of the heart's internal structure. For the efficiency of the RHD data acquisition process, it is important to collect imagery in the views with the highest rate of RHD diagnosis [2]. During echocardiographic examinations, the imagery can be taken looking at different parts of the heart, which are called views. These views include the Parasternal Long-Axis view (PLAX) which allows for the measurement of the mitral valve leaflets midpoints and tips [3][4], the Parasternal Short-Axis view (PSAX) which allows for the measurement of the right coronary cusp and non-coronary cusp and Apical four-chamber (A4C) which allows for measurement of the anterior mitral valve leaflet midpoint. The PLAX view is obtained by placing the sonographic probe next to the sternum showing both the aortic and mitral valves. The PSAX view is obtained from the PLAX view by rotating the probe at 70°-110° clockwise obtaining a view of the aortic valve, mitral valve, papillary and apical left ventricular levels. The A4C view is obtained with the patient lying on their left side, left lateral position and the probe is placed on the heart apex. These many views can be distinguished using AI-based view classifiers. This problem has attracted the attention of several researchers [3][5][6][7][8][9].

Since view classification is important to diagnosing RHD, professionally labelled images and videos after sonography exams are obtained according to view and disease. The end goal of this would be to design a system that correctly groups echocardiographic data according to the morphological view. Well-trained sonographers and cardiologists can label the echocardiographic data we have using a data annotation application that we designed. Unfortunately, there was a high associated labelling cost with this process and the amount of unlabelled data we obtain is more than the rate at which it can be annotated reasonably. We explored unsupervised learning, which does not require a label to learn patterns within data.

We believe that since each image in a particular view shares a similar shape with other views of that particular shape and is dissimilar with other views, we can compute the level of similarity amongst the files, having accounted for the patient, colour, machine configuration and machine setting details. We can find a relationship between the data and view without the labelled views. The machine configuration is important as it is even set out in the WHF guidelines [1].

In this paper, we look at unsupervised data analysis processes to group echocardiographic videos according to view without any prior knowledge of the view. These are exploratory steps that could potentially reduce the workload of specialists. The parasternal short axis (PSAX), aortic four-chamber (A4C) and parasternal long axis (PLAX) are views of interest.

2. Objectives

2.1 Main Objective

Explore echocardiographic data using unsupervised methods to find what relationship will be found between the imagery in the echocardiographic data and the echocardiographic views.

2.2 Specific Objectives

1. Use PCA to perform dimensionality reduction on the video dataset.

2. Use agglomerative clustering and euclidean distance to group similar videos.
3. Compare dendrogram results using visual methods

3. Methodology and Technology Description

To obtain a pattern indicative of the echocardiographic view from the unlabelled dataset we explored dimensionality reduction and data clustering methods. Finding patterns from unsupervised learning will provide an avenue for increasing the labelling efficiency of echocardiographic data annotation. Principal component analysis and agglomerative clustering methods were explored. Image and video pre-processing techniques were also used.

3.1 Preprocessing methods

Before performing dimensionality reduction and clustering, it was necessary to preprocess the data by converting each video to its constituent frames using OpenCV. These frames were grayscaled, cropped, removing patient information and focusing on the area of interest, and binary thresholded as a normalisation technique. Binary thresholding reduced the noise in the dataset while maintaining the morphological features in each frame. Figure 1 below shows the preprocessing steps and their outcome. The critical morphological features are seen from visual inspection.

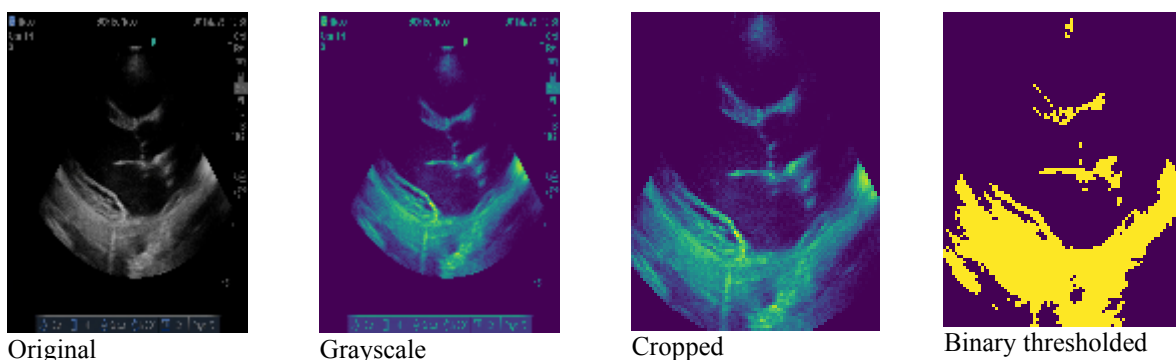


Figure 1: Preprocessing steps on a PLAX view ultrasound ; steps performed left to right

3.2 Data

The data was used in three sets. The first set contained 15 videos that had all come from one source and all with labels. The second set consisted of 77 labelled and unlabelled videos, all from a single source. 21 videos were discarded (27% of the dataset) due to having visual peculiarities such as the one seen in Figure 2, causing obstructions of the valves for currently unknown reasons. The third set contained 848 videos from multiple different sources, none of these videos was discarded. Each video contained only one view. The videos were also renamed for anonymity purposes. The importance of the first set was to check whether a pattern was to be found and to confirm whether or not there was any relationship between the clustering of each video and its echocardiographic view. We then tested whether the pattern found in the first set would extend to a larger set of data, in the second set. All labels in the data were obtained from a Paediatric Cardiologist.



Figure 2: Peculiar view

3.3 Principal Component Analysis (PCA)

PCA is a technique that is widely used for dimensionality reduction, lossy data compression, feature extraction and data exploration [10][11]. Working with high-dimensional data such as images comes with difficulties in analysis, interpretation and visualization because it contains many unnecessary dimensions which are often correlated. The data possesses an intrinsic lower-dimensional structure which PCA can exploit to reduce the size of the representation [10][12]. To ensure that the outlier frames in each video are compensated for, we used the median operation to obtain the final video PCA values from PCA values for each video frame.

3.4 Agglomerative Hierarchical Clustering

Agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. The algorithm treats each object as a single cluster and computes dissimilarity between each pair of objects in the dataset. Agglomerative clustering works by grouping objects into a hierarchical cluster tree, based on the distance generated by a linkage function [13]. The distance is measured between two and more data points. Options for linkage are Complete, Single, Average, Centroid and Ward. Here, Ward's linkage was used. Ward's linkage is a method which uses Ward's linkage criterion to determine the similarity between clusters and iteratively merges them until the desired number of clusters is reached. Ward's linkage criterion is based on the sum of the squared Euclidean distances between the points in the two clusters.

4. Results

The procedure documented in the methodology was performed on the three sets of data. The results from the first set further guided the unsupervised exploration of the second and third datasets. The first dataset contained 3 A4C videos, 4 PLAX videos and 8 PSAX videos. The 1st dataset had class imbalance but the method we used circumvented this issue. The videos were clustered into 3 groups. The first with five videos, 3 A4C and 2 PSAX, the second with six videos, all of the PSAX view and the third with 4 videos all of the PLAX view seen in Figure 3. The dataset was clustered according to their 2 component PCA values as can be seen from the dendrograms in Figure 4.

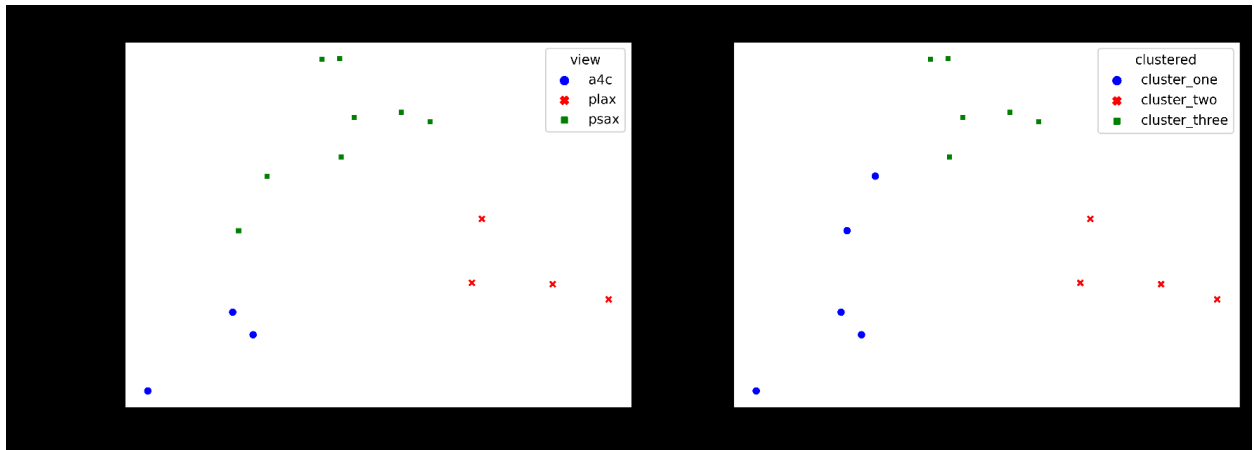


Figure 3: Clustering results on 15 videos all with known view on the left after agglomerative clustering on the right, 3 clusters emerged

The clusters in Figures 3 and 4 are based on the agglomerative hierarchical clustering algorithm. The dendrogram and similarity measures were later compared to give some understanding as to why the results were being obtained. They show how the algorithm worked out allocating videos to clusters. The shorter the height, the more similar the videos are.

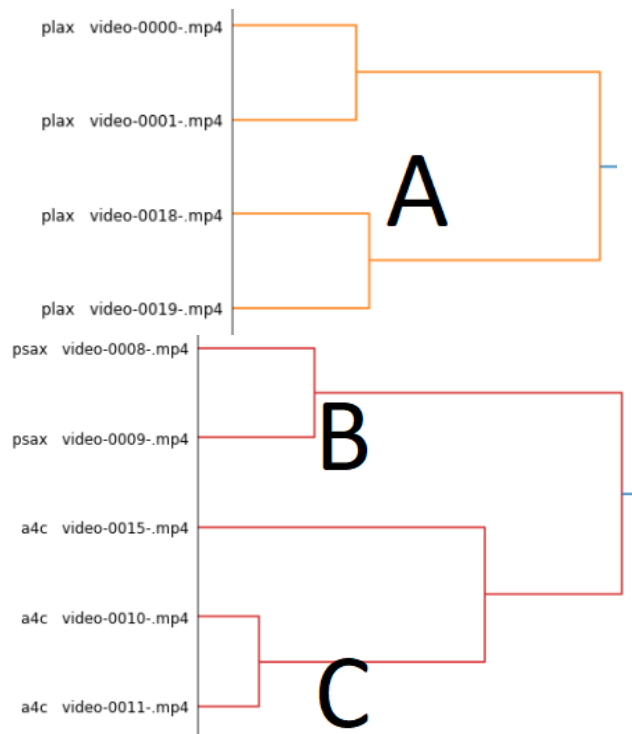
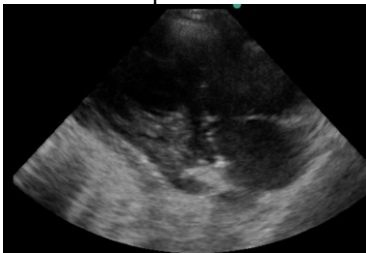
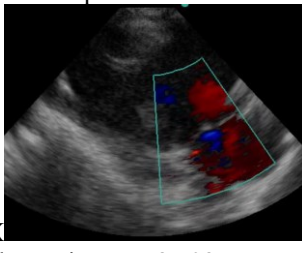
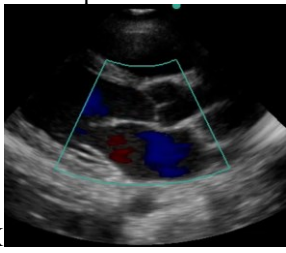
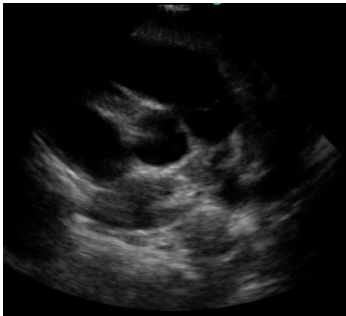
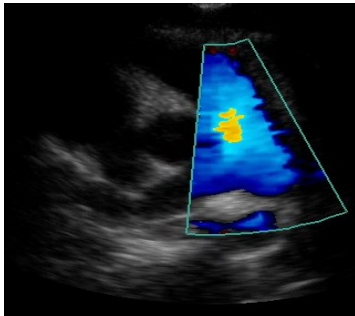
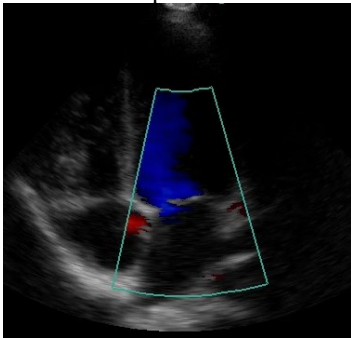
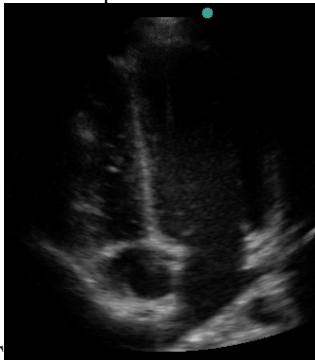
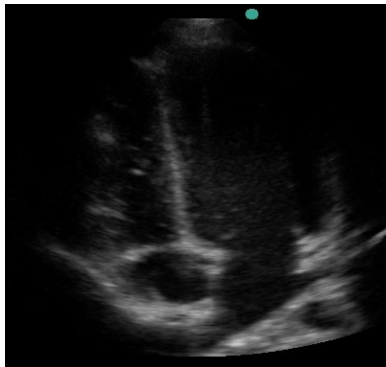


Figure 4: Some sections of the dendrogram of the above-clustered videos (Sectioned to concentrate on per video comparison)table

The similarity in Figure 4 is checked against their euclidean distances in Table 1. Of all the files in section A, Figure 4, all were PLAX, the dendrogram computed video-0000 and video-0001 as most similar and video-0018 and video-0019 were also similar, both clusters were found to be similar but with higher linkage. The high linkage can be explained by the euclidean distances in Table 1. For sections B and C, the dendrogram computed video-0008 and video-0009 as similar, which held true as both were in the PSAX view, video-0010, video-0011 and video-0015 were all A4C were computed as similar, there is a high linkage between section B and C as is expected of different views, further analysis of that is seen in Table 1. From Table 1, sections A, B and C, visually it was seen that the videos look

similar, the blue-red colour may have affected the clustering even after grayscaling, controlling for this would allow for a clearer comparison.

Table 1: Comparison of videos, euclidean distances based on dendrograms

section	video 1	video 2	video 3
A	video-0000.mp4 PLAX  Euclidean Distance : 0 (same)	video-0001.mp4  PLAX Euclidean Distance: 8.508 (similar)	video-0018.mp4  PLAX Euclidean Distance: 66.01 (similar)
B	video-0009.mp4 PSAX  Euclidean Distance : 0 (same)	video-0008.mp4 PSAX  Euclidean Distance: 8.595 (similar)	Not available
C	video-0010.mp4 A4C  Euclidean Distance : 0 (same)	video-0015.mp4  A4C Euclidean Distance: 32.97 (similar)	video-0011.mp4 A4C  Euclidean Distance: 28.40 (similar)

For the second dataset, a 56-leaf dendrogram was generated, sections of it shown in Figure 6 as well as scatterplots of agglomerative clusters seen in Figure 5. In Figure 5, on the left is a plot of the two principal components according to their expert-labelled view and on the right are the same principal components agglomerative-ly clustered.

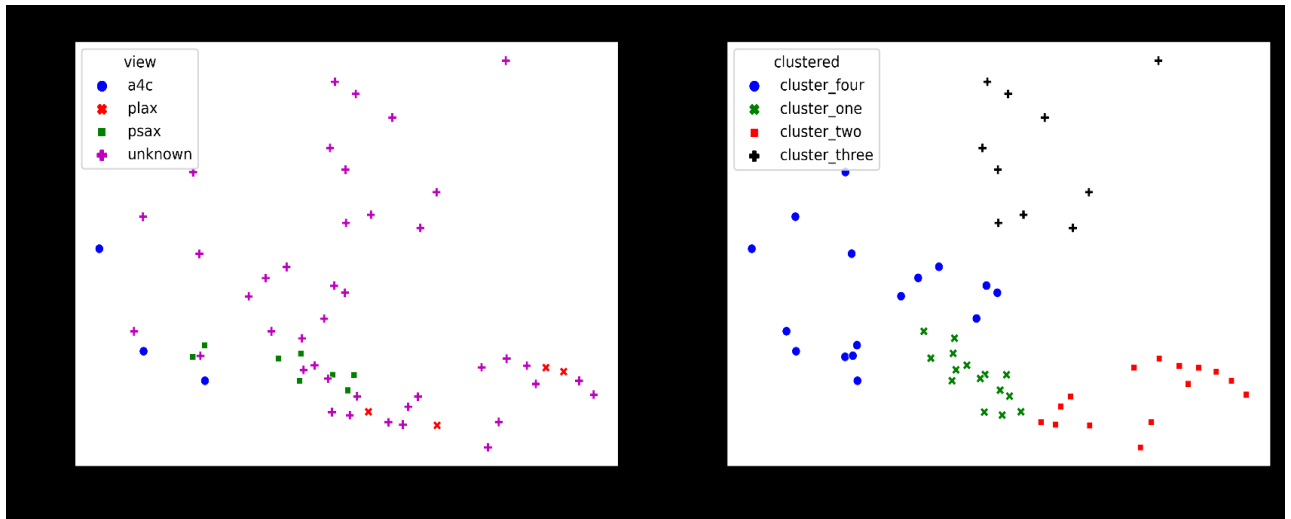


Figure 5: Clustering results on 56 videos some (15) with known and unknown view on the left and agglomeratively clustered views on the right

In Figure 5, it was seen that many of the A4C, PLAX and PSAX points on the left were clustered into cluster four, cluster two and cluster one on the right respectively. Cluster three points were all generated from unknown points. Next was a comparison of the dendrogram obtained for a sample of the videos and their similarity measures. For the second set, we also computed the euclidean similarity measure among the videos to determine why the dendrogram trees in Figure 6 were obtained, these measures are seen in Table 2.

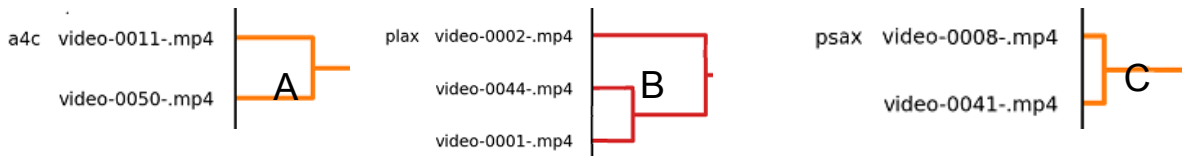

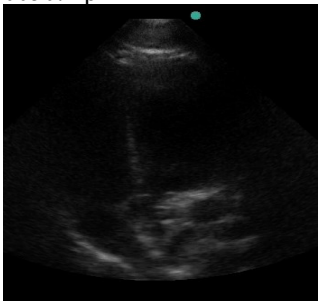
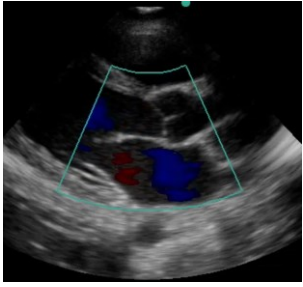
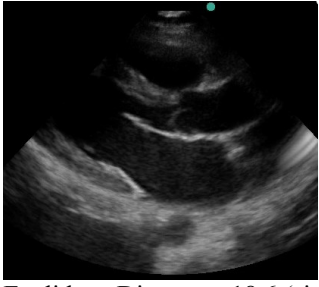


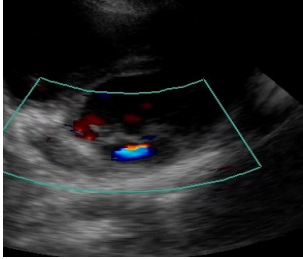


Figure 6: Some sections of the dendrogram of the above-clustered videos (Sectioned to concentrate on per video comparison)

Table 2: Comparison of videos, euclidean distances based on dendrograms

section	video 1	video 2	video 3
A	video-0011.mp4  Euclidean Distance : 0 (same) Known A4C	video-0050.mp4  Euclidean Distance : 12.19 (similar)	Not available
B	video-0002.mp4	video-0044.mp4	video-

	 <p>Euclidean Distance : 0 (same) Known PLAX</p>	 <p>Euclidean Distance : 18.6 (similar)</p>	 <p>0001.mp4 Euclidean Distance : 17.14 (similar)</p>
C	<p>video-0008.mp4</p>  <p>Euclidean Distance : 0 (same) Known PSAX</p>	<p>video-0041.mp4</p>  <p>Euclidean Distance: 3.13 (similar)</p>	<p>Not available</p>

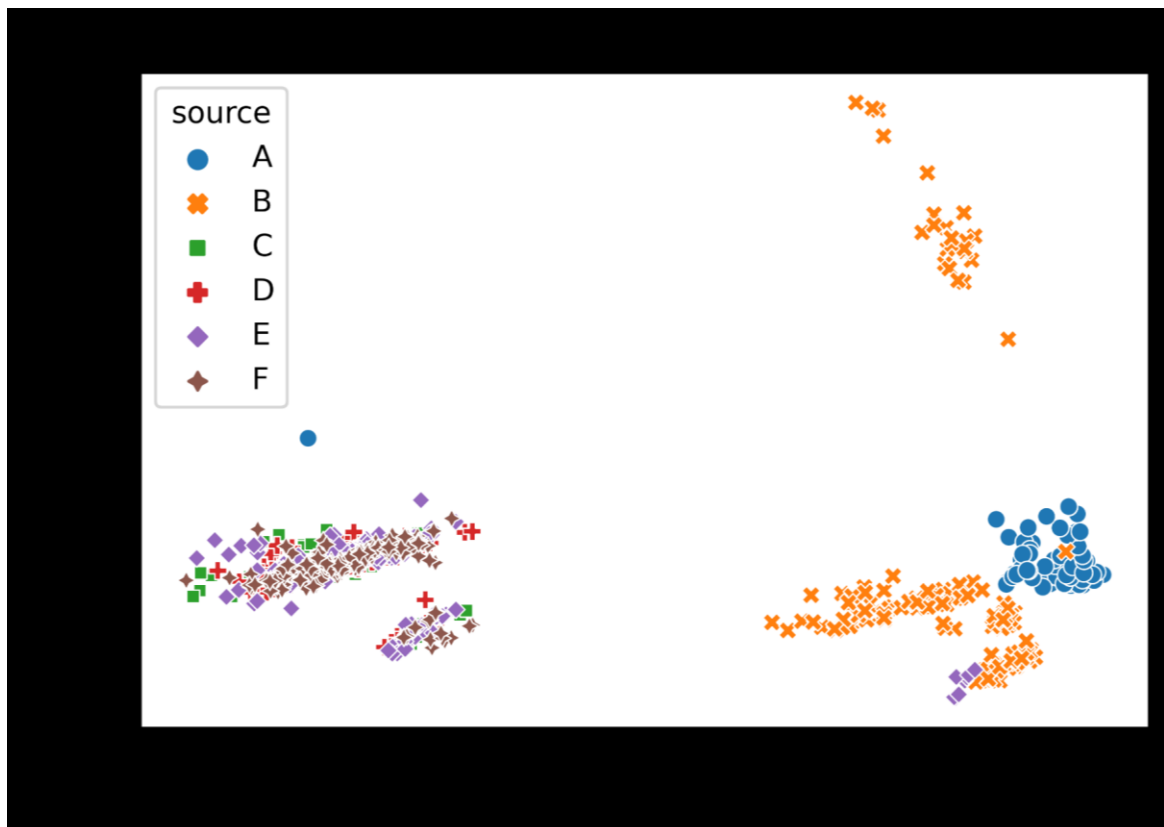


Figure 7: Effect of sonography machine settings

For the third dataset, 848 videos, we performed the same operations as above, we obtained Figure 7. We expected the files to be clustered similarly to the 1st and 2nd datasets, according to view. Upon inspection of the major clusters, we found that this was not the case. We found that all labelled videos were clustered together and were part of source F in Figure 7. After visual inspection, we found that the videos had been clustered

according to their sources. This may be because the different sources looked different visually. With some having patient electrocardiogram signals within the bounds of the ultrasounds. It was interesting to find that the machine configuration actually affected the results.

5. Business Benefits

Diagnostic ultrasound is a popular imaging method used for screening several soft tissues within the human body such as RHD screening, pregnancy monitoring, thyroid screening, breast cancer screening among others. Most of these examinations are performed by highly-trained clinicians and may take up to 1 hour even for skilled professionals. Automating different parts of this sonographic process would be crucial and very beneficial to sonography. Reducing the economic burden of diseases can be detected using sonography and ease the workload on clinicians while keeping the ultimate responsibility for diagnosis with human experts. A high-potential business case would be the design of an ultrasound device that uses embedded AI to aid sonographers in locating the view of interest quicker. This will cut across the several above fields.

6. Conclusions

Here we focused on finding a solution to the expert labelling time and cost problem. We were not looking to screen the echocardiogram for diagnosis but we are looking to locate the PLAX, PSAX and A4C views before they are seen by a clinician for annotation. We found that in the expert-labelled and unlabelled datasets, the methods very promisingly group similar videos. The potential to cluster images and videos with similar views together was clearly present. In the third dataset, it was confirmed that we must control for machines when performing analysis. Using unsupervised learning, we can reduce the time and resources required to annotate medical data for echocardiographic view classification. 2-component PCA for the videos in this study only explained 20%-30% of the variance. The low variance captured may be the reason why the sets of data did not completely define boundaries between the different views. If we grouped these videos into twos or threes, according to their similarity, we would potentially reduce the overall dataset being labelled by a factor of 2 or 3.

We recommend this approach for dataset size reduction, especially when used with small-size clusters. For future work, we will implement the above methods into a view annotation web application and access the impact on dataset size reduction. We will also explore using more n-components rather than just two. More n-components may explain a larger portion of the variance. We will look at using T-distributed Stochastic Neighbor Embedding (TSNE) for dimensionality reduction. Performance in more tests on larger datasets will determine how well the method can scale after accounting for the machine's configuration.

Acknowledgement

We would like to thank Data Science Africa for support through the Affiliated Centre Program, Google for funding this work through a Research Award and NVIDIA Corporation for a hardware grant to the Centre for Data Science and Artificial Intelligence (DSAIL).

References

- [1] Reményi, B., Wilson, N., Steer, A. et al. World Heart Federation criteria for echocardiographic diagnosis of rheumatic heart disease—an evidence-based guideline. *Nat Rev Cardiol* **9**, 297–309 (2012)

- [2] Soni, P. N., Shi, S., Sriram, P. R., Ng, A. Y., & Rajpurkar, P. (2022). Contrastive learning of heart and lung sounds for label-efficient diagnosis. *Patterns*, 3(1), 100400. <https://doi.org/10.1016/J.PATTER.2021.100400>
- [3] L. Mugambi, L. ZÜHLKE and C. W. Maina, "Towards AI Based Diagnosis of Rheumatic Heart Disease: Data Annotation and View Classification," 2022 IST-Africa Conference (IST-Africa), 2022, pp. 1-8, doi: 10.23919/IST-Africa56635.2022.9845657.
- [4] R. H. Webb, N. Culliford-Semmens, K. Sidhu, and N. J. Wilson, "Normal echocardiographic mitral and aortic valve thickness in children," *Heart Asia*, vol. 9, no. 1, pp. 70–75, Mar. 2017
- [5] Chartsias, A., Gao, S., Mumith, A., Oliveira, J., Bhatia, K., Kainz, B., & Beqiri, A. (2021). Contrastive Learning for View Classification of Echocardiograms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12967 LNCS, 149–158. https://doi.org/10.1007/978-3-030-87583-1_15/COVER
- [6] Kusunose, K., Haga, A., Abe, T., & Sata, M. (2019). Utilization of Artificial Intelligence in Echocardiography. *Circulation Journal*, 83(8), CJ-19-0420. <https://doi.org/10.1253/CIRCJ.CJ-19-0420>
- [7] Vaseli, H., Liao, Z., Abdi, A., Girgis, H., Behnami, D., Hooman Vaseli, al, Abdi, A. H., Luong, C., Taheri Dezaki, F., Dhungel, N., Rohling, R., Gin, K., Abolmaesumi, P., & Tsang, T. (2019). Designing lightweight deep learning models for echocardiography view classification. <https://doi.org/10.1117/12.2512913>, 10951(8), 93–99. <https://doi.org/10.1117/12.2512913>
- [8] Østvik, A., Smistad, E., Aase, S. A., Haugen, B. O., & Lovstakken, L. (2019). Real-Time Standard View Classification in Transthoracic Echocardiography Using Convolutional Neural Networks. *Ultrasound in Medicine & Biology*, 45(2), 374–384. <https://doi.org/10.1016/J.ULTRASMEDBIO.2018.07.024>
- [9] Madani, A., Arnaout, R., Mofrad, M., & Arnaout, R. (2018). Fast and accurate view classification of echocardiograms using deep learning. *Npj Digital Medicine* 2018 1:1, 1(1), 1–8. <https://doi.org/10.1038/s41746-017-0013-1>
- [10] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. In *Information Science and Statistics*. Springer-Verlag New York. <https://www.springer.com/gp/book/9780387310732>
- [11] Graphical Representation of Data Using Principal Components. (2002). *Principal Component Analysis*, 78–110. https://doi.org/10.1007/0-387-22440-8_5
- [12] Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
- [13] Agglomerative Hierarchical Clustering - Datanovia. (n.d.). Retrieved October 19, 2022, from <https://www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering/>