PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

EASy EXAM: Artificial Intelligence for Pattern Recognition of UltraSound Final Report

[redacted]

Activity Report

Abstract—Quickly diagnosing and administering aid to people experiencing cardiac arrest is difficult and has a low success rate. Traditional echo cardiograms take about 20 windows(different views of the heart), the EASy exam simplifies the diagnostic process with only one window, but requires significant training. The EASy exam categorizes a heart into one of seven phenotypes each with a treatment plan. In order to broaden accessibility of this technique, and minimize training, this project seeks to use AI and machine learning to analyse and categorize echo cardiogram images. The software would accept an ultrasound video and correctly place it in the appropriate phenotype. The end product is intended for use with mobile ultrasound technology allowing this test to be performed anywhere.

Index Terms—AI, Machine Learning, Ultrasound, Pattern Recognition, EASy Exam, Echo-cardiograms, Neural Networks, Image Classification, Image Segmentation

✦

1 The Problem

AROUND 600,000 people suffer from cardiac arrest events each year and the survival rates are low. One of the leading factors in the mortality rate is the difficulty in quickly assessing the correct course of treatment. It takes a significant amount of training to perform an echo-cardiogram and even more training to accurately diagnose the issue. Even in a large hospital there are only a few individuals able to perform the necessary cardiac assessment. Traditional examinations require around 20 different angles of the heart. The EASy exam (Echocardiographic Assessment using Subxiphoid-only view), developed by Dr. Nibras Bughrara, simplifies the process to only one angle. This image can obtained using a portable ultrasound allowing the test to be performed anywhere. The goal of the project is to utilize the simplicity of the test and machine learning to train an AI to identify one of seven phenotypes which are used to determine the treatment based on the EASy Exam Medical protocol. This will allow rural Hospitals without cardiologists to better provide cardiac care and even allow EMTs to administer immediate emergency aide.

[redacted]

2 Inspiration

Machine learning is a growing field that extends into many applications. In recent years, deep learning was used in various medical applications. Notably the segmentation capabilities which proved useful in diagnosing biological abnormalities. The EchoNet-LVH is a deep learning model that managed to quantify heart features such as wall thickness using AI. The model was capable to quantifying ventricular hypertrophy with precision equal to human experts [1]. Multiple other projects aim for similar goals with different heart diseases such as Aortic Stenosis from the Echocardiograms. The reason AI models are be-g adopted is the speed and efficiency they provide. Once the reliability is proven, these new tools will help with the diagnoses burden. [2], [3]. These projects show viability for using machine learning to train an AI to recognize heart features. Though similar projects exist, this particular application is unique and will need to be trained independently.

3 Design Constraints

- **Unorganized Data Samples** : Data samples may not be organized in their corresponding phenotype. Organizing data samples require prior medical knowledge and in order to train the system the data needs to be correctly categorized.
- **Undecipherable Data Samples**: The pixels of samples may be obscure. Samples may be in different angles which could lower the accuracy of the system. Depending on the operator the echo images can vary significantly.
- **Limited Data Samples**: Due to the confidentiality of the patients, the number of organized data samples may be constrained. Data from the ultra-sound may have personal information which must be removed in order to use in a research capacity.
- **Licensed Programs**: The required neural network may run on software that must be purchased and could be costly.
- **Network Access**: A large cloud network may be required depending on the sample size. Currently, the data samples are in the Mp4 and JPEG formats which would require a large networks if there are thousands of samples of such formats.
- **Time Constraint**: The project design period is 2 semesters which limits the complexity of the design as extensive testing is required and data is becoming available at a slow rate.

4 Our Solution

Our design aims to provide users with accurate predictions based on ultrasound videos of the subcostal view of the heart. To achieve this, we employed two AI models: one for image segmentation and the other for image classification. We used the U-net to segment the videos by highlighting the left and right ventricles in the echocardiogram. Using custom python code the video is run through a Frame averaging algorithms to process the video and convert it into a single image. The second AI model utilizes the convolutional neural network (CNN) to extract relevant features during the training process and produce the desired output. The model provides metrics for training and validation accuracy ensuring that the model's outputs are sufficiently reliable.

4.1 High-Level System Diagram

The process relies on the application of the EASy Exam protocol. Trained professionals conduct the examination by taking ultrasound images of the subcostal view of the heart. After that, these images are processed through the feature recognition AI to "highlight" the ventricles and pass that to the classification AI to identity the Phenotype and provide the confidence of the assessment in the form of the likelihood the image belongs to the given phenotype. After identifying the phenotype, the physician can use the EASy Exam protocol suggest the necessary treatment plan.

4.2 System Requirements

4.2.1 System Users

The users of this system are physicians and emergency workers. First of all, certified trainers would instruct physicians on how to utilize this system (EASy Exam Training). After the proper training, a physician would perform a subcostal cardiac scan using a portable ultrasound. Then, the ultrasound video with the correct view is produced and loaded into the system. The system will then process the video and feed the resultant data into the trained model. The output of the system is the suggested phenotype based on the model's assessment.

4.2.2 Functional Requirements

1) **Objective:** To have a physician input ultrasound video into the AI system and

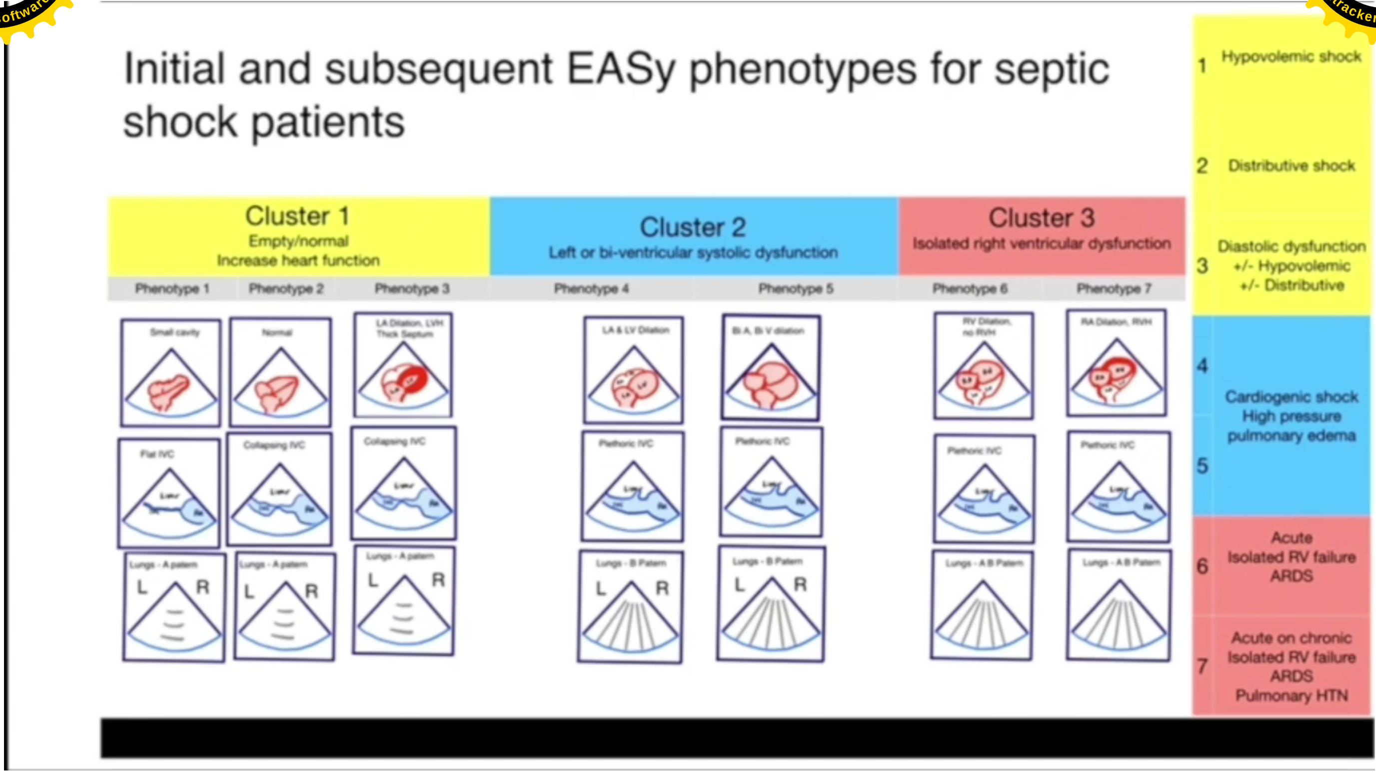

Figure 1. EASy Exam: Phenotypes.

have the system output the images into its corresponding EASy phenotype

2) **Features:** Some features of the system are processing AI(image annotation), feature detection AI(image segmentation), and classification AI(image classification).

4.2.3 Non-Functional Requirements

1) **Performance:** Accurately classify a heart image
2) **Usability:** More physicians will be able to use the ultrasound with the guidance of this system.
3) **Metrics and Measurements:** Training accuracy, validation accuracy, training loss, and validation loss are the relevant metrics. They are measured as percentage and quantity of data.

4.3 Justification

The main advantage of deep learning is adaptability which is a crucial factor when it comes to biological analysis. Adaptability is essential, as all organs are not the same exact size and shape from person to person and organs can change as a disease progresses. It's possible to account for such variables with traditional algorithms, however, it becomes an issue of complexity. Deep learning simplifies the process as the AI comes to recognize varied structures as long as there is a sufficient data set to train with. In the long run, data can be accumulated and given to the AI to train for a more robust model, this will effectively improve the accuracy and the system's adaptability in classifying organs.

4.4 Ethical Concerns

- Data Privacy Issues

The Images required to train the AI contain patient information which adds complications in getting them for research purposes. The images need to be de-identified and even then there are concerns with sharing hospital data.

- Misdiagnosis

This system would not be implemented if there was not significant confidence in its efficacy, but the implications of a misdiagnosis are troubling. Any automated process that has medical applications require substantial rigor in verifying their functionality.

5 Our System

The initial prototype AI will aim to implement the system features using widely available data to set the benchmark for the deep learning model's performance.

The main requirement for the data is correct representation, as due to lack of phenotype images different data sets will have to be used. Each set comes with its set of challenges such as the need for feature extraction and image processing. This will also affect the complexity of the model.

Currently the system exists as two separate AIs, one for segmentation and one for classification. Each uses a separate data set for the purposes of developing the model. With available data sets of substantial size both models work with a high degree of accuracy. Applying these models to heart images is only a matter of time and critical mass of available images.

6 System Design

The proposed solution aims to utilize the AI deep learning capabilities and available medical data to create an AI model to classify Ultrasound Heart images into their corresponding phenotype. The system was divided into 3 major functions that each need to be developed, they are Image Processing, Feature Extraction, and Classification.

6.1 Video Processing

To enable our classification AI to function effectively, we need to process the video inputs into a single image that captures all the relevant information. To accomplish this, we utilize our segmentation AI to extract the necessary information, along with a Frame averaging algorithm. By combining these tools, we can create a comprehensive image that allows our classification AI to make accurate and informed predictions.

6.1.1 Segmentation AI

Videos are composed of individual frames or images. In the case of the Easy exam, we focus on identifying differences in chamber sizes of the heart and fluids around it. These are the key features that our classification AI needs to detect autonomously. To enhance the process, we will train a segmentation AI to analyze the frames of the video and apply masks to highlight the chambers and fluids around the heart. By utilizing this approach, We will make our segmentation AI act as a filter for our classification AI. This filter will help make even lower quality scans useful for our classification AI.

6.1.2 Frame Averaging

Once the segmentation process is complete, we obtain multiple images or frames with the highlighted chambers and fluids. However, our final input format requires a single image. To achieve this, we apply a frame averaging algorithm, which calculates the average value of each pixel across all frames. With a sufficiently long video input, the frames will average out evenly, providing a clear representation of the heart's movement. This step generates the final input required for the classification process to commence.

6.1.3 Splitting the Data set

Our image data needs to be divided into two sets: a training set and a validation set. The training set will help us change the way the Model views an Image while the validation set will be used to confirm functionality for images the model did not see. The data set is split into 80% for the training set and 20% for the validation set.

6.2 Feature Extraction

Just like a human, an AI classifies images based on the features it sees. But for a more robust model it should be able to detect these feature or pattern of features independent of their location on the image. This process is called Local Feature Extraction which the Convolutional Neural Network (CNN) works best for.

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

6.2.1 *Re-scaling*

The CNN starts by re-scaling all the values of the Tensor (once at the start). This usually means the ”squishing of the values” in such a way so that all values are between 0 and 1 (ideally) at the start. This is done to normalize the data preventing the values from blowing up and hitting large numbers which could lead to a single pixel dominating our Convolutions step. This step effectively stops a single data point from being disproportionately weighted.

6.2.2 *Convolution and Pooling*

The CNN detects features by applying a mask to an area (usually in 2x2 or 3x3 pixel area) then squishing this area into a single pixel to represent it. The most common ways of representations are through the average or through the max value in this area as shown in figure 1.

Figure 2. Convolutions and Pooling example

6.3 Classification

Once all our features have been extracted into a single unique image that represents our whole video. The next step is to classify said image, we will allow each block of data to cast a vote on how likely this value belongs to a certain class. This voting process is represented through the Neural Network.

6.3.1 *Flattening*

The Neural Network starts by changing our data structure from a tensor to a simple array of data. This allows it to then connect each value of the array to a layer of a node known as a Dense Layer. It is called dense because every node is connected to every value in the array. This layer is then connected to another hidden layer or more until we reach the final output layer which has nodes equal to the number of classes we're classifying at this Neural Network level.

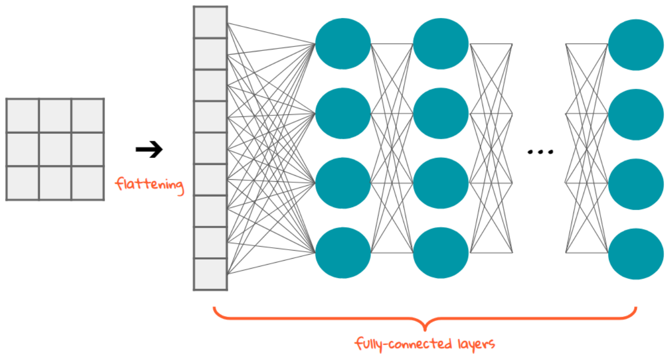

Figure 3. Flattening example

6.3.2 *Classification Algorithm*

There are multiple ways to represent each vote. They are represented through, ”How much do we value each node in our Neural Network?” or ”How will we reprelsent the final Result? as a binary (yes or no)? or a percentage of how close it is to a certain class?”. The first question is answered by the Loss function which compares the result to the predication the AI made and changes the vote weights in a way that drives the prediction closer to their right result. The second question depends on the size of the data set. Due to our small data set size, we will use Binary Classification.

Parts List & Budget

500 GB External SSD: The 500 GB SSD will be enough to store around 40,000 AVIs. The stakeholder (Albany Med.) will reimburse the purchaser (See Table 1).

Software Dependencies

1) **Anaconda:** Anaconda is an open-source distribution of software languages such as Python and R. In addition, Anaconda

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

Item	Part #	Supplier	Cost	Description
500 GB External SSD	Samsung T7	newegg	$75	Data Storage
TOTAL			**$75**	

Table 1
System Parts List & Budget

contains Jupyter Notebook. These applications are needed to create an AI system.

2) **Jupyter Notebook:** Jupyter Notebook is a web development that contains live code, equations, visualizations, and text. This interface can be used with the Python language.
3) **Python:** The AI system is created using the software language, Python. Python is a high-level and object-oriented programming language. Python has a variety of libraries and frameworks such as MATLAB and TensorFlow that are used for data-analysis.
4) **TensorFlow:** TensorFlow is a Python library for fast numerical computing, machine learning, deep learning, and algorithms. Some functionality of TensorFlow are acquiring data, training neural networks, creating predictions, and refining future results.
5) **Keras:** Keras is an open-source software library for implementing neural networks. Keras allows ease of neural network creation and supports multiple back-end neural network computations.
6) **V7 Darwin Labs:** An end-to-end computer vision development framework for teams creating image and video annotation products.

System Integration & Standards

1) **Agile Methodology (SCRUM):** The project follows an agile design approach. The project consists of sprints building off previous sprints. Each sprint follows these steps: plan, design, build, test, review, and launch. Each sprint supports test-ability, feasibility, and focus in critical points of the system.

7 Experimental Design

7.1 Objectives of the Experiment

This experiment sought to explore if AI models could be a viable solution for recognizing patterns in medical imaging. Our project aimed to demonstrate the feasibility of utilizing AI models for organic materials in medical applications. Specifically, we sought to evaluate the functionality of our AI model in analyzing lung images and applying masks to cell images. The ultimate goal was to provide healthcare professionals with a reliable tool that could assist in their diagnoses and treatments of various medical conditions.

7.2 Experiment Evaluation Parameters

As mentioned we tested two AI models. We assessed them through: Training Accuracy, Validation Accuracy, Training loss, and Validation loss.

- **Training Accuracy:** This metric represents the percentage of correctly predicted images from the training data set. It provides insight into how well the model is performing during training, but is less indicative of its overall predictive capability.
- **Validation Accuracy:** This metric represents the percentage of correctly predicted images from the validation data set. It is the most reliable indicator of the model's accuracy, as the validation data set is not used to alter model parameters, only to validate its predictions.
- **Training Loss:** This metric indicates the optimization of the loss function for the training data set. It measures how confident the model was when making incorrect predictions. A high training loss is indicative of over-fitting. While the training data set is being used to alter the model parameters, this metric is primarily used to track the training process.

Activity Report

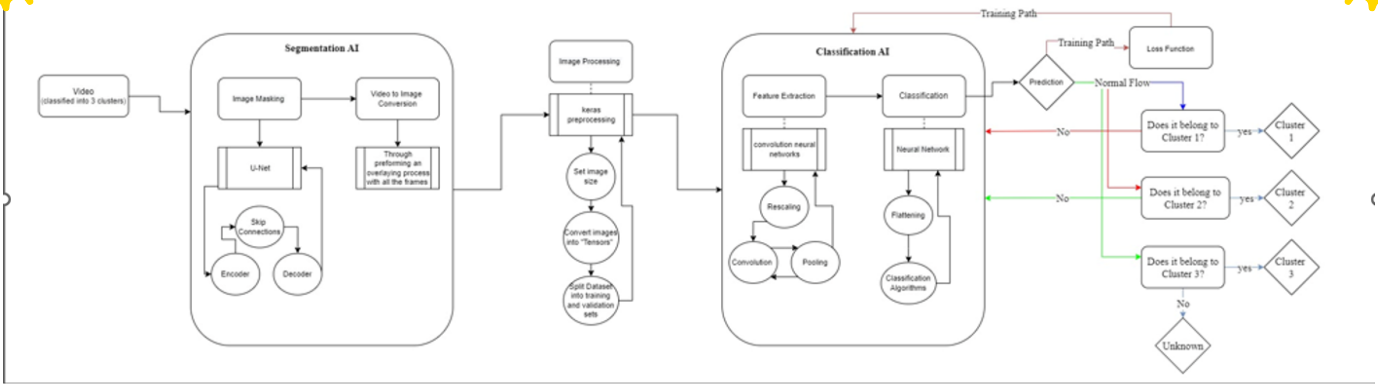

Figure 4. EASy Exam: Artificial Intelligence for Pattern Recognition of Ultrasound Images Physical Design use case diagram.

- **Validation Loss:** This metric indicates the optimization of the loss function for the validation data set. It is similar to the training loss, but provides a more accurate representation of the model's performance and over-fitting. The training of the model will end if the Loss function is unable to improve/lower our Validation loss.

7.3 Experiment Details

We assessed the impact of varying data set sizes on both AI models. The lung image classification model had an ample data set of around 5,000 images of two classes. We performed a rough comparison by gradually reducing the data set size by half each time. Conversely, the image segmentation model had a smaller data set of 670 images, which we considered a medium-sized data set. To compensate for this, we reduced subsequent datasets to 100 images and began reducing dataset sizes by half from there. The data was divided into a training dataset and a validation dataset in a 4:1 ratio, which is a common practice for training AI models. Our training code was designed such that if the loss function was unable to improve the validation loss anymore, the training would come to an end. Since the datasets being used were not the final datasets of interest, we conducted a proof of concept analysis. Our analysis adopted a qualitative approach rather than a quantitative approach, as these values were unlikely to remain the same in the final model.

8 Results

8.1 General Observations

There is a clear correlation between the number of images and the system's reported accuracy, as shown in figures (1, 2, 5, 6). Specifically, as the number of images used for training or validation increased, the accuracy of the system also increased. Conversely, as the number of images used for training or validation decreased, the accuracy of the system decreased. There is a risk of over-fitting the data by training the model too many times. There is a point at which the model will become less reliable as it "memorizes" images rather than identifying significant features.

8.2 Validation loss Ceiling

As previously discussed, the validation loss is a crucial indicator of the model's over fitting. If the loss increases while training, the model may become less adaptable and start to develop a bias. Therefore, in order to create an AI model capable of handling natural variability in human organs, it is crucial to have a significant number of distinct images to train the model. This approach prevents training the model with the same images repeatedly, which can cause the validation loss to increase as shown in figures (4, 8). This is why not all data sets finish at the same epochs, once the validation loss increases, the training stops. This is because we seek to have the best model accu-

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

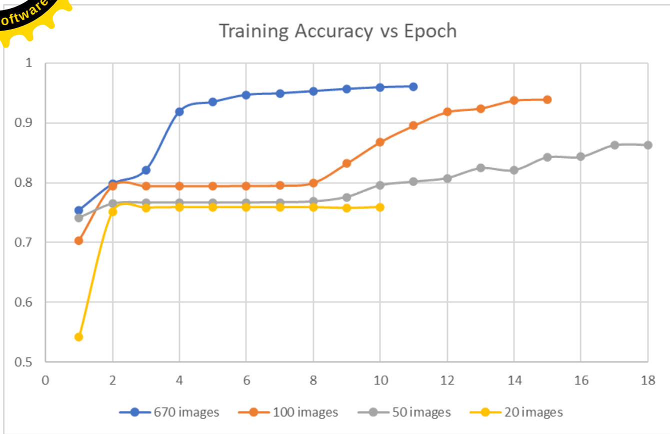

Figure 5. EASy Exam: Image segmentation AI Training Accuracy results

racy while maintaining the lowest validation loss possible, and avoiding over fitted models.

8.3 Impact of Small Data Size on Model Accuracy

The accuracy figures (1, 2, 5, 6) reveal that many small data sets experience breakthroughs in accuracy with a significant improvement in just one epoch. Conversely, some data sets demonstrate little to no progress over multiple epochs. The root cause of both phenomena can be traced back to the model training process. During training, the model makes assumptions based on the training data and validates them using the validation data set. With limited data, there is a delay in uncovering faulty assumptions, as there is insufficient data to disprove these assumptions during the training phase before validation.

9 Discussion

9.1 Data Quality Assessment

In that experiment, we aimed to test two AI models with different functionalities: image classification and image segmentation. Although these processes seemed distinct, they were closely related as image segmentation involved classifying each pixel within an image, as opposed to the entire image as a whole. Therefore, both AI models were evaluated based on the same assessment criteria.

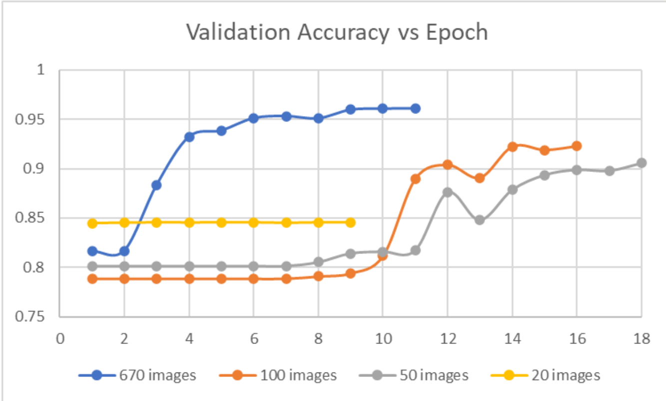

Figure 6. EASy Exam: Image segmentation AI Validation Accuracy results

Figure 7. EASy Exam: Image segmentation AI Training Loss results

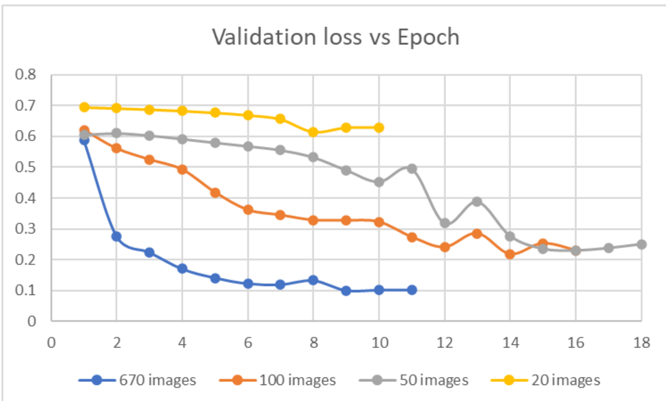

Figure 8. EASy Exam: Image segmentation AI Validation loss results

PDF-XChange Editor Click to BUY NOW! www.tracker-software.com

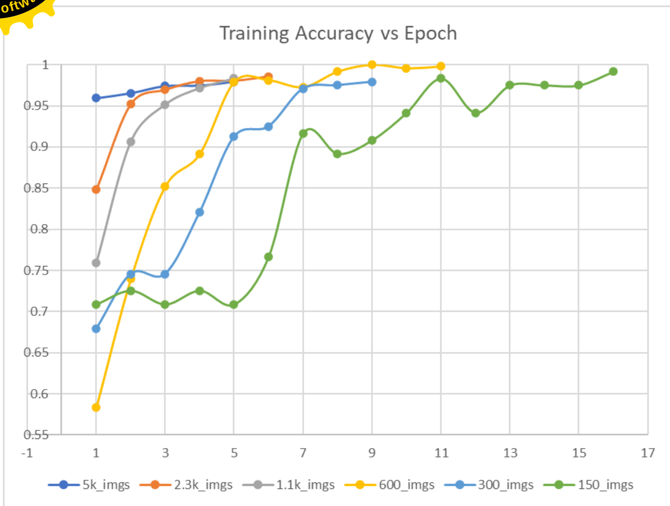

Figure 9. EASy Exam: Classification AI Training Accuracy results

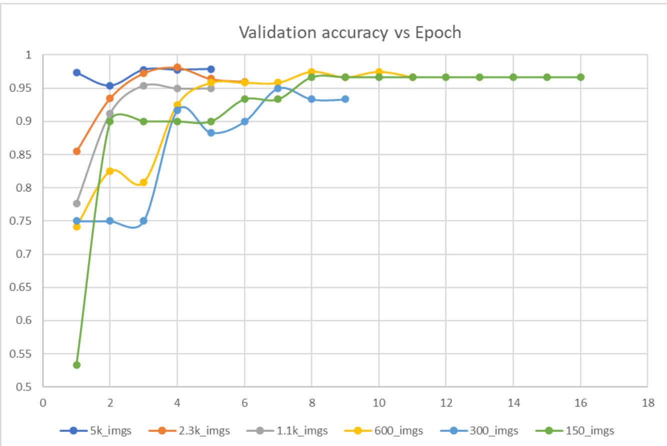

Figure 10. EASy Exam: Classification AI Validation Accuracy results

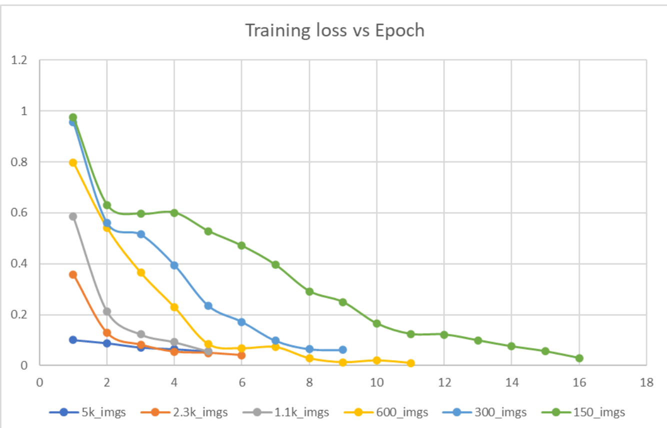

Figure 11. EASy Exam: Classification AI Training Loss results

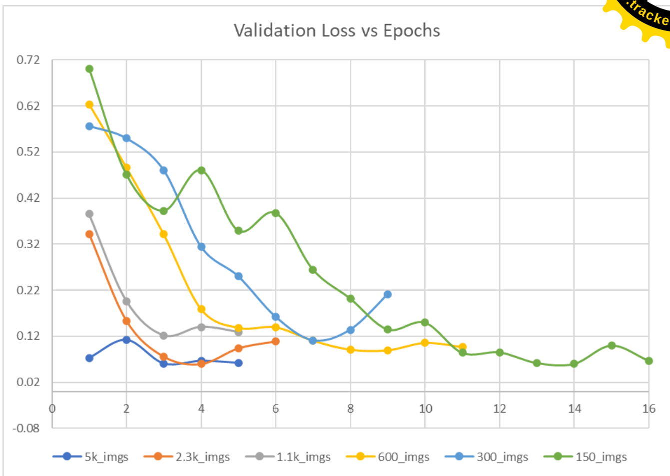

Figure 12. EASy Exam: Classification AI Validation loss results

9.2 Limitations & Future Work

- **Acquiring Data:** The limitations of acquiring data for AI include the availability and quality of the data, which can affect the accuracy and performance of the model. Future work could address this limitation by collecting a larger and more diverse data set, using data augmentation techniques, or developing methods for generating synthetic data. Regarding the quality of data, some work has already been done to create a program that segments the image, making relevant structures easier for the AI to parse.
- **Number of Phenotypes:** When faced with many categories that have specific differences, an AI model may struggle to learn effectively, particularly with a small data set, which can cause over-fitting. Essentially, the model will try to build correlations with the limited data set, and those correlations may not be relevant. To overcome this issue, it is important to collect a robust data set for each category. Each category should be well-represented as the model is less effective overall if it struggles to identify any single element. Similarly, it is important that the input format is similar or the same for each category so that the model differentiates for the right reason. For example, if some images of a category have a watermark or something

of the like, the model may use that to identify objects in that category rather than relevant structures.

- **Not User friendly:** Currently, our system's user-friendliness is limited by the use of multiple AI systems with different software programs to test its functionality. Additionally, the interface is purely in Python and would not be easily usable by someone not familiar with reading and using code. To address these limitations, future work could focus on unifying the system's software language and environment, utilizing cloud computing infrastructure, and developing a more user-friendly interface.
- **Not Portable:** Our system is currently reliant on powerful hardware that largely restricts its use to a desktop environment. Future work could include exporting a trained model onto custom hardware designed to specifically handle the needs of the AI model. Though the training process can have tremendous requirements, the trained model does not require az much overhead.

10 Conclusion

Our study aims to explore the potential of AI under controlled test conditions. Specifically, we aim to assess the influence of data size on the model's performance to better understand its capabilities and determine the viability of AI in identifying organic structures in medical imaging. Our results reveal that even a small number of images can produce a functional model; however, models produced from insufficient data do not generate the level of accuracy necessary for medical applications. Large amounts of input data (exceeding 1000 images) are necessary to create a high degree of confidence in the model's assessments. The significant advancements made by the model with larger data sets are an encouraging indication of its potential to handle more complex images with adequate training. The model's ability to excel with a relatively small data set implies that it has a robust feature extraction process and can generalize well to previously unseen data. Further testing has demonstrated that with a sufficient amount of data, the model exhibits remarkable potential in identifying and distinguishing features in medical imaging. As more training data becomes available for subcostal echocardiograms, we believe that the AIs outlined in this paper will be able to realize our goal of accurately differentiating heart images into their respective phenotypes.

References

[1] Z. Huang, G. Long, B. Wessler, and M. C. Hughes, *A new semi-supervised learning benchmark for classifying view and diagnosing aortic stenosis from echocardiograms*, 2021.

[2] G. Duffy, P. Cheng, B. He, N. Yuan, J. Ebinger, J. Zou, J. K. Patel, R. M. Witteles, S. Cheng, and D. Ouyang, "Precision phenotyping of left ventricular hypertrophy with echocardiographic deep learning," *Circulation*, vol. 144, no. Suppl_1, pp. A12 669–A12 669, 2021.

[3] J. Zhou, M. Du, S. Chang, and Z. Chen, "Artificial intelligence in echocardiography: detection, functional evaluation, and disease diagnosis," *Cardiovascular Ultrasound*, vol. 19, no. 1, pp. 1–11, 2021.