Machine Learning Model to Evaluate Endotracheal Tube Placement

Progress Report

Tina Gholami
Department of Electrical and Computer Engineering Western University London, Canada

Kyle Lacroix
Department of Electrical and Computer Engineering Western University London, Canada

Spencer Vecile
Department of Electrical and Computer Engineering Western University London, Canada

***Abstract*—More people than ever before are being ventilated due to COVID-19. Since proper tube insertion is essential for an optimal outcome, tube placement is verified using x-rays. A radiologist must read the x-ray. Due to mounting caseloads caused by COVID-19, automation of the x-ray reading process would result in higher turnaround times and improved patient care. The National Institutes of Health Clinical Center provided the chest x-ray data of patients that were ventilated. The x-rays of the chest images were first prepared through preprocessing. The images were reduced in size since the larger images required increased processing time but did not offer any additional information. Google's EfficientNetB0 was used as the model for the data. During training, the dataset was split: 70% training, 15% validation and 15% test. After the initial tuning, the model achieved a training accuracy of 100%. Although the results are promising, further tuning of the model is warranted to improve the metrics of the model.**

Keywords—machine learning, illness

I. Introduction

As COVID-19 continues to spread around the world, hospitals are becoming overwhelmed. Doctors and nurses cannot spend as much time with their patients as they used to which could lead to an increase in human errors. When COVID enters the body, it can slowly lower the oxygen saturation levels. If the blood oxygen level is too low, the blood cannot provide organs with enough oxygen needed for life. To deal with this, hospitals put patients on forced ventilation to provide sufficient oxygen. The patient must be intubated, a procedure in which a doctor or nurse inserts an endotracheal tube (ET) into the throat. To ensure the tubes are placed correctly, a physician or radiologist must manually look at an x-ray to verify that they are placed in the optimal position. Not only does this leave room for human error, but delays are also common as radiologists can be busy reading other scans. Implementing an automated process that can provide instant feedback on tube placement could help patient care and reduce hospitals' workload. The hospitals would considerably benefit from this process even when the COVID crisis is over. This project will examine the possibility of using machine learning to read x-rays to determine if ET tubes are correctly inserted.

II. Related work

A recent study by Borkowski et al. [1], investigated using AI to diagnose COVID-19 reliably. These researchers used publicly available CXR images for patients with COVID-19 pneumonia, pneumonia from other etiologies, and normal CXRs as a dataset to train Microsoft CustomVision. Pneumonia is an infection that inflames your lungs' air sacs. The researchers achieved a trained model with 92.9% recall and precision. While their metrics were not as high as some ML projects, it does show promise. Utilizing more data that captures more information would yield an even better result. Another study by Jain et al. [2], took a closer look at pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. The researchers created six models. The first two models had two and three convolutional layers respectively. The remaining four models were pre-trained models. By the end of the study, they had six models that had decent validation accuracy ranging from 70.99-92.31%. This study furthers the proof of concept**.**

III. Dataset

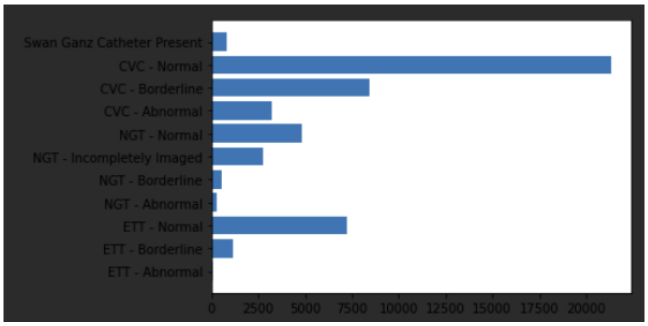
The dataset that will be used for this project contains x-ray images of catheter lines inside the patient's lungs [3]. This dataset was provided by the National Institutes of Health Clinical Center and was labelled by a team of over 30 doctors. This dataset is broken down into a set containing 3582 test images and 30083 training images; unfortunately, the test images are not labelled as they were for the competition so they will be unusable. The images are not uniform in size and range from 2000-3000 pixels in width and 2000-3000 pixels in height. Also included in the dataset is a train.csv file which has 13 columns including StudyInstanceUID (unique ID for each image), ETT - Abnormal (endotracheal tube placement abnormal), ETT - Borderline (endotracheal tube placement borderline abnormal), ETT - Normal (endotracheal tube placement normal), NGT - Abnormal (nasogastric tube placement abnormal), NGT - Borderline (nasogastric tube placement borderline abnormal), NGT - Incompletely Imaged (nasogastric tube placement inconclusive due to imaging), NGT - Normal (nasogastric tube placement borderline normal), CVC

- Abnormal (central venous catheter placement abnormal), CVC - Borderline (central venous catheter placement borderline abnormal), CVC - Normal (central venous catheter placement normal), Swan Ganz Catheter Present and PatientID (unique ID for each patient in the dataset). This CSV file has columns 2-11 in a binary encoding format which is the true labels.

IV. Methodology

A. Exploratory Analysis

The first thing revealed during the dataset exploration was that although there are 30083 images, there are only 3255 unique patients so some patients have multiple x-rays in the dataset. At maximum, a patient has 172 x-rays and at minimum, a patient has one x-ray. Next was the number of observations for each label which is summarised in the figures below:



Malpositions	Number of Observations
ETT - Abnormal	79
ETT - Borderline	1138
ETT - Normal	7240
NGT - Abnormal	279
NGT - Borderline	529
NGT - Incompletely Imaged	2748
NGT - Normal	4797
CVC - Abnormal	3195
CVC - Borderline	8460
CVC - Normal	21324
Swan Ganz Catheter Present	830

Fig. 1 & 2. Dataset

From the chart and table, it can be seen that the classes are not balanced with the majority of them being CVC-normal. It can also be seen that each image can have multiple labels.

B. Preprocessing

During the preprocessing step, first, the training labels are loaded and the columns StudyInstanceUID and PatientID are dropped. A new column is created called combined_label. In this column, first, the binary encodings are concatenated into a single binary label and then encoded using sklearns LabelEncoder class. This results in 211 unique labels in the dataset. Once this is complete, the images are loaded, resized to 224x224 and paired with their corresponding labels. Now that all the data is in a single list, it is split into 75% training data and 15% test data.

C. Model

The first model tried was a simple architecture that had very poor results which led to the adoption of a pre-built architecture. The model used is Google's EfficientNetB0 [4] with randomized initial weights, an input size of 224x224x1, sigmoid activation function on the output layer and 211 output classes. The model is compiled with sparse categorical cross-entropy loss, Adam's optimizer and accuracy as the metric. Finally, during training, a validation split of 15% is used so the final dataset split is 70% training, 15% validation and 15% test.

D. Hyperparameter Tuning

Thus far, there hasn't been a lot of hyperparameter tuning since most of the time has been spent on preprocessing the dataset and finding an appropriate model to use. One part that has been tuned is the input image size. To start, the images were shrunk to 650x650 but this created problems with the dataset being too large. It was then scaled down to 450, 350 and finally 224. There was no difference between classification accuracy that could be seen with the different image sizes; however, with a smaller sized image, training was much quicker and the whole dataset could fit in memory. Batch size was also tuned. First, 64 was used but training was very slow so it was lowered to 32 which cut training time in half while maintaining epochs constant.

V. Results and Analysis

A. Metrics

TABLE I. Training

Epoch	***Loss***	***Accuracy***	***Validation Loss***	***Validation Accuracy***
1	0.0591	0.9942	1.9559e-4	1.0000
2	1.1200e-4	1.0000	4.8334e-5	1.0000
3	4.8205e-5	1.0000	2.3847e-5	1.0000
4	2.5315e-5	1.0000	1.2900e-5	1.0000
5	1.51193-5	1.0000	8.29933e-6	1.0000

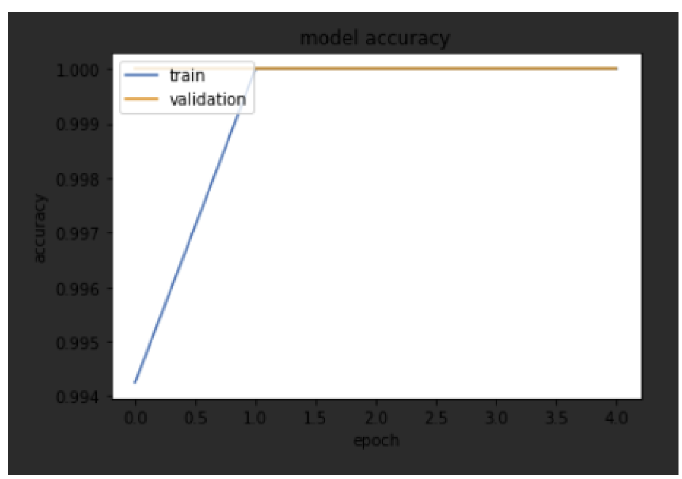


Fig. 3. Metrics of the model during training

Table 1 and Fig 3 shows the values that were recorded at the end of 5 epochs of training and are accurate in evaluating the model's performance. After two epochs, the accuracy went to 100%. The loss continued to decrease as additional epochs were added. The validation loss decreased as well demonstrating that overfitting was not an issue. The same trend can be seen with accuracy and validation accuracy. Although the model works well while training data, further tuning is warranted to ensure the model performs well with the testing data.

B. *Expected Results*

In this project, the goal is to predict the tube emplacement as accurately as possible. Towards that end, the number of both false positives and false negatives should be kept at a minimum. The former case requires unnecessary operations just to later realize that the tube was placed correctly, not to mention the pain and anxiety the patient must go through during the physical process. Similarly, in the latter case, the wrong prediction of the displacement of the tube causes long-term effects on the patients' lungs and trachea. Furthermore, since the data is imbalanced, using accuracy as the only metric cannot give us a real perspective on the model's performance. For such skewed classification problems, it is suggested to use a variety of metrics such as accuracy, f1-score, recall, etc. altogether to yield better performance evaluation results.

VI. Next Steps

Different architectures and different types of algorithms will be tried to improve the metrics. Convolutional Neural Network model has been partially constructed. The Convolutional Neural Network will be concurrently developed with the current model. Once both models are completed, the current model will be compared with the performance obtained from the Convolutional Neural Network model to see which model is superior. As well, to improve the metrics further, the hyperparameters will be tuned over the upcoming weekend. The various hyperparameters will be considered for tuning. As the algorithm is tuned, the changes in results will be documented for the Final Project Report.

Our model accuracy seems very high and it could be due to some data leakage between the training test and validation sets. The theory is that because some patients have multiple records, they are most likely spilling into the three different sets causing artificially high accuracy. To fix this a custom splitting function must be created to ensure a patients records are only in one of the three sets.

On March 24th, the group will meet to discuss what will be included in the presentation. During the meeting, each group member will be given a task for the presentation. From 25-27th of March, the slides with accompanying text will be created. On March 28th, the group will put all of the slides together. From March 29-31st, the group will practise and record the presentation. On April 1st, the presentation will be submitted to OWL. A Google doc and a GitHub repository have already been created for the Final Project Report and Code Submission. As the results are obtained from the model, they will be entered into the Google doc. Once all of the results are obtained, they will be analyzed for trends. As code is created for the project, it will be uploaded to the GitHub repository. On April 7th, a group meeting will occur to discuss what needs to be included in the Final Project Report and to divide the work among group members. On April 14th, the group will meet to finalize the report.

References

[1] A. A. Borkowski, N. A. Viswanadhan, L. B. Thomas, R. D. Guzman, L. A. Deland, and S. M. Mastorides, "Using Artificial Intelligence for COVID-19 Chest X-ray Diagnosis," Federal practitioner : for the health care professionals of the VA, DoD, and PHS, Sep-2020.

[2] R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," Measurement, 01-Jul-2020.

[3] "RANZCR CLiP - Catheter and Line Position Challenge," Kaggle. [Online]. Available: https://www.kaggle.com/c/ranzcr-clip-catheter-line-classification.

[4] Tan, Mingxing, and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in Int. Conf. on Machine Learning, 2019, pp. 5.