Predicting Clinical Parameters And Abnormalities From Biomedical Lung Dataset Using Machine Learning Algorithms

Bastin Joseph
University of Wisconsin - Madison
bjoseph5@wisc.edu

Apurbaa Bhattacharjee
University of Wisconsin - Madison
bhattachar26@wisc.edu

Abstract

The project focuses on solving three biomedical image analysis tasks on Lung CT and COVID-19 datasets namely lung segmentation, predicting clinical parameters of the lungs and predicting clinical abnormalities. We have implemented, analysed and benchmarked multiple image segmentation algorithms to segment the lung area in the images. We are able to obtain Dice coefficient value of 0.9341 and IoU value of 0.8767 using UNet++ on Lung CT Dataset and Dice coefficient value of 0.9156 and IoU value of 0.8445 using UNet+ on COVID-19 Dataset. For the image classification task on COVID-19 Dataset we are able to achieve an accuracy of 0.7544 using a convolutional neural network. We are also able to predict lung area and lung volume for the Lung CT Dataset with R2 score of 0.9866 and 0.9851 respectively. From the experiments performed as a part of this project, we have gained sound knowledge of machine learning algorithms for biomedical image analysis.

1 Introduction

Biomedical image analysis is one of the core research area in the machine learning and deep learning domain. With the increased availability of computational power in the form of GPUs it is now possible for researchers to train deep neural network models. This has led to the widespread popularity of deep learning algorithms in solving computer vision and medical image analysis problems. The key challenge in solving medical image analysis problems is that often the dataset is difficult to procure or not released to public. Also for generating the labels on the ground truth medical images expert intervention is often required which becomes a difficult and time consuming process. Researchers and scientists from the medical community have been constantly helping computer scientists to create the datasets needed for analyzing and solving problems in medical image analysis using deep learning.
Medical image analysis problems can be related to brain, lungs, kidneys, heart, liver etc. and predicting abnormalities and/or clinical parameters in those organs. With advances in the field of computer vision techniques we can get better accuracy in predictions performed on the medical images. Usually biomedical image analysis tasks require researchers to segment out the shape of the organs using segmentation strategies and then perform some regression or classification task on the segmented feature. It is of utmost importance to have low False Negative(FN) rate in medical image analysis because if we predict some patient's data as negative, which was actually positive then it could lead to delayed or wrong treatment which might even be lethal. So we should keep this in mind while designing our model.
Image augmentation is a pre-processing step often used in biomedical image analysis. This is because of the limited data availability of medical images. So researchers often add augmentation to the dataset to increase the number of training data for the model.
Lung diseases are the one of the most widely spread diseases in the medical community. Our project

deals with several medical image analysis tasks related to lung diseases. We are trying to come up with a model(s) which can predict clinical parameters and abnormalities of human lungs. Our motivation for this project is the fact that predicting necessary clinical parameters and abnormalities by our model will be useful in the medical community for analysing lung diseases like Emphysema, Lung Cancer, Chronic Obstructive Pulmonary Disease(COPD), Tuberculosis, COVID-19.

2 State-of-the-Art

There are several machine learning and deep learning algorithms which are considered state of the art for biomedical image analysis.

Image Segmentation Image segmentation lies at the core of biomedical image analysis problems where the key challenge is to segment out the organ under inspection using the ground truth image and the mask image. Deep models are used to train the ground truth images and mask images for segmenting out the organs. UNet (5), UNet+ (6), UNet++ (6), Fast R-CNN (7), Faster R-CNN (8), Mask R-CNN (9) are some state of the art deep neural network models used for segmentation tasks. Using the output of the segmentation model several image regression and classification tasks can be achieved using Convolutional Neural Networks.

3 Problem Statement

Our aim here is to understand various machine learning algorithms that can be applied to practical problems in biomedical image analyis. We place our focus on two datasets in our study namely - COVID-19 Dataset (2), (1) and Lungs CT Dataset (3). We will be presenting and comparing multiple machine learning models to segment lungs, classify images and find clinical parameters.

3.1 COVID-19 Segmentation and Classification

The dataset we have here has chest X-ray (2) of patients along with ground truth masks (1). These are the X-rays of patients and has label of whether the patient is COVID-19 positive or not. So our goal here is two-fold. The first task is to segment the X-ray to obtain the region of masks. The second task is to classify the patient as COVID-19 positive or negative based on the X-ray.

3.2 Lungs CT Segmentation and Clinical Parameter Prediction

We have taken a labelled dataset of lung CT (3) with ground truth mask and clinical parameters of the lungs. Our goal here is to segment the lung area in CT scan image as well as predict clinical parameters. Here we will be performing two tasks - image segmentation and regression. Lung Densitometry test on CT images of lungs is used to assess the attenuation level of the pulmonary tissues inside the lungs. The attenuation levels are expressed in Percentile Density (PD) 5th to 95th percentile Hounsfield units (HU) which reflects both the degree of inflation inside the lungs and the structural abnormalities (12) (13). PD05 through PD95 measures are often used in classifying the level of lung cancer. Case studies on smokers suggest that often for smokers the PD value is higher as compared to that of non-smokers. (3)

4 Method

4.1 Image Segmentation (Both Datasets)

Let us look at the Image Segmentation task first. This is a problem posed in both of the dataset that we are looking at. We have implemented and compared multiple state-of-the-art image segmentation algorithms namely, UNet (5), UNet+ (6), UNet++ (6), Fully Convolutional Network (FCN) (10) and a Convolutional Neural Network (DownUpNet) which uses downsampling followed by upsampling.

We have implemented each of the algorithms mentioned here on our own (UNet implementation is based on the baseline (4)). We have trained and tested our implementation of the segmentation algorithms on each of the two datasets and will be conclusively reporting the results.

We will discuss further about the implementation and observation of each of these algorithms.

Network Architecture Each of the 5 models are implemented with architecture as described below

1. UNet (implementation based on (4))
   We have used 4 levels of convolutional layers. On the downscaled path from input we have $8, 16, 32, 32$ filters at each level$(0, 1, 2, 3)$ with filter size of (3×3). On the upscaled path we added $32, 32, 24, 16$ filters at each level$(3, 2, 1, 0)$ with filter size of (2×2). Downsampling and Upsampling is done between each level with a factor of 2. The final upsampled layer is followed by a convolutional layer with 64 filters size (1×1). This layer uses dropout with a factor 0.5. Final output layer follows this layer. Each convolution operation is performed and padded to obtain same size output. We use $ReLU$ activation throughout except for output layer. Output layer uses $sigmoid$ activation function.
2. UNet+, UNet++
   These models are built on top of UNet implementation with skip connections made as in (6).
3. Fully Convolutional Network (FCN)
   This model has 8 convolutional layers followed by output layer, all using padding of the same size. Convolutional layers have $8, 16, 32, 32, 32, 24, 16, 64$ filters of size $3, 3, 3, 1, 2, 2, 2, 1$ respectively. $Sigmoid$ activation is used for output layer. All other layers use $ReLU$ activation.
4. DownUpNet
   This model is derived from UNet by removing the skip connections.

4.2 Binary Classification for Abnormality Detection (COVID-19 Dataset)

For the COVID-19 X-ray dataset we classify X-rays into COVID-19 positive or negative. We use a simple Convolutional Neural Network model to perform binary classification of the images.

Network Architecture We use a CNN for performing regression task. It has a set of convolutional layers followed by fully connected layers. There are 3 convolutional layers with $8, 16, 16$ filters with max pooling layer (2×2) after each layer. Each convolutional layer uses filter of size (3×3) with $ReLU$ activation and same size padding. There are 6 fully connected layers with $256, 128, 64, 32, 16, 8$ units respectively each using $ReLU$ activation. Output layer uses $sigmoid$ activation and outputs a single output value.

4.3 Linear Regression on the Segmented Data (Lung CT Dataset)

The Lung CT dataset has labels of lung area, lung volume and lung percentile density(PD) in Hounsfield Units. We use a simple CNN model for predicting these values using linear regression.

Network Architecture We use a CNN for performing regression task. The convolutional layers are same as that of the classification model described above There are 3 fully connected layers with $256, 128, 16$ units respectively each using $ReLU$ activation. Output layer outputs a single output value.

4.4 Evaluation Metrics

For evaluating the performance of the segmentation models we have used several evaluation metrics like **Validation Loss, Dice Coefficient, IoU** (18).

Dice Coefficient Dice coefficient is a statistical measure for computing the similarity between two samples. (19)

$$D = \frac{2 * |\hat{Y} \cap Y|}{|\hat{Y}| + |Y|}$$

We compute the Dice coefficient between the predicted masks($\hat{Y}$) and ground truth masks(Y). The range of Dice coefficient is 0-1 and higher the value, higher is the segmentation model accuracy.

Intersection over Union (IoU) The Intersection over Union(IoU) is also known as Jaccard Index (19). It is also a statistical measure for computing the similarity between two samples.

$$IoU = \frac{|\hat{Y} \cap Y|}{|\hat{Y} \cup Y|}$$

We compute the IoU between the predicted masks($\hat{Y}$) and ground truth masks(Y). The range of IoU is 0-1 and higher the value, higher is the segmentation model accuracy.

5 Results and Discussion

5.1 Training

We have performed different experiments on the Lung Dataset (CT, X-ray). Given below are the training parameters used for our model training and the details of the experiments performed by us as a part of the project.

1. **Platform** - TensorFlow Keras, scikit-learn
2. **Training Epochs** - We have trained each of the segmentation models for 100 epochs. We observed that the model was converging well at/before 100 epochs, so we did not train the model further as it would lead to over-fitting of the training data.
3. **Train-Validation Split** - The data is split into 80% training and 20% validation parts randomly sampled.
4. **Model Optimizer** - We have used **Adam optimizer** with learning rate = $2e^{-4}$.
5. **Loss Functions** -
   (a) **Segmentation Model** - Binary Cross Entropy Loss
   (b) **Classification Model** - Binary Cross Entropy Loss
   (c) **Regression Model** - Mean Squared Error(MSE)
6. **Evaluation Metrics** -
   (a) **Segmentation Model** - Validation Loss, Dice Coefficient, IoU
   (b) **Classification Model** - Accuracy, Precision-Recall, F1-score, Confusion Matrix
   (c) **Regression Model** - R^2 Metric

5.2 Image Segmentation

We have tried segmenting both our datasets with 5 different image segmentation models as mentioned before in the Method section - UNet, UNet+, UNet++, Fully Convolution Network and DownUpNet. For getting the segmentation results we have trained each of these models for 100 epochs using Adam optimizer having learning rate of $2e^{-4}$.

Figure. 1 shows the results of our models on the Lung CT Dataset. We have randomly sampled five images from the test dataset and used them on each of the five models - FConvNet, DownUpNet, UNet, UNet+, UNet++. Figure. 2 shows the results of our models on the COVID-19 Dataset. We have randomly sampled five images from the test dataset and used them on each of the five models - FConvNet, DownUpNet, UNet, UNet+, UNet++. From the results we can see that the segmentation results of the FConvNet is not good. DownUpNet, UNet, UNet+, UNet++ are able to segment out the lungs satisfactorily with UNet++ results being the most accurate. Table. 1 and Table. 2 shows the evaluation metrics(Dice coefficient, IoU (18)) obtained using different segmentation models on the Lung CT and COVID-19 validation dataset respectively. The results for all the models have been summarized in Figure. 4 and Figure. 5.

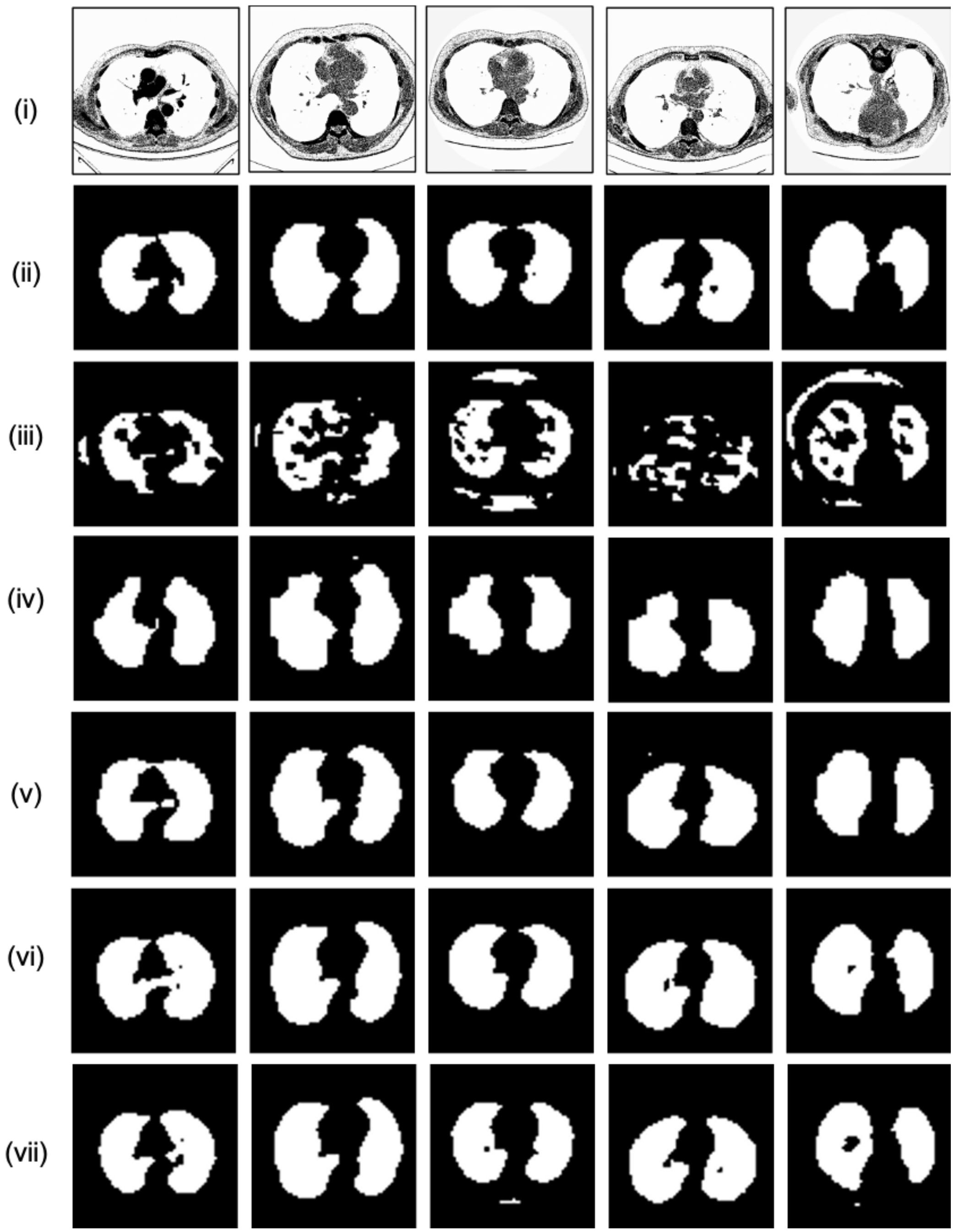

Figure 1: Lung CT Dataset Segmentation Results - (i) corresponds to the ground truth images, (ii) corresponds to the ground truth mask images, (iii) segmentation results using Fully Convolution Network(FConvNet), (iv) segmentation results using DownUpNet, (v) segmentation results using UNet, (vi) segmentation results using UNet+, (vii) segmentation results using UNet++.

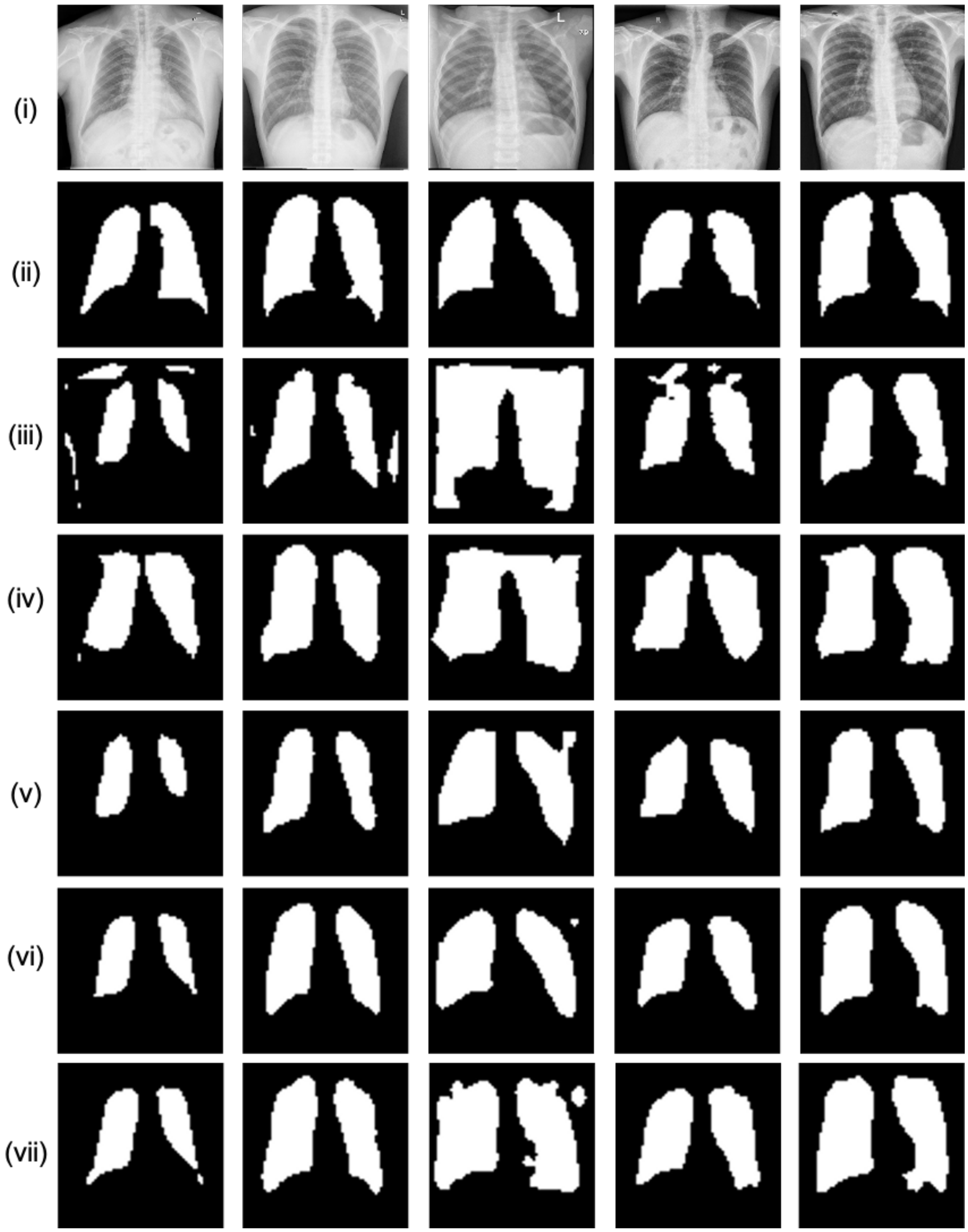

Figure 2: COVID-19 Dataset Segmentation Results - (i) corresponds to the ground truth images, (ii) corresponds to the ground truth mask images, (iii) segmentation results using Fully Convolution Network(FConvNet), (iv) segmentation results using DownUpNet, (v) segmentation results using UNet, (vi) segmentation results using UNet+, (vii) segmentation results using UNet++.

Model	Dice Coefficient	IoU
FConvNet	0.7784	0.6397
DownUpNet	0.8667	0.7650
UNet	0.9185	0.8504
UNet+	0.9340	0.8766
UNet++	0.9341	0.8767

Table 1: Evaluation metrics and results for segmentation models on Lung CT validation dataset

Model	Dice Coefficient	IoU
FConvNet	0.8549	0.7470
DownUpNet	0.8743	0.7769
UNet	0.8966	0.8130
UNet+	0.9156	0.8445
UNet++	0.9047	0.8302

Table 2: Evaluation metrics and results for segmentation models on COVID-19 validation dataset

5.3 Image Classification

For the COVID-19 dataset, the task was to classify the patients as COVID-19 positive or negative. For this task we have used a simple CNN of depth 3 with $ReLU$ as the activation function for hidden layers. $Sigmoid$ activation function has been used for the output layer. For checking the performance of the classification model we have used several evaluation metrics like **Accuracy, Precision-Recall, F1-score, Confusion Matrix** (14), (15), (16), (17). Figure. 3 shows the classification accuracy of our model trained for 100 epochs. We have observed that the generalization gap between the training and validation loss is high, so the model is $overfitting$, the reason being that our dataset is not large. To reduce the generalization gap, we should train the model with more training instances. Table. 3 shows the confusion matrix on our classification results. The following is the classification summary report for our model -

1. Training accuracy = 100%
2. Validation accuracy = 75.44%
3. Precision = 0.73
4. Recall = 0.80
5. F1-score = 0.71

		Actual	
		Positive	Negative
Predicted	Positive	49	12
	Negative	18	35

Table 3: Confusion Matrix for COVID-19 lung X-ray image classification

5.4 Image Regression

For the Lung CT dataset, the task was to predict clinical parameters like lung area, lung volume, lung percentile density(PD). The dataset consisted of 2D images and masks and also 3D images. We have used a simple CNN model for the regression task. There were only four 3D images(without labels) provided in the dataset, so due to lack of data and labels we could not train a model with the 3D data. Initially we predicted the regression values for all the features provided to us, but we observed that the prediction for the PD features were not accurate. So our inference from the initial experiment is that the parameters related to lung PD values depends on the features of the 3D structure of lungs. R^2 metric has been used as an evaluation metric for the regression model. The scores have been summarized in the Table. 4 and we can clearly observe that the predictions for PD values are not accurate. Hence we have predicted the lung area and lung volume fraction from our regression model. Figure. 6 visualizes the regression loss for the predictions of the two features lung area and lung volume. In practice the lung area can be used to assess the attenuation level of the pulmonary tissues inside the lungs. Abnormalities like Emphysema, Chronic Bronchitis, Chronic Obstructive Pulmonary Disease(COPD) are caused by inflammation in the lung alveolar tissue (11) which can be analysed using the lung area feature predicted by our model.

Our implementation of the models can be found at https://github.com/tobastin/CS760-Project-Lungs

Feature	R^2 Metric
Lung Area(mm^2)	0.9866
Lung Volume Fraction	0.9851
Lung Mean PD(HU)	0.5584
Lung PD95(HU)	0.4356
Lung PD05(HU)	0.1322

Table 4: R^2 Metric for the predicted features using regression on Lung CT Dataset

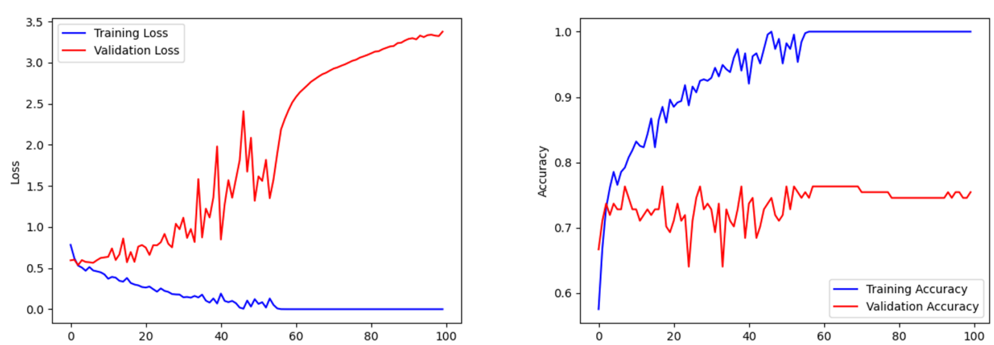

Figure 3: COVID-19 Dataset Classification Evaluation Metrics - Loss and Accuracy

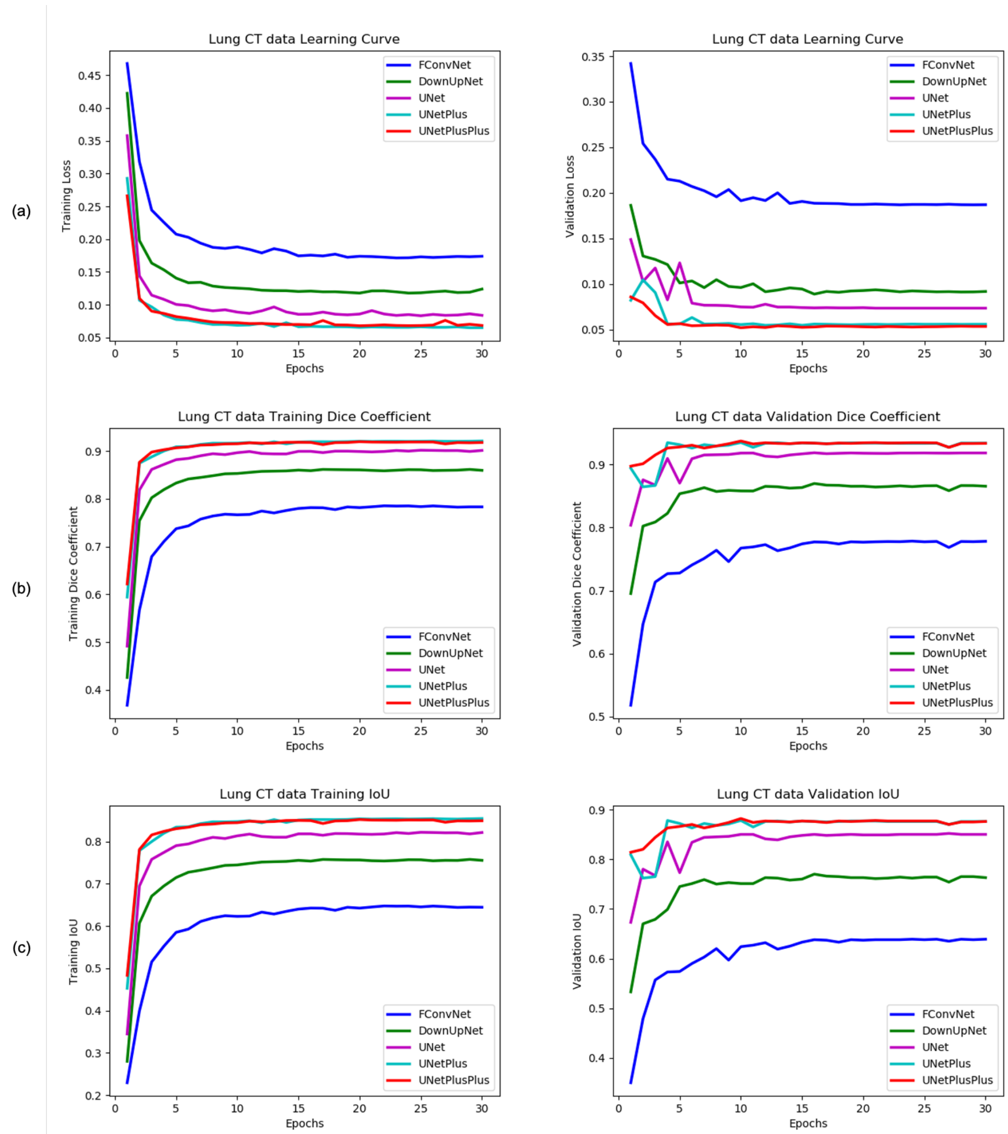

Figure 4: Lung CT Dataset Segmentation Evaluation Metrics - We have trained our model on 100 epochs but after 30 epochs the model was saturated. So we have plotted the metrics Loss(a), Dice Coefficient(b), IoU(c) for 30 epochs to visualize the learning curves.

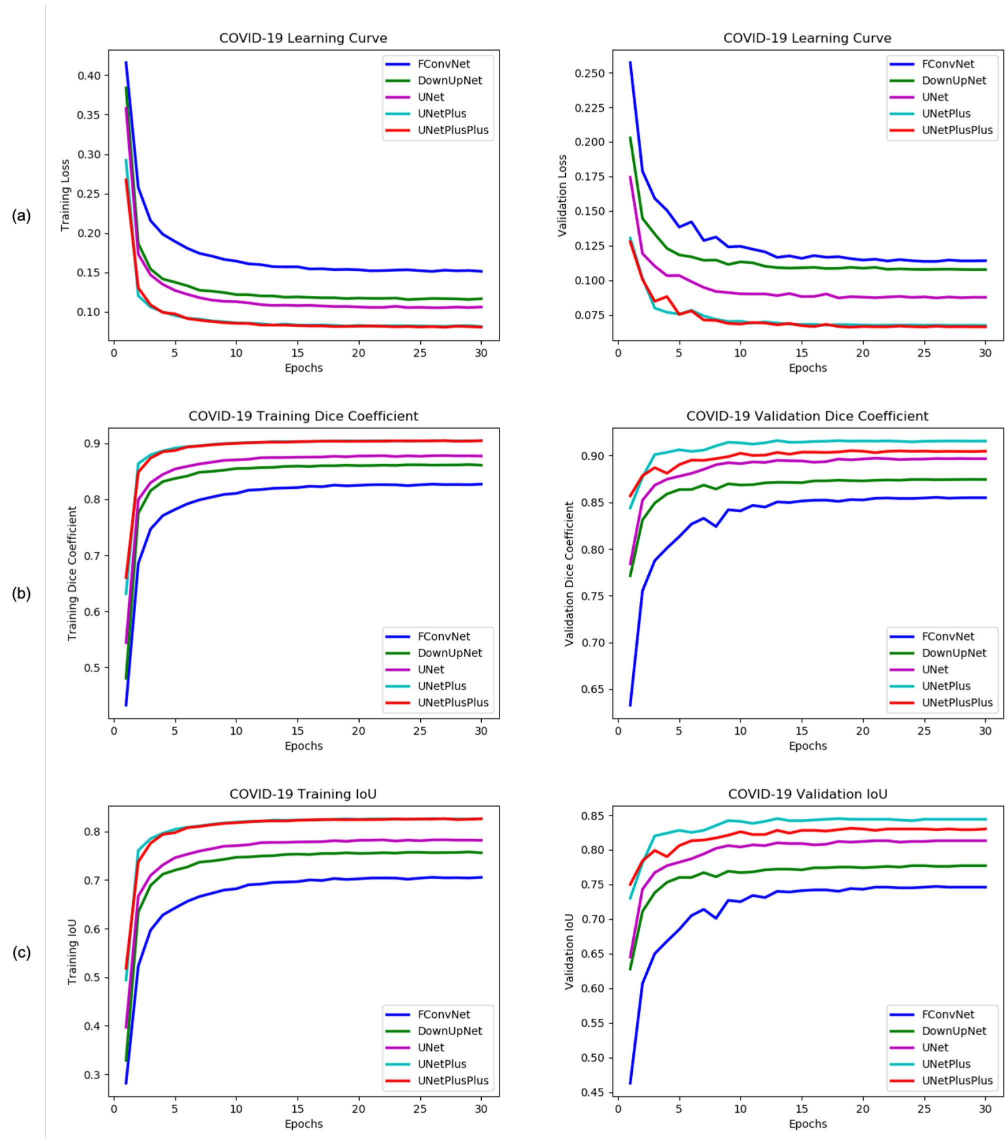

Figure 5: COVID-19 Dataset Segmentation Evaluation Metrics - We have trained our model on 100 epochs but after 30 epochs the model was saturated. So we have plotted the metrics Loss(a), Dice Coefficient(b), IoU(c) for 30 epochs to visualize the learning curves.

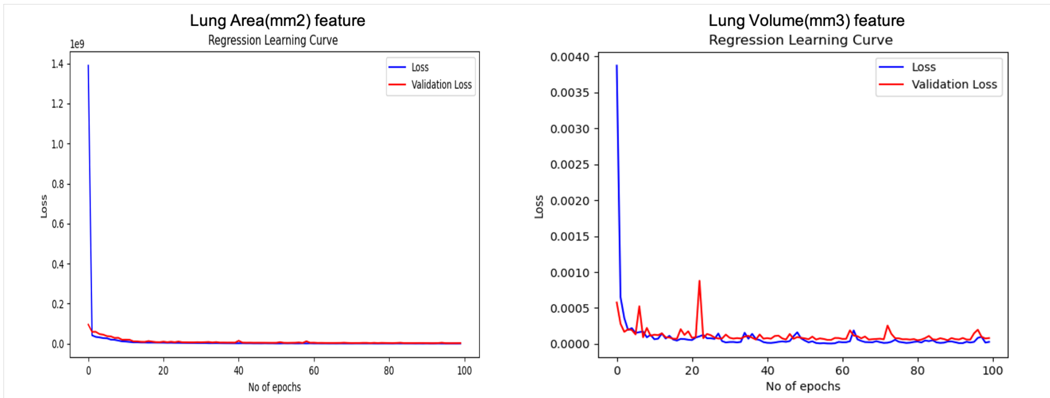

Figure 6: Lung CT Dataset Regression Evaluation Metrics - Loss

6 Conclusion

In this project we have tackled multiple biomedical image analysis problems on two different datasets. We have implemented and compared the performance of five different models for image segmentation tasks. We have also extended our study to solving applications of the segmented data like classification and regression problems. We have benchmarked our results on several state-of-the-art evaluation metrics. From the results we can conclusively state that we are able to obtain satisfactory results.

References

[1] https://www.kaggle.com/yoctoman/shcxr-lung-mask

[2] https://lhncbc.nlm.nih.gov/publication/pub9931

[3] https://www.kaggle.com/kmader/finding-lungs-in-ct-data

[4] https://www.kaggle.com/toregil/a-lung-u-net-in-keras

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[6] Zhou, Zongwei, et al. "Unet++: A nested u-net architecture for medical image segmentation." Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018. 3-11.

[7] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[8] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

[9] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[10] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[11] https://www.mayoclinic.org/diseases-conditions/copd/symptoms-causes/syc-20353679

[12] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5341876/

[13] https://erj.ersjournals.com/content/40/5/1142

[14] https://towardsdatascience.com/understanding-data-science-classification-metrics-in-scikit-l

[15] https://muthu.co/understanding-the-classification-report-in-sklearn/

[16] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html

[17] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

[18] https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb9963

[19] Tustison, N. J., and J. C. Gee. "Introducing Dice, Jaccard, and other label overlap measures to ITK." Insight J 2 (2009).