

Medical Tube Abnormality Detection in Chest X-Rays using Deep Learning

*Thesis submitted to
Indian Institute of Technology Kharagpur
for the award of the degree
of
Bachelor of Technology (Hons.)*

*in
Instrumentation Engineering*

*by
Kumar Tanmay
18IE10038*

Under the supervision of
Prof. Pabitra Mitra



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
April 2022

DECLARATION

I certify that

- a. the work contained in the thesis is original and has been done by me under the guidance of my supervisor;
- b. the work has not been submitted to any other institute for any other degree or diploma;
- c. I have followed the guidelines provided by the Institute in preparing the thesis;
- d. I have conformed to ethical norms and guidelines while writing the thesis;
- e. whenever I have used materials (data, models, figures and text) from other sources, I have given due credit to them by citing them in the text of the thesis, and giving their details in the references, and taken permission from the copyright owners of the sources, whenever necessary.

Kumar Tanmay



Department of Electrical Engineering
Indian Institute of Technology Kharagpur
Kharagpur, India-721302

Certificate

This is to certify that this thesis entitled **Medical Tube Abnormality Detection in Chest X-Rays using Deep Learning** submitted by **Kumar Tanmay** (I8E10038), to the Indian Institute of Technology Kharagpur, is a record of bona fide research work under my supervision and I consider it worthy of consideration for the award of the degree of Bachelors of Technology (Hons.) in Instrumentation Engineering of the Institute.

Prof. Pabitra Mitra

Department of Computer Science and
Engineering
Indian Institute of Technology Kharagpur
India - 721302.

I.I.T. Kharagpur
April 2022

Acknowledgement

I would like to take this opportunity to express my heartiest thanks and immense gratitude to all who supported me and inspired me to come so far and made this journey successful. I would begin by expressing my gratitude to my supervisor, Dr. Pabitra Mitra for his continuing support, valuable feedbacks and important lessons throughout my thesis. I would also like to thank all the Professors of Department of Electrical Engineering and other Departments at IIT Kharagpur, without whose guidance it would not have been possible.

Last but not the least, all this would not have been possible without constant support and love from my mother and father.

Abstract

Serious and life-threatening complications may occur due to the mispositioning of catheters and tubes in patients. Doctors must look at Chest X-Rays manually to verify that the tubes and catheters are in their optimal position. It is a very time-consuming process, and any delay may result in a life-threatening situation for intubated patients. Deep learning algorithms can help us to recognize these abnormal positions automatically and in very little time. The Endotracheal Tube, Nasogastric Tube, Central Venous catheter, and Swan Ganz catheter are some of these medical devices that we are going to focus on in this project. Different anatomical regions like Carina, Gastro-Esophageal junction, Cavo-Atrial junction, Pulmonary-Arteries, Superior Vena Cava, etc., seen on the Chest X-Rays can help us to find the correct positioning of the tube tips. We will be using Convolution neural networks for semantic segmentation of different anatomical structures, tubes, and their tip ends seen on the Chest X-Ray. The segmentation models for these tubes and anatomical structures will be trained and tested on Qure.ai CXR annotated dataset. The relative distance between the centroid of these segmented tube tips and their reference anatomical regions can help to predict whether the tube tip is in its optimal position or not. Also, the segmentation of the ideal tube tip regions annotated on the Chest X-Ray by the radiologists can help us to train their segmentation models which will further help us to compare the distance of the tube tip to be examined from the ideal tip location to find any abnormality in the tube placement.

Contents

Abstract	iv
1 Introduction	1
2 Prior Art	3
3 Scope and Objective	5
4 Binary Classification of Chest X-Ray Images to Detect the Presence and Absence of a Tube or Catheter	6
4.1 Dataset	6
4.2 Data Preprocessing	7
4.2.1 RANZCR Classification Dataset	7
4.2.2 Qure.ai Classification Dataset	7
4.3 Data Augmentation	7
4.4 Methodology	8
4.4.1 Binary Classification	8
4.5 Results	9
4.6 Discussion	9
5 Multi-Label Classification of Chest X-Ray Images to Detect the Presence and Absence of any Tube or Catheter	12
5.1 Dataset	12
5.2 Data Preprocessing	12
5.3 Data Augmentation	13
5.4 Methodology	13
5.4.1 Multi-Label Classification	13
5.5 Results	14
5.6 Discussion	15
6 Semantic Segmentation of Tube Lines and Tube Tips	17
6.1 Dataset	17
6.2 Data Pre-Processing	17

6.3	Data Augmentation	18
6.4	Methodology	18
6.4.1	Semantic Segmentation using UNET	18
6.4.2	Semantic Segmentation using UNET++	18
6.5	Results	21
6.6	Discussion	21
7	Semantic Segmentation Of Anatomical Regions, and Ideal Tip Location	24
7.1	Dataset	25
7.2	Data Pre-Processing	25
7.3	Data Augmentation	26
7.4	Methodology	26
7.5	Results	27
7.6	Discussion	27
8	Abnormal Tube Tip Position Detection in Chest X-Rays using Student-Teacher Approach	33
8.1	Dataset	33
8.2	Proposed Methodology	34
8.3	Results	35
8.4	Discussion	36
9	Abnormality Detection by Comparing the Euclidean Distance between the Actual Tip and Centroid of the Reference Anatomical Structures	38
9.1	Methodology	39
9.2	Results	39
9.3	Discussion	41
10	Abnormality detection by Comparing the Euclidean Distance between the Actual Tip and Ideal Tip	42
10.1	Methodology	42
10.2	Results	42
10.3	Discussion	43
Future Work		45
Conclusion		46
References		48

Chapter 1

Introduction

The classification of biomedical images effectively plays an important role in improving the clinical health care system as well as treatment. The practicing and trained radiologist describes accurately the presence and placement of an ETT (Endotracheal Tube), NGT (Nasogastric Tube), CVC (Central Venous Catheter), SGC (Swan Ganz Catheter) on frontal chest radiographs. However, it takes a lot of time and may sometimes be error-prone [4]. Traditional machine learning techniques like support vector methods (SVMs), in biomedical image classification, have been there for a long time. There are however some following disadvantages of the traditional methods. Their performance for accurate prediction is quite low, time-consuming, and far from the practical standard [11]. Deep Convolutional Neural Network has now been firmly established as a strong and efficient tool in 2-D and 3D medical imaging [7]. Biomedical semantic segmentation plays a vital role in finding small patches of information from medical images. There is a widespread application to separate homogeneous areas which have become the most important and critical component of treatment and disease diagnosis and in the detection of boundaries and segmentation of anatomical structures [10].

Serious and life-threatening complications may occur due to the mispositioning of catheters and tubes in patients. If the ETT tube is inserted into the main stem bronchus, it causes pneumothorax and hyperinflation of one lung and hypoxemia in a non-ventilated lung. It is essential to examine the nasogastric tube tip position using a Chest X-Ray before giving food or any medication through the nasogastric tube to any patient to avoid unnecessary complications. The Chest X-ray may also be useful for the identification of any CVC malposition to help reduce the risk of serious complications such as pericardial tamponade or erosion of the pulmonary vessel wall. The Swan-Ganz catheter must be placed very carefully. In case any misplacement of its tube tip like its entry into the pulmonary artery may lead to various complications like pneumothorax, lung infarction, heart block, balloon perforation, blood thrombosis, catheter knotting, infection, or any damage to the valvular muscle. Doctors must look at Chest X-Rays manually to verify that the tubes and catheters are in their optimal position. It is a very time-consuming process, and any delay may result in a life-threatening situation for intubated patients.

Deep learning algorithms can help us to recognize these abnormal positions automatically and in very little time. This can reduce the delay in the diagnosis and any serious complications arising due to improperly placed tubes.

Chapter 2

Prior Art

Deep learning has recently seen enormous success in challenging problems of computer vision in natural images. This success has prompted a surge of interest in applying CNNs (Convolutional Neural Networks) to medical imaging. The advent of the UNET [15] and UNET++ [9] for Biomedical Segmentation has remarkably improved the performance of deep neural networks in medical imaging. There has been an extensive study of deep neural networks for the assessment of endotracheal tube position [12]. Since the positions of the tubes need to be regularly checked for detecting any misplacement to prevent any serious complications from occurring, computer-aided detection (CAD) methods were generally used to facilitate the detection of ET tubes, NG Tube, CV catheters, SG catheters [23]. Deep neural networks (DNN), most generally the convolutional neural networks (CNNs), have widespread use in changing image classification tasks. These have achieved great and appreciable performance since 2012 [32]. Some research on medical image classification by CNN (convolutional neural networks) has achieved performances better than the human experts. The CheXNet, which is a CNN with 121 layers that have been trained on a dataset containing more than 100,000 frontal chest radiographs (Chest X-ray 14), performed better than the average performance of four radiologists [21].

There have been some previous works related to the prediction of any malpositioning of the tubes and catheters. Lakhani et al. [11] has conducted experiments for evaluating the efficacy of deep convolutional neural networks in frontal chest radiographs and abdominal radiographs for classifying the presence vs absence of the endotracheal tube as well as their low vs normal positions seen on the radiographs. They trained different CNN models like Pre-Trained (on ImageNet) and Untrained AlexNet_U and GoogleLeNet for their analysis. They obtained the best AUC score of 0.809 for low vs normal endotracheal tube position and 0.943 for presence vs absence of this tube.

Frid-Adar et al. [7] has proposed a method for efficient and accurate detection and segmentation of the endo-tracheal tube in chest radiographs using both synthetic and real data. They used a combined CNN architecture for endotracheal tube detection and segmentation in Chest X-Rays. They named this architecture as ETT-Net which is built

from a VGG16 style encoder having two paths from this encoder end one is the decoder of the U-Net for segmentation of this tube and the other to dense layers for detection of the ETT (endotracheal tube) on the Chest X-Rays. They obtained the best AUC score of 0.96 for the presence vs absence of an ETT tube.

Singh et al. [10] has worked on the assessment of the critical feeding tube (nasogastric tube) malpositions on radiographs using deep convolutional neural networks. They trained different DCNN models like Inception V3, ResNet50, DenseNet 121 each pre-trained on ImageNet to classify critical vs noncritical placement of the nasogastric tube on 5475 de-identified HIPAA (Health Insurance Portability and Accountability Act of 1996) compliant frontal view chest and abdominal radiographs. They obtained the best AUC score of 0.87 for critical vs noncritical placement of the tube.

Chapter 3

Scope and Objective

The aim of this project is to develop deep learning algorithms, especially deep convolutional neural networks that can help automatically detect the abnormal positions of the tubes and catheters seen on Chest X-Rays. The following objectives have been set to achieve this aim.

- Classification of Chest X-Rays images to detect the presence and absence of any tube or catheter on Chest X-Rays.
- Localization and Segmentation of Tubes, Anatomical Regions, and Actual Tip Position.
- Detection of any abnormal Tube Tip position.

Chapter 4

Binary Classification of Chest X-Ray Images to Detect the Presence and Absence of a Tube or Catheter

Image classification is the process to categorize and label groups of vectors or pixels in an image on the basis of some specific rules. It involves the feature extraction from the image to observe some patterns in the dataset. Using Artificial Neural Network (Multi-Layer Perceptron) for the image classification task would end up being very computationally expensive since the trainable parameters become enormously large.

The Convolutional Neural Network-based deep neural system has seen an immense growth in the classification task in the medical domain. Almost all of the state-of-the-art methods for image classification involve Convolution Neural Networks. Being an excellent feature extractor, they are being utilized in the classification of medical images which can prevent the complicated and expensive task of feature engineering. Research advancements in computer vision-based deep learning led to a great improvement in the CNN architecture.

4.1 Dataset

Our study involved working on two datasets, the first is the data collected by Qure.ai and the second is the data collected by The Royal Australian and New Zealand College of Radiologists (RANZCR). The Qure.ai dataset comprises 51011 Chest X-Rays having 1400-pixel resolution which are labeled with ETT present, NGT Present, and CVC Present. If the tube is present in the given Chest X-Ray, it is marked as ‘1’ otherwise ‘0’. The RANZCR dataset has 30083 Chest X-Rays each of more than 2000- pixel resolution with the labels marked as abnormal, normal, and borderline condition of the tube

Tube	# Chest X-Rays which have the tube	# Chest X-Rays which does not have the tube
ETT	28993	52424
NGT	22651	58766
CVC	52007	29410
SGC	873	80544

Table 4.1: No. of data samples for each tube in combined Ranzcr and Qure.ai Datasets.

position. These datasets are combined together resulting in 81417 Chest X-Rays. These Chest X-Rays are grayscale images. Table 4.1 shows the distribution of the Chest X-Rays containing the tube and not containing the tube.

4.2 Data Preprocessing

4.2.1 RANZCR Classification Dataset

This dataset contains 30083 Chest X-Rays with 12 columns. The ‘Study Instance ID’ column specifies the unique id of the Chest X-Rays and the ‘Patient ID’ specifies the id of the patient examined. There may be one or more than one Chest X-Rays for each patient. The other columns are defined as ‘ett-normal’, ‘ett-borderline’, ‘ett- abnormal’, ‘ngt -normal’, ‘ngt -borderline’, ‘ngt- abnormal’, ‘cvc -normal’, ‘cvc- borderline’, ‘cvc- abnormal’, ‘swan ganz catheter present’. This dataset is pre-processed to add three columns of ‘ETT Present’, ‘NGT Present’, and ‘CVC Present’. When any of the abnormal, normal, borderline labels are labeled as ‘1’, then the tube is present (marked as ‘1’) or else absent (marked as ‘0’).

4.2.2 Qure.ai Classification Dataset

This dataset contains 51334 Chest X-Rays with 4 columns named ‘Chest Id’, ‘ETT Present’, ‘NGT Present’, and ‘CVC Present’. We merged Qure.ai and RANZCR classification datasets together. This final dataset now contains 81417 Chest X-Rays. This combined dataset is divided into a training set, validation set, and test set in the proportion of 70:20:10. These images are resized to 768x768 pixels resolutions before feeding them to the models.

4.3 Data Augmentation

Because of the lack of sufficient data for training the classifier models, various augmentation techniques are used to solve the problem of overfitting. Random Horizontal Flipping, Random Rotation, Random Brightness, Random Equalization, Gaussian Noise, Gaussian Blur, Median Blur, Motion Blur, CLAHE [35] (Contrast-Limited Adaptive

Histogram Equalization). These images are also normalized with mean=0.5 and standard deviation=0.5.

4.4 Methodology

4.4.1 Binary Classification

Binary classification [29] refers to those classification tasks in which there are two class labels. These class labels can be 'positive' (label = 1) and 'negative' (label = 0). In our case, the presence of the tube is considered as label=1, and absence is considered as label=0. These models have been individually trained for each tube. Their input layer which initially has three channels and 224x224 input dimensions have been modified in such a way that the input layer now contains only 1 channel and 512x512 (or 768x768) input dimensions. The output layer is also modified to a vector of size 1 and the sigmoid activation function is applied to the model logits. After training this classifier, the output neuron predicts the probability of whether the tube is present or not in the given input Chest X-Ray image.

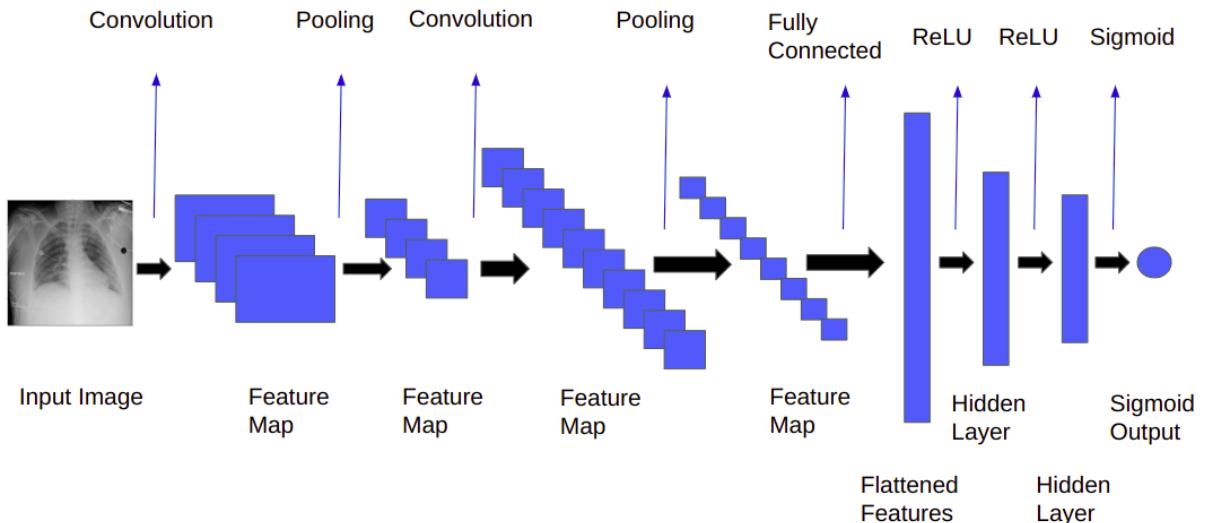


Figure 4.1: Binary Classification using CNN Architecture.

These models are trained from scratch i.e with random weights initially. Also, we have tried some ImageNet Pre-trained models that were also used as a part of experimentation. Weighted Binary Cross-Entropy Loss with different weights for each tube was used as the loss function [1]. The final classification dataset is highly imbalanced. Weighted Binary Cross Entropy is used so that more penalization is done for the wrong prediction of the smaller classes. Adam was used as an optimizer. The weights (β) assigned to ETT, NGT, CVC, and SGC are 1.80, 2.20, 1.00, and 25 respectively. The optimum threshold for prediction for each tube is set to 0.5. Auto Learning Rate Finder helps to find the

initial learning rate for training [33]. We also used a One-Cycle Learning Rate Scheduler to adjust the learning rate after a fixed number of epochs [25]. The probability of positive prediction and Weighted Binary Cross-Entropy Loss function is defined respectively as

$$\hat{p} = \text{sigmoid}(\text{logit}(X)) \quad (4.1)$$

$$WBCE(\hat{p}, p) = -\frac{1}{N} \sum_{i=1}^N (\beta \cdot p \cdot \log(\hat{p}) + (1 - p) \cdot \log(1 - \hat{p})) \quad (4.2)$$

where X is the input to the model, β is the weighing hyper-parameter, p is the probability of the presence of the tube in the input image. \hat{p} is the ground truth label and N is the batch size. These model performances are evaluated using AUC-ROC scores on the test set. Sensitivity is a metric that tells how well a machine learning model can detect positive instances whereas Specificity measures the proportion of true negatives that are correctly identified by the model. The negative and positive predictive values (NPV and PPV respectively) are the proportions of positive and negative results in statistics which are true negative and true positive results, respectively. The Sensitivity, Specificity, Negative Predictive Values (NPV), positive predictive values (PPV), and Jaccard Index (Accuracy) curves are also plotted against various thresholds.

4.5 Results

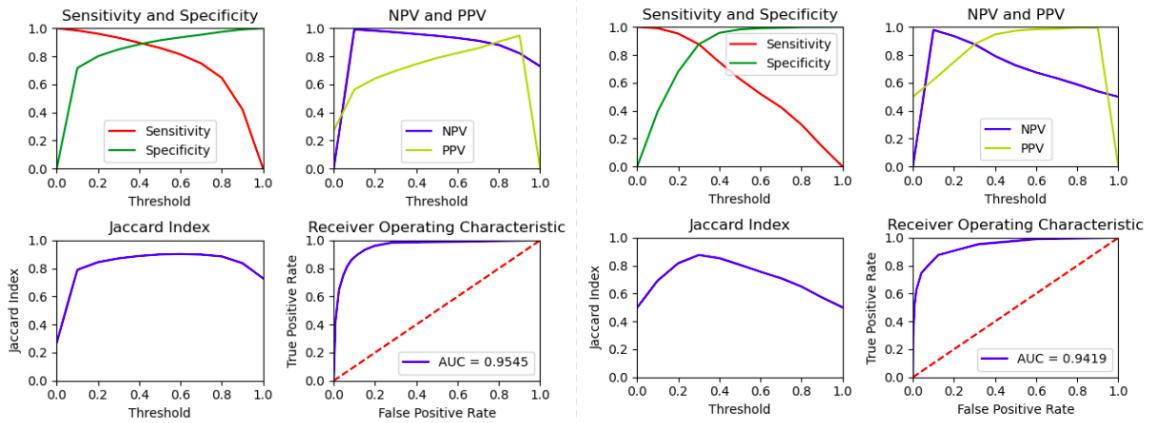
The experiments are performed for binary classification to identify whether a tube is present or not in the Chest X-Ray image. Different CNN models like Resnet50 [6], Semnasnet_050 [28], RegNetx_002 [31], etc., were trained accordingly to classify the frontal chest radiographs. Table 4.2 shows the results for Binary Classification on individual tubes for the respective CNN model. Table 5.1 shows the results of a multi-label classifier for the SEMNasNet-050 model. The Figures (4.2a, 4.2b, 4.2c, 4.2d) shown below contain the Sensitivity and Specificity plot, Negative and Positive Predictive Values (NPV and PPC) plot, Jaccard Index plot against various thresholds, and AUC-ROC Curve for all tube classifiers having SEMNasNet based CNN backbone (best performance model).

4.6 Discussion

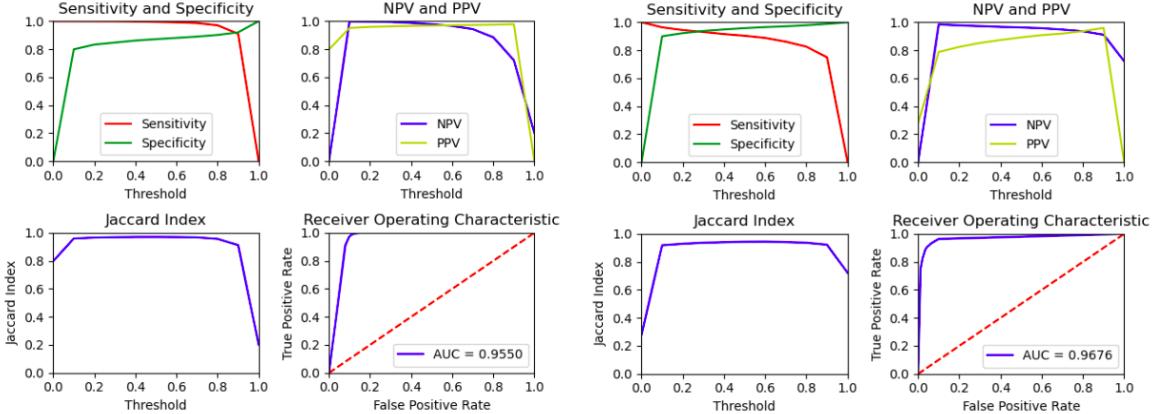
The Binary Classification works best when the SEMNasNet_050 is used as the backbone. We also observed that the model having smaller CNN architecture works better than their larger architectural counterparts. For example, ResNet18 has a higher AUC-ROC score compared to ResNet34 and ResNet50. Similarly, Regnetx_002 performs better than RegNetx_016. The pre-trained version of the model performs poorly compared to the untrained version. This may be due to the different data distributions in the ImageNet

CNN Models/Tube	ETT	NGT	CVC	SGC
ResNet 18	0.9236	0.9117	0.9289	0.9414
Resnet 34	0.9102	0.8910	0.9196	0.9388
ResNet 50	0.8913	0.8823	0.9045	0.9243
EfficientNet_b0	0.9280	0.9145	0.9374	0.9482
RegNetx-002	0.9312	0.9213	0.9368	0.9588
RegNetx-016	0.9002	0.8919	0.9102	0.9227
SEMNasNet_050	0.9545	0.9419	0.9550	0.9672
ResNeXt-50(32x4d)	0.9328	0.9302	0.9392	0.9473
Pre-trained(ImageNet) ResNet50	0.8412	0.8239	0.8610	0.8823

Table 4.2: AUC-ROC Scores of Tube Binary Classifier Model with different CNN backbones.



(a) Plots for ETT Binary classifier with SEM-NasNet backbone. (b) Plots for NGT Binary classifier with SEM-NasNet backbone.



(c) Plots for CVC Binary classifier with SEM-NasNet backbone. (d) Plots for SGC Binary classifier with SEM-NasNet backbone.

Figure 4.2: Plots for various Binary classifiers with SEMNasNet backbone.

dataset which have real-life RGB images which are completely different from grayscale Chest X-Rays. The important low-level features extracted by the pre-trained network may not be suitable for application in the Chest X-Rays datasets. These images are resized to 512x512 pixels resolutions before feeding them to the models. The results for

SGC classifier is better than others because of its distinctive loop structure which helps the model to identify it easily. In some of the Chest X-Rays, we observed that there are some cases where the CVC looks similar to ETT hence making it difficult for model to identify them correctly resulting in more false positive and false negative cases.

Chapter 5

Multi-Label Classification of Chest X-Ray Images to Detect the Presence and Absence of any Tube or Catheter

Almost all of the state-of-the-art methods for image classification involve Convolution Neural Networks. Being an excellent feature extractor, they are being utilized in the classification of medical images which can prevent the complicated and expensive task of feature engineering. Research advancements in computer vision-based deep learning led to a great improvement in the CNN architecture.

5.1 Dataset

The same datasets as discussed previously in Binary Classification is utilized for multi-label classification. The Qure.ai dataset comprises 51011 Chest X-Rays having 1400-pixel resolution which are labeled with ETT present, NGT Present, and CVC Present. If the tube is present in the given Chest X-Ray, it is marked as ‘1’ otherwise ‘0’. The RANZCR dataset has 30083 Chest X-Rays each of more than 2000- pixel resolution with the labels marked as abnormal, normal, and borderline condition of the tube position. These datasets are combined together resulting in 81417 Chest X-Rays. These Chest X-Rays are grayscale images.

5.2 Data Preprocessing

The dataset is preprocessed in the Binary Classification stage for adding four columns of ‘ETT Present’, ‘NGT Present’, ‘CVC Present’, and ‘Swan Ganz Catheter’. This preprocessed dataset will be used here as well. When any of the abnormal, normal,

borderline labels are labeled as ‘1’, then the tube is present (marked as ‘1’) or else absent (marked as ‘0’) for any tube. This combined dataset is divided into a training set, validation set, and test set in the proportion of 70:20:10. These images are resized to 768x768 pixels resolutions before feeding them to the models.

5.3 Data Augmentation

Because of the lack of sufficient data for training the classifier models, various augmentation techniques are used to solve the problem of overfitting. Random Horizontal Flipping, Random Rotation, Random Brightness, Random Equalization, Gaussian Noise, Gaussian Blur, Median Blur, Motion Blur, CLAHE [35] (Contrast-Limited Adaptive Histogram Equalization). These images are also normalized with mean=0.5 and standard deviation=0.5.

5.4 Methodology

5.4.1 Multi-Label Classification

Multi-Label Classification is the task of assigning multiple labels to each instance simultaneously [19]. Mathematically, it is the problem of finding a model that maps inputs x to binary vectors y (assigning a value of 0 or 1 for each label in y). Since we have four different classes possible for each instance, the output layer of the CNN model is modified to four neurons where the i -th neuron signifies the probability of the presence of the i -th tube in the Chest X-Rays.

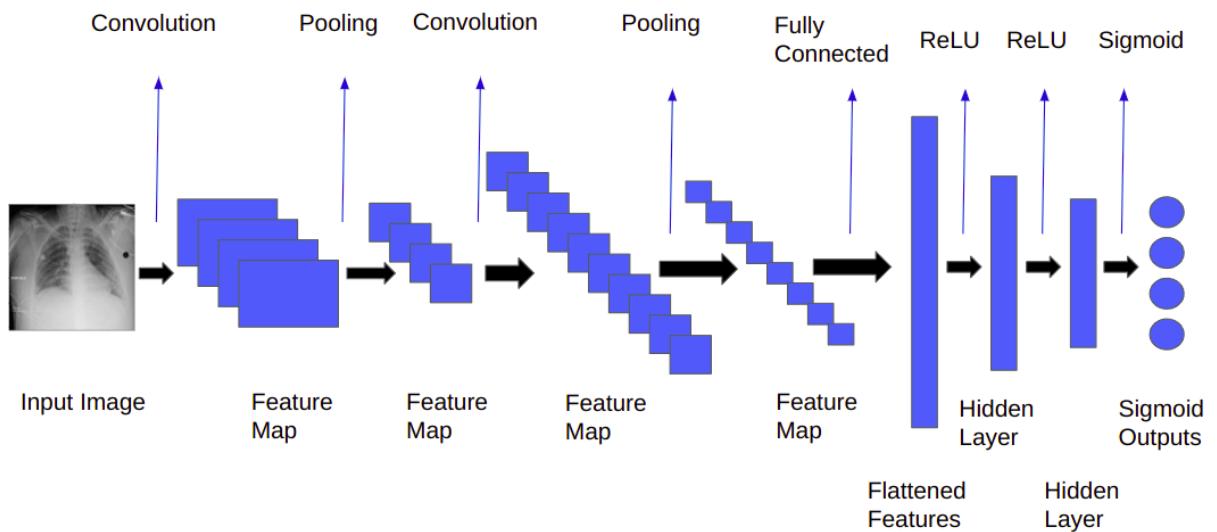


Figure 5.1: Multi-Label Classification using CNN Architecture.

Similar to the Binary Classification task, their input layer which initially has three channels and 224x224 input dimensions has been modified in such a way that the input layer now contains only 1 channel and 512x512 or (768x768) input dimensions. The output layer is modified to a vector of size 4 and the sigmoid activation function is applied to each of the 4 neurons. After training this classifier, the output layer predicts the probabilities of the presence of each tube in the given input Chest X-Ray image. These models are trained from scratch i.e with random weights initially. Weighted Binary Cross-Entropy Loss with different weights for each tube was used as the loss function. The final classification dataset is highly imbalanced. Weighted Binary Cross Entropy is used so that more penalization is done for the wrong prediction of the smaller classes. Adam was used as an optimizer. The weights (β) assigned to ETT, NGT, CVC, and SGC are 1.80, 2.20, 1.00, and 25 respectively. The optimum threshold for predictions of each tube is set to 0.5. Auto Learning Rate Finder helps to find the initial learning rate for training. We also used a One-Cycle Learning Rate Scheduler to adjust the learning rate after a fixed number of epochs. The Weighted Binary Cross-Entropy Loss function is defined as

$$\hat{p}_j = \text{sigmoid}(\text{logit}(X)) \text{ (for jth neuron)} \quad (5.1)$$

$$\text{WBCE}(\hat{p}, p) = -\frac{1}{4N} \sum_{i=1}^N \sum_{j=1}^4 (\beta_j \cdot p_j \cdot \log(\hat{p}_j) + (1 - p_j) \cdot \log(1 - \hat{p}_j)) \quad (5.2)$$

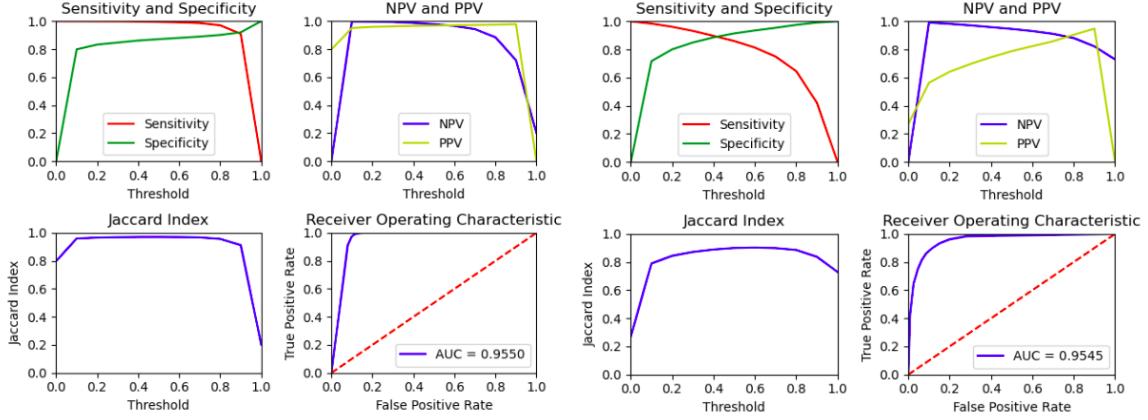
where X is the input, β_j is the weighing hyper-parameter, p_j is the probability of the presence of the jth tube in the input image. \hat{p}_j is the ground truth label for that jth tube and N is the batch size. These model performances are evaluated using AUC-ROC Scores on the test set.

5.5 Results

The experiments are performed for multi-label classification to identify whether any tube is present or not in the Chest X-Ray Image. The experiments are performed on SEM-NASNET_50 because of its good performance in Binary Classification for all tubes and are trained accordingly to classify the frontal chest radiographs. Table 5.1 shows the results of a multi-label classifier for the SEMnasNet-050 model. The Figures (5.2a, 5.2b, 5.2c, 5.2d) shown below contain the Sensitivity and Specificity plot, Negative and Positive Predictive Values (NPV and PPC) plot, Jaccard Index plot against various thresholds, and AUC-ROC Curve for all tube classifiers having SEMnasNet based CNN backbone.

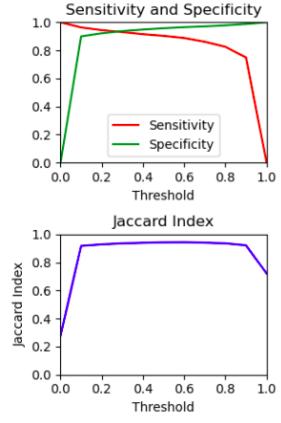
Tube	ETT	NGT	CVC	SGC
AUC-ROC	0.9550	0.9545	0.9676	0.9851

Table 5.1: AUC-ROC Scores of Multi-Label Classifier model having SemNasNet_50 backbone.

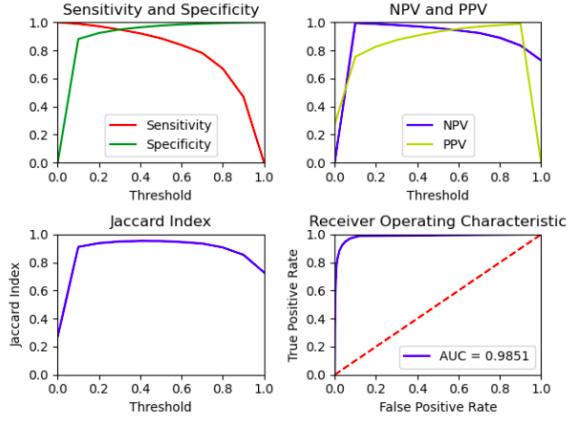


(a) Plots for Multi-Label classifier with SEM-NasNet backbone for ETT.

(b) Plots for Multi-Label classifier with SEM-NasNet backbone for NGT.



(c) Plots for Multi-Label classifier with SEM-NasNet backbone for CVC.



(d) Plots for Multi-Label classifier with SEM-NasNet backbone for SGC.

Figure 5.2: Plots for Multi-Label classifiers with SEMNasNet backbone.

5.6 Discussion

As evident in the Table 5.1 and Table 4.2, the multi-label classifier works better than the individual tube classifiers. There is an average improvement of 2-3 % in performance for multi-label classifiers compared to binary classifiers. The Multi-label classifier having shared backbones for all tubes seemed to perform better than the individual classifiers for each tube. There may be more than one tube present in the Chest X-Ray. The multi-label classifier can capture the distinctive features of each tube which help them to distinguish between them. Hence it may be the reason behind its better performance over single tube binary classifiers. Also, only a single multi-label model is required compared

to four different models for individual tube. This also reduces the computational cost and inference time.

Chapter 6

Semantic Segmentation of Tube Lines and Tube Tips

Semantic image segmentation is a computer vision task of clustering regions or parts of an image together belonging to the same object class. It is basically a form of pixel-level classification as each pixel in an image is classified according to a category/class [27]. The output is the binary mask image where the masked regions belonged to a particular category. Our task is to individually train the segmentation model which can produce a masked image of a particular tube/tip/anatomical structure.

6.1 Dataset

The tube lines and actual tube tips seen on the frontal chest radiographs are carefully annotated by a group of radiologists. RANZCR annotated dataset contains 18000 tube line annotations done for 9025 Chest X-Rays. The Qure.ai annotated dataset contains annotations for tube lines and actual tube tips. This annotated dataset will be required for the semantic segmentation to generate the image masks which will be used for the tube tip abnormality detection. The data samples in the combined Qure.ai and RANZCR datasets are as given in the Table 6.1.

6.2 Data Pre-Processing

The annotations for the tube-lines in the Ranzcr Dataset are given in the form of X and Y coordinates of points. These X and Y coordinates denote a point in the tube line in Chest X-Rays. A corresponding binary masked Image is created by joining all

ETT	NGT	CVC	SGC
12137	9663	17287	152

Table 6.1: Number of Annotations against each Tube.

points in the tube lines. Another binary image is created for the tube tip which is the extreme point of the tube line closer to the central vertical line of the image. The Qure.ai dataset already contains the binary mask images of the tube lines, tips, and anatomical regions for the corresponding Chest X-Rays. After the required binary mask images are created, both the datasets are merged and divided into training, validation, and test sets in proportion 70:20:10. These binary images are going to be used as labels for training the segmentation models.

6.3 Data Augmentation

Because of the lack of sufficient data for training the segmentation models, various augmentation techniques are used to solve the problem of overfitting. Random Horizontal Flipping, Random Rotation, Random Brightness, Random Equalization, Gaussian Noising, Gaussian Blur, Median Blur, Motion Blur, CLAHE [20] (Contrast-Limited Adaptive Histogram Equalization). These images are also normalized with mean=0.5 and standard deviation=0.5.

6.4 Methodology

6.4.1 Semantic Segmentation using UNET

Olaf Ronneberger et al. [21] developed UNET (Figure 6.1) for BioMedical Segmentation in images. The UNET architecture has two pathways. First is the encoder path (contraction path) which captures the context in the input image. Its encoder is basically a traditional stack containing convolutional and max pooling layers. Second is the decoder path (symmetric expanding path) which helps to enable precise localization using transposed convolutions. The skip pathways from encoder maps to symmetric maps at the decoder helps the decoder to extract more information from the encoder which gets lost while downsampling to a feature vector. Hence it is an end-to-end fully convolutional network (FCN), as it only contains Convolutional layers and does not contain any Dense layer because of which it can accept images of any size.

6.4.2 Semantic Segmentation using UNET++

The UNET++ [35] is the modified version of UNET. The aim of UNET++ (Figure 6.2) is to improve segmentation accuracy compared to the UNET. This can be done by adding a Dense block and convolution layers between the encoder and decoder. UNet++ has three additions to the original U-Net:

- Redesigned Skip Pathways (shown in green): The purpose of these convolutions layers in skip pathways is to reduce the semantic gap between the feature maps of

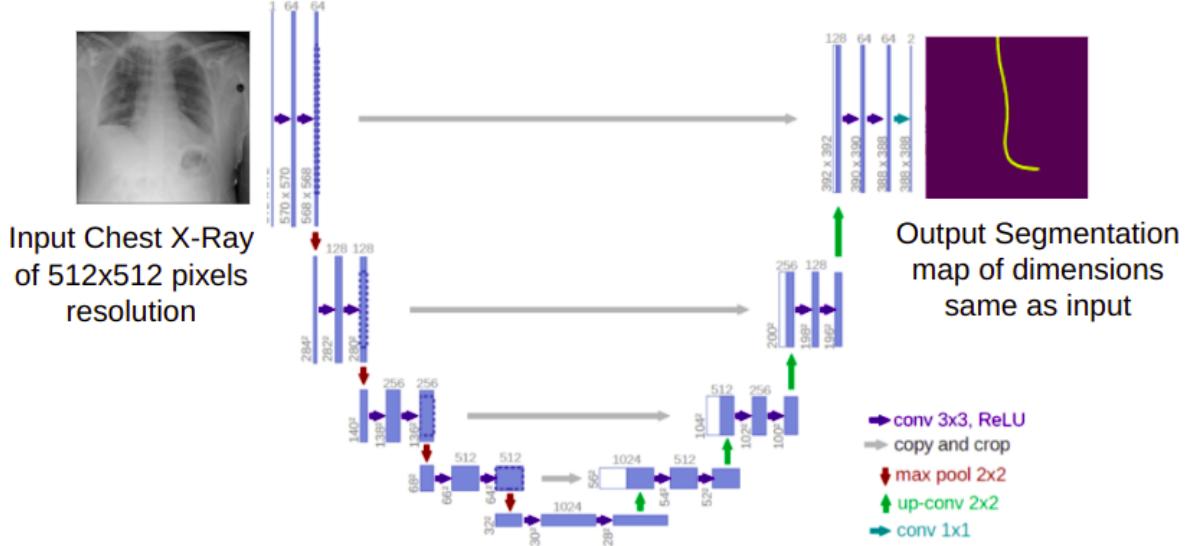


Figure 6.1: UNET Architecture.

the encoder and decoder networks. Thus making it easier to optimize better.

- Dense Skip Connections (shown in blue): Dense skip connections help to accumulate feature maps so that they can arrive at the current node because of the dense convolution block along each skip pathway. This helps in the generation of feature maps having a full resolution at various semantic levels.
- Deep Supervision (shown in red) : This is added in order to prune the model to manage the model complexity. This can help to provide a balance between performance and speed (inference time).

The following Segmentation task involves:

- **Medical Tube Localization and Segmentation on Chest X-Rays.** As discussed before, UNET++ and UNET are extremely useful CNN architectures for biomedical semantic segmentation. Different UNET++ and UNET-based Segmentation models with three different CNN backbones were trained to generate the segmentation tube masks as output for each Chest X-Rays. These Segmentation Tube masks are the binary images in which each pixel of the image is classified either as ‘1’ if the pixel is in the tube line region or ‘0’ if it is a background pixel.
- **Segmentation of Actual Tip position.** Similarly, to obtain the segmentation masks for each ideal tube tip and actual tube tip, different UNET++-based segmentation models with three different backbones were trained to generate the segmentation tip masks.

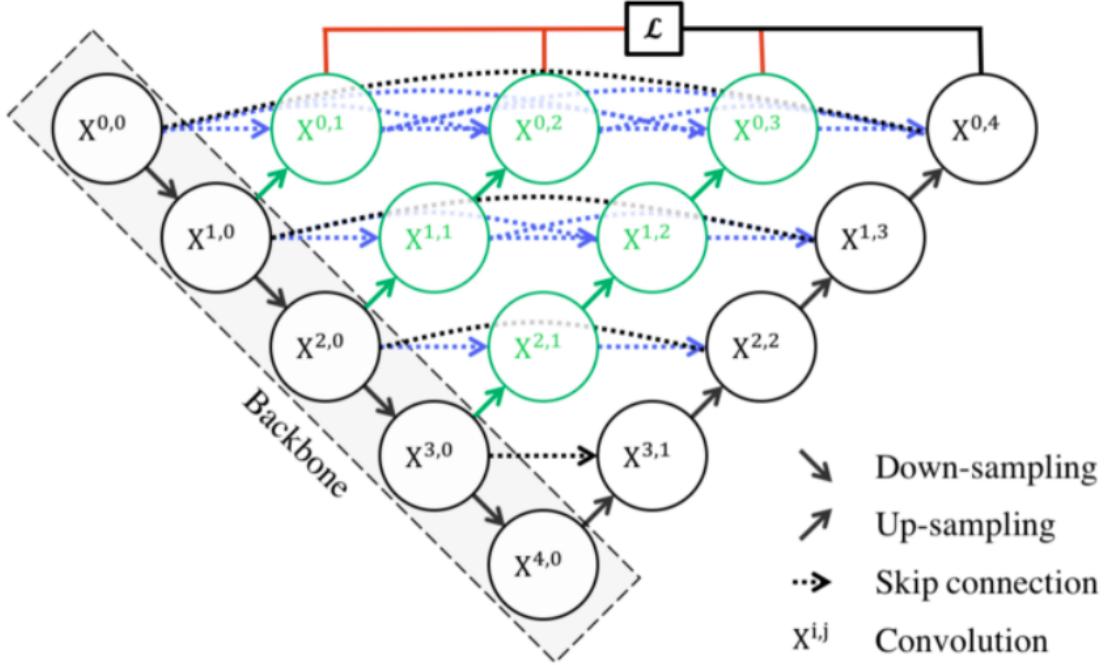


Figure 6.2: Architecture of UNET++.

All the segmentation models are trained using Combo loss [18] which is defined as a weighted sum of modified binary cross-entropy and Dice loss. It leverages the flexibility of Dice loss [8] to solve the problem of class imbalance and also use cross-entropy for curve smoothing. It's defined as

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1}$$

$$L_{m-bce} = -\frac{1}{N} \sum_i \beta(y - \log(\hat{y})) + (1 - \beta)(1 - y) \log(1 - \hat{y}) \quad (6.1)$$

$$CL(y, \hat{y}) = \alpha L_{m-bce} - (1 - \alpha)DL(y, \hat{y})$$

where β is the weighing hyper-parameter for binary cross-entropy, α is the weighing hyperparameter for the combo loss, \hat{p} is the probability of prediction, y is the ground truth label of the presence of the pixel in the required region in Chest X-Rays, N is the sample size. Their input layer which initially has three channels and 224x224 input dimensions have been modified in such a way that the input layer now contains only 1 channel and 512x512 input dimensions. The output is modified to a map of the equivalent dimensions as that of the input. The sigmoid activation function is applied to the output map to convert the values into probabilities. After training this classifier, the output layer predicts the probabilities of the presence of pixels in the required region in the given input Chest X-Ray image. These models are trained from scratch i.e with random weights initially. As defined above, Combo Loss with different weights for each tube, tip,

Tube Line/Tip	Model	Average IoU Score	Best IoU Score
ETT	ResNest14d	0.64	0.90
NGT	ResNest14d	0.61	0.88
CVC	ResNest14d	0.68	0.91
SGC	EfficientNet_b0	0.54	0.77
ETT Tip	EfficientNet_b0	0.22	0.42
NGT Tip	RegNety_002	0.16	0.41
CVC Tip	EfficientNet_b0	0.32	0.42
SGC Tip	EfficientNet_b0	0.12	0.31

Table 6.2: Average IoU Scores of different Segmentation models having UNET architecture.

and anatomical structure is used as the loss function. Since the pixels in these regions are very less compared to that in the background, weights are defined as $\beta = 2$ or tube lines and anatomical structures and $\beta = 10$ for tube tips. This adds penalization for the wrong prediction of the foreground pixels which are smaller in number compared to background pixels. α is set to 0.5. Adam was used as an optimizer with default parameters. The optimum threshold for each pixel’s prediction is set to 0.5. Auto Learning Rate Finder helps to find the initial learning rate for training. We also used a One-Cycle Learning Rate Scheduler to adjust the learning rate after a fixed number of epochs.

6.5 Results

The UNET and UNET++-based segmentation models are trained for each tube line and actual tube tips. ResNest14d [34], Regnet-002 [31], EfficientNet_b0 [27] are the CNN backbones which are used. Table 6.2 shows the average Intersection over Union (IoU) for each tube and their tip seen in Chest X-Ray using UNET segmentation architecture with the best CNN backbone. Table 6.3 shows the average Intersection over Union (IoU) for each tube and their tip seen in Chest X-Ray using UNET++ segmentation architecture with the best CNN backbone. Figures (6.3, 6.4, 6.5, 6.6) show the results for each tube segmentation models having UNET++ Architecture with the best CNN backbone mentioned in Table 6.3. These figures show the Chest X-Ray(left), Actual Mask of the Tube (middle) and Predicted Mask of the Tube (right).

6.6 Discussion

The UNET++ architecture-based models perform better than that of UNET Architecture. The mean IOU scores for the segmentation models for tube lines lie well above 0.6. The mean IoU scores for segmentation models for predicting tube tips lie between 0.12 to 0.35. Although the predicted mask region and actual mask region lie close to each other, they may not be overlapping much resulting in a low IoU score. Hence Euclidean

Tube Line/Tip	Model	Average IoU Score	Best IoU Score
ETT	ResNest14d	0.64	0.90
NGT	ResNest14d	0.61	0.88
CVC	ResNest14d	0.68	0.91
SGC	EfficientNet_b0	0.54	0.77
ETT Tip	EfficientNet_b0	0.22	0.42
NGT Tip	RegNety_002	0.16	0.41
CVC Tip	EfficientNet_b0	0.32	0.42
SGC Tip	EfficientNet_b0	0.12	0.31

Table 6.3: Average IoU Scores of different Segmentation models having UNET++ architecture.

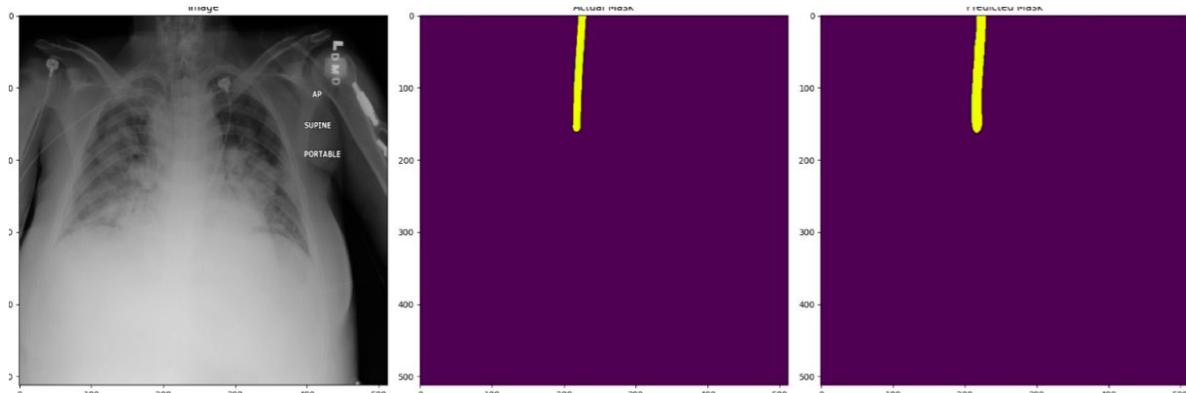


Figure 6.3: Chest X-Ray (left), Actual Mask (middle) and Predicted Mask (right) - ETT segmentation using UNET++ architecture with ResNest14d backbone.

distance between their respective centroids can be used to evaluate the model performance in a better way. The SGC segmentation model show lower average IoU scores as the annotated dataset for SGC is not sufficient to train the model and hence can lead to overfitting problems.

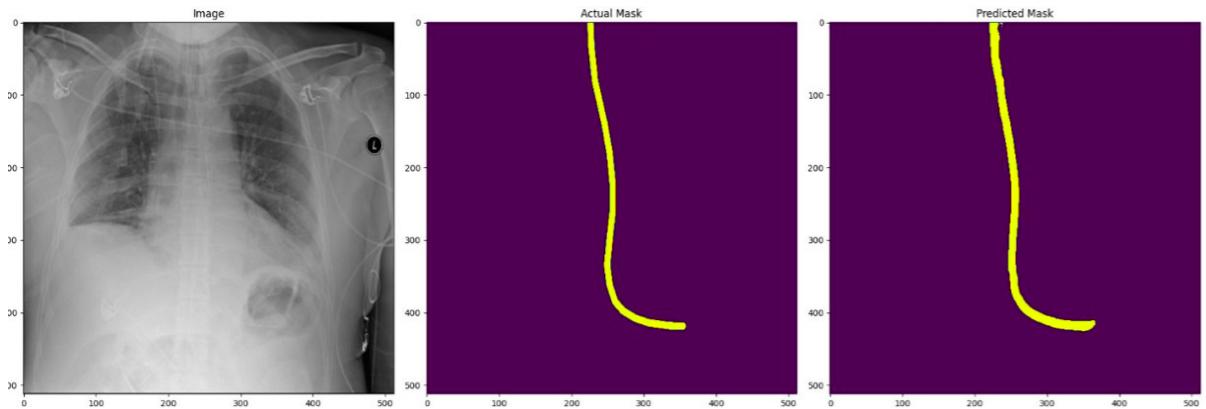


Figure 6.4: Chest X-Ray (left), Actual Mask (middle) and Predicted Mask (right) - NGT segmentation using UNET++ architecture with ResNest14d backbone.

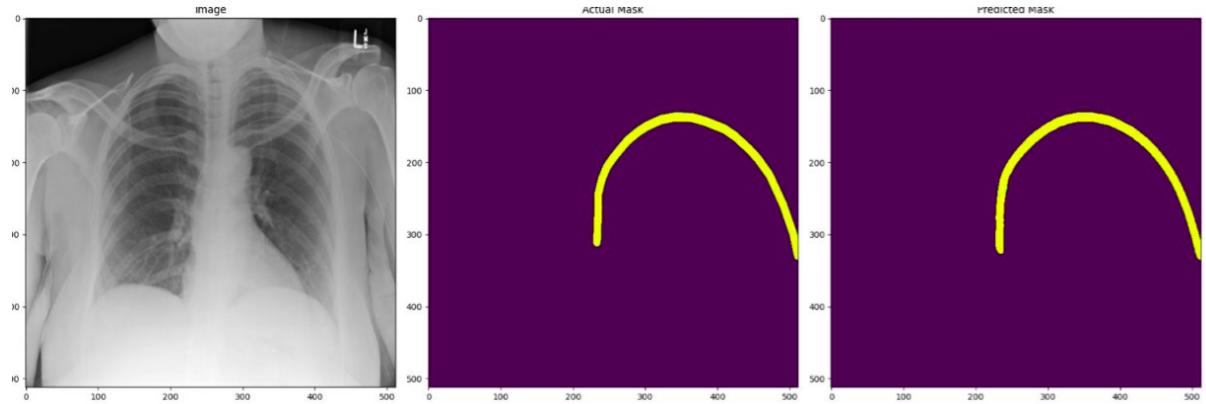


Figure 6.5: Chest X-Ray (left), Actual Mask (middle) and Predicted Mask (right) - CVC segmentation using UNET++ architecture with ResNest14d backbone.

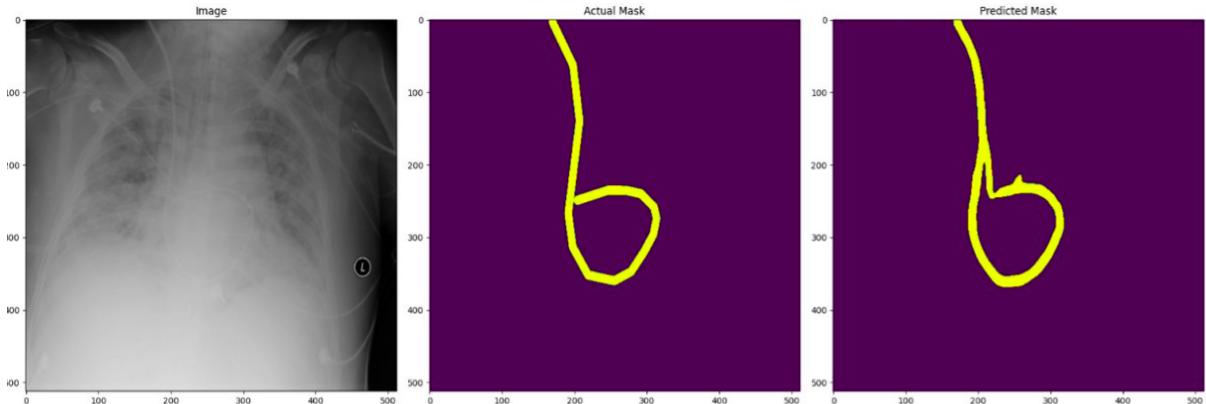


Figure 6.6: Chest X-Ray (left), Actual Mask (middle) and Predicted Mask (right) - SGC segmentation using UNET++ architecture with EfficientNet-b0 backbone.

Chapter 7

Semantic Segmentation Of Anatomical Regions, and Ideal Tip Location

The Chest Anatomical structures/regions are those areas that are defined by the landmarks provided by evident structures that can be visible in Chest X-Rays. These regions are often useful to the radiologists for the prediction of any abnormalities and diseases while analyzing the Chest X-Rays. Some of the most common regions visible in Chest X-Rays are Carina, Heart, Gastro-Esophageal Junction, Cavo-atrial Junction, Lungs, Thorax, Bones, etc.

The following anatomical structures are going to be useful for predicting the abnormality in Tube tip positions:

- Carina or Tracheal Bifurcation is a ridge at the base of the windpipe or trachea separating the openings of the left and right main bronchi. This region acts as a reference point for detecting both the ETT and CVC tip malpositions. [24]
- Gastro-Esophageal Junction is an anatomic region/Junction lying between the proximal stomach and the distal esophagus. This region acts as a reference region for the detection of NGT Tube Tip malposition [3].
- Superior Vena-Cava is a large, valveless vein located in the thorax, more specifically, in the anterior right, superior to the mediastinum that carries venous blood from the upper half of the body and returns it to the right atrium [14].
- Pulmonary arteries are the arteries that carry deoxygenated blood from the right side of the heart to the lungs during pulmonary circulation. The main pulmonary artery or pulmonary trunk gets divided after exiting the bottom right ventricle of the Heart. The right pulmonary artery and the left pulmonary artery go to the right lung and left lung respectively [2].

- Pulmonary Hilum is located on the medial aspect of both lungs. It is the anatomical region that is hard to find on Chest X-Rays [17].
- Mediastinum Margin is located within the thorax. It is encapsulated on the right and left by pleurae and surrounded by the lungs to the sides, the chest wall in front, and the spine at the back [22].
- The Cavo-Atrial Junction is the junction between the superior border of the right atrium and the right lateral border of the superior vena cava. It is an important landmark to recognize as it marks an optimum site for CVC tip placement [26].

UNET++ and UNET are extremely useful CNN architectures for biomedical semantic segmentation. Different UNET++ and UNET-based Segmentation models with three different CNN backbones were trained to generate the segmentation masks as output for each Chest X-Rays. These Segmentation masks are the binary images in which each pixel of the image is classified either as ‘1’ if the pixel is in the anatomical region/ ideal tip region or ‘0’ if it is a background pixel.

7.1 Dataset

The anatomical structures like carina, Cavo-atrial junction, etc. and ideal tip Target seen on the frontal chest radiographs are carefully annotated by a group of radiologists. The Qure.ai annotated dataset contains annotations for ideal tube tip region for all tubes, Carina (Tracheal Bifurcation), Heart, Superior Vena Cava, Cavo-Atrial Junction, Gastro-Esophageal Junction, Left Pulmonary Artery, Right Pulmonary Artery, Mediastinal Margin, Pulmonary Hilum. This annotated dataset will be required for the semantic segmentation to generate the image masks which will be used for the tube tip abnormality detection. The data samples in the Qure.ai datasets are as given in Table 7.1.

Carina	GE Junction	Cavo-Atrium Junction	Pulmonary Hilum	Superior Vena Cava	Heart	Left Pulmonary Artery	Right Pulmonary Artery
8389	445	589	325	567	5301	672	892

Table 7.1: Number of annotated data for each anatomical regions.

7.2 Data Pre-Processing

The Qure.ai dataset already contains the binary mask images of the tube lines, tips, and anatomical regions for the corresponding Chest X-Rays. After the required binary mask images are created, this dataset is merged and divided into training, validation, and test sets in proportion 70:20:10. These binary images are going to be used as labels for training the segmentation models.

7.3 Data Augmentation

Because of the lack of sufficient data for training the segmentation models, various augmentation techniques are used to solve the problem of overfitting. Random Horizontal Flipping, Random Rotation, Random Brightness, Random Equalization, Gaussian Noise, Gaussian Blur, Median Blur, Motion Blur, CLAHE [35] (Contrast-Limited Adaptive Histogram Equalization). These images are also normalized with mean=0.5 and standard deviation=0.5.

7.4 Methodology

All the segmentation models are trained using Combo loss [18] which is defined as a weighted sum of modified binary cross-entropy and Dice loss. It leverages the flexibility of Dice loss to solve the problem of class imbalance and also use cross-entropy for curve smoothing. It's defined as

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1}$$
$$L_{m-bce} = -\frac{1}{N} \sum_i \beta(y - \log(\hat{y})) + (1 - \beta)(1 - y) \log(1 - \hat{y}) \quad (7.1)$$
$$CL(y, \hat{y}) = \alpha L_{m-bce} - (1 - \alpha) DL(y, \hat{y})$$

where β is the weighing hyper-parameter for binary cross-entropy, α is the weighing hyperparameter for the combo loss, \hat{p} is the probability of prediction, y is the ground truth label of the presence of the pixel in the required region in Chest X-Rays, N is the sample size. Their input layer which initially has three channels and 224x224 input dimensions have been modified in such a way that the input layer now contains only 1 channel and 512x512 input dimensions. The output is modified to a map of the equivalent dimensions as that of the input. The sigmoid activation function is applied to the output map to convert the values into probabilities. After training this classifier, the output layer predicts the probabilities of the presence of pixels in the required region in the given input Chest X-Ray image. These models are trained from scratch i.e with random weights initially. As defined above, Combo Loss with different weights for each anatomical structure is used as the loss function. Since the pixels in these regions are very less compared to that in the background, weights are defined as $\beta = 2$ for anatomical regions and anatomical structures and $\beta = 10$ for ideal tube tip location. This adds penalization for the wrong prediction of the foreground pixels which are smaller in number compared to background pixels. α is set to 0.5. Adam was used as an optimizer with default parameters. The optimum threshold for each pixel's prediction is set to 0.5. Auto Learning Rate Finder helps to find the initial learning rate for training. We also used a One-Cycle Learning Rate Scheduler to adjust the learning rate after a fixed number of epochs.

Tip/Anatomical region	Model	Average IoU Score	Best IoU Score
Ideal ETT Tip	ResNest14d	0.34	0.47
Ideal NGT Tip	ResNest14d	0.28	0.57
Ideal CVC Tip	ResNest14d	0.31	0.52
Ideal SGC Tip	EfficientNet_b0	0.18	0.45
Carina	EfficientNet_b0	0.56	0.87
Heart	RegNety_002	0.91	0.99
GE-Junction	EfficientNet_b0	0.08	0.28
Left PA	EfficientNet_b0	0.49	0.78
Right PA	EfficientNet_b0	0.51	0.82
Mediastinal Margin	EfficientNet_b0	0.47	0.68
Superior Vena Cava	RegNety_002	0.59	0.87
Cavo-atrial Junction	EfficientNet_b0	0.48	0.72
Pulmonary Hilum	EfficientNet_b0	0.28	0.47

Table 7.2: Average IoU Scores of Different Segmentation Models.

Similarly, to obtain the segmentation masks for each anatomical structure like Carina, Cavo-Atrial Junction, Superior Vena Cava, Heart, Pulmonary Arteries, Gastro-Esophageal Junction, Mediastinal Margin, different UNET++ based Segmentation models with different backbones were trained to generate the segmentation masks for each of these structures.

7.5 Results

The UNET++-based segmentation models are trained for these anatomical structures. Regnet-002 [31], ResNest14d [34] , EfficientNet_b0 [27] are the CNN backbones which are used. Table 7.2 shows the average Intersection over Union (IoU) for each anatomical regions and ideal tip location seen in Chest X-Ray using UNET++ segmentation architecture with the best CNN backbone. Figures (7.1 - 7.12) show the results for each their respective segmentation models having UNET++ Architecture with the best CNN backbone mentioned in Table 7.2. These figures shows the Chest X-Ray(left), Actual Mask of the Tube (middle) and Predicted Mask of the Tube (right). Table 7.2 shows the results for each required structure such as Carina, Heart, GE-Junction (Gastro-Esophageal Junction), Left and Right PA (Pulmonary Artery), Mediastinal Margin, Superior Vena Cava, Cavo-atrial Junction, and Ideal Tip locations seen in Chest X-Ray.

7.6 Discussion

The mean IOU scores for the segmentation models for some (large) anatomical regions lie well above 0.6. The mean IoU scores for segmentation models for predicting tube tip and small anatomical region lie between 0.07 to 0.5. Although the predicted mask region and

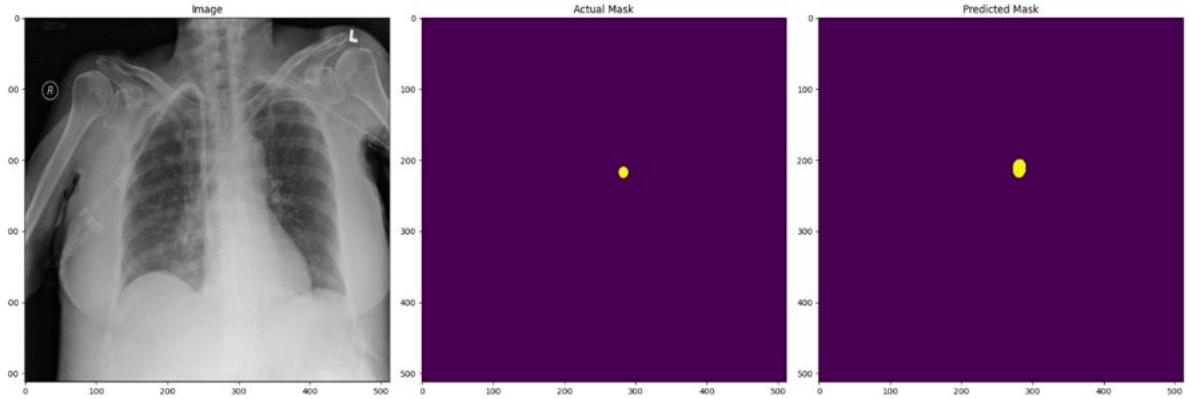


Figure 7.1: Actual (middle) and Predicted Mask (right) of Carina predicted using UNET++ Architecture with EfficientNet-b0 backbone.

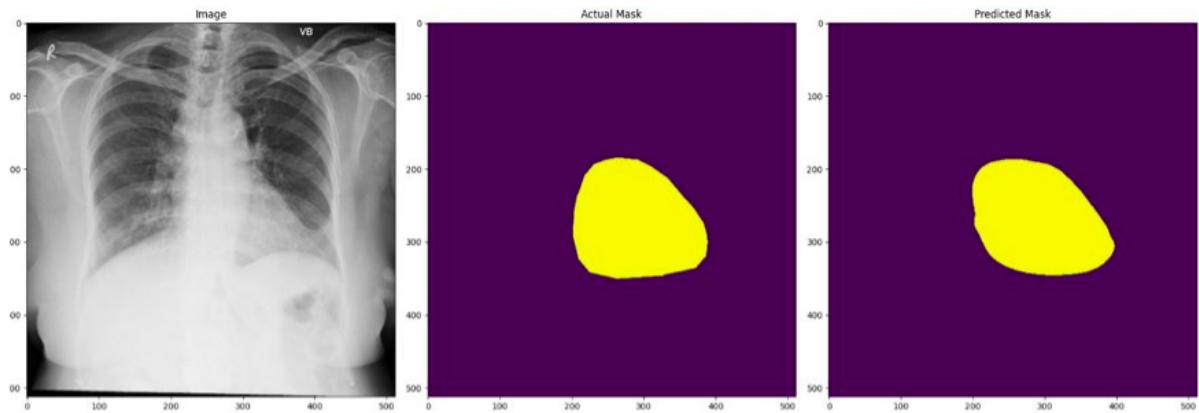


Figure 7.2: Actual (middle) and Predicted Mask (right) of Heart predicted using UNET++ Architecture with RegNety-002 backbone.

actual mask region lie close to each other, they may not be overlapping much resulting in a low IoU score. Hence Euclidean distance between their respective centroids can be used to evaluate the model performance in a better way.

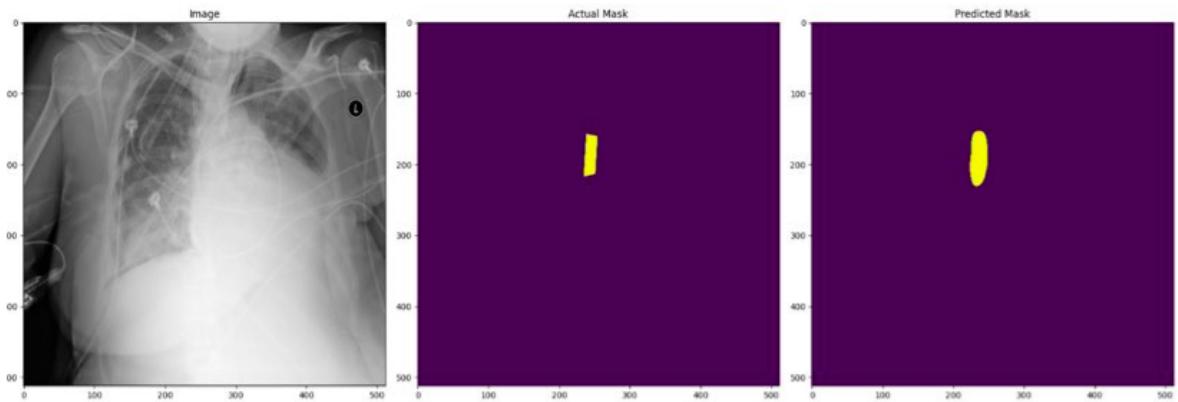


Figure 7.3: Actual (middle) and Predicted Mask (right) of Mediastinal Margin predicted using UNET++ Architecture with EfficientNet-b0 backbone.

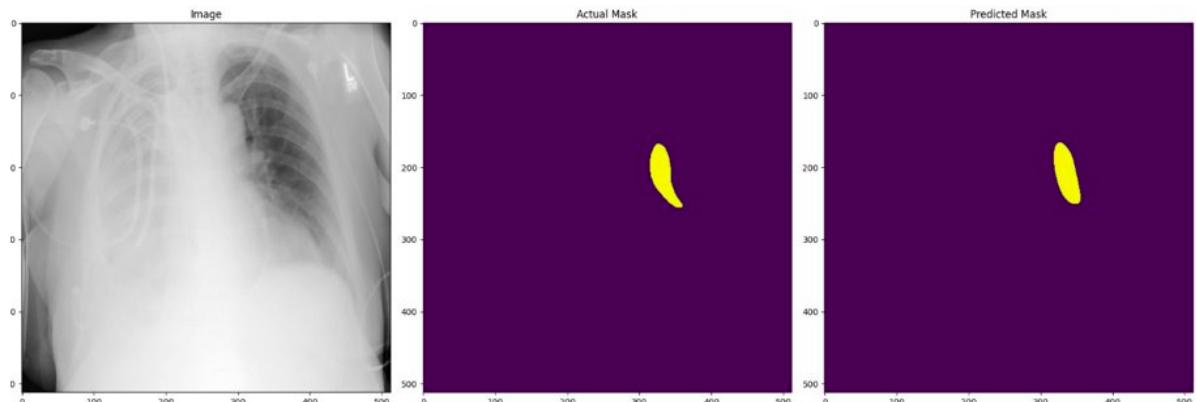


Figure 7.4: Actual (middle) and Predicted Mask (right) of Left Pulmonary Artery predicted using UNET++ Architecture with EfficientNet-b0 backbone.

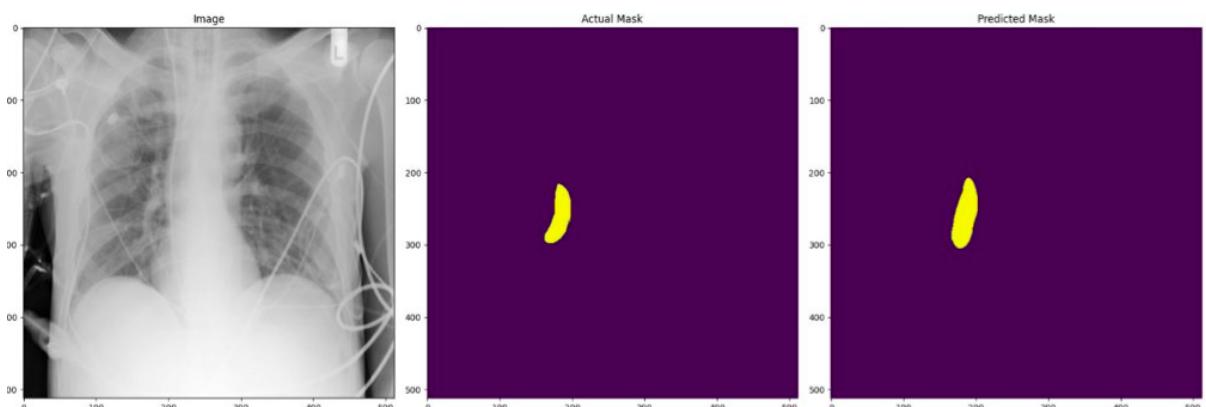


Figure 7.5: Actual (middle) and Predicted Mask (right) of Right Pulmonary Artery predicted using UNET++ Architecture with EfficientNet-b0 backbone.

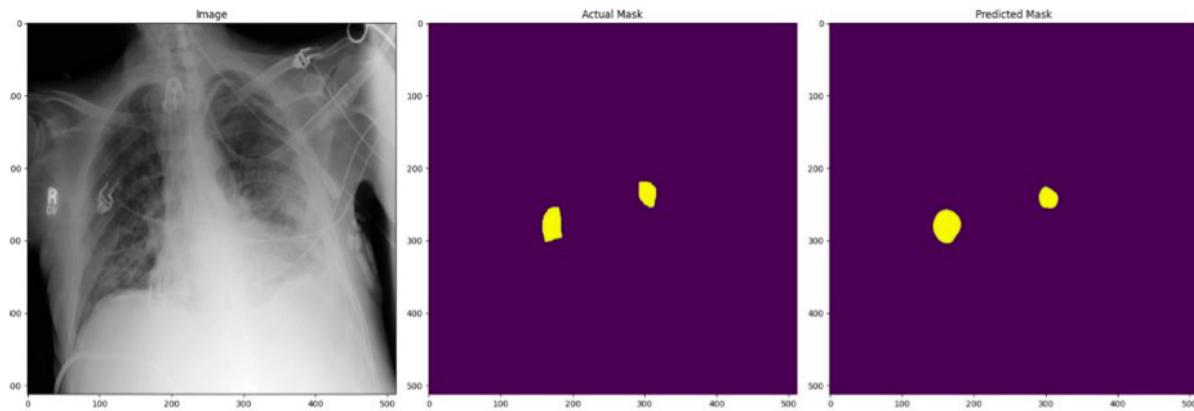


Figure 7.6: Actual (middle) and Predicted Mask (right) of Pulmonary Hila predicted using UNET++ Architecture with EfficientNet-b0 backbone.

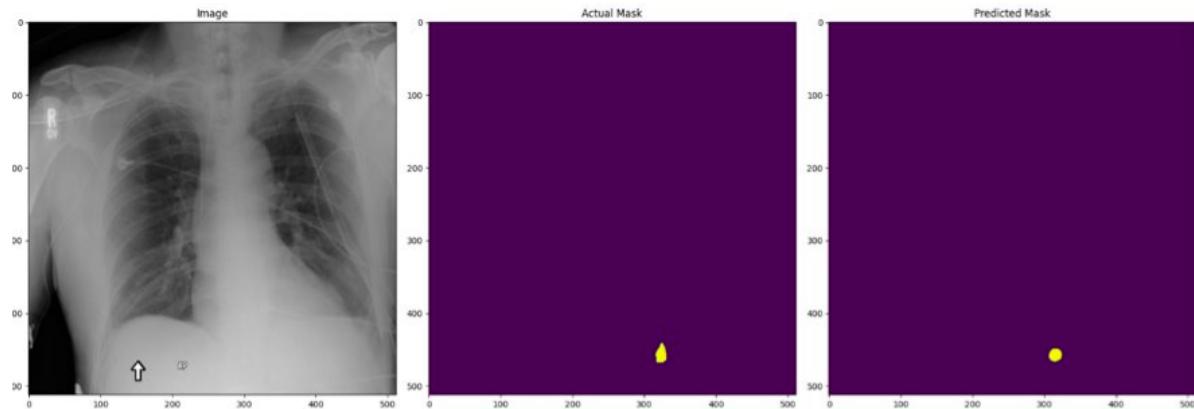


Figure 7.7: Actual (middle) and Predicted Mask (right) of Gastro-Esophageal Junction predicted using UNET++ Architecture with EfficientNet-b0 backbone.

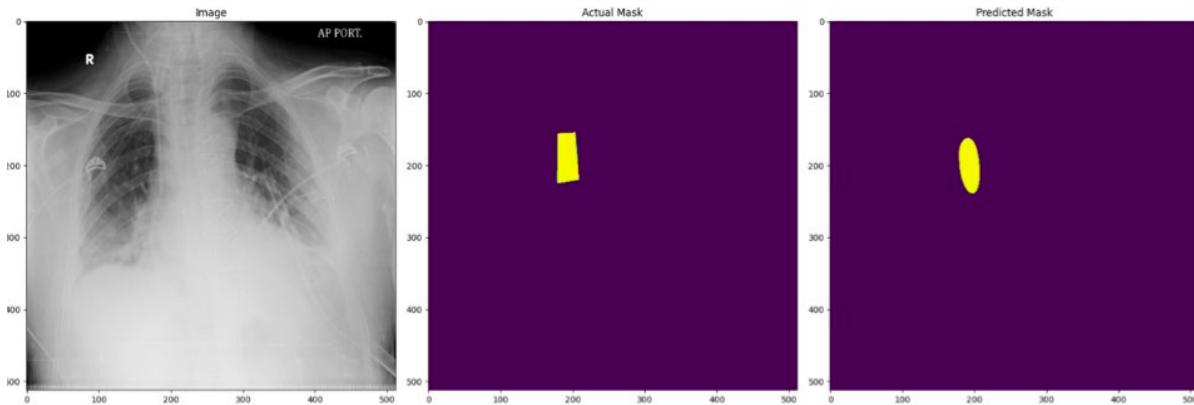


Figure 7.8: Actual (middle) and Predicted Mask (right) of Superior Vena-Cava predicted using UNET++ Architecture with RegNety-002 backbone.

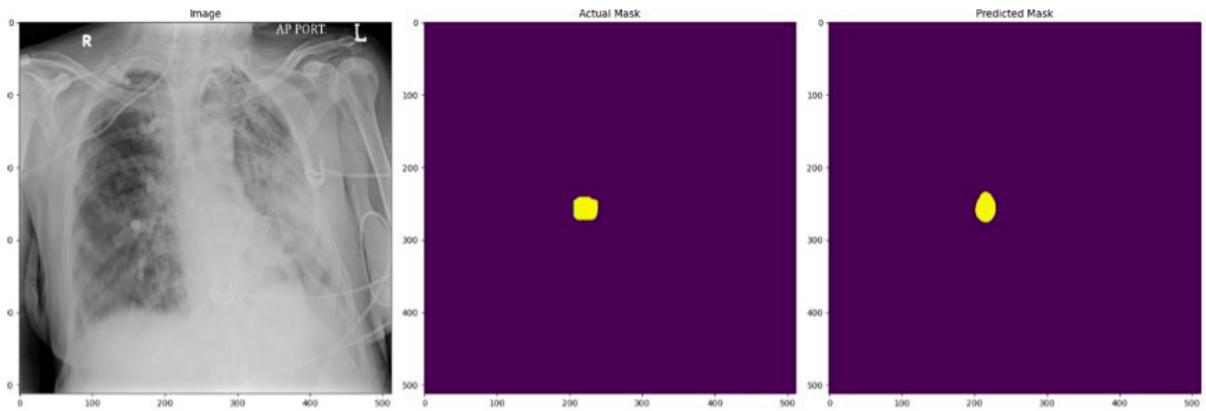


Figure 7.9: Actual (middle) and Predicted Mask (right) of Cavo-Atrial Junction predicted using UNET++ Architecture with EfficientNet-b0 backbone.

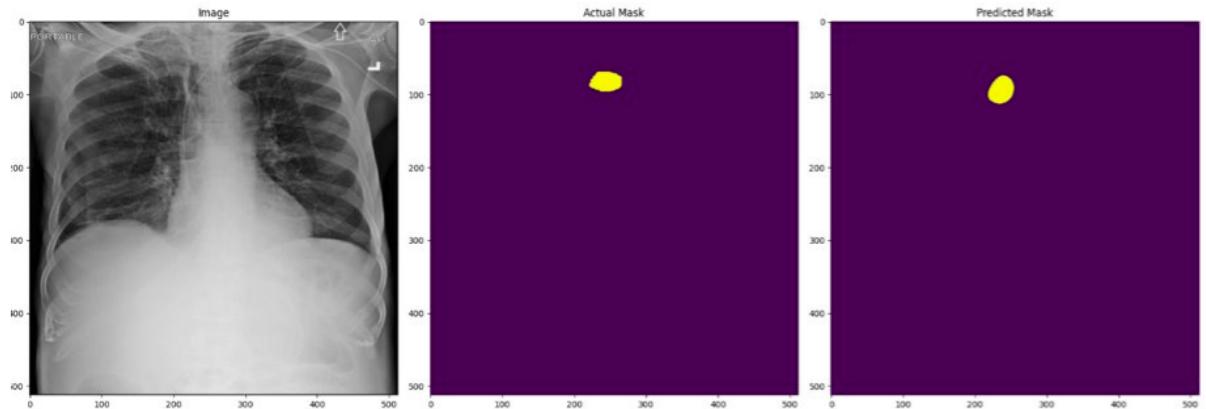


Figure 7.10: Actual (middle) and Predicted Mask (right) of Ideal ETT Tip Location predicted using UNET++ Architecture with ResNest14d backbone.

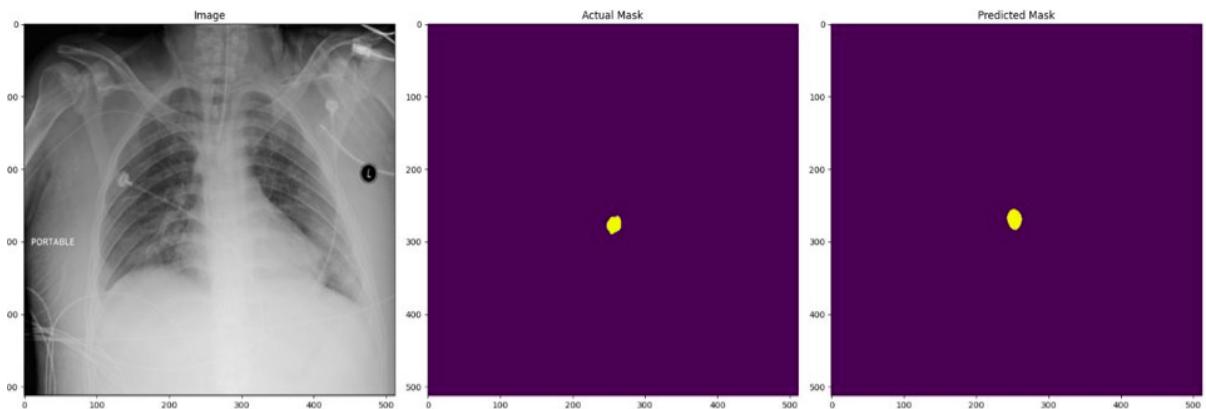


Figure 7.11: Actual (middle) and Predicted Mask (right) of Ideal CVC Tip Location predicted using UNET++ Architecture with ResNest14d backbone.

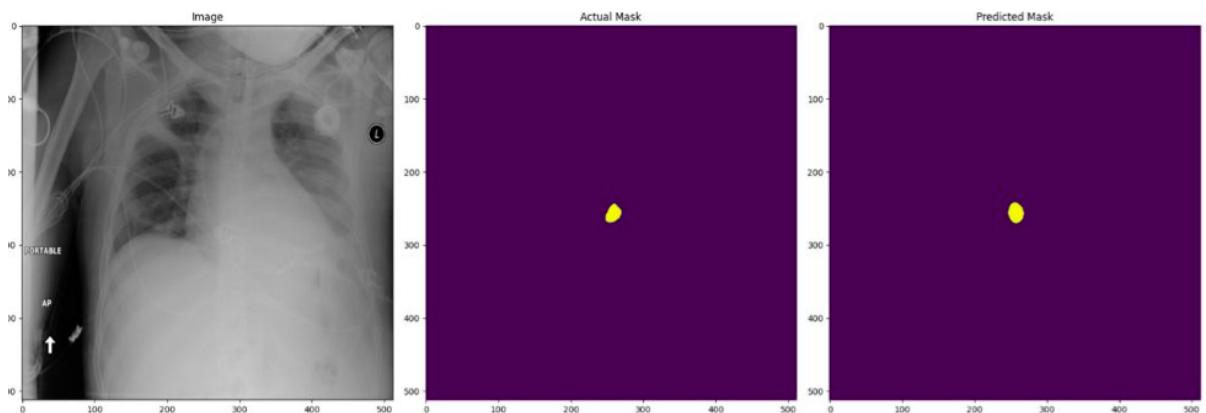


Figure 7.12: Actual (middle) and Predicted Mask (right) of Ideal SGC Tip Location predicted using UNET++ Architecture with ResNest14d backbone.

Chapter 8

Abnormal Tube Tip Position Detection in Chest X-Rays using Student-Teacher Approach

The abnormality detection of tube tips is the final and most important part of the project. This involves the use of the previous stages sequentially starting from the classification, segmentation, and then the detection of malpositions. We have proposed three techniques which can help to detect or classify abnormal positioning of the tubes. The first is the Abnormality detection using Student-Teacher Approach, second is the Abnormality Detection using reference anatomical regions for each tube, third is the Abnormality Detection using Ideal Tip Location. In this chapter, the student-teacher is described.

8.1 Dataset

The RANZCR dataset used in the classification stage will be used here. As described earlier, this dataset contains 30083 Chest X-Rays with 12 columns. The ‘Study Instance ID’ column specifies the unique id of the Chest X-Rays and the ‘Patient ID’ specifies the id of the patient examined. There may be one or more than one Chest X-Rays for each patient. The other columns are defined as ‘ett-normal’, ‘ett-borderline’, ‘ett- abnormal’, ‘ngt-normal’, ‘ngt-borderline’, ‘ngt-abnormal’, ‘cvc-normal’, ‘cvc- borderline’, ‘cvc- abnormal’, ‘swan ganz catheter present’. The abnormal labels for Swan Ganz Catheter are not available. Hence our focus is restricted to Endotracheal Tube, Nasogastric tubes, and Central Venous catheters. The borderline cases are considered normal. If the tube tip is in an abnormal position in the given Chest X-Ray, it is marked as ‘1’ otherwise ‘0’.

8.2 Proposed Methodology

The first stage involves the classification of the Chest X-Rays on the basis of the presence and absence of individual tubes. This step will help to filter out those Chest X-Rays which do not contain any of the four tubes. This can be done in two ways. In a first way, the Chest X-Rays are fed to trained classifier models for each individual tube. Alternatively, these Chest X-Rays can be input to the trained multi-label classifier for the tubes.

In the second stage, the Chest X-Rays are fed to those segmentation models for which the tube is classified as visible by the classifier model/models. This results in the semantic segmentation of tube lines, tube tips, and reference anatomical regions. The output is the binary mask images of these tube lines, tube tips and anatomical regions separately. In the third stage, the detection problem is solved using Student-Teacher Network Approach.

Knowledge distillation [5] is a model compression technique with the help of which a smaller model is trained to imitate a pre-trained, larger model, or multiple(ensemble) models. This is often sometimes referred to as the "teacher-student", where the smaller model is the student and the larger model is the teacher. However, we have used a modified version of Knowledge Distillation in which the teacher and student model have the same CNN Architecture but a different number of input channels. A teacher model is a Convolutional Neural Network having three input channels - a grayscale Chest X-Ray, a binary image containing the tube mask, and a binary image of the reference anatomical structure mask. A student model is the Convolutional Neural Network with the backbone same as that of the teacher model but having only one channel as input grayscale Chest X-Ray. The output layer of both the models contains single neuron denoting the probability of the tube being abnormal. Here the knowledge is transferred from the teacher model to the student model by the minimization of the loss function between the feature maps of the student and teacher so that the student behaves like a teacher model.

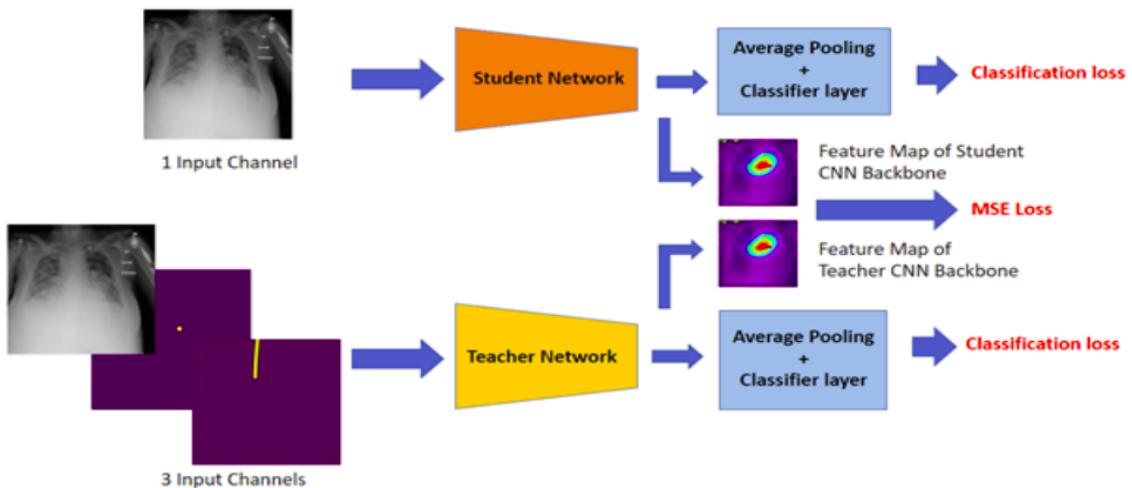


Figure 8.1: Schematic Diagram of Student-Teacher Approach.

We use a Feature Loss which is defined as the Mean Square Error Loss function calculated between the teacher’s feature map and student feature map. Also, the student loss function will also help to train the distilled model to produce the correct labels. Hence, the student loss function is defined as the cross-entropy loss between the predicted class probabilities of the student model and the ground-truth labels known as “hard labels”. The weighted average of Feature loss (FL) [30] and Student Loss will be used to train the student network. The combined loss is defined as

$$L(S(X)) = a * FL(Z_T, Z_S) + (1 - a) * CE(p, \hat{p})$$

$$CE(p, \hat{p}) = -\frac{1}{N} \sum_{i=1}^N (p \cdot \log(\hat{p}) + (1 - p) \cdot \log(1 - \hat{p})) \quad (8.1)$$

$$FL(Z_T, Z_S) = \text{MSE}(Z_T, Z_S)$$

where Z_T and Z_S are the teacher feature vector and student feature vector respectively, N is the size of the output samples, α is the weighting parameter, p is the probability of the tube detected as abnormal in the input image, \hat{p} is the ground truth label. We trained the teacher model first. Then the student model is trained by using the knowledge transferred to it by the teacher model. The student model is trained using the weighted average of feature loss and student loss. A base model has also been trained to compare the performance of the teacher and student network with it. The base model has the same CNN backbone as the student and teacher but has only one input channel. This base model is trained from scratch.

For training the teacher model for detecting ETT abnormal positions, three input channels are required. These are the Chest X-Ray, ETT line mask, and Carina mask. Similarly, the teacher model for NGT requires Chest X-Ray, NGT line, and Gastro-Esophageal Junction mask as input channels. The CVC teacher model requires five input channels. The Chest X-Ray, the CVC tube line, Carina, Superior Vena Cava, and Cavo-Atrial Junction.

8.3 Results

The student, teacher, and the baseline model all have the same CNN backbone architecture. SEMNasNet_050 is used as the CNN backbone for all. As discussed above the teacher model is trained first. Then the student model is trained with the help of the features map from the teacher model. The baseline model is also trained. The models predict the probability of the current tube position being abnormal. Table 8.1, 8.2 and 8.3 shows the AUC-ROC scores for student, teacher and the baseline models for ETT, NGT, CVC tube respectively. The Figure 8.2 contains the Sensitivity plot, Specificity plot, Jaccard Index plot, Negative and Positive Predictive Value plots against threshold and AUC-ROC curve which evaluates the model performance in a better way.

Models	AUC-ROC scores
Student	0.7158
Teacher	0.7213
Baseline	0.5704

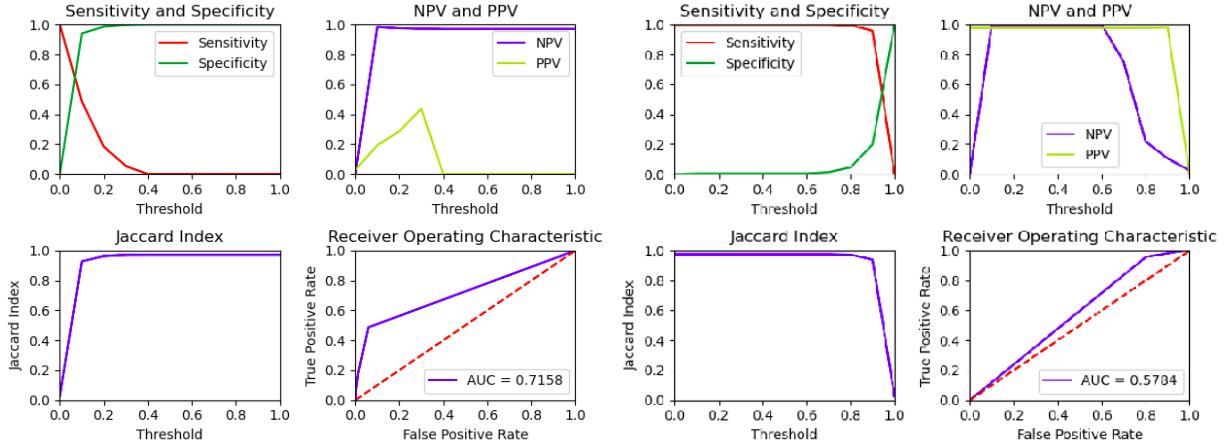
Table 8.1: AUC-ROC scores for the models for ETT Abnormality classifier.

Models	AUC-ROC scores
Student	0.6876
Teacher	0.7011
Baseline	0.5515

Table 8.2: AUC-ROC scores for the models for NGT Abnormality classifier.

Models	AUC-ROC scores
Student	0.7458
Teacher	0.7813
Baseline	0.6174

Table 8.3: AUC-ROC scores for the models for CVC Abnormality classifier.



(a) Plots for Student Model for ETT Abnormality classification having SEMNasNet_050 Backbone. (b) Plots for Baseline Model for ETT Abnormality classification having SEMNasNet_050 Backbone.

Figure 8.2: Plots for Student and Baseline Models for ETT Abnormality classification having SEMNasNet_050 Backbone.

8.4 Discussion

As evident from the results, the teacher model performs the best compared to the student and the baseline models. This is due to the more information contained in the input stage while training the teacher model. The student model performs comparably to the teacher model while using only one input channel. The student model is independent of the segmentation masks for the anatomical regions, tube lines, and tips. This makes it easier

to use compared to the teacher model which needs additional input before inferring on the Chest X-Rays. Using Knowledge Distillation, the student model now contains dark information from the teacher model which helps it to perform better than the baseline model. The performance of the student and teacher models are comparatively the same for each tube. Although their results are not satisfactorily good as very few abnormal cases are there for each tube. Hence there is a high imbalance in the training data.

Chapter 9

Abnormality Detection by Comparing the Euclidean Distance between the Actual Tip and Centroid of the Reference Anatomical Structures

The Chest Anatomical structures/regions are those areas that are defined by the landmarks provided by evident structures that can be visible in Chest X-Rays. These regions are often useful to the radiologists for the prediction of any abnormalities and diseases while analyzing the Chest X-Rays. Some of the most common regions visible in Chest X-Rays are Carina, Heart, Gastro-Esophageal Junction, Cavo-atrial Junction, Lungs, Thorax, Bones, etc.

The conditions for malpositions with respect to reference anatomical regions helps to deduce the abnormal position of the tip.

- The ETT tip position is considered as safe if it is at a distance of 5-7 cm vertically above the carina. If the tip is inserted any further, it may enter into the left or right main bronchus. which can lead to ventilation of a single lung and the contralateral lung or a lobe of the intubated lung will collapse.
- The NGT Tube tip is correctly placed below the diaphragm line down the Gastro-Esophageal Junction. Some possible malpositions may occur when the tube gets looped around the carina region or its misplacement in the right main bronchus. In both cases, the tip lies above the gastroesophageal junction.
- The CVC tip should ideally be present in the Superior Vena Cava region below the Carina and above the Cavo-Atrial Junction. Any Malposition of a CVC means it is lying outside of Superior Vena Cava. There are many chances of misplacement

of the CVC tip. The most common misplacement is that the CVC may go much deeper into the atrium below the Cavo-Atrial Junction. The CVC may also take an incorrect course and end up with their tip in completely wrong place. The CVC tip projected over the left brachiocephalic vein rather than the Superior Vena Cava is considered abnormal.

- The tip of the Swan Ganz catheter needs to be in the right or left pulmonary artery. It should be precisely located not more than 1 cm lateral to the mediastinal margin to avoid any complications, The Catheter should not extend beyond the pulmonary hilum in the Chest X-Rays.

9.1 Methodology

The trained segmentation models for each reference anatomical structure and actual tube tip will be used to generate the respective segmentation masks. The Euclidean distance between the centroids of the anatomical region and the actual tip position will help to detect any abnormality in the tube position. For example, the endotracheal tube is considered abnormally placed if the tip lies below the carina or 5 cm above the carina. Using the constraints for each tube positions defined above for the optimal distancing of the tube tip from their respective anatomical structures, it can be possible to detect the abnormality of tube position in a well defined manner. If the tip lies satisfactorily within the maximum safe distance from the anatomical region, then the prediction is ‘0’ (normal position) else ‘1’ (abnormal position). This approach is evaluated on the RANZCR dataset.

The Blob Coordinates finding Algorithm helps to find the centroid of any blob/patch in the masked image. Using this we can get the coordinates of the centroid of the actual tip as well as the reference anatomical regions. Each Chest X-Ray has pixel-to-distance in cm spacing data. The pixel distance is calculated between the reference anatomical region and the actual tip using Euclidean Distance. This pixel distance is converted to distance in cm using the data available. If the tip lies within a certain distance from the reference anatomical region, it is considered normal else it is considered abnormal. The method is then evaluated against the ground truth label for abnormality.

9.2 Results

The distance between the centroids of the actual tips and the reference anatomical regions can easily be found out. If the actual tip lies within a well defined distance from the those regions, this technique predicts the tip positioning as normal. If the prediction is positive but the ground truth label is negative, it is considered as false positive. If the prediction is negative but the ground truth label is positive, it is considered as false negative. Since

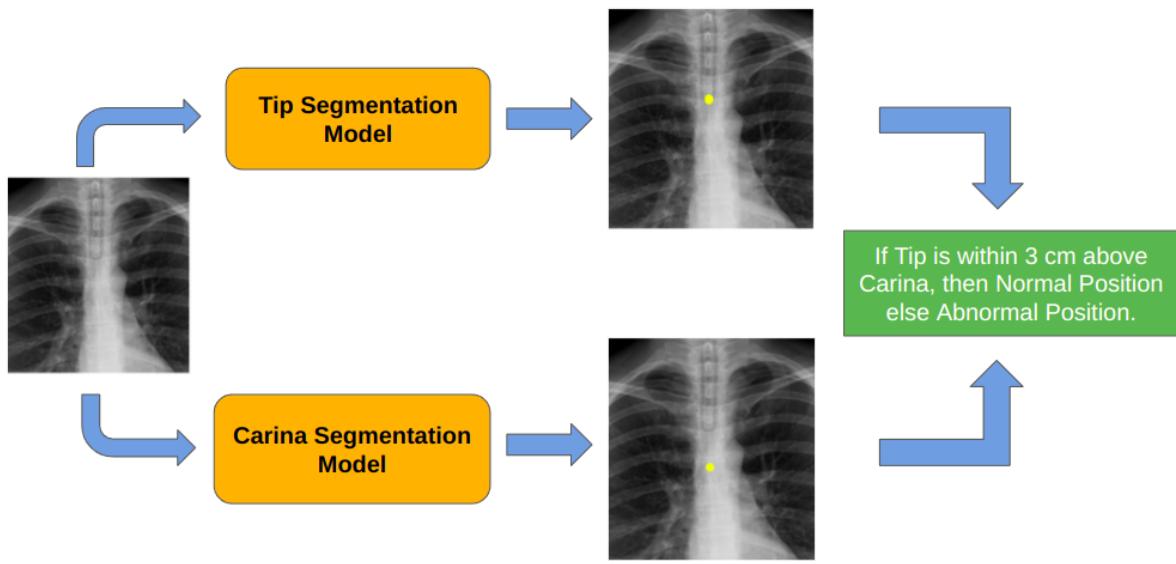
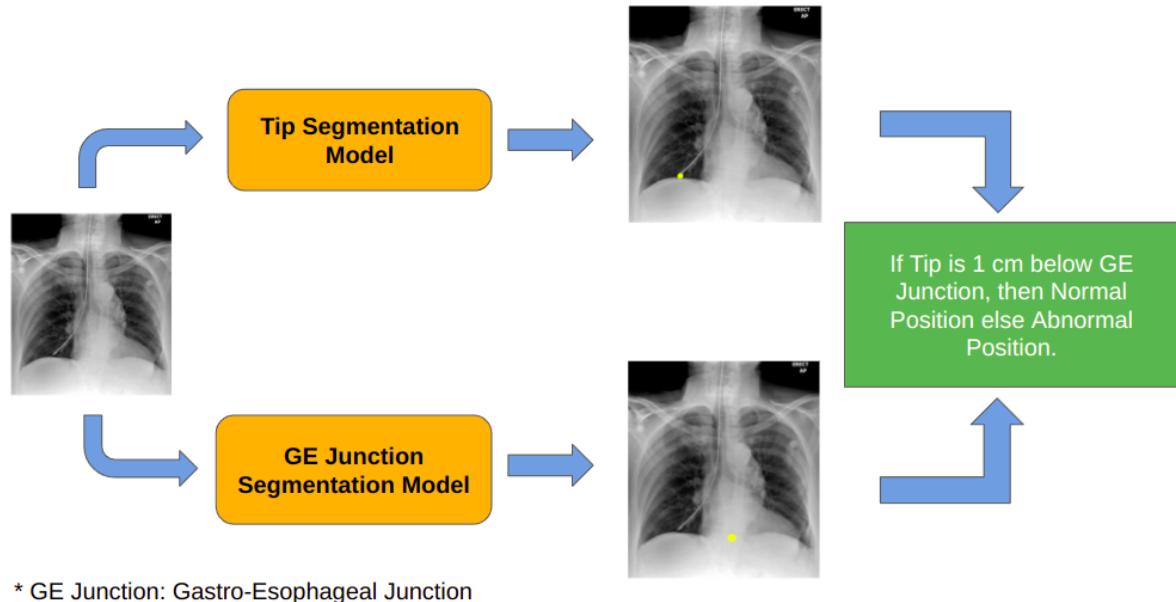


Figure 9.1: Schematic Diagram of Abnormality detection of ETT using trained segmentation models for anatomical regions.



* GE Junction: Gastro-Esophageal Junction

Figure 9.2: Schematic Diagram of Abnormality detection of NGT using trained segmentation models for anatomical regions.

the training is done on the dataset which has very less abnormal data samples compared to the normal cases. This may lead to more false negative cases. Hence accuracy will be not be good metric in that case. Therefore we used precision and recall instead. Using precision and recall values we can easily calculate the F1 score for evaluating the technique. Table 9.1 shows the Precision, Recall, F1 scores for each tube case.

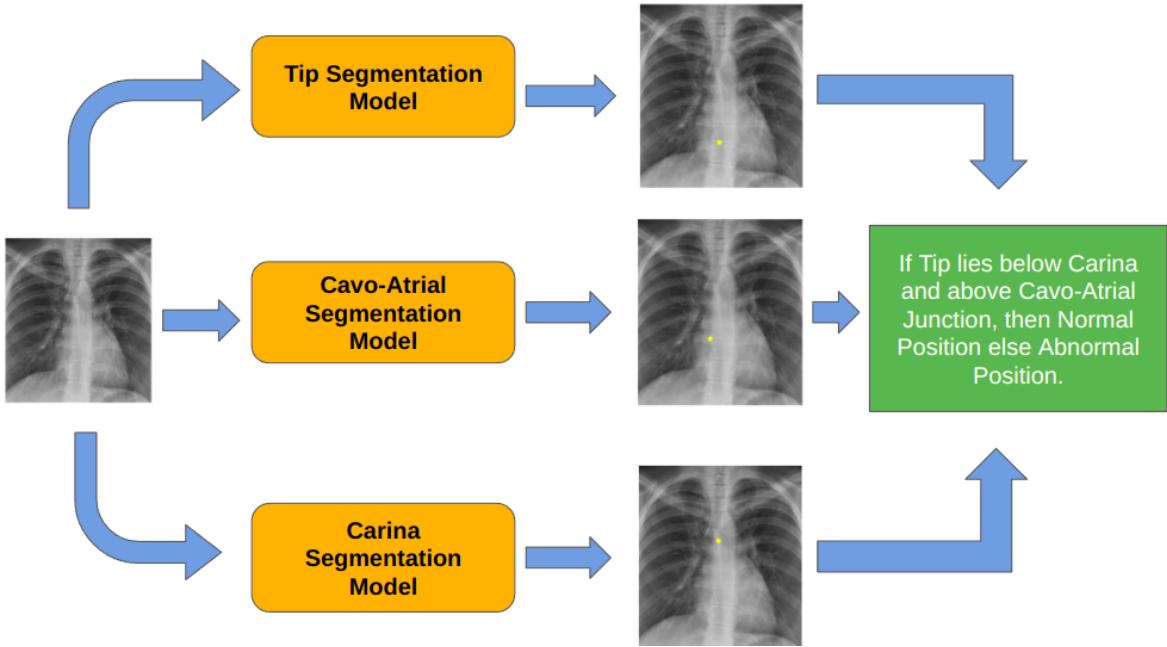


Figure 9.3: Schematic Diagram of Abnormality detection of CVC using trained segmentation models for anatomical regions.

Tube	Precision	Recall	F1 Score
ETT	0.97	0.78	0.86
NGT	0.98	0.77	0.86
CVC	0.94	0.75	0.82

Table 9.1: Precision, Recall, and F1 scores for each tube abnormality detector (with different thresholding for different tube based on their anatomical regions).

9.3 Discussion

The dataset contains higher normal cases compared to abnormal cases. This leads to higher precision for each of the tube cases but lower recall values. This method works better than the student-teacher approach. There is an average performance boost of 10% compared to the former approach. The F1 scores for ETT and NGT are better compared to the CVC. This is because the optimal distance condition for ETT and NGT requires only vertical distance from carina and Gastro-Esophageal Junction respectively whereas for CVC it requires individual Euclidean distance from three anatomical regions making it complex and prone to error.

Chapter 10

Abnormality detection by Comparing the Euclidean Distance between the Actual Tip and Ideal Tip

The Ideal tip location are those target locations where the tip of the tube placed should ideally be present. The radiologists can easily infer from Chest X-Rays where the ideal tip location should be so that they can judge the abnormal condition of tube. If the ideal tip location is known, it becomes easier to find any malpositioning in tube tip.

10.1 Methodology

The segmentation models for predicting binary masks of actual and ideal tips are already trained in previous stages. On feeding the Chest X-Ray into both the segmentation models, we get the segmentation mask for the actual tip and ideal tip location respectively. Using the Blob Coordinates finding algorithm, we can get the centroid position of both locations. The centroidal distance between the ideal tip and the actual tip is calculated using pixel spacing in cm. If the distance between both of them is less than a certain threshold say any value between 0.5-1 cm , the tip is considered to be normally placed.

10.2 Results

The distance between the centroids of the ideal and actual tips can easily be found out. If the actual tip lies within 1 cm of distance from ideal tip, this technique predicts the tip positioning as normal. If the prediction is positive but the ground truth label is negative, it is considered as false positive. If the prediction is negative but the ground truth label is positive, it is considered as false negative. Since the training is done on the

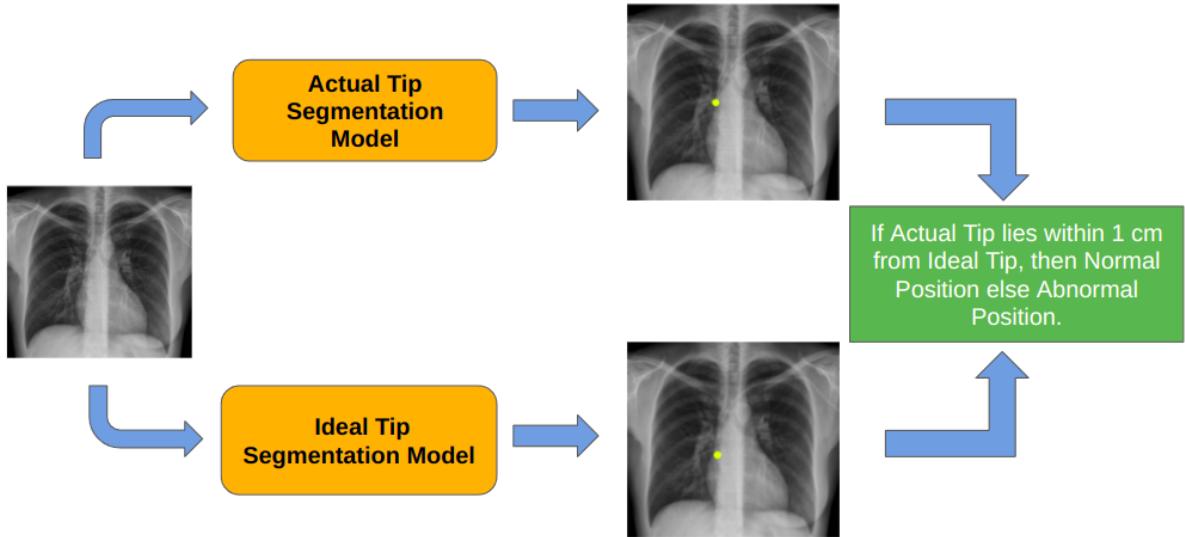


Figure 10.1: Schematic Diagram of Abnormality detection of CVC using trained segmentation models.

Tube	Precision	Recall	F1 Score
ETT	0.98	0.85	0.90
NGT	0.98	0.81	0.88
CVC	0.95	0.83	0.88

Table 10.1: Precision, Recall and F1 scores for each tube abnormality detector (thresholding at 1 cm from ideal tip).

dataset which has very less abnormal data samples compared to the normal cases. This may lead to more false negative cases. Hence accuracy will not be a good metric in that case. Therefore we have precision and recall instead. Using precision and recall values we can easily calculate the F1 score for evaluating the technique. The Table 10.1 shows the Precision, Recall, F1 scores for each tube case.

10.3 Discussion

The dataset contains higher normal cases compared to abnormal cases. This leads to higher precision for each of the tube cases but lower recall values. This method works better than both the student-teacher approach and the abnormality detection using reference anatomical regions. There is an average performance boost of 14% compared to the student-teacher approach and 3% to the second method.

Future Work

This thesis proposes novel methods for accurate prediction of any abnormal position of the tube in the Chest X-Rays. With the good performance already achieved with the proposed method, there is a scope for further improvement. The main focus is to reduce the false-positive and false-negative cases so that the method can be widely accepted in healthcare services. There are some ways which can help improve the performance further. In our proposed approach, we are using the ideal tip location as means of predicting abnormality. If the dataset having ideal target area labels is available then it can be used as a better reference for judging the abnormality. In this case, instead of using the distance metric with a certain threshold, we can use intersection over union for predicting malpositions. As evident from the multi-label classifier's better performance compared to individual tube binary classifiers, a common model for predicting segmentation labels can be used for all tubes/anatomical regions simultaneously. A combined architecture for classification and segmentation can also be experimented with using Y-Net [13]. As the medical datasets are generally smaller in size, the models directly trained on these datasets face the problem of overfitting and poor generalization over the test set. Pre-trained models on large Chest X-Ray datasets can also be used so that better features can be learned. This pre-trained model can then be fine-tuned on a smaller target dataset. This can help reduce the problem of overfitting. Also, pretraining using self-supervised learning can also be used in case labeled data is not available in abundance [16].

Conclusion

Deep convolutional neural networks perform well in distinguishing Chest X-Rays in detecting the presence/absence and normal/abnormal position of tubes since the training dataset is sufficient to obtain a good AUC for all prevalence of datasets. The abnormal cases are fewer in number compared to normal cases. The teacher-student approach faces the problem of overfitting and poor results because of this. In healthcare, it is very important to deal with false positive and false negative cases, and extreme care is taken to prevent such wrong results. Our proposed approaches work well but there is still a lot of scope for improvement. This work can help reduce the massive manual workload of doctors and clinicians, especially in peak times like COVID-19 waves. The automatic early detection of malpositions can save many lives and reduce further damage to the organs by timely alerting the clinicians to correct the wrong positions.

Bibliography

- [1] Yuri Sousa Aurelio et al. “Learning from imbalanced data sets with weighted cross-entropy function”. In: *Neural processing letters* 50.2 (2019), pp. 1937–1949.
- [2] Antonio Bozzani et al. “Iatrogenic pulmonary artery rupture due to chest-tube insertion”. In: *Texas Heart Institute Journal* 37.6 (2010), p. 732.
- [3] Mervyn D Cohen et al. “Accurate localization of the position of the tip of a naso/orogastric tube in children; where is the location of the gastro-esophageal junction?” In: *Pediatric radiology* 41.10 (2011), pp. 1266–1271.
- [4] Lawrence R Goodman et al. “Radiographic evaluation of endotracheal tube position”. In: *American Journal of Roentgenology* 127.3 (1976), pp. 433–434.
- [5] Jianping Gou et al. “Knowledge distillation: A survey”. In: *International Journal of Computer Vision* 129.6 (2021), pp. 1789–1819.
- [6] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [7] Mohammad Hesam Hesamian et al. “Deep learning techniques for medical image segmentation: achievements and challenges”. In: *Journal of digital imaging* 32.4 (2019), pp. 582–596.
- [8] Shruti Jadon. “A survey of loss functions for semantic segmentation”. In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE. 2020, pp. 1–7.
- [9] Meredith Kells and Susan Kelly-Weeder. “Nasogastric tube feeding for individuals with anorexia nervosa: an integrative review”. In: *Journal of the American Psychiatric Nurses Association* 22.6 (2016), pp. 449–468.
- [10] Daniel S Kermany et al. “Identifying medical diagnoses and treatable diseases by image-based deep learning”. In: *Cell* 172.5 (2018), pp. 1122–1131.
- [11] Chien-Shun Lo and Chuin-Mu Wang. “Support vector machine for breast MR image classification”. In: *Computers & Mathematics with Applications* 64.5 (2012), pp. 1153–1162.

- [12] Mary B Markovich. “Central Venous Catheter Tip Placement: Determination of Posterior Malposition—A Case Study”. In: *Journal of the Association for Vascular Access* 11.2 (2006), pp. 85–89.
- [13] Sachin Mehta et al. “Y-Net: joint segmentation and classification for diagnosis of breast biopsy images”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 893–901.
- [14] L Das Narla et al. “Evaluation of umbilical catheter and tube placement in premature infants.” In: *Radiographics* 11.5 (1991), pp. 849–863.
- [15] Marc Noppen and Tom De Keukeleire. “Pneumothorax”. In: *Respiration* 76.2 (2008), pp. 121–127.
- [16] Cheng Ouyang et al. “Self-supervision with superpixels: Training few-shot medical image segmentation without annotation”. In: *European Conference on Computer Vision*. Springer. 2020, pp. 762–780.
- [17] John Pfitzner, Heather J Stevens, and David G Lance. “Identifying imminent displacement of a double-lumen tube caused by surgical traction at the pulmonary hilum”. In: *Journal of cardiothoracic and vascular anesthesia* 21.5 (2007), pp. 776–777.
- [18] Pranav Rajpurkar et al. “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning”. In: *arXiv preprint arXiv:1711.05225* (2017).
- [19] Jesse Read and Fernando Perez-Cruz. “Deep learning for multi-label classification”. In: *arXiv preprint arXiv:1502.05988* (2014).
- [20] Ali M Reza. “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement”. In: *Journal of VLSI signal processing systems for signal, image and video technology* 38.1 (2004), pp. 35–44.
- [21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [22] Claudio Sandroni et al. “Unusual central venous catheter malposition into the left internal mammary vein: a case report”. In: *Intensive Care Medicine* 29.12 (2003), pp. 2338–2339.
- [23] William G Schultheis and Paras Lakhani. “Using Deep Learning Segmentation for Endotracheal Tube Position Assessment”. In: *Journal of thoracic imaging* 37.2 (2022), pp. 125–131.
- [24] M Schuster et al. “The carina as a landmark in central venous catheter placement”. In: *British Journal of Anaesthesia* 85.2 (2000), pp. 192–194.

- [25] Leslie N Smith. “Cyclical learning rates for training neural networks”. In: *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2017, pp. 464–472.
- [26] YG Song et al. “Use of vertebral body units to locate the cavoatrial junction for optimum central venous catheter tip positioning”. In: *British journal of anaesthesia* 115.2 (2015), pp. 252–257.
- [27] Mingxing Tan and Quoc Le. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 6105–6114.
- [28] Mingxing Tan et al. “Mnasnet: Platform-aware neural architecture search for mobile”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2820–2828.
- [29] Ahsan Bin Tufail, Yong-Kui Ma, and Qiu-Na Zhang. “Binary classification of Alzheimer’s disease using sMRI imaging modality and deep learning”. In: *Journal of digital imaging* 33.5 (2020), pp. 1073–1090.
- [30] Weitao Wan et al. “Rethinking feature distribution for loss functions in image classification”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 9117–9126.
- [31] Jing Xu et al. “RegNet: Self-regulated network for image classification”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [32] Samir S Yadav and Shivajirao M Jadhav. “Deep convolutional neural network based medical image classification for disease diagnosis”. In: *Journal of Big Data* 6.1 (2019), pp. 1–18.
- [33] Jun Yang and Fei Wang. “Auto-ensemble: An adaptive learning rate scheduling based deep learning model ensembling”. In: *IEEE Access* 8 (2020), pp. 217499–217509.
- [34] Hang Zhang et al. “Resnest: Split-attention networks”. In: *arXiv preprint arXiv:2004.08955* (2020).
- [35] Zongwei Zhou et al. “Unet++: A nested u-net architecture for medical image segmentation”. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.