**ADS2002 Final Report**

**Catheter Placement Project**

**Group 5 Thursday Studio**

**Authors:**

Sneha Binu, Charis Lai, Shekinah Robins, Sanugi Fernando, Ashlyn Tiwari, Edmund Chau

**Table of Contents**

[**Executive Summary 2**](#_heading=h.3dy6vkm)

[Project Brief 2](#_heading=h.1t3h5sf)

[Our Data 2](#_heading=h.e3b9f9hmfxuc)

[Findings 3](#_heading=h.17dp8vu)

[**Body 4**](#_heading=h.3rdcrjn)

[Introduction 4](#_heading=h.lnxbz9)

[Data Quality & Preprocessing 4](#_heading=h.35nkun2)

[Exploratory Data Analysis 5](#_heading=h.1ksv4uv)

[Model Development 8](#_heading=h.44sinio)

[Binary Classification of Presence of Tube 9](#_heading=h.of1istbzw6ch)

[EfficientNetB0 CNN Model 9](#_heading=h.3snfhwepnx7a)

[MobileNetV2 CNN Model 10](#_heading=h.qomapwc4aegq)

[Classification of Positioning of Tube  
ResNet50 CNN Model 10](#_heading=h.c35tcp8n1w7e)

[Results 10](#_heading=h.tw6hjxrrnktj)

[EfficientNetB0 CNN Model 10](#_heading=h.3jzu93aygzig)

[MobileNetV2 CNN Model 13](#_heading=h.t61j5mmtaya0)

[ResNet50 CNN Model 16](#_heading=h.odc104bwz9fg)

[Conclusions 22](#_heading=h.ff2h4f31fbwp)

[**Bibliography 24**](#_heading=)

# 

# **Executive Summary**

## Project Brief

In hospital settings, accurate placement of medical devices such as catheters are critical to ensure effective patient care and avoid severe health complications. Commonly inserted devices, such as catheters and medical tubes are used for essential treatments but can be prone to malpositioning, especially in high stress environments like overburdened hospitals. Misplacement of these devices can lead to serious health risks, including infections or impaired functionality (Mauri et al., 2018). Our project focuses on using data science to assist medical professionals in identifying whether catheters are positioned correctly in chest X-ray images. By leveraging machine learning techniques, this project aims to automate the detection process, potentially reducing human error and enhancing the speed of diagnosis in medical settings.  
Our Data

The data is a large, labelled dataset of chest radiographs provided by the National Institute of Health, containing 40,000 images. The dataset includes coordinates for each tube, specifying whether each type of tube is normally positioned, borderline, or abnormal. This consists of thousands of chest X-ray images with corresponding labels indicating correct or incorrect placement of various catheters and tubes. Each image is multi-labeled with binary classifications indicating the presence or absence of anomalies. The data is split into training and validation sets.

Main Objective

The aim of this project is to develop deep learning models that accurately classify medical images to detect both the presence and proper placement of various catheter tubes in chest radiographs. By leveraging deep learning techniques, the model will identify key medical devices such as central venous catheters (CVC), nasogastric tubes (NGT), and endotracheal tubes (ETT), and assess whether these tubes are correctly positioned or misaligned. This will provide clinicians with critical support in identifying potentially harmful catheter misplacements.   
Findings

In our research, we utilised deep learning models to visually explain how the models predict normal, borderline and abnormal catheter placements in chest X-rays. Our analysis revealed that the model's findings closely match the expected position of the central venous catheter (CVC). We should expect the catheter tip to ideally lie within the superior vena cava, following the vein’s natural curve while avoiding proximity to the pericardial sac (the outermost layer of the heart) to prevent complications (Gibson & Bodenham, 2013).

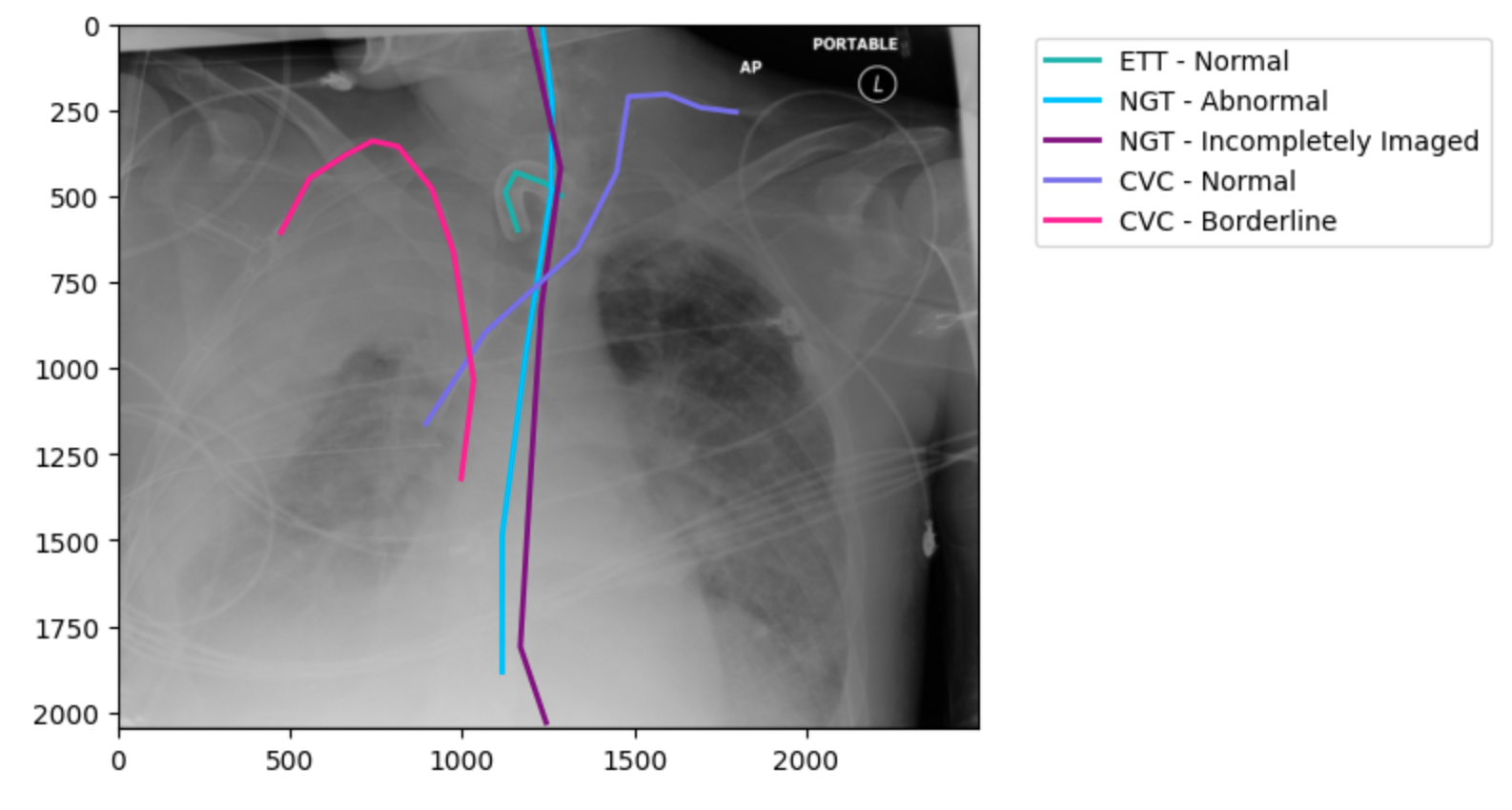
Our data analysis further revealed that CVCs were the most frequently placed catheters across all placement categories, particularly dominating in normal placements, while ETTs and NGTs also showed significant counts.   
Upon further research we discovered that the COVID-19 pandemic had led to an increase in central line complications due to the overwhelming demand on healthcare systems and the involvement of less experienced practitioners(Ilonzo et al., 2020). The urgency of treating critically ill patients and the full capacity of intensive care units have made rapid central line insertions more prone to errors. Although ultrasound-guided access can help reduce risks, the high-pressure environment has still led to a rise in complications. This not only affects patient outcomes but also increases healthcare workers' exposure to the virus during the management of these complications.  
Ultimately, our models' success demonstrates how they can improve clinical judgement when it comes to catheter placement. These models can help medical professionals reduce errors and enhance patient safety by precisely recognising and categorising catheter placements. In the long term, incorporating these prediction models into clinical practice can improve patient outcomes, make better use of available resources, and lessen complications—especially in high-stress scenarios like the COVID-19 pandemic. Continued refinement of these models is essential to improving predictive capabilities.   
Out of three deep learning models explored, ResNet50 showed the most balanced performance with moderate accuracy. The intended process for our approach to determining the placement of any one particular type of catheter as seen in an X-ray image requires output from two different models. First, an image will be inputted into the binary classification model for one of either CVC, ETT or NGT catheters. This will return whether the chosen type of catheter is present in the image or not. If present, this image can be inputted into the next model which would classify the positioning of the given catheter as either normal, borderline or abnormal. Due to various challenges and limitations, this report details the preliminary findings in this project. It is hypothesised that with more resources and time, the deep learning models shown in this report can be refined to yield higher accuracies and increased reliability.

# **Body**

## **Introduction**

The problem of malpositioned lines and tubes in hospital patients is a significant concern, as improper placement can lead to serious complications. To address these challenges, this project aims to develop a solution leveraging deep learning for automatically detecting the presence and placement of lines and tubes in chest radiographs. By using medical imaging, the goal is to assist medical professionals in quickly identifying malpositioned tubes, reducing the risk of human error. For this project, we are using a dataset derived from chest radiographs, which includes catheter coordinates to identify the placement of several types of medical lines and tubes. These labels indicate whether a device is present in the image and, if present, whether it is positioned correctly or incorrectly. We anticipate finding correlations between the presence of specific tubes and their respective rates of malposition. Furthermore, there may be patterns indicating which tubes are more prone to misplacement under particular conditions, such as those inserted in emergency situations or by less experienced staff. We expect that our model will be able to detect these relationships and offer accurate predictions based on image data alone. A key challenge in this project is the significant time required to train the model, which can slow down experimentation and model development. Furthermore, the multi-label nature of the task which requires predictions about the presence of multiple types of lines and tubes, as well as whether they are malpositioned adds an additional layer of difficulty.

## **Data Quality & Preprocessing**

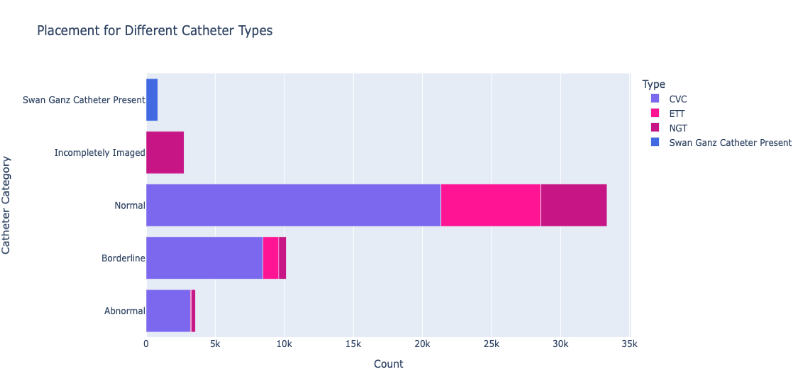
The preprocessing stage of our project involved an initial examination of the dataset, focusing on its structure and identifying any missing values. We utilised the train.csv and train\_annotations.csv files for the analysis, removing rows containing all zeros (no tubes present), incomplete NGT records, and Swan Ganz catheter placements. Upon further inspection, we confirmed there were no null entries. The train.csv file consists of 13 columns—two identifying columns (image IDs) and 11 binary labels, which indicate the presence or absence of different lines and tubes. We mapped the chest radiograph images to their corresponding records using the StudyInstanceUID. To assist in visualising the data, we developed functions to display the images and overlay annotations, which included the coordinates of tube placements. 

**Figure 1 : X-ray image overlaid with annotations and different catheters**

These annotations were colour-coded to distinguish between normal, borderline, and abnormal placements as shown in Figure 1. This step was critical in preparing the data for the modelling phase.

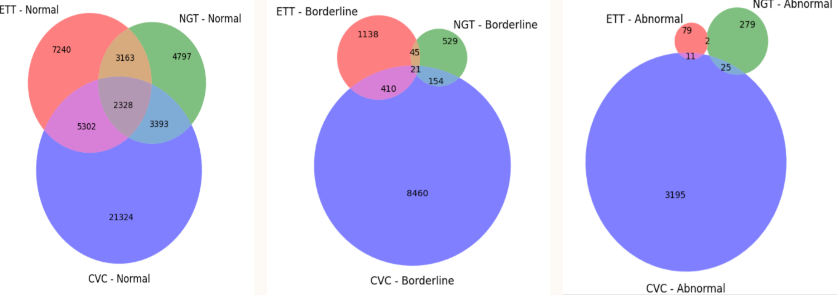
## **Exploratory Data Analysis**

The following data visualisations present our exploratory data analysis of catheter placements across different types, including ETT, NGT, and CVC, divided into three categories: normal, borderline, and abnormal placements.



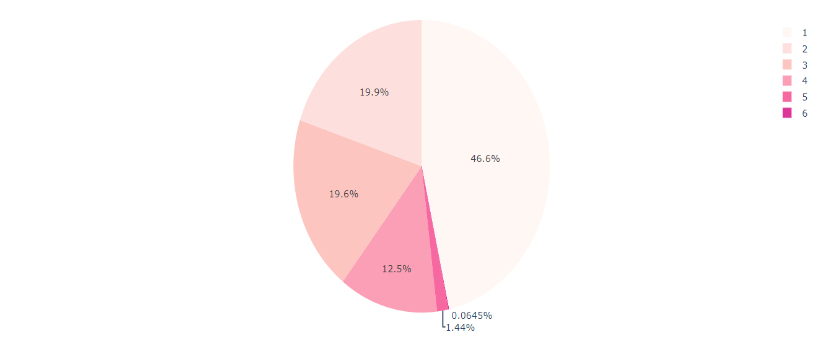
**Figure 2 : Placement for Different Catheter Types**

In Figure 2, we can see that the CVC catheters account for the highest number of placements across the board, especially in the normal and borderline categories, while NGT catheters are notably affected by incomplete imaging, a unique challenge for this type. ETT catheters, though significant in normal placements, contribute less to borderline and abnormal categories, suggesting better overall placement performance compared to other catheter types.

****

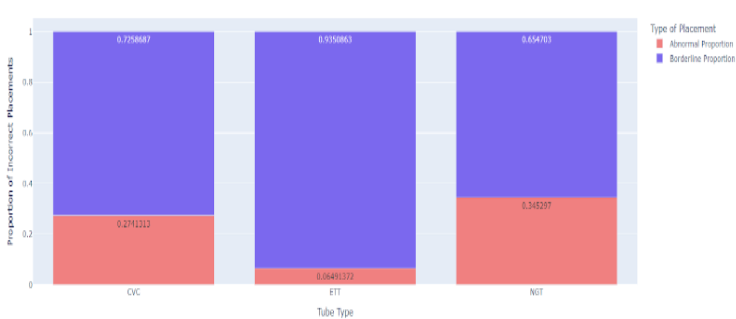
**Figure 3: Distribution of Catheter Types by Placement**

Furthermore, we used Venn diagrams to illustrate placement overlaps as shown in Figure 3, which reinforces some of the previously stated findings. It is clear that CVC catheters dominate in terms of normal placement, with over 21,000 cases, followed by ETT and NGT catheters. Significant overlap is visible in the normal placement category, indicating that many patients had more than one type of catheter correctly positioned. In contrast, the borderline placement category shows a reduction in overlap, suggesting fewer cases where multiple catheter types were borderline positioned simultaneously. The abnormal placement category reveals minimal overlap between catheter types, particularly for ETT and NGT catheters, showing that abnormal placement tends to be isolated to one catheter type per patient.



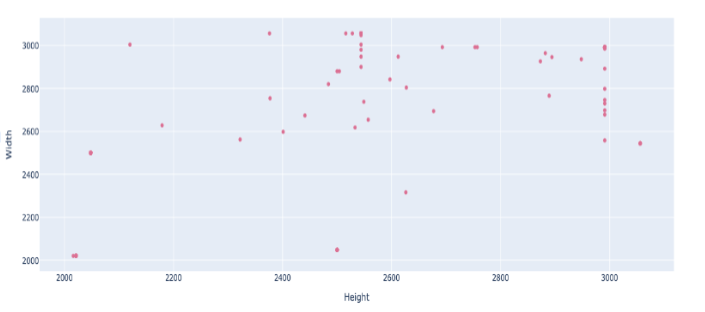
**Figure 4 : Frequency of the Number of Tubes**

The pie chart in Figure 4 illustrates the frequency of the number of tubes placed in patients, revealing that the majority (46.6%) had only one tube inserted, indicating that most patients required minimal intervention. A significant portion of patients, around 19.9% and 19.6%, had two or three tubes placed, respectively, showing that multiple tube placements were also relatively common. A smaller percentage, 12.5%, had four tubes placed, while very few patients needed more complex interventions, with 1.44% having five tubes and only 0.0645% requiring six tubes. Overall, the chart indicates that while most patients had one or two tubes, there are cases requiring higher numbers of catheter placements, though they are much less frequent.



**Figure 5 : Proportion of Abnormal and Borderline Placements by tube type**

Following this, we analysed the distribution of abnormal and borderline tube placements by catheter type as illustrated in Figure 5. The analysis revealed that 72% of CVC placements were classified as borderline, with the remaining 28% marked as abnormal. For ETT placements, a significant 93% were borderline, while only 7% were abnormal, indicating that most ETT placements were closer to the correct positioning. In contrast, 65% of NGT placements were identified as borderline, with 35% being abnormal. This breakdown highlights a trend where borderline placements are more common across all catheter types, with ETT having the highest proportion of borderline placements, suggesting fewer significant errors in placement for this tube type.



**Figure 6 : Sample of Image Dimensions**

Finally, we showcased a sample of the image dimensions used in our analysis as shown in Figure 6, which ranged from 2000 to 3000. This provided a clear view of the variability in image sizes, allowing us to assess how dimension differences might influence the accuracy and quality of the tube placement analysis

Findings:

NGTs faced challenges with incomplete imaging, which affected placement verification. Abnormal placements were generally isolated to one catheter type per patient, with minimal overlap between types.

Regarding tube frequency, 47% of patients had only one tube placed, while others required multiple tubes, though more complex interventions with four or more tubes were rare. Analysing borderline and abnormal placements showed that the majority of CVC (72%) and ETT (93%) placements were borderline, whereas NGTs had a higher rate of abnormal placements (35%). Finally, image dimensions used for the analysis ranged from 2000 to 3000, showcasing a clear view of the variability in image sizes

## **Model Development**

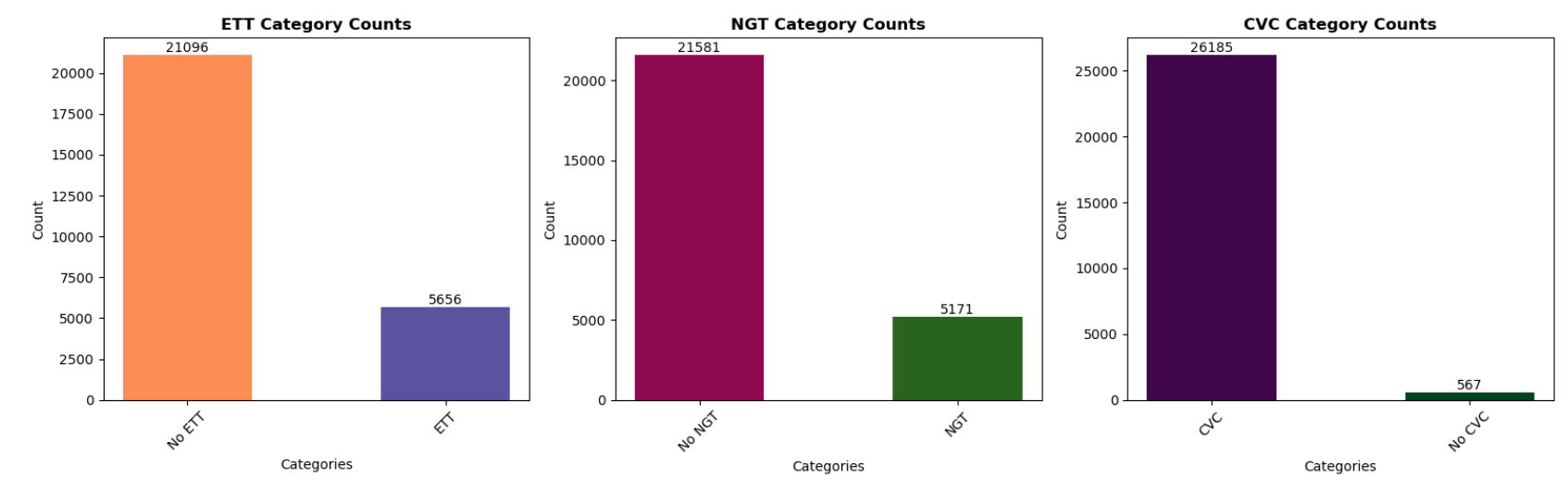
To model the placement of catheters in X-rays, a technique called transfer learning was used which allowed us to adapt a pre-trained model that had been trained on a larger dataset and use it on our image classification problem. This is useful because it allows us to leverage generalised features that the model has already learned.

We used transfer learning because we're working with a smaller dataset that has class imbalances, which could cause overfitting if we trained a model from scratch. This is because, in our data set, of the classes which we intended to predict, there are some classes that occur significantly more frequently than others. Pre-trained models are also already familiar with a diverse set of images which helps with better generalisation to new data. Additionally, the complex architecture of these pre-trained models are often too challenging to train effectively from scratch on smaller datasets.  
To apply this technique, we began by freezing the layers and weights of the pre-trained model to preserve the general features learned by the model and added new trainable layers on top and trained using a relatively high learning rate, to help the model learn patterns in our catheter images. We then unfroze a small portion of the pre-trained layers and trained them with a lower learning rate, allowing the model to fine-tune its weights specifically to our set of images without significantly altering the previously learned features.

For the task of modelling, we broke up the task into two main subsections. First, we decided to create three different binary classification models to determine whether an image contained a specific type of catheter or not. Each binary classification model was intended to predict whether one specific type of catheter, either a CVC, ETT or NGT, was present in the image or not. Secondly, we attempted to determine whether a given catheter’s positioning was normal, abnormal or borderline.

We initially used images with the overlaid tube annotations to train our models. However, we realised that despite achieving high training, validation, and testing accuracies, the model was only predicting the majority class for all images without the annotations, even after handling class imbalances. So we instead switched to using the unannotated images and while this led to a slight decrease in overall accuracy, the model was able to learn more generalised features relevant to both classes, resulting in improved reliability of predictions.

Before training our models we needed to deal with the class imbalances in our data, as shown in Figure 7, to reduce risk of overfitting. This was done by downsampling the instances in the classes ‘No ETT’ and ‘No NGT’ down to roughly 5000 to match the number of instances in the classes ‘ETT ’ (present) and ‘NGT ’ (present) respectively. As shown in Figure 7, the class imbalance was reversed and much more significant in the case for CVC tubes with roughly 52 instances of ‘CVC’ for every instance of ‘No CVC’. Hence, ‘CVC’ (present) was downsampled and ‘No CVC’ was upsampled for both classes to match at approximately 2000 instances. However, we hypothesised that as the upsampling of ‘No CVC’ was done through replication of existing images, risk of overfitting would still be present.



**Figure 7 : Frequency graphs of binary classes of each model**

### **Binary Classification of Presence of Tube**

We evaluated three different convolutional neural networks—EfficientNetB0, MobileNetV2, and ResNet50—to thoroughly assess their performance and appropriateness for our specific classification task. These models were noted in our initial research into model architectures best suited for image classification. Our main objective was to identify which model would be most effective for classifying the types of catheters in our dataset.

#### **EfficientNetB0 CNN Model**

This model was explored as it scales all dimensions of an image including depth, width and resolution using a simple yet highly effective compound coefficient. (Tan, M., & Le, Q., 11 Sept 2020) After looking into the competition for catheter placement in Kaggle, it was observed that competitors with high rankings used this architecture for predicting the types of catheters.

#### **MobileNetV2 CNN Model**

This model is a lightweight 53-layer Deep Convolutional Neural Network model that uses a smaller number of parameters. It has lower computational cost which is desirable for training as the process will be more time efficient. It can be used when computational resources are limited, such as image classification and object detection on mobile devices.  
  
 **ResNet50 CNN Model**This model is a Deep Convolutional Neural Network with 50 layers which uses skip connections to address vanishing gradients, enabling it to learn complex patterns from the data.It is of interest in this project as it has been successful in other medical imaging tasks such as brain tumour classification and detection of pulmonary infections in COVID-19 CT scans.

### **Classification of Positioning of Tube** **ResNet50 CNN Model**

After training three models for the binary classification of the presence of a specific tube in an image, it was found that this model was the best suited to our dataset. Hence, it was hypothesised that this model would be the most suitable option for the multi-label classification problem of determining the positioning of a tube.

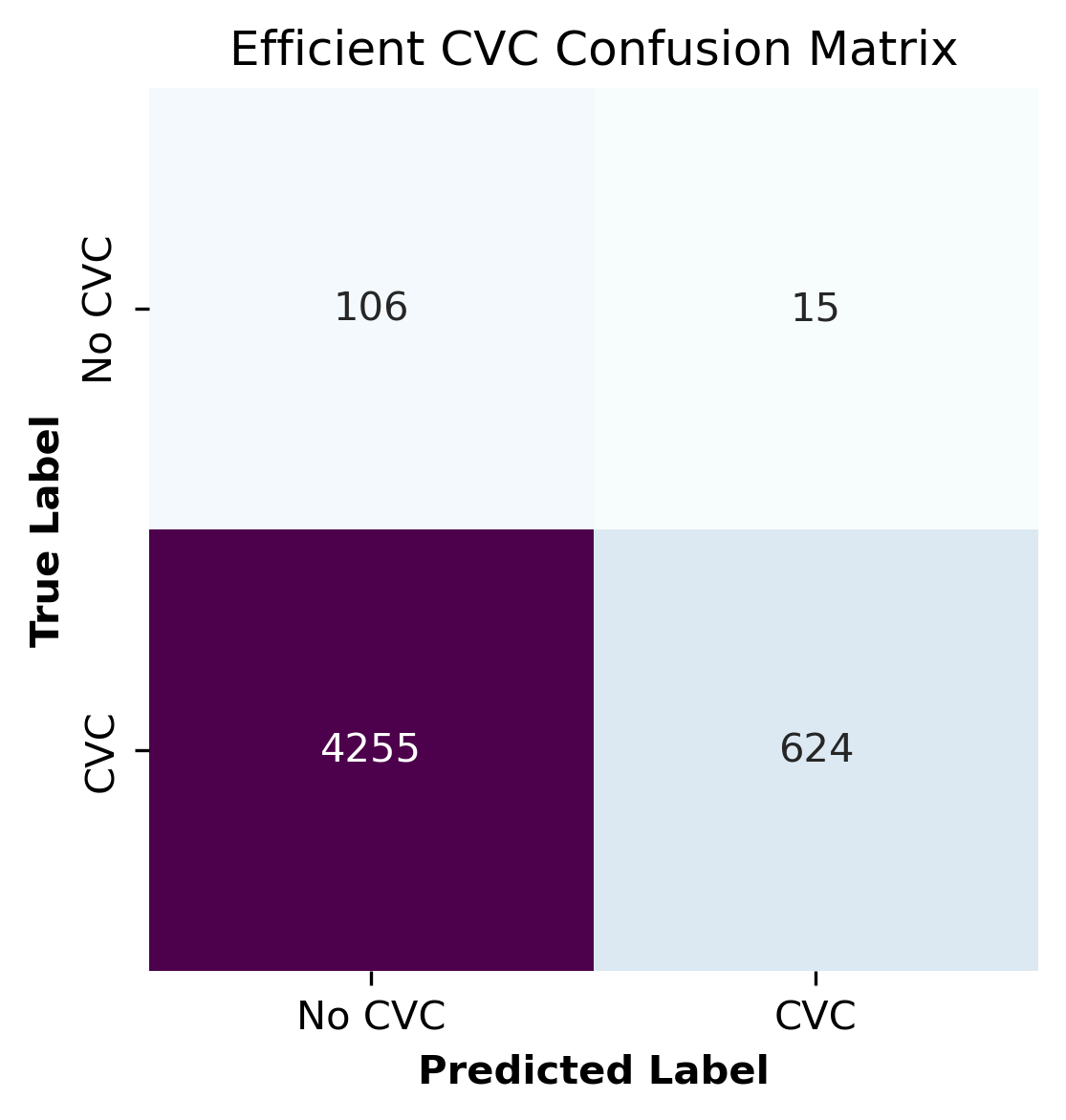
## **Results**

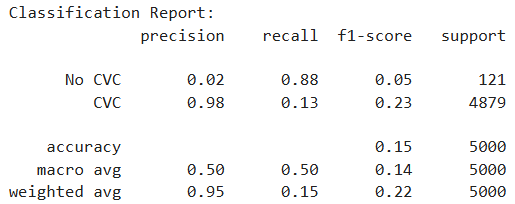
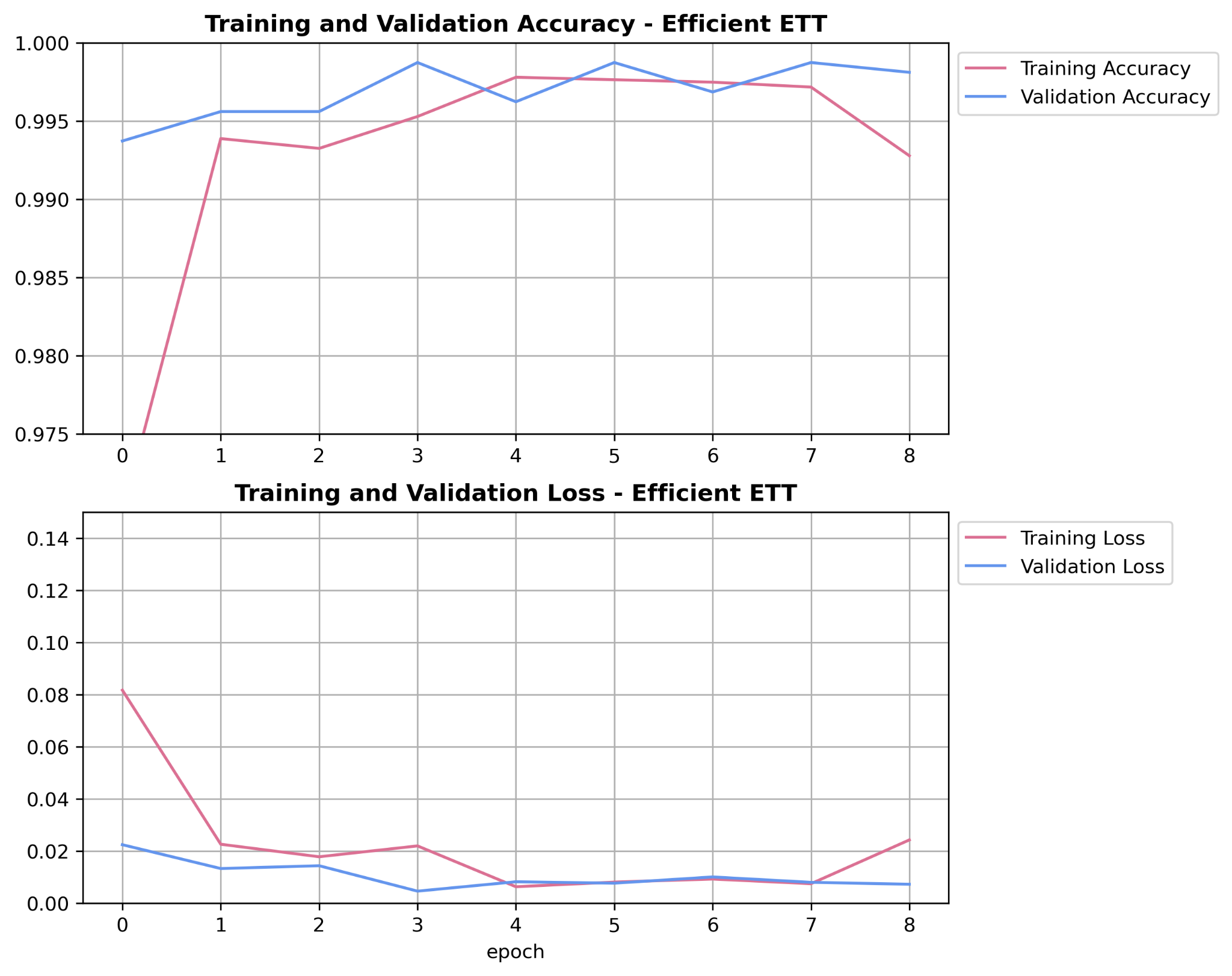
### **EfficientNetB0 CNN Model**

As evidenced by the graph shown in Figures 8, the training and validation accuracies for the three model types consistently range between 0.99 and 1, while the training and validation losses fall between 0 and 0.1. These metrics suggest ideal model performance. However, a closer look revealed a different interpretation.

Classifying CVC Tubes

**Figure 8 : Graphs of training and validation accuracy and loss of CVC EfficientNet Model for epochs trained**

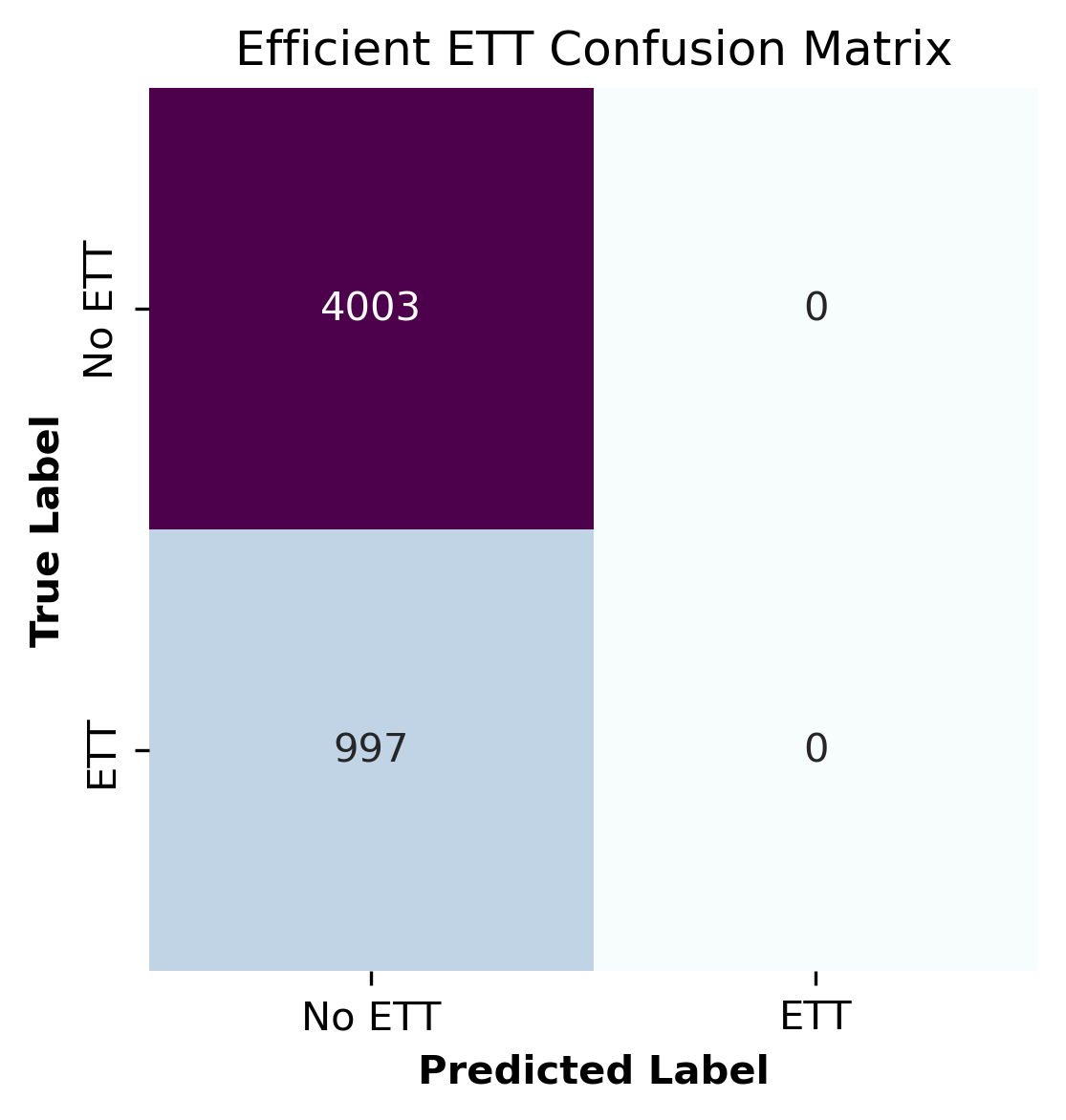
**Figure 9 : Classification report and confusion matrix of CVC EfficientNet Model**

Upon examination of the classification report and confusion matrix as shown in Figure 9, compared to the testing set, the accuracy for CVC tubes is notably poor at 0.15. The precision for images without CVC presence is 0.02, whereas the precision for images with CVC presence is 0.98. This aligns with the confusion matrix, which shows only 106 images predicted as True Positives.  


Classifying ETT Tubes

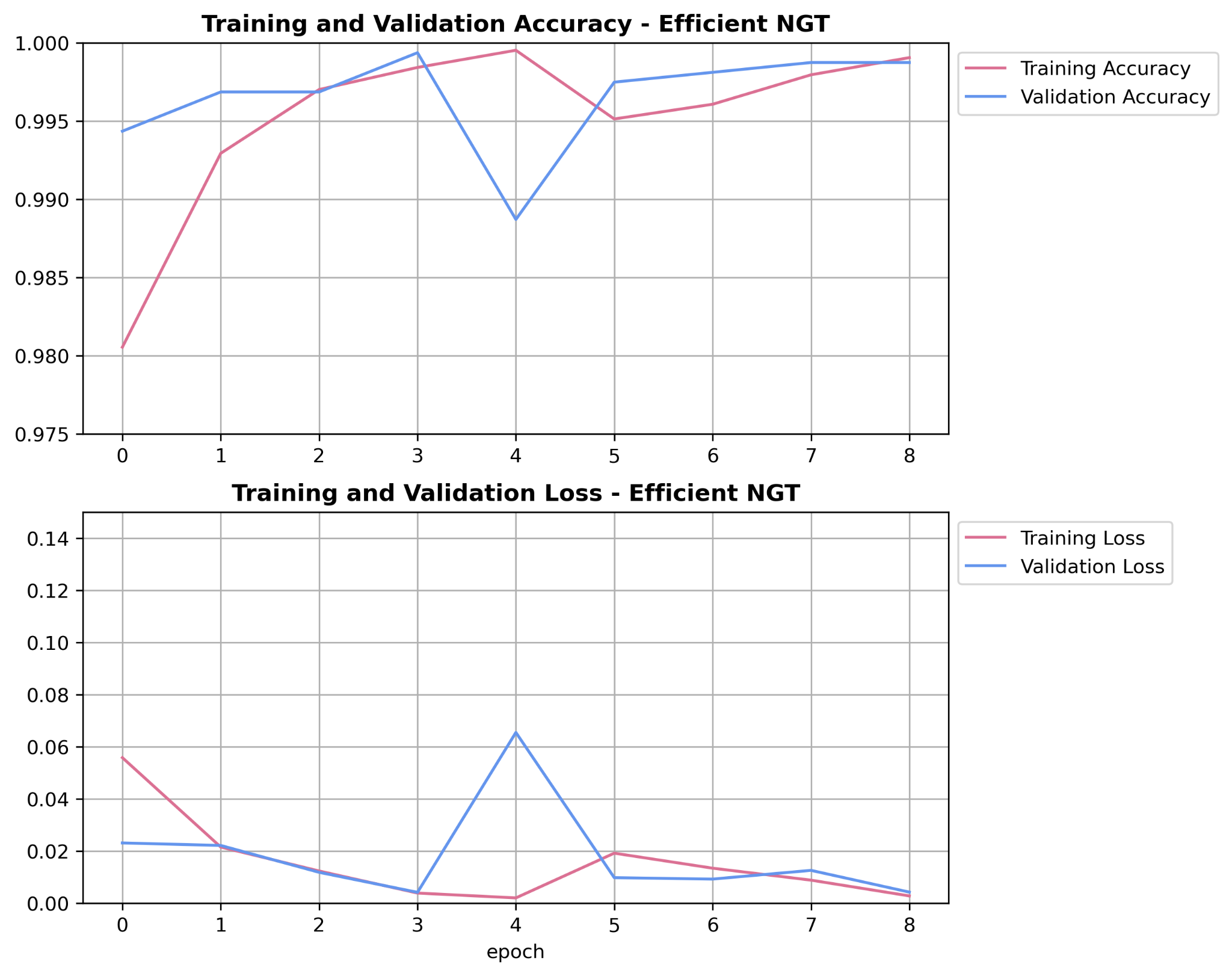
**Figure 10 : Graphs of training and validation accuracy and loss of ETT EfficientNet Model for epochs trained**

As evidenced in Figure 11, the model for ETT tubes demonstrates significantly better accuracy at 80%. However, the precision for images without ETT presence is 0.8, while the precision for images with ETT presence is 0.

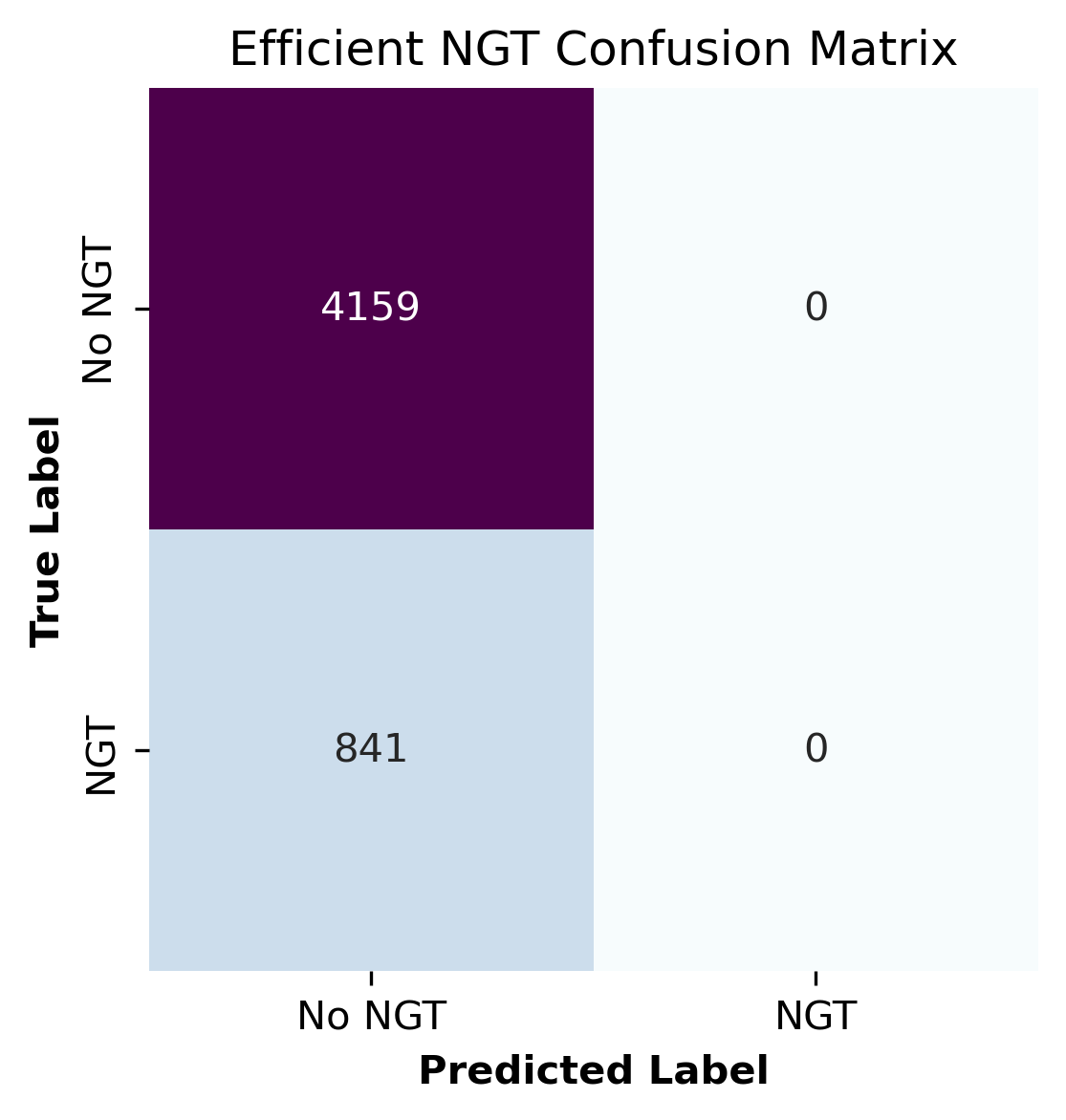
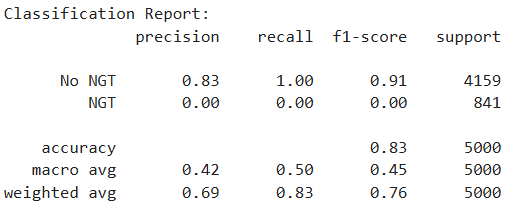


**Figure 11 : Classification report and confusion matrix of ETT EfficientNet Model**

Classifying NGT Tubes



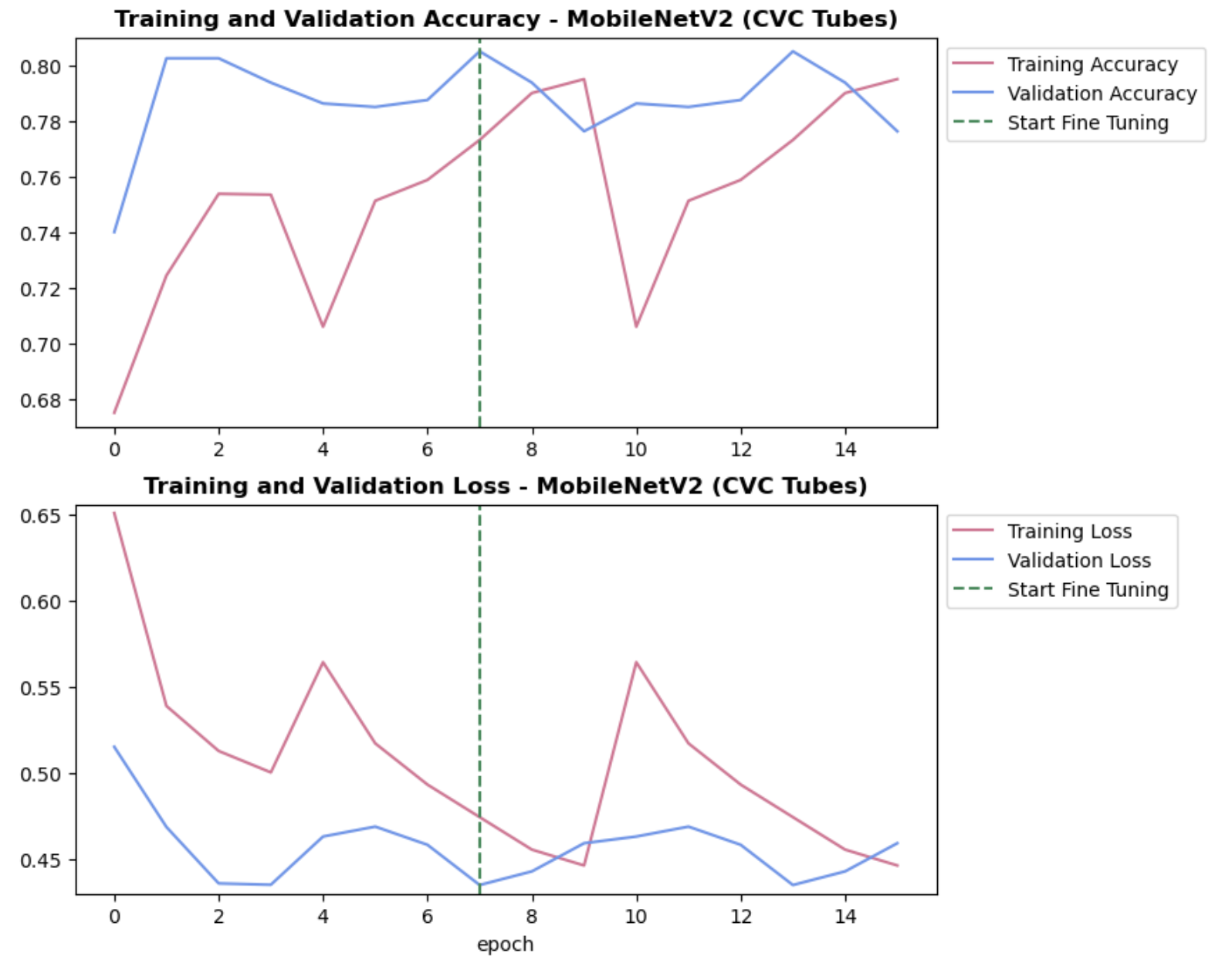
**Figure 12 : Graphs of training and validation accuracy and loss of NGT EfficientNet Model for epochs trained**



**Figure 13 : Classification report and confusion matrix of NGT EfficientNet Model**

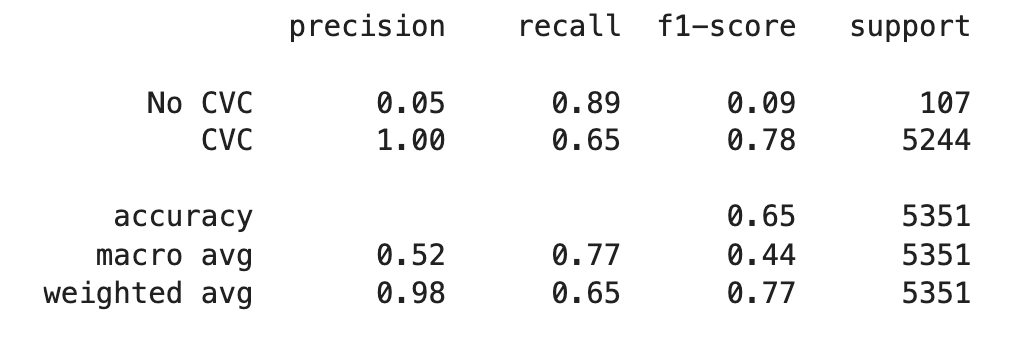
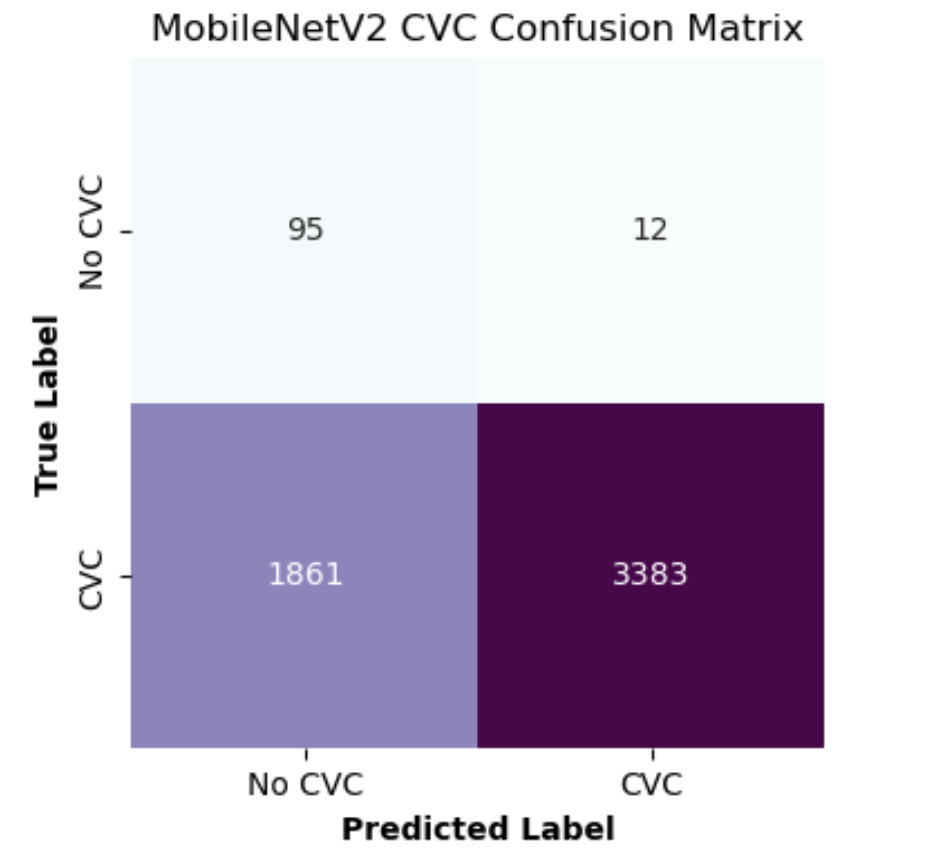
Similar to the ETT model, the NGT tube model exhibits an accuracy of 0.83, with a precision of 0.83 for images without NGT presence and 0 for images with NGT presence.  
It can be concluded that the models for ETT and NGT fail to predict the presence of these tubes in any images, consistently classifying all images as having no tubes. Their high accuracy is attributable to data imbalance, with 80% of the data in both datasets lacking the presence of the respective tube types. Consequently, the models can achieve high accuracy by predicting the absence of tubes in all cases.  
The inverse performance of the CVC-type model, which shows high precision for images with CVC presence, can be attributed to the differing structure of the CVC dataset. Unlike the ETT and NGT datasets, the CVC dataset contains a majority of images with CVC tube presence.  
To address the issue of imbalanced data proportions, an attempt was made to resample the data, equalising the number of images with and without tube presence. Unfortunately, this approach did not yield improved accuracy. Thus, EfficientNet is deemed unsuitable for identifying tube presence in these scenarios.  
In light of these findings, we propose exploring alternative models such as MobileNet and ResNet to potentially achieve better performance in predicting tube presence.

### **MobileNetV2 CNN Model**

Classifying CVC Tubes

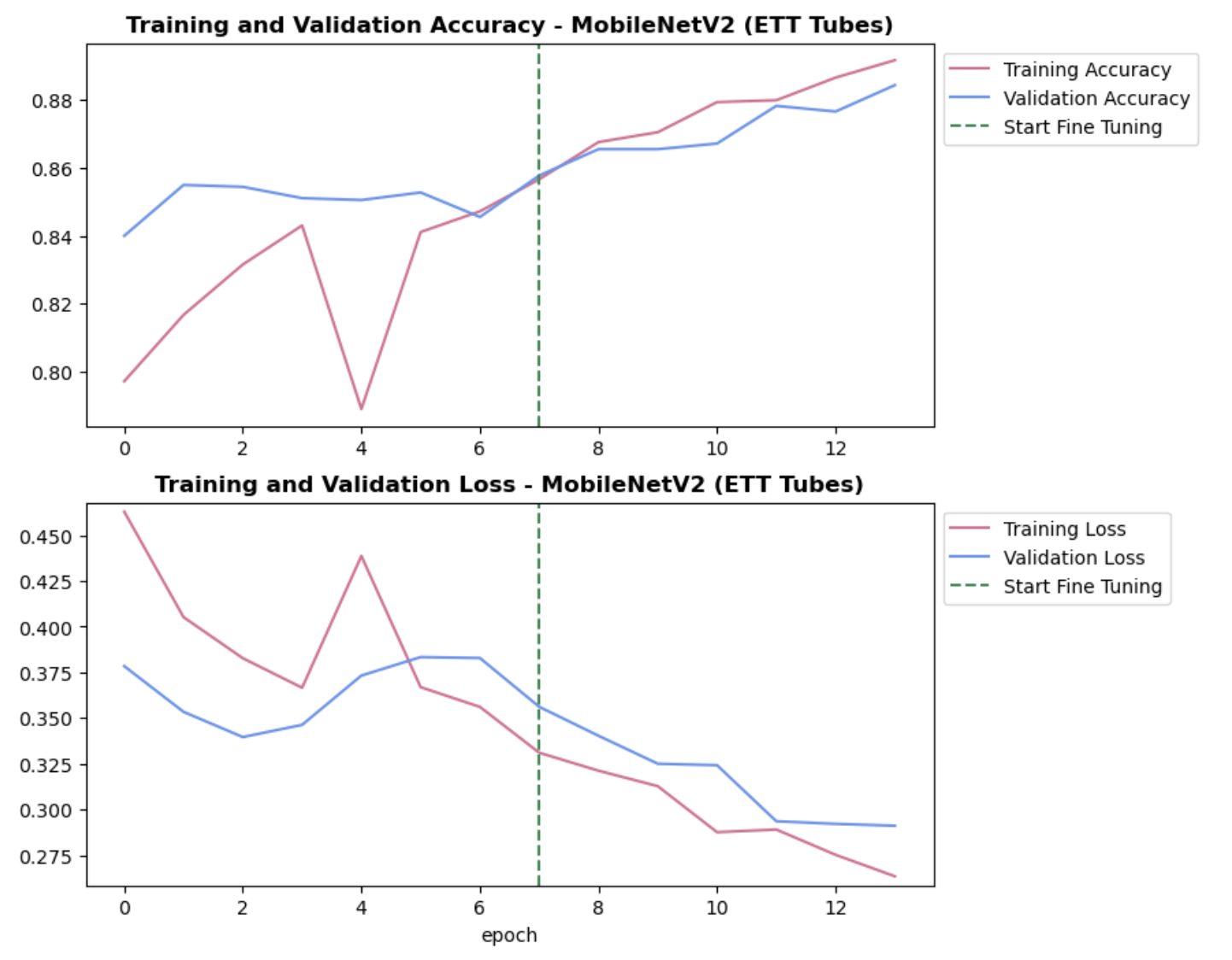
**Figure 14 : Graphs of training and validation accuracy and loss of CVC MobileNet Model for epochs trained**

Inspecting how the model performs classifying the CVC tubes, Figure 14 shows the validation accuracy is consistently high at around 0.8, and the training accuracy goes in an increasing trend with slight dips at 4 and 10 epochs. Additionally, training loss on a decreasing trend with peaks at 4 and 10 epochs, while the validation loss is consistently below 0.5.



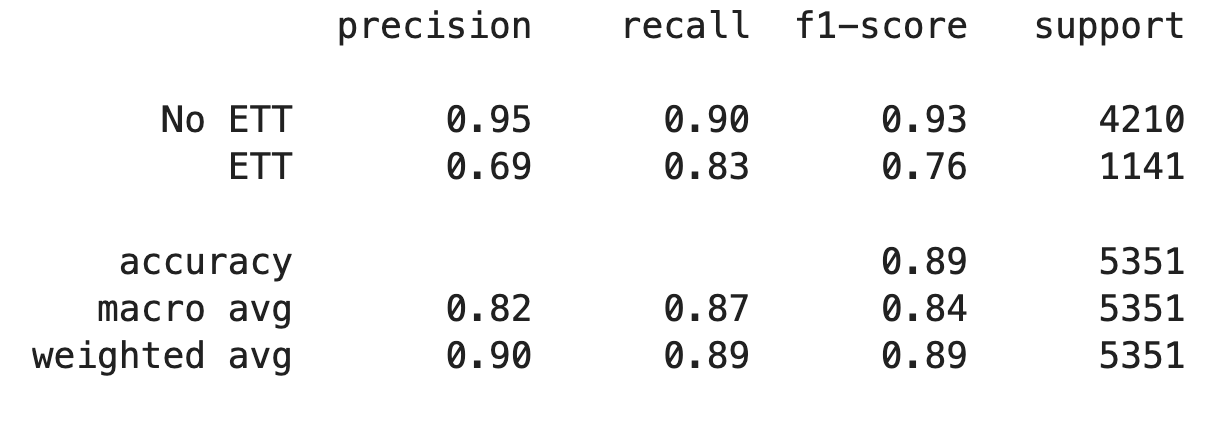
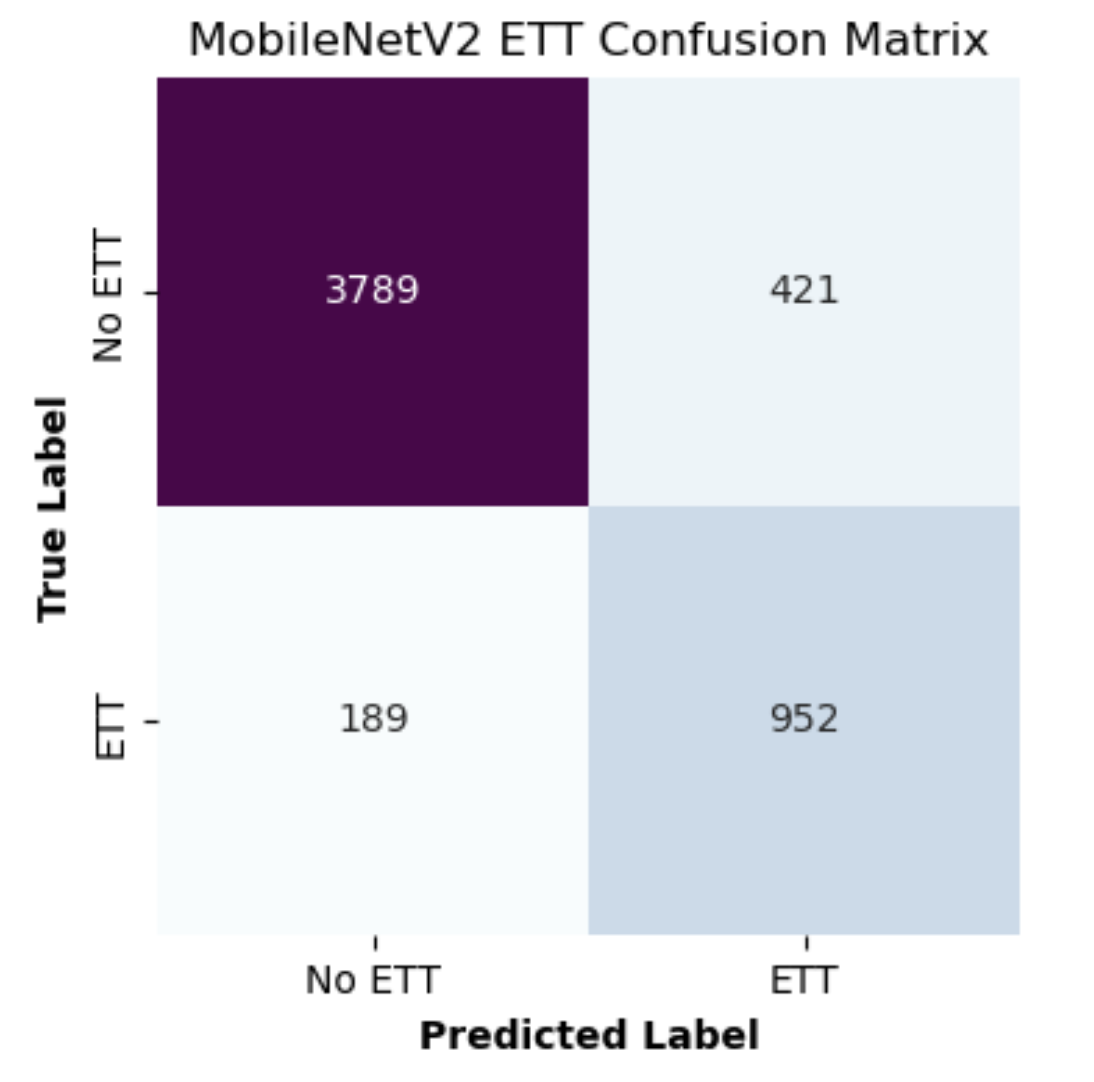
**Figure 15 : Classification report and confusion matrix of CVC MobileNet Model**

Looking at Figure 15, by inspecting the classification report, the CVC class achieved a precision of 1.00, while the No CVC class had a precision of 0.05, meaning the model was easily able to identify the CVC images while struggling at identifying the No CVC images. This can also be seen by the confusion matrix, as there are 1861 false positive cases and only 12 false negative cases, so the high amount of misclassified No CVC images is what made the precision low. The overall testing accuracy is 0.65 meaning that 65% of the dataset was correctly classified, so adjustments could be made to further optimise this model when classifying the CVC images.

Classifying ETT Tubes

**Figure 16 : Graphs of training and validation accuracy and loss of ETT MobileNet Model for epochs trained**

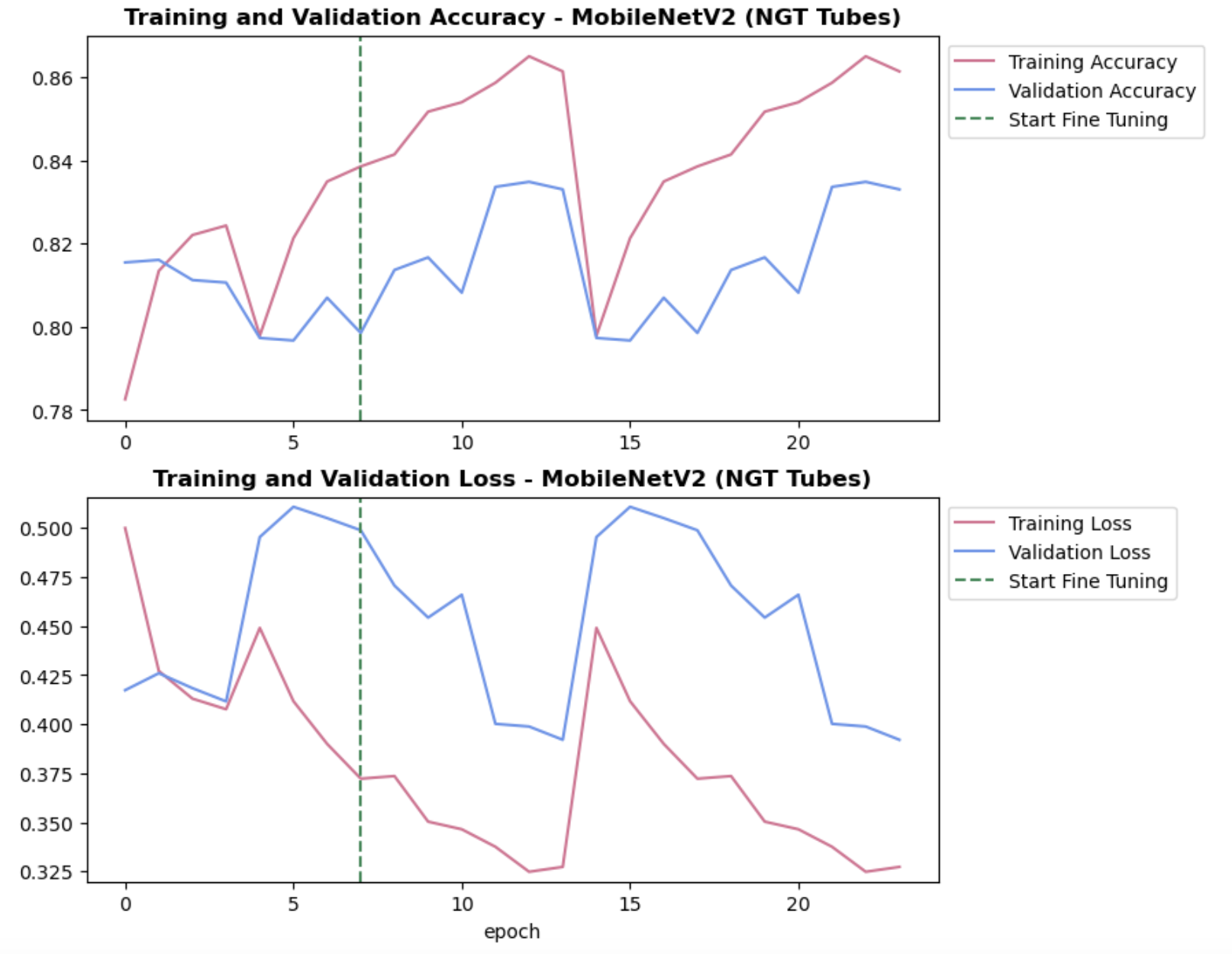
Looking at how the model performs when classifying the ETT tubes, both the training accuracy and validation accuracy go on an increasing trend both going towards a value of 0.88, however, the training accuracy has a dip at 4 epochs (Figure 16). Moreover, the training and validation loss goes on a decreasing trend both going towards a value of 0.275, with the training accuracy having a peak at 4 epochs.



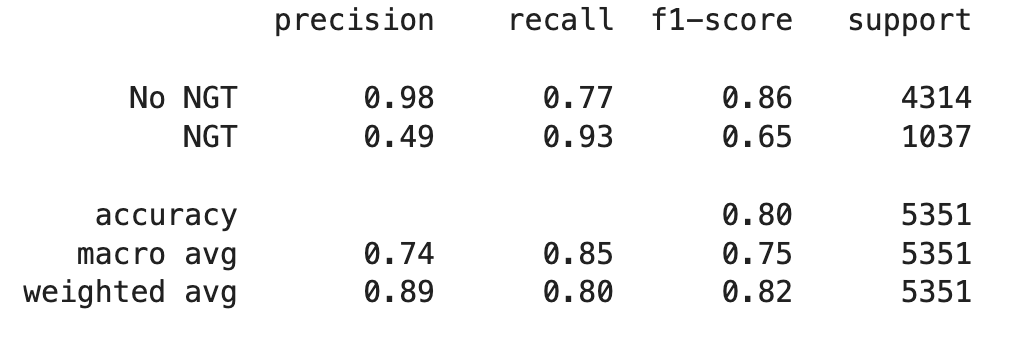
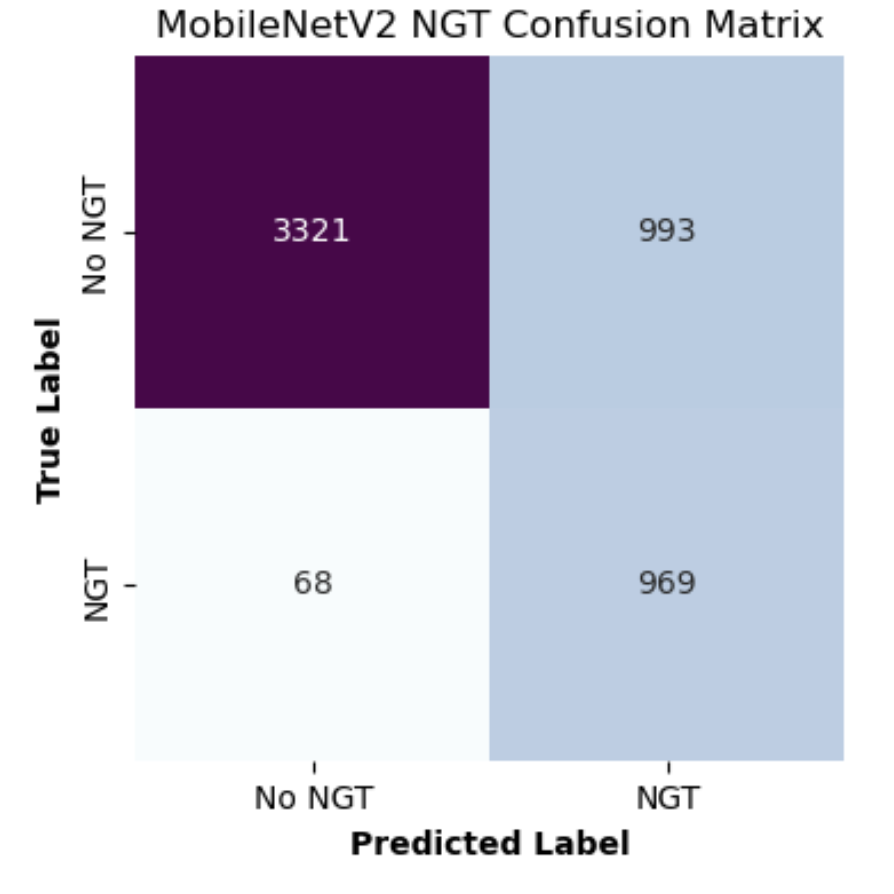
**Figure 17 : Classification report and confusion matrix of ETT MobileNet Model**

The classification report highlights that the ETT class has a precision value of 0.69, while the No ETT class has a precision value of 0.95 (Figure 17). This means that the model struggled more on identifying the ETT tubes, as we have 421 false negative cases by looking at the confusion matrix, while the model was more able to be more accurate on the No ETT class with only 189 false positive cases. The model was able to achieve a testing accuracy of 0.89, meaning that 89% of the dataset was correctly classified, so the model was more accurate with classifying the ETT tubes.

Classifying NGT Tubes



**Figure 18 : Graphs of training and validation accuracy and loss of NGT MobileNet Model for epochs trained**

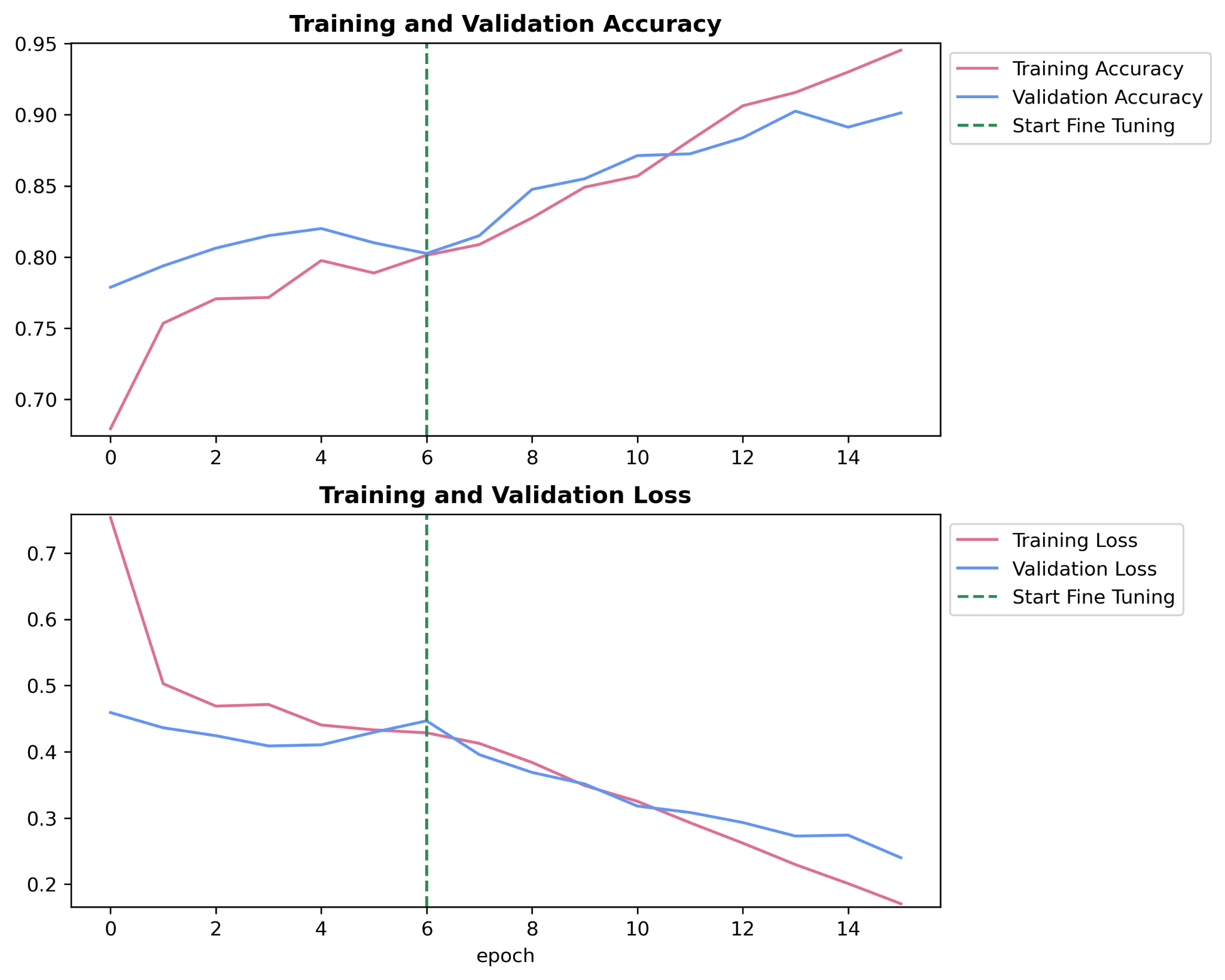


**Figure 19 : Classification report and confusion matrix of NGT MobileNet Model**

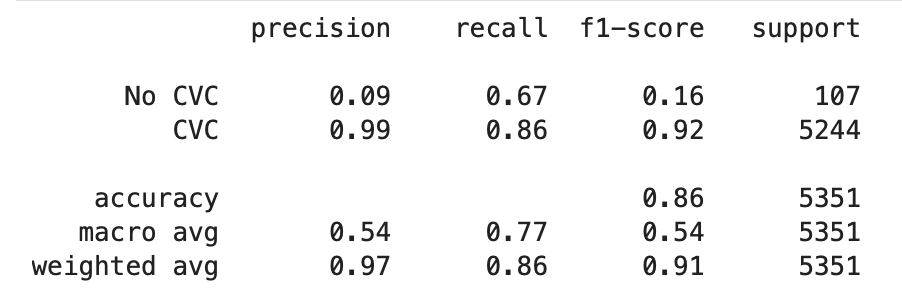
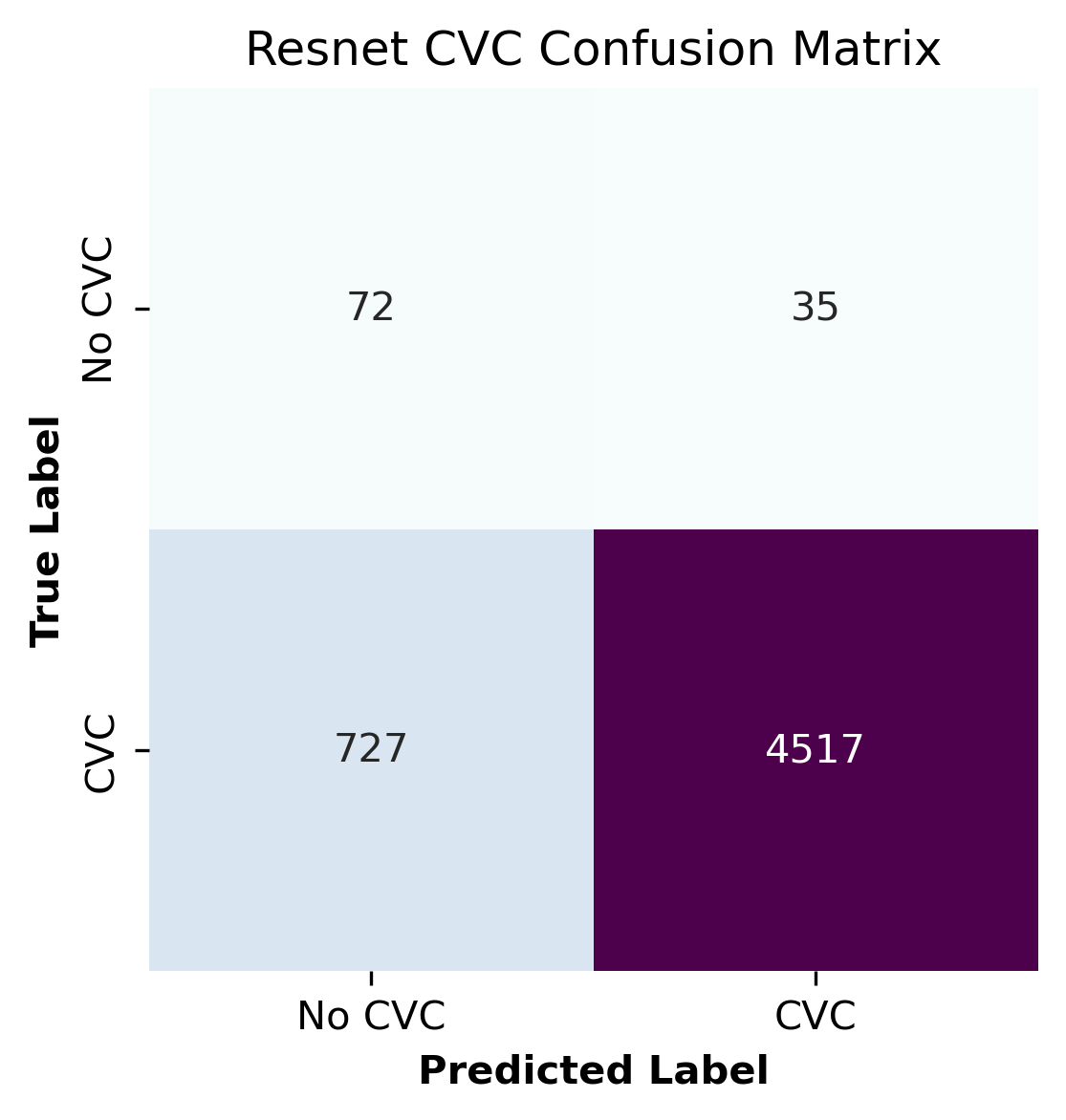
For the model performing on the NGT tubes, both the training and the validation accuracy are both on an increasing trend, however, the training accuracy has dips at around 4 and 14 epochs and peaks at around 0.86, while the validation accuracy has more fluctuation throughout and peaks at approximately 0.84 (Figure 18). The training loss is on a decreasing trend, however does have some peaks, but does steadily approach a value of 0.325 occasionally. The validation loss somewhat does on a decreasing trend, however does approach a value of 0.4. Looking at the classification report in Figure 19, the No NGT class has a precision value of 0.98, which means that the model was accurate in predicting the NGT images, with only 68 false positive cases by inspecting the confusion matrix. While the NGT class has a precision of only 0.49, meaning that the model struggled more on identifying the NGT images, as there are 993 false negative cases by looking at the confusion matrix. The model was able to get a training accuracy of 0.8, meaning that 80% of the dataset was correctly classified, so the model did better than the CVC tubes, but not as good as the ETT tubes.

Overall, the models were able to achieve a moderately high accuracy, however, the model does seem to struggle with some of the classes, such as the No CVC, ETT and NGT class. This could be due to the model having low computational cost meaning it’s not as complex as the other models, so it might not be able to pick up the smaller details of the images, thus the precision is not very high.

### **ResNet50 CNN Model**

Classifying CVC Tubes

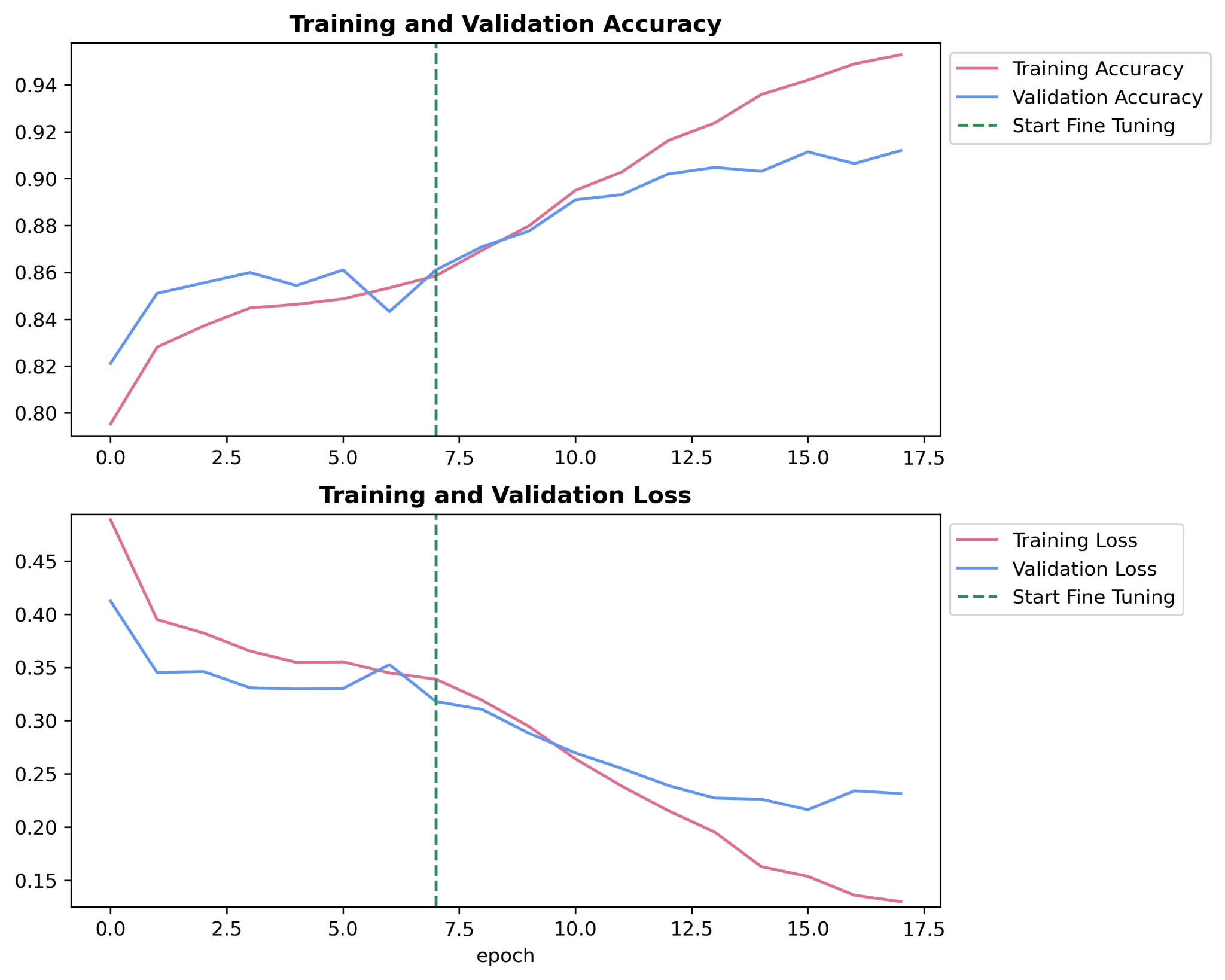
**Figure 20 : Graphs of training and validation accuracy and loss of CVC ResNet Model for epochs trained**



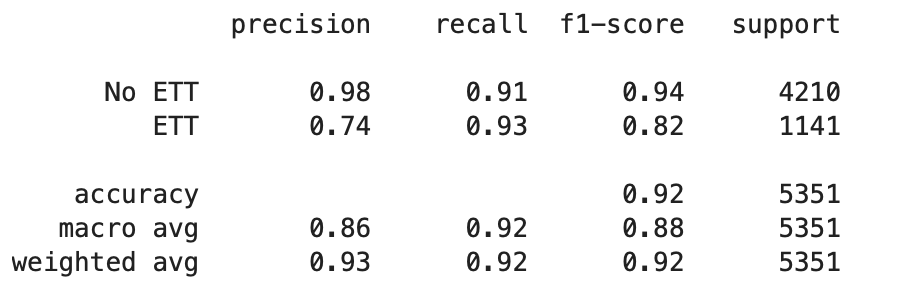
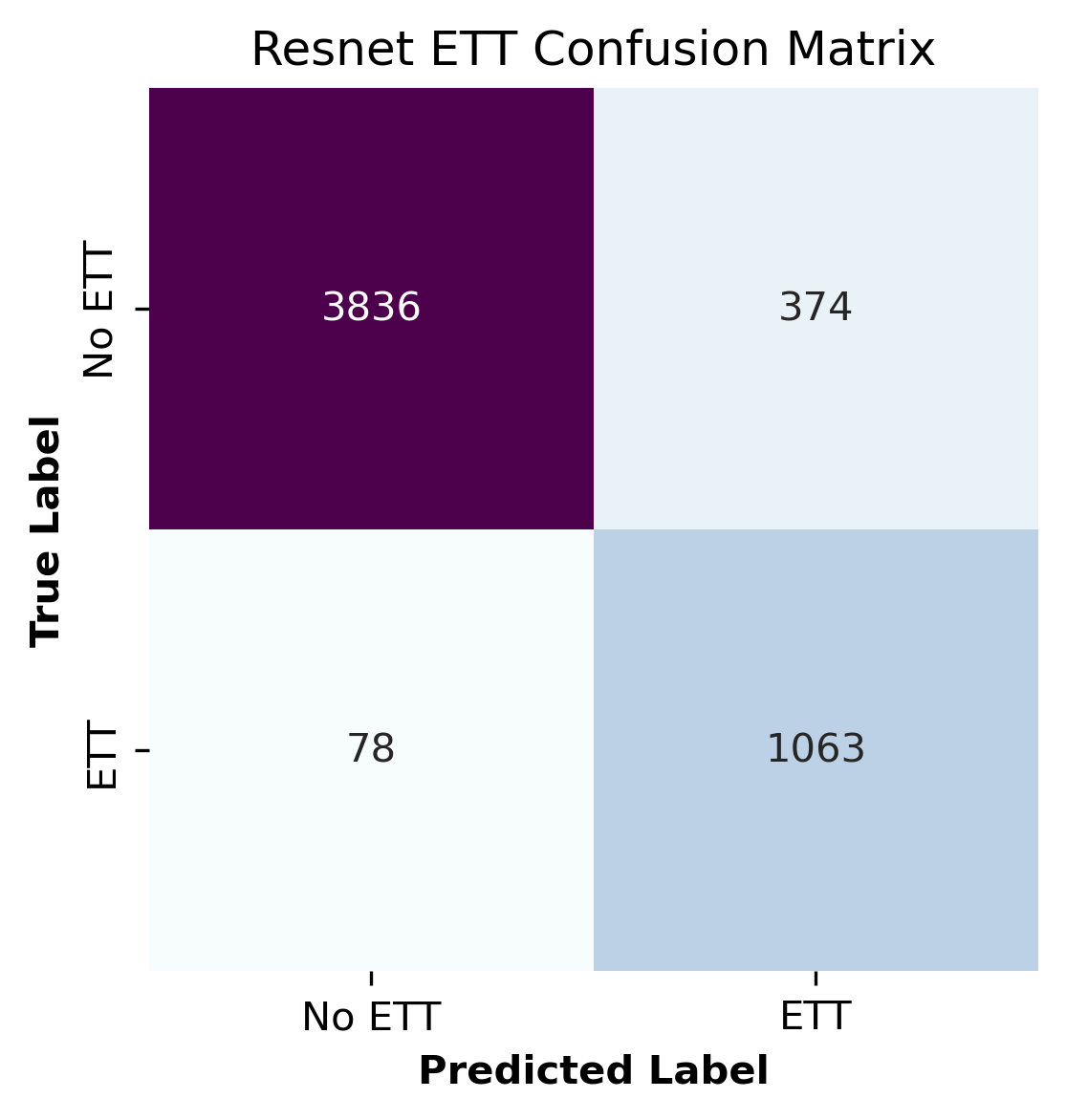
**Figure 21: Classification report and confusion matrix of CVC ResNet Model**

The CVC model demonstrated high accuracies, with the training accuracy increasing from approximately 0.70 to 0.95 and validation accuracy peaking at around 0.89 (Figure 20). After fine-tuning by unfreezing 20 layers, both training and validation accuracies showed significant improvement. In terms of performance on testing data, a testing accuracy of 0.86 was achieved. The model attained a high precision of 0.99 for the CVC class, however, struggled with the No CVC class, reflected in a lower precision of 0.09. This disparity suggests that while the model is good at recognizing CVC images, it may be overfitting to this class due to the lack of diversity and representation in the No CVC images even after upsampling.

Classifying ETT Tubes



**Figure 22: Graphs of training and validation accuracy and loss of ETT ResNet Model for epochs trained**

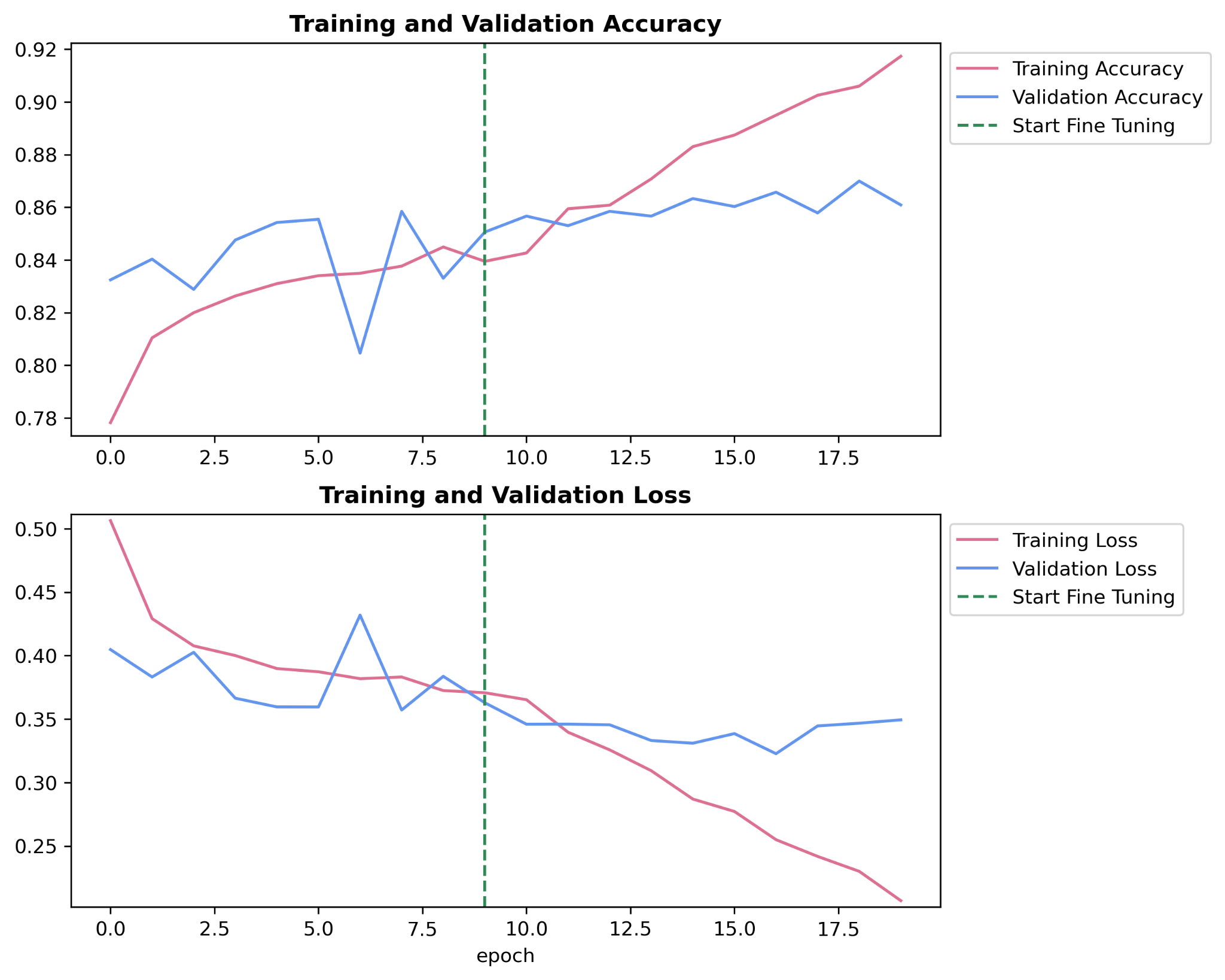


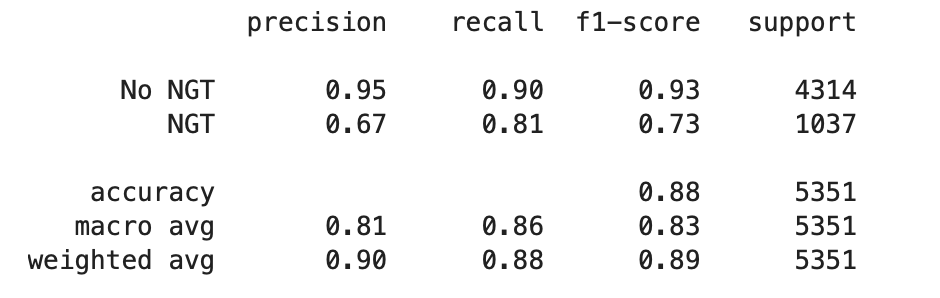
**Figure 23 : Classification report and confusion matrix of ETT ResNet Model**

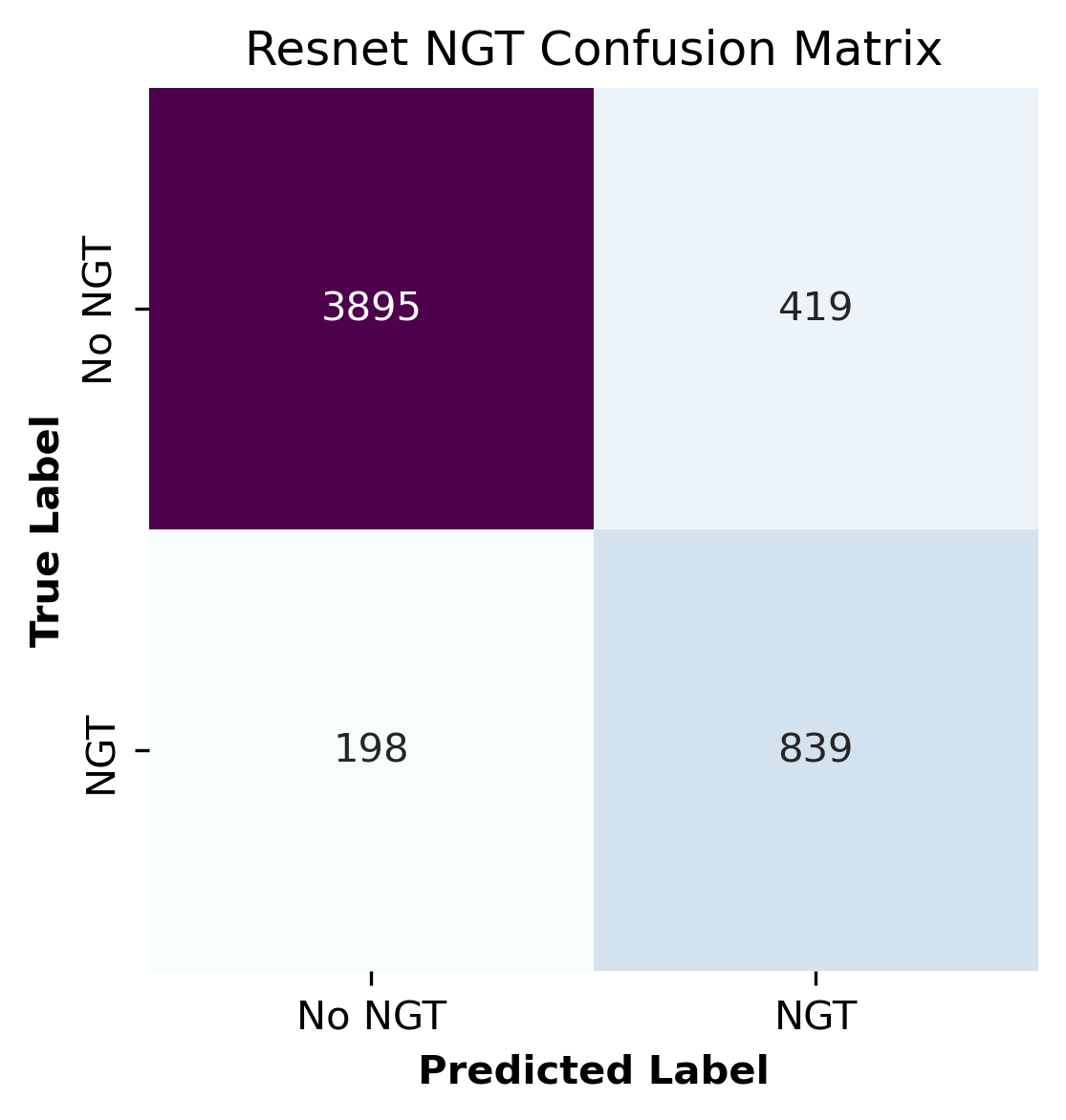
For the ETT model, as shown in Figure 22, the training accuracy improved from around 0.80 to 0.94, with validation accuracy closely following at approximately 0.94. Both training and validation losses steadily decreased, indicating effective model training. As illustrated in Figure 23, the testing accuracy reached 0.92, suggesting that the model was successful in classifying the catheter images. In terms of class performance, the ETT class showed a precision of 0.74 and a recall of 0.93, showing reasonable accuracy in identifying ETT images. In contrast, the No ETT class performed better, with a precision of 0.98 and a recall of 0.91. This highlights that while the model excels in identifying No ETT images, it shows a slight insufficiency in predicting the ETT class.

Classifying NGT Tubes

As shown in Figure 24, the NGT model’s training accuracy improved from around 0.78 to 0.92 and the validation accuracy peaked at 0.85. As illustrated in figure 25, the testing accuracy for the NGT model was 0.88, suggesting strong overall performance. In terms of class performance, the NGT class achieved a precision of 0.67 and a recall of 0.81, indicating a reasonable ability to identify true positives. However, the No NGT class performed better, with a precision of 0.95 and a recall of 0.90. This suggests that the model is much better at accurately classifying No NGT images compared to NGT images, similar to the performance trends observed in the CVC and ETT models.

**Figure 24 : Graphs of training and validation accuracy and loss of NGT ResNet Model for epochs trained**

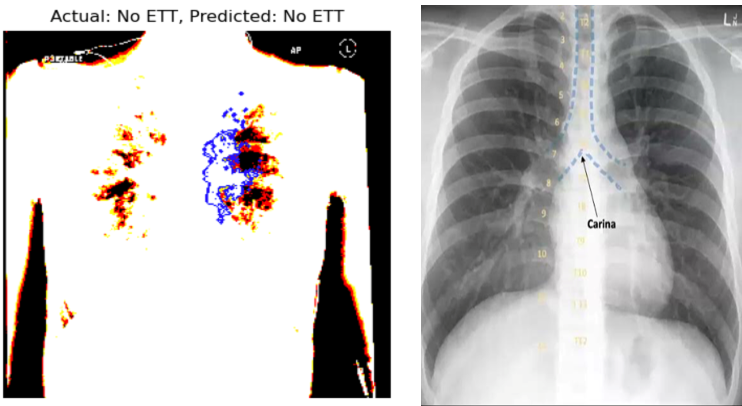


**Figure 25 : Classification report and confusion matrix of NGT ResNet Model**

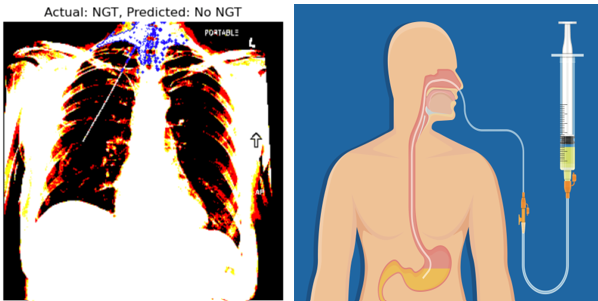
The results from the CVC, ETT, and NGT models illustrate the limitations of relying solely on upsampling to address class imbalance in the dataset. While upsampling was implemented to increase the number of minority class samples, the performance metrics revealed that this approach was insufficient to enhance the model’s generalisation capabilities. Therefore, future strategies should incorporate data augmentation techniques and varied examples to create a more diverse dataset that can better support the model in learning the essential features for accurate classification across all classes.

### **Explainable AI: LIME Interpreter**

To further evaluate model performance we used Local Interpretable Model-Agnostic Explanations (LIME). This technique is popular for its ability to provide interpretable explanations for the predictions of any machine learning model regardless of its underlying algorithm. LIME works by adding slight perturbations to the existing data and recording the changes in consequent prediction made by the relevant model. These changes are weighted by the similarity of the sampled instance to the existing instance. A linear model can then be fitted to the perturbed instances and its coefficients will reflect feature importance in model prediction. Consequently, the feature importance for the original data can be found and an explanation can be given of the model’s prediction.



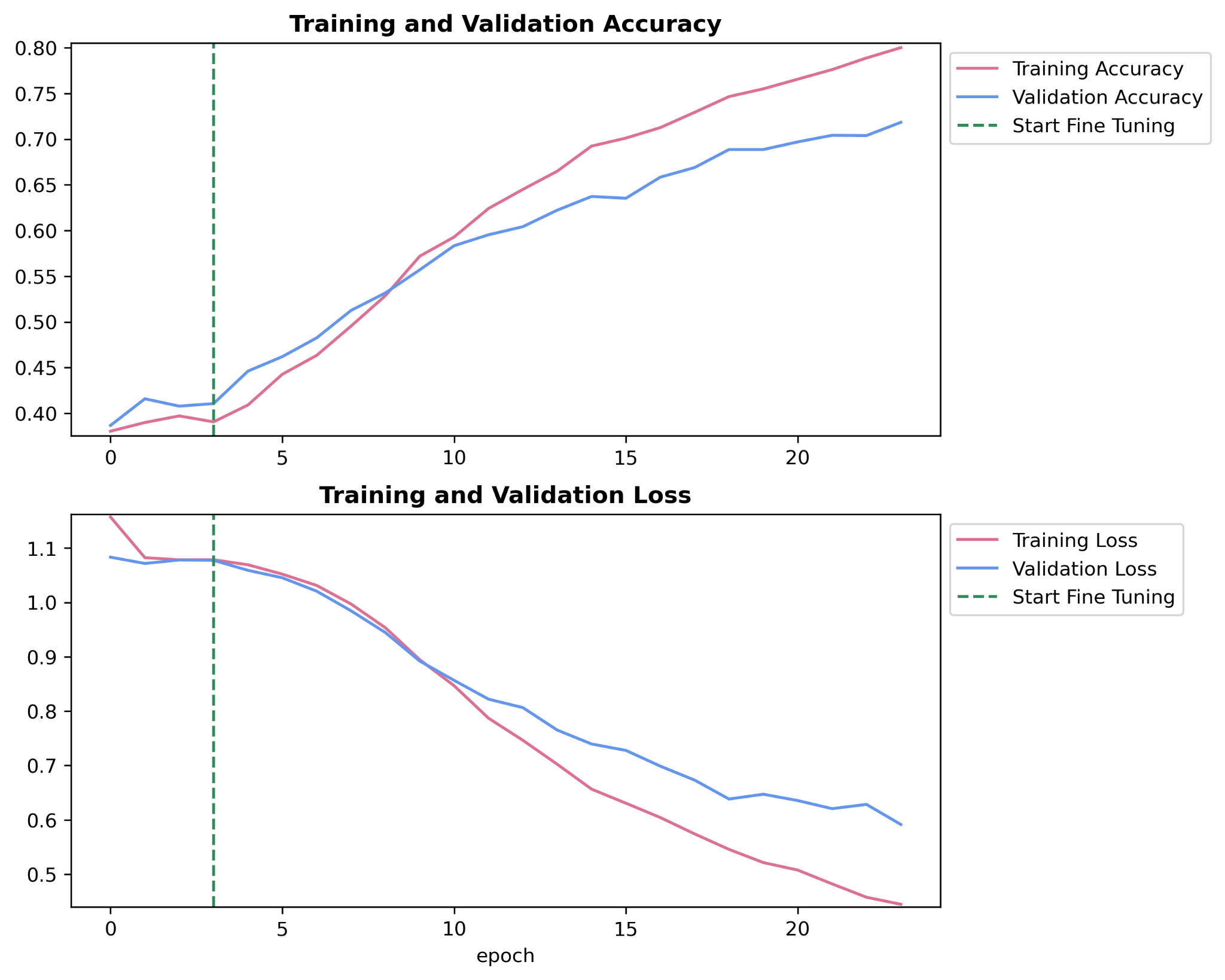
**Figure 26 : Comparison of LIME interpreter of correctly predicted X-ray with no ETT (left) and diagram of expected placement area of ETT tube (right)**

In Figure 26, the image on the left shows the explanation given by the LIME interpreter for the ETT classifier ResNet model, for an image that was correctly classified by the model to contain no ETT tube. The blue region in the image denotes the area in the input image found to have the highest feature importance. Comparing the two images above, it is clear that the area found by the model to have highest importance in making a prediction is consistent with the expected area of placement of an ETT tube. This shows that the model’s predictions are based on relevant features in input images.  


**Figure 27 : Comparison of LIME interpreter of incorrectly predicted X-ray with NGT present (left) and diagram of expected placement area of NGT tube (right)**

In Figure 27, the image on the left shows the explanation given by the LIME interpreter for the NGT classifier ResNet model, for an image that was incorrectly classified by the model to contain no NGT tube. The diagram on the right shows that theoretically an NGT tube would extend from the throat down to the stomach. As shown by the blue region in the input image on the left, the model we created allocated high importance to the throat area in the X-ray.Although the prediction was wrong, the model did attribute importance to a relevant area in the input image.

With more time to train the three ResNet models created, it is reasonable to hypothesise that model accuracy will increase and a LIME interpreter would reveal similar results as the model familiarises more with the data.



**Classification of Positioning of Tube**As we were able to create models with the purpose of detecting the presence of a specific type of catheter in an X-ray, the next step was to create a model that could determine the positioning of a given tube. For this, we decided to create a ResNet50 model for CVC tubes. The instance counts of CVC tubes with positions normal, borderline and abnormal were roughly 10 thousand, 3 thousand and 1 thousand respectively. Through upsampling and downsampling we equalised the instance counts to 6 thousand for each before training.

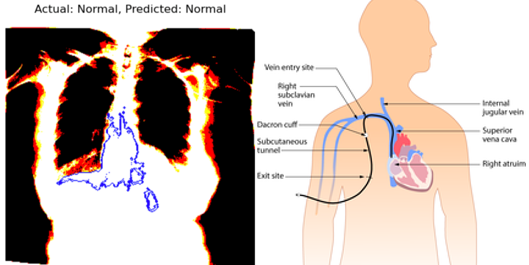
**Figure 28 : Graphs of training and validation accuracy and loss of CVC positioning ResNet Model for epochs trained**

## 

##### 

**Figure 29 : Classification report and confusion matrix of CVC positioning ResNet Model**

Training and validation accuracy increased as the number of epochs increased and for the amount of training done, these values approached approximately 0.8 and 0.7 respectively (Figure 28). Training and validation loss decreased as expected. An accuracy of 0.53 was achieved with the limited training as expected for a harder classification problem. The confusion matrix also shows that normally positioned CVCs were accurately predicted the most. This is most likely due to the fact that the other two classes had to be upsampled through replication. It is hypothesised that with more training and lowered class instance counts, resulting in reduced replication for minority classes, the accuracy could be improved. With more time, these techniques could have been applied to make models for ETTs and NGTs.



**Figure 30 : Comparison of LIME interpreter of correctly predicted X-ray with normal positioning of CVC (left) and diagram of expected placement area of CVC tube (right)**

This model also accurately predicted that the image on the left in Figure 30 had a CVC tube positioned correctly and using LIME we noted that the most important feature in making this prediction as denoted by the blue region was consistent with where the tube should be present in the body theoretically as shown in the diagram above. This signifies that the model is looking for relevant features and with more training could show higher accuracy scores.

## Conclusions

Through this project we have shown that it is possible to model the presence and proper positioning of catheters in X-ray images using deep learning algorithms. By comparing three different models with differing architectures, we concluded that the state of the images inputted, annotated as opposed to raw, and their diversity in terms of class were significant factors in model results.   
The given data set contained approximately 40 thousand images that were of satisfactory quality. The given data frame had no missing values and was easy to interpret and manipulate. Through data preprocessing we removed low quality images and other image instances that did not represent relevant information. Through exploratory data analysis, we better understood the spread of the data and it was made clear there were many class majorities which could lead to risk of overfitting when training models. To combat this, we downsampled majority classes while upsampling minority classes. It was hypothesised that as upsampling was done through replication of existing images, this process did not necessarily increase the diversity of the images intended to train the model. It was understood that this would limit the model’s performance as we were unable to further mitigate risk of overfitting.   
Our approach to modelling of creating two different models, one to detect the presence of a catheter and another to determine the precise positioning of a catheter was found to be highly efficient. This was because otherwise, for images containing multiple tubes which constitutes around half the dataset, a multi-classification model would be required. This is a more complex problem that would require a more complex model architecture. This was avoided both for simplicity’s sake and to adhere to time constraints.

The three models explored; EfficientNetB0, MobileNetV2 and ResNet50, had similar underlying architectures. Efficient NetB0 showed poor precision and accuracy with predictions made being skewed to the majority class. MobileNetV2 showed increased accuracy but was not reliably precise. ResNet50 was best suited to our data and promised reliable results. Even with low accuracy, through LIME we were able to determine

The project faced several challenges and limitations that impacted the efficiency and accuracy of the model. One significant challenge was the extended time required to train the deep learning models. Given the large dataset of around 40 thousand images, training time became a major bottleneck, especially when fine-tuning hyperparameters and testing different models. This extended the time needed for experimentation and iterations, making it harder to quickly assess model performance and refine it further.

Another prominent challenge was overfitting, particularly on annotated images. Any training done on manually labelled, annotated images, were found to lead to overfitting. This issue was worsened by the lack of diversity in the images, especially after upsampling the minority classes. Although upsampling increased the number of images for underrepresented classes, it did not add new variations, so the model often learned repetitive patterns from replicated images, further increasing the risk of overfitting.

Class imbalance was also a major limitation. The dataset showed a clear dominance of certain types of catheters, especially in normal positioning, with fewer cases of abnormal or borderline placements. This imbalance posed a risk of bias in the model, where it could become better at predicting the majority classes while performing poorly on the minority classes.

To overcome these challenges, our team has identified several key strategies for improvement moving forward. One key strategy is to train the model with "all-zero-data," meaning the introduction of negative samples or data points with no catheter present. This would help the model learn more diverse conditions, improving its ability to distinguish between the presence and absence of a catheter more accurately. Additionally, fine-tuning the model will be necessary to improve its generalisation capabilities, reducing overfitting and ensuring it can handle unseen data effectively. Finally, implementing a user interface (UI) for the model would allow users, such as clinicians, to interact with the model's outputs in a practical and intuitive way. This step aims to enhance the model's real-world applicability, ensuring that the predictions are not only accurate but also accessible for end-users in medical settings. The UI would need to focus on user experience to ensure smooth navigation, clarity in presenting results, and ease of integration with medical workflows.

# **Bibliography**

Fu, Y. (2023). Image classification via fine-tuning with EfficientNet. *Keras.* <https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/>

Gibson, F., & Bodenham, A. (2013). Misplaced central venous catheters: applied anatomy and practical management. *British Journal of Anaesthesia*, *110*(3), 333–346. <https://doi.org/10.1093/bja/aes497>

Ilonzo, N., Rao, A., Soundararajan, K., Vouyouka, A., Han, D., Tadros, R., Kim, S. Y., Love, B., Ting, W., Marin, M., & Faries, P. (2020). The importance of a centralized line service during the COVID-19 pandemic. *Journal of Vascular Surgery*, *72*(2), 403–404. <https://doi.org/10.1016/j.jvs.2020.05.016>

Mauri, D., Zafeiri, G., Tsali, L., Chalkidou, A., Zarkavelis, G., Papadaki, A., Filis, P., & Pentheroudakis, G. (2018). Identification of catheter misplacement in early port CVC dysfunction. *Współczesna Onkologia*, *22*(2), 129–134. <https://doi.org/10.5114/wo.2018.77044>

Transfer learning and fine-tuning. *TensorFlow*. (2024). <https://www.tensorflow.org/tutorials/images/transfer_learning>

Yashvi Chandola, Virmani, J., H.S. Bhadauria, & Kumar, P. (2021). Lightweight

end-to-end Pre-trained CNN-based computer-aided classification system design for chest radiographs. Elsevier EBooks, [https://www.sciencedirect.com/topics/computer-science/mobilenetv2#:~:text=MobileNetV2%20is%20a%20lightweight%2053,output%20using%20point%2Dwise%20convolutions.](https://www.sciencedirect.com/topics/computer-science/mobilenetv2#:~:text=MobileNetV2%20is%20a%20lightweight%2053,output%20using%20point%2Dwise%20convolutions)