Catheter final project

Byz: Jae, Aden, Byron, Yohan, Nic, Charles, Avish

Table of contents:

Executive summary

This report examines whether catheters and lines are present and where they're located in chest x-rays. The project seeks to harness machine learning methods for X-ray data analysis, specifically concentrating on catheter placement categorization. The goal is to improve the accuracy of classification that aims for more precise catheter, which could hopefully lead towards a reasonable outcome and prevent unnecessary choices.

Project background

Catheters are essentially flexible tubes that can be inserted into the body to deliver or drain fluids, drugs and in some cases they give access directly for example blood vessels or cavities. In veins or arteries anywhere in the body, all tubes must be inserted precisely to work and avoid complications. It is imperative to identify the presence and location of catheters-lines are one of key elements when reading chest X-rays in medical imaging due improper placement could lead to related complications. Through the use of deep learning models, this project hopes that an image analysis pipeline will assess 30k images to recognize catheters, determine their position in a subject and evaluate whether they are appropriately placed or not. Furthermore, various feature extraction techniques like grayscale will be employed to improve visual perception of abnormalities in X-rays. One of the challenges is identifying whether catheters are misplaced, and prompt recognition has implications for prevention or early surveillance can minimise their consequences: pneumothorax, infections, vascular injury or inadequate drug delivery. --Malpositioned central venous catheter which could lead to fluid entering the wrong anatomical geometry and can be fatal. Hence, for avoiding these undesired complications and safe care of the patient accurate identification and evaluation regarding placement of catheter is warranted.

<u>Catheter Types</u>
CVC - Central Venous Catheter
NGT - Nasogastric Tube
ETT - Endotracheal Tube

Initial EDA uncovered that CVCs were the most common catheter, prompting additional background research to learn more about them.

Central venous catheters, commonly called central line catheters, get their name from the fact that they are inserted into a vein (venous) and from their position running down the centre of the body (central/central line). CVCs are commonly inserted in one of three places; the subclavian vein (under the collarbone), the internal jugular vein (in the neck) or a vein in the arm (called a PICC-peripherally inserted central catheter). It is then threaded through until it is positioned correctly. The intended position for the tip of this catheter is the superior vena cava, the large vein just before the right atrium, see figure 1. Its primary purpose is to deliver drugs, fluids or blood to a patient for emergency or long term treatment. The positioning in the vena cava allows for drugs to be delivered straight to the heart. This means the drug is dispersed around the body straight away, allowing for treatments to occur as soon as possible, which is paramount in emergency situations.

There are a few complications which may arise when inserting a CVC. The most common include damage of the veins, lung complications, heart complications, device dysfunction and infection which aren't always related to the positioning of a CVC. Usually, incorrect placement of a catheter won't cause any major issues. As all veins eventually lead back to the heart, all that will happen is the medication will take longer to get to the heart and around the body. However, malposition can risk complications such as blood clots, a collapsed lung or an irregular heartbeat. Ultrasound can be used for real time imaging of the location of the tip of the catheter as it is inserted, which helps guide it to the correct position. The use of ultrasound greatly improves the safety and quality of CVCs inserted in the neck, but isn't as effective for other insertion sites.

CVC placement is classified as either normal, abnormal or borderline. A normal placement occurs when the tip of the catheter is in the vena cava (top left at the edge of the heart when looking at an X-ray). It is abnormal if it's in the completely wrong position, which includes if it is too deep inside the heart. A borderline placement is when the catheter tip is either just before the vena cava or just after, which is not ideally positioned but not completely wrong either.

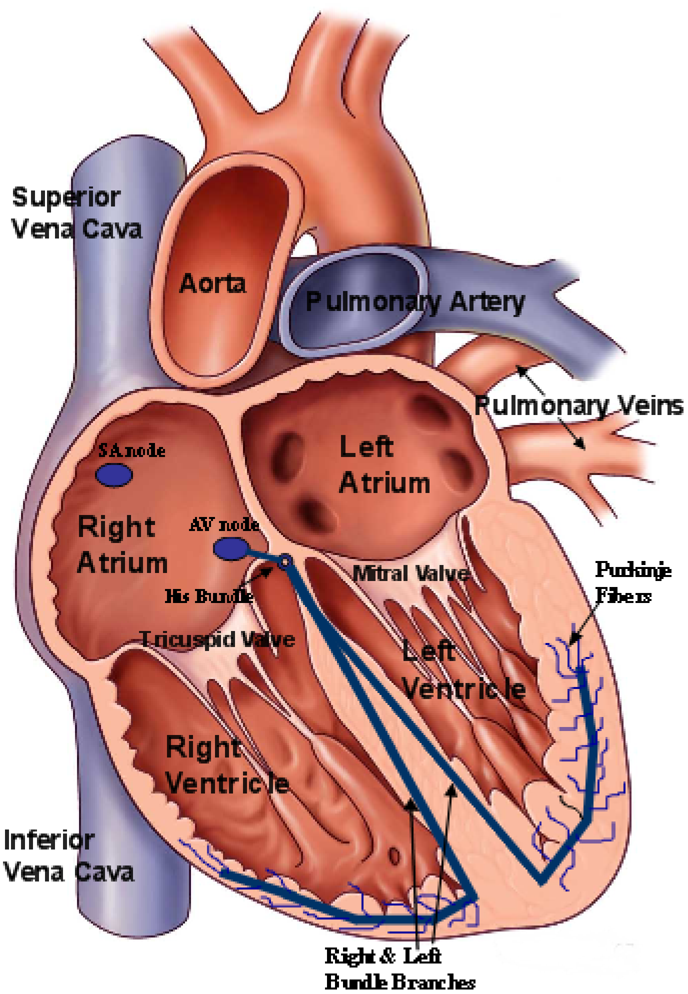


Figure 1, diagram of the heart

Catheter Classification:

ETT - Normal

Normal ETT placement means that the tube is positioned correctly, typically in the mid-trachea, ensuring effective ventilation without causing harm to the airway or lungs. Correct positioning allows for adequate oxygen delivery and patient safety.

ETT - Abnormal

An abnormal endotracheal tube (ETT) placement occurs when the tube is not correctly positioned in the trachea. Mispositioning could result in the tube being inserted too far (into a bronchus), not far enough, or outside the airway. Improper placement can lead to complications such as lung collapse, inadequate ventilation, or airway trauma, making accurate detection critical.

ETT - Borderline

A borderline ETT placement refers to a situation where the tube's position is near the acceptable limits but still raises concerns. This could indicate potential risks of the tube slipping into an abnormal position during patient movement, necessitating close monitoring and possible repositioning.

NGT - Normal

A normal NGT placement indicates that the tube is correctly placed, usually passing through the nose, oesophagus, and into the stomach. Proper placement ensures safe delivery of food, medication, or drainage of stomach contents.

NGT - Abnormal

An abnormal nasogastric tube (NGT) placement indicates that the tube, which is intended to deliver nutrition or remove stomach contents, is positioned incorrectly. Misplacement can result in the tube entering the lungs, leading to aspiration pneumonia or other complications.

NGT - Borderline

A borderline NGT placement suggests that the tube is positioned in a suboptimal but not entirely incorrect location. This placement might require careful monitoring as it poses a risk of migration into an unsafe position.

NGT - Incompletely Imaged

In some cases, the imaging of the NGT might be incomplete, meaning that the entire tube's path is not visible on the x-ray. This makes it difficult to conclusively determine whether the tube is properly placed or not.

CVC - Normal

A normal CVC placement occurs when the catheter is positioned in a large central vein, such as the superior vena cava, ensuring efficient delivery of medications, fluids, or blood samples. Proper placement is crucial for critically ill patients who need rapid, reliable venous access.

CVC - Abnormal

An abnormal central venous catheter (CVC) placement refers to the catheter being in the wrong location, such as outside a central vein or in a small peripheral vein. This can lead to ineffective central venous access, increased risk of thrombosis, or injury to nearby structures like the heart or lungs.

CVC - Borderline

A borderline CVC placement means that the catheter is near the acceptable range but may be at risk of slipping out of its ideal position. Close monitoring is necessary to ensure it remains in place for proper central venous access.

Swan-Ganz Catheter

The Swan-Ganz catheter, also known as a pulmonary artery catheter, is a type of CVC used to measure pressures in the heart and lungs, particularly in critically ill patients. It is inserted through a large central vein and advanced into the pulmonary artery. This catheter helps monitor heart function, guide fluid management, and assess conditions like heart failure. Proper placement is essential for obtaining accurate hemodynamic measurements; malposition can lead to inaccurate readings or damage to the blood vessels.

Data collection

The data that is provided by kaggle contains 4 files: train.csv - contains image IDs, binary labels, and patient IDs, sample_submission.csv - a sample submission file in the correct format, test - test images and train - training images.The dataset comprises of over 30,083 annotated chest X-ray images which are focused on catheter placements, including types such as central venous catheters and endotracheal tubes. The dataset consists of the PatientID for 3,255 patients and Expert annotations from over thirty doctors on 9,095 studies. To identify the different types of catheters the data set uses the different classifications: StudyInstanceUID - unique ID for each image, ETT - Abnormal - endotracheal tube placement abnormal, ETT - Borderline - endotracheal tube placement borderline abnormal, ETT - Normal - endotracheal tube placement normal, NGT - Abnormal - nasogastric tube placement abnormal, NGT - Borderline - nasogastric tube placement borderline abnormal, NGT - Incompletely Imaged - nasogastric tube placement inconclusive due to imaging, NGT - Normal - nasogastric tube placement borderline normal, CVC - Abnormal - central venous catheter placement abnormal, CVC - Borderline - central venous catheter placement borderline abnormal, CVC - Normal - central venous catheter placement normal, Swan Ganz Catheter Present- to show whether or not this case is present in the x ray.

Objectives and methods

The project will involve creating a range of binary classification models, each designed to detect and assess the placement of specific catheter types, ensuring coverage of various tube categories.

To tackle the challenge of identifying poorly matched tubes in chest X-rays, we plan to leverage machine learning by training and testing our model on a dataset of 40,000 chest radiographs. This will allow us to develop a robust model capable of generalising well across a wide range of images. For the classification task, we will employ a Convolutional Neural Network (CNN) architecture, which is highly effective for image-based tasks due to its ability to automatically extract relevant features from raw pixel data. CNNs excel at identifying complex patterns, making them ideal for classifying both the presence of medical tubes and determining whether they are correctly placed. Thus leading us to have two direct approaches which were a binary classification model and a multi label cvc classification.

Key findings

1 Binary Classification Models:The initial CVC binary classification model showed high accuracy due to data imbalance, where the model predominantly predicted the "CVC present" category, as most images contained CVC catheters. Solutions like adding class weights and upsampling the minority class (non-CVC images) resulted in more balanced predictions. For the ETT classification, the imbalance was handled through downsampling. Similarly, the NGT binary classification model performed well without the need for resampling. Overall, the main challenge in the binary models was the class imbalance, which was mitigated through upsampling and downsampling techniques.

2 Multi-label CVC Classification: After upsampling the minority classes (Borderline and Abnormal), the model showed some improvement but still heavily favoured the "Normal" class, misclassifying all samples in the other two categories. The multi-label classification task aimed to distinguish between CVC catheter placements (Normal, Borderline, and Abnormal). Using ResNet50 with a pre-trained architecture and custom layers, the initial model struggled due to class imbalance, favouring the "Normal" class, and failed to differentiate between "Borderline" and "Abnormal" placements.

Conclusion

The binary classification models were trained to identify the presence of CVC, ETT, and NGT catheters in chest X-rays using the ResNet50 architecture. Initial models showed high accuracies, but this was due to severe class imbalance, leading to biased predictions towards the majority classes. By applying resampling techniques such as upsampling and downsampling, the models produced more balanced results with accuracies of around 60-67%.

For multi-label classification, the focus was on differentiating CVC catheter types. Despite using techniques like upsampling and class weighting, the models struggled to identify minority classes, heavily favouring the "Normal" class. Low scores in metrics such as precision and recall, along with ROC analysis, revealed the model’s inability to distinguish between minority classes like "Abnormal" and "Borderline." These results highlight the challenges of class imbalance and suggest that more sophisticated approaches are needed to achieve reliable and unbiased classification in medical imaging tasks.

Introduction

Catheters and lines are essential medical devices used in critical care for the delivery or removal of liquids. They are inserted into blood vessels to administer fluids or medications into the body. However, improper placements of these tubes can lead to serious complications such as re-surgery, vascular injury and improper delivery of medications. The objective of this project was to try to solve the issue of malpositioned by early detection. This can be done by creating a deep learning model that can automatically detect the presence and assess the placement of catheters and lines in chest radiographs. This project aims to leverage modern image analysis techniques to classify the position and accuracy of catheter placement which will aid clinicians identify improperly positioned tubes to prevent adverse patient outcomes

Data origin

The data for this project is sourced from a kaggle competition hosted by the Royal Australian and New Zealand college of radiologists. The dataset consists of X-ray imaging studies accumulated in the picturing archive and communication systems (pacs) which store and retrieve medical images for hospital use. These imaging studies are annotated with detailed metadata, making them acceptable for deep learning tasks providing faster diagnostic processes.

Data collection

The dataset provided contains over 30,000 chest X-rays with corresponding labels that classify the placement of different types of catheters and tubes. Key classifications included:
StudyInstanceUID: A unique identifier for each image in the dataset.
ETT (Endotracheal Tube): The placement of the endotracheal tube, categorised as normal, abnormal, or borderline.
NGT (Nasogastric Tube): The placement of the nasogastric tube, classified similarly as normal, abnormal, borderline, or inconclusive due to incomplete imaging.
CVC (Central Venous Catheter): The classification of the central venous catheter's placement.
PatientID: A unique identifier for each of the 3,255 patients represented in the dataset.
Furthermore, dataset also included expert annotations

Objectives

The primary objective for this project is to develop a deep learning model that can automatically identify the presence and assess the positioning of catheters and lines in chest radiographs. The following consists of our approach to this task,

1. Binary classification on cvc catheter, as they were most prevalent
2. Binary classification on other type of catheters
3. Categorical Multilabel cvc catheter classification
4. Categorical Multilabel all catheter classification
5. Model interpretation using LIME

Train the model over the image data and aim to categorise tubes that are poorly positioned and aid clinicians in early detection and thus improving patient outcomes by reducing human error

Hypothesis/research question

Our research question involves developing a deep learning model to automatically detect the presence and malposition of catheters and tubes in chest radiographs. Achieving this can have significant potential to enhance clinical decision-making and workflow efficiency in hospitals. Automating this process with a robust model could reduce diagnostic errors, minimise the burden on radiologists, and accelerate the treatment process, especially in high-demand healthcare environments. This technology could be integrated into hospital systems to provide real-time alerts, improving both patient outcomes and operational efficiency.

Data quality/ exploratory data analysis

Data quality

To determine the quality of the data, we first identified any missing or NaN values in the data. which ensured that we had complete records for our analysis.There were no NAN values present. Next, we checked for duplicates to eliminate any redundancy, as duplicated entries could skew the training process and model evaluations, luckily there were no dupes as well. Identifying and resolving any data quality issues early on was crucial to ensure that the model was trained on accurate and reliable information.

Data preprocessing

During this phase, we addressed the identified issues of possible duplicate or NAN values. We then compared the main train dataset with the corresponding annotation file to verify consistency between labelled information and x-ray imaging. We also structured the data into a clean and organised dataframe which was helpful for later exploratory data analysis and modelling.

Exploratory data analysis

Our eda focused on helping us understand the distribution of catheter conditions. To distinguish between positioned and malpositioned catheters we made a pie chart, which highlighted 65.3% of catheters in the dataset were correctly positioned, while 34.7% were malpositioned. This suggests a substantial proportion of cases where catheter placement is incorrect

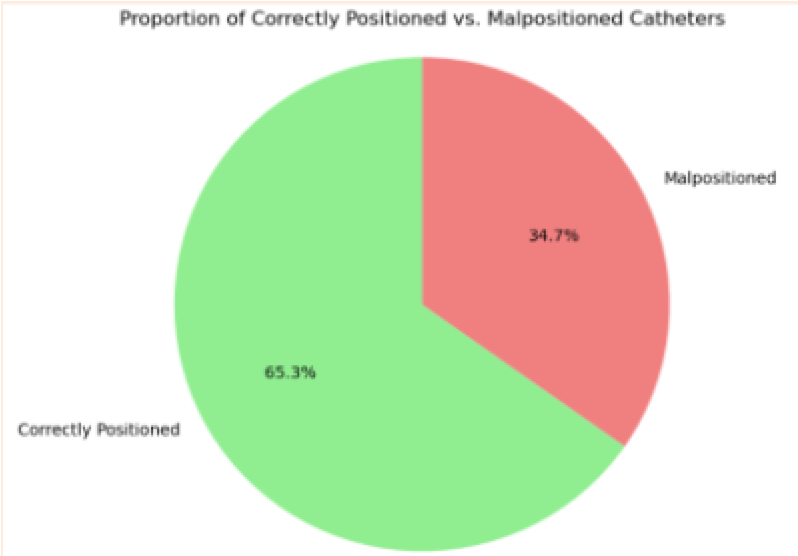


Figure 1 - pie chart

We also wished to get an idea of distribution of number of observations, conducted this by making a histogram which revealed that most patient IDs have only a few associated observations, with a noticeable drop after the initial counts. Out of 3,255 unique patients, 2,858 patients have more than one observation, indicating repeat imaging.

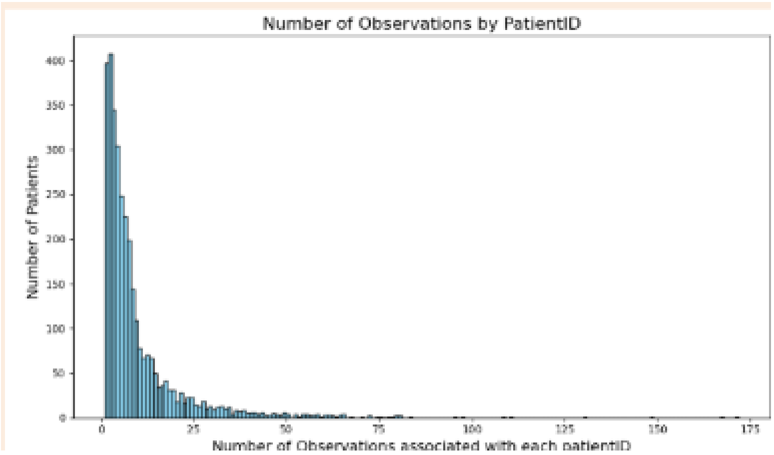


Figure 2 - histogram of number of observations

```
Total number of unique patients: 3255
Number of patients with more than one observation: 2858
```

Figure 3 - unique patients

Next we moved onto looking at the distribution of catheters/catheter conditions, we represented this by a bar chart which showed that "CVC - Normal" is the dominant class, with 21,324 instances, while other conditions like "CVC - Abnormal" (3,195) and "ETT - Abnormal" (79) are underrepresented. This imbalance poses a challenge for classification models as other catheters could be undertrained. This also shifted our approach to use CVC catheter as it was the most prevalent, into our classification task in our modelling.

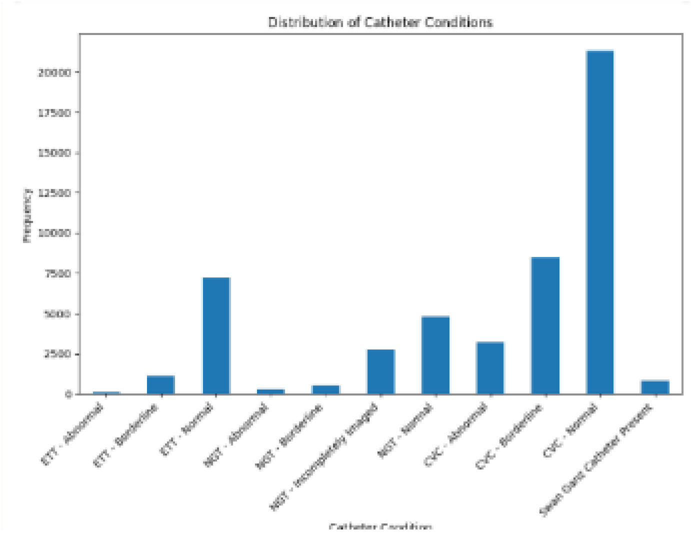


Figure 4 - bar chart of distribution of catheter conditions

	0
ETT - Abnormal	79
ETT - Borderline	1138
ETT - Normal	7240
NGT - Abnormal	279
NGT - Borderline	529
NGT - Incompletely Imaged	2748
NGT - Normal	4797
CVC - Abnormal	3195
CVC - Borderline	8460
CVC - Normal	21324
Swan Ganz Catheter Present	830

Figure 5 - results of barchart

Finally we looked for concentration of the placement of the catheter tip and to see what an outlier would look like. The density plot displays the concentration of catheter tip coordinates. Catheter tips - end points of catheters location are represented in the density plot, where most of the tips are concentrated in the central darker region. revealing common regions within X and Y coordinates. However, the scatter plot on the right shows significant outliers, especially in the extreme X and Y positions. This likely corresponds to CVC-normal since the dataset consists of 70% of that, while the lighter shades represent areas with less tips, which corresponds to malpositioned CVCs or other catheters.

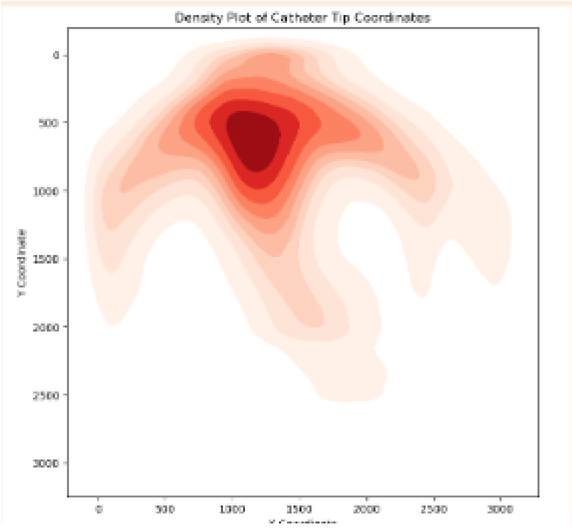


Figure 6 - density plot of catheter tip coordinates

Figure 7 - scatter plot of tip coordinates

Model development + Results

A classification model is a subset of machine learning that predicts the category in which an input belongs. Through labelled training dataset, it learns patterns and can then map input features to specific labels giving seeds of unseen data. Classification models are used in various applications, from spam detection to medical diagnosis and image recognition. In classification tasks, you will find the following common algorithms: Decision Trees Support Vector Machines (SVM) k-Nearest Neighbors (KNN) Neural Networks. Most of the models are focused on being very accurate as to what category a new piece of data should fall into.

Advantages of Classification models There predictions are relatively simple, they merely assign data to some categories built in beforehand. They are similarly applicable in many areas (healthcare, finance or customer service) and you can usually make decisions quickly about unseen data after training the model. In addition, classification models can be well evaluated by Performance metrics that are accuracy, precision, recall and F1-score which gives clear understanding of its effectiveness.

However, classification models also present certain limitations. They are constrained by the fact that they can only assign data to predefined categories, which may be insufficient for problems that require more nuanced or continuous output. Additionally, these models are prone to overfitting, where they perform well on the training data but fail to generalise effectively to new data, especially when the training data is noisy or imbalanced. Another challenge is their dependency on the quality and quantity of labelled data, which directly influences the performance of the model. Furthermore, classification models can introduce or amplify bias present in the training data, raising concerns regarding fairness and accuracy, particularly in sensitive domains like healthcare or criminal justice.

ResNet (Residual Network) is a deep learning architecture designed to address the problem of vanishing gradients in very deep networks. Developed by researchers at Microsoft in 2015, ResNet incorporates residual learning by allowing certain layers to skip connections. Rather than learning the full mapping from input to output, ResNet layers learn the residual, or difference, between the input and desired output, enabling deeper networks without losing critical information.

This architecture has several key advantages. First, it facilitates the training of extremely deep networks—some with over 100 layers—without a significant drop in performance. Second, the skip connections prevent the vanishing gradient problem by preserving the flow of gradients during backpropagation. This design also allows for improved accuracy in tasks such as image classification and object detection, as the network is able to learn residuals more effectively than traditional architectures. Furthermore, pre-trained ResNet models are highly adaptable and can be easily fine-tuned for transfer learning, making them versatile tools for various machine learning tasks.

However, ResNet is more computationally complex than simpler architectures and requires significant resources to train and execute. Additionally, despite its residual connections, deep ResNet models are still susceptible to overfitting, particularly when applied to small datasets without sufficient regularisation. Tuning hyperparameters and determining the optimal depth for the network can also be challenging and resource-intensive.

EfficientNet is a family of convolutional neural networks (CNNs) introduced by Google in 2019. It aims to improve both accuracy and efficiency in image recognition tasks by using a compound scaling approach that uniformly scales depth, width, and resolution. Traditional models typically scale these dimensions individually, leading to inefficiencies, but EfficientNet balances all three, optimising the trade-off between performance and computational cost.

EfficientNet has several distinct advantages. It achieves excellent accuracy with fewer parameters and floating-point operations (FLOPs), making it faster and more resource-efficient than many earlier architectures. By scaling depth, width, and resolution together, it achieves a balanced model design that consistently outperforms older architectures, such as ResNet, across a range of tasks. Moreover, EfficientNet models are highly effective for transfer learning, allowing for quick adaptation to new tasks through fine-tuning.

Despite these strengths, EfficientNet’s architecture is more complex than simpler models like ResNet or VGG, making it more difficult to implement and debug. While it reduces computational costs compared to other models, larger EfficientNet variants still demand substantial hardware resources, particularly when trained on large-scale datasets. Additionally, the model's success is highly dependent on correctly scaling its depth, width, and resolution; errors in scaling can undermine its efficiency and accuracy.

DenseNet (Densely Connected Convolutional Networks) was introduced in 2017 to improve the flow of information and gradients between layers in a deep learning model. Unlike traditional architectures, where each layer only receives input from the previous one, DenseNet connects every layer to all its preceding layers. This dense connectivity ensures that each layer has access to the feature maps and gradients from all earlier layers, improving gradient flow and reducing redundancy.

DenseNet offers several advantages, starting with its ability to mitigate the vanishing gradient problem. The dense connections help ensure that gradients are effectively propagated back through the network, allowing for the successful training of very deep models. DenseNet also uses parameters more efficiently by reusing feature maps from earlier layers, which leads to fewer parameters overall, even though the model has dense connections. This architecture also improves feature propagation, as each layer can leverage information from all preceding layers, which enhances performance in image classification and other tasks. Additionally, DenseNet’s reuse of features helps prevent overfitting, especially in smaller datasets, by discouraging the network from learning redundant or unnecessary features.

However, DenseNet's architecture is memory-intensive, as the dense connections lead to increased memory usage, which can be a limitation when scaling up to larger models or high-resolution inputs. DenseNet also tends to have longer training times due to the complexity of its dense connections, and its implementation can be more challenging compared to simpler architectures, such as ResNet or VGG, because of the need to manage dense feature reuse effectively.

Binary Classification:

Overview:

We decided to begin modelling with a binary classification problem due to the high volume of data we had, which would create issues if a more complex model was chosen. The objective was to train a model that would classify catheters as ‘CVC’ or ‘non-CVC’ based on features in the image. Multiple approaches were taken for this binary classification problem, however they were all done through the prebuilt architecture ResNet50. After completing the CVC classification, the same was also done for ETT and NGT catheters.

CVC binary classification:

To begin with, the prebuilt architecture ResNet was used to create a model that was trained on all 30083 images, with 80% of the images (24067) being used to train the model and 20% of the image (6016) being used as a validation set.

```
Found 24067 validated image filenames belonging to 2 classes.
Found 6016 validated image filenames belonging to 2 classes.
```

Due to the large size of the dataset, we were not able to fully train the model, however, we were getting very high accuracies for each training step, at around 0.9830 (98.30%). As this was our baseline model, we did not expect the accuracy to be this high, and assumed that there was some sort of error or inaccuracy in the training.

```
Epoch 4/10
753/753 [==============================] - ETA: 0s - loss: 0.1133 - accuracy: 0.9759
```

To investigate this issue, we did some manipulation on the dataset by filtering the images to only those that have annotations (the catheters are plotted out). This would make it easier to visualise the catheters in these images, as well as reducing the size of the data to make it more digestible for the model (9095 images remaining). Upon training this model, and looking at the confusion matrix, we saw that the model was only predicting catheters to be CVC and none to be non-CVC. This indicated that the issue was a result of the imbalance in the data. Specifically, there were 8826 images that contained a CVC catheter, and 229 images that did not contain a CVC catheter. This led to the model only predicting ‘CVC present’ and the extremely high accuracy.

```
[9]: unique_annotated_images =  train_annot['StudyInstanceUID'].unique()
     annotated = train_data[train_data['StudyInstanceUID'].isin(unique_annotated_images)]
     annotated.head(1)
```

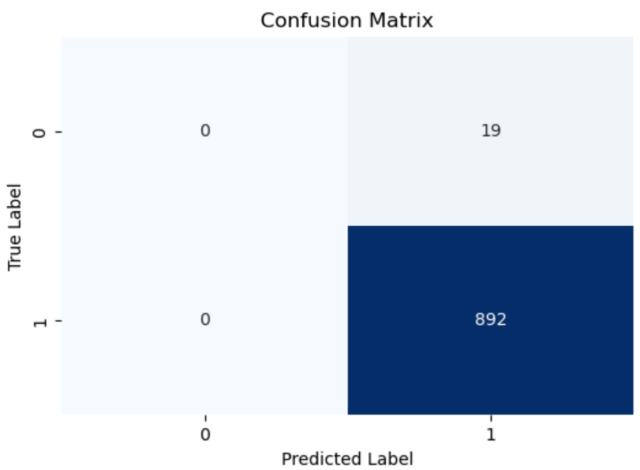
[9]:

	StudyInstanceUID	ETT - Abnormal	ETT - Bord
2	1.2.826.0.1.3680043.8.498.23819260719748494858948050424870692577	0	

```
df['CVC_numeric'].value_counts()
```

```
CVC_numeric
1    8866
0     229
Name: count, dtype: int64
```

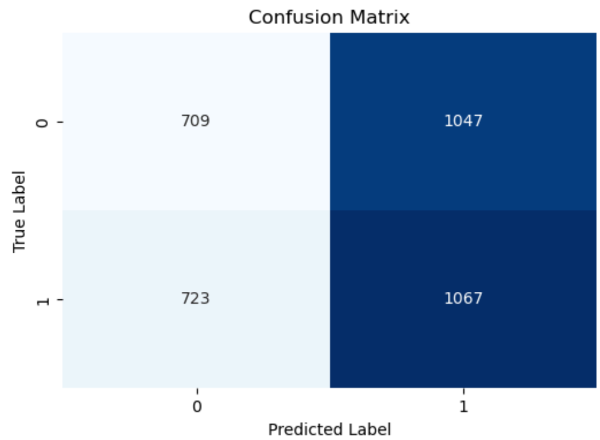
There are 8866 CVC type, 229 non-CVC type



Moving forward, we looked at 2 solutions to the data imbalance problem. The first solution was to add weights to the training of the model. This would penalise the accuracy more when the model predicted ‘CVC present’ incorrectly. However, upon training this model, the accuracy remained extremely high at around 97%. This led us to our second solution, which was to upsample the number of ‘non-CVC’ images in the data using the resample method. This gave us a balanced dataset with 8826 CVC images, and 8826 non-CVC images to train the model on. This model gave us an accuracy of 0.596 (59.6%) which was much closer to our expectation of the model.

```
non_CVC = df[df['CVC_numeric'] == 0]
CVC = df[df['CVC_numeric'] == 1]


non_CVC_bal = resample(non_CVC,
                       replace=True,
                       n_samples=int(len(CVC)),
                       random_state=42)
CVC_bal = resample(CVC,
                   replace=True,
                   n_samples=int(len(CVC)),
                   random_state=42)

bal_df = pd.concat([CVC_bal, non_CVC_bal])
bal_df = bal_df.sample(frac=1, random_state=42).reset_index(drop=True)
bal_df['CVC_numeric'].value_counts()
```

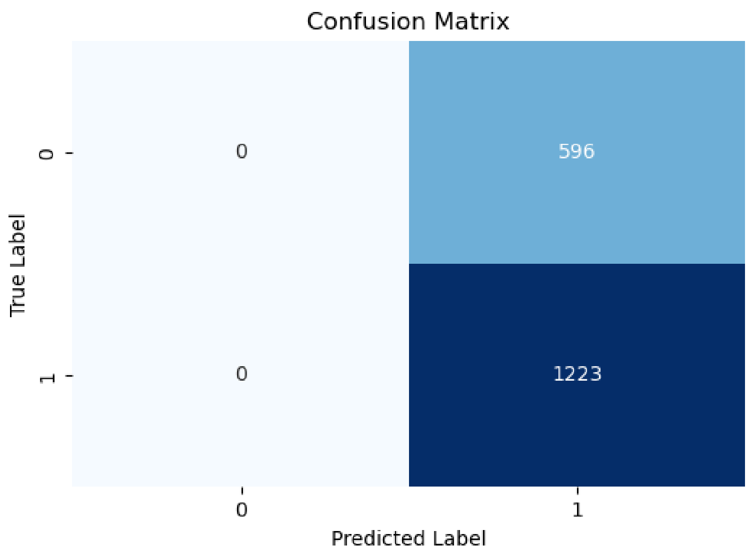
```
CVC_numeric
0    8866
1    8866
Name: count, dtype: int64
```



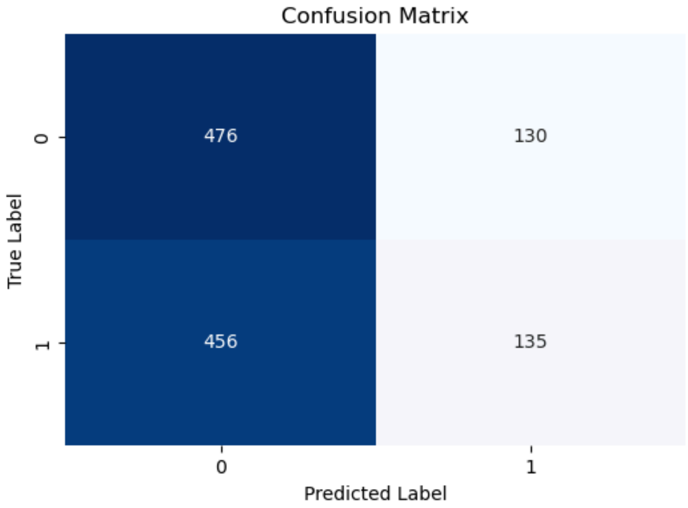

```
111/111 [==============================] - 442s 4s/step - loss: 0.6705 - accuracy: 0.5964
Validation Loss: 0.6705045700073242
Validation Accuracy: 0.596446692943573
```

ETT binary classification:

With the results obtained from the CVC binary model, it was decided that we would apply the same method to see if we could classify ETT catheters as either ETT present or ETT not present. The same setup was used with a 80% training and 20% validation split on the 9095 images. Out of these 9095 images, 6101 images did not have an ETT catheter present, while the other 2994 images did. As ETT catheters were significantly less prevalent to CVC, we expected that the model may return an even lower accuracy. However, after training the model, an accuracy of 67% was returned. However, similarly to the original CVC model, we saw that the model was simply predicting ETT present due to the imbalance in the data.



Hence, a solution of similar fashion was applied. Instead of upsampling like the CVC model, we decided to downsample the number of ‘ETT not present’ images to 2994. This was mainly done to reduce the amount of time it took to train the model, but to also train the model on fewer images where the ETT catheter was not present.



As expected, this reduced our accuracy to 0.618 (61.8%). However, we believe that this was a more accurate representation of our model’s capabilities.

```
38/38 [==============================] - 146s 4s/step - loss: 0.6691 - accuracy: 0.6182
Validation Loss: 0.6691277027130127
Validation Accuracy: 0.6182122230529785
```

NGT binary classification:

The final catheter type that we tried to binary classify was the NGT catheter. Upon inspection, the balance between NGT present (3177) and NGT not present (5918) catheters was relatively equal. This is why we tried modelling it without resampling the data. The same setup was used with a 80% training and 20% validation split on the 9095 images.
This model gave us a 0.654 (65.4%) accuracy, which we were pleased with as we did not have to do any resampling.

```
57/57 [==============================] - 193s 3s/step - loss: 0.6139 - accuracy: 0.6537
Validation Loss: 0.613928496837616
Validation Accuracy: 0.6536558270454407
```

Binary Classification summary:

A binary classification model was trained for each of the CVC, ETT and NGT catheter types to try and determine if each was present or not in all 9095 annotated images. In doing this, a persisting issue was the imbalance in the data between ‘catheter type present’ and ‘catheter type not present’, which caused the models to predict in a biassed manner. This was solved primarily through resample (upsampling and downsampling) which balanced the data and allowed the model to predict more accurately.

Categorical: Multi-label CVC Catheter Classification Before Upsampling

Setup:

The focus of this project was exclusively on CVC catheter data. Although ResNet50 is known for being memory-efficient, the dataset contained 17,999 X-ray images, which caused software crashes due to memory limitations. To address this, we removed instances where multiple catheters were present within the same study. Furthermore, we only worked images that had corresponding annotations, reducing the data to 8,866 x-ray images. We believed that this would allow the model to learn to distinguish between different catheters.

Next, we created categorical columns to classify the CVC catheters.

```
cvc_class
normal        5147
borderline    2563
abnormal      1156
Name: count, dtype: int64
```

To ensure consistency across the dataset, we used an image data generator to preprocess and augment the training and validation X-rays so that all images were the same size and shared the same properties. We set class mode to categorical for our multiple catheter type classification

Modelling:

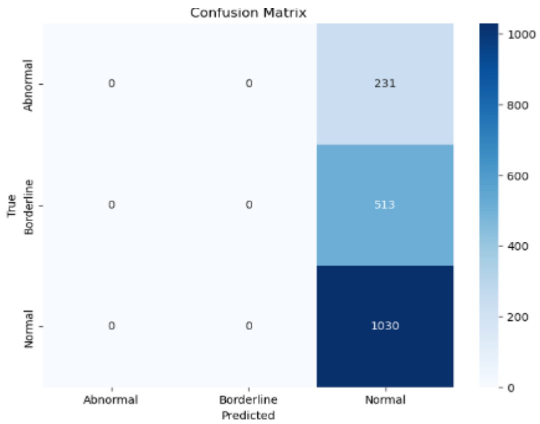
The model architecture was based on ResNet50, utilising its 50 pre-trained layers as the backbone. On top of ResNet50, we added custom layers that included a global average pooling layer, a dense layer with 128 units, a dropout layer to prevent overfitting, and a final dense output layer for multi-class classification. In total, the model had 50 layers from ResNet50 and 4 custom layers. The pre-trained layers were frozen to retain the weights learned during training on the ImageNet dataset, expecting that this would help prevent overfitting.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resnet50 (Functional)       (None, 7, 7, 2048)        23587712

 global_average_pooling2d (G  (None, 2048)             0
 lobalAveragePooling2D)

 dense (Dense)               (None, 128)               262272

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 3)                 387

=================================================================
Total params: 23,850,371
Trainable params: 262,659
Non-trainable params: 23,587,712
_________________________________________________________________
```

Modelling evaluations:

During training, the model achieved a validation loss of 0.9419 and a validation accuracy of 58%.

However, upon evaluating the confusion matrix, it became clear that the model was heavily biassed towards predicting the "Normal" class for all cases, including the "Abnormal" and "Borderline" classes. This indicated a likely issue with class imbalance, as the predictions were predominantly concentrated in the "Normal" column, suggesting that the model had not effectively learned to distinguish between the different classes.



Further analysis using the classification report reinforced these findings. The model completely failed to classify any "Abnormal" or "Borderline" samples correctly, as indicated by the 0 scores across all metrics for these minority classes. This poor performance suggests that the model was unable to learn the distinguishing features of these classes. Additionally, the model demonstrated a strong bias towards predicting the "Normal" class, as evidenced by a perfect recall (1.00) for this class. This bias was likely caused by the imbalance in the dataset, with significantly more "Normal" samples compared to "Abnormal" and "Borderline" samples.

	precision	recall	f1-score	support
abnormal	0.00	0.00	0.00	231
borderline	0.00	0.00	0.00	513
normal	0.58	1.00	0.73	1030
accuracy			0.58	1774
macro avg	0.19	0.33	0.24	1774
weighted avg	0.34	0.58	0.43	1774

The evaluation of the model strongly suggested that it was suffering from both overfitting and class imbalance, despite the measures taken to prevent overfitting, such as adding dropout layers and freezing pre-trained weights. This surprising result led us to explore upsampling the minority classes in an attempt to mitigate the class bias and imbalance issue.

Categorical: Multi-label CVC Catheter Classification After Upsampling

Setup:

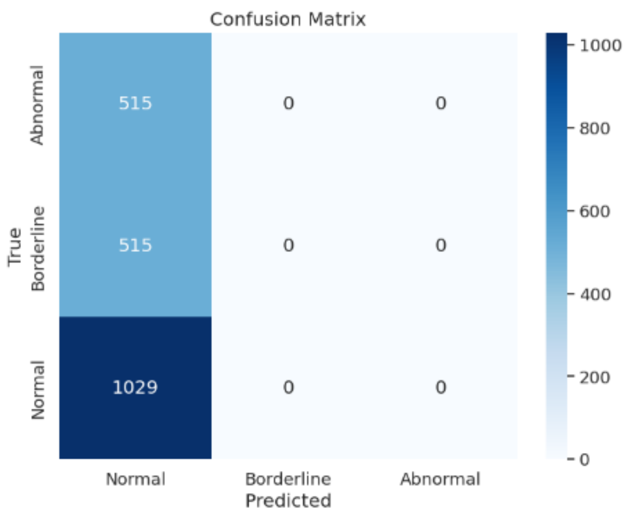
To address the class imbalance in our dataset, we upsampled the "CVC_Abnormal" and "CVC_Borderline" classes to create a balanced dataset. After upsampling, we worked with three CVC classes of X-ray images: "Normal," "Borderline," and "Abnormal," with 5147 images per class. We then used the same process for training and testing the model

```
cvc_class
normal        5147
borderline    5147
abnormal      5147
Name: count, dtype: int64
```

Modelling:

During training, the model achieved a validation loss of 0.636 and a validation accuracy of 67%.

After upsampling, we evaluated the model's performance using a confusion matrix. The results revealed significant challenges in distinguishing between the three classes. For the **Abnormal Class** (515 samples), all 515 samples were predicted as "Normal," meaning the model completely misclassified this class. Similarly, for the **Borderline Class** (515 samples), all 515 samples were also misclassified as "Normal." In contrast, for the **Normal Class** (1029 samples), all 1029 samples were correctly classified as "Normal." Unfortunately, this was the only class for which the model made accurate predictions.



When evaluating the model's performance through the classification report, we observed the following:

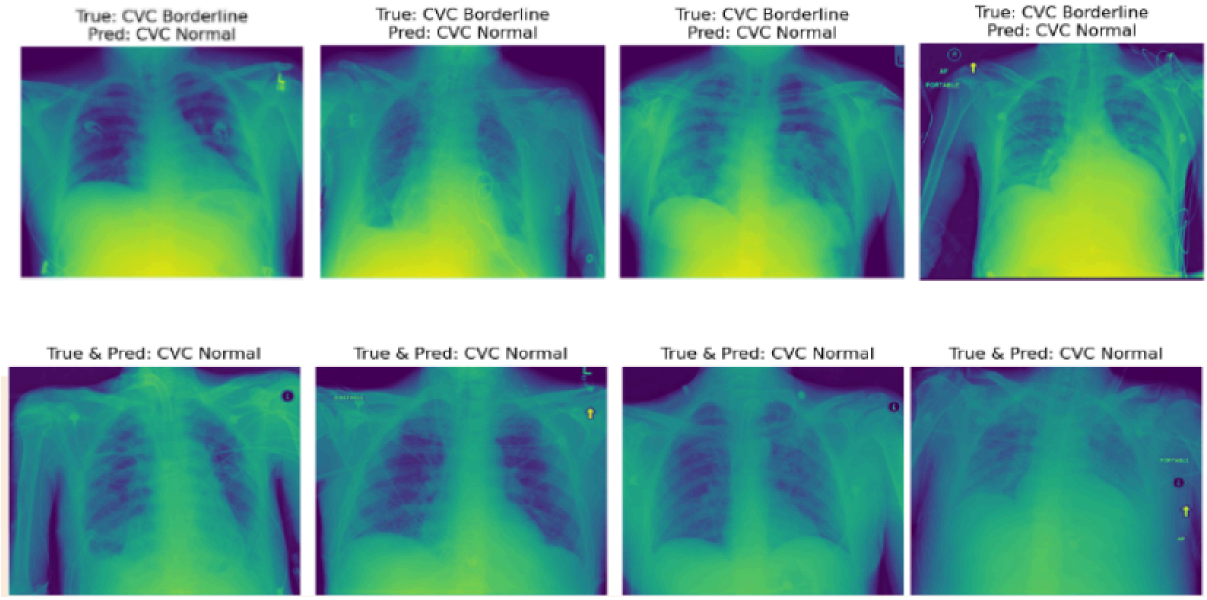
- **Normal Class**: The model achieved a precision of 0.25, meaning that 25% of the samples predicted as "Normal" were actually correct. The recall for this class was 1.00, indicating that all "Normal" samples were correctly predicted. The F1-score for the "Normal" class was 0.40, reflecting relatively strong performance compared to the other classes.
- **Borderline Class**: The model completely failed to classify any samples in the "Borderline" class, with a precision, recall, and F1-score of 0.00 across the board.
- **Abnormal Class**: Similarly, the model showed no ability to classify the "Abnormal" class. The precision, recall, and F1-score for this class were all 0.00, as all "Abnormal" samples were predicted as "Normal."

	precision	recall	f1-score	support
normal	0.25	1.00	0.40	515
borderline	0.00	0.00	0.00	515
abnormal	0.00	0.00	0.00	1029
accuracy			0.25	2059
macro avg	0.08	0.33	0.13	2059
weighted avg	0.06	0.25	0.10	2059

Overall, this model still fails to classify "Borderline" and "Abnormal" Classes, despite upsampling. All samples from these classes were misclassified as "Normal," indicating that the model has not learned the distinguishing features of these minority classes. This therefore portrays that there still is the continued bias towards predicting "Normal", suggesting that class imbalance remains a significant issue, even after upsampling. The model seems to struggle to learn the features of the minority classes ("Borderline" and "Abnormal"), resulting in poor performance for these categories.

Out of options, we plotted the images of misclassified x-rays and correctly classified ones to try understand why our model is predicting poorly:

True vs Predicted Catheter Classification results



Just from observation, the key difference between the images is the brightness of the x-rays at the organs. This high heat area indicates fatty organs or body fluids present in the patient's body during that time. This leads us to believe that the status of the patient, weight, fat distribution and fluid levels, cause the model to misclassify the catheter present in the patient. The layers our model uses are not sufficient enough to break down the x-ray images due to the high saturation of the x-ray.

Although adding more layers to the model might improve the models learning capabilities, we would have to reduce the data size due to limitations of software and time. Balancing model complexity and data size remains a challenge in this context.

Categorical: Multi-label Catheter Classification Setup

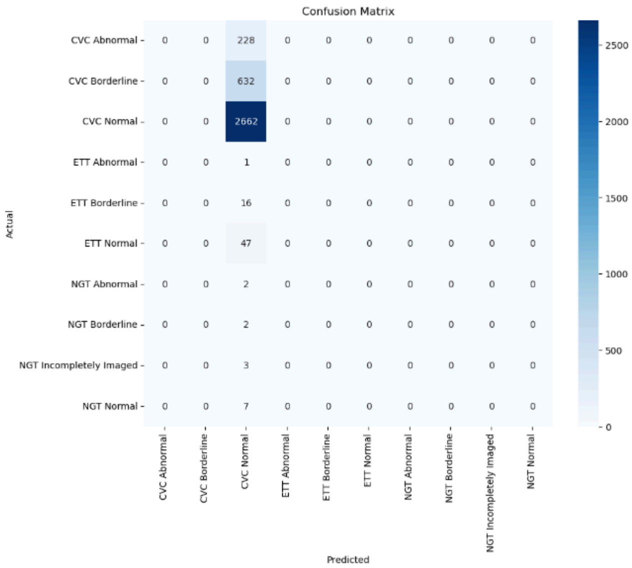
Setup:

We worked with 17,999 studies for catheter classification, focusing only on X-ray images where a single catheter was present. Due to software limitations, significant undersampling was applied to the dataset. To address the class imbalance, we also applied class weights during the modelling process. Additionally, we created categorical columns to classify the different catheter types.

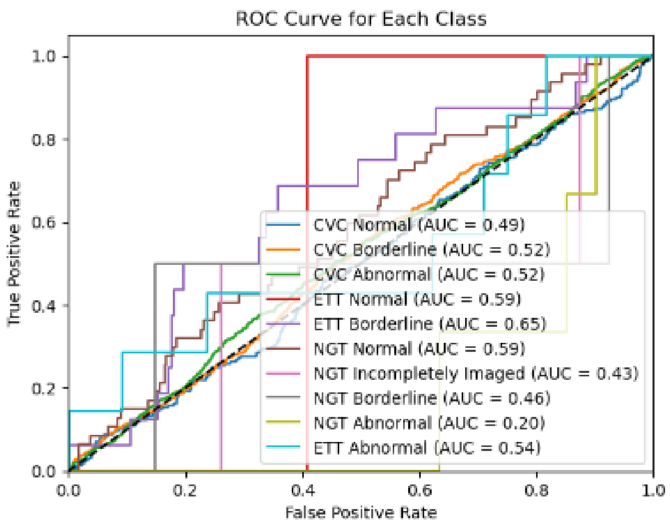
```
Primary_Catheter_Type
CVC Normal                 13312
CVC Borderline              3159
CVC Abnormal                1139
ETT Normal                   233
ETT Borderline                80
NGT Normal                    37
NGT Incompletely Imaged       17
NGT Borderline                 9
NGT Abnormal                   8
ETT Abnormal                   5
Name: count, dtype: int64
Found 14399 validated image filenames belonging to 10 classes.
Found 3600 validated image filenames belonging to 10 classes.
```

Modelling:

We used the same process as in the multi-label classification of CVC catheters. The model was trained for several hours, but the results were not as expected.



	precision	recall	f1-score	support
CVC Abnormal	0.00	0.00	0.00	228
CVC Borderline	0.00	0.00	0.00	632
CVC Normal	0.74	1.00	0.85	2662
ETT Abnormal	0.00	0.00	0.00	1
ETT Borderline	0.00	0.00	0.00	16
ETT Normal	0.00	0.00	0.00	47
NGT Abnormal	0.00	0.00	0.00	2
NGT Borderline	0.00	0.00	0.00	2
NGT Incompletely Imaged	0.00	0.00	0.00	3
NGT Normal	0.00	0.00	0.00	7
accuracy			0.74	3600
macro avg	0.07	0.10	0.09	3600
weighted avg	0.55	0.74	0.63	3600



Overall, the same issue arises in this problem. Class Imbalance: The model is heavily biassed towards predicting the "CVC Normal" class, which likely dominates the dataset. The model also has not learned any meaningful features for the minority classes, such as "CVC Abnormal," "CVC Borderline," or any of the ETT and NGT classes. This could be due to insufficient examples of these classes or the model's inability to differentiate between them.

This time we used a new metric to evaluate our model in the ROC plot. From observing the ROC plot, it shows that the AUC (Area Under the Curve) for CVC Abnormal and CVC Borderline are both quite low (around 0.52). The low AUC indicates that the model struggles to distinguish between these classes and others. The AUC for ETT Normal (0.59) and ETT Borderline (0.65) is higher compared to CVC categories, which suggests that the model has slightly better, though still weak, discriminatory power for these classes. The very low AUC for NGT Abnormal (0.20) indicates that the model has almost no ability to correctly classify this class.

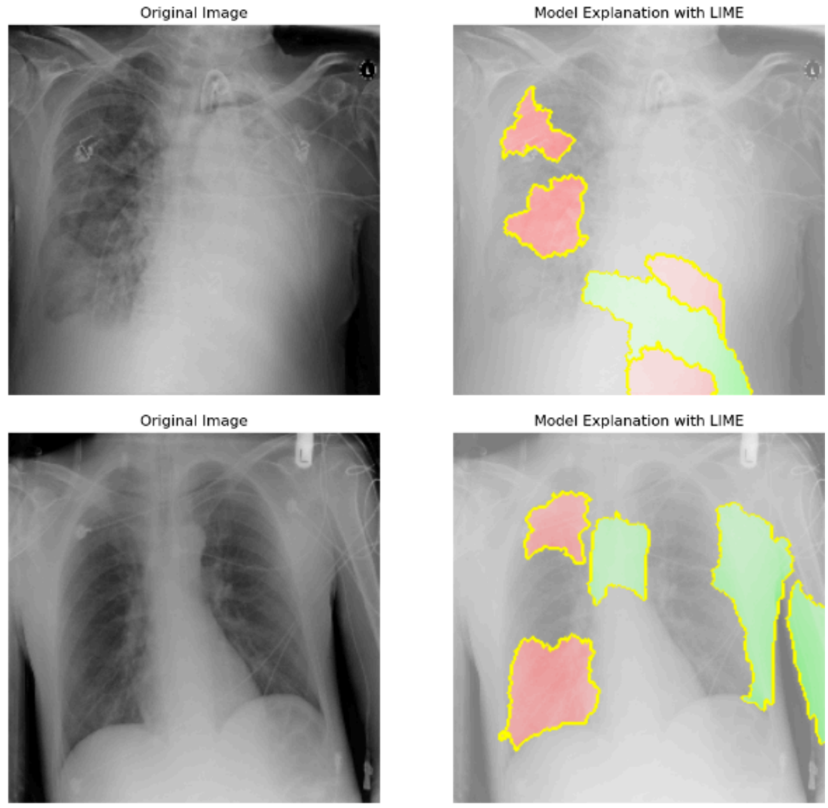
The CVC Normal AUC (0.49) is low despite this being the dominant class, suggesting that the model is biassed towards this class but not performing well overall. This aligns with the earlier confusion matrix, where the model was heavily biassed towards predicting "CVC Normal" even for other classes.

Overall, although the model achieves an accuracy of 74%, this is primarily driven by its success in predicting the majority class, "CVC Normal." The poor macro average from the classification report reveals that the model is performing poorly overall when considering all classes equally.

Given these results, it is likely that the model will perform similarly to the multi-label CVC type classifier when upsampled. Furthermore, the large number of data points and classes present significant challenges for the software, making it difficult to handle this complexity effectively.

Model explainability

Having arrived at the final model for our project, the next phase was to explain the model using the techniques learned about explainable AI (XAI). It is essential for models to be explainable to users. AI interpretability reveals what is happening within these systems and helps identify potential issues such as information leakage, model bias, robustness, and causality. The main approach of the group was to use a data science technique known as LIME (Local Interpretable Model-Agnostic Explanations). LIME is a technique designed to explain the predictions of complex, black-box machine learning models, which are often difficult to interpret directly. The key idea behind LIME is to create a simplified, local model around a specific prediction to offer insights into how the original model arrived at that particular result



.
LIME was used on the correctly identified images and the above output was generated. In the above image, the highlighted spots show the important regions of the image. The red areas represent parts of the image which had a negative impact on the model. These regions can cause issues such as imbalance and outliers. The green regions were the points which had positive and caused the model results to be better.

Conclusion

Our project explored various deep learning models to classify and identify different catheter types in chest X-rays, focusing primarily on binary and multi-label classification using ResNet50 as the backbone. While classification models offer significant advantages in simplicity and efficiency, their effectiveness is heavily influenced by data quality and balance. This was evident in our binary classification models for CVC, ETT, and NGT catheters, where initial high accuracy was misleading due to imbalanced datasets. We addressed this by employing resampling techniques and adding weights, which led to more realistic accuracies for each binary classification model.

For the multi-label classification task, we encountered persistent issues with class imbalance, which led to biased predictions towards the dominant "Normal" class. Despite attempts to mitigate this bias through upsampling and class weighting, the model struggled to correctly classify minority classes like "Abnormal" and "Borderline." The analysis of ROC plots and confusion matrices highlighted that the model failed to learn meaningful distinguishing features, particularly for minority classes, reflecting the challenge of handling imbalanced datasets effectively. In which if we wanted to move on this is the area we'd have to spend more time on

Overall, the results emphasise the importance of addressing class imbalance and refining model architectures when working with complex medical datasets. Although our models showed limited success in correctly identifying minority classes, they provided valuable insights into the limitations of standard deep learning architectures like ResNet50 in such applications. Future work should focus on enhancing model complexity, improving feature extraction, and exploring alternative architectures that are better suited for imbalanced, high-dimensional medical imaging data.

References

Ahmed, R. (Jul, 2023). *Endotracheal Tube.* National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK539747/

Azura Vascular Care. (Feb, 2018). *5 Common Central Venous Catheter (CVC) Complications.* Azura Vascular Care. https://www.azuravascularcare.com/infodialysisaccess/central-venous-catheter-complications/#:~:text=Central%20Venous%20Catheter%20Complication%20%231,can%20occur%20during%20CVC%20placement

Bansal S, Bansal S, Dogra M, Khan I. (n.d.). *A Malpositioned Central Venous Catheter.* Apicare online. https://www.apicareonline.com/index.php/APIC/article/view/312/1221#:~:text=Although%20usually%20without%20any%20serious,complications%20like%20erosion%20and%20thrombosis

Hosny, A. (Aug, 2018). *Artificial intelligence in radiology.* National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6268174/

Kolikof, J. (Jul, 2023). *Central Venous Catheter.* National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK557798/

Saugel, B. (Aug, 2017). *Ultrasound-guided central venous catheter placement: a structured review and recommendations for clinical practice.* National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC5572160/

Shim, J. (2024). 'How to identify a catheter'. Interview by Byron Shim, 14 September.

Viso. (n.d.). *EfficientNet: Scaling neural networks for improved accuracy and efficiency*. Viso. https://viso.ai/deep-learning/efficientnet/

Google AI. (2019, June 7). *EfficientNet: Improving accuracy and efficiency through AutoML and model scaling*. Google Research Blog. https://research.google/blog/efficientnet-improving-accuracy-and-efficiency-through-automl-and-model-scaling/

GeeksforGeeks. (n.d.). *DenseNet explained*. GeeksforGeeks. https://www.geeksforgeeks.org/densenet-explained/

Zhang, Y., & Zhu, X. (2021). A review of DenseNet for image classification and segmentation. *SN Computer Science, 2*(6), 1-10. https://doi.org/10.1007/s42979-021-00782-7

GeeksforGeeks. (n.d.). *Advantages and disadvantages of different classification models*. GeeksforGeeks. https://www.geeksforgeeks.org/advantages-and-disadvantages-of-different-classification-models/

Choudhury, A., Singh, S., & Kumar, R. (2022). A comprehensive review on machine learning applications for healthcare management. *Journal of Healthcare Informatics Research, 6*(3), 213-230. https://doi.org/10.1007/s41666-022-00108-4

Wang, J., Zhang, Y., & Qiao, H. (2022). A deep learning model for detecting nasogastric tube placement in chest radiographs. *Journal of Medical Imaging and Radiation Sciences, 53*(2), 101-109. https://doi.org/10.1016/j.jmir.2021.12.006

Databasecamp. (2021, December 16). *ResNet: An introduction to residual neural networks*. https://databasecamp.de/en/ml/resnet-en

Signon, F. (Oct, 2022). *Nasogastric Tube*. National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK557798/