Medical Imaging Classification: Catheter Placement

Thursday Studio - Group 2

Oct 2023

Authors:
Jiaqi Dong - 33051089
Noha Jung - 33154678
Oscar Brazzale - 33154376
Rory Thomas - 33154368
Yijia Xue - 33051178

Executive Summary

The study aimed to explore the feasibility of using neural network-based image classification to assess catheters, particularly in X-ray images. The primary research question focused on whether these machine learning models could reliably detect the presence and assess the quality of various catheters, including Endotrachial Tubes (ETT), Nasogastric Tubes (NGT), and Central Venous Catheters (CVC).

The research used extensive image datasets provided by the Royal Australian & NZ College of Radiologists, along with corresponding labels indicating catheter presence and quality. The study initially anticipated that applying neural networks to this task might yield varying results, given the complexity of the issue and the diverse quality of the images. The expectation was that CVC catheters would be relatively easier to detect due to their distinctive appearance. Data preprocessing and cleaning were essential to make the dataset suitable for neural network applications. This included addressing issues like the structure of coordinate data and the imbalanced distribution of "normal" catheter placements, which posed challenges for model training.

To tackle the complexity of the catheter classification problem, the research proposed a two-step model. The "finder" models were designed to identify catheter presence, while the "grader" models aimed to assess the quality of catheters. This separation of tasks allowed for more efficient training, mitigation of data imbalance, and manual tuning of model components. In terms of model implementation, convolutional neural networks (CNNs) were chosen due to their established effectiveness in image classification tasks. These CNNs were fine-tuned to fit the specific requirements of the catheter assessment model. Additionally, image resizing was employed to standardize input dimensions, a crucial step in preparing the images for the neural network.

Results from the study revealed varying model performance. The NGT finder models demonstrated high accuracy, especially when working with highlighted images, showcasing significant promise in catheter detection. In contrast, the NGT grader models faced challenges related to data imbalance and predominantly predicted "normal" classifications. On the other hand, ETT and CVC models, trained on image data, exhibited notably high accuracy in catheter detection.

In conclusion, the study determined that while it did not definitively answer the research question, it provided valuable insights into the potential of neural networks for catheter assessment. The research demonstrated promise, particularly when working with highlighted images and considering various aspects for future research.

The study highlighted several areas for potential future research. These include further investigation into image straightening techniques to address image quality issues, exploring

alternative neural network architectures, and conducting more comprehensive testing for ETT and CVC models.
Overall, the study suggested that using neural networks for catheter assessment is a promising avenue for research, with opportunities for refinement and improvement in subsequent studies.

Introduction

Overview

Defined by Oxford Languages as “a flexible tube inserted through a narrow opening into a body cavity for removing, [inserting or replacing] fluid”, catheters are a vital tool for doctors and can be seen everywhere throughout hospitals administering aid to patients in need (Oxford Languages, 2023). Although the process of placing a catheter is something that a doctor will do countless times with great success, anomalies do occur and catheters are malpositioned. As catheters interact with some of the most delicate parts of the body, abnormal placement can cause complications from infection all the way to “death” (RANZCR, 2020). As a results of these consequences it is vital that abnormal catheter placement can be picked up and dealt with immediately and without error. Although doctors can conduct this analysis, this is time consuming, prone to human error and realies heavily on highly skilled labour. If a machine was capable of undertaking catheter examination this would solve all these issues thereby reducing pressure on an already strained hospital system. As a result the main research question this report seeks to answer is formed:

“Can we use neural network based image classification to reliably assess catheters”

Context

Throughout this project the placement of three different kinds of chest catheters was analysed a diagram of which is provided in Figure 1. The first of these three catheters is the endotrachial tube (ETT). Inserted through the mouth and extending no further than the collar bone, ETT catheter is a fairly low-risk procedure that aids patient ventilation by “clearing secretions” and “maintain airway patency” (The Royal Childrens Hospital, 2020). The next type of catheter is the nasogastric tube, also known as, an NGT. Translating to “nose to stomach” the aptly named nasogastric tube is inserted through the nose where it continues all the way to the patients stomach. It often serves the purpose of providing medication on nutrition to patients when they can’t ingest it orally (Cleveland Clinic, 2022). The third and final catheter is a Central Venous Catheter (CVC). This catheter is “neccesary for all critically ill patients” and is inserted into a one of a patient major veins where it can perform tasks such as dialysis or administration of medication (Kolikof, Peterson, & Baker, 2023).

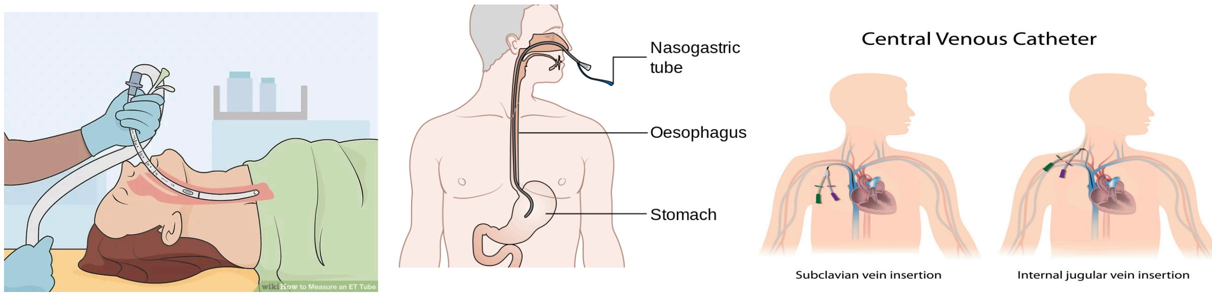
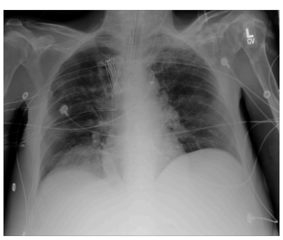


Figure 1: (From left to Right) ETT Catheter, NGT Catheter and CVC Catheter

Data Resources Available

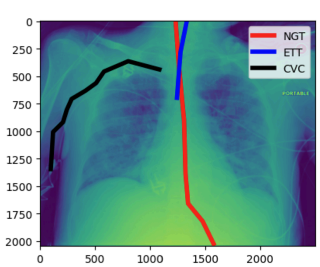
To complete the project a collection of datasets created by the Royal Australian & NZ College of Radiologists was provided. Among these data sets was a set of 30,000 images such as the one in Figure 2 that contained x-rays from patients that had between 1 and 6 catheters at a time. This image data was accompanied by the data set named “train.csv” which is also shown in Figure 2. This CSV file contained binary labels for each image stating whether each kind of catheter was in the image and whether it was normally placed, abnormally placed or borderline between the previous two options.



	StudyInstanceUID	ETT - Abnormal	ETT - Borderline	ETT - Normal	NGT - Abnormal
438	1.2.826.0.1.3680043.8.498.33289872132944517059...	0	0	1	0
512	1.2.826.0.1.3680043.8.498.58815494070381425972...	0	0	1	0
888	1.2.826.0.1.3680043.8.498.18423562395321282202...	0	0	1	0
1103	1.2.826.0.1.3680043.8.498.80010847260848855758...	0	0	1	0
1206	1.2.826.0.1.3680043.8.498.68620140014200110095...	0	0	0	0

Figure 2: Example of “train” image and “train.csv” labels

Alongside this main data set, a CSV file named “train_annotations.csv”. This additional file contained a set of coordinates corresponding to 18,000 of the original 30,000 images. Later investigation uncovered that these coordinates graphed the position of the catheter in the image as shown. The result of superimposing these coordinates over an image and the data set are shown in Figure 3.



	StudyInstanceUID	label	data
0	1.2.826.0.1.3680043.8.498.12616281126973421762...	CVC - Normal	[[1487, 1279], [1477, 1168], [1472, 1052], [14...
1	1.2.826.0.1.3680043.8.498.12616281126973421762...	CVC - Normal	[[1328, 7], [1347, 101], [1383, 193], [1400, 2...
2	1.2.826.0.1.3680043.8.498.72921907356394389969...	CVC - Borderline	[[801, 1207], [812, 1112], [823, 1023], [842, ...
3	1.2.826.0.1.3680043.8.498.11697104485452001927...	CVC - Normal	[[1366, 961], [1411, 861], [1453, 751], [1508,...
4	1.2.826.0.1.3680043.8.498.87704688663091069148...	NGT - Normal	[[1862, 14], [1845, 293], [1801, 869], [1716, ...

Figure 3: Example of Contents and Application of “train_annotations.csv”

A set of test images was also provided, however, as it did not have any labels to go with it, it was not used in the project. The presence of SwanGanz catheters and patient IDs were also recorded in “train.csv”, however, these variables weren’t used during the project

Expectations/Hypothesis

It was expected that although the the neural network would show some ability identify and distinguish catheters, it would struggle to do so reliably and only receive accuracy scores around 75%. The reason behind this expectation is that the neural network should naturally struggle on images that are very noisy and are poor quality. This is problematic as some of the images appeared quite noisy and the clarity in the x-rays varied drastically. Therefore, it is expected that the neural network will fail to classify these images thus not meeting the requirements to be titled as “reliable”.

Furthermore, it was also expected that the model would struggle differently from catheter to catheter. Although the CVC is the most complex procedure, they are quite long and therefore, they are easy to see and have large potential to have many tells that show it is malpositioned. For this reason it is believed neural network based image classification models will see most success with this catheter type.

Data Quality

The data provided was fairly complete and easy to work with, however, it was not without issues. From unexpected data types to unbalanced composition of datasets, significant work needed to be performed on the data to make it usable for neural network applications.

Data Cleaning

Many analytical techniques and modelling functions required for this project need data to be in a well-structured format that follows a set of logic conventions. The first area where the data provided didn’t meet this mark was in the “train_annotations.csv” file. Logically, a set of coordinates should be stored as a list of tuples thus allowing iteration to be used intuitively and easily. Unfortunately, the limitations of CSV files meant that this was not the case for the coordinates in “train_.annotations.csv” as shown in Figure 4 and consequentially this had to be fixed. These coordinates, formally stored in a string, were only used once and thus were separated externally into temporary lists of x coordinates and y coordinates immediately before their usage. In hindsight, these coordinates should have been reinserted into the data frame in the aforementioned ideal format for later use, however, time restrictions classed this as unnecessary work.

StudyInstanceUID	label	data
1.2.826.0.1.3680043.8.498.12616281126973421762...	CVC - Normal	[[1487, 1279], [1477, 1168], [1472, 1052], [14...
1.2.826.0.1.3680043.8.498.12616281126973421762...	CVC - Normal	[[1328, 7], [1347, 101], [1383, 193], [1400, 2...

```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   StudyInstanceUID  17999 non-null  object
 1   label             17999 non-null  object
 2   data              17999 non-null  object
dtypes: object(3)
```

Figure 4: Coordinates in the wrong data structure (object is a string)

From the start of the project, the general idea was to use TensorFlow to build an image-classifying neural network. As TensorFlow requires inputted ADTs to be of a specific and complex structure, CSV files weren't going to accommodate the organisation and portability demands of using TensorFlow in a group setting. As a result, the "pickle" file type was used extensively throughout the project. Picklel Files preserve complex data structures such as TensorFlow objects and NumPy arrays within data frames at the expense of greater file size as illustrated by Figure 5. Yet again, although an alternative such as a JSON file would have been a more efficient use of space, this would have been more complicated and time constraints forced the prioritisation of simple solutions.

Figure 5: Comparison of Pickle and CSV File Types

Other miscellaneous data cleaning was required, however, this was a result of work done rather than data received.

Missing and Incomplete Data

Missing or incomplete data may confuse neural networks or cause analysis to fail entirely and thus they have to be removed, imputed or dealt with in some other way. Although no NAs were present in the data sets many of the images depicted incompletely imaged catheters or catheters with no labels. Examples of these occurrences are shown in Figure 6. For simplification purposes, images that fell into either of these categories were dropped as well. Initially, images with multiple of the same type of catheter were also excluded due to complexity concerns, however, this decision was later revoked as complexity issues would be dealt with as a consequence of the training process.

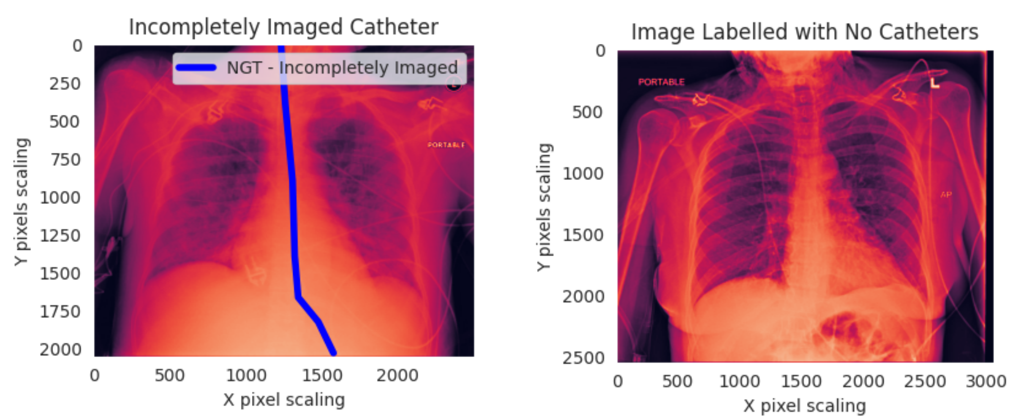


Figure 6: Examples of Missing or Incomplete Data

Balancing

In a classification problem, if one class of data significantly outnumbers the others, it may lead the model to be biased toward the majority class and exhibit poor performance on the minority class.

As shown in Figure 7, the data is highly imbalanced, with a much larger number of "normal" placements compared to "abnormal" and "borderline" placements for all three catheter types (ETT, NGT, and CVC). This can pose challenges when performing data analysis or modelling, as the imbalance may lead to biased results or reduced accuracy in estimating "abnormal" or "borderline" cases. Notably, the NGT category introduces an additional class, "incompletely imaged," which accounts for a substantial proportion of the NGT data (about 30%).

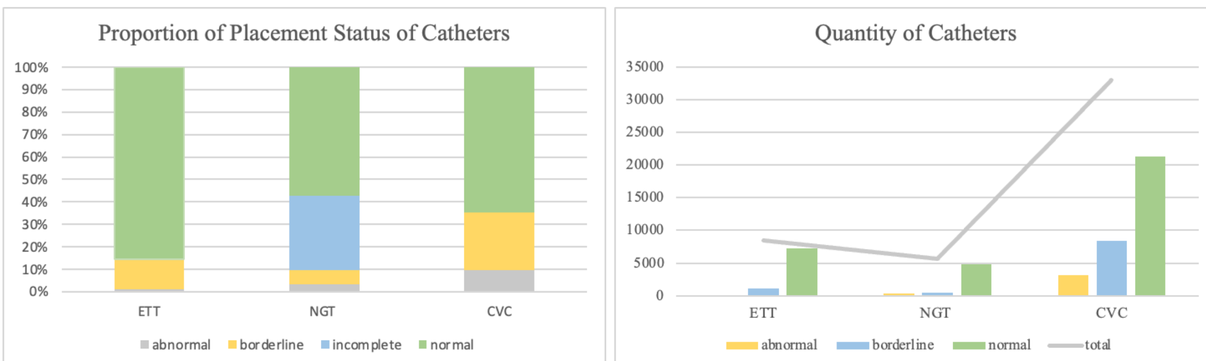


Figure 7: Analysis of Data Compostion by Type

Even though our data analysis and modelling would not consider this classification and remove it, it is evident that the number of correctly placed catheters still represents the most significant portion of the data for all three catheter types (see Figure 7). Moreover, the overall imbalance in the data may still pose challenges for certain analyses or modelling tasks. In a comprehensive comparison, the proportion of placement states in CVC is more balanced than in the other two types, and the sample size is larger, which can significantly enhance modelling accuracy. Both NGT and ETT data present unique challenges with smaller sample sizes and extremely unbalanced placement state proportions, increasing the likelihood of bias in data analysis and modelling.

Although data balancing had potential cause havoc in model training with the time and data resources available it was unfeasible to deal with this. Oversampling techniques such as SMOTE mischaracterised the real world distribution of catheters portrayed in the data while failing to accurately capture the complexity of the data at hand. Conversely, undersampling techniques restricted data sets thereby reducing the ability to train models. As a result of both data balancing management techniques failing to relieve this issues, the data was used as is. For now this section serves to provide a potential cause for later issues in model training.

Model Development

In its initial state, the data was comprised of many images that suffered from low quality, or lot of noise. As this obstructed the clarity of catheters in the images some preprocessing of the images was attempted for use in the model. The main techniques investigated were Gaussian smoothing and edge detection.

Gaussian Smooting

Gaussian smoothing is a technique which removes noise and grain from an image and essentially smooths it out increasing clarity in the process. This is advantageous for reliably classifying catheters in an x-ray as any unnecessary noise and artefacts are removed, leading to a clearer general image for the model to pick up on the most significant features. Figure 8 illustrates an example of the effects of this process where Gaussian smoothing is applied the image, it becomes clearer, and the noise is reduced. It can be seen that as Gaussian smoothing is applied the noise and artefacts are removed, however, the image also becomes blurrier. Due to this fact, there is a tradeoff between applying Gaussian smoothing to smooth out an image and not applying too much, as it would result in an unusable image where all clarity is lost to blur.

Figure 8: Examples of Gaussian Smoothing

Edge Detection

The other technique used was edge detection. Edge Detection detect the edges of an image and displays and image entirely comprised of these outline. This technique works by identifying discontinuities in the brightness of the image, which is then traced to produce an “edge only” image. This technique can potentially highlight features of interest, such as the catheters which were to be detected by the model. This advantage of highlighting important features has the potential to greatly improve the model's accuracy, as the said features would be easier to see. Figure 9 depicts an example of edge detection, where the edges of the image are more clearly outlined. Although this technique shows great promise Figure 9 also clearly illustrates the issues with this technique. Because there are also a lot of other parts in the background of the image, those get detected as well, causing a mess of edges and lines. Because of this issue, edge detection was not used for our image processing, as the improvement in clarity of the catheter wasn't big enough for us to see any noticeable differences. In the future, however, it would be worth investigating further, as with further processing it could potentially become a powerful tool in detecting catheters.

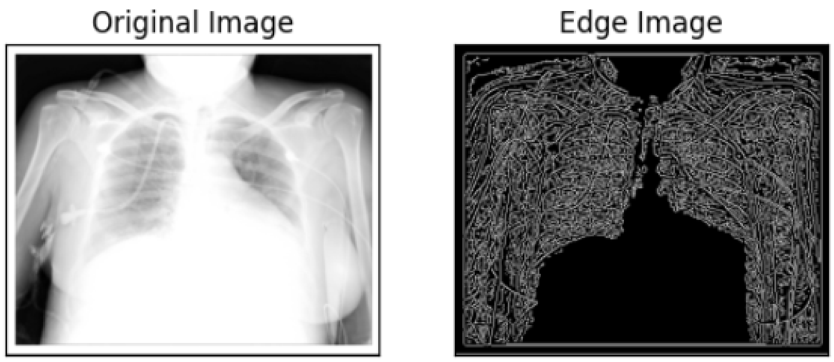


Figure 9: Examples of Edge Detection

Model Idea Outline

The initial plan was to create some kind of model that would take an X-ray image as an input and output a list of catheters present in the image accompanied by their quality. In theory, a model like this could be created by using the raw images available, creating a list of expected outputs for each image and training a multilabel image classification model on this data. This process is not very practical though as it will take a very long time, require a lot of data and the complexity of the problem will cause poor accuracy. These facts meant that an alternative method of achieving a model like this needed to be investigated.

The Components

As the aforementioned problem was in a complex state, the natural choice was to use divide and conquer to break a large multilabel classification problem into several binary subproblems. Logically the model can be broken down into asking two questions. First, “What catheters are in this image?” and secondly “are they any good?”. Each of these questions was assigned a class of model defined by the group as either a “finder model” or a “grader model”, the logic of which is visualised in Figure 10. For the sake of explanation, say that the only concern is NGT catheters. The “NGT Finder model” would take in an image and return a value expressing its confidence that an NGT catheter is in the image. Then the “NGT Grader model” takes images with NGTs in them as input and returns a value between 0 and 1 on how abnormal it thinks the catheter is.

Figure 10: Finder and Grader Models Logic

Putting it Together

These aforementioned model classes can then be fit together fairly intuitively with algorithmic help as the “finder” model is just dictating whether the “grader” model will run (i.e. if a user-chosen threshold is met by the finders output, the grader will run). This flow between the two models allows us to detect NGT catheters in images with their classifications. From here this process can be repeated for each type of catheter to form the web of models depicted in Figure 11. This web then achieves the group aim as the top stream detects NGT catheters in the image with their classifications, the middle stream detects ETT catheters in the image with their classifications and the bottom stream detects CVC catheters in the image with their classifications. Therefore, if the results are collated the results the aim of: “creating a model that takes in an image and returns a list of catheters in the image with their classifications” can be achieved.

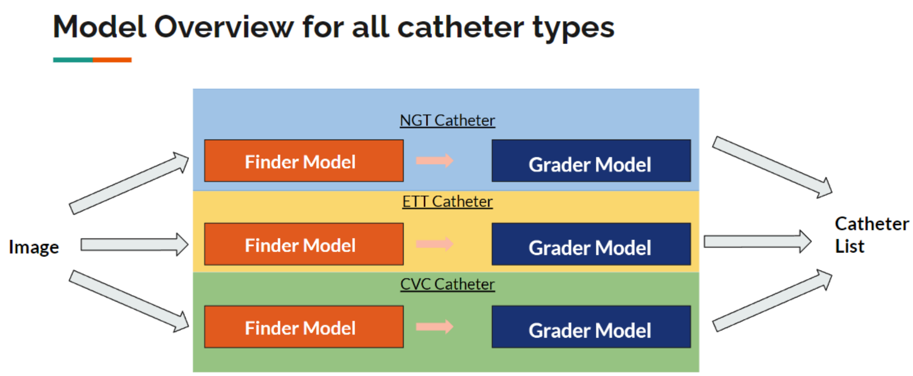


Figure 11: Final Web of Models for Tagging Images

The benefits of separating the classification task this way include:

- Reduced run time as simpler models require fewer parameters.
- Allowing each part of the problem to be trained on a different data set.
- Providing more intuitive insights as to where the model is failing in the event that it does fail. (i.e. “Is the model struggling to grade or find”, “Is the model struggling with ETTs”).
- Allows manual tuning of sections of the problem in the event that there is a weak link.

Model Training

The “web-structured” model allowed for each part of the classification problem to be trained separately. Thus it would be a shame if significant thought was not put into the training process.

Consequentially, each model had several training sets crafted for it that included modified versions of the original image data subsetted specifically for that model type.

Creating New Training Images

Although a model could be trained purely on the unedited images, it seemed naive not to use catheter coordinates in “train_annotations.csv” to aid the training process. Resultantly, these coordinates were incorporated into the training process by way of synthesizing new images that included them. Alongside the original set of images, two more sets were created. These images followed the form depicted in Figure 12. The first set of images (named “just highlights”) included just an outline of the catheter on a black background and then the second set (named “Highlights and Images”) included images where the outline of the catheter had been superimposed on its corresponding image.

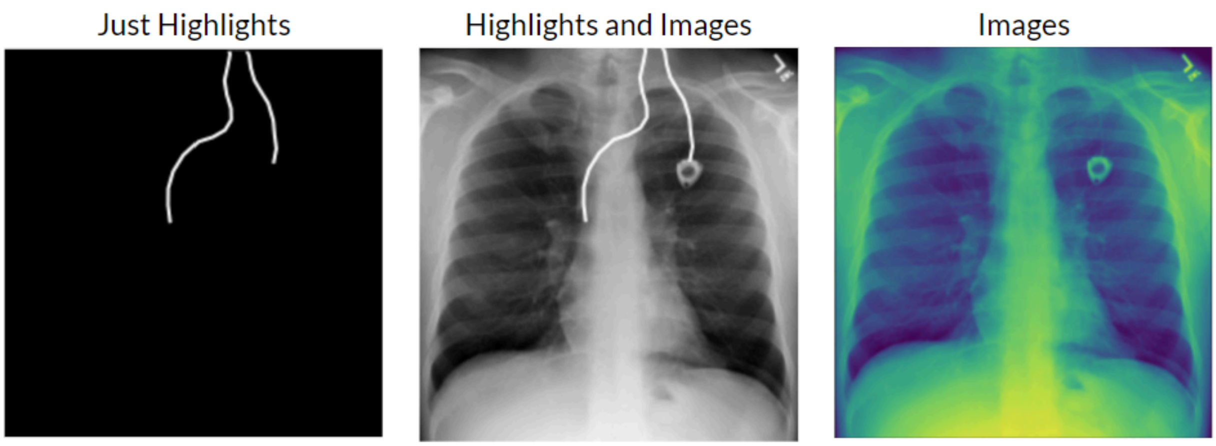


Figure 12: Examples of each kind of Training Image

The general idea behind creating these data sets is that a model could be first trained on the “just highlights” data set as this would give the model some idea of which pixels are important. It could do this as the catheter is in white and the background is black, therefore, making the only pixels with non-zero RGB values those that contained a catheter. This initial training would then provide the model with a set of initial weights. This added step is vital to the training process as the complexity of the images means that it is highly likely a model trained on unbalanced data, such as the data provided, would give importance to sections of the image that are irrelevant to the issue (an issue the groups colleagues observed).

A second stage of training would then be conducted on these initialised weights where the “Highlights and Images” dataset would train the model to deal with added noise. In hindsight, this may not have been as beneficial as it may have taught the model to ignore anything other than a white line. This being said, if a model was able to find success in this stage it would still be a practical finding as an unskilled worker could trace the catheter in the images prior to analysis and then use the model to great success. Furthermore, if the image is low enough quality that a human can’t find the catheter, it is unlikely that a machine could anyway.

After these two initial training stages, the weights would then be saved, and loaded to a new model of the same structure and then this new model would be trained on just the images. Success in this training stage would mean that the model could work entirely autonomously. Therefore, the need to train the model on just images is clear as it is acquainting the model with its real-life, intended use case.

Choosing Subsets

Each model was then only trained on an appropriate subset of the data. The main distinction in training the two model classes is that the finder models were trained on all images with many catheters in them while the grader models were only trained on images that included their given catheter type (i.e. the NGT grader was only trained on images with an NGT catheter in them). Furthermore, the initial training stages of grader models only included images where a single catheter was highlighted as depicted in Figure 13, with the main idea being to avoid confusion. The reasoning behind the only training grader models on images that included their designated catheter is clear as it dramatically simplifies the problem and is congruent with the holistic model structure. Simultaneously, the reasoning behind only training the grader models on images where a single catheter was highlighted is that the image is due to it only being designed to grade one catheter type. Therefore, if it is trained initially on images that include many catheters it may get confused. In hindsight, it may have been beneficial to train it on both the single catheter and multiple catheter training sets, however, this idea did not come to the group prior to model training.

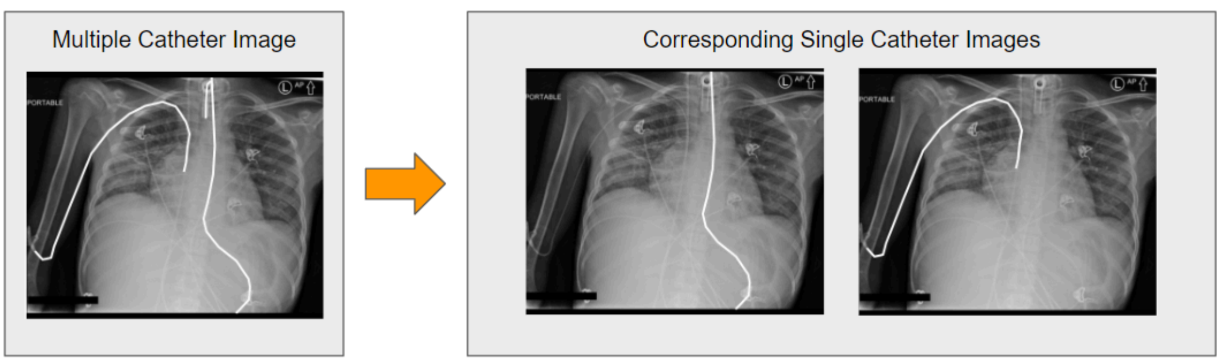


Figure 13: Examples of Single-Catheter and Multi-Catheter Images

Model Implementation

To answer the research question, “Can we use neural network-based image classification to reliably assess catheters” we needed to implement this model with some kind of neural network. With many neural network types available, an appropriate style of neural network needed to be chosen to best suit the problem. Furthermore, when implementing the neural network, inputs and outputs would have to be changed appropriately to fit the final model.

Neural Network Structure

As stated by Samir S. Yadav and Shivajirao M. Jadhav in their paper titled: “Deep convolutional neural network based Medical image Classification for disease diagnosis”, convolutional neural networks provide the “best results on varying image classification tasks” (Yadav & Jadhav, 2019). Therefore, it is logical to use some kind of convolutional neural network for this task. At a high level, a CNN such as the one depicted in Figure 14 is a neural network that uses a “feature learning section” prior to the application of a standard neural network. The “feature learning section” uses filters to recognise patterns in various segments of the image and then pool each segment's results together to generate a set of pattern similarity scores. As the layers of the network are traversed, this pattern recognition changes from recognising a specific kink or bend in a catheter’s outline to being able to recognise the type and classification of the entire catheter. The pattern similarity scores are then passed on to a neural network that processes the results to classify the image (IBM Technology, 2021).

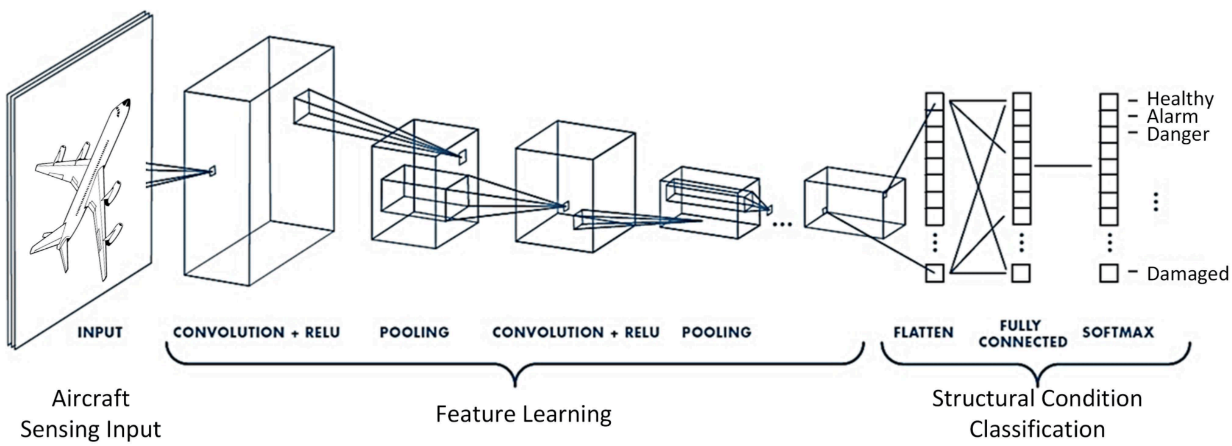


Figure 14: Diagram of General CNN Structure

There are many kinds of neural networks such as U-net or Resnet which handle the convolution process in special ways to specialise in biomedical imaging. This being said, the regular CNN was used as time restraints meant simple ways to provide a proof of concept and answer the research question were favourable.The general structure of this neural network was sourced from TensorFlow’s tutorial on creating CNNs with modifications made to fit the specifications of the model envisioned in this report’s previous section. The model structure used for both grader and finder models had 3 convolutional layers separated with 2 pooling layers for feature learning the section and a “flatten” layer to transition into a standard neural network section comprised of 2 dense layers and a dense output layer using a sigmoid function. A sigmoid output layer was used as the models were intended to return a confidence value that a certain binary condition (whether the catheter is there or whether the catheter is abnormal) was met and this is what a sigmoid function is intended to do.

Image Resizing

Neural networks require inputs to have the same dimensions as their input layer is a fixed size. This was an issue as the images provided took on values in a range of varying sizes and aspect ratios. The solution used was to use TensorFlow's inbuilt resize function to resize all images to 250px by 250px as shown in Figure 15. Although this would cause images to stretch differently based on the aspect ratio of the original image, the other option was to crop the image and potentially lose data which wasn't really an option. Another benefit of this reduced input size was that model training would be much quicker as a smaller input size would require fewer parameters.

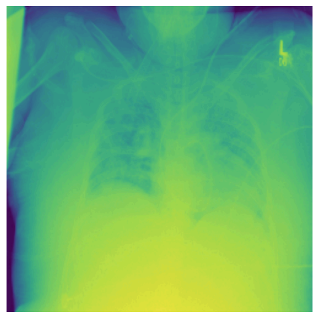


Figure 15: Resized Image

Results

The pressure of time constraints meant that rather than completing the entire model, varying aspects of the hypothesised holistic model were created, trained and tested as a means of finding where the model may struggle. This would inturn tell us if a neural network based image classification could be used to reliably classify catheters in x-ray images as it would explore all aspects of this process. Throughout the model testing phase a NGT finder and grader where taken through the entire 3 stage training process to provide a means to evaluate the training process. Simultaneously, ETT and CVC Finders were trained on just image data to identify which catheters the CNN methodology struggle to work with.

Scores Discussion

NGT Models

The first models created were the finder models for each of the 3 different types of catheters that were explored the first of which being the NGT finder. As mentioned, the NGT catheter models were taken through the intended intended training methodology saw training on all 3 different sets of images. The first set had the catheter highlights on their own, the second set had the catheter highlights over the X-ray images, and the final set had just the X-ray images. As the model progressed through the training stages it saw an initial accuracy of 0.9856 which then dropped dropped to 0.9851 before dropping to a final accuracy of 0.828 when classifying images on their own. These accuracies correspond to the confusion matrices shown in Figure 16.

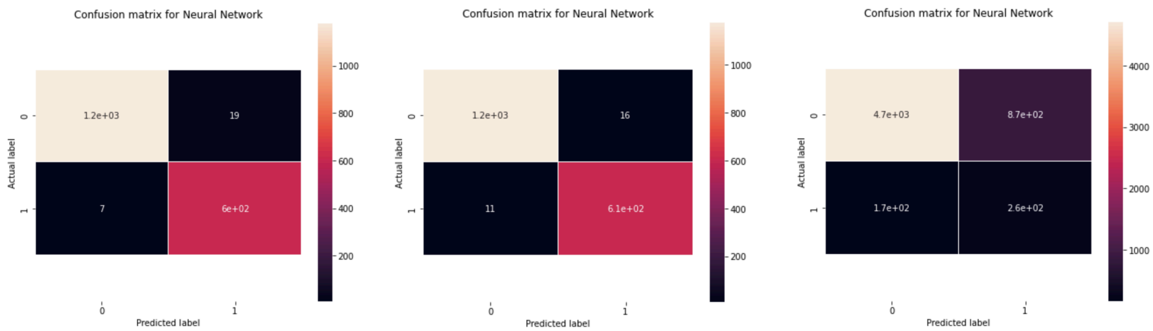


Figure 16: Confusion Matrices for NGT Finder Training. (Left to right) Just Highlight training, Highlight and Image Training, Just Image Training. 1 = present, 0 = not present

The high number of entries along the diagonal of the first two confusion matrices tell us that we saw great success in finding NGT catheters when the catheter is highlighted as most of the predictions were correct. Contrastingly, when the model was tested on just images it did not see such success, with a large portion of its accuracy being due to data imbalance confirming its tendency to label images as not having an NGT present. Although it may seem as though the image struggles with just images, if we investigate the model further this doesn't appear to be the case. Figure 17 depicts the spread of confidence values output by the model separated by the true label of the image being classified. This graph indicates although the model struggled to classify the images, it was still making some distinction between images containing NGTs and images that don't contain them. This fact is shown by the distribution of confidence values (the models confidence that an NGT is in the image) stretchs higher for images that contained NGTs that images that did not contain them.

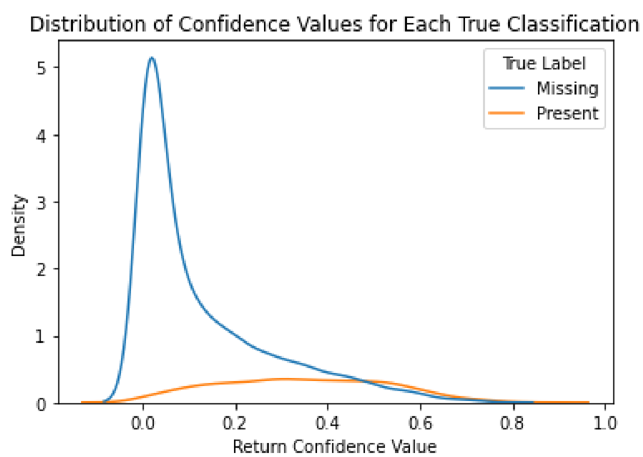


Figure 17: Distribution of model results for each image classification ("What confidence did the model tend to output for images with NGT's and images without?")

The next step was training the grader models for the NGT stream of the final model. As the 3 options for a catheters classification formed a spectrum from well placed to malpositioned, normal NGTs were labeled 0, borderline NGTs were labeled 0.5 and abnormal NGTs were labeled 1. This way the output of the model would essentially be a value from between 0 and 1 that gave the level of abnormality in the catheter. This, however, is the stage of the modelling process where data balancing became an issue. The main issue across all the catheter types is the large number of normally placed catheters which led to an inherent bias towards the prediction being normal. As the model was further trained, the accuracy decreased and the number of normal predictions increased. This is shown in Figure 18 below where apparent accuracy increases across the training process but it is evident that this is the result of unbalanced data as the model only predicted normal classifications. If this model were to be used, the first training step would be used as the final model. This had an accuracy of 0.66 and the real positive is that only 10% of the time did the model predict that an image had a worse reading than what was actually present. Although an aim of this process should be to reduce False negatives due to the medical domain, such poor results meant that this was an aspect of model tuning that was to advanced to consider.

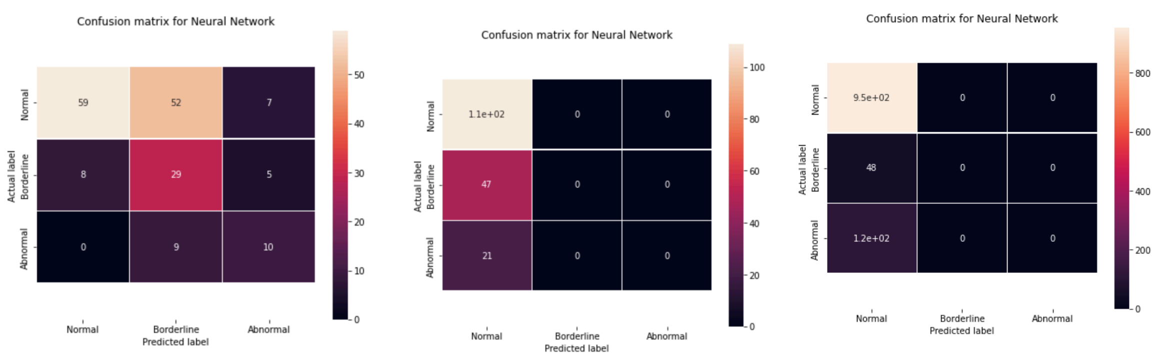


Figure 18: Confusion Matrices for 1st, 2nd and 3rd training of NGT Grader model (left to right)

Figure 18 does show that although unsuccessfully, the initial training stage saw the model make an attempt to classify images. Further investigation generated Figure 19 which, using the same format as Figure 17, shows that some distinction is being made during the original training stages of the model.

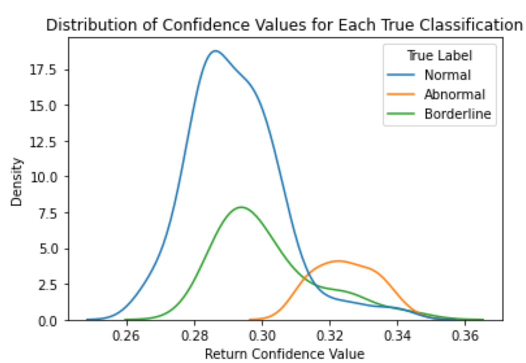


Figure 19: Distribution of Model Output Seperated by True Label

Exploring Performance of ETT and CVC detection

The ETT classifier and the CVC finder were then trained on just image data as a means of testing whether neural networks struggled to detect a particular kind of catheter. This process yielded incredibly high accuracy of 0.97 and 0.93 respectively. The confusion matrices for the Binary Models are shown below in Figure 20. These binary models performed extremely well and showed promise in the modelling process that was outlined above.
‘

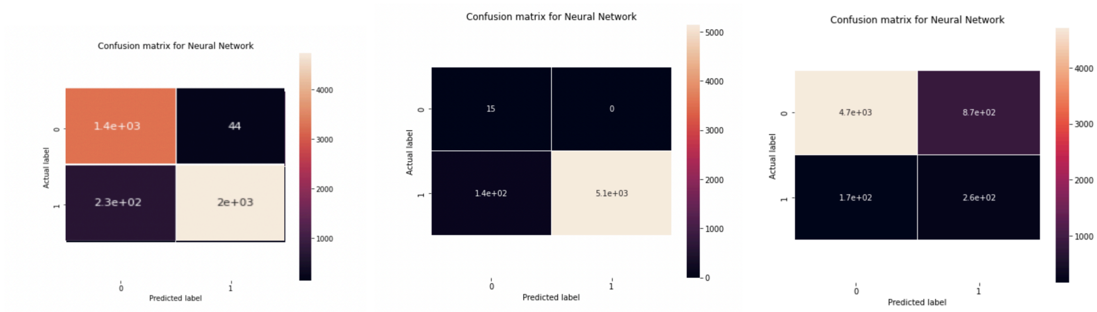


Figure 20: Confusion Matrices for ETT, CVC and NGT models respectively

Overall the best two-step model built accuracies of 0.99 followed by an accuracy of 0.66. Due to it being a two-step model, multiplying the accuracies together gives the model an accuracy of 0.65, albeit this was from a smaller sample size.

Using the less ideal models can lead to an accuracy as low as 0.54 but this is also with every prediction in the grading model being normal which can lead to high recall and does not help to solve the problem in a real sense of identifying catheters that are misplaced to see whether they need to be fixed. Minimising false negatives is important in ensuring a patient is not sent away with a reading that says they are normal when really they have a misplaced catheter that can cause issues regarding their health.

Conclusions

The main goal of this project was to assess if neural networks can be used to reliably find and grade catheters from X-ray images.

The above process and results show that it may be possible to use neural networks to reliably assess catheters.

Although the results found weren’t extremely high, strong scores in the NGT finder models early stages, and the ability of the NGT grader model to make distinctions between catheter classifications show promise. Although the report didn’t provide strong evidence that reliable classification was possible it indicated that working towards this goal is not a dead end and that it may be possible.

Further Research Questions and Areas of Future Study

Altough the result of this study didn’t directly answer the research question, they did provide an indication of ways this could be achieved given future study. Aspects to look at in future study are listed below:

Greater Success when Working with Highlighted Images

As shown in the results section, highlighted images performed phenomenally for finder models and showed great promise for grader models. This indicates that great success might be found from creating a model that can highlight catheters accurately and then using this to pre-process images. Additionally, if a hospital was to implement these models they could use unskilled workers to highlight catheters for model use, thus taking the pressure of doctors to classify the catheters themselves.

Image Quality

There are many reasons for the lack of accuracy in some of the models. One of the main reasons is the lack of consistency in the angle and size of the X-Ray imaging. This has a major effect when using an array of data pulled from an image. When the difference in positioning from the 3 different grades of the catheter is so small, having minor alterations on the angle and sizing of someone's chest in each image makes it very hard to build an overly accurate model using the X-ray images. Future research should involve investigating image straightening techniques to solve this issue.

Data Balance

The next issue with the model was the lack of variety in the levels of catheters imaged. There were over 80% of catheters were labelled normal which makes building a model that correctly predicts all 3 classifications difficult. The more trained the model became, it started to predict normal which gave it the highest accuracy but was useless when applying the model to a real-world scenario. Even attempts at data balancing led to undertrained models which had the opposite problem. Overall the data provided from the images made it difficult to build a high accuracy, high recall, two-step model.

Alternate Model Architectures

As discussed in the “model development section”, alternative neural network architectures were not investigated. Looking further into these may present promising results as CNNs such as Unet are designed for biomedical imagery.

Further Research Questions and Areas of Future Study

This project presented plenty of challenges and whilst some were able to be worked through, there were still lots of further questions and research opportunities that arose from the modelling. Firstly, due to time constraints and technical difficulties, the full CVC and ETT models were not able to be run and this is something that can be further investigated to see if improved results are yielded.

References:

Ahmed, R. A., & Boyer, T. J. (2023, July 24). Endotracheal tube. https://www.ncbi.nlm.nih.gov/books/NBK539747/

Bediwy, A. S., & Amer, H. G. (2012, January 31). Pigtail catheter use for draining pleural effusions of various etiologies. https://www.hindawi.com/journals/isrn/2012/143295/

Brownlee, J. (2021, March 16). Smote for imbalanced classification with python. https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/

Elhamraoui, Z. (2020, July 11). Introduction to convolutional neural network. https://medium.com/analytics-vidhya/introduction-to-convolutional-neural-network-6942c189a723

Gehrke, S. (2023, March 22). How to measure an ET tube: 14 steps. https://www.wikihow.com/Measure-an-ET-Tube

IVC. (2023). Treatment / central venous catheter placement. https://ivcnorthwest.com/treatment/central-venous-catheter-placement/

Kaggle. (2020). RANZCR clip - catheter and Line Position Challenge. https://www.kaggle.com/competitions/ranzcr-clip-catheter-line-classification/overview

Kendrick, A. (2020). Endotracheal tube suction of ventilated neonates. https://www.rch.org.au/rchcpg/hospital_clinical_guideline_index/endotracheal_tube_suction_of_ventilated_neonates/

Kolikof, J., Peterson, K., & Baker, A. M. (2023, July 26). Central venous catheter . https://www.ncbi.nlm.nih.gov/books/NBK557798/

Oxford Languages. (2023). catheter . https://www.google.com/search?sca_esv=575564128&rlz=1C1CHBF_en-GBAU727AU727&sxsrf=AM9HkKnMrJdcpWG2zITYjBRQVGOmNvDWxg%3A1697962995031&q=catheter&si=ALGXSlaYxylIm14_NEvUA9w95SVcXDHroIECKDc31uY0pP-rM6zH8mUDL8VKBRg0kK6prKsUp68y0X4MwnvIEINXGBrlUl-txN_k0NDIb_y6OpuAgY0pVzs%3D&expnd=1&sa

=X&ved=2ahUKEwjbzJ6JnYmCAxXaMHAKHRGvBQ8Q2v4IegQIFhAT&biw=1920&bih=931&dpr=1

Oxford Medical Education. (2016, April 18). Nasogastric (NG) tube placement. https://oxfordmedicaleducation.com/clinical-skills/procedures/nasogastric-ng-tube/

Verma, A. (2022, December 22). Indwelling pleural catheter. https://iplungclinic.com/service/indwelling-pleural-catheter/

What are Convolutional Neural Networks (CNNs)? (2021). Retrieved October 22, 2023, from https://www.youtube.com/watch?v=QzY57FaENXg.

Yadav, S. S., & Jadhav, S. M. (2019, December 17). Deep convolutional neural network based medical image classification for disease diagnosis - journal of big data. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0276-2