

CAB420 Group Assignment

COVID-19 Chest X-ray Image Classifier

Group 29: Koralalage Wanigasekera– N11226757, Celine Blumer – N100012711,
Bailey Nugent N10003703, Wei Jian Ho – N109161712

Introduction

Acting as the causal agent behind one of the most impactful pandemic events experienced in the last decade, COVID-19 is the infectious disease whose reach has spread globally. It is an acute respiratory disease that can be spread even by asymptomatic patients and has recorded an unprecedented rate of spread between individuals through infected droplets entering the body. The most effective method to control the virus from spreading as suggested by the World Health Organization (WHO) was early detection and quarantine. For this method to be fully effective in the society, there was a demand for accurate and efficient techniques for screening COVID-19 infected patients, to both provide the necessary treatment as well as for isolation from the public to minimize the spreading of the virus.

The most widely used technique to detect the COVID-19 virus is the RT-PCR test (reverse transcript polymerized chain reaction test) which measures the antibodies produced in reaction to the virus in discussion. While the main issue faced by many when utilizing the PCR kits was the limited availability of test kits, the other issue was timing constraints; it takes anywhere from a few hours to three days to receive the results of a PCR test. This is a time-consuming task which often resulted in infected patients spreading the virus in the public unknowingly. The other technique for detecting the virus is the Rapid Antigen test (RAT). While being able to produce the results a lot faster than a PCR test, RATs proved to be less accurate most of the time. It is due to these reasons that a more accurate, reliable, and faster COVID-19 screening method is required.

It is a known fact that various imaging techniques have made a significant difference in many fields. A few such fields are agriculture, medical, remote sensing etc. for many years these imaging techniques were used in the field of medicine for the classification and detection of different diseases such as skin diseases, types of ulcers and cancer. Soon after the spread of the COVID-19 virus began, it was revealed that patients suffering from COVID-19 tend to show abnormalities that are unique in COVID patients on chest x-ray images as well. This opened a new door allowing doctors to analyze CXR (chest x-ray) imaging of patients to detect COVID-19 rather than waiting 3 days for a PCR test result. The only concern when analyzing CXR's of patients is that it extremely requires domain specific knowledge of the medical staff. Given that the clinical environments were already severely burdened by COVID-19, analyzing individual CXR's added more weight to the situation.

This is where machine learning and deep learning comes into play. Being able to use a machine learning and /or deep learning model to classify chest x-ray images and accurately detect the presence of COVID-19 virus could help relieve these pressing issues in the following ways:

- i. Since imaging equipment is more widely available and accessible it resolves the issue of limited availability of PCR tests.
- ii. It resolves the time consumption issue caused by PCR testing
- iii. If able to classify with high accuracy, it resolves the issue of RATs providing less accurate results.
- iv. It assists with triaging patients and relieves the medical staff and clinical environments of some of the stress caused by time consuming PCR tests and the scarcity of domain specific resources needed to manually analyze a CXR.

Related Work

Given the extreme global impact of COVID, much effort and interest has been seen in literature proposing solutions to detection and classification of the virus using machine learning. All with a wide variety of approaches in methodology and architecture selection.

Due to the rapid emergence and novelty of the virus many papers cite issues sourcing large samples of COVID positive images. This produced imbalanced datasets when integrated with established imaging datasets such as pneumonia and healthy patients. However, as the pandemic has progressed and more cases diagnosed, a greater number of samples have been incrementally added to publicly available datasets. As such related works may have been constructed using smaller & different variations of dataset. This does make direct comparisons difficult however should still allow reasonable evaluations to be drawn.

Convolutional Neural Networks (CNNs) was by far the most prevalent method existing in literature. Primarily due to their automatic feature extraction and ease of use when using pretrained networks.

One of the first attempts at applying machine learning to the diagnosis of COVID occurred very early on in the pandemic in early 2020 by Chinese researchers. (Xu et al., 2020) Collated a small sample size of 219 x-ray images of positive cases and applied a from scratch CNN to achieve 86.7% detection rate. Whilst this initial result proved promising, the small sample size hinders validity. To overcome this small dataset, (Gupta et al., 2022) proposed the use of capsule networks, increasing classification performance to 95.7%.

Other methods to overcome class imbalance include (Li et al., 2021) who proposed a stacked convolutional autoencoder which reduced overfitting when compared to CNNs by providing greater control in layer by layer feature extraction.

(Jia et al., 2021), investigated performance of industry standard CNN architectures such as VGG16, ResNet, DenseNet & MobileNet. They found little performance difference between architectures but demonstrated DenseNet121 slightly edging ahead with an average accuracy of 98.8% primarily due to slight overfitting from other networks.

(Abbas et al., 2021) includes a ‘class decomposition layer’ in their CNN which finds common features in a per class basis and splits them further into sub classes. Whilst they claim the model is more robust to data irregularity, little benefit can be noted in performance when compared to standard CNNs.

Likewise (Keidar et al., 2021) explored standard CNN architectures but generated segmentation masks of the lungs using a separate pretrained UNet in pre-processing. This resulted in a 4% accuracy improvement over baseline CNNs. Masking allows networks to only pay attention to regions of interest hence increasing learning ability. Interestingly, (Yoo et al., 2020) also utilises a ResNet but integrates a decision tree mechanism. This essentially splits each task into a binary classification. Whilst this method did see improvement in classifying normal/abnormal the approach struggled to differentiate between COVID and non-COVID cases.

(A. et al., 2021) First applies a Contrast-Limited Adaptive Histogram Equalization and Butterworth filters to increase contrast and reduce noise respectively. These processed images are then passed to a CNN for classification. This method boasts the highest accuracy of 99.3% and highlights the apparent benefits in pre-processing suggesting that contrast may be a key factor in diagnosis.

Whilst more limited compared to deep learning, exploration of traditional machine learning algorithms has also proved promising. (Nazish et al., 2021) used HoG feature extraction followed by a Support Vector Machine and Logistic regression achieving 96% average accuracy in the binary classification of COVID positive and COVID negative subjects. This is comparable to results from deep learning methods. Similarly, (Ismael & Şengür, 2021) found SVMs to provide satisfactory results however used ResNet as a feature extractor.

More recent image classification approaches such as Visual Transformers have also been applied. (Shome et al., 2021) compared a standard ViT to baseline CNNs and reported ViTs beating out all comparable evaluated CNNs such as Resnet & Mobilenet.

All methods present with high accuracy, suggesting this is a task in which machine learning has a high ability to learn separation between classes and shows potential for real world use in diagnostics.

This paper proposes four classification methods two using traditional machine learning and two deep learning:

- Histogram of Gradients (HOG) Feature Extraction + SVM
- LDA + Random Forest Classifier
- Convolutional Neural Network
- Convolutional Visual Transformer

Dataset

The data used is the COVID-QU-Ex dataset publicly available on Kaggle and was compiled by researchers from the University of Qatar. The dataset consists of 14,853 chest X-ray (CXR) images from a range of age and genders. The classes and number of samples are:

- 3316 COVID-19 infections,
- 1345 Non-COVID infections (Viral or Bacterial Pneumonia)
- 10,192 Normal (Healthy).

A collection of sample images from each class are presented in figure 1:

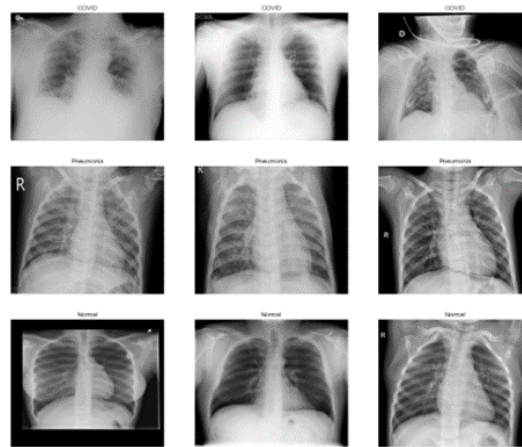


Figure 1: Chest medical images from dataset of COVID, Pneumonia & Normal

All cases have been independently labeled by medical professionals with expert domain knowledge. Therefore, the ground truth is believed to be accurate however could not be verified.

Preprocessing

Samples were divided using a 70,10,20% split ratio to construct the train, validation and testing datasets respectively. The class distribution is quite heavily unbalanced with significantly more samples for the 'Normal' class and only 22.3% and 6.4% of samples belonging to the 'COVID' & 'Pneumonia' classes respectively. Figure 2 below shows the number of samples per class for each data subset :

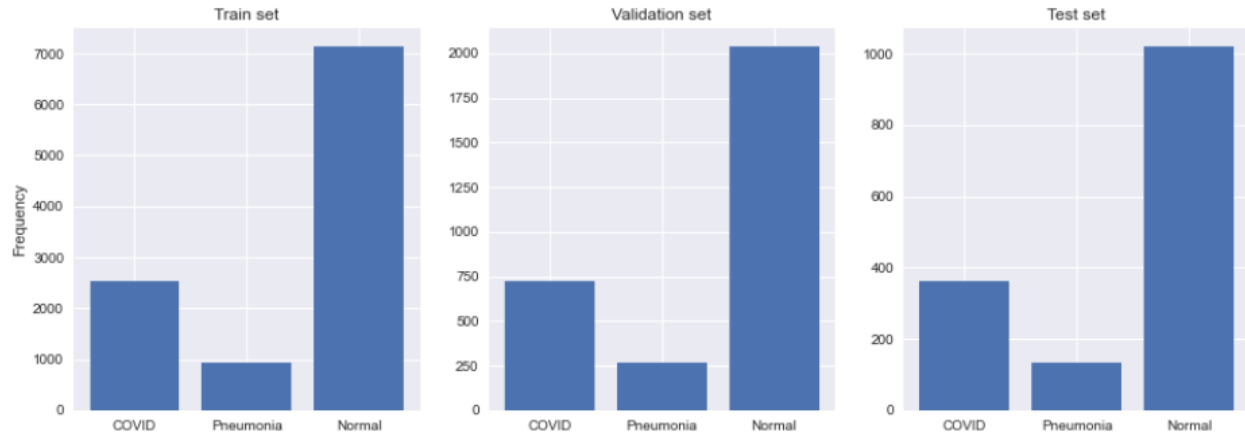


Figure 2: Histogram of class distribution for all data subsets

From this imbalance it is likely models will be biased towards the Normal class but this will be evaluated during development.

Whilst over and undersampling of classes to combat class imbalance will not be pursued, image augmentation will be used where possible to increase the number of samples. Class weighting will also be explored.

Images contain 3 channels and have a native size of 300x300. Within the loading procedure all images are reduced to single channel (greyscale) as they are effectively natively greyscale due to the problem domain. These extra channels are seen to contain redundant information and as such were discarded to reduce computational complexity.

Pixel value normalization was the only base/common preprocessing applied with each proposed method containing separate further feature extraction and preprocessing specific to the architecture.

Labels are strings by default so are numerically encoded categorically using one-hot representation.

Performance of each method is evaluated using standard classification metrics such as accuracy, precision, recall & F1 Score. These metrics are also most suitable for imbalance datasets such as this. Confusion matrices will be utilized to easily demonstrate per class performance. As with most medical based tasks, Recall should be prioritized as positive cases need to be correctly identified rather than attempting to minimize false positives.

Methodology

1 Support Vector Machine With Histogram Oriented Gradients

To build the image classification model to fit the COVID-QU-Ex Radiography Dataset, a total of 4 approaches were carried out: two non-deep and deep learning classification methods.

For the non-deep methods, two ensemble learning procedures were used, the first being a set of support vector machine multi-class classifiers with the Histograms of Oriented Gradients (HOG) feature descriptor applied to the input image data.

Firstly, support vector machines (abbreviated to SVMs) are supervised learning models that achieve classification by generating “linear decision boundaries” to correctly classify images by class, determined by their data point’s distance from a given hyperplane. SVMs are one of the most robust, and widely used methods for image classification and pattern recognition for the several benefits it offers data scientists such as its ability “to generalize well ... with limited training samples”, its use of regularization algorithms to mitigate overfitting and flexibility to allow for both binary and multi-class applications (Ogole, Im, & Mountrakis, 2011). Another useful feature of SVM is how it accommodates for fine tuning hyperparameters such as kernel and C choice, which can significantly improve performance with any given set of images. Due to these advantageous properties, however, the limitations that exist due to the SVMs effectiveness with few samples is its susceptibility to noise (common in large datasets), “large memory requirement” (due to its algorithms computing quadratically) and predisposition towards lengthier run times (Jian-Pei Zhang, 2005).

Onto the topics of pre-processing, the histogram of oriented gradients (HOG) is a popular feature extraction technique for image data from counting “the occurrence of gradient orientation” (Alhindi, Kalra, Ng, Afrin, & Tizhoosh, 2018). The benefit of this process is the ability to capture important structural parts of an image, reduce dimensionality to decrease noise and assist in making “the classification process less prone to overfitting” (O. Déniz, 2011). The main disadvantage for HOG may be that with the reduction of an image to meaningful ‘patches’ of the image, it may not be the most applicable to a highly detailed specific image classification problem. Also, as HOG converts images using a sliding window procedure so may cause larger computational time to complete.

To discuss the appropriateness of the SVM method and HOG feature extraction to the COVID-QU-Ex Radiography Dataset, while SVMs are not usually best for large datasets (the number of samples in the entire training dataset being 21715 images), due to the abundance of similar image classification studies supporting the excellent empirical effectiveness of SVMs alongside the suitability SVM’s strong capability with classifying between multiple classes from an image, support vector machines were a clear choice for one approach used in this report. It is hypothesized using HOG as a pre-processing technique alongside the SVM will be beneficial to the performance of the classifier considering the uniform, similar looking image nature of the data. It would achieve this by simplifying the most meaningful differences between images, so was selected as the feature descriptor of choice.

To implement these techniques into a method for this dataset, a multi-class SVM will be built and fit to the image data (also known as train_X and test_X in the python implementation) – with each of the three types of radiography lung images being represented as different classes in the label data (also known as train_Y or test_Y programmatically). A one versus one, one versus

rest and grid search will then be carried out to determine the best hyperparameters for the model.

```
{'C': 100, 'degree': 3, 'kernel': 'poly'}  
{'C': 10, 'gamma': 0.1, 'kernel': 'rbf'}
```

Figure 3: Optimal Hyperparameters Output By Grid Search for SVM with HOG (Top Image) and SVM without HOG (Bottom Image).

To integrate HOG into the image data, the hog function imported from the skimage library was used to convert all test, train and validation image data into images with key features extracted in a HOG format to be inputted into the SVMs.

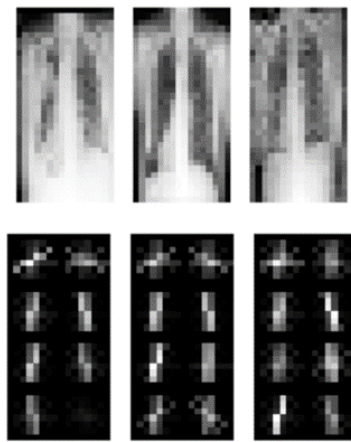


Figure 4: Images of Same Three Images; Before and After HOG Processing.

2 Random Forest With Local Binary Patterns and Linear Discriminant Analysis

For the 2nd non-deep learning approach, a random forest classifier is proposed to classify chest x-ray images using local binary pattern (Abbreviated to lbp) to extract useful features and linear discriminant analysis (Abbreviated to lda) is used to reduce dimensions.

The classifier is constructed using a decision tree based ensemble technique. The structure of each tree in the random forest is binary, created in a top-down manner, multiple decision trees are built and merged together. During the training of the model, the random forest iteratively splits feature vectors based on a feature, obtaining classification results at the leaves of the tree. At each branch, a variable and threshold is selected to split data based on a criteria, in this case gini impurity is used. There are 2 conditions that can end the iterative training. The first condition occurs when no more information gain is possible. The second condition occurs when the training process reaches a leaf node (Maximum depth of the tree).

LBP, is a gray-scale invariant texture descriptor measure for classification. A binary code is generated at each pixel by thresholding its neighborhood pixels to either 0 or 1 based on the value of the center pixel. This are the general steps involved in finding LBP for an image: (Pankajpatra, 2020)

1. Set a pixel value as the center pixel
2. Collect its neighborhood pixels (A 3x3 matrix is used resulting in a total neighborhood pixel of 8)
3. Neighborhood pixel is threshold to 1 if its value is greater than or equal to center pixel value otherwise threshold it to 0.
4. After thresholding, threshold values are collected from the neighborhood either clockwise or anticlockwise, producing a 8-digit binary code. The binary code is then converted into decimal.
5. Lastly, the center pixel value is replaced with the resulted decimal, repeating the same process for all the pixel values present in the image

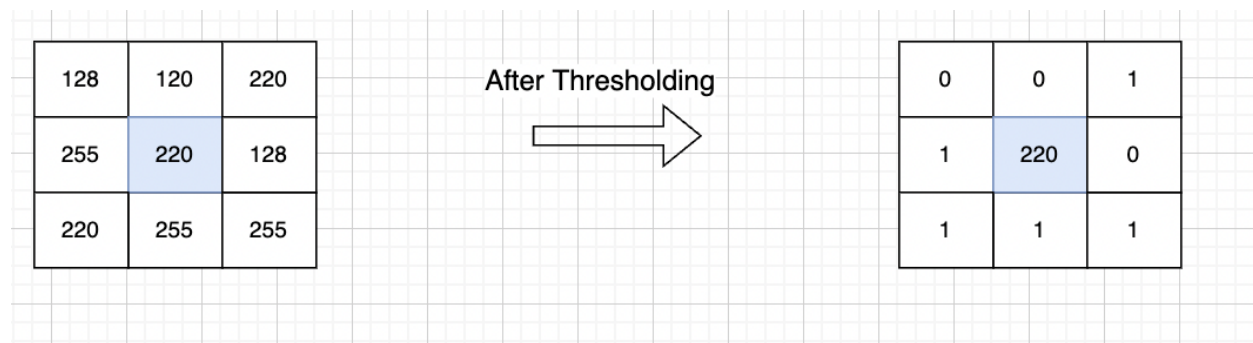


Figure 5: Results of 3x3 matrix pixels after thresholding

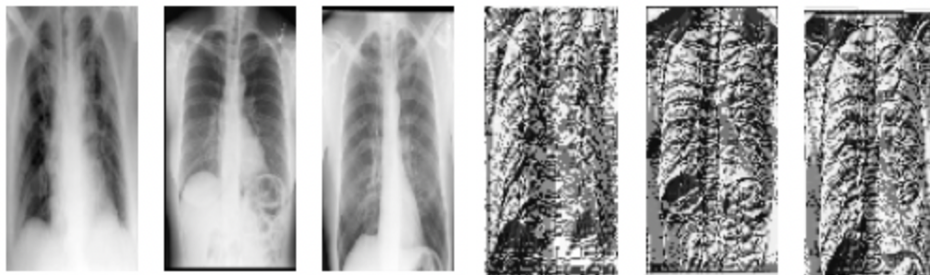


Figure 6: Pair of the same Chest X-ray before and after LBP processing

Lastly, using the image processed by lbp, lda is applied on the images. LDA seeks to find a projection that maximizes discriminative information and find the combination feature that separates the best between classes. Samples in the same class are pushed closer together and samples in different classes are pushed further away. (M.Farhan Tandia, 2020)

LBP is a simple yet efficient texture operator, it is invariant to monotonic changes in the grayscale domain, enabling it to present texture descriptors therefore it is useful in dealing with

images where there are very little local illumination changes. Its low computational requirement is helpful when you have a huge set of data that requires image processing. The dimensions in the images are reduced using LDA to further decrease training time. Hyper-parameter tuning is done using grid search for the following model: Normal Random forest and LBP+LDA image processed random forest, looking for the optimal number and maximum depth of trees.

(Hashem Davarpanah, S, Khalid, Fatimah, Abdullah, Lili Nurliyana, Golchin, Maryam, 2015)

Methodology	N_estimators	Max_depth	Class_Weights
Normal RF	102	11	None
RF + LBP + LDA	203	61	None

Figure 6: Results of grid search hyperparameter tuning

3 Convolutional Neural Network

Data augmentation has been used when training the deep convolutional neural network. This was performed so that it would help generalize the model more and therefore address the issue of class imbalance up to a certain extent. In terms of transformations that were used in this model; images were rotated in a range of zero to one hundred and eighty degrees, the width shift range and height shift range were both set to 20 percent, horizontal flips were activated and a validation split of 10 percent. These data augmentation transformations were performed only on the training and validation data sets.

Instead of putting together layers and coming up with a custom network architecture of our own, we have fine-tuned two already existing pre-trained convolutional neural networks for this task. These two networks are VGG16 and ResNet50. These models were chosen based on proven previous experiments where they provided good performance. And also because with both these models we had the ability to use different sized input images than the image sizes they were trained on. The weights used in both the VGG16 model and the ResNet50 model were pre-trained on the ImageNet dataset (J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, 2009). Both these models were trained using images sized 224x224x3.

The hyperparameters and such that was used for this task are as follows;

The input image size we have used for the task was constrained by severe computational restrictions. Various image sizes were tested to find the best that would suit the computational restraints and therefore, an input image size of 64x64x3 was used. Images were loaded into memory using a data generator with a batch size of 32. The batch size of 32 was maintained throughout training because any smaller than that would not have helped the model generalize better. During model training “categorical cross entropy” was used as the loss function, “softmax” as the activation function and “RMSprop” as the optimizer with a 0.00001 learning rate. A smaller learning rate was used with a higher number of inputs in order to lower the risk of overfitting and avoid the model from converging too quickly and reaching a suboptimal solution.

The VGG16 model was trained twice; once with Adam and once with RMSprop as the optimizer and the model with RMSprop outperformed the one with Adam.

When fine-tuning both these pre-trained models, we have frozen the last 3 layers of the network. The number of layers to be frozen was decided after trial and error runs and the architecture that gave the highest accuracy was chosen. Then one layer each of GlobalAveragePooling2D, BatchNormalization was added before two Dense + BatchNormalization blocks and then the final Dense layer with three neurons for the three classes in the classification output.

The fine-tuned ResNet50 model had a significant drop in performance when compared to the VGG16 model. This was likely caused by class imbalance, so class weights were utilized to try and improve the model. But it resulted in worse performance than the fine-tuned ResNet50 with no class weights.

In the discussion and evaluation section below only the model trained on the pre-trained VGG16 model will be considered, since it proved to be the more reliable and accurate model.

4 Visual Transformer

Transformers have recently Visual Transformers and have recently taken as the state of the art architecture for image classification.

As transformers were originally developed for seq-seq tasks to perform image tasks, images must be transformed into a sequence. It achieves this by splitting images into patches. These patches are flattened into a linear projection and combined with positional embedding. This vector input into a standard transformer encoder which classifies the image. The encoder is comprised of multiple layered blocks containing multiheaded attention, MLP units and layer normalisation. Using these self-attention layers, the transformer is able learn both local and global spatial information across the entire image. It achieves this all without the use of convolutional layers. Given enough data, ViTs are capable of outperforming similar state of art CNN architectures but requiring fewer trainable parameters and layers. The figure below provides an overview of the ViT process discussed:

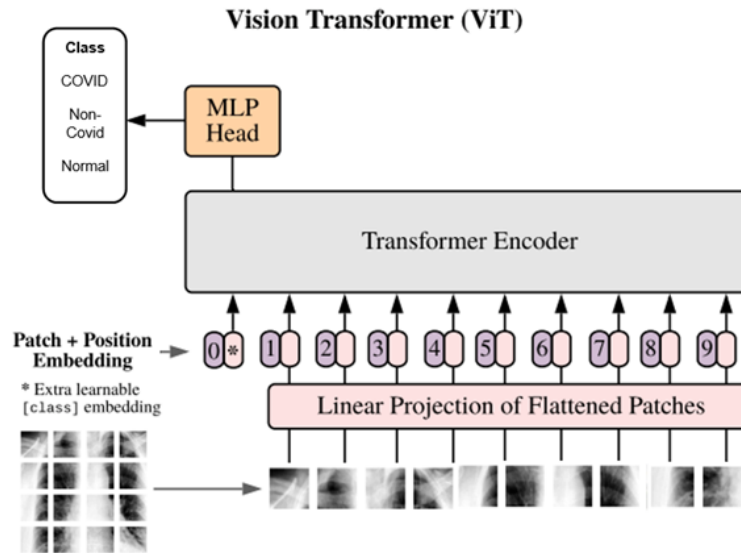


Figure 7: Vision Transformer Architecture

As an initial pre-processing step, images are down sampled to 125x125. This was a necessary trade off as it allows images to fit within memory constraints whilst keeping a respectable batch size of 32. Too small of a batch size can greatly degrade the model's ability to generalize.

As found in (A. et al., 2021), increasing contrast and reducing noise in images yielded considerable benefits in CXR image classification. To implement a similar pre-processing pipeline all images were passed through an adaptive contrast equalization filter before being fed into the network. This stretches out the image's pixel intensity range and allows for greater localized contrast.

Data augmentation was utilized to increase sample sizes, this consisted of small horizontal and vertical translations, as well as $\pm 10\%$ zoom.

The exact transformer architecture used is heavily constructed from (Khalid, 2021) which in turn is a keras implementation of the original ViT paper (Dosovitskiy et al., 2021). Rather than using a standard ViT it is proposed to incorporate convolution layers into the transformer to form a Con-ViT. This replaces the default patch extractor and encoder with a convolutional based tokenizer. Generally, the inclusion of convolutional layers allows for better spatial feature extraction and has seen performance benefits over standard ViTs in other image classification tasks. Whilst the standard ViT architecture has seen some limited and promising use (Shome et al., 2021), to our knowledge is the first time a Convolutional ViT has been applied to this COVID classification task.

Testing conducted comparing a standard ViT to Convolutional ViT resulted in a 4.3% accuracy increase on the validation set. Although a trade-off to this performance increase is that the Con-ViT tripled the number of trainable parameters yielding longer training times.

Due to the vast number of hyper parameters and architecture choices it is not feasible to perform an exhaustive fine tuning of the network given the available computational resources and time permitted to conduct experiments. As such hyper parameters such as training parameters (optimizer, learning rate ect.) remain the same. Whilst these cannot be determined optimal, the following architecture design choices were selected from limited experimentation:

- Projection Embedding Size: 36
- Number of attention heads: 5
- Number of transformer layers: 3
- Output dense layers: 256,128

A smaller limited search was also conducted to determine the optimum patch size. The results of this search are presented in table {}:

Patch Size	Average Accuracy on Validation Set
8x8	91%
16x16	90%
32x32	89%
64x64	87%

Accounting for stochastic randomness, patch sizes of either 32x32 or 64x64 produced the best performance. It would seem that too small of a patch size reduces and also too large of a patch size does not provide enough nuanced localized information.

Models were trained for 50 epochs and allowed to converge. Early Stopping was utilized to minimize overfitting. Class weights were experimented with to combat class imbalance however all weight combinations resulted in poorer performance on the validation set than the default.

Evaluation & Discussion

	<u>COVID</u>			<u>Pneumonia</u>			<u>Normal</u>		
Methodology	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
HOG - SVM	0.59	0.59	0.59	0.59	0.79	0.77	0.84	0.84	0.84
RF + LBP + LDA	0.77	0.72	0.75	0.81	0.87	0.84	0.9	0.91	0.9
CNN - VGG16	0.96	0.82	0.88	0.93	0.94	0.93	0.94	0.98	0.96
ViT	0.91	0.90	0.90	0.97	0.78	0.86	0.94	0.97	0.96

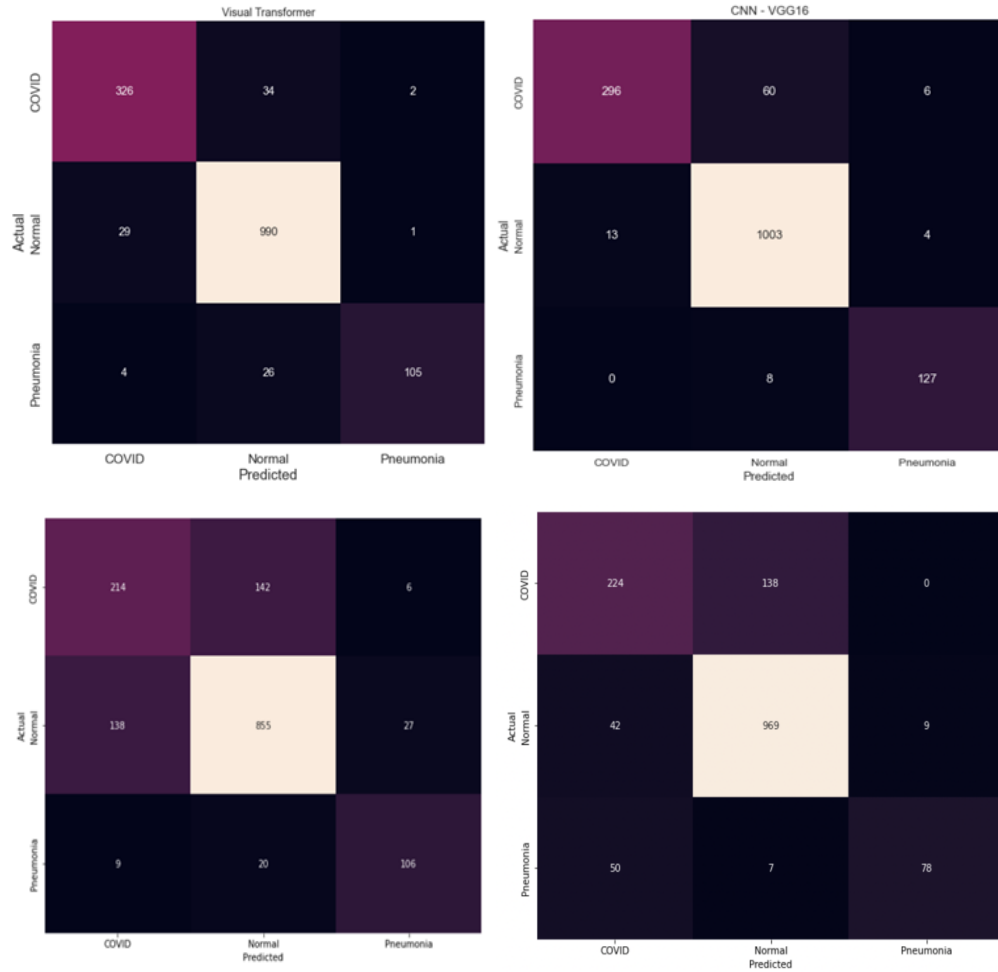


Figure 8: Four Testing Data Confusion Matrices From the Visual Transformer (Top Left), CNN (Top Right), SVM with Hog (Bottom Left) and RF with LDA and LBP (Bottom Right).

Overall

All proposed methods present high recall in detection of COVID. A good distinction between COVID and pneumonia exists with almost no pneumonia samples being incorrectly classified as COVID. These results support the decision to choose these methods, and suggest they are sufficient classification methods for multi class detection, further reasoning behind these choices included in the earlier methods section of this report.

Non Deep Learning Methods

In regards to the two non-deep learning methods - the RF with LBP and LDA alongside the SVM with HOG preprocessing - the random forest model produced a notably higher performance classifying between the 3 classes, approximately 5 to 10 percent higher in precision, recall and f1 score. Both the SVM and the RF were most effective at classifying normal radiography

images, followed by pneumonia images then COVID positive images. This result is logically sound as the higher performing approach is RF, a model that is more “robust against the overfitting problem” and less prone to class imbalance than SVMs that are highly susceptible to noise (Sheykhmousa, 2020).

Isolating the SVM performance only, a grid search and model fit for an SVM was completed with and without HOG being integrated into the image data. The difference in f1 value between the testing sets with HOG applied were roughly 10 percent lower in performance than the unfiltered image data. A similar pattern can be inferred from the RF without adding LBP or LDA yielding higher precision, recall and f1 values compared to the same data utilizing these additional preprocessing approaches. The probable cause of this result going against what was expected - the HOG and LDA/LBP increasing performance - is that feature extraction methods, while capturing meaningful components of images, can remove parts of the image that may be key to its classification. In this case, each lung depicted in each radiograph image on the surface are very similar looking, so this would be an example where the original image may provide more useful detail and information to a model than a HOG or LDA converted image could. In addition to this, while HOG and LDA/LBP are often state-of-the-art tools to use to improve classification success, there are a few key features related to the nature of the chosen dataset that prevented this improvement while using them. Firstly, when viewing the HOG images for example, the output images depict several markings that seem to outline the edges of each subject's lungs. Secondly, viewing the raw image data shows each subject not being completely aligned or in the same position in each image. Thirdly, the HOG imagery does not seem to include any features being highlighted within the inside of the lung, the details of how each lung looks potentially containing the vital distinctions between lungs with the class of COVID, normal or pneumonia. These observations can suggest that either higher resolution images need to be used so that HOG can pick up the detail in the lungs or that another feature extraction approach - one that captures the key differences between classes - is applied in future.

Deep Learning Methods

The VGGNet and ViT possess very comparable performance with very similar per class performance. The ViT does however have a slightly higher recall and true positive rate at detecting COVID. Though, the VGGNet has a lower occurrence of false positives with the ViT incorrectly classifying 33 persons as being COVID positive. Again, being a pre-screening tool, whilst false positive predictions are not ideal, they are certainly preferred to false negatives which would mean patients presenting with COVID would be incorrectly identified and not get treatment.

Both the ViT and VGG networks align somewhat closely with similar experiments carried out in literature. Differences in performance are most likely due to small variations in the dataset and fine tuning of hyperparameters.

As an example of class separation the T-SNE plot shown by Figure 9 below shows the learned embeddings of the VGG network. We can see clusters are somewhat grouped into distinct

clusters particularly with the pneumonia class being most distinct. Quite some overlap exists between the COVID and Normal images meaning the network sees these as more closely related.

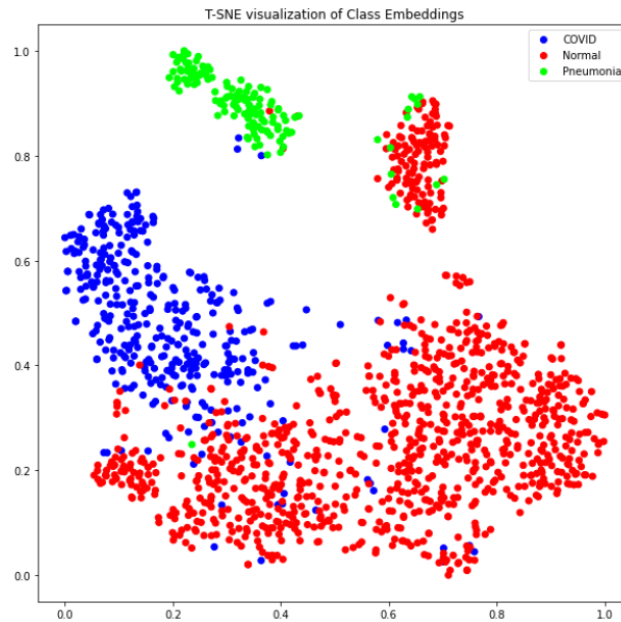


Figure 9: T-SNE plot for learned embeddings of VGG

Interestingly it would seem class imbalance has not had a large impact on the results with no large discrepancies in per class performance (only affected the ResNet50). Whilst the 'Normal' class presented with a much higher number of samples there is little bias towards it. Although when predictions are incorrect the models do favor 'Normal'.

Between the two convolutional neural network models: ResNet50 model and VGG16 model, a clear uniformity of performance can be observed. This makes logical sense as these neural networks have a range of built in features that allow them to produce desirable classification results. For one, as these pre-trained deep neural nets have been trained extensively - such as VGG16 - they still produce excellent outcomes from smaller sets of data. Additionally, the pre-trained ResNet model features identity mapping, an addition that "helps in avoiding the overfitting problem to the training set", making it more robust (Theckedath, 2020).

Given a larger sample size or ability to use transfer learning it is expected that the difference in performance between CNN & ViT would increase. It is likely both models have not been able to reach their full potential due to the small sample sizes.

Comparing Non Deep and Deep Learning Methods

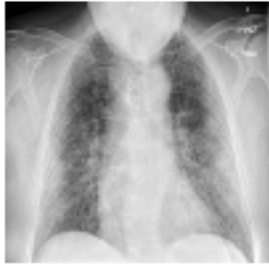


While both non deep and deep methods produced promising outcomes in correctly differentiating between classes based on radiography imagery, the distinctive feature that

separates these two categories is the superior performance in detecting COVID positive radiographs - the key aim of this study - in the ViT and CNN models over the SVM and RF traditional methods. Referencing the literature, a pattern can be observed that while binary SVMs have shown comparably better results for image classification over CNNs, CNNs have been found to surpass multi class SVMs (Jawale, 2019). Another potential factor in the success of the CNN and ViT models over the others is the increased performance with larger sets of ground truth data, such as the one that is the focus of this report.

1.1 Failure Cases

An in-depth determination of failure cases and their exact image characteristics leading to said failure requires expert domain knowledge which is not possible. However, an examination into perceived patterns within the test set can be performed. From this we can determine if models are classifying particular images correctly/incorrectly and identify ‘troublesome’ samples.

The table below provides an individual sample images and each models corresponding prediction:

	COVID 	COVID 	Pneumonia 
SVM	Normal	COVID	Normal
RF	COVID	COVID	COVID
VGG	Normal	COVID	Pneumonia
ViT	Normal	COVID	Pneumonia

As can be seen, models when incorrectly guessing most often class cases as normal. Again, this is most likely due to issues with class imbalance as normal is the most dominant class.

1.2 Computational Requirements & Training Times

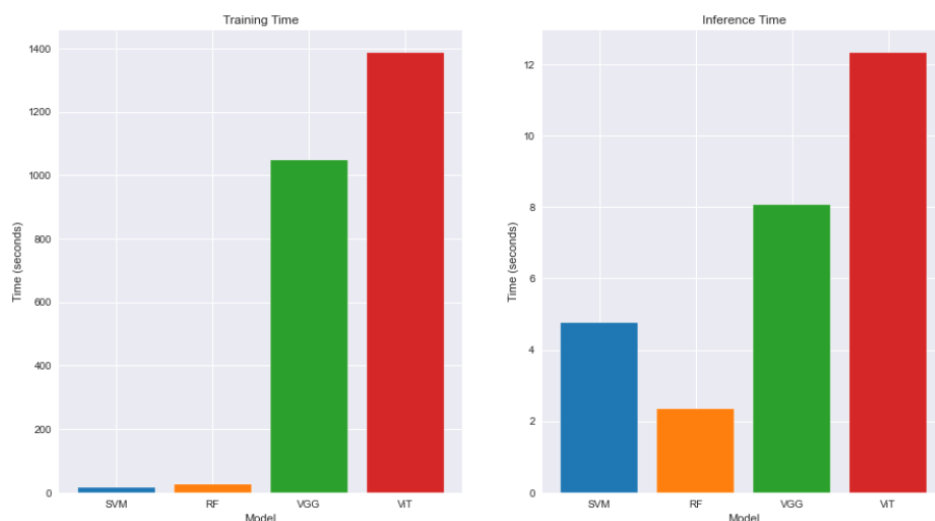


Figure 10: Train & Inference times for all developed models

Figure 10 above depicts both the training and inference times for each model. As expected, the training times for both the SVM and Random Forest are significantly lower than the deep learning methods (CNN & ViT). Whilst the overall computational requirements of the non-deep learning methods are lower, the entire training set must be loaded into memory at once which produces challenges with RAM overloading. Whereas deep learning benefits from dynamic batching in the form of data generators reducing RAM requirements.

Although the VGG has more than 4 times the number of trainable parameters than the ViT it is interesting to note that the training time is higher. This is likely caused by the use of a smaller batch size when training the ViT.

Inference times of all models are acceptable with the SVM and RF being slightly quicker to predict the training set. The longer inference times of the ViT are expected to be caused by the tokenization process. Computationally, all models are perfectly suitable for use cases in a hospital environment and could possibly even be used on handheld devices.

Conclusion & Future Works

To conclude, the non-deep and deep learning methods proposed do a fairly good job of pre-screening patients differentiating between COVID and non-COVID cases, giving doctors an

added layer of COVID19 pre-processing, allowing them to provide medical care to patients who are on a higher spectrum of urgency. Non-COVID patients can be turned away after preprocessing, saving more time for doctors. Predicting COVID19 cases being the key objective. The proposed deep learning methods has >90% success in predicting COVID19 cases

Even though the performance of non-deep learning methods with image processing are not as good compared to using it with original chest x-ray images, the performance of predicting COVID19 cases is still above 80%, providing fairly reliable results. Other forms of image processing technique can be applied to further increase the performance of non-deep learning techniques.

For future works, expanding the dataset to include more samples would be highly beneficial as this will further evaluate the developed models ability to generalize and allow for a better understanding of the real world performance with a wider selection of patients. Further samples would also be acquired to fix class imbalance issues found within the current dataset, increasing the number of samples for COVID and pneumonia cases.

Interpretability visualizations like Gradient-weighted class activation mapping (Abbreviated to Grad-CAM) can be used on deep learning methods. It uses the gradients of any target concept flowing into the final convolutional layer producing a coarse localization map highlighting specific important regions in the image used for prediction. This helps us to better understand what the model is focusing on leading to false classification, in our case a chest x-ray image expertise is required to provide information on which area of the image should our model focus on to increase accuracy. This is a good starting point, providing a direction in which we can head to improve the models.

Appendix

Group Contributions

Member	Contributions	Percentage
Celine - n10012711	<ul style="list-style-type: none"> - SVM - Literature Review - Discussion 	25%
Wei Jian - N10916172	<ul style="list-style-type: none"> - Random Forest - Conclusion - Discussion 	25%
Koralalage - N11226757	<ul style="list-style-type: none"> - CNN - Introduction - Discussion 	25%
Bailey - N10003703	<ul style="list-style-type: none"> - ViT - Literature Review - Discussion 	25%

Additional Non-Deep Learning Model Performance

	<u>COVID</u>			<u>Pneumonia</u>			<u>Normal</u>		
Methodology	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
RF	0.88	0.74	0.8	0.91	0.78	0.84	0.89	0.96	0.93
RF + LBP	0.86	0.32	0.46	0.91	0.16	0.27	0.74	0.99	0.85
SVM	0.88	0.74	0.80	0.93	0.92	0.93	0.91	0.96	0.93

References

Alhindi, T., Kalra, S., Ng, K., Afrin, A., & Tizhoosh, H. (2018). Comparing LBP, HOG and Deep Features. 1-7.

Jawale, A. &. (2019). Comparison of Image Classification Techniques : Binary and Multiclass using Convolutional Neural Network and Support Vector Machines.

Jian-Pei Zhang, Z.-W. L. (2005). A parallel SVM training algorithm on large-scale classification problems. 2005 International Conference on Machine Learning and Cybernetics, 1637-1641.

K. Zhang, J. D. (2011). Segmenting human knee cartilage automatically from multi-contrast MR images using support vector machines and discriminative random fields,. IEEE International Conference on Image Processing, 721-724.

M. Sheykhmousa, M. M. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6308-6325.

O. Déniz, G. B. (2011). Face recognition using Histograms of Oriented Gradients,. Pattern Recognition Letters, 1598-1603.

Ogole, C., Im, J., & Mountrakis, G. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 247-259.

Statnikov, A. W. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics, 319 .

Panka, Jpatra. (2020, March 6). *Create a local binary pattern of an image using opencv-python*. GeeksforGeeks. Retrieved June 11, 2022, from <https://www.geeksforgeeks.org/create-local-binary-pattern-of-an-image-using-opencv-python/>

Tandia, M. F. (2020, July 18). *Dimensionality reduction: PCA vs Lda for Face Recognition*. LinkedIn. Retrieved June 11, 2022, from <https://www.linkedin.com/pulse/dimensionality-reduction-pca-vs-lda-face-recognition-m-farhan-tandia/>

Hashem Davarpanah, S, Khalid, Fatimah, Abdullah, Lili Nurliyana, Golchin, & Maryam. (n.d.). *A texture descriptor: Background local binary - griffith university*. Retrieved June 11, 2022, from <https://research-repository.griffith.edu.au/bitstream/handle/10072/101354/GolchinPUB430.pdf;sequence=1>

Theckedath, D. S. (2020). Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. SN COMPUT. SCI., 1-79.

Yuan Yao, G. L. (2003). Combining flat and structured representations for fingerprint classification with recursive neural networks and support vector machines. Pattern Recognition, 397-406.

J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.