Using Lung ultrasound images for building a reliable Point-of-care Covid-19 testing system

By: Sanket Sanjay Bhosle (0776136), Mandeep Kaur (0779091),
Anjaliraj Baburaj (0781499), Simranjeet Singh (0783612),
Rishika Shivani Adulla (0784445)
Students, Data Analytics for Business, St. Clair College

Abstract

Covid-19 is a global health crisis that continues to have devastating effects on the whole world. Everyone has faced its impact and consequences in one way or another. It has not only challenged the health systems worldwide but has also shaken the social and economic systems. Since the virus spreads from person to person. Healthcare providers have been stressing about the importance of early diagnosis of people with the virus to control its spread. Even though RT-PCR and Rapid Tests are the widely used methods for Covid-19 diagnosis but the limited supplies of RT-PCR and long waiting time in getting results remain an issue. This study supports the concept of using machine learning for diagnosing COVID-19 using lung ultrasound images which would facilitate rapid clinical decision support. Using deep learning for a Point-of-care testing system would enable health care providers to conduct bedside examination of patients allowing quick tracking of virus among infected patients.

Introduction - Coronavirus disease is an infectious disease caused by SARS-CoV-2 virus. Most people infected with the virus experience mild to moderate respiratory illness and are able to recover without requiring any special treatment. But, some do require serious medical attention. The virus puts older people at risk along with those having medical conditions like heart diseases, diabetes, chronic respiratory illness or cancer. Anyone can catch the virus and become seriously ill)(World Health Organization,2022).

A person infected with coronavirus even if they show no symptoms may emit aerosols while talking or breathing. Aerosols are defined as infectious viral particles that can float around in the air for up to three hours. Another person breathing in these aerosols can become infected with the coronavirus (Harvard Medical School, 2022). Isolating infected patients and testing people they met are the primary steps taken after Covid-19 diagnosis.

Viral Tests are a way of diagnosing the virus, a viral test gives the test results based on samples collected from nose or mouth. Types of viral tests include Rapid tests and laboratory tests (Centers for Disease Control and Prevention, 2022).

- Laboratory tests :- Reverse Transcription-Polymerase Chain Reaction (RT-PCR) is a common example, nasal swab or saliva are taken as a sample. Such tests usually take 24-72 hours to return the results, which are reliable for people with or without symptoms (Centers for Disease Control and Prevention, 2022).

- Rapid tests :- Antigen tests, self-testing kits are common types of rapid tests. Nasal swabs are taken as a sample, the results are returned within 15-30 minutes. However, results are less reliable for people not experiencing any symptoms (Centers for Disease Control and Prevention, 2022).

The gold standard polymerase chain reaction (PCR) test used to diagnose COVID-19 requires laboratory processing that takes 1-3 days to return the results, due to backlogs and supply shortages. So an individual could be negative when tested but positive by the time the result is returned (Rubin, 2020). Moreover, such a delay is not ideal in case of emergencies. Rapid tests on the other hand provide quick results but they are not as reliable as laboratory tests, such tests have a high risk of giving False Negatives. The same goes for at home self-testing kits which are not as accurate as desired, False Positives are rare but False Negatives are very frequent. (Yetman, 2022). Rapid tests return a positive result 72% of the time on an average, with the 95% confidence interval being between 63.7% - 79% (Cochrane Database of Systematic Reviews, 2021) .

The need for rapid clinical decision support in COVID-19 can't be denied, making point-of-care diagnosis system imperative. Bedside examination of patients is important in case of emergencies to plan the course of care for them in time. Using lung ultrasound images for developing a point-of-care diagnosis system provides these advantages. Nevertheless, it is cost effective as the ultrasound equipment can be reused unlike the testing kits. They are also fast and easy to disinfect (Wong, 2022)

Solving the Problem - A quick and reliable Point-of-care Covid diagnosis system that returns the results instantly and is highly accurate. A machine learning based diagnosis system can be developed that classifies patients into right category of results – Positive or Negative using their lung ultrasound images. Such a model would help the healthcare professionals by saving a great deal of their time as the classification model would return the results momentarily. Moreover, the final deployed model's promising accuracy would save the trouble of conducting a series of follow up tests. The solution would change the clinical decision support as this would not only save care providers and patients from unnecessary delays and waiting times but will also enable healthcare professionals to take the necessary actions timely.

Research Questions - Following are the research questions based on the problem statement:

(I) Which machine learning classification algorithm works the best for diagnosing Covid-19 using lung ultrasound images evaluated using accuracy, precision, recall and F1 score?

(II) Which technique helps the most in achieving the success criteria by optimizing model performance – Transfer Learning, Data Augmentation, Class weights balancing, Stratification, SMOTE (Synthetic Minority Oversampling Technique) , Near Miss Under Sampling ?

Dataset - The largest curated collection of point of care lung ultrasound (POCUS) images has been led by Ashkan Ebadi at National Research Council Canada and Alexander MacLean at Vision and Image Processing Group (University of Waterloo), with Pengcheng Xi and Dr. Florea at McGill University and is named 'COVIDx-US' (Wong, 2022).

COVIDx-US is an open access benchmark dataset of ultrasound imaging related to COVID-19. The dataset has been curated from 9 different sources - Butterflynetwork, GrepMed, PocusAtlas, LITFL, Radiopaedia, CoreUltrasound, Papers, UF, Claurius and its most recent version consists of 242 lung ultrasound videos and 29,651 processed lung ultrasound images of COVID-19 infected patients, normal cases, pneumonia infected patients and those with other lung conditions. (National Research Council Canada,2022)

192 lung ultrasound videos and 18,628 images have been extracted from 8 sources - GrepMed (20 videos) , PocusAtlas (32 videos), Clarius (6 videos) , LITFL (63 videos), CoreUltrasound (18 videos) , Radiopaedia (5 videos), Papers (22 videos), UF (26 videos).

The lung ultrasound videos have been captured with one of the two probes – Convex (14,239 images) and Linear (4389 images). The length of the images range from 197 pixels to 1350 pixels, whereas the width of the images range from 198 pixels to 1920 pixels. All the images were stored in .JPG format. The images had variations in dimension since these were extracted from videos that have been collected from different sources and captured with different probes.

The dataset has lung ultrasound images belonging to four classes – COVID (4003 images), Normal (2201 images), Other (7975) and Pneumonia (4449).

Methods of Analysis – Convolutional neural networks have been used for developing classification models capable of predicting COVID-19 and differentiating it from other conditions. However, a series of data pre-processing steps have been performed on the image dataset to make it ready before using it for the classification task.

Data Pre-processing –

1. Duplicate images were removed from each class – Covid, Normal, Pneumonia and Other.

   1.1 Algorithm md5 has been used for converting each image into its corresponding 32 character hash value.
   1.2 Images with exact same hash value were considered duplicates of one another.
   1.3 All duplicates of an image except one, were removed from the dataset.
   1.4 This reduced the number of images in the dataset to 15,798.

2. All images in the dataset were normalized i.e. the pixel values were rescaled to be between 0-1.

3. Dimension variation in images were addressed by resizing all images to have a length of 197 pixels and width of 198 pixels.

4. Images were stored in RGB mode.

5. Finally, the entire dataset was split up into three parts namely training, validation and testing with a ratio of 7:1:2

   5.1 Training set consisted of 70% of images – 9691 images.
   5.2 Validation set had 10% of images – 1384 images.
   5.3 Test set contained the rest 20% of images – 2771 images.

Results and Discussion -

Two kinds of classifiers have been developed using the lung ultrasound images – Binary and Multi-class classifiers.

Binary Classification **–** The entire image dataset was organized into two classes namely Covid (3995 images) and Non-Covid (11803 images) , the latter consisted of all the three classes i.e. Normal, Pneumonia and Other.

Baseline Model **-** Pre-trained convolutional neural network VGG-16 has been used as a baseline for two class classification. The model was used in combination with Image Data Generators for inputting data as batches with every batch of 512 images of dimension (197,198,3). The parameters of layers in feature extractor part of VGG-16 were retained as learned from the 'imagenet' dataset. The model returned a prediction accuracy of 55% on the unseen set of images in the test dataset. However, the model's performance was fairly better on the Non-Covid class as it was on the other. The same can be attributed to the major class imbalance in the dataset i.e. the number of images in the two image classes were not comparable which in turn lead to the model's biased performance towards the majority Non-Covid class.

Based on the literature conducted, a set of techniques for handling class imbalance along with machine learning models for implementation were identified. The performance returned by these models has been documented in Table 1.

The techniques used have been summarized below :-

(i) ***Offline Data Augmentation*** - Since the class imbalance is really prominent and we can't expect the model to generalize well on covid class if it did not get to learn from Covid class as much as it did from Non-Covid class. Performing offline data augmentation on Covid class in training dataset i.e., transforming the existing images for creating new, increased the number of images in the training dataset.

For the image transformation, rotation range was set to 40 degrees, shear range to 2 degrees, zoom range to 2 degrees with the horizontal flip enabled and brightness range set to 0.5 to 1.5. The augmentation generated around 4800 images, increasing the number of images to 7600 in Covid class for the training set.

The overall accuracy went up by 23%.

(ii) ***Near Miss Under Sampling*** **–** This under sampling technique for 'Imbalanced Learn' library uses K Nearest neighbours (KNN) for under sampling the minority class. It selects the Non-Covid samples for which the average distance to the N closest samples of the Covid class is the smallest and removes them. Support Vector Machine has been trained for the classification task after performing this technique. This resulted in a prediction accuracy of 86%.

(iii) ***Synthetic Minority Oversampling Technique*** – SMOTE is an oversampling technique from 'Imbalanced Learn' library. The technique has been applied while fitting Random Forest Classifier on the training data. It ensured that the minority class (Covid) gets oversampled by replicating random images and both of the classes (Covid and Non-Covid) end up with equal number of images. It improved the model's performance by 34% .

(iv) ***Class weights balancing*** – This method balances weights of both classes in such a manner that the model does not get biased towards a certain class regardless of the number of instances in that class.

Compute_class_weights() method of class_weight class has been used for calculating weights of both classes. The minority class i.e. Covid as expected is given a greater weight than the majority class Non-Covid to make these two balanced. The weights are – Covid : 1.71 , Non-Covid : 0.70

Training ResNet50V2 with balanced class weights gave the best performing model.

Model	**Technique**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
VGG-16 (Baseline)	-	55%	66%	67%	66%
VGG-19	Offline Data Augmentation	78.40%	76%	98%	86%
Support Vector Machine	Near Miss Under sampling	86.50%	90%	84%	85%
Random Forest Classifier	SMOTE	89.02%	92%	86%	88%
Logistic Regression	-	84.37%	89%	83%	83%
ResNet50V2	Balanced Class Weights	93.17%	99%	90%	94%

Table 1. Machine learning models along with the key performance indicators.

Considering all of the performance metrics, ResNet50V2 outperformed all other models.

- **Accuracy** – The model correctly identifies 93% of the total images in the test dataset.
- **Precision** – When the model classifies an image as of COVID class, it is correct 99% of the time
- **Recall** – Of all the COVID positive patients in the test dataset, the model can correctly identify 90% of them.
- **F1-Score** – The harmonic mean of precision and recall is 94%.

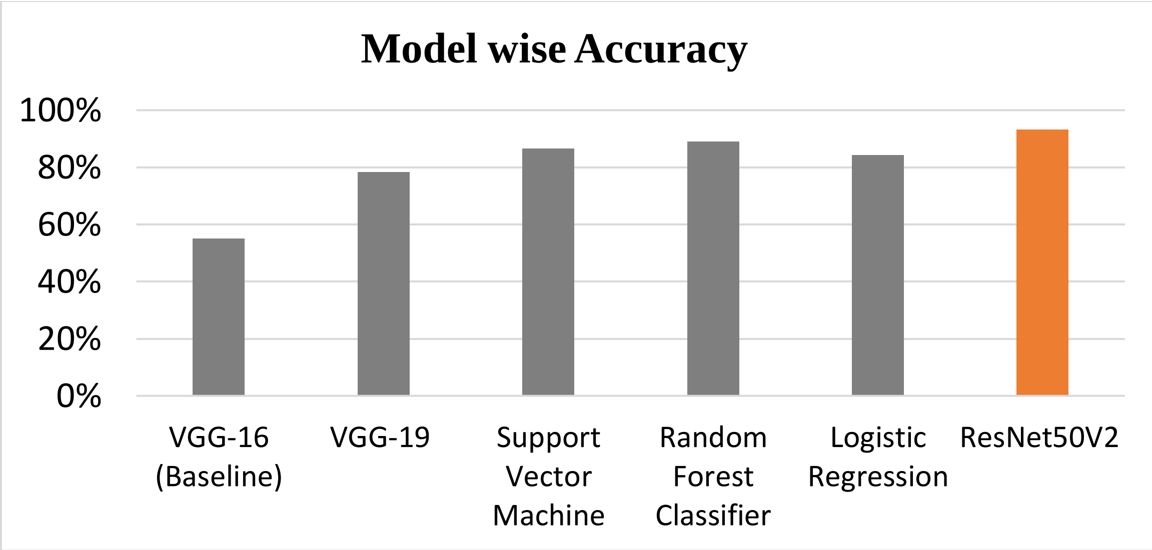

Fig1. Model wise accuracy of binary classifiers

The confusion matrix for the model shows that it is competent in identifying COVID-19 (99.8% accuracy) from lung ultrasound images and it is fairly good at the Non-Covid class as well (90.3% accuracy). The can be compared to the baseline model's confusion matrix for the performance improvement.

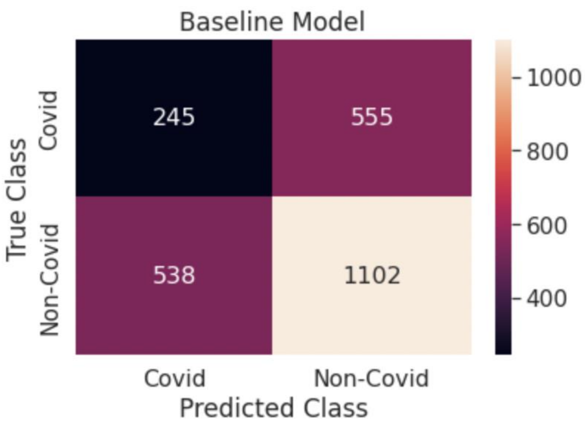

Figure 2. Confusion matrix for Baseline Model

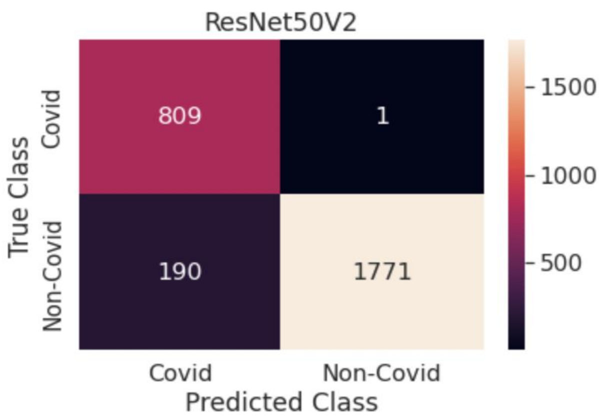

Figure 2. Confusion matrix for ResNet50V2

Multi class classification – Four class classifiers have been trained and tested to classify lung ultrasound images into four classes – Covid (3995 images), Normal (1611 images) , Pneumonia (2855 images) and Other (5335 images).

Four different machine learning models along with the class imbalance techniques have been used for the task. The performance metrics are reported in Table 2.

Their process has been explained as follows :-

(i) ***Balanced Bagging Classifier :-*** The bagging method creates multiple subsets of the training dataset and makes predictions on each subset, final prediction is an aggregate of the predictions from each subset. It resamples the imbalanced classes while training the model, so that all classes have comparable number of instances. Resampling strategy used is 'Not Majority' which resampled all other classes except majority (Other) so that they have number of images closer to the majority class.

Using balanced bagging classifier with decision tree as a base estimator returned good accuracy of 95.37%.

(ii) ***Stratification*** – Stratified sampling of training, testing and validation dataset ensures that all of the classes in each of these sets gets images in the same proportion as it was present in the original dataset. This has been performed at the time of splitting the dataset into three parts. Initially, the proportion of images in all the four classes was as follows :-

Covid : 0.29 , Normal : 0.11 , Pneumonia : 0.21 , Other: 0.39

All four classes were distributed in the same proportion in training, test and validation set. Multinomial logistic regression model along with stratified sampling returned an accuracy of 94%.

Model	Technique	Accuracy	Precision	Recall	F1-Score
ResNet50V2	Balanced class weights	87.8%	89%	87%	85.6%
Xception	Balanced class weights	96.90%	93%	99%	97%
Decision Tree Classifier	Balanced Bagging Classifier	95.37%	95%	93%	96%
Multinomial Logistic Regression	Stratification	94.42%	96%	96%	96%

Table 2. Machine learning models along with the key performance indicators.

As per the performance metrics and the time taken for returning predicted class, pre-trained model Xception turns out to be the best multi classifier. The same can be inferred from the Fig. 4. Moreover, the model took the least time for training and returned the prediction classes instantly.

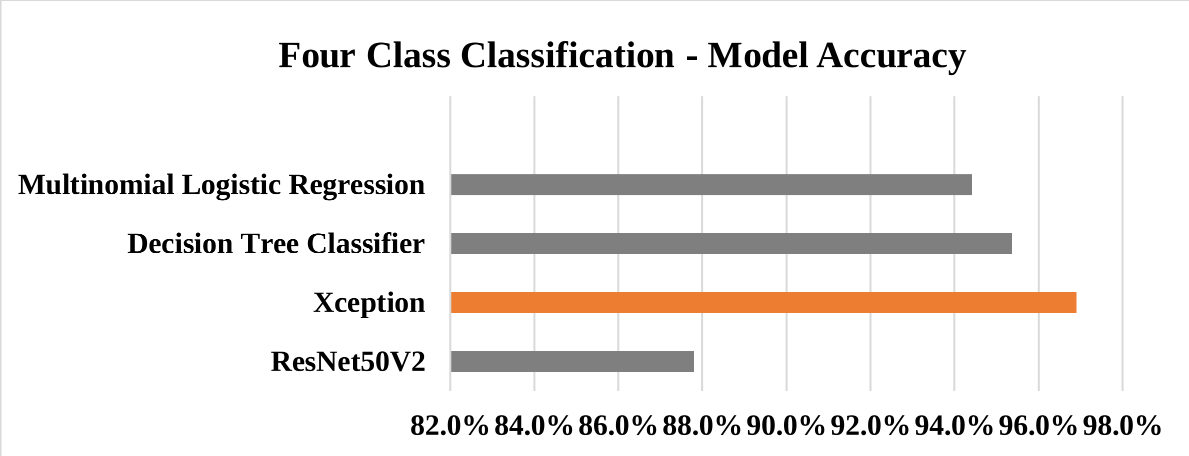

Fig 4. Model wise accuracy

The model has returned an exceptionally great performance on the COVID class, with a value of 1.0 for precision as well as recall. This means that when the model classifies a lung ultrasound image as that of COVID positive it is correct 100% of the time. Moreover, the model is able to correctly identify 100% of the COVID positive cases.

The class wise performance of Xception has been shown by the classification report (Figure 5) as well as the confusion matrix (Figure 6). The class representation is as follows :
0 – Normal, 1- Covid, 2- Pneumonia, 3 – Other. The classification report summarizes the model performance for all the four classes.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	322
1	0.93	1.00	0.96	1068
2	1.00	0.86	0.92	568
3	1.00	0.99	1.00	800
accuracy			0.97	2758
macro avg	0.98	0.96	0.97	2758
weighted avg	0.97	0.97	0.97	2758

Fig 5. Classification report of Xception

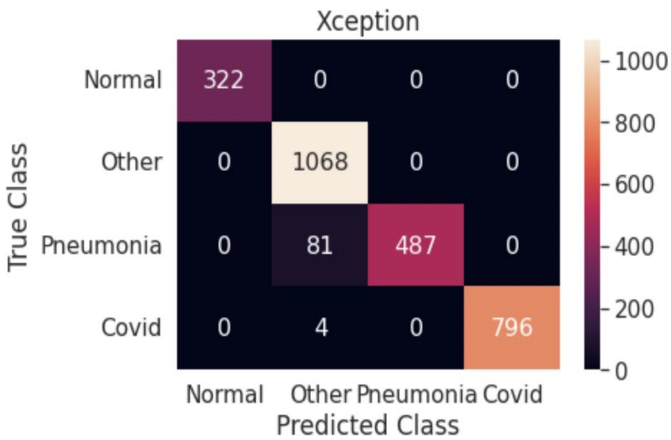

Fig 6 Xception model’s confusion matrix

Conclusions – The study showed that Convolutional neural networks in combination with Image data generators and transfer learning have the capability of being an alternative approach for lung ultrasound image classification for predicting COVID-19 and differentiating it from Non-Covid cases. Deep learning architectures like ResNet50V2 and Xception along with algorithms like Support Vector Machine, Random Forest Classifier and Balanced Bagging Classifier can be used for classifying COVID positive cases and other classes. Using balanced weights for each class, stratified sampling and data augmentation are some of the techniques which can make the models robust and unbiased towards all classes. With a good proportion of images belonging to each class, models can be trained to achieve desirable prediction performance.

Recommendations – Based on the findings and conclusions, the following recommendations can be made: -

- More lung ultrasound images can be collected to enhance classification model's performance. Training classification model on a greater number of images will make it possible for the model to extract more features and learn better than before.

- The scope can be expanded by including more cases other than COVID, Normal and Pneumonia. As there are only three well defined categories of lung ultrasound images as of now, adding more classes will expand the model's scope of classification.
- The 'Other' category can be researched to understand the kind of cases contained in this class. This is the only class of lung ultrasound images which is not clearly defined. Additionally, it has the highest number of images among all classes, it'd make sense to study this class.

- New attempts can be made to gather more images for COVID class. Even though handling the class imbalance while model training has helped but having more images from the class would help the model learn better.

- The model's performance can be validated with help of RT-PCR tests.

- An application capable of identifying and differentiating COVID-19 from other conditions using lung ultrasound images should be developed. The same must be tested and validated in a clinical setting.

References -

Centers for Disease Control and Prevention (2022, February 25). How to protect yourself and others. CDC 24/7: Saving Lives, Protecting People
https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html

Harvard Medical School (2022, February 28). Preventing the spread of Coronavirus. Harvard Health Publishing
https://www.health.harvard.edu/diseases-and-conditions/preventing-the-spread-of-the-coronavirus

Laguarta, J., Puig, H.F., Subirana,B. (2020). COVID-19 Artificial Intelligence Diagnosis using only cough recordings.
https://www.researchgate.net/publication/344930557_COVID-19_Artificial_Intelligence_Diagnosis_using_only_Cough_Recordings

National Research Council Canada (2022). COVIDx-US : An open-access benchmark dataset of COVID-19 related ultrasound imaging. Github, Inc.

https://github.com/nrc-cnrc/COVID-US

Tamal, M., Alabdullah, M., Hourani, R., Alshammari, M.N. (2021). An integrated framework with Machine Learning and Radiomics for accurate and rapid early diagnosis of COVID-19 from Chest X-ray.

https://www.researchgate.net/publication/351337385_An_Integrated_Framework_with_Machine_Learning_and_Radiomics_for_Accurate_and_Rapid_Early_Diagnosis_of_COVID-19_from_Chest_X-ray

Walid, H., Narin, A. (2021). Deep neural networks for COVID-19 detection and diagnosis using images and acoustic-based techniques : a recent review (Review) https://www.researchgate.net/publication/354102369_Deep_neural_networks_for_COVID-19_detection_and_diagnosis_using_images_and_acoustic-based_techniques_a_recent_review

Wong A. (2022, January). *Please share with anyone interesting in AI and healthcare*With the rise of the Omicron variant and the world continuing [Post]. LinkedIn https://www.linkedin.com/posts/alexander-wong-90650216_icu-ai-edgeai-activity-68921045583411527682P2h/?utm_source=linkedin_share&utm_medium=member_desktop_web

World Health Organization (2022, January 29). Coronavirus disease (COVID-19). who.int https://www.who.int/emergencies/diseases/novel-coronavirus-2019

Zoabi,Y. Deri-Rozov, S., Shomron,N. (2021). Machine Learning based prediction of COVID-19 diagnosis based on symptoms. https://www.researchgate.net/publication/348205307_Machine_learningbased_prediction_of_COVID-19_diagnosis_based_on_symptoms

Link for GitHub repository -

https://github.com/Sanketb77/Point-of-care-COVID_detection