**Business Problem Formulation**

**Sanket Sanjay Bhosle Rishika Shivani Adulla**

(0776136) (0784445 )

**Mandeep Kaur Anjaliraj Baburaj**

(0779091) ( 0781499)

**Simranjeet Singh**

(0783612)

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

**Problem Statement:-** Covid-19 has had and continues to have devastating effects on the whole world. Everyone has faced its impact and consequences in one way or another. 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).

The coronavirus spreads mainly from person to person. A person infected with coronavirus even if they show no symptoms they 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).Healthcare providers have been stressing about the importance of early diagnosis of people with the virus to control its spread. 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](https://www.cdc.gov/coronavirus/2019-ncov/testing/diagnostic-testing.html) 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 takes 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).

However, 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. Also, to help control its spread quick tracking of evolution of virus among infected patients will prove helpful .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)

1. We hypothesized that COVID-19 subjects,
2. especially including asymptomatics, could be accurately
3. discriminated only from a forced-cough cell phone recording using
4. Artificial Intelligence. To train our MIT Open Voice model we
5. built a data collection pipeline of COVID-19 cough recordings
6. through our website (opensigma.mit.edu) between April and May
7. 2020 and created the largest audio COVID-19 cough balanced
8. dataset reported to date with 5,320 subjects.
9. We hypothesized that COVID-19 subjects,
10. especially including asymptomatics, could be accurately
11. discriminated only from a forced-cough cell phone recording using
12. Artificial Intelligence. To train our MIT Open Voice model we
13. built a data collection pipeline of COVID-19 cough recordings
14. through our website (opensigma.mit.edu) between April and May
15. 2020 and created the largest audio COVID-19 cough balanced
16. dataset reported to date with 5,320 subjects.
17. We hypothesized that COVID-19 subjects,
18. especially including asymptomatics, could be accurately
19. discriminated only from a forced-cough cell phone recording using
20. Artificial Intelligence. To train our MIT Open Voice model we
21. built a data collection pipeline of COVID-19 cough recordings
22. through our website (opensigma.mit.edu) between April and May
23. 2020 and created the largest audio COVID-19 cough balanced
24. dataset reported to date with 5,320 subjects.
25. Machine learning-based prediction of COVID-19 diagnosis
26. based on symptoms

**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 proposed: -

1. Which machine learning classification algorithm works the best for diagnosing Covid-19 using medical data of patients (Lung ultrasound images) evaluated using accuracy, precision, sensitivity, F1 score, ROC Curve – AUC Score, Log loss ?
2. Which technique helps the most in achieving the success criteria by optimizing model performance – Transfer learning, Data augmentation, Mix-up augmentation, Progressive resizing, Deep learning libraries like fastai, Hyperparameter tuning using Keras Tuner, fine tuning ?

**Review of related studies:-** Several attempts have been made to develop alternative COVID-19 testing systems, some of them have been reviewed below :-

1. A free, real time, non-invasive AI speech processing framework has been developed that makes use of acoustic feature extractor for detecting COVID-19 using cough recordings. The framework was trained on the largest audio COVID-19 cough dataset with 5320 subjects (Laguarta, Puig, Subirana , 2020). However, the model does not account for differences in cough due to age and cultural/regional differences. Moreover, people coughing for getting tested would spread aerosols in air that tend to float around for up to three hours, making it an safe testing method.
2. A comprehensive review of deep learning based neural network for COVID-19 detection and diagnosis using chest X-ray and CT- Scan images revealed that deep learning algorithms are a time saving alternative to gold standard of COVID testing (RT-PCR) and an efficient way of differentiating pneumonia or other lung infections from coronavirus (Walid, Narin, 2021).But, CT scan images have too many cross sections of the same person making it difficult for the algorithm to extract features. Also, such models fail to incorporate manual/ground truths with the images like age, gender or other medical information of patients.
3. An Integrated Framework with Machine Learning and Radiomics for Accurate and Rapid Early Diagnosis of COVID-19 from Chest X-ray has been developed and it’s performance has been evaluated. The solution uses a combination of machine learning and radiomics for diagnosing coronavirus using chest X-ray images in presence of other diseases like pneumonia, bacterial infection and at different severity levels (Tamal, Alabdullah, Hourani, Alshammari ,2021). The system does not require manual intervention, however it takes up to 2 minutes for returning the results. Also, the sensitivity drops with increase in severity level as the patterns that act as radiomics features of COVID-19 vanish and become less distinguishable .
4. A machine learning based prediction model diagnosing COVID-19 based on symptoms is a predictive model that classifies person as COVID positive or negative as per symptoms inputted, this can be used as quick and alternative diagnosis technique in cases of limited healthcare resources (Zoabi, Deri-Rozov, Shomron ,2021). But, the model lacks robust training data; also since the input given is as per the questions asked and self-reporting of answers is subject to bias, making it an unreliable option.

**Comparison of solution with existing studies: -** Ultrasound machines are readily available, reusable, and easy to disinfect. The procedure is easy and does not expose patients to harmful X-rays for prolonged periods. There is no risk of bias as clinicians use the equipment themselves for taking lung ultrasound images. Ultrasound machines are globally accessible and the solution would not require intensive computational power making it useful in areas with limited resources. Due to the complexities in CT- scan images, radiologists have to manually analyze the images along with a diagnosis system, there is no such case for the solution. Moreover, abundance of data is available as lung ultrasound images have been collected for years before COVID as well, for non-COVID cases and other lung infections.

**Knowledge, skills, technology needed: -** In order to develop the solution and derive answers for the research questions, the team needs to have Python programming skills, knowledge of image classification machine learning models, model performance evaluation metrics, performance optimization techniques, feature engineering, transfer learning, data augmentation. Moreover, the team needs to have hands on experience with Jupyter notebook , Google Colab, technology for deploying the model as a web based interface.

**Success of solution :-** A model providing results instantly with high values of accuracy, precision, recall, minimal false positives and false negatives, minimal log loss would be considered a successful solution.

**Solution Evaluation:-** The solution will be evaluated as per its performance on unseen lung ultrasound images that would not be a part of the image set used for training. Accuracy, precision, sensitivity, F1 score, ROC Curve – AUC Score, Log loss are the metrics to be used for performance evaluation. A model returning high values for these and takes the least time for returning the results would be considered for final deployment . This is a feasible approach as it ensures that the model generalizes well on real world data extracting features from the data instead of just memorizing it.

**Data Pipeline: -**

1. Defining the business problem **->**
2. Review of related studies **->**
3. Data collection:- Data curation from data sources, Data cropping , Ultrasound image extraction **->**
4. Data Processing **:-** Cleaning image frames , normalization, image resolution, dataset verification **->**
5. Splitting dataset :- Using hold out method/K-fold cross validation/Iterative K-fold cross validation **->**
6. Training the baseline model ->
7. Model performance optimization :- Transfer learning, data augmentation, feature engineering, Keras tuner, fine tuning ->
8. Model evaluation ->
9. Model deployment

**Ethical Concerns:-** The analysis must be respectful of privacy and anonymity of the people who provided the data, and it is not a concern as that the data does not have any kind of identifiable information about the data providers. Regarding the conscious data usage, clear instructions are available, before proceeding with the data usage for licensed sources, it has to verified that the usage is allowed as per the Creative Commons (CC) license. For unlicensed data sources, the intended usage has to be verified (National Research Council Canada ,2022).

**Estimated impact on stakeholders :-**

* Having a Point-of-care Covid prediction model would help the healthcare providers in cases of emergency and in ICU’s (intensive care units) by reducing the time for getting results. They can utilize the same time in planning their course of care for the patient. The infected patient can be isolated quickly. Moreover, it could save a lot of resource wastage in terms of equipment, time and personnel as this diagnosis system would not require conducting follow up tests for reliability.
* It’d also ease up things for patients as they would not have to go through excruciating waiting time and delays. They’d also have the peace of mind having received a reliable and accurate result. It’d become a viable option for people because of its cost and mobility. Also, patients would feel comfortable as such a test is not invasive and also allows for social distancing.

**Data sources :-** The largest curated collection of point of care lung ultrasound (POCUS) images has been led by [Ashkan Ebadi](https://ca.linkedin.com/in/ashkanebadi?trk=public_post_share-update_update-text) at National Research Council Canada and [Alexander MacLean](https://ca.linkedin.com/in/alexander-maclean?trk=public_post_share-update_update-text) at [Vision and Image Processing Group](https://www.linkedin.com/company/vision-and-image-processing-group?trk=public_post_share-update_update-text) ([University of Waterloo](https://www.linkedin.com/company/uwaterloo?trk=public_post_share-update_update-text)), with [Pengcheng Xi](https://ca.linkedin.com/in/pengcheng-xi?trk=public_post_share-update_update-text) and Dr. Florea at [McGill University](https://www.linkedin.com/company/mcgill-university?trk=public_post_share-update_update-text) 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, non COVID-19 infected people, normal cases and those with other lung conditions. (National Research Council Canada,2022)

**Challenges :-**

1. The lung ultrasound images that would form the dataset are to be collected from 9 different sources, scraping thousands of images from ultrasound videos might be computationally and time intensive.
2. The images might not be balanced among all classes making it harder for efficient utilization of dataset for predictive model.
3. Normalizing all of images collected into a standard resolution, scaling and formatting pose a challenge too.

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-6892104558341152768-2P2h/?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_learning-based_prediction_of_COVID-19_diagnosis_based_on_symptoms>