Mean Teacher Model with Consistency Regularization for Semi-Supervised Detection of COVID-19 using Cough Recordings

Dinesh Kumar M R[1], K S Paval[1], Shreya Sanghamitra[1], Shrish Surya N T[1], Jyothish Lal G[1], and Vinayakumar Ravi[2]

[1] Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India.
g_jyothishlal@cb.amrita.edu
[2] Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia.

Abstract. This study provides a novel approach for detecting COVID-19 using cough recordings. The method employs a Mean Teacher model with consistency regularization, utilizing both labeled and unlabeled data. The model is made up of a student network and a teacher network, with the teacher network guiding the training of the student network. Generalization is improved by maintaining consistency in the predictions of the student network. We used a collection of cough recordings from COVID-19 patients and healthy people, with balanced labeled and unlabeled parts. Accuracy, loss, precision, recall, and AUC are all measured using k-fold cross-validation. The results suggest that semi-supervised COVID-19 identification can be successful. When labeled data is scarce, the findings highlight the value of cough recordings and semi-supervised learning.

Keywords: COVID-19, semi-supervised learning, Mean Teacher model

1 Introduction

The COVID-19 epidemic has created considerable obstacles to public health systems across the world. Effective COVID-19 containment and mitigation methods need timely and precise detection of patients. Traditional diagnostic procedures, such as polymerase chain reaction (PCR) testing, have scalability, cost, and turnaround time limits. As a result, there is an increasing interest in creating non-invasive, cost-effective, and quickly deployed alternative detection technologies.

There has been a flurry of study in recent years into the prospect of applying machine learning and signal processing approaches for COVID-19 detection [16, 14]. Coughing is a frequent sign of respiratory disorders, particularly COVID-19; hence one technique is to analyze cough recordings. Cough recordings can capture key auditory elements that provide discriminating information about the disease's existence.

In this study, we offer a novel approach for semi-supervised COVID-19 identification using cough recordings. Due to the lack of labeled data in many real-world contexts, semi-supervised learning approaches have gained popularity. To train a Mean Teacher model with consistency regularization, we use both labeled and unlabeled data. The Mean Teacher model comprises a student network and a teacher network, with the teacher network providing goal predictions to steer the student network's training. Consistency regularization motivates the student network to make consistent predictions on expanded versions of the same input, hence improving its generalization skills.

We use a dataset of cough recordings gathered from COVID-19 patients and healthy persons to assess the efficacy of our suggested technique. To achieve equal representation of the classes, the dataset's features are carefully separated into labeled and unlabeled sections, with stratified sampling. To get reliable and robust performance estimations, we use K-fold cross-validation.

The contributions of this study include the development of a novel semi-supervised learning approach for COVID-19 detection using cough recordings and an extensive evaluation of its performance. The results demonstrate the effectiveness of our proposed method in achieving high accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC). These findings highlight the potential of utilizing cough recordings as a non-invasive and accessible modality for COVID-19 detection and emphasize the importance of semi-supervised learning techniques in scenarios with limited labeled data.

2 Related Works

The widespread outbreak of coronavirus illness (COVID-19) has made identification and diagnosis difficult. In recent years, there has been an increase in interest in applying machine learning approaches for COVID-19 identification, notably through cough recording analysis [9, 13, 19].

Zhang et al. [20] presented a semi-supervised learning technique for detecting COVID-19 in chest CT images. Their technique performed well in distinguishing COVID-19 patients from healthy controls. Adhikari et al. [1] conducted a detailed scoping analysis of the epidemiology, clinical manifestations, and prevention of COVID-19 during the early epidemic phase. Their findings add to our knowledge of the importance of reliable COVID-19 detection methods. Tarvainen and Valpola's Mean Teacher Model [18] has exhibited increased performance in semi-supervised deep learning tasks. Their concept of weighted consistency objectives improves the learning process by utilizing unlabeled data. Furthermore, Lu et al. [10] suggested a semi-supervised audio classification approach with consistency-based regularization that may be used to detect COVID-19 from cough recordings.

Several articles concentrated on cough feature extraction approaches. Gupta et al. [6] investigated the use of Mel-frequency cepstral coefficients (MFCC) for feature extraction. Miron et al. [11] examined volumetric and slice-based techniques for COVID-19 identification in chest CT images. This research shed light on how to effectively characterize cough recordings for later categorization tasks. COVID-19 detection relies heavily on classification methods. Using cough samples, Imran et al. [7] built an AI-enabled preliminary diagnostic method for COVID-19. Convolutional neural networks (CNN) were used for classification based on audio MFCC characteristics. Chaudhari et al. [4] introduced the Virufy app, which uses crowdsourcing and clinical datasets for AI identification of COVID-19 from cough recordings.

Large-scale cough analysis datasets are critical for training and assessing detection systems. Orlandic et al. [12] introduced the COUGHVID dataset, which is a significant resource for evaluating large-scale cough analysis methods. In sound event categorization tasks, semi-supervised learning approaches have been shown to be successful. Rasmus and Valpola [15] proposed ladder networks that combine unsupervised pre-training with supervised fine-tuning. This method may be modified to improve COVID-19 detection from cough recordings. Other respiratory disorders have also benefited from cough sound analysis. Amrulloh et al. [2] studied cough sound analysis in pediatric populations for pneumonia and asthma categorization. Machine learning algorithms were used to categorize cough noises, with encouraging results. Crowdsourcing has been used to meet the need for large-scale datasets. Brown et al. [3] offered the COUGHVID crowdsourcing dataset, which provides a corpus for the research of large-scale cough analysis algorithms.

3 Materials & Methodology

3.1 Dataset

Cough audio signal categorization has been used successfully to identify several respiratory illnesses, and there is a lot of interest in using Machine Learning (ML) to enable universal COVID-19 screening. The COUGHVID [12] collection contains over 20,000 crowdsourced cough recordings from people of all ages, genders, geographic areas, and COVID-19 statuses. Furthermore, skilled pulmonologists labeled over 2,000 recordings to detect medical anomalies in the coughs, resulting in one of the biggest expert-labeled cough datasets available for application in a variety of cough audio classification tasks. As a result, the COUGHVID dataset [12] provides a plethora of cough recordings for training machine-learning models to solve the world's most pressing health concerns.

3.2 Computing requirements

The following computer setup was used for the experiments and analysis in this study:

1. **Hardware:**
   Processor: Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz (8 CPUs), 2.5GHz

   RAM: 16GB

   GPU: NVIDIA Tesla P100

2. **Software:**
   Programming Language: Python 3.9.13

   (a) Libraries and Packages:
       ffmpeg: Version 2023-05-25
       Librosa: Version 0.9.2
       TensorFlow: Version 2.10.1
       Other standard Python packages (e.g., numpy, matplotlib, etc.)

During the experimental and analysis stages, particular versions of the software packages specified above were utilized to assure repeatability and compatibility.

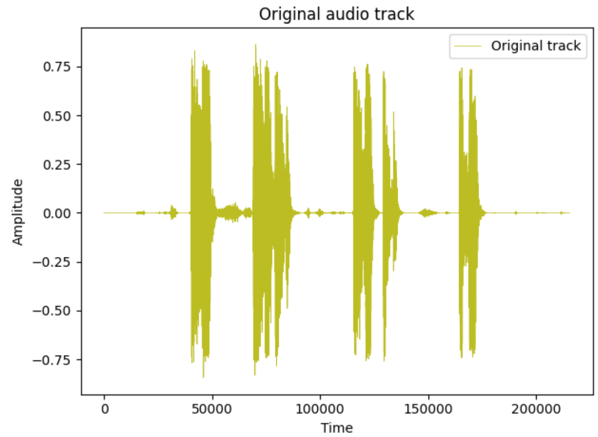
3.3 Data preprocessing

Several preprocessing processes were used to assure the quality and applicability of the cough recordings dataset [12] for training the Mean Teacher model with consistency regularization. These stages include noise reduction, silence removal, and audio file downsizing.

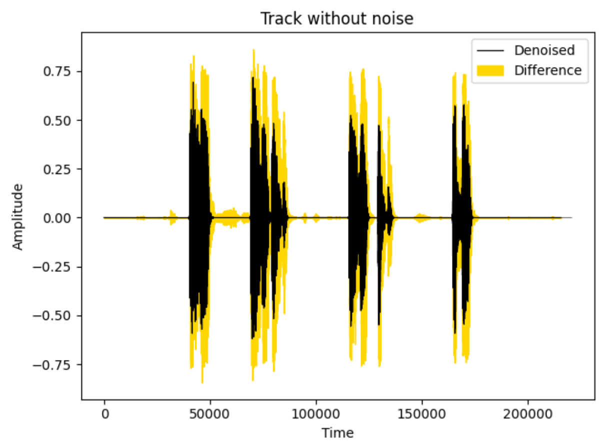
To begin, noise reduction was used to improve the quality of the cough recordings. This step aids in the removal of background noise and artifacts that may interfere with COVID-19 pattern identification.

Then, silence removal was used to remove parts of the audio recordings where there was no substantial cough signal. The elimination of silence helps to focus the model's attention on the essential cough segments, enhancing its capacity to recognize COVID-19 patterns.

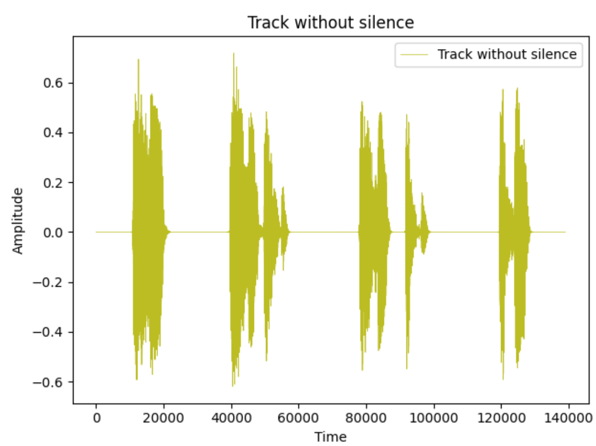
Finally, the audio files were resized to guarantee uniformity in the length of the cough recordings. This stage entailed altering the length of the cough recordings to a predetermined length. Resizing the audio files ensures that the input size is uniform during the training phase, allowing the model to learn from consistent temporal patterns. An example can be visualized in fig 1.



(a)



(b)



(c)

Fig. 1. Cough audio preprocessing, **a.** Original audio track, **b.** Track without noise where the black region signifies the denoised audio, **c.** Track without silence

3.4 Data Augmentation

Various strategies were used to supplement the cough recordings dataset [12] in the data augmentation section. Data augmentation is a typical method for artificially increasing the number and diversity of training data, which can aid in improving the model's generalization capacity. The following data enhancement methods were used:

Time stretching alters the speed of an audio transmission by either extending or compressing its duration. Variations in cough speed and length were added by randomly stretching the cough recordings, replicating different persons' coughing tendencies.

Pitch Shifting is the process of altering the pitch or frequency of an audio transmission without affecting its length. Random pitch shifting was performed on cough recordings, resulting in fluctuations in cough pitch comparable to those generated by people with varying vocal characteristics.

White noise injection is the process of adding random noise signals to cough recordings. The use of white noise makes the model more resistant to background noise and environmental fluctuations that may occur during the testing process.

These strategies boost the dataset's variety and diversity, allowing the model to learn from a greater range of cough patterns and improve its capacity to generalize to previously unreported data.

3.5 Feature extraction and feature selection

Mel-frequency cepstral coefficients (MFCC) are a typical form of feature used in speech recognition and other audio processing applications [5, 17]. MFCCs are retrieved from a speech signal's spectrogram, which is a visual depiction of the signal's frequency components across time [8].

Depending on the application, the amount of MFCCs extracted might vary. However, because the typical amount of MFCCs is 13, we choose to use 13 in our work as well. MFCCs are a versatile and powerful collection of properties that may be used in a variety of audio-processing applications. Because they are noise-resistant, discriminative, and compact, they are ideal for speech recognition applications.

3.6 Proposed Framework for COVID-19 Classification

We discuss the proposed framework for identifying COVID-19 utilizing cough recordings in this section. A semi-supervised learning strategy is used in the framework, which includes a Mean Teacher model with consistency regularization. To increase classification performance, the model uses both labeled and unlabeled data.

1. **Mean Teacher Model:** The Mean Teacher model consists of two neural networks: a student network and a teacher network. The architecture of both the models are same and can be viewed in fig 2.

```
Layer (type)                Output Shape              Param #
=================================================================
input_1 (InputLayer)       [(None, 422, 13)]         0

conv1d_1 (Conv1D)          (None, 418, 64)           4224

max_pooling1d_1 (MaxPoolin   (None, 139, 64)         0
g1D)

conv1d_2 (Conv1D)          (None, 135, 128)          41088

max_pooling1d_2 (MaxPoolin   (None, 45, 128)         0
g1D)

flatten (Flatten)          (None, 5760)              0

dense (Dense)              (None, 256)               1474816

dropout (Dropout)          (None, 256)               0

dense_1 (Dense)            (None, 128)               32896

dense_2 (Dense)            (None, 1)                 129

=================================================================
Total params: 1,553,153
Trainable params: 1,553,153
Non-trainable params: 0
```

Fig. 2. The Architecture of the Student-Teacher model

The student network is trained on labeled data with the goal of making accurate predictions.

The teacher network, which has the same weight as the student network, makes consistent predictions to help steer the student's learning.

The predictions of the teacher are produced by applying the trained student network to unlabeled data.

2. **Consistency Regularization:** It is used to ensure that the predictions of the student and teacher networks on unlabeled data are consistent.

   It promotes the model to generate comparable outputs for perturbed variants of the same input, resulting in better generalization.

   The consistency loss function computes the difference in predictions between the student and instructor networks.

3. **Implementation:** The features of the cough recordings are split into two parts: labeled data (Labels of the cough recordings) and unlabeled data (MFCC features).

   Unlabeled data is used for both student training and teacher supervision, whereas labeled data is used for supervised training of the student network.

   The Adam optimizer with a fixed learning rate is used to train the student network.

   Training is carried out with a batch size for a set number of epochs.

   Checkpoints in the model are used to store the optimal weights depending on validation accuracy.

   The trained student network is evaluated on both the validation set and the test set to assess its performance. The overall flow of our methodology can be visualized in fig 3.

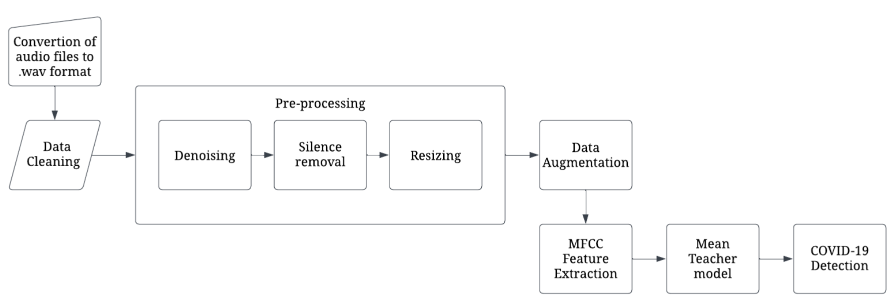


Fig. 3. Proposed methodology for COVID-19 detection

3.7 Performance Metrics

Accuracy : The proportion of correctly classified cough recordings, measuring the overall correctness of the model's predictions.

The formula for calculating accuracy is:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}} \tag{1}$$

Loss : The mean squared error between the predicted and true labels is used to calculate the consistency loss.

The formula for calculating the loss(MSE) is:

$$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}(\text{predicted label}_i - \text{true label}_i)^2 \tag{2}$$

where N is the total number of samples.

Precision : The proportion of true positive predictions (correctly identified COVID-19 cases) among all positive predictions (both true positives and false positives), indicating the model's ability to minimize false positives.

The formula for calculating precision is:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3}$$

Recall (Sensitivity) : The proportion of true positive predictions among all actual positive cases, measuring the model's ability to detect COVID-19 cases.

The formula for calculating recall is:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4}$$

AUC-ROC : (Area Under the Receiver Operating Characteristic Curve) It evaluates how well the model can differentiate between COVID-19 positive and negative cases based on the predictions made by the model. A higher AUC-ROC score indicates better performance, with values closer to 1 indicating a more accurate and reliable model.

The formula for calculating the AUC-ROC score is:

$$\text{AUC-ROC} = \int_{-\infty}^{\infty} \text{True Positive}(t) \cdot \text{False Positive}'(t)\, dt \tag{5}$$

4 Result & Discussion

We describe the results of our tests with the Mean Teacher model with consistency regularization for COVID-19 classification based on MFCC characteristics in this section. We next divided the dataset into 80:10:10 training, validating, and testing groups (see table 1) and ran 10-fold cross-validation to evaluate the model's performance, yielding the following results.

Table 1. Data-sets have been split for training, validation, and testing by 80:10:10 ratio, respectively

Dataset Type	Number of Data points	Number of features
Train	7,364	
Validation	920	422
Test	921	

4.1 Cross-Validation Results

We evaluated the model's performance using several measures such as accuracy, loss, precision, recall, and ROC AUC. To offer an overall sense of the model's performance, average values over all folds are presented in table 2.

Table 2. Average evaluation metrics for Cross-Validation results

Metric	Value
Accuracy	88.46%
Loss	0.68
Precision	87.39%
Recall	72.78%
ROC AUC	88.55%

The model had an overall accuracy of 88.46% across all folds, demonstrating that it could properly diagnose COVID-19 patients. The model's precision and recall were determined to be 87.39% and 72.78%, respectively, showing a reasonable balance between recognizing positive instances properly and minimizing false positives. The area under the receiver operating characteristic curve (AUC) was determined to be 88.55%, suggesting that the model has an excellent overall discriminative capacity. The validation set's learning curves in fig 4 & 5 showed a decrease in loss over epochs and a gain in accuracy, precision, and recall.

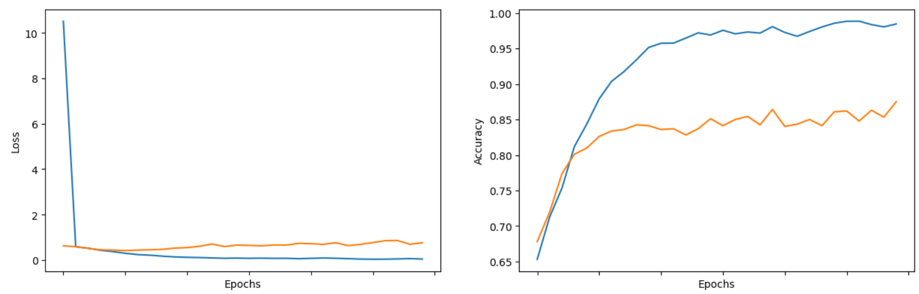


Fig. 4. Loss & Accuracy plot for one fold

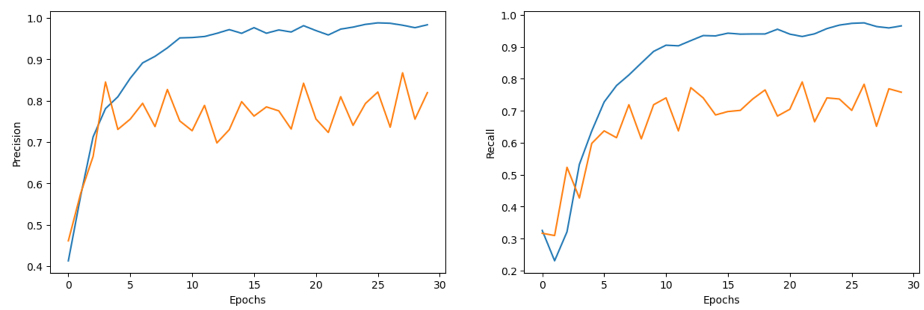


Fig. 5. Precision & Recall plot for one fold

4.2 Test Set Performance

The trained model was then tested on an independent test set to see how well it was generalized. The model achieved the following performance on the test set can be seen in table 3.

Table 3. Performance on test

Metric	Value
Accuracy	89.14%
Loss	0.56
Precision	90.22%
Recall	72.24%
ROC AUC	88.95%

The confusion matrix for the test set indicated the model's performance in terms of true positives, true negatives, false positives, and false negatives, proving its efficacy even further. The confusion matrix for the test set is presented in Figure 6. According to the test set's confusion matrix, the model correctly detected 618 negative instances and 203 positive cases, with 22 erroneous negatives and 78 false positives.

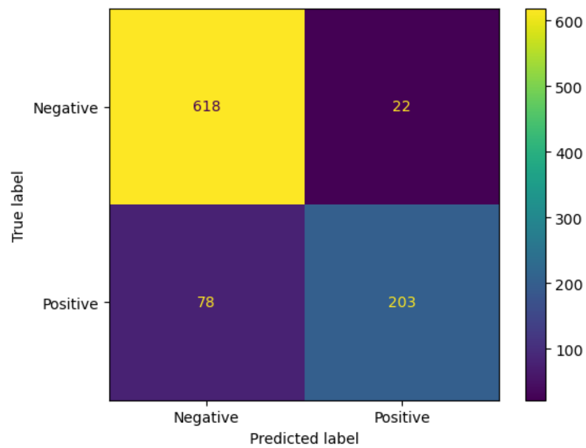


Fig. 6. Confusion matrix for the test set

The AUC score of 0.9 in fig 7 means a higher AUC indicates that the model can efficiently differentiate between positive and negative instances, which is critical for illness detection accuracy.

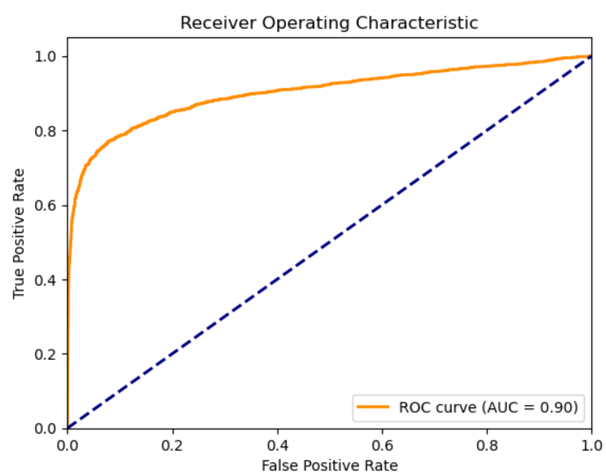


Fig. 7. ROC curve

Overall, the validation and test results show that the suggested Mean Teacher model with consistency regularization is effective in detecting COVID-19 using cough recordings. The model performs well in terms of loss reduction, accuracy, precision, recall, and AUC, making it a useful strategy for COVID-19 screening and diagnosis.

5 Conclusion

We developed a unique method for detecting COVID-19 using cough recordings in this work. The Mean Teacher model with consistency regularization, which we propose, takes advantage of the benefits of semi-supervised learning to efficiently categorize COVID-19 and non-COVID-19 patients. Our work makes significant contributions to data preparation techniques such as noise reduction, silence removal, and audio file resizing, which improve the quality and consistency of the input data.

We discovered promising results through thorough testing and assessment on a big dataset, as seen above in Result & Discussion section. These performance metrics demonstrate the ability of our proposed methodology to correctly detect COVID-19 instances from cough recordings.

Overall, our findings show that cough recordings have the potential to be a non-invasive and easily accessible technique for COVID-19 identification. The Mean Teacher model with consistency regularization suggested here offers a promising strategy for accurate and efficient screening, which can assist in the early identification and management of COVID-19 patients.

References

1. Adhikari, S. P., Meng, S., Wu, Y. J., Mao, Y. P., Ye, R. X., Wang, Q. Z., ... et al. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infectious diseases of poverty*, **9**(1), 1–12. https://doi.org/10.1186/s40249-020-00646-x
2. Amrulloh, Y., Abeyratne, U., Swarnkar, V., & Triasih, R. (2015). Cough sound analysis for pneumonia and asthma classification in pediatric population. In *2015 6th international conference on intelligent systems, Modelling and Simulation* (pp. 127–131). IEEE. https://doi.org/10.1109/ISMS.2015.41
3. Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., ... & Mascolo, C. (2020). Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. *arXiv preprint arXiv:2006.05919*. https://doi.org/10.1145/3394486.3412865
4. Chaudhari, G., Jiang, X., Fakhry, A., Han, A., Xiao, J., Shen, S., & Khanzada, A. (2020). Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. *arXiv preprint arXiv:2011.13320*. https://doi.org/10.48550/arXiv.2011.13320
5. Darshana, S., Theivaprakasham, H., Lal, G. J., Premjith, B., Sowmya, V., & Soman, K. P. (2022). MARS: A Hybrid Deep CNN-based Multi-Accent Recognition

System for English Language. In *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)* (pp. 1–6). IEEE.
6. Gupta, S., Jaafar, J., Ahmad, W. F., & Bansal, A. (2013). Feature extraction using MFCC. *Signal & Image Processing: An International Journal*, **4**(4), 101–108. https://doi.org/10.5121/sipij.2013.4408
7. Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, M. S., Ali, K., ... & Nabeel, M. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, **20**, 100378. https://doi.org/10.1016/j.imu.2020.100378
8. Kumar, C. S. A., Maharana, A. D., Krishnan, S. M., Hanuma, S. S. S., Lal, G. J., & Ravi, V. (2022). Speech Emotion Recognition Using CNN-LSTM and Vision Transformer. In *International Conference on Innovations in Bio-Inspired Computing and Applications* (pp. 86–97). Springer.
9. Kvapilova, L., Boza, V., Dubec, P., Majernik, M., Bogar, J., Jamison, J., ... & Karlin, D. R. (2020). Continuous sound collection using smartphones and machine learning to measure cough. *Digital Biomarkers*, **3**(3), 166–175. https://doi.org/10.1159/000504666
10. Lu, X., Foo, C. S., Teh, K. K., Tran, H. D., & Chandrasekhar, V. R. (2019). Semi-Supervised Audio Classification with Consistency-Based Regularization. In *INTERSPEECH* (pp. 3654–3658). https://doi.org/10.21437/Interspeech.2019-1231
11. Miron, R., Moisii, C., Dinu, S., & Breaban, M. E. (2023). Evaluating volumetric and slice-based approaches for COVID-19 detection in chest CTs. In *International Conference on Medical Imaging with Deep Learning* (pp. 383–392). PMLR. https://openaccess.thecvf.com/content/ICCV2021W/MIA-COV19D/papers/Miron_Evaluating_Volumetric_and_Slice-Based_Approaches_for_COVID-19_Detection_in_Chest_ICCVW_2021_paper.pdf
12. Orlandic, L., Teijeiro, T., & Atienza, D. (2021). The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific Data*, **8**(1), 156. https://doi.org/10.1038/s41597-021-00937-4
13. Pahar, M., Klopper, M., Warren, R., & Niesler, T. (2021). COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*, **135**, 104572. https://doi.org/10.1016/j.compbiomed.2021.104572
14. Radha, D., et al. (2021). Analysis of COVID-19 and pneumonia detection in chest X-ray images using deep learning. In *2021 International Conference on Communication, Control and Information Sciences (ICCISc)* (Vol. 1, pp. 1–6). IEEE.
15. Rasmus, A., Berglund, M., Honkala, M., Valpola, H., & Raiko, T. (2015). Semi-supervised learning with ladder networks. *Advances in neural information processing systems*, **28**. https://proceedings.neurips.cc/paper_files/paper/2015/file/378a063b8fdb1db941e34f4bde584c7d-Paper.pdf
16. Sathvik, P. G., Kumar, M. R., Neeli, G. H., Narasimha, I. Y., Singh, T., & Duraisamy, P. (2022). RESNET-50, CNN and HNN Medical Image Registration Techniques For Covid-19, Pneumonia and Other Chest Ailments Detection. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–7). IEEE.
17. Shraddha, S., Kumar, S., & others. (2022). Child Speech Recognition on End-to-End Neural ASR Models. In *2022 2nd International Conference on Intelligent Technologies (CONIT)* (pp. 1–6). IEEE.
18. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Ad-*

vances in neural information processing systems, **30**. https://proceedings.neurips.cc/paper_files/paper/2017/file/68053af2923e00204c3ca7c6a3150cf7-Paper.pdf
19. Valdés, J. J., Xi, P., Cohen-McFarlane, M., Wallace, B., Goubran, R., & Knoefel, F. (2021). Analysis of cough sound measurements including COVID-19 positive cases: A machine learning characterization. In *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 1–6). IEEE. https://doi.org/10.1109/MeMeA52024.2021.9478714
20. Zhang, Y., Su, L., Liu, Z., Tan, W., Jiang, Y., & Cheng, C. (2022). A semi-supervised learning approach for COVID-19 detection from chest CT scans. *Neurocomputing*, **503**, 314–324. https://doi.org/10.1016/j.neucom.2022.06.076