Detection of Cough from Telephonic Conversation via Artificial Intelligence to Predict the COVID Infections

R. KEERTHIYA
Ph.D Research Scholar
Department of Computer Applications
St. Peter's Institute of Higher Education and Research
Chennai, Tamil Nadu, India.

LATHA R
Head, Department of Computer Applications
St. Peter's Institute of Higher Education and Research
Chennai, Tamil Nadu, India.

***Abstract*---In this paper, an AI model was developed using machine learning algorithm that tends to predict the coughs from the telephonic conversations. The machine learning algorithm is trained with the sounds of coughs that enables the model to track the presence of covid in telephonic conversation records. The model is trained with the datasets and allowing the model to detect the cough sounds. The test is conducted with different set of conversation, thereby detecting the presence of asymptotic covid-19 infections in a novel way. The simulation is conducted in python environment to test the efficacy of the model against various recorded voice calls. The results of simulation shows that the proposed method achieves higher degree of accuracy in detecting the cough tones from the input datasets than other methods.**

Keywords---*Detection, Cough Detection, Artificial Intelligence, Covid Infection.*

I. INTRODUCTION

As of December 2020, more than 73 million cases of COVID-19 had been reported throughout the world [1]. Isolation of COVID-19-infected individuals, as well as a comprehensive range of tests, is required for infection control and resource optimization in the healthcare setting [2-3]. A direct connection between a patient and a practitioner is required for RT-PCR testing, which is now the gold standard. The turnaround time for this testing is variable, and it is both expensive and difficult to obtain for the majority of the world population. The availability of affordable and accessible testing is critical, especially given the continued rise in cases and the fact that vaccine approval and distribution are still pending.

Cough recordings from covid-19 have been obtained and analyzed by many research organizations in order to train machine learning algorithms for the identification of covid-19 virus. These models, on the other hand, have been trained utilizing data that has been collected in a variety of forms and recording types. Various other methods [10] exclusively collect cough recordings, in contrast [8], who also collects extra recordings of counting and vowels in addition to coughs. These datasets are also derived from a number of sources, including clinical settings, crowd sourcing, and extraction frominterviews conducted in the public media [9, 10]. Because recordings of COVID-19 cough data are compressed in a number of ways, there is currently no recognized standard for collecting COVID-19 cough data. Because datasets come from a diverse range of sources, contain a diverse range of contents, and are presented in a variety of formats, machine learning model developers face a significant challenge.

In order to ensure good performance in all locations and situations, a globally deployable machine learning solution must be trained and validated utilizing data from across the world [8]. This is the first time that a multimodal deep learning model has showed generalizability in detecting the presence or absence of COVID-19 in global cough datasets generated from a variety of sources, including both crowd sourced and clinically acquired information [14]-[20].

In this study, a machine learning technique was employed to develop an artificial intelligence model that can anticipate coughs based on phone calls. Cough noises are utilized to train the machine learning algorithm, which then allows the model to detect the presence of COVID in telephone conversations that have been captured on tape.

II. BACKGROUND

COVID-19 causes a wide range of symptoms that damage the human airway and lungs, despite the fact that the illness's clinical presentation can be extremely variable [4, 5]. Dry cough is the most prevalent symptom, which can be associated with other more serious symptoms [6]. In this approach, the sound of a cough may be used to detect the presence of a disease. With many audio features, machine learning algorithms have already exhibited excellent results in recognizing cough sounds [6]-[7].

Various research initiatives throughout the world have been the subject of machine learning models that are based on coughing. Several models [9, 10] have been developed to simulate coughing and other human audio sounds. Principal component analysis (PCA) is then performed on the features before they are fed into a simple binary classifier (SBC) [13]. Using the Mel-spectrogram as a model input is still another option to consider.

ResNet-18 [15] is a deep convolutional neural network that has been pre-trained, and ensembles of shallow and deep networks [14] are also being researched. The ResNet-18 network is in use. The researchers at MIT employed ResNet-50 [16]-based models to greatly improve the performance of their model [14]. They employed MFCC coefficients as input and transfer learning to build three separate biomarkers, which were then fed into three independent ResNet-50 convolutional neural networks in parallel [14] to achieve the best results.

As a result, based on the findings of research that has been published to date, it is hard to determine whether these models can reliably identify COVID-19 from arbitrary cough recordings. Using clinical data as training data, models have a high risk of failing to generalize. Transfer learning and pretraining have been proposed in the past to increase the size of the training dataset. This research demonstrates that our approach is capable of dealing with huge datasets from a variety of different sources.

III. PROPOSED METHOD

In this section, Deep Auto encoder (DAE) is used for the extraction of speech features after the pre-processing. The pre-processing involves removal of noise and empty signals from the input voice signals.

A. *Feature Extraction*

After feature extraction, the classification is conducted using the deep auto encoder output is of the same dimension as the input, it can be used to learn how to efficiently encode or represent the original data at hidden layers. It is important to note that the auto encoder is a feature extraction approach that does not rely on class labels to extract features. It is for this reason that the purpose of feature extraction is not to perform classification tasks, but rather to preserve information.

Back propagation is a technique that is commonly employed while training an auto-encoder. Using back-propagation to train networks with a large number of hidden layers has a number of intrinsic faults that must be taken into consideration. Unless the flaws have propagated beyond the initial few layers of the neural network, training will be rendered ineffective. This results in extremely slow learning and lousy answers, even when using more advanced approaches such as reinforcement learning. This problem can be addressed through the use of the DBN pretraining algorithm and other unsupervised pretraining techniques. Deep auto encoding has been employed in conjunction with other techniques to encode coughs, spectrogram-like speech characteristics, and images into a compact binary code for content-based image retrieval.

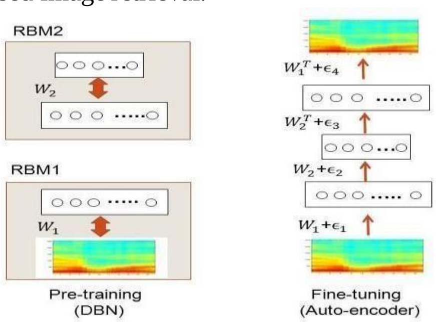

Fig 1: Deep Autoencoder Architecture for the Extraction of Speech Features

The discrete representations of this model in binary code can be used to obtain voice information or as bottleneck characteristics in speech recognition systems, depending on the application.

It is necessary to unroll the three hidden layers of the deep auto encoder in order to utilize the weight matrices of the DBN (Deep Belief Network). Encoding takes place in the bottom layers of this deep auto encoder, which employs matrices, while decoding takes place in the upper layers, which operate in the opposite direction. Through the use of error back- propagation, we are able to fine-tune the deep auto encoder depicted in Figure 1 in order to keep the reconstruction error as low as feasible.

Upon completion of the learning process, the following procedure can be used to encode and reconstruct any variable-length spectrogram that exists. Deep auto encoders are designed to take as input log power spectra from *N* consecutive overlapping frames that have been normalized using the zero-mean and unit-variance normalized log power spectrums. It then uses a logistic function to generate real-valued activations in the first hidden layer, which are subsequently processed by the second hidden layer. These real values are sent on to the following coding layer in order to generate codes, which are then converted into binary.

In the coding layer, the activations of hidden units are quantized so that they can be either 0 or 1, with 0.5 serving as a cutoff point. Last but not least, using the overlap-and-add technique, a full-length speech spectrogram is reconstructed by taking the outputs of the deep auto encoder applied to every accessible window of N consecutive frames and adding them together.

B. *Classification using RBM*

An RBM (Restricted Boltzmann Machine), or a specific Markov random field, is made up of one layer of stochastic hidden units and one layer of stochastic visible or observable units, which are both stochastic. There are no links between visible and hidden units in RBMs because all of the visible and hidden units are connected to the RBM network. RBMs can be represented as bipartite graphs, which are a type of graph.

RBMs are used to define the joint distribution $\mathrm{p}(\mathrm{v}, \mathrm{h}; \theta)$ over visible and hidden units v and h in terms of an energy function $\mathrm{E}(\mathrm{v}, \mathrm{h}; \theta)$ with model parameter θ.

$$\mathrm{p}(\mathrm{v}, \mathrm{h}; \theta) = \frac{exp\,(-\mathrm{E}(\mathrm{v}, \mathrm{h}; \theta))}{Z},$$

$$Z = \sum_{\mathrm{v}} \sum_{\mathrm{h}} exp(-\mathrm{E}(\mathrm{v}, \mathrm{h}; \theta))$$

where Z is either a normalization factor or a partition function, a model assigns a marginal probability to an observer visible vector v.

$$\mathrm{p}(\mathrm{v}; \theta) = \frac{\sum_h exp(-\mathrm{E}(\mathrm{v}, \mathrm{h}; \theta))}{Z}.$$

When considering a Bernoulli with visible and concealed RBM, the energy function is defined as

$$E(v,h;\theta) = -\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\, v_i h_j - \sum_{i=1}^{I} b_i v_i - \sum_{j=1}^{J} a_j h_j,$$

where
w_{ij} - interaction term between visible v_i and hidden unit h_j,
b_i and a_j - bias, and
I and J - visible and hidden units.
The study estimates the conditional probability is straightforward:

$$p(h_j = 1|v;\theta) = \sigma\left(\sum_{i=1}^{I} w_{ij}\, v_i + a_j\right),$$

$$p(v_i = 1|h;\theta) = \sigma\left(\sum_{j=1}^{J} w_{ij}\, h_j + b_i\right),$$

where $\sigma(x) = 1/(1+exp(x))$.
For a Gaussian (visible)-Bernoulli (hidden) RBM, where the energy is

$$E(v,h;\theta) = -\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\, v_i h_j - \frac{1}{2}\sum_{i=1}^{I}(v_i - b_i)^2 - \sum_{j=1}^{J} a_j h_j,$$

The conditional probabilities are given as below:

$$p(h_j = 1|v;\theta) = \sigma\left(\sum_{i=1}^{I} w_{ij}\, v_i + a_j\right),$$

$$p(v_i|h;\theta) = N\left(\sum_{j=1}^{J} w_{ij}\, h_j + b_i, 1\right),$$

where
v_i - real values and Gaussian distribution is followed with mean $\sum_{j=1}^{J} w_{ij}\, h_j + b_i$, where the variance is unity.

Real-valued and Bernoulli RBMs can transform real-valued random variables into binary random variables, and then Bernoulli and Gaussian RBMs can further process the binary random variables that have been converted.

The Gaussian conditional distribution (for continuous data) and the binomial conditional distribution (for non-continuous data) are two of the most widely used conditional distributions in RBM (for binary data). The RBM can also make use of distributions that are more generic in nature.

The gradient of the log likelihood log $p(v;\theta)$ can be used to generate the RBM weights update rule, which can be expressed as follows:

$$\Delta w_{ij} = E_{data}(v_i h_j) - E_{model}(v_i h_j),$$

where
$E_{data}(v_i h_j)$ - observed value in training set and
$E_{model}(v_i h_j)$ - expectation value under the model.
Figure 2 depicts samples acquired during the RBM learning procedure.

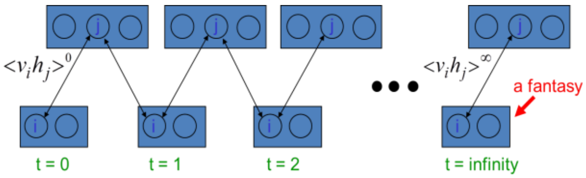

Fig 2: A Sampling during RBM Learning

RBMs and related deep learning technologies can only be applied to practical scenarios if RBMs have been extensively trained before they are employed. Using hidden variables to characterize the distribution of input data, the RBM described above can be used to characterize the distribution of input data without relying on any label information. When both the data and the label information are available, it is possible to generate a labeled data set. However, this is not always the case. As a result, it is conceivable to employ the same CD learning to maximize the approximate generative objective function associated with the data probability while maintaining the same CD learning. Furthermore, a discriminative objective function can be expressed in terms of the conditional probability associated with the label identity.

IV. RESULTS AND DISCUSSIONS

In this section, we validate the entire model on various input speech datasets that includes ESC-50 dataset for testing and initially it is trained with COUGHVID dataset [22]. A total of 80% of the ESC-50 dataset is used for training the classifier after it is trained with COUGHVID dataset and the remaining 20% is used for testing.

AI solutions are theoretically viable, but this does not ensure their practical viability, as the final output is dependent on both the quantity and quality of data as well as the level of sophistication of the machine learning algorithm used. The four classifications of cough, bronchitis, and covid-19 are used here to examine the practical viability of a cough-based covid-19 diagnosis solution.

The study employs the accuracy, specificity and sensitivity performance measures, as well as the F1-score on the validation set and cross-validation of the models, to evaluate it. The model's total accuracy is what is being measured here. It is possible to test the performance of machine learning models with limited data by utilizing the k-fold cross validation approach.

Using mean confusion matrices derived from cross-validation, we have calculated these performance indicators. Regularization techniques, such as DAE's regularization parameter, were also utilized to avoid the problem of over-fitting. Cross-validation accuracy is used to adjust the various hyper-parameters of deep neural network models.

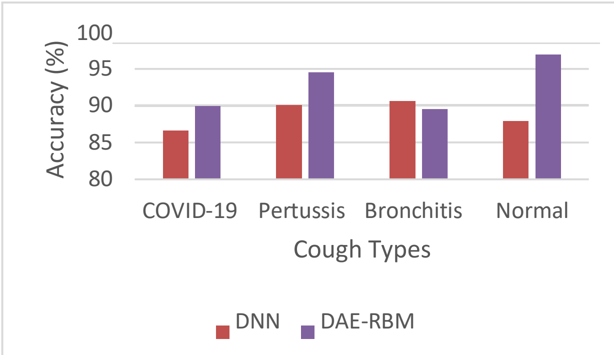

Fig 3: Accuracy

Figure 3 shows the accuracy of the proposed DAE-RBM with existing deep neural network (DNN) model. The result of simulation shows that the proposed method achieves higher degree of accuracy than other methods over all cough types from the input dataset.

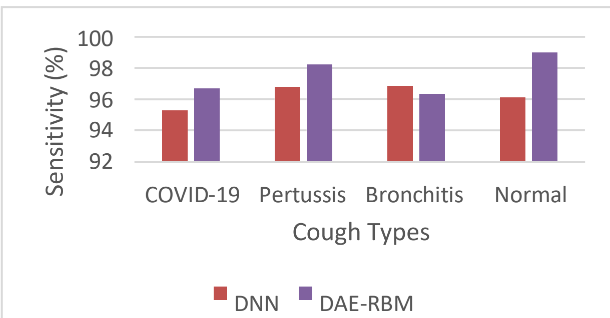

Fig 4: Sensitivity

Figure 4 shows the sensitivity of the proposed DAE-RBM with existing DNN model. The results of simulation shows that the proposed method achieves higher sensitivity than other methods over all cough types from the input dataset.

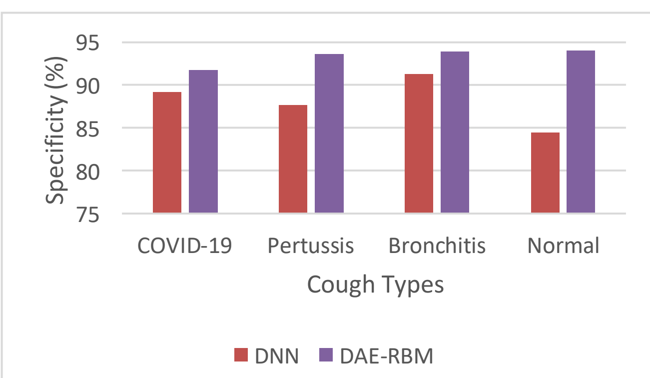

Fig 5: Specificity

Figure 5 shows the specificity of the proposed DAE-RBM with existing DNN model. The result of simulation shows that the proposed method achieves higher specificity than other methods over all cough types from the input dataset.

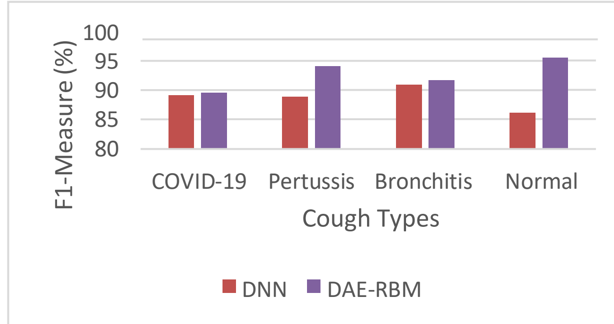

Fig 6: F1-Measure

Figure 6 shows the f1-measure of the proposed DAE-RBM with existing DNN model. The result of simulation shows that the proposed method achieves higher f1-measure than other methods over all cough types from the input dataset.

V. Conclusions

In this paper, an AI model was developed using machine learning algorithm that tends to predict the coughs from the telephonic conversations. The machine learning algorithm is trained with the sounds of coughs that enables the model to track the presence of covid in telephonic conversation records. The model is trained with the datasets and allowing the model to detect the cough sounds. The test is conducted with different set of conversation, thereby detecting the presence of asymptotic covid-19 infections in a novel way. The simulation is conducted in python environment to test the efficacy of the model against various recorded voice calls. The result of simulation shows that the proposed method achieves higher degree of accuracy in detecting the cough tones from the input datasets than other methods.

References

[1] 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 Audio Samples.

[2] Khadidos, A., Khadidos, A. O., Kannan, S., Natarajan, Y., Mohanty, S. N., & Tsaramirsis, G. (2020). Analysis of COVID-19 Infections on a CT Image Using DeepSense Model. Frontiers in Public Health, 8.

[3] Maheshwari, V., Mahmood, M. R., Sravanthi, S., Arivazhagan, N., ParimalaGandhi, A., Srihari, K., ... & Sundramurthy, V. P. (2021). Nanotechnology-Based Sensitive Biosensors for COVID-19 Prediction Using Fuzzy Logic Control. Journal of Nanomaterials, 2021.

[4] Szankin, M., Kwasniewska, A., Glowacka, N., Ruminski, J., Nicolas, R., & Gamba, D. (2021, July). Fully Automated AI-powered Contactless Cough Detection based on Pixel Value Dynamics Occurring within Facial Regions. In 2021 14th International Conference on Human System Interaction (HSI) (pp. 1-7). IEEE.

[5] 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), 1-10.

[6] Rao, S., Narayanaswamy, V., Esposito, M., Thiagarajan, J., & Spanias, A. (2021, July). Deep Learning with hyper-parameter tuning for COVID-19 Cough Detection. In 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-5). IEEE.

[7] Ahmed, T., Ahmed, M. Y., Rahman, M. M., Nemati, E., Islam, B., Vatanparvar, K., ... & Gao, J. A. (2020, October). Automated time

synchronization of cough events from multimodal sensors in mobile devices. In Proceedings of the 2020 International Conference on Multimodal Interaction (pp. 614-619).

[8] Coppock, H., Jones, L., Kiskin, I., & Schuller, B. (2021). COVID-19 detection from audio: seven grains of salt. The Lancet Digital Health, 3(9), e537-e538.

[9] Khanzada, A., Hegde, S., Sreeram, S., Bower, G., Wang, W., Mediratta, R. P., ... & Rameau, A. (2021). Challenges and opportunities in deploying COVID-19 cough AI systems. Journal of Voice.

[10] Mohammed, E. A., Keyhani, M., Sanati-Nezhad, A., Hejazi, S. H., & Far, B. H. (2021). An ensemble learning approach to digital corona virus preliminary screening from cough sounds. Scientific Reports, 11(1), 1-11.

[11] Petrović, N., & Kocić, Đ. (2021, September). IoT for COVID-19 Indoor Spread Prevention: Cough Detection, Air Quality Control and Contact Tracing. In 2021 IEEE 32nd International Conference on Microelectronics (MIEL) (pp. 297-300). IEEE.

[12] Bagad, P., Dalmia, A., Doshi, J., Nagrani, A., Bhamare, P., Mahale, A., ... & Panicker, R. (2020). Cough against COVID: Evidence of COVID-19 signature in cough sounds. arXiv preprint arXiv:2009.08790.

[13] Preethi, S. R., Revathi, A. R., & Murugan, M. (2020). Exploration of Cough Recognition Technologies Grounded on Sensors and Artificial Intelligence. In Internet of Medical Things for Smart Healthcare (pp. 193-214). Springer, Singapore.

[14] Coppock, H., Jones, L., Kiskin, I., & Schuller, B. (2021). Bias and privacy in AI's cough-based COVID-19 recognition–Authors' reply. The Lancet Digital Health, 3(12), e761.

[15] Chuma, E. L., & Iano, Y. (2020). A movement detection system using continuous-wave Doppler radar sensor and convolutional neural network to detect cough and other gestures. IEEE Sensors Journal, 21(3), 2921-2928.

[16] Deshpande, G., Batliner, A., & Schuller, B. W. (2022). AI-Based human audio processing for COVID-19: A comprehensive overview. Pattern recognition, 122, 108289.

[17] Wu, D. X., Chen, Q., Yao, K. H., Li, L., Shi, W., Ke, J. W., ... & Shen, K. L. (2019). Pertussis detection in children with cough of any duration. BMC pediatrics, 19(1), 1-9.

[18] Lella, K. K., & Pja, A. (2022). Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath. Alexandria Engineering Journal, 61(2), 1319-1334.

[19] Ponomarchuk, A., Burenko, I., Malkin, E., Nazarov, I., Kokh, V., Avetisian, M., & Zhukov, L. (2022). Project Achoo: a practical model and application for COVID-19 detection from recordings of breath, voice, and cough. IEEE Journal of Selected Topics in Signal Processing.

[20] Mouawad, P., Dubnov, T., & Dubnov, S. (2021). Robust Detection of COVID-19 in Cough Sounds. SN Computer Science, 2(1), 1-13.

[21] 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), 1-10.

[22] COVID-19, Open Research Dataset Challenge (CORD-19): Available online at: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge