Grenze International Journal of Engineering and Technology, January Issue

Updated Model Deep Learning Methods for Classification of Respiratory Diseases using Different Medical Imaging and Cough Sounds

Aswin. A[1] and Dr. G. Ramesh[2]
[1-2]Research Scholar, Research Supervisor, School of Computer Science, Takshashila University, Ongur, Tindivanam, Villupuram District, India

***Abstract*—The respiratory disease has caused irreparable damage to the world. In order to prevent the spread of pathogenic, it is necessary to identify infected people for quarantine and treatment. The use of artificial intelligence and deep learning approaches can lead to prevention and reduction of treatment costs. The purpose of this study is to create deep learning models in order to diagnose people with the respiratory disease through the sound of coughing. Background: This study aimed to propose an automatic prediction of respiratory disease like Cough using cough sound based on deep learning models algorithms. Result: With the data here it is collected (about 254 people) in different networks, we have reached acceptable accuracies. Conclusions: To the best of our knowledge, this study is the first to test the potential of deep learning methods for chronic cough investigation, which also shows a great advantage over existing algorithms for patient data clustering.**

***Keywords*— Coughing, Deep learning, Ensemble learning, Resnet50, Respiratory conditions, Cough, Classification, Diagnosis.**

I. INTRODUCTION

Rapid and self-testing technologies for COVID-19 are therefore more crucial than ever in order to confirm cases and stop the virus from spreading by implementing preventative measures like self-isolation and quarantine, among others.

The RTPCR test, also known as the swab test, is currently the primary COVID-19 diagnostic tool. However, it is a costly and invasive procedure that frequently necessitates taking the patient to a testing facility, which may not be practical in many situations due to the severity of the case, the patient's mobility, etc.

A. DL Model for Diagnosis of Chest Diseases using Cough Sounds

The work done to distinguish cough from cough sounds using a range of DL techniques is described in this section. Hemdan et al. [18] offer a hybrid architecture they call CR-19 for quickly identifying and diagnosing COVID-19 using a range of machine learning (ML) techniques that are produced from cough audio signals. They accomplish this by taking advantage of coughing sounds. The purpose of this architecture is to achieve the above described objective. The accuracy of this framework has significantly increased as a result of the application of ML techniques and the genetic algorithm. Their CR-19 framework has an accuracy level of 92.19 percent.

Grenze ID: 01.GIJET.11.1.21_1

Six different classifiers that have previously been trained were employed in the study [19] to the COVID-19 cough sounds were classified using a total of six different classifiers that had previously undergone training. Among these classifiers were GoogleNet, ResNet-18, ResNet-50, MobileNet-V2, and ResNet-101, in addition to Nas-Net-Mobile. They first transformed the audio data into a visual representation using the spectrogram method. These pre-trained models were then used to extract the sound's attributes and classify it as either COVID-19 or non-COVID-19 based on which category it belonged to. With an accuracy rate of 94.91 percent, the ResNet-18 outperforms other classifiers according to the data collected.
The primary goal is to provide an overview of all the significant algorithms that have been covered in the literature thus far and that provide a trustworthy level of accuracy when predicting a disease based just on cough sounds. Consequently, the researchers are encouraged to locate the facts with ease while working in the healthcare industry.

The following is a summary of our primary contributions to this work:

- Here, the mechanism of coughing is described, along with its distinguishing characteristics and the forms of cough that are thought to be indicative of particular diseases. We offer the rationale for characterizing subtle or nuanced cough features using CAD and ML technologies in order to improve diagnosis.
- Here, common respiratory illnesses are identified, cough kinds are characterized by duration, and cough characteristics are compared.
- In this case, it is offering the crucial steps required to create a strong ML/DL-based framework for effective and precise detection and diagnosis. This covers feature engineering, data collection, preprocessing, and model training.
- This article summarizes the most recent AI-based technologies that successfully identify, monitor, and diagnose various respiratory disorders by taking advantage of special latent cough traits.
- Here is a quick summary of the machine learning methods for diagnosis and detection that were trained on non-cough noises.

Types of coughs according to duration Cough can be broadly divided into three groups according to its duration: acute (less than three weeks), subacute (three to eight weeks), and chronic (more than eight weeks) [20]. This section provides a detailed explanation of the respiratory disorders are the primary cause of death in developing nations, while fewer people die from lower respiratory tract infections now than they did twenty years ago. Pneumonia and lower respiratory tract infections kill around three million children each year, making infants and young children especially vulnerable.

B. Anatomy of Respiratory System

As air moves through the airways and interacts with various respiratory system structures, these noises are created. As illustrated in Fig. 1, these noises are normally produced in the lungs, lower respiratory tract (trachea and bronchial tubes), and upper respiratory tract (nose and throat). A thorough description of the human respiratory system can be found in [5] and [1]. By analysing respiratory sounds, medical professionals and technicians can spot anomalies such fluid buildup, inflammation, or obstructions in the airways and decide on the best course of action.

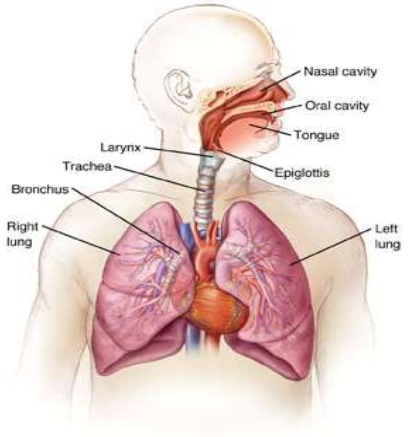

Fig 1. Anatomy of the Respiratory System

Coughing, breathing, and speaking are among the several respiratory sound kinds. A common sign of respiratory disorders, coughing is the body's way of protecting itself against inadvertently breathing in foreign objects or infection-produced substances [6]. Sounds associated with secretions in the airways initiate it. Allergies, respiratory infections, and irritants including dust, smoke, and other substances are common causes. Through the inhaling and exhale of air, breathing entails the respiratory system's exchange of carbon dioxide and oxygen. Exhaling is the release of air from the lungs, and inhaling is the drawing of atmospheric air into the lungs. But while abnormal noises like wheezing or crackling are different from normal breathing, normal respiratory sounds are clear and unhindered. Speech is produced by the controlled movement of air from the lungs through the vocal tract, which is facilitated by the vocal cords, lips, tongue, and jaw. The two categories of speech sounds are vowels and consonants. The vocal cords tremble when air passes through them during spoken speech. Unvoiced speech is speech that is made without the vibration of the vocal cords. As a result, random or periodic sound patterns are produced. Different vocal tract geometries are used to make vowel sounds. These sounds can be classified as either normal or aberrant respiratory sounds, which can reveal important details about a person's respiratory health. During a physical examination, a stethoscope can be used to hear these noises.

C. Organization of the Paper

The next Section II discusses the mechanism that produces cough and its distinctive features. Section III presents the types of cough based on duration, while Section IV describes the data acquisition methods. Section IV briefly discusses about the detection and diagnosis leveraging non-cough sounds. Different open research issues and future recommendations are outlined in Section V. Finally, we conclude the paper in Section VI.

II. MOTIVATION

Many researchers investigated bio-markers to identify COVID-19 based on biomedical signals such as chest X-rays, CT scans [15], and speech sounds. This work focuses on methods based on audio signals that comprise cough, breath, and vowel sounds. The development of smartphones has created new opportunities for automated sound-based respiratory health monitoring and diagnostics from the convenience of one's own home. Smartphones with built-in, high-quality microphones present a special chance for this. The use of machine learning and audio signal processing methods, which have attracted a lot of interest from researchers, is crucial to the success of such applications. Pitch, duration, intensity, Harmonic-to-Noise Ratio (HNR), jitter, and shimmer are examples of temporal and prosodic characteristics that are frequently used to identify unhealthy sounds. Numerous pertinent applications have made use of spectral properties obtained from the log Mel spectrogram. Therefore, HOS's ability to record coughing and breathing sound signals even when there is Gaussian noise serves as the primary driving force for the investigation of HOS-based methods for COVID-19 sound detection.

III. PROBLEM STATEMENT

The development of straightforward, affordable, and successful techniques for COVID-19 screening and monitoring utilising remotely recorded speech signals is desperately needed. There is currently little research in this field, which calls for more investigation and advancement.

- A number of issues must be resolved, such as the tiny percentage of cough data, the lack of objective information about health state, and the existence of biassed data, which makes it challenging to correctly identify false alarms.
- Most of the research on cough diagnostics that is now available focusses on aspects such as spectral-related features, Zero-Crossing Rate (ZCR), and Mel-frequency Cepstral Coefficients (MFCC). However, these algorithms usually perform worse when dealing with noisy data.
- Features that rely on MFCC and lower-order statistics miss crucial elements like phase information and signal nonlinearities. To improve the precision of cough detection techniques, these factors are essential and ought to be taken into account. Analysing respiratory sounds to identify cough sickness is the goal of this research work.

IV. OBJECTIVE

This study aims to address the urgent need for a straightforward and affordable testing procedure for COVID-19 diagnosis using remotely recorded speech signals. The creation of such approaches is essential for enhancing medical results and making effective illness screening and monitoring possible.

- To identify the biomarkers of the disease in the acoustics of these sounds.
- To provide a simplistic and cost-effective tool for diagnosing cough using breath, cough, and speech sounds.
- To employ the oversampling approach for achieving efficient training for class imbalance in the database and to facilitate to less complex deep learning model for the classification of respiratory sound signals.
- To explore the utility of HOS for capturing cough signal biomarkers
- To compare and analyze the detection performance of the proposed HOS-based methods with that obtained from other methods available in the literature.

V. Proposed Methods

This paper proposes a deep neural network-based algorithm that can quickly, remotely, and without any negative side effects identify the corona virus based just on coughing sounds.
Two unseen (used solely for testing) clinical and non-clinical datasets are utilised to test the system after it has been trained on publically accessible crowdsourced datasets.
According to experimental results, the suggested system performs better on datasets than the six well-known deep learning architectures because of its superior generalisation capability.
For test datasets that have not yet been viewed, the accuracy of the suggested approach is 92.4% in NoCoCoDa and 66.5% in Virufy.

VI. Methods

Here, the potential benefits of longitudinally modelling cough, breath, and voice biomarkers are examined, as well as the potential for precise and timely illness progression monitoring.

A. Data Set Preparation and Statistics

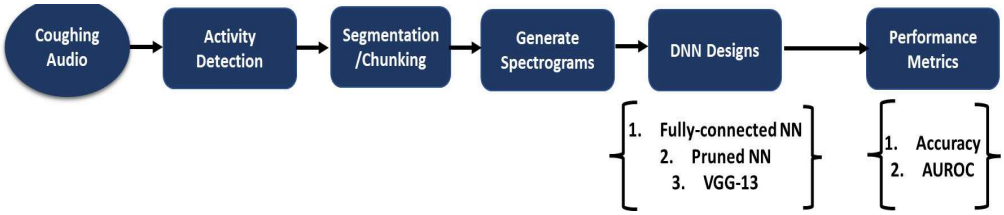

Fig. 2. A block diagram depicting an overview of the system

B. Data Processing

To effectively develop the deep learning model, we ensured a sufficient amount of appropriately processed data available for modeling by performing audio preprocessing, sequence generation, and data augmentation.

C. Audio Preprocessing

First, audio recordings were transformed to a mono channel and resampled to 16 kHz. After being normalised to a maximum amplitude of 1, these audio recordings were preprocessed by eliminating the silences at the start and finish. The real values of the coefficients that comprise a wave's Continuous Wavelet Transform (CWT) are represented graphically by the wave's scalogram [24]. Both of the measurements in this study are made using the scalogram technique. Together with the father signal, the scale parameter is used in CWT to calculate the cough sound signal [25].

$$CWT(a,b) = \frac{1}{|a|^{0.5}} \int_{-\infty}^{\infty} T(S)\, \theta\left(\frac{s-b}{a}\right) ds \text{ ----------(1)}$$

VII. Experimental Setup

The model that was proposed was constructed with open-source TensorFlow (TF)version (v) 2.12.0, whereas the seven DL models i.e., Vgg-19, ResNet-101, ResNet-50, DenseNet-121, Efficient NetB0, DenseNet-201, and inception-V3 are implemented with TF version(v) 1.8. Additionally, the Keras library was leveraged as the backbone for each of their implementations. Python language [23] is also employed in the construction of processes that are unrelated to the creation of convolution networks. The experiment was carried out on a workstation that utilized the Windows operating system and featured 32 gigabyts of RAM in addition to an 11-gigabyte graphics processing unit (GPU) from NVIDIA.

A. Enhancement of Preprocessing

The dataset includes images of eight distinct lung disorders as well as images of normal conditions. These images come in the form of CXRs, CT scans, and cough sounds. D_1 represents the collections of datasets, while d1(a) ∈ D_1, a = 1,2,3, 4, ,|D| = 600 represents each image inside those lung disease datasets. We have D_1 = [$d_1(1)$, $d_1(2)$,. . . ., $d_1(a)$, ..., $d_1(|D|)$]. The size of each CXR, CT scan, and CSI is [$d_1(a) = W_1 \times H_1 \times M_1$]. Here $W_1 = H_1 = 600$ *and* $M_1 = 3$, where *W* represents the width, *H* represents the height, and *M* shows the 3-channel RGB (red, green, and blue). Raw CXR, CT scans, and CSI cannot be used to train the proposed model or the baseline models because of the redundant information present in all three color channels, inconsistent contrast, and excessive sizes of the images.

First, we took the CXR, CT scan, and CSI and converted them to grayscale by keeping the luminance information solely. Using the following Eqs (2 and 3), gave us the grayscale image set D_2.

$D_2 = G(D2)$----------(2)

$$G(D2) = \{d_2(1), d_2(2), d_2(3), \ldots\ldots, d_2(a), \ldots\ldots, d_2(|D|)\} \text{ ----------(3)}$$

where *G* presents the grayscale process. Now the size of the image is $[d_2(a) = W_2 \times H_2 \times M_2]$. Here $W_2 = H_2 = 600$ *and* $M_2 = 1$.

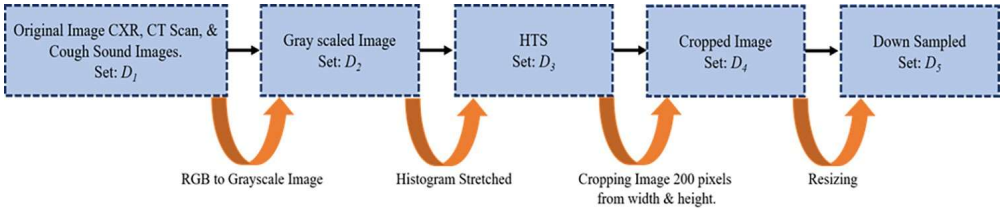

Fig 3. Enhancement of Preprocessing Techniques

The second approach histogram stretching (HTS) was utilized to improve the contrast of each CXR, CT scan, and CSI. For a_{th} image $d_2(a)$, $a = 1,2,3, 4,$..... , $|D|$, we begin by utilizing

Eqs (4 and 5) to get their minimum and maximum grayscale values, which are denoted by the notations $G_{min}(a)$ and $G_{max}(a)$, respectively

$$G_{min}(a) = min_{p=1}^{W_1} \times min_{p=2}^{H_1} \times d_2(a|p_1,p_1) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$G_{max}(a) = max_{p=1}^{W_1} \times max_{p=2}^{H_1} \times d_2(a|p_1,p_1) \text{ ----------(5)}$$

where (p_1, p_2) denotes the coordinates of a pixel in the CXR, CT scans, and CSI $d_2(a)$. Using Eq (6), $d_3(a)$ represents the new HTS image.

$$d_3(a) = \frac{d_2(a) - G_{min}(a)}{G_{max}(a) - G_{min}(a)} \text{ ----------(6)}$$

After that, we obtain the HTS image set $D_3 = HTS\ (C_2) = \{d_3(1), d_3(2), \ldots. d_3(a),\ldots.., d_3(|D|)\}$.

Third, the CXR, CT scans, and CSI were cropped to remove the text and patient information before training the proposed model. Thus, we get the cropped lung disease dataset D_4 using Eqs (7 & 8).

$$D_4 = R(D_3, [z_t, z_b, z_l, z_r]) \text{ ---------- (7)}$$

$$D_4 = \{d_4(1), d_4(2), d_4(3), \ldots\ldots, d_4(a), \ldots\ldots, d_4(|D|)\} \text{ ----------(8)}$$

where *R* represents the cropping process. The parameters z_t, z_b, z_l, z_r represents top, bottom, left, and right, respectively, values of the CXR, CT scans, and CSI. These parameters are used to crop the image in a unit of the pixel. Thus, we set $z_t = z_b = z_l = z_r = 200$. After this, the size of each image is $[D_4(a) = W_4 \times H_4 \times M_4]$. Now, we can have $W_4 = H_4 = 400$ and $M_4 = M_2 = 1$. where ↓: $O \longrightarrow I$ means the down sampling (DS) function. The parameter *I* is a down sampled CXR, CT scan, and CSI of original image *O*. For the present work, the images were down- sampled to the fixed size of resolution, $W_5 = H_5 = 299$, $M_5 = 1$. There are several advantages to DS, such as it can reduce storage space and a smaller- size dataset can assist in preventing the proposed classification system from overfitting. The approach of trial and error is the justification for why we decided to set $W_5 = H_5 =$

299. We find that a lower size would make the images blurry, which will also result in a decline in the classifier's performance. On the other hand, greater size will result in overfitting, which will hamper the performance.

B. Sequence Generation

Each participant's audio samples were divided into brief sequences of a set number of five samples in order to enhance the number of audio sequences available for model building. An additional restriction was put in place to limit the maximum time interval between two consecutive samples to 14 days in order to guarantee that the five samples in an audio sequence featured sufficient and effective audio dynamics in COVID-19 progression. Sequences that didn't follow this restriction were eliminated. As a result, sequence lengths that covered the course of the sickness ranged from 5 to 56 days.

C. Proposed model

The conventional approaches to DL produced remarkable results in illness diagnosis. The CNN is an innovative form of artificial neural network. The proposed CNN is made up of convolutional layers (ConvLs), pooling layers (PLs), non-linear activation methods (NLAMs), and fully connected layers (FCLs). The primary function of the proposed CNN model is to convolute information. Convolution in two dimensions, in the width and height directions, is executed by ConvLs [10]. It is important to note that proposed model weights start as random values and are later learned from the data itself through the process of network training. The proposed model takes three steps during a ConvLs operation: i) Kernel-based convolution (KBC); (ii) Stack; and (iii) NLAMs. The proposed model takes an input matrix I, kernels K_p, $\forall p \in [1,2,3\ldots, P]$, and an output O, (here O refers to the result of the full three- step convolution layer, as opposed to the result of a single convolution). A layer's ability to conduct convolution is denoted by the presence of ConvLs, and the phrase "complete convolu- tion layer" refers to the combination of ConvLs, a stack, and NLAMs. In addition, we have used the same color to symbolize both the input and the output of the ConvLs because the out- put would be utilized as the input for the ConvLs that come after it.

For each kernel K_p, the results of the convolution are calculated by using Eq (11).

$$f(p) = I \otimes K_p, \forall p \in [1,2,3, \ldots P] \text{ ----------(9)}$$

where "convolution operation" is denoted by$\otimes$. After that, each *f(p)* matrix is stacked with the others to form the three-dimensional matrix D by using Eq (12).

$$D=[f(1), f(2), f(3), \ldots\ldots, f(P)] \text{ ----------(10)}$$

Where P refers to the operation on the stack. Finally, matrix D is input into the NLAM, which then produces the final matrix. Eq (13) is used to measure this final matrix.

$$Y=NLAM(D) \text{ ----------(11)}$$

As demonstrated in Eq (14), we can compute the respective sizes (Z), of the three primary components (input, kernel, and output).

$$Z(s) = \begin{cases} W_I \times H_I \times M_I, s = I \\ W_K \times H_K \times M_K, s = K_P, \forall p \epsilon [1,2,3, \ldots\ldots, P] \\ W_Y \times H_Y \times M_Y, s = Y \end{cases} \text{ ----------(12)}$$

$$\begin{cases} W_Y = 1 + \frac{2 \times U_P + W_I - W_K}{U_s} \\ H_Y = 1 + \frac{2 \times U_P + H_I - H_K}{U_s} \\ M_Y = P \end{cases} \text{ ----------(13)}$$

$$NLAM_{ReLU}(f_{ab}) = ReLU(f_{ab})$$

$$ReLU(f_{ab}) = max(0, f_{ab}) \text{ ----------(14)}$$

$$\underbrace{\{L\_a, a = 1,2,3, \ldots\ldots, |I|\}}_{\dot{I}} \text{ I } \underbrace{\{B_a, a = 1,2,3, \ldots\ldots, |I|\}B}_{\dot{B}}$$

$$M_e = \frac{1}{|I|}\left(\sum_{a=1}^{|I|} L_a\right) \text{ ----------(15)}$$

$$V_e = \frac{1}{|I|}\left(\sum_{a=1}^{|I|} L_a\right)(L_a - M_e)^2 \text{ ----------(16)}$$

$$L_a = \frac{L_a - M_e}{\sqrt{V_e + d_s}} \text{ ----------(17)}$$

$$b_a = P_1 \times L_a + P_2, a = 1,2,3, \ldots, |I| \text{ ----------(18)}$$

$$b_a = P_1 \times \frac{(L_a - M_p)}{\sqrt{(V_p + \Delta)}} + P_2 \text{ ----------(19)}$$

$$L_{rp} = \frac{|R - N|}{N} \text{ ----------(20)}$$

D. Comparative Study

The outcomes of the five runs (R) using P(1) through P(4). P(1) is BM and P(2) contains the BANL and dropout layers. In P(3) we replace the MPL layer of P(2) with RBAP. Finally, P(4) is developed by adding the MWDG with P(3). The above values it very evident that the P(1) model resulted in the seven performances listed: $m^1 = 91.22 \pm 2.23$, $m^2 = 91.93 \pm 2.19$, $m^3 = 91.83 \pm 2.15$, $m^4 = 91.91 \pm 2.03$, $m^5 = 91.91 \pm 2.09$, $m^6 = 91.86 \pm 2.15$, and $m^7 = 91.91 \pm 2.10$.

TABLE I. RESULTS COMPARISON OF PROPOSED MODEL P (1) TO P (4) OVER VALIDATION SET OF 5 RUNS.

Models	Runs	m^1	m^2	m^3	m^4	m^5	m^6	m^7
P (1)	1	91.16%	91.85%	91.68%	91.78%	91.92%	91.64%	91.89%
	2	91.29%	91.89%	91.78%	91.89%	91.96%	91.81%	91.91%
	3	91.42%	91.91%	91.84%	91.99%	91.94%	91.83%	91.92%
	4	91.56%	91.94%	91.91%	91.89%	91.98%	91.88%	91.94%
	5	91.69%	91.96%	91.99%	91.85%	91.96%	91.91%	91.92%
	ME±SD	91.22±2.23	91.93±2.19	91.83±2.15	91.91±2.03	91.91±2.09	91.86±2.15	91.91±2.10

The detailed definitions of m^1 to m^7 are discussed in section 3.5. For P(2), the model enhanced the outcomes as $m^1 = 93.50 \pm 1.20$, $m^2 = 93.94 \pm 1.09$, $m^3 = 93.63 \pm 1.18$, $m^4 = 93.94 \pm 1.09$, $m^5 = 93.94 \pm 1.11$, $m^6 = 93.80 \pm 1.05$, and $m^7 = 93.98 \pm 1.14$. Additionally, the P (3) model produced the significant performance as $m^1 = 96.89 \pm 2.19$, $m^2 = 96.94 \pm 2.23$, $m^3 = 96.63 \pm 2.18$, $m^4 = 96.91 \pm 2.09$, $m^5 = 96.89 \pm 2.22$, $m^6 = 96.80 \pm 2.05$, and $m^7 = 96.84 \pm 2.14$. Comparing the results of seven evaluation metrics between P(2) contains the BANL and P(3) with RBAP. This study concluded that RBAP gives significantly better performance than using MPL in P(2). Finally, the P(4) model achieved the tremendous results as compared to P(1) to P (3) in terms of $m^1 = 99.01 \pm 0.97$, $m^2 = 98.98 \pm 1.01$, $m^3 = 99.63 \pm 0.18$, $m^4 = 99.01 \pm 0.97$, $m^5 98.99 \pm 0.99$, $m^6 = 98.98 \pm 1.01$, and $m^7 = 98.96 \pm 1.02$. The increase in performances of MWDG with P(4) compared to P(3) indicates that MWDG can help improve the performance of the model.

VIII. CONCLUSION

In conclusion, audio-based diagnostics with longitudinal data is suggested here as a reliable method for tracking the course of coughs by modelling audio biomarkers longitudinally using sequential machine learning algorithms. This demonstrates how our system can track the course of an illness, particularly an individual's recovery. In addition to offering a fast, cost-effective, and adaptable technique for tracking COVID-19, this work demonstrates the potential of telemonitoring for monitoring respiratory illnesses more broadly.

REFERENCES

[1] Menni C, Sudre CH, Steves CJ, Ourselin S, Spector TD (2020) Quantifying additional COVID-19 symptoms will save lives. Lancet 394(10241):e107–e108

[2] Rothan HA, Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 109:102433

[3] Larsen JR, Martin MR, Martin JD, Kuhn P, Hicks JB (2020) Modeling the onset of symptoms of COVID-19. Front Public Health 8:473

[4] Deshpande G, Batliner A, Schuller BW (2022) AI-based human audio processing for COVID-19: a comprehensive overview. Pattern Recogn 122:108289

[5] Kiziloluk S, Sert E (2022) COVID-CCD-Net: COVID-19 and colon cancer diagnosis system with optimized CNN hyperparameters using gradient-based optimizer. Med Biol Eng Comput :1–18

[6] Amyar A, Modzelewski R, Li H, Ruan S (2020) Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. Comput Biol Med 126:104037

[7] Wang S et al (2021) A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). Eur Radiol :1–9

[8] Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. Pattern Anal Appl :1–14

[9] Ilhan HO, Serbes G, Aydin N (2021) Decision and feature level fusion of deep features extracted from public COVID-19 data-sets. Appl Intell :1–21
[10] Brown C et al (2020) Exploring automatic diagnosis of COVID- 19 from crowdsourced respiratory sound data, 3474–3484. https://doi.org/10.1145/3394486.3412865
[11] Imran A et al (2020) AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 20:100378
[12] Mohammed EA, Keyhani M, Sanati-Nezhad A, Hejazi SH, Far BH (2021) An ensemble learning approach to digital corona virus preliminary screening from cough sounds. Sci Rep 11(1):1–11
[13] Xia T et al (2021) COVID-19 sounds: A large-scale audio dataset for digital respiratory screening
[14] Mallol-Ragolta A, Cuesta H, G´omez, E, Schuller BW (2021) Cough-based COVID-19 detection with contextual attention convolutional neural networks and gender information. In: 22nd Annual Conference of the international speech communication association, INTERSPEECH 2021, pp 4236–4240
[15] Dash TK, Mishra S, Panda G, Satapathy SC (2021) Detection of COVID-19 from speech signal using bio-inspired based cepstral features. Pattern Recogn 117:107999
[16] Laguarta J, Hueto F, Subirana B (2020) COVID-19 artificial intelligence diagnosis using only cough recordings. IEEE Open J Eng Med Biol 1:275–281
[17] Soltanian M, Borna K (2022) COVID-19 recognition from cough sounds using lightweight separable-quadratic convolutional network. Biomed Sig Process Control 72:103333
[18] Coppock H et al (2021) End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. BMJ Innovations 7(2)
[19] Suppakitjanusant P et al (2021) Identifying individuals with recent COVID-19 through voice classification using deep learning. Sci Rep 11(1):1–7
[20] Chaudhari G et al (2020) Virufy: global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. arXiv:2011.13320
[21] Akg¨un, D, Kabakus¸, AT, S¸ent¨urk, ZK, S¸ent¨urk, A, K¨uc¸ ¨ukk¨ulahlı, E (2021) A transfer learning-based deep learning approach for automated COVID-19 diagnosis with audio data. Turk J Electr Eng Comput Sci 29(8):2807–2823
[22] Shuja J, Alanazi E, Alasmary W, Alashaikh A (2021) COVID- 19 open source data sets: a comprehensive survey. Appl Intell 51(3):1296–1325
[23] Davis S, Mermelstein P (1980) Comparison of parametric representations formonosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Sig Process 28(4):357–366
[24] Sahidullah M, Saha G (2012) Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Comm 54(4):543–565
[25] Ghahramani P, Hadian H, Povey D, Hermansky H, Khudanpur S (2020) An Alternative to MFCCs for ASR,1664–1667. https://doi.org/10.21437/Interspeech.2020-2691
[26] McFee B et al (2015) librosa: Audio and music signal analysis in python. In: Proceedings of the 14th python in science conference, vol 8, pp 18–25
[27] Arias-Vergara T et al (2021) Multi-channel spectrograms for speech processing applications using deep learning methods. Pattern Anal Appl 24(2):423–431
[28] Meghanani A, Anoop CS, Ramakrishnan AG (2021) An exploration of log-mel spectrogram andMFCC features for Alzheimer's dementia recognition from spontaneous speech. In: 2021 IEEE spoken language technology workshop (SLT), pp 670–677
[29] Yu Y-B et al (2021) Attentive deep CNN for speaker verification. In: 12th international conference on signal processing systems, vol 11719. International Society for Optics and Photonics, p 117191U
[30] Chen X, Zahorian SA (2021) Improving speaker verification in reverberant environments. In: ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing
[31] (ICASSP). IEEE, pp 5854–5858
[32] Michelsanti D et al (2021) An overview of deep-learning-based audio-visual speech enhancement and separation. IEEE/ACM Trans Audio Speech Lang Process 29:1368–1396. https://doi.org/10.1109/TASLP.2021.3066303
[33] Hasannezhad M, Yu H, Zhu W-P, Champagne B (2022) PACDNN: a phase-aware composite deep neural network for speech enhancement. Speech Comm 136:1–13
[34] Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Bengio Y, LeCun Y (eds) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015 Conference Track Proceedings. http://arxiv.org/abs/1409.1556
[35] Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105