Applying Convolutional Neural Network and Transfer Learning to Mel-Spectrograms for COVID-19 Diagnosis via Cough Sounds.

Mrs Faiza.M.Saleh[*], Ms Raja.A.M.Moftah[**], Dr Kenz A. Bozed [**]

University of Benghazi

DOI: 10.29322/IJSRP.13.06.2023.p13805
http://dx.doi.org/10.29322/IJSRP.13.06.2023.p13805

Paper Received Date: 13th April 2023
Paper Acceptance Date: 20th May 2023
Paper Publication Date: 6th June 2023

Abstract- COVID-19, an extremely contagious disease caused by the coronavirus, has led to a global pandemic. Among its prominent symptoms is a persistent cough. Previous studies have suggested that cough sounds carry specific signatures associated with COVID-19. This study highlights the potential of utilizing AI models, specifically deep learning techniques, by collecting a large amount of audio data and divided into training, transfer learning, and testing datasets. The data is pre-processed to ensure a consistent length and transformed into Mel-spectrograms. A CNN model is trained using open-source datasets with different categories they are wind, breath, and cough to create a pre-trained model. Deep transfer learning is then applied by adding new layers to the pre-trained model and adapting it to the new task of binary classification for COVID-19 (positive or negative). Finally, the performance of the model is evaluated using a separate test dataset. The NoCoCoDa dataset, curated by M. Cohen-McFarlane, proved to be a valuable resource in this study. It offered diverse coughing samples and included an additional 73 positive cough sound records and 100 randomly selected negative cough sound records from other datasets to evaluate the model's effectiveness. The proposed model achieves impressive results and exhibits an area under the curve (AUC) value of 0.98 on the previously unseen test set. and illustrates that the model achieved an accuracy of approximately 98%, with a precision of 95% for negative samples and 100% for positive samples. These results highlight the model's high accuracy and precision in identifying both positive and negative cases of COVID-19 based on cough sounds.

Index Terms- COVID-19 diagnosis, deep transfer learning, Convolutional Neural Networks, Mel-spectrogram, cough sounds.

I. Introduction

A highly contagious illness called COVID-19 is brought on by a coronavirus. One of the most common symptoms of the virus is a persistent cough, and as of early 2021, there were over 100 million cases of COVID-19 and millions of deaths worldwide [1]. This global pandemic has made it imperative to develop rapid, accurate, and scalable diagnostic tools [2]. One promising tool is the analysis of cough sounds, which has demonstrated potential in diagnosing respiratory diseases [3]. Recent studies have posited that cough sounds may carry specific signatures associated with COVID-19, thereby providing a possible method for preliminary screening [4]. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown high efficiency in processing and classifying audio data in general and covid-19 audio data(cough and breath) in specific such as [5], presents an artificial intelligence (AI) model designed to help diagnose COVID-19 from cough samples, and their model was able to distinguish COVID-19 coughs from other types of coughs with an accuracy of more than 90%.in another study [6], the authors developed an AI framework that uses cough recordings to screen for COVID-19. The cough recordings were transformed into Mel-frequency cepstral coefficients (MFCCs), a common feature used in speech and audio processing, to be used as input for AI models. The model achieved 98% of accuracy. Other studies used Transfer learning using a pre-trained model (most of them using a model available online such as VGG19, ResNet50, DenseNet, MobileNet, etc.) as a starting point and then fine-tuning it on a specific task. It's a powerful technique that can often achieve high performance with less data than training a model from scratch. In the study [6], This research proposes models that use various pre-trained deep neural networks, including VGG19, ResNet50V2, DenseNet121, and MobileNet, for diagnosing COVID-19 based on cough sounds. These models were evaluated using the Cambridge dataset, The results showed that the model that employed MobileNet through transfer learning provided the best accuracy (86.42%). And in [7], their approach utilizes transfer learning by leveraging knowledge from a pre-trained model called VGGish. then fine-tuning the VGGish model specifically for COVID-19 detection using the COUGHVID dataset. And the study results are the experimental assessment of the proposed approach shows unweighted average recalls of 72.3% and 69.3% on the development and test sets, respectively. we propose an AI model for covid-19 diagnosis using Convolutional Neural Network (CNN), deep transfer learning and a Mel-spectrogram. First, we collect a big amount of audio data to train, test and applied deep transfer learning for the proposed model, these data were divided into many sections training dataset, transfer learning dataset, and testing dataset. Then, we pre-processed the data to

ensure the consistency of data length and convert it into a Mel-spectrogram, after that, build a CNN model was specifically trained by a huge open-source dataset with different categories (wind, breath, and cough) to create a pre-trained model, then we used this model to apply deep transfer learning by adding new layers and adapt it to the new task(binary classification positive covid-19 or negative covid-19), finally, Evaluated the model performance by a test dataset.

II. METHODOLOGY

1.1 Data Collection and Pre-Processing

Data plays a crucial role in the success of any AI model, and for training deep learning models, a large dataset is necessary to achieve good results. However, when it comes to COVID-19 cough sound records, the availability of datasets, especially positive coughs from COVID-19 patients, is limited. Therefore, multiple datasets from different sources were utilized in this project to compensate for this limitation. To ensure the quality and reliability of the collected data from various resources, data pre-processing was performed. This step involved converting all audio formats to the WAV format. The dataset was then divided into two folders: one containing positive COVID-19 cough samples and the other containing negative COVID-19 cough samples. The positive and negative folders were balanced, meaning they had the same size and nearly the same number of samples. To handle audio files with large durations, a segmentation process was conducted, splitting them into homogeneous parts, where each part contained 2 seconds of audio. For example, a file originally lasting 4 seconds would be split into two files, each consisting of 2 seconds. All audio files used in this study were converted to the 32-bit format with a sampling rate of 48.0 kHz and a fixed length of 2 seconds. The next section will discuss the conversion of audio samples into Mel spectrograms. Silence removal techniques were applied to all datasets in two ways: removing silence from the beginning and end of records and removing silence from the entire audio signals. However, neither approach resulted in significant improvements in the results. The data utilized in this project was divided into three parts, as depicted in Figure 1. These datasets were obtained from various open-source datasets. Detailed descriptions of these datasets will be provided in the following sections.

1.2 Dataset for Training

The CNN model was trained using a large dataset consisting of three categories of sound records: wind sound, breath sound, and cough sound. These records were obtained from various online sources, ensuring a diverse and comprehensive training dataset. During this stage, the model was trained using a total duration of 5 hours and 3 minutes of audio recordings.

Table 1 explains the three sound categories that are used in a training phase.

Data set	Items number	Resource link
Wind dataset	999	https://zenodo.org/record/5136592 - .Y5tDm-xBw-R
Cough dataset	937	296 https://zenodo.org/record/5136592#.ZCPehexBw-S 641 https://github.com/covid19-cough/dataset
Breath dataset	920	https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database

1.3 Dataset for Deep Transfer Learning

In accordance with the work of [8], transfer learning proves to be advantageous when the new task involves limited data or when collecting a significant amount of data is prohibitively costly. In this project, a separate dataset was utilized specifically for deep transfer learning. This dataset comprised 50 minutes (533.9 megabytes) of cough recordings, which were further divided into 26 minutes of negative cough records (282.8 megabytes) and 23 minutes of positive cough records (251.1 megabytes). It is important to note that obtaining several positive cough samples posed a challenge due to their limited availability. To address this issue, positive cough samples were sourced from multiple datasets. All the mentioned datasets are openly accessible online, as illustrated in Table 2. The first dataset, known as the "cleaned dataset," was created by compiling samples obtained from highly reliable sources such as COSWARA and Virufy. The second dataset comprises a substantial number of positive and negative cough samples; however, in this project, only positive samples were utilized.

Table2: explains the datasets used in the deep transfer learning phase.

Dataset	Positive covid-19	Negative covid-19	resource
Cleaned dataset	193 (82.1 MB)	469 (280.9 MB)	https://www.kaggle.com/datasets/pranaynandan63/covid-19-cough-sounds
Raw dataset	681 (169 MB)	0 (0 MB)	https://github.com/covid19-cough/dataset

1.4 Dataset for Testing

In this stage, the project utilizes the NoCoCoDa dataset [9], which has been curated to provide reliable and comprehensive data for researchers and policymakers to better understand the virus and its impact on public health. The dataset consists of audio samples specifically collected to represent real-world coughing sounds, utilizing various microphones, and recording devices. Moreover, the data collection process ensured diversity among the participants, incorporating different demographic groups. This diversity makes the NoCoCoDa dataset an excellent resource for researchers interested in developing algorithms for the accurate detection and classification of coughing sounds. M. Cohen-McFarlane, the creator of the NoCoCoDa dataset, has included 73 positive cough sound records and randomly selected 100 negative

cough sound records from various datasets for testing the model's performance.

1.5 Mel Spectrogram and Mel scale

In this project, one of the crucial data pre-processing stages involves converting audio signals into 2D images using Mel-spectrograms and the Mel scale, as illustrated in Figure 2. Mel-spectrograms and Mel scale are highly effective representations for audio data. These representations treat the results of the transformations as images, which can be used as inputs for convolutional neural networks (CNNs) or other deep learning models. A Mel-spectrogram provides a two-dimensional depiction of an audio signal, showcasing how its spectral content evolves over time. It is derived from the short-time Fourier transform (STFT) of the audio signal and undergoes further processing to mimic the human auditory system's frequency perception [10]. Mel-spectrograms are chosen in this project as effective representations due to their compatibility with deep learning techniques.

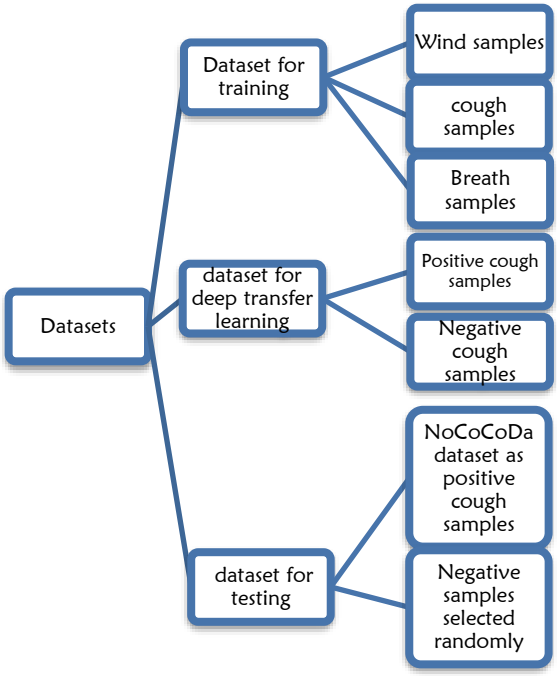

Figure 1: The project dataset parts.

The subsequent step involves converting the Mel-spectrograms into the Mel scale. The Mel scale is a perceptual scale of pitches that closely approximates the human ear's frequency perception. It is designed to account for the non-linear nature of human pitch perception, where the perceived distance between two frequencies varies at different frequency levels. The representations align more closely with human auditory perception by converting the Mel-spectrograms into the Mel scale.

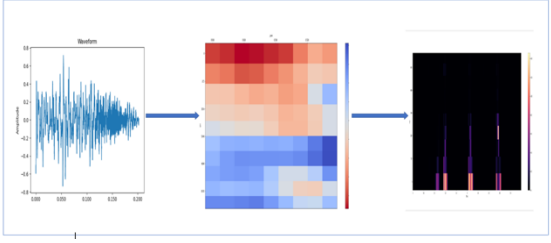

Figure 2: Sample record from the project dataset convert from audio signal to Mel spectrogram to Mel scale by Librosa python library.

1.6 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a type of artificial neural network that is commonly used for many tasks.

The main building block of a CNN is the convolutional layer, which applies a set of learnable filters to the input data. And results in a new feature map that highlights certain patterns in the input data. By stacking multiple convolutional layers on top of each other, CNNs can learn increasingly complex and abstract representations of the input.

CNNs also include pooling layers, which downs ample the feature maps by taking the maximum or average value.
Finally, CNNs often end with one or more fully connected layers, which take the output from the convolutional and pooling layers and perform classification or regression tasks based on the learned features [11].

An overview of the used CNN structure is shown in Fig 3, convolutional neural network architecture with four convolutional layers and three fully connected layers (Linear layers), designed for deep_audio_features task [12].

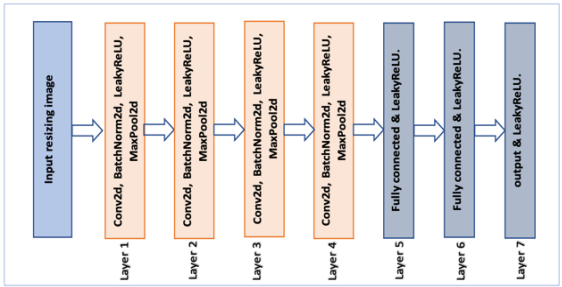

Figure 3: The architecture of the project CNN: Convolution, max-pooling, Batch Normalization, and fully connected layers.

Below is a high-level description of the convolutional layers' architecture:

The first layer takes a grayscale image with one channel as input, applies 32 filters of size 5x5 pixels to produce 32 output channels, performs batch normalization to normalize the output values, applies a LeakyReLU activation function with a negative slope of 0.01, and then applies max pooling with a kernel size of 2x2 pixels and stride of 2.

The subsequent three layers follow the same structure as first layer, but with increasing numbers of filters and output channels. Specifically, the second layer takes 32-channel output from the first layer and applies 64 filters, the third layer takes 64-channel output from second layer and applies 128 filters, and the fourth layer takes 128-channel output from the third layer and applies 256 filters.

After the four convolutional layers, the output is flattened and fed into three fully connected layers. The first linear layer applies a dropout rate of 0.75 to prevent overfitting, followed by a fully connected layer with 1024 output neurons and a LeakyReLU activation function. The second linear layer applies a dropout rate of 0.5, followed by a fully connected layer with 256 output neurons and a LeakyReLU activation function. Finally, the third

linear layer is a fully connected layer with editable output neurons and a LeakyReLU activation function.

The editable output neurons dimension can calculate by the number of the dataset classes number, in this project the first use of the CNN the output dimension was three classes wind, breath, and cough, and the second time when using the pre-trained CNN (transfer learning approach) the output dimension was two classes, positive Covid-19 cough, or negative Covid-19 cough.

1.7 Loss function

CrossEntropyLoss function is used in this project, and it's commonly used in machine learning, particularly for classification tasks. It is designed to measure the dissimilarity between two probability distributions, typically the ground-truth distribution and the predicted distribution. In essence, CrossEntropyLoss quantifies how well the model's predictions match the true labels. Mathematically, CrossEntropyLoss can be defined as follows:

Loss = - (1/N) * Σ(Σ(y_ij * log(p_ij))) **(1)**

As shown in (1), N is the number of data, y_ij is the truth label, and p_ij is the predicted label.

1.8 Deep Transfer learning

Deep transfer learning is a powerful machine learning technique that involves utilizing pre-trained deep learning models to address a new task that is related to the original task for which the model was initially trained. The underlying concept of transfer learning is to leverage the knowledge and learned representations of the pre-trained model, which was trained on a large dataset, and apply it to a new task that may have a smaller dataset [13]. In this project, the model is initially trained on three distinct categories: wind, breath, and cough, as described in Section 2.1. This initial training results in a well-pre-trained model capable of effectively classifying these three classes. Subsequently, a process called fine-tuning is performed, which involves making slight adjustments to the pre-trained model using a new dataset and a new task, aimed at enhancing its performance [14]. In other words, fine-tuning enables the model to learn task-specific features that may not have been present in the original dataset on which the pre-trained model was trained [13]. This step refines the model's capabilities, allowing it to adapt and improve its predictions for the new task at hand.

2. results

2.1 Experiment Approach

The CNN was trained by an AdamW optimizer, whose learning rate is 0.002. The max epoch number is 500 and the batch size was 5 with the early stop technique to stop the training when obtaining the highest accuracy, respectively. The CNN was implemented by Pytorch as a popular machine intelligence library that has gained significant attention in recent years, especially for convolutional neural network (CNN) models. This library provides a flexible and efficient platform for building and training deep learning models [15]. The model was also trained and tested on MacBook Air 2020 Apple M1 Chip CPU & GPU, 8 GB memory chips (DDR4).

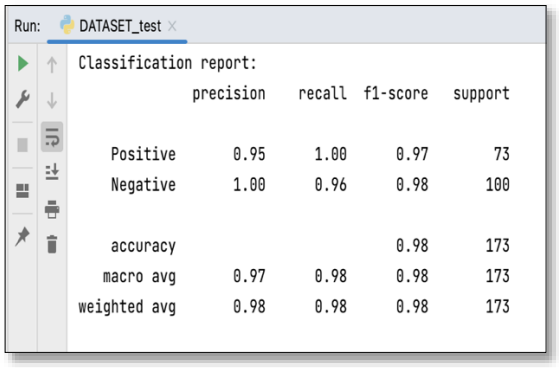

Figure 4: Screenshot from the project dataset test.

2.2 Performance Measurements

AI performance measurements are essential for evaluating the effectiveness and efficiency of AI models, systems, and algorithms. In this project these measurements are applied by **sklearn.metrics** as a module in the Scikit-Learn library in Python. **sklearn.metrics** contains a lot of functions for evaluating predictive models**Invalid source specified.**, such as

Classification metrics: it's a common function includes:

accuracy: It computes the accuracy classification score.

Precision: It computes the precision of the classification, i.e., the ratio of true positive predictions to the total positive predictions.

Recall: It computes the recall of the classification, i.e., the ratio of true positive predictions to the total actual positives.

f1_score: It computes the F1 score, which is the harmonic mean of precision and recall.

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\textbf{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\textbf{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\textbf{F1\_Score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Recall}+\text{Percision}} \quad (5)$$

where TP (True Positive) denotes positive cough samples that are correctly classified by the model. FP (False Positive) denotes positive cough samples that are incorrectly classified by the model. TN (True Negative) Negative cough samples that are correctly classified by the model. FN (False Negative) Negative cough samples that are incorrectly classified by the model.

As shown in Figure (5), TP equals 73 of 73 positive samples, and the FP is Zero, on the other side, the TN equals 96 and FN is only 4.

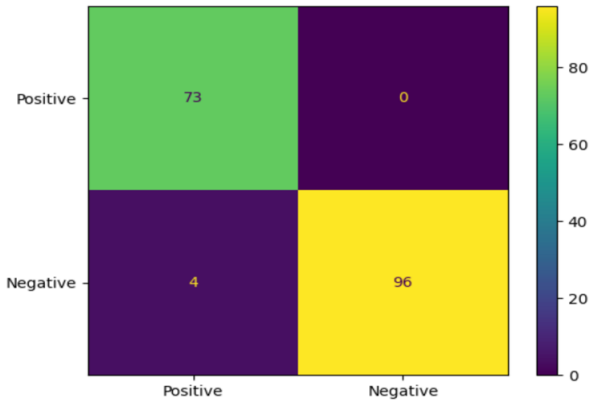

Figure 5: confusion matrix of the proposed model result

As a result of the proposed model, the ROC curve is shown in Figure (6) below, with an AUC of 0.98 on the test set had never been seen before, and as shown in Figure (4), the accuracy is about 98%, and precision 95% for positive samples and 100% for negative samples, also the recall is achieved 100% for positive and 96% for negative, finally the F1-Score is equal 97% for positive and 98% for negative.

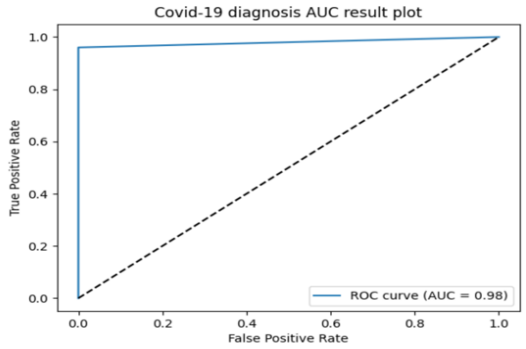

Figure 6: ROC curve for covid-19 classification result.

III. CONCLUSION & DISCUSSION

Developing and evaluating an AI model for COVID-19 diagnosis based on Mel-spectrogram, Convolutional Neural Network (CNN), and deep transfer learning have shown promising results. The model leverages the power of deep learning to analyze cough sounds, transforming them into Mel-spectrogram representations, and uses a pre-trained CNN for feature extraction and classification.

The use of transfer learning provides the advantage of leveraging pre-existing knowledge from large-scale datasets, which aids in achieving high accuracy even with limited COVID-19 specific data. The CNN's ability to process these spectrograms and identify intricate patterns allows the model to distinguish between coughs of COVID-19 positive individuals and others.

However, while these results are encouraging, several points should be noted. Firstly, the model's performance is dependent on the quality and diversity of the training data. The more varied and comprehensive the data, the better the model's generalizability. Therefore, future work should focus on collecting more diverse data, including different types of coughs, backgrounds, and recording conditions, also as an important finding of our study is that Convolutional Neural Networks (CNNs) trained on task-specific datasets achieve the highest accuracy compared to models trained on more general datasets like VGG19, ResNet50, DenseNet, and MobileNet. This highlights the significance of training CNNs with domain-specific knowledge for optimal performance in specific tasks.

Secondly, it's important to keep in mind that AI models like this should serve as a supplementary tool for healthcare professionals, not a replacement. While the model can provide a fast and non-invasive preliminary screening method, the final diagnosis should still rely on established medical procedures.

In conclusion, this research shows the potential of using AI, and particularly deep learning methods, in helping the fight against COVID-19. By further refining these models and incorporating more diverse data, we can move towards more reliable, non-invasive, and rapid diagnostic tools to aid in pandemic management and potentially other respiratory diseases.

REFERENCES

[1] [1] Q. Zhou, J. Shan, W. Ding, C. Wang and S. Yuan, "Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network," Frontiers in Robotics and AI, p. 11, 2021.

[2] [2] K. U. Islam and J. Iqbal, "An Update on Molecular Diagnostics for COVID-19," Frontiers in Cellular and Infection Microbiology, p. 10, 2020.

[3] [3] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood , M. S. Riaz, K. Ali , C. N. John, M. I. Hussain and M. Nabeel, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," ELSEVIER, vol. 20, p. 20, 2020.

[4] [4] P. Bagad, A. Dalmia, J. Doshi, A. Nagrani and P. Bhamare, "Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds," p. 12, 2020.

[5] [5] J. Laguarta, F. Hueto and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings," IEEE, vol. 1, p. 7, 2020.

[6] [6] D. AKGÜN, A. TALHA KABAKUŞ, Z. K. ŞENTÜRK, A. ŞENTÜRK and E. KÜÇÜKKÜLAHLI, "A transfer learning-based deep learning approach for automated COVID-19diagnosis with audio data," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 29, p. 18, 2021.

[7] [7] R. S.-U. na, C. Botelho, F. Teixeira, T. Rolland, A. Abad and I. Trancoso, "Transfer Learning-Based Cough Representations for Automatic Detection of COVID-19," INTERSPEECH 2021, p. 6, 2021.

[8] [8] F. Zhuang, Z. Qi, K. Duan and D. Xi, "A Comprehensive Survey on Transfer Learning," IEEE, vol. 109, no. 1, p. 10.

[9] [9] M. Cohen-McFarlane, R. Goubran and F. Knoefel, "Novel Coronavirus Cough Database: NoCoCoDa," IEEE, vol. 8, 2020.

[10] [10] H. Xiao, F. and Yu, "Speech Recognition Using Deep Learning Algorithms Based on Spectrograms and Mel-Spectrograms," p. 11, 2019.

[11] [11] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," Nature, 2015.

[12] [12] T. Giannakopoulos, "Github," 1 2023. [Online]. Available: https://github.com/tyiannak/deep_audio_features. [Accessed 1 3 2023].

[13] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, "How transferable are features in deep neural networks?," In NeurIPS, p. 27, 2014.

[14] M. Huh, P. Agrawal and A. A. Efros, "What makes ImageNet good for transfer learning?," 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1608.08614 .

[15] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury and G. Chanan, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," Advances in Neural Information Processing Systems (NeurIPS), p. 11, 2019.

[16] H. X. a. F. Yu, "Speech Recognition Using Deep Learning Algorithms Based on Spectrograms and Mel-Spectrograms," IEEE , p. 12, 2019.

[17] Y. E. Erdog˘an and A. Narin, "COVID-19 detection with traditional and deep features on cough acoustic signals," ELSEVIER, p. 11, 2021.

AUTHORS

First Author – Author name, qualifications, associated institute (if any) and email address.

Second Author – Author name, qualifications, associated institute (if any) and email address.

Third Author – Author name, qualifications, associated institute (if any) and email address.

Correspondence Author – Author name, email address, alternate email address (if any), contact number.