Dicle University
Journal of Engineering

https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Compare the classification performances of convolutional neural networks and capsule networks on the Coswara dataset

Abdulaziz Muhammad[1*], Muhammet Ali Arserim[2], Ömer Türk[3]

[1] Dicle Üniversitesi, Mühendislik Fakültesi, hassan.f.o@hotmail.com,ORCID: https://orcid.org/0000-0002-0673-1408
[2] Dicle Üniversitesi, Mühendislik Fakültesi,marserim@dicle.edu.tr,ORCID: https://orcid.org/0000-0002-9913-5946
[3] Mardin Artuklu Üniversitesi, Meslek Yüksekokulu, omerturk@artuklu.edu.tr,ORCID: https://orcid.org/0000-0002-0060-1880

ARTICLE INFO

Article history:

Received 24 March 2023
Received in revised form 24 April 2023
Accepted 25 April 2023
Available online 20 June 2023

Keywords:

COVID-19,CNN,CapsNet,
k-fold,leave-one-out

Doi: 10.24012/dumf.1270429

* Corresponding author

ABSTRACT

Since the beginning of the COVID-19 pandemic, researchers have developed numerous machine learning models to distinguish between positive and negative COVID-19 sounds. The aim of this study is to compare the classification performances of convolutional neural networks (CNN) and capsule networks (CapsNet) on the Coswara dataset, which includes 1404 healthy subjects and 522 COVID-19 positive subjects, each containing nine different types of sounds. The dataset was preprocessed by using oversampling and normalization techniques after feature extraction. k-fold cross-validation was used (where k=10) to train and evaluate the models. The CNN classifiers achieved a 94% ACC, while the CapsNet classifiers achieved an 90% ACC.

Furthermore, when using leave-one-out cross-validation, the CNN classifier achieved an ACC of 99%. we also compared the performance of the CNN and CapsNet networks on the Coswara dataset without preprocessing. Without oversampling techniques, the CNN classifiers achieved an 93% ACC, compared to 54% for the CapsNet classifiers. When normalization techniques were not applied, the CNN classifiers achieved an 86% ACC, while the CapsNet classifiers achieved a 26% ACC.

Introduction

COVID-19 is caused by the highly contagious SARS-CoV-2 virus [1]. Many people get infected and recover without the need for special medical care, but the elderly and people with cancer and chronic respiratory diseases can experience serious complications. As of February 21, 2023, there were reportedly (757,264,511) confirmed cases and (6,850,594) confirmed deaths of COVID-19 globally [1]. Most common symptoms are fever (98.6%), cough (59.4%), tiredness (69.6%), and loss of taste or smell [1,2].

The COVID-19 pandemic has had a significant impact on the world, and one of the biggest challenges has been accurately detecting the virus in patients. Deep learning techniques, such as convolutional neural networks and recurrent neural networks, have been used in COVID-19 detection to improve accuracy and speed up the process.These deep learning models can analyze medical images, such as X-rays and CT scans, to identify COVID19 pneumonia patterns. They can also process audio samples of coughs to detect COVID-19 sounds. With the help of large datasets and advanced machine learning algorithms, these models can achieve high accuracy rates in detecting COVID-19.

Several studies have been conducted using artificial intelligence to aid in the detection and diagnosis of COVID-19 based on chest X-ray and CT scan images. For example, COVID-19 was detected from chest X-ray images using a CoroNet, a Deep Convolutional Neural Network with 89.6% accuracy [3]. The ResNet50 architecture achieved 96.23% accuracy using computed tomography (CT) images [4]. Many studies utilized small data sets, and it should be noted that codes and data are unavailable for several studies. This poses difficulties in assessing the efficacy of the methods in clinical settings [5].

Other studies have used cough sounds to detect COVID-19. For example, Bagad et al. achieved 72% accuracy using the ResNet18 architecture [6]. Pahar et al. reported achieving 98% accuracy using the ResNet50 architecture with leave-p-out cross-validation, as well as 73%, 81%, 89%, 95%, and 94% accuracy with the Logistic regression (LR), support vector machines (SVM), multilayer perceptrons (MLP), convolutional neural networks (CNN), and long-short term memory (LSTM) architectures, respectively [7]. Aly et al. achieved 96% accuracy using simple binary classifiers by averaging the predictions of multiple models trained and evaluated separately on different sound types, such as cough, breathing, and speech [8].

Convolutional neural networks (CNN) and capsule networks (CapsNet) are two popular types of neural networks that have been widely used in image and sound recognition tasks. In particular, researchers have developed numerous machine learning models to distinguish between positive and negative COVID-19 sounds since the onset of the COVID-19 pandemic. The aim of this study is to compare the classification performances of CNN and CapsNet on a COVID-19 sound dataset called Coswara. By evaluating the performance of these two types of neural networks on the same dataset, this study aims to shed light on their respective strengths and weaknesses in detecting COVID-19 sounds.

Convolutional Neural Networks (CNN) is a feedforward artificial neural network consisting of layers. In general, a CNN structure comprises convolutional, pooling, and activation layers. CNN transmits the important weight values it has obtained to other layers. The pooling layer in the CNN structure reduces the data through downsampling. However, this can result in the loss of features such as depth and angle of the obtained features relative to each other. To preserve these properties, capsule neural networks (CapsNet) that use a dynamic routing algorithm are recommended. One of the most important goals of this study is to compare the classification performance of these two architectures

The aim of this study is to compare the effectiveness of simplified versions of convolutional neural networks (CNN) and capsule networks (CapsNet), while also examining the impact of preprocessing and cross-validation techniques on both. The goal of this study is to understand the strengths and weaknesses of each neural network when applied to the Coswara dataset. As neural networks are increasingly popular in various applications, it is crucial to compare their performance and identify their optimal use cases to advance the field of machine learning.

Coswara dataset was used because it is diverse and available online [9]. The Coswara dataset is a collection of audio recordings that includes sounds from healthy individuals and those diagnosed with COVID-19. It consists of nine distinct sound types, such as coughing, breathing, and vowel sounds. The dataset has been widely used by researchers to investigate the potential of machine learning algorithms for COVID-19 detection. The sound of coughing was also utilized to extract features, and we applied preprocessing techniques to train both CNN and CapsNet to compare their performances.

Material and methods

In this section, we will mention the dataset in detail and methods used in the study. Figure 1 shows the flow chart of the model that we followed. The dataset was cleaned to remove empty and too-short audio files. Features were extracted, and preprocessing techniques like SMOTE and Normalization was applied, followed by the application of the Convolutional Neural Network (CNN) and capsule network (CapsNet).

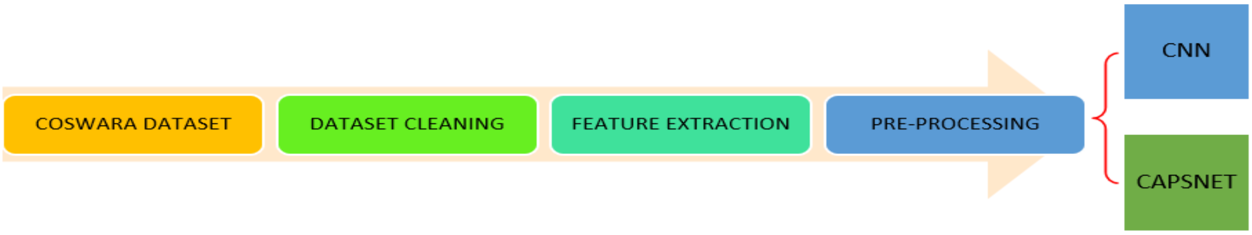


Figure 1. flow chart of the model

Dataset

This study utilized the Coswara dataset, which was curated by the Indian Institute of Science (IISc) [10]. The dataset consists of audio recordings from both healthy individuals and those diagnosed with COVID-19. Each audio sample

contains nine distinct types of sounds, including cough-heavy, cough-shallow, breathing-deep, breathing-shallow, counting-fast, counting-normal, and three vowel sounds: “a”, “e”, and “o”. Table 1 presents the different types of COVID-19 cases within the dataset.

The dataset contains (1,404) healthy cases, and those who were diagnosed with COVID-19 were divided according to symptoms into three categories: positive_mild (325), positive_moderate (127), and positive_asymp (70). All three categories were collected during feature extraction. Thus, the total number of cases in the dataset is (522) infected with COVID-19 compared to (1,404) healthy cases. The imbalance between the number of healthy cases and COVID-19 cases in the dataset can lead to biased training results [7]. This is because the model may tend to predict the majority class (healthy cases) more frequently, leading to lower accuracy in predicting the minority class (COVID-19 cases). To address this issue, we applied the Synthetic Minority Oversampling Technique (SMOTE) [11]. In addition, we used the normalization technique to reduce overfitting and improve accuracy. For this purpose, we used the Scikit-learn Python library [12].

Table 1 covid-19 cases

Covid-19 cases	Numbers of samples
healthy	1404
positive_mild	325
no_resp_illness_exposed	192
resp_illness_not_identified	153
positive_moderate	127
recovered_full	103
positive_asymp	70
Total	2374

Feature extraction

To extract features from the audio files, we utilized the librosa Python package [13] to compute Mel-frequency cepstral coefficients (MFCCs). Before feature extraction, we resampled the raw audio samples to 22.50 kHz. Figure 2 provides a visual representation of the feature extraction process.

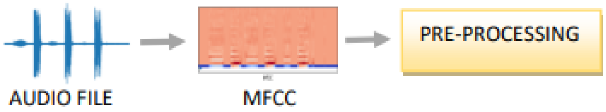


Figure 2 feature extraction

Convolutional Neural Network

CNNs are a class of deep neural networks that are commonly used for image and video recognition, classification, and segmentation tasks. In a CNN, the input is processed through a series of layers, each of which applies a set of filters or kernels to the input. These filters extract important features, such as edges or patterns, from the input, which are then used to classify or segment the data [14]. A convolutional neural network architecture contains several layers that work together. These layers are:

- Convolution layer: applies a set of filters or kernels to the input data, which extracts important features such as edges or patterns.
- The pooling layer is responsible for down-sampling the feature maps created by the convolutional layer to reduce their spatial dimensions. The most popular type of pooling is max pooling.
- A fully connected layer, which is also referred to as dense layer, is a type of neural network layer in which each input is connected to every output neuron by means of a learnable weight [14].
- Activation layer or activation function: applies a mathematical function to the output of a previous layer. Common activation functions include sigmoid, tanh, ReLU, and softmax. Figure 2 shows an example of a simple CNN classifier.

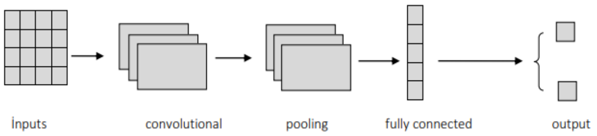


Figure 2. CNN Classifier

For the CNN classifier, we utilized two-dimensional convolutional layers and max-pooling with a dropout rate of 0.2, followed by two dense layers.

Capsule network

Capsule Network (CapsNet) is a type of neural network architecture proposed by Geoffrey Hinton and his team at the University of Toronto in 2017 [15].

The aim of Capsule networks is to overcome the limitations of traditional Convolutional Neural Networks (CNNs) such as the inability to handle variations in object position, size, and orientation.

The basic building block in a Capsule Network is a capsule, which is a group of neurons that represent a specific part of an object and encode various properties of that part, such as its pose, size, and deformation, capsule's inputs and output are vectors [15], Capsules are organized in layers, where higher-level capsules represent more complex parts of the object [3]. Figure 3 shows Structure of the capsule network in Dynamic Routing Between Capsules,2017.

In a capsule network, the convolutional layer performs feature extraction, and its output is fed into the primary capsule layer. Dynamic routing connects the primary capsules to higher-level capsules, allowing the network to learn how to combine and transform pose vectors for more complex and abstract input features. The DigitCaps layer is a fully connected layer with ten 16D capsules, each of which receives input from all the capsules in the previous layer. The final layer determines the length of each capsule in the previous layer, which is necessary to obtain the probability that the object or entity is present in the input [15]. To build the capsent classifier, we utilized the code that was shared on GitHub [16].

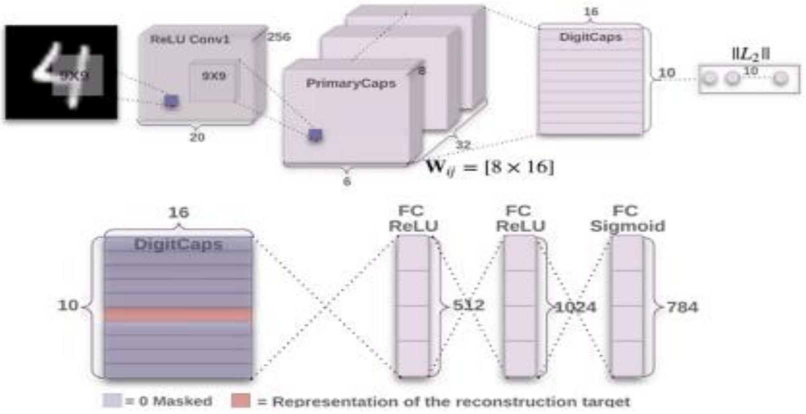
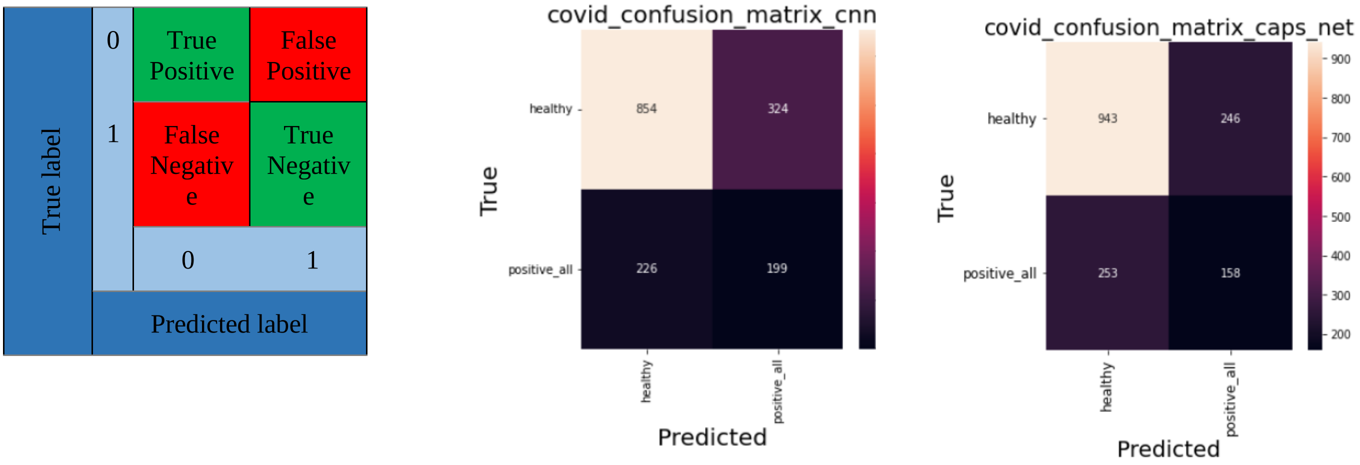


Figure 3 shows Structure of the capsule network in Dynamic Routing Between Capsules,2017.

Results

After extracting features from the Coswara dataset, we used convolutional neural network and capsule network models to perform a 2-class classification of COVID-19 and health by utilizing the train-test-split cross-validation. Figure 4(a) provides a visual representation of the original confusion matrix. Additionally, Figure 4(b) shows the confusion matrix for the CNN classifier, and Figure 4(c) shows the confusion matrix for the CapsNet classifier.



(a)

(b)

(c)

Figure 4(a)shows original confusion matrix (b) confusion matrix for the CNN classifier(c) the confusion matrix for the CapsNet classifier

We used state-of-the-art evaluation metrics, such as Accuracy, Recall, Precision, and F1-Score, to assess the performance of our classification task. Table 2 displays a detailed description and mathematical expression of these performance metrics used in our study [17].

Table 2 performance metrics

Metrics	Description	Mathematical Expression
Accuracy	The degree of correctness of COVID-19 measurements was measured by evaluating the number of true values against all instances that were assessed.	$\text{Acc} = \dfrac{tp + tn}{tp + tn + fp + fn}$
Sensitivity or Recall	The proportion of true positive COVID-19 cases was calculated by dividing the number of correctly identified	$\text{Sen} = \dfrac{tp}{tp + fn}$

	individuals with the disease by the actual number of people who have the disease.	
Precision	The proportion of true positive COVID-19 cases was calculated by dividing the number of correctly identified individuals with the disease by the predicted number of people who were expected to have the disease.	$\text{prec} = \frac{tp}{tp + fp}$
Specificity	The proportion of true negative non-COVID-19 cases was calculated by dividing the number of correctly identified individuals without the disease by the actual number of people who do not have the disease.	$\text{Spec} = \frac{tn}{tn + fp}$
F1_Score	The weighted average of precision and recall is a performance metric that combines precision and recall using a weighted average, where the weight corresponds to the proportion of true positives for a given class.	$\text{F1\_score} = 2 * \frac{sen * spec}{sen + spec}$

Table 2 presents the performance comparison of two deep learning models, CNNs and CapsNets, which were trained and evaluated using the train-test-split cross-validation method

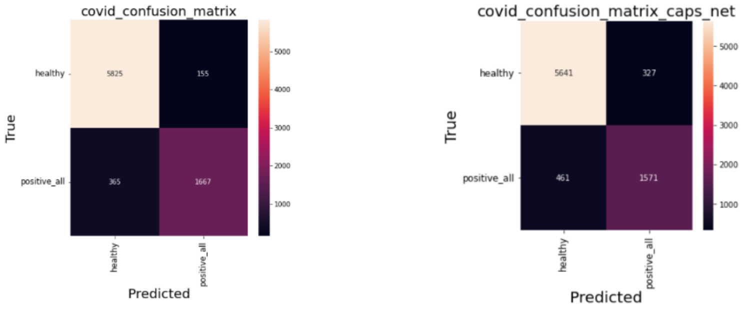
table 3 performance of CNN and CapsNet with train-test-split cross validation

Classifiers	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)	F1-Score (%)
CNN	69	58	59	38	58
CapsNet	69	59	59	39	59

In our study, we compared the performance of two deep learning models, Convolutional Neural Network (CNN) and Capsule Network (CapsNet), in a specific classification task Our initial evaluation using train-test-split cross-validation showed that the two classifiers achieved comparable performance. To validate these results, we conducted k-fold cross-validation with k=10. The evaluation results revealed that the CNN classifier consistently outperformed the CapsNet classifier in terms of Accuracy, Recall, Precision, and F1-Score. We have presented the detailed performance metrics of the CNN and CapsNet classifier using k-fold cross-validation in Table 3. In addition, we generated confusion matrices for both classifiers, with Figure 5(a) displaying the confusion matrix for the CNN classifier and Figure 5(b) showing the confusion matrix for the CapsNet classifier. Based on these findings, we can conclude that the CNN model may be a better choice for the classification task than CapsNet.

table 4 performance of CNN and CapsNet with K-Fold (where k=10) cross vlidation

Classifiers	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)	F1-Score (%)
CNN	94	90	93	91	91
CapsNet	90	86	88	82	87



(a)

(b)

Figure 5(a) the confusion matrix for the CNN classifier 5(b) the confusion matrix for the CapsNet classifier

We observed that the training time for CNNs was 222 seconds, while Capsule Networks took 1417 seconds to train on the same dataset, indicating a significant disadvantage of Capsule Networks in terms of training time. Furthermore, using leave-one-out cross-validation with the CNN network, we achieved the highest accuracy of 99%, but it took 17,412 seconds to complete the process. As a result, we did not apply the leave-one-out cross-validation technique with CapsNet.

In addition, the effect of preprocessing techniques appears differently on two networks, where the capsule network is affected greatly as shown in Table 4.

table 5 effect of preprocessing techniques

	CNN(ACC)	CapsNet(ACC)
With normalization, oversampling and train-test-split	0.69	0.69
With normalization, oversampling and k-fold (k=10)	0.94	0.90
Without normalization	0.86	0.26
Without oversampling	0.93	0.54

Table 6 presents the results of some previous research conducted on the Coswara dataset.

Table 6 presents the results of some previous research on the Coswara dataset

Research	Dataset	Model/classifiers	Results ACC
Chaudhari et al. [18]	Coswara Coughvid [19]	Ensemble Deep Learning Mode	77.1%
Bagad et al. [6]	Cough Against Covid Coswara	ResNet-18 Shallow classifiers	72%
Aly at el. [8]	Coswara Virufy [19]	Deep Model Shallow classifiers	96.4%
Pahar et al. [7]	Coswara SARCOS	MLP CNN LSTM Resnet50	87% 94% 94% 95%
Alli et al. [17]	Coswara	CNN based on DeepShufNet	90.1
Verde et al. [20]	Coswara	Naive Bayes (NB) Bayes Net (BN) Support Vector Machine (SVM)	90% 89% 97%
This Study	Coswara	CapsNet (k-fold where k=10) CNN (k-fold where k=10) CNN (leave-one-out)	90% 94% 99%

Discussion and Conclusions

In this study, the Coswara dataset was selected and processed. Then, convolutional neural networks and capsule networks were applied using three types of cross-validation techniques. The results show that CNNs achieved better results, are more flexible and easier to apply, and are less affected when preprocessing techniques such as oversampling and normalization are not applied. However, capsule networks require more training time.

References

[1] WHO, https://www.who.int/health-topics/coronavirus.

[2] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong et al, “Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China”, JAMA, vol. 323, no. 11, pp. 1061– 1069, 2020.

[3] A. I. Khan , J. L.Shah , M. M. Bhat, “CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images”,2020.

[4] S.Walvekar,D. Shinde, “Detection of COVID-19 from CT Images Using resnet50”, 2020.

[5] P.Aggarwal , N. K. Mishra, B.Fatimah , P. Singh , A. Gupta , S. D. Joshi ,”COVID-19 image classification using deep learning: Advances, challenges and opportunities”, 2022, 105350.

[6] P.Bagad,A.Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A.Mahale,S.Rane, N. Agarwal, R.Panicker, “Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds |”,2020.

[7] M.Pahar, M.Klopper, R. Warren, T.Niesler, “COVID-19 Cough : Classification using Machine Learning and Global Smartphone Recordings” ,2021,104572.

[8] M.Aly, K.H. Rahouma, S. M. Ramzy,” Pay attention to the speech: COVID-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings”, pp 3487-3500, 2022.

[9] https://www.kaggle.com/datasets/janashreeananthan/coswara.

[10] https://coswara.iisc.ac.in/?locale=en-US.

[11] https://imbalanced-learn.org/stable.

[12] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html.

[13] https://librosa.org/doc/main/generated/librosa.feature.mfcc.html.

[14] R. Yamashita & M. Nishio & R.Gian Do & K. Togashi,” Convolutional neural networks: an overview and application in radiology”, pp.611–629, 2018.

[15] M. Patrick, A. Adekoya a , A.Mighty , B.Edward : “Capsule Networks – A survey”, pp1295-1310, 2022.

[16] https://github.com/XifengGuo/CapsNet-Keras.

[17] O. Abayomi-Alli, R.Damaševičius ,A.Abbasi ,R.Maskeliūnas s ,” Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques”, 2022.

[18] G. Chaudhari, X.Jiang, A.Fakhry, A. Han, J. Xiao, S. Shen, A. Khanzada,” Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough”,2020.

[19] L. Orlandic, T. Teijeiro, D. Atienza, The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms, 2020.

[20] L.VERDE, G.DE PIETRO, A. GHONEIM, M. ALRASHOUD , K. N. AL-MUTIB , G. SANNINO ,” Exploring the Use of Artificial Intelligence Techniques to Detect the Presence of Coronavirus Covid-19 Through Speech and Voice Analysis”, 2021.3075571.