



Regular Article

# COVID-19: respiratory disease diagnosis with regularized deep convolutional neural network using human respiratory sounds

Lella Kranthi Kumar<sup>a</sup> and P. J. A. Alphonse<sup>b</sup>

Department of Computer Applications, NIT Tiruchirappalli, Tiruchirappalli, Tamil Nadu 620015, India

Received 14 March 2022 / Accepted 26 July 2022 / Published online 10 August 2022

© The Author(s), under exclusive licence to EDP Sciences, Springer-Verlag GmbH Germany, part of Springer Nature 2022

**Abstract** Human respiratory sound auscultation (HRSA) parameters have been the real choice for detecting human respiratory diseases in the last few years. It is a challenging task to extract the respiratory sound features from the breath, voice, and cough sounds. The existing methods failed to extract the sound features to diagnose respiratory diseases. We proposed and evaluated a new regularized deep convolutional neural network (RDCNN) architecture to accept COVID-19 sound data and essential sound features. The proposed architecture is trained with the COVID-19 sound data sets and gives a better learning curve than any other state-of-the-art model. We examine the performance of RDCNN with Max-Pooling (Model-1) and without Max-Pooling (Model-2) functions. In this work, we observed that RDCNN model performance with three sound feature extraction methods [Soft-Mel frequency channel, Log-Mel frequency spectrum, and Modified Mel-frequency Cepstral Coefficient (MFCC) spectrum] for COVID-19 sound data sets (KDD-data, ComParE2021-CCS-CSS-Data, and NeurIPs2021-data). To amplify the models' performance, we applied the augmentation technique along with regularization. We have also carried out this work to estimate the mutation of SARS-CoV-2 in the five waves using prognostic models (fractal-based). The proposed model achieves state-of-the-art performance on the COVID-19 sound data set to identify COVID-19 disease symptoms. The model's learnable parameter gradients have vanished in the intermediate layers while optimizing the prediction error which is addressed with our proposed RDCNN model. Our experiments suggested that  $3 \times 3$  kernel size for regularized deep CNN (without max-pooling) shows 2–3% better classification accuracy compared to RDCNN with max-pooling. The experimental results suggest that this new approach may achieve the finest results on respiratory diseases.

## 1 Introduction

The issue of Human Respiratory Sound Auscultation (HRSA) and the identification of SARS-CoV-2 have gotten a lot of attention from clinical researchers and public health care scientists [1, 2]. The World-Health-Organization (WHO) announced that COVID-19/SARS-CoV-2 is the largest pandemic in the unbroken society in the year 2020. This COVID epidemic claims over 6,029,852 lives in the world as of 12th March 2022. As far as 12th March 2022, there have been 452,201,564 confirmed cases, 6,029,852 deaths, and 10,704,043,684 people have been vaccinated [3]. According to medical experts, the tracking of SARS-CoV-2 disease symptoms (for different variants) is very difficult to put an end to the spreading of SARS-CoV-2

The Dynamics of the COVID-19 Pandemic—Nonlinear Approaches on the Modelling, Prediction and Control. Guest editor: Santo Banerjee.

disease. In this context, a variety of Artificial Intelligence (AI) techniques have been introduced into the actual world to deal with such issues [4–7].

The seriousness of the SARS-CoV-2 disease is identified into 3 major types such as mild COVID-19 symptoms, moderate COVID-19 symptoms, and COVID-19 with extreme symptoms. The existing research focuses on diagnostic tools and predictive models to identify the severity of the COVID-19 pandemic with the first wave, second wave, third wave, and beyond [8–11]. The researchers have been focused on diagnostic tools, predictive models, and statistical tools to identify SARS-CoV-2 situations in various countries for different variants. The researchers and scientists are focused on various physics models to estimate the risk factors in different waves and analysis of how the SARS-CoV-

<sup>a</sup> e-mail: kranthi1231@gmail.com (corresponding author)

<sup>b</sup> e-mail: alphonse@nitt.edu

2/COVID-19 is spreading between different species. The researchers are also focused on mathematical models, non-linear dynamic models, and prediction models to identify the seriousness of the virus, how the vaccination is distributed to the people, how vaccination gives immunity to our body, and the death rate in various countries [59–63].

In recent, the SARS-CoV-2/COVID-19 pandemic with different variants (Alpha, Beta, Gamma, Delta, Delta+, and Omicron) is wreaking havoc on the real world to make people afraid to contact physically [12–16]. The SARS-CoV-2 disease can be identified with different human-generated data, such as X-ray, CT-scan, RTPCR, patient tweets, and respiratory sounds. Medical researchers have implemented several deep learning (DL), machine learning (ML), and signal processing (SP) models to diagnose SARS-CoV-2 with the human pulmonary sounds (cough, breathing, and voice) and images (X-ray and CT-scan) [17–23]. Therefore, there are numerous options available to identify SARS-CoV-2, and one of the approaches is human respiratory sound auscultations (HRSA). Medical scientists have employed different human pulmonary sounds like food absorption, breath, body vibration, heartbeat sound, lung sounds, cough, and voice to diagnose the SARS-CoV-2 disease. Nowadays, a manual examination is used to collect such signs during scheduled visits of the patients [24].

Clinical scientists and technological researchers have started their research by collecting respiratory sound auscultations using digital stethoscopes from the human body, and it performs automatic analysis of human respiratory sounds data [25, 26] (e.g., wheezing sound detection in asthma). Clinical scientists are also experimented with using the voice audio data to aid in the automatic recognition of a variety of illnesses: Alzheimer's disease [27] (it can be identified with a slur in the voice, stammer in the voice, repetition of words, and incomplete words in a sentence), Parkinson's disease (PD) has a variety of consequences with voice data [28] (various peoples with PD will speak close-mouthed, and they do not express sufficient emotions in the single timbre sound, breathy voice, and slur words), coronary artery disease [29, 30] (voice disorder, neck pain, and fatigue), and various other invisible disorders [31–33] with voice frequency and rhythm, brain trauma from the voice, battle fatigue, etc. The utilization of respiratory system sound as a screening tool for many illnesses has an enormous potential for timely screening and inexpensive remedies that can be provided freely accessible to society if included in main commodities.

In the last 2 years, various machine learning (ML) methods and DL frameworks such as support vector machine (SVM), convolutional neural network (CNN), multivariate linear regression (MLR) models, learning vector quantization (LVQ), and ensemble pre-trained approaches have been used to ameliorate the effects of

human pulmonary sound disease diagnosis on the KDD-data, ComParE2021-CCS-CSS-Data, and NeurlPs2021-data. Digital stethoscopes can obtain human respiratory sounds, and it is reliable and robust to collect respiratory sounds. A recent study has been initiated to examine how the human respiratory system generated sounds [breathing sound (inhale and exhale), single sentence voice sound, and cough sound] are collected by smartphone and web-based environments from patients to verify SARS-CoV-2 symptoms with healthy human respiratory sound signs. To use a set of 48 users with SARS-CoV-2/COVID-19 signs (collected using smart devices) and other diagnostic coughs trained in a series of models that have been used initially to classify COVID-19 symptoms from the healthy symptoms [34–38]. The utilization of the human respiratory auscultations on crowdsourced COVID-19 unregulated data is used in this part of our research.

The research is focused on the human respiratory sound recognition system to classify various respiratory diseases (COPD, Asthma, Bronchitis, Pertussis, and COVID-19). Therefore, there should be an adequate model is necessary to diagnose respiratory sounds. This research introduces Soft-Mel-frequency spectrum, Log-Mel frequency spectrum, and MMFCC feature extraction methods that are owned in the RDCNN Model-1 (with Max-Pooling), or RDCNN Model-2 (without Max-Pooling). The research also includes augmented data with the COVID-19 sounds dataset to amplify the model performance with a mix of L2 regularization to reduce overfitting due to a lack of KDD-data, ComParE2021-CCS-CSS-Data, and NeurlPs2021-data datasets. The major contribution of this research is focused on designing two RDCNN models (with max-pooling and without max-pooling) with three different feature spectrums [Soft-Mel frequency channel, Log-Mel frequency spectrum, and Modified Mel-frequency Cepstral Coefficient (MMFCC) spectrum].

In this research, we used various depths sizes and receptive fields to obtain relevant disease features that should aid in the classification of human respiratory sound data. The study also highlights the usefulness of Neural Architecture Search (NAS) in identifying superior framework classification. We have also carried out this work to analyze and estimate the mutation of Coronavirus in the five variants using a prognostic model (fractal-based). The datasets have been collected from Cambridge University for research purposes with mutual agreement. We have done statistical analysis and pre-processing on the collected datasets and prepared the final respiratory COVID-19 sounds dataset consisting of 10-classes [001—COPD, 002—Asthma, 003—Pertussis, 004—Bronchitis, 005—COVID-19 Variant\_1 (Alpha), 006—COVID-19 Variant\_2 (Beta), 007—COVID-19 Variant\_3 (Gamma), 008—COVID-19 Variant\_4 (Delta), 009—COVID-19 Variant\_5 (Omicron), and 010—Healthy Symptoms]. Also, feature extraction methods are applied with Model-1 and

Model-2 on these expanded data. Our study shows that using RDCNN without max-pooling function (Model-2) and the MFCC feature spectrum and using augmented data applied for model training can produce fantastic and state-of-the-art results for the diagnosis of respiratory diseases.

This research work is organized into different sections to understand the best manner of the presentation. The background works and analysis of the existing works are illustrated in Sect. 2. The dataset description, statistical analysis of datasets, the augmentation approach, experimental setup, the RDCNN model creation, model analysis of RDCNN with the implementation process, and Model-1 and Model-2 architectural frameworks with functional analysis are presented in Sect. 3. Section 4 focuses on result analysis on three benchmark datasets, performance testing of the proposed model concerning to accuracy, and  $F_1$  score by comparing the obtained results with the existing approaches. And finally, we concluded this research summary by describing this research's significant findings and study in the last section.

## 2 Background works

Sound has long been acknowledged as a possible human respiratory health screening medium by researchers and experts. For instance, individual pulmonary audio sounds were employed with a digital stethoscope to recognize sounds from the human respiratory system [39]. It requires highly skilled professionals to listen and understand the disease, so it is being rapidly supplanted by newer technologies like magnetic resonance imaging (MRI) and ultrasounds, which are easy to examine and analyze. Recent advances on the other hand in automatic sound analysis and design may be able to overcome these methods and provide low-cost, widely disseminated alternatives in the form of respiratory sound. Microphone-based sound signals have recently been used to process sounds on commodity and device types of machinery like iOS or Android-based tools and wearable technology.

Brown et al. [40, 47, 48] suggested web-based, iOS, and Android applications to record pulmonary COVID-19 sound data (KDD-data, ComParE2021-CCS-CSS-Data, and NeurIPs2021-data) from crowdsourcing sound data of around 2 k COVID-19-positive cases (with different variants) from about 36 k unique users. Authors have launched one COVID-19 sounds app from Cambridge University to collect pulmonary sound data [single sentence voice audio data, breathing (inhale and exhale) sound data, and cough sound data]. The authors have considered a few general parameters (previous medical history, cigarette smoking status, and demographics) while collecting the samples. The authors have considered asthma, pertussis, COVID-19, and COPD diseases. The authors have done binary classification for 3 various tasks (Task-1 is Asthma Vs. COVID-19, Task-2 is COVID-19 Vs. Non-COVID-19,

and Task-3 is COVID-19 Vs. Healthy sounds) using the SVM model to identify the disease symptoms; after that, they have used VGG-Net (Visual Geometry Group) model to improve the performance for Task-2 and Task-3.

The existing research focused on diagnostic tools and predictive models to identify the severity of the COVID-19 pandemic with the first wave, second wave, third wave, and beyond. Easwaramoorthy et al. proposed the prognostic models (fractal-based) to estimate the risk factors, virus spreading, and the death rate in the first and second waves. And also, the authors have applied statistical tools and performance measures to test the model performance [8]. A Gowrisankar et al. implemented a fractal interpolation method to predict the Omicron variant cases of six countries (Denmark, UK, South Africa, India, Germany, and The Netherlands) in the month of December 2021, and January and February 2022. The authors also identify the transmission rate of the Omicron variant in these six countries [56]. Gopal et al. introduced a susceptible exposed infectious removed (SEIR) model to analyze vaccine strategies and virus spreading in the lockdown situation for the first and second wave [57]. Natiq and Saha implemented a non-linear dynamical model by combining of susceptible infected and recovered (SIR) model and Lotka–Volterra model to identify the virus transmission between the various species [58]. The researchers are also focused on mathematical models, non-linear dynamic models, and prediction models to determine the seriousness of the virus, how the vaccination is distributed to the people, how vaccination gives immunity to the human body, and the death rate in various countries [59–63].

Han et al. [41] presented the COVID-19 voice data intelligence system based on four different parameters: (1) quality of sleep, (2) fatigue, (3) anxiety, and (4) severity. The authors have launched one Corona-Voice-Detect app from Mellon University to collect human respiratory sounds (Cough, voice). Teijeiro et al. [42] introduced the “COUGHVID” app for collecting cough (around 20 k recordings) samples to examine COVID-19 symptoms. They have trained a neural network (NN) model with 120 cough samples and 95 non-coughs samples. Among them, 25% of COVID-19 samples are symptomatic, 35% are symptomatic samples with voice, 25% are healthy samples, and 15% are asymptomatic. Wang et al. [2] developed an approach (GRUs—gated recurrent units) for categorizing substantial screening of infected people with COVID-19 in various ways. The work identifies the different voice vocal patterns and can put this instrument to use in the real world. A powerful new respiratory simulation (RS) approach is proposed in this study to cover the variation between such an enormous volume of training data and insufficient available test data to consider the features of natural respiratory sound patterns from the respiratory sounds. Direct human-to-human transmission mechanism is also considered in this analysis to estimate the spreading of the virus [43].

Later, the researchers have been implemented AI-based screening models (AI4COVID-19, CNN, DCNN, RNN, etc.) using feature extraction methods (SSP—speech signal processing, MFCC, MMFCC—modified MFCC, mel-frequency, and GFCC—gamma-tone frequency cepstral coefficients, etc.) to identify COVID-19 disease symptoms from the lung sounds data. The summary of the previous examinations on human respiratory sound disease analysis is represented in Table 1. According to the above analysis, there is no accurate and sensitive framework is not available to diagnose human respiratory disease on large datasets to detect COVID-19 symptoms with various variants (Alpha, Beta, Gamma, Delta, and Omicron) and other respiratory disease symptoms.

### 3 Dataset description and experimental setup

#### 3.1 Dataset preparation and analysis

Three datasets (KDD-data, ComParE2021-CCS-CSS-Data, and NeurlPs2021-data) have been collected from the University of Cambridge for research purposes with the mutual agreement. Cambridge University's ethical committee has accepted these three datasets to do the research. The authors [40] have implemented a web-based application and iOS/Android application to collect the respiratory sounds. Along with this dataset, the authors have collected past medical history (any symptoms—fever, running or blocked nose, loss of taste and smell, sore throat, dry cough and wet cough, muscle aches, dizziness, the difficulty of breathing or feeling shortness of breathing, headache, confusion or vertigo, chills, and tightness in the chest) from the users. Along with that, cigarette smoking status, age, gender, current COVID-19 status, previous COVID-19 status (Never, 14 days and above, and at least 14 days), and the current status of hospitalization. After collecting all these parameters from the users, the authors have collected respiratory sounds like breath sound (five times), cough (three times), and voice for a single sentence (I hope that these data can help to manage the virus pandemic.) Furthermore, this Android application suggests that users give additional respiratory sounds every 2 days (breath, cough, and voice). It provides a unique opportunity to investigate the progress in sound-based clinical well-being. This information is secured on University of Cambridge databases, and the acquired data are retained locally until the device is connected to WiFi, at which point it is transferred. If the necessary is received and past medical data are captured, then the users' personal information will be deleted based on the user requests.

The Cambridge University medical researchers started collecting the data in April 2020, and they released the first COVID-19 crowdsourced dataset (KDD-data) in May 2020. This dataset

consists of ten different data folders (asthma-android-with-cough, asthma-web-with-cough, covid-android-no-cough, covid-android-with-cough, covid-web-no-cough, covid-web-with-cough, healthy-android-no-symptoms, healthy-android-with-cough, healthy-web-no-symptoms, healthy-web-with-cough). They collected around 12,000 samples from different countries, and around 302 users tested with COVID-19 positive among them. The second COVID-19 crowdsourced dataset (ComParE2021-CCS-CSS-Data) is released in June 2021 [47]. This dataset consists of two metadata (.csv) data files which are used to classify two various classes (Task-1: COVID symptoms, Task-2: No COVID symptoms). The third dataset (NeurlPs2021-data) was released in November 2021, and it consists of two different data folders to predict two tasks (Task-1: other respiratory disease symptoms prediction, Task-2: COVID-19 symptoms prediction). It is a huge dataset that contains 53,500 respiratory sound samples collected (above 550 h in total) from 36,117 unique users (62% males, 36% females, and 2% unidentified data). The COVID-19 virus screening self-reported for individual users is also provided with 2105 individuals testing positive. As of our knowledge, this COVID-19 crowdsourced dataset is a huge dataset that consists of three different modalities (voice, breath, and cough).

The statistical analysis of this large COVID-19 crowdsourced sounds dataset (NeurlPs2021-data) is represented in Fig. 1. In this dataset, a smaller group of participants are current and ex-smokers, and most of the users are (20–50 year age adults) non-smokers. The smoking status of the participants is 19,776 users who have never smoked cigarettes, 8718 cigarette smokers (407 will smoke more than 20 cigarettes per day, 3017 will smoke 11–20 cigarettes every day, 3995 users smoke around 1–10 cigarettes, and 1245 users will smoke less than one cigarette daily) and these COVID-19 sounds data have been collected from the various countries (Greece—393, Russia—407, Spain—522, Iran—987, Brazil—1016, Germany—1191, UK—2019, US—2829, Italy—4698, etc.). Along with this dataset availability of existing datasets [42, 47, 53–55], we have collected various other COVID-19 respiratory sounds datasets (Virufy—121 samples, COVID-19-Cough—1324 samples, Coswara—2030 samples, Tos COVID-19—5867 Samples, and COUGHVID—27,550 samples).

The systematic analysis of the various datasets, modalities, and medical history is presented in Table 2. Along with these data, we have collected a few COVID-19 samples from the local hospitals and prepared one COVID-19 respiratory disease dataset (with a total of 90,341 samples, around 270 GB Data) with ten classes (001—COPD, 002—Asthma, 003—Pertussis, 004—Bronchitis, 005—COVID-19 Variant\_1 (Alpha), 006—COVID-19 Variant\_2 (Beta), 007—COVID-19 Variant\_3 (Gamma), 008—COVID-19 Variant\_4 (Delta), 009—COVID-19 Variant\_5 (Omicron), and 010—Healthy Symptoms). We have collected COVID-19 large-scale sounds (breath, voice, and cough) data from trusted repositories and cited sources in references [40, 53–55]. The central part of

**Table 1** The previous analysis to the identification of COVID-19 disease signs from human respiratory/pulmonary sound data

Authors	Dataset	Year	Model	Classification		Accuracy (%)
				Tasks	Modalities	
Alsabek et al. [44]	Own dataset (14 patients)	2020	SVM + MFCC	Positive-COVID vs negative-COVID	Breath	43
				Positive-COVID vs negative-COVID	Cough	42
				Positive-COVID vs negative-COVID	Breath	58
				Positive-COVID vs negative-COVID	Cough	65
					Voice	79
				Anxiety	—	49
				Sleep quality		55
				Fatigue		42
				COVID-19 or healthy signs	Cough	86
				COVID-19 vs non-COVID-19	Cough	83.20
				Binary classification (COVID-19 positive vs negative COVID-19)	Speech	82.5
				AB		78.9
Quian et al. [41]	COVID-19 audio data	2020	SVM	Binary classification (COVID-19 positive vs negative COVID-19)	Speech	80.3
				AB		79.4
						81.2
				Asthma/COVID-19 cough	Breath, cough, voice	80
					Cough	87
					Cough	88
				COVID-19 vs healthy signs	Cough and speech	82.2
				COVID-19 vs healthy signs		83.9
				COVID-19 vs healthy signs		78.8
				COVID-19 vs healthy signs		84.6
				COVID-19 vs healthy signs		45.3
				COVID-19 vs healthy signs		70.8
Orlandic et al. [42]	COUGHVID	2020	Signal processing + CNN	COVID-19 vs healthy signs		87.5
				Task-1: other respiratory disease symptoms prediction	Breath	60
						52
						65
					Cough	70
						66
						74
Laguarda et al. [45]	MIT-open-voice	2020	CNN + MFCC	Task-1: other respiratory disease symptoms prediction	Breath	60
						52
						65
					Cough	70
						66
						74
M. Al Ismail et al. [46]		2020	LR NL-SVM DT RF AB	Task-1: other respiratory disease symptoms prediction	Breath	82.5
						78.9
						80.3
						79.4
						81.2
Brown et al. [40, 47, 48]	KDD-data	2020	SVM and VGG-Net	Task-1: other respiratory disease symptoms prediction	Breath, cough, voice	80
					Cough	87
					Cough	88
ComParE2021-CCS-CSS-Data	ComParE2021-CCS-CSS-Data	2021	COMPARE functionals + SVM COMPARE BoAW + SVM DEEPSPECTRUM + SVM S2SAE + SVM Transformer + SVM CNN + LSTM RNN Fusion of best	COVID-19 vs healthy signs	Cough and speech	82.2
						83.9
						78.8
						84.6
						45.3
						70.8
						87.5
NeurlPs2021-data	NeurlPs2021-data	2021	OpenSMILE + SVM Pre-trained VGGish Fine-tuned VGGish OpenSMILE + SVM Pre-trained VGGish Fine-tuned VGGish	Task-1: other respiratory disease symptoms prediction	Breath	60
						52
						65
					Cough	70
						66
						74

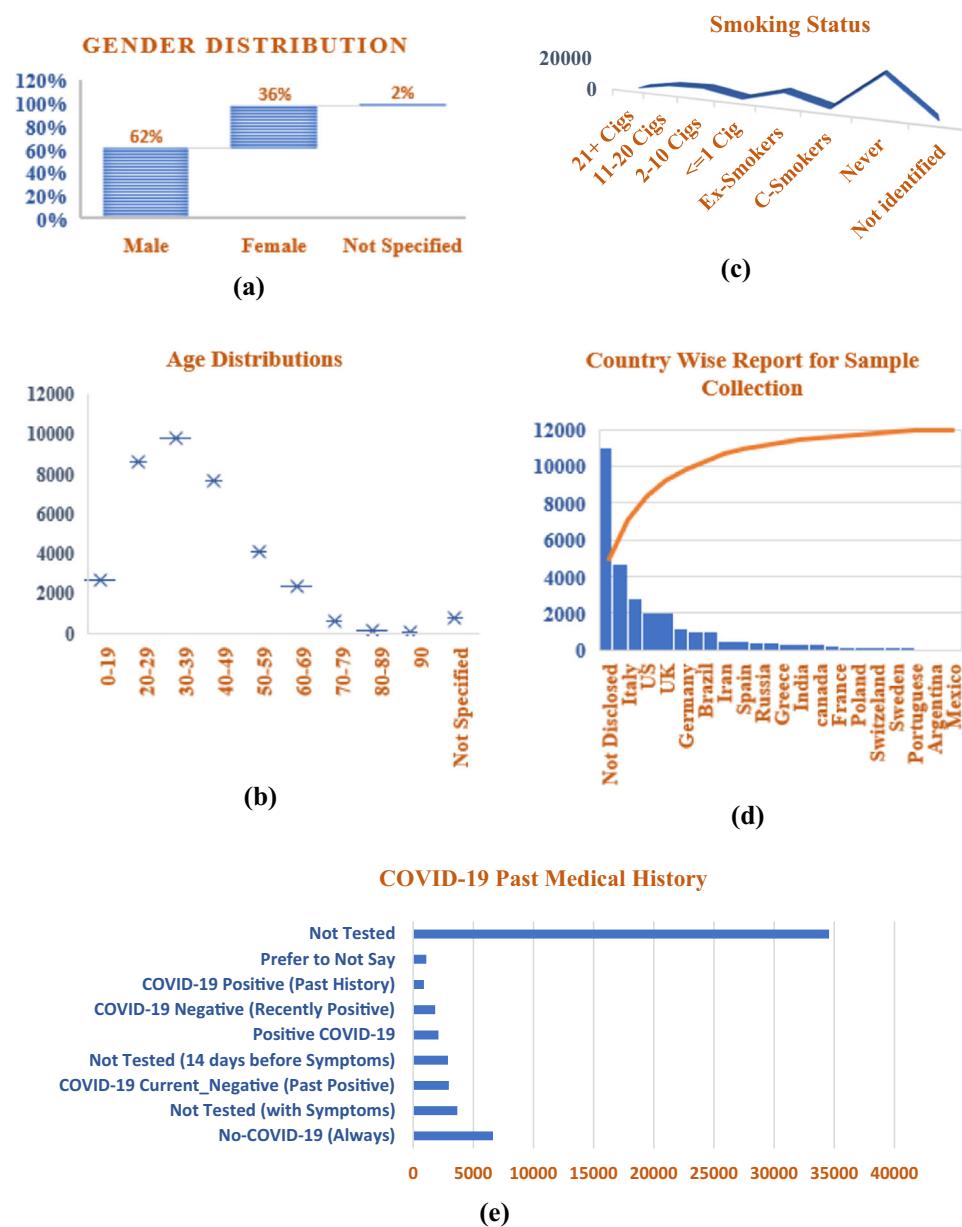
**Table 1** (continued)

Authors	Dataset	Year	Model	Classification		Accuracy (%)
				Tasks	Modalities	
Lella and Pja [35]	COVID-19 Sounds + Augmented Data	2021	OpenSMILE + SVM	OpenSMILE + SVM	Voice	63
				Pre-trained VGGish		59
				Fine-tuned VGGish		69
				OpenSMILE + SVM	Fusion	74
				Pre-trained VGGish		67
				Fine-tuned VGGish		75
				OpenSMILE + SVM	Task-2: COVID-19 symptoms prediction	Breath
				Pre-trained VGGish		59
				Fine-tuned VGGish		62
				OpenSMILE + SVM	Cough	62
Lella and Pja [23]	COVID-19 Sounds + Augmented Data	2022	1D CNN + DDAE + Data Augmentation	Pre-trained VGGish		62
				Fine-tuned VGGish		66
				OpenSMILE + SVM	Voice	52
				Pre-trained VGGish		61
				Fine-tuned VGGish		61
				OpenSMILE + SVM	Fusion	64
				Pre-trained VGGish		64
				Fine-tuned VGGish		71
				COVID-19 vs healthy signs	Breath + cough	89.36
				COVID-19 vs healthy signs	Cough	88.48
Lella and Pja [36]	COVID-19 Sounds + Augmented Data	2022	Light-Weight CNN	COVID-19 vs asthma	Cough	88.86
				Asthma vs healthy	Breath and speech	84.96
				Asthma vs healthy	Cough	85.89
				Asthma disease	Cough, breath, and voice	89.78
				COVID-19 disease		92.32
				Pertussis disease		89.69
				Bronchitis disease		88.74
				Healthy symptoms		93.65
				COVID-19 vs asthma	Breath + cough	93.43
				COVID-19 vs other respiratory sounds	Cough + breath + voice	96.96

**Fig. 1** Statistical analysis of COVID-19 sounds dataset (NeurlPs2021-data): **a** analysis of gender distribution (male—62%, female—36%, not specified—2%), **b** analysis of age for COVID-19 sounds dataset (most of the users are from 20 to 50 years old), **c** smoking status of the participants is 19,776 users never smoked cigarettes, 8718 cigarette smokers (407 will smoke more than 20 cigarettes per day, 3017 will smoke 11–20 cigarettes every day, 3995 users smoke around 1–10 cigarettes, and 1245 users will smoke less than one cigarette daily), **d** analysis of the COVID-19 sounds data from country-wise (Greece—393, Russia—407, Spain—522, Iran—987, Brazil—1016, Germany—1191, UK—2019, US—2829, Italy—4698, etc.,), **e** COVID-19 past medical history have been categorized nine main groups [no-COVID-19 (always), not tested (with symptoms), COVID-19 current negative (past positive), not tested (14 days before symptoms), COVID-19 positive, negative COVID-19 (recently positive), positive COVID-19 (past history), prefer to not say, and not tested]

these data was collected from Cambridge University (covid-19-sounds@cl.cam.ac.uk) with an agreement for research purposes, and the remaining data were collected from publicly available sources and local hospitals. The dataset includes individual participants' self-reported COVID-19 status and another broad spectrum of health issues and demographics. The newly prepared dataset consists of 7261 COVID-19-positive samples in 90,341 with 595 h 33 m duration [48].

We have classified the total respiratory sound data (breath, voice, and cough) into 10 different classes, which are 001—COPD (111 samples),

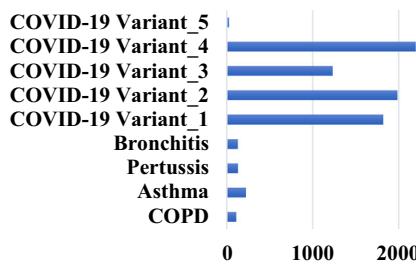


002—Asthma (224 samples), 003—Pertussis (131 samples), 004—Bronchitis (129 samples), 005—COVID-19 Variant\_1\_Alpha (1820 samples), 006—COVID-19 Variant\_2\_Beta (1985 samples), 007—COVID-19 Variant\_3\_Gamma (1232 samples), 008—COVID-19 Variant\_4\_Delta (2198 samples), 009—COVID-19 Variant\_5\_Omicron (26 samples), and 010—Healthy Symptoms (82,485 sound samples), and the analysis of this respiratory disease data from human respiratory sounds is represented in Fig. 2. We applied the proposed model to this benchmark dataset and achieved state-of-the-art results. The authors believe that this proposed model offers a lot of promise for speeding up research

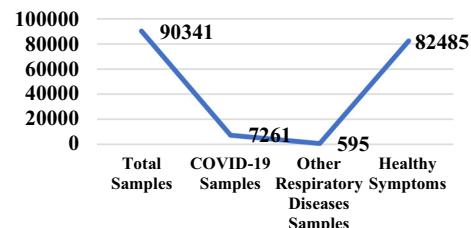
**Table 2** The Systematic analysis of various data samples, modalities, participants, and COVID-19-positive cases

Dataset	#Part	#Sam	#Posi	#Dur (h/m)	#Mod	#A	#G	#S	#C	#Other	#Anno
Virufy	16	121	48	5 m	C					ST, DT	CV
COVID-19-cough	N/A	1324	682	58 m	C	—	—	—	—	N/A	CV*, SR
Coswara	N/A	2030	343	50 h	B, C, V					ST, DT	SR
Tos COVID-19	2758	5867	2926	7 h 30 m	C	—	—	—	—	DT	CV
COUGHVID	N/A	27,550	1156	35 h	C		*	*		DT, CN	SR, EL
COVID-19 Sounds	36,116	53,449	2106	552 h	B, C, V					ST, DT, L, CN	SR
Respiratory COVID-19 Data*	N/A	90,341	7261	595 h 33 m	B, C, V					ST, DT, L, CN	SR, EL
(combined data and a few new samples collected from local hospitals)											

#Part participants, #Sam samples, #Posi positive, #Dur duration (h—hours/m—minutes), #Mod modalities (B—breath, C—cough, V—voice), #A age, #G gender, #S symptoms, #C comorbidities, #Other (ST smoking status, DT date and time, CN country name, L language), # Anno annotation (CV—clinically validated, SR—self-reported, EL—expert labeled, \*—partially validated), with, \* with or without, N/A not applicable

**Respiratory Disease Samples**

(a)

**Comparison of COVID-19 Disease, Other Respiratory Disease, and Healthy Symptoms**

(b)

**Fig. 2** Analysis of human respiratory disease from the dataset [(a) a comparative analysis of COVID-19 disease (variant\_1\_Alpha, variant\_2\_Beta, variant\_3\_Gamma, variant\_4\_Delta, and variant\_5\_Omicron) with other respiratory diseases (COPD, Asthma, Pertussis, and Bronchitis), and (b) comparative analysis of total samples with COVID-19 disease and other respiratory diseases]

in the rapidly emerging field of respiratory sound-based Artificial Intelligence (AI) framework for human health.

### 3.1.1 Data augmentation

The model has been tested with 5 different augmentation pairs, and the results are listed below. Each distortion is applied directly to the respiratory sound signal before being translated into the input feature spectrogram, which is frequently applied to train neural networks [49, 50]. It is important to note that we set the acceleration factors for every augmentation carefully to ensure that better operational validity is retained. The deformations and augmentation sets are defined in the

following: (1) time stretch—it is used to change the running pitch of a sample sound signal without changing it (1.06, 0.80, 1.24, and 0.94 factors are used in this package). (2) Pitch Shift1—without affecting the pitch, the respiratory sound samples can be increased or decreased (1, 2, -2, -1 factors have been used in this package). (3) Pitch Shift2 (PS2)—we intended to create a 2nd augmentation factors' set by considering our first test on Pitch Shift1 revealed that shifting of the pitch tone was very useful to boost. Four maximum values were added to each sample's pitch (2.50, 3.50, -3.50, -2.50). (4) Range comparison dynamically (RCD)—these factors (2.50, 3.50, -3.50, -2.50) were compressed online, one from the 'ICECAST' broadcasting server, and the other three from normal Dolby E.

It is a stereo pair of digital soundtracks that can be processed as a digital sound stream. (5) Background noise—combine the sample with another additive noises in a series of several kinds of sound files; for each sample, 4 background noised audio sounds were merged into one (with collected respiratory sound data—the authors blended a few ambient sound noises and other human respiratory generated sounds). ‘ $C_n$ ’ is the generated as mixed or combined value, which is illustrated in Eq. (1)

$$C_n = ((1 - r_w) \times s) + (r_w * b_n), \quad (1)$$

where  $C_n$  is the combined or mixed value after adding background noise with the original respiratory sound signal,  $r_w$  is the random weighted parameter with the 0.1, 0.5 factors, ‘ $b_n$ ’, is the bias value, and ‘ $s$ ’ is the respiratory original sound sample. The augmentations were applied using the MUDA library, which is something the user is directed for further details on how every distortion is carried out. MUDA picks the respiratory sound file, and the JAMS (A JSON Annotated Music Specification) setup list that goes with it, and then generates the deformed audio as well as better JAMS data that includes all of the wrapping factors. In this work, we have converted the actual observations provided with the assessment data into a JAM (A JSON Annotated Music) file, which is freely available online (open-source library).

### 3.2 Experimental setup

All of the studies were carried out in the Python 3.6.8 language environment. Several API packages were utilized for training two RDCNN models from scratch (Model-1 and Model-2). This research focuses on implementing two RDCNN models (Model-1 and Model-2), applying data augmentation methods to balance the data, and examining how the model performs on augmented and actual data. To extract the deep sound features to test the model in this experiment, we have used three different feature channels (Soft-Mel frequency spectrum, Log-Mel frequency spectrum, and MMFCC—Modified Mel-frequency Cepstral Coefficient spectrum). Various libraries are extracted from the open source in our implementation, which is briefly described below.

#### a. ANACONDA

Python is implemented using the open-source ANACONDA module. NumPy, pandas, and other ML and data science packages have already been loaded. Several popular visualization packages are also included (matplotlib package). It can run on either a Windows or a Linux operating system. It also allows the creation of numerous environments, each with its own set of packages to complete the work.

#### b. LIBROSA

LIBROSA is another essential python package used in this investigation. The library [51] has

been accessed to assess the recordings of respiratory sound signals. Three audio feature extraction approaches are employed in this study, and they can all be found in one package. This library is also running on the offline data pre-processing in this investigation. It can perform complex respiratory sound file interpretation without a visual dataset.

#### c. KERAS

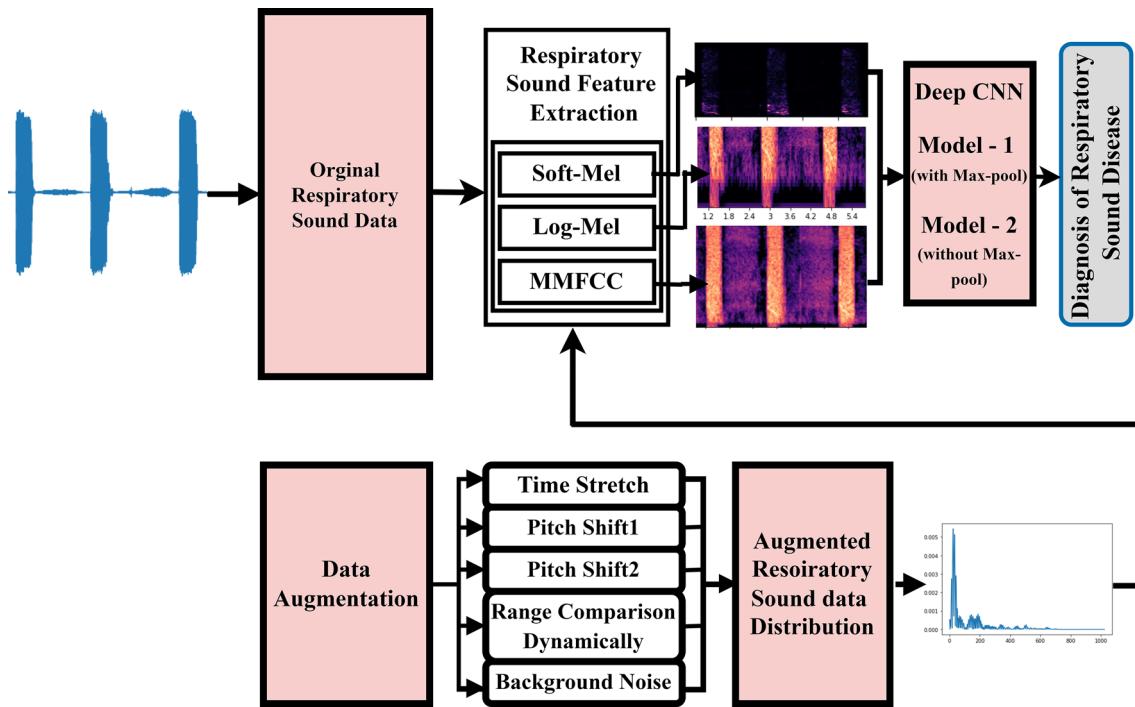
In this research, the regularized deep CNN models (Model-1 and Model-2) are devolved with the KERAS library [52]. It helps to easily apply max-pooling, activation function, drop or create new layers for the model. It helps to classify the respiratory sound disease with extracted features from the pre-trained data.

### 3.3 Methodology

Three respiratory disease audio sound recording datasets are used in this experiment. Two Regularized Deep CNN (RDCNN) models are developed in the study to classify pulmonary sounds. The max-pooling operation is applied for Model-1, and Model-2 is implemented with and without max-pooling operation. In last year, the use of RDCNN to identify unique audible sounds has grown significantly in popularity. Many researchers used RDCNN to create their sound classification models using various methodologies [36, 50]. The concept of pulmonary sound features classification is also addressed in this study, which employs three audio feature extraction methods: Soft-Mel frequency spectrum, Log-Mel frequency spectrum, and MMFCC frequency spectrum using RDCNN. Five distinct audio data augmentation techniques (the data augmentation is described in Sect. 3.1.1 of this experiment) are applied to boost the accuracy of the RDCNN framework on COVID-19 data. Model-1 and Model-2 of this RDCNN (with and without max-pooling function) are applied to these supplemented datasets in the same scenario. The resulting experimental results reveal significant improvements in testing accuracy. The research framework is depicted in Fig. 3 as a broad block diagram.

#### 3.3.1 Regularized deep convolutional neural network (RDCNN)

The RDCNN model is designed with max-pool (Model-1) and without max-pool (Model-2) operations that are significant to identify respiratory disease symptoms from the respiratory sound features. These two models have been implemented using three major important feature extraction methods (Soft-Mel frequency spectrum, Log-Mel frequency spectrum, and MMFCC) for every dataset. It will help to extract the deep respiratory sound features from the audio sound files. The audio data augmentation technique is also used in our implementation to increase the efficiency and reliability of the suggested models after applying feature



**Fig. 3** The proposed methodology block diagram with various feature extraction techniques and data augmentation approaches

extraction spectrums to three datasets. The input vector size of the MMFCC spectrogram is (20,128), and the remaining two spectrums (Soft-Mel frequency spectrum, Log-Mel frequency spectrum) input vector size is (128,128). The representation of the input layer is ‘ $X$ ’, and  $\mathbb{F}(\cdot|\Theta)$  is a component non-linear operation ‘ $\Theta$ ’ represents its parameter function. Introducing of this non-linear operation is to map ‘ $X$ ’ with output estimated value is ‘ $Y$ ’. The functionality of the dense layer and convolutional layer can be represented as Eq. (2)

$$Y = \mathbb{F}(\cdot|\Theta) = f_l(\dots f_3(f_2(X|\Theta_2)|\Theta)|\Theta), \quad (2)$$

where,  $\mathbb{F}_l(\cdot|\Theta_l)$  is used to identify the RDCNN layer of the Model-1 and Model-2. The expression of the dense layer and convolutional layers are shown in Eqs. (3) and (4)

$$Y_l = f_l(X_l|\Theta_l) = Z(K_l X_l + b_i), \quad \Theta_l = [K_l, b_i], \quad (3)$$

$$Y_l = f_l(X_l|\Theta_l) = Z(K_l * X_l + b_i), \quad \Theta_l = [K_l, b_i], \quad (4)$$

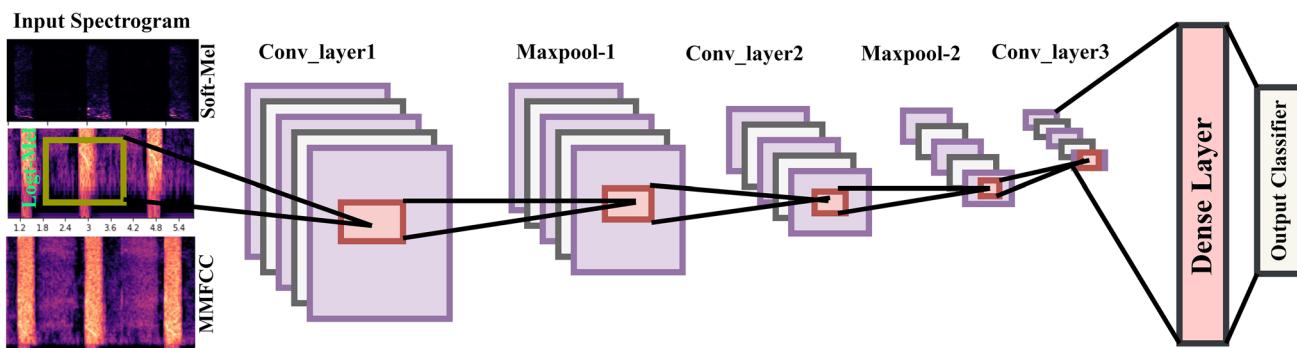
where  $X_l$  is the input vector, its filter or kernel is represented as  $K_l$ , convolutional operation is represented as ‘ $*$ ’, the non-linear function is  $\Theta_l$ , the activation function with pointwise is represented as  $Z(\cdot)$ ,  $Y_l$  is for output prediction,  $b_i$  is vector bias.

The major parameters that have been used in this analysis are the optimization parameter (Adam Optimizer), the uniform batch size is 32, the number of epochs has been used in this are 69 (we have been

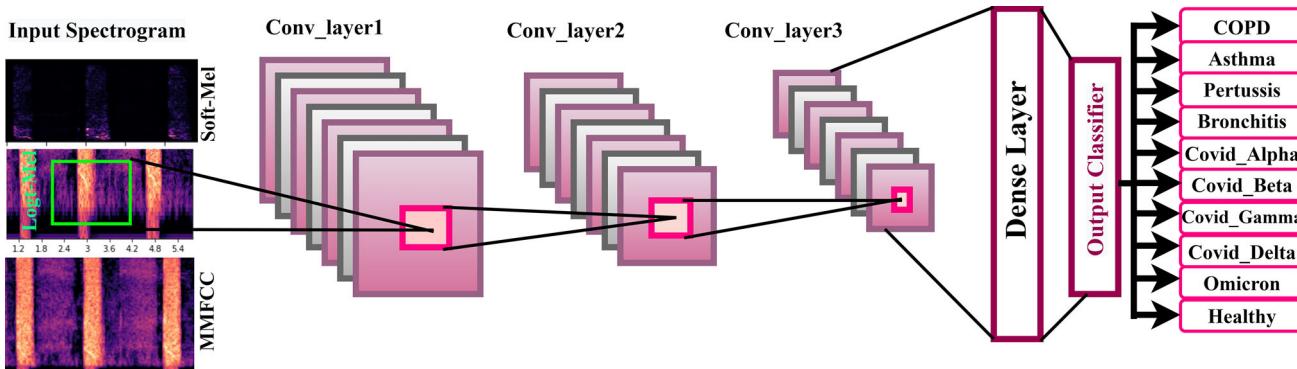
tested with different epochs, but the network is saturating at 69th epoch with 32 batch size), the activation function rectified linear unit (ReLU) has been used to this network for first 4—layers and last layer have been used with ‘softmax’ function, and we have designed the RDCNN model using 0.001 L2 norm regularization in this analysis. The general architecture of Model-1 (with max-pooling) is represented in Fig. 4 and Model-2 (without max-pooling) is illustrated in Fig. 5. A detailed description of Model-1 and Model-2 (with max-pool and without max-pool) has been explained in Sect. 3.3.2.

### 3.3.2 Functional description of RDCNN framework with and without max-pool function

The RDCNN backbone is a convolutional network process. The convolutional filter is applied to the input spectrum in this stage. The convolution is done between the input spectrum and the filter that is being used. Edge detection (ED), pattern recognition (PR), and other similar techniques are used to detect convolved features of the input sound spectrogram. The total convolutional outcome of the filter values with their corresponding input spectrogram values yields these convolved features. After then, this filter moves over the input spectrogram until it reaches the terminus. The value of the stride determines how this filter slides across the image spectrogram. In our Model-2 framework, the deep convolutional feature values are the input to the activation function for the ReLU. The output of this layer is set to be zero if the input value is



**Fig. 4** The convolutional architecture of Model-1 (with max-pooling) with different input spectrograms



**Fig. 5** The convolutional architecture of model-2 (without max-pooling) using various input spectrograms along with ten output classes (COPD, Asthma, Pertussis, Bronchitis, COVID-19\_Alpha\_variant, COVID-19\_Beta\_variant, COVID-19\_Gamma\_variant, COVID-19\_Delta\_variant, COVID-19\_Omicron\_variant, and Healthy sounds class)

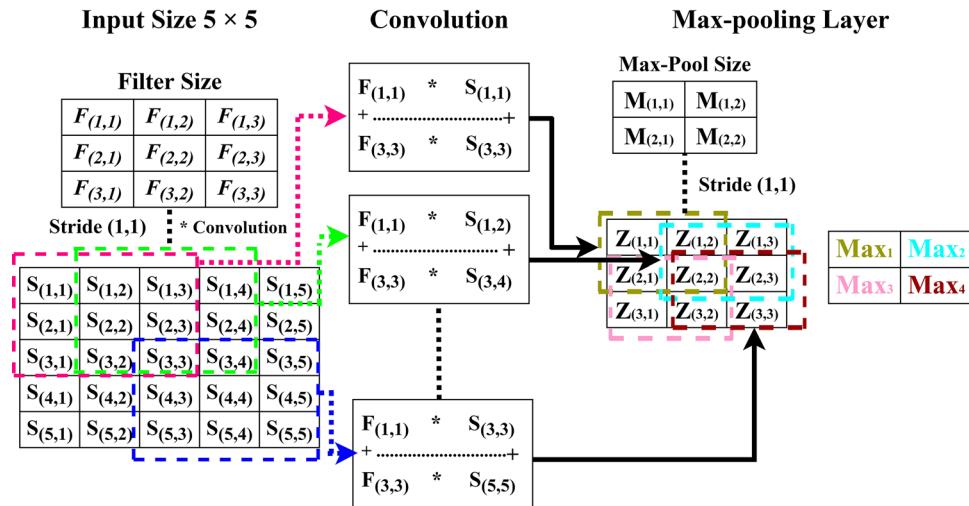
zero. There is no max-pool operation done before the ReLU function and after the convolution operation in Model-2.

The pooling feature is primarily used to aid in the extraction of smooth and sharp features. Dimensionality, computation, and variance can all be reduced. Pooling operations are usually divided into two categories. The most prevalent is max-pooling, which is used to derive features such as points, edges, and so on. Average-pooling is another sort of pooling that aids in the extraction of smooth features. After the convolution process, Model-1 evaluates the max-pool operation in our implementation. The goal of this research is to determine the impact of applying max-pool for Model-1 and not using max-pool for Model-2 on RDCNN without taking into consideration of time limitations. We clearly illustrate the operational functionality of max-pooling and convolutional operations in Fig. 6. The input respiratory sound spectrogram size is  $5 \times 5$  which is represented by 'S', 'F' denoted  $3 \times 3$  filter size, the predicted value is 'Z', the stride is considered is (1,1), and 'M' represents max-pooling with size  $2 \times 2$ .

We are considering breathing, cough, and voice sounds are the primary three input modalities for the RDCNN in both the models (Model-1 and Model-2). Here, the RDCNN models (Model-1 and Model-2) are automatically extracting deep respiratory sound

features from the sound spectrogram. The spectrograms have been generated using three various feature channels (Soft-Mel, Log-Mel, and MFCC). We are considering breathing, cough, and voice sounds are the primary three input modalities for the RDCNN in both the models (Model-1 and Model-2). The RDCNN models (Model-1 and Model-2) derive numerous respiratory sound characteristics such as respiratory sound intensity, respiratory sound frequency, respiratory sound pitch quality, vocal resonance, loudness, voice activity detection (VAD), voice pitch, signal duration, cough peak-flow rate, SNR—signal to noise ratio, SLE—strength of Lombard effect, subglottic pressure, peak velocity-time, acoustic signal, cough expire volume, and air volume from the input modalities. To get an accurate diagnosis, it is essential to separate normal respiratory sounds from abnormal ones such as crackles, wheezes, and pleural rub. In our work, the RDCNN models (Model-1, Model-2) outperformed the existing DL models to regularize the unbalanced dataset of COVID-19 sounds. The model learnable parameter's gradients have vanished in the intermediate layers while optimizing the prediction error, and it leads to the early stopping of the model to train. It is addressed using our proposed RDCNN models. The existing DL models are used for binary classification of the COVID-19 disease. Our proposed model exhibits the best results in the multiclass classification of respiratory diseases.

**Fig. 6** The convolutional neural network operational framework with max-pool function



### 3.3.3 Model-1 architecture (CNN with max-pool function)

- Layer-1: The layer-1 of Model-1 consists of 24 kernels with a  $5 \times 5$  filter size. In this layer, we used L2 norm regularization with a 0.001 value. The activation function used in this layer is ReLU and the max-pooling stride is  $3 \times 3$ .
- Layer-2: The second layer of Model-1 consists of 36-filters with a  $4 \times 4$  kernel size. In this layer, we used L2 norm regularization with a 0.001 value. In this layer, padding is the “same” for the MMFCC feature extraction technique and “valid” for both the Soft-Mel spectrum and Log-Mel spectrum. Then the activation function used in this layer is ReLU, and the Max-pooling stride is  $2 \times 2$ .
- Layer-3: The 3rd layer of Model-1 consists of 48-filters with a  $3 \times 3$  kernel size. In this layer, padding is the “same” for the MMFCC feature extraction technique and “valid” for both the Soft-Mel spectrum and Log-Mel spectrum. Then, the activation function used in this layer is ReLU, and there is no Max-pooling function has been used in this layer.
- Layer-4: The 4th layer (1st dense layer) of this Model-1 framework is made up of 60 hidden units with the ReLU activation parameter. The model is implemented with a 0.5 rate of dropout to avoid the problem of overfitting.
- Layer-5: The layer-5 (second dense layer) of this Model-1 comprises output units that are the same as the variety of diseases classes in the dataset. We have used the “Softmax” classifier in this layer as an activation function.

### 3.3.4 Model-2 architecture (CNN without max-pool operations)

- Layer-1: Layer-1 of Model-2 (without max-pooling) consists of 24 kernels with a  $5 \times 5$  filter size. In this layer, we used L2 norm regularization with a 0.001

value. ReLU activation function has been used in this layer to transform actual data in a non-linear formation.

- Layer-2: The second layer of Model-2 consists of 36-filters with a  $4 \times 4$  kernel size. In this layer, we used L2 norm regularization with a 0.001 value. In this layer, padding is the “same” for the MMFCC feature extraction technique and “valid” for both the Soft-Mel spectrum and Log-Mel spectrum. Then, the activation function used in this layer is ReLU, and no max-pooling function is used at this level.
- Layer-3: The 3rd layer of model-2 consists of 48-filters with a  $3 \times 3$  kernel size. In this layer, padding is the “same” for the MMFCC feature extraction technique and “valid” for both the Soft-Mel spectrum and Log-Mel spectrum. Then, the activation function used in this layer is ReLU, and there is no Max-pooling function has been used in this layer.
- Layer-4: The 4th layer (1st dense layer) of this Model-2 architecture is made up of 60 hidden units with the ReLU activation parameter. The model is implemented with a 0.5 rate of dropout to avoid the problem of extracting the same features from the data (overfitting).
- Layer-5: The layer-5 (second dense layer) of this Model-2 is made up of output units which are the same as the variety of diseases classes in the dataset. We have used the “Softmax” classifier in this layer as an activation function.

## 4 Results and discussion

We have discussed the RDCNN model performance on three datasets (KDD-data, ComParE2021-CCS-CSS-Data, and NeurlPs2021-data) using Model-1 and Model-2 frameworks with respect to the accuracy Eq(5),  $F_1$  score Eq(8), precision Eq(6),

and recall Eq(7) measurements. These measurements can be calculated for each class of ten classes [001—COPD, 002—Asthma, 003—Pertussis, 004—Bronchitis, 005—COVID-19 Variant\_1 (Alpha), 006—COVID-19 Variant\_2 (Beta), 007—COVID-19 Variant\_3 (Gamma), 008—COVID-19 Variant\_4 (Delta), 009—COVID-19 Variant\_5 (Omicron), and 010—Healthy Symptoms]. We have used k-fold validation methods to measure the proposed model performance which is well suited to evaluate the DL and ML techniques

Accuracy

$$= \frac{\text{True Positive Value} + \text{True Negative Value}}{\left( \begin{array}{l} \text{True Positive Value} + \text{False Negative Value} \\ + \text{False Positive Value} + \text{True Negative Value} \end{array} \right)}, \quad (5)$$

Precision (Class = COVID – 19)

$$= \frac{\text{True Positive Value} (\text{Class} = \text{COVID} - 19)}{\left( \begin{array}{l} \text{True Positive Value} (\text{Class} = \text{COVID} - 19) \\ + \text{False Positive Value} (\text{Class} = \text{COVID} - 19) \end{array} \right)}, \quad (6)$$

Recall (Class = COVID – 19)

$$= \frac{\text{True Positive Value} (\text{Class} = \text{COVID} - 19)}{\left( \begin{array}{l} \text{True Positive Value} (\text{Class} = \text{COVID} - 19) \\ + \text{False Negative Value} (\text{Class} = \text{COVID} - 19) \end{array} \right)}, \quad (7)$$

$$F_1 - \text{Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \quad (8)$$

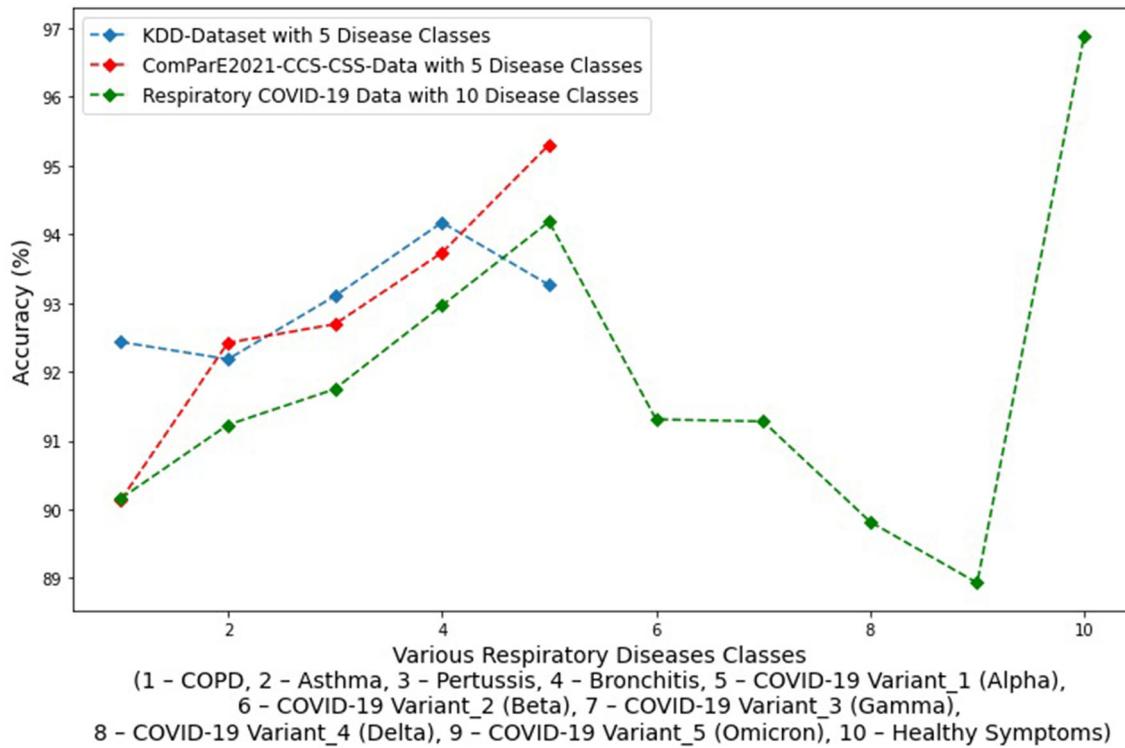
We have deployed the  $L_2$  norm regularization method for both the models (Model-1 and Model-2) to avoid the overfitting problem by fine-tuning the convolutional model parameters such as regulating the number of hidden layers, adjusting dropout rate, applying different learning rates, and implementing different activation functions in different layers. In addition to that, we have examined the number of epochs (1–69 because loss is varying between 1st epoch and 69th epoch) to decrease the model loss. This observation has proven that the proposed model (without max-pooling) gives good accuracy when compared with the existing models on 3 different datasets [1. KDD-data, 2. ComParE2021-CCS-CSS-Data, and 3. Respiratory COVID-19 (newly prepared data)]. The Respiratory COVID-19 (newly prepared data) data are prepared based on NeurlPs2021-data raw dataset, collected data from local hospitals, and extracted from internet sources].

The Model-1 (with max-pooling) analysis and performance on various datasets is shown in Table 3. The model performance has been tested on 3 datasets (1. KDD-data, 2. ComParE2021-CCS-CSS-Data, and 3. Respiratory COVID-19) with respect to the accuracy and  $F_1$  Score. We have balanced KDD-data and ComParE2021-CCS-CSS-data according to the disease class. These observations identify 92.43% accuracy for Asthma disease, 92.18% accuracy for Pertussis, 93.10%

accuracy for Bronchitis, 94.17% accuracy for COVID-19 diseases, and 93.26% accuracy for Healthy respiratory sound symptoms on KDD-data. The accuracies for Asthma, Pertussis, Bronchitis, COVID-19, and Healthy symptoms is 90.14%, 92.42%, 92.69%, 93.73%, and 95.30% on ComParE2021-CCS-CSS-Data. The performance of Model-1 (with max-pooling) for different classes is 90.16% accuracy for COPD, 91.23% accuracy for Asthma, 91.75% accuracy for Pertussis, 92.96% accuracy for Bronchitis, 94.18% accuracy for COVID-19 Variant\_1 (Alpha), 91.31% accuracy for COVID-19 Variant\_2 (Beta), 91.28% accuracy for COVID-19 Variant\_3 (Gamma), 89.82% accuracy for COVID-19 Variant\_4 (Delta), 88.93% accuracy for COVID-19 Variant\_5 (Omicron), and 96.89% accuracy for Healthy respiratory sound symptoms on newly extracted dataset (Respiratory COVID-19 Data). The result evaluation analysis is represented in Fig. 7 on different datasets for various diseases.

**Table 3** The Model-1 (with max-pooling) analysis and performance on various datasets

Dataset	Respiratory disease name	Accuracy (%)	$F_1$ score (%)
KDD-data	Asthma	92.43	93.12
	Pertussis	92.18	93.32
	Bronchitis	93.10	93.98
	COVID-19	94.17	95.26
	Healthy symptoms	93.26	94.89
ComParE2021-CCS-CSS-data	Asthma	90.14	91.56
	Pertussis	92.42	93.32
	Bronchitis	92.69	93.14
	COVID-19	93.73	94.15
	Healthy symptoms	95.30	96.42
Respiratory COVID-19 data	COPD	90.16	90.89
	Asthma	91.23	92.56
	Pertussis	91.75	92.38
	Bronchitis	92.96	93.87
	COVID-19 Variant_1 (Alpha)	94.18	95.50
	COVID-19 Variant_2 (Beta)	91.31	92.18
	COVID-19 Variant_3 (Gamma)	91.28	91.96
	COVID-19 Variant_4 (Delta)	89.82	90.53
	COVID-19 Variant_5 (Omicron)	88.93	89.42
	Healthy symptoms	96.89	97.63



**Fig. 7** The result analysis Model-1 (with max-pooling) for various diseases on different datasets

**Table 4** The Model-2 (without max-pooling) analysis and performance on various datasets

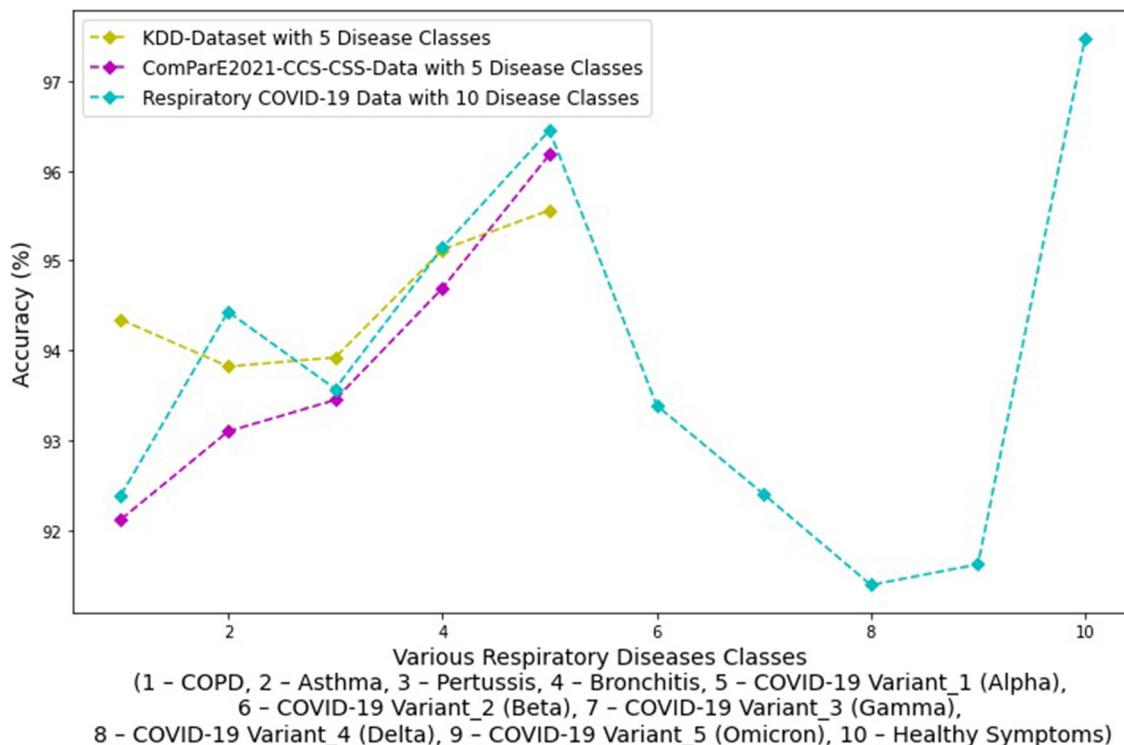
Dataset	Respiratory disease name	Accuracy	F <sub>1</sub> Score (%)
KDD-data	Asthma	94.34	95.56
	Pertussis	93.82	94.64
	Bronchitis	93.92	94.63
	COVID-19	95.12	96.37
	Healthy symptoms	95.56	96.36
ComParE2021-CCS-CSS-Data	Asthma	92.12	93.35
	Pertussis	93.10	94.16
	Bronchitis	93.45	94.62
	COVID-19	94.69	95.13
	Healthy symptoms	96.18	97.20
Respiratory COVID-19 Data (large scale data + COVID-19 data)	COPD	92.39	93.84
	Asthma	94.43	95.69
	Pertussis	93.57	94.43
	Bronchitis	95.14	95.86
	COVID-19 Variant_1 (Alpha)	96.45	97.96
	COVID-19 Variant_2 (Beta)	93.39	93.94
	COVID-19 Variant_3 (Gamma)	92.40	93.82
	COVID-19 Variant_4 (Delta)	91.39	92.12
	COVID-19 Variant_5 (Omicron)	91.62	92.83
	Healthy Symptoms	97.48	98.12

The Model-2 (without max-pooling) analysis and performance on various datasets is shown in Table 4. The model performance has been tested on 3 datasets (1. KDD-data, 2. ComParE2021-CCS-CSS-data, and 3. Respiratory COVID-19) by concerning the accuracy and  $F_1$  Score. We have balanced KDD-data and ComParE2021-CCS-CSS-Data according to the disease class by applying data augmentation technique. These observations identify 94.34% accuracy for Asthma disease, 93.82% accuracy for Pertussis, 93.92% accuracy for Bronchitis, 95.12% accuracy for COVID-19 diseases, and 95.56% accuracy for Healthy respiratory sound symptoms on KDD-data. The accuracies for Asthma, Pertussis, Bronchitis, COVID-19, and Healthy symptoms is 92.39%, 93.10%, 93.45%, 94.69%, and 96.18% on ComParE2021-CCS-CSS-Data. The performance of Model-2 (with max-pooling) for different classes is 92.12% accuracy for COPD, 94.43% accuracy for Asthma, 93.57% accuracy for Pertussis, 95.14% accuracy for Bronchitis, 96.45% accuracy for COVID-19 Variant\_1 (Alpha), 93.39% accuracy for COVID-19 Variant\_2 (Beta), 92.40% accuracy for COVID-19 Variant\_3 (Gamma), 91.39% accuracy for COVID-19 Variant\_4 (Delta), 91.62% accuracy for COVID-19 Variant\_5 (Omicron), and 97.48% accuracy for healthy respiratory sound symptoms on newly extracted dataset (Respiratory COVID-19 Data). The result evaluation analysis is represented in Fig. 8 on different datasets for various diseases. These observations show that the Model-2 (without max-pooling) is giving

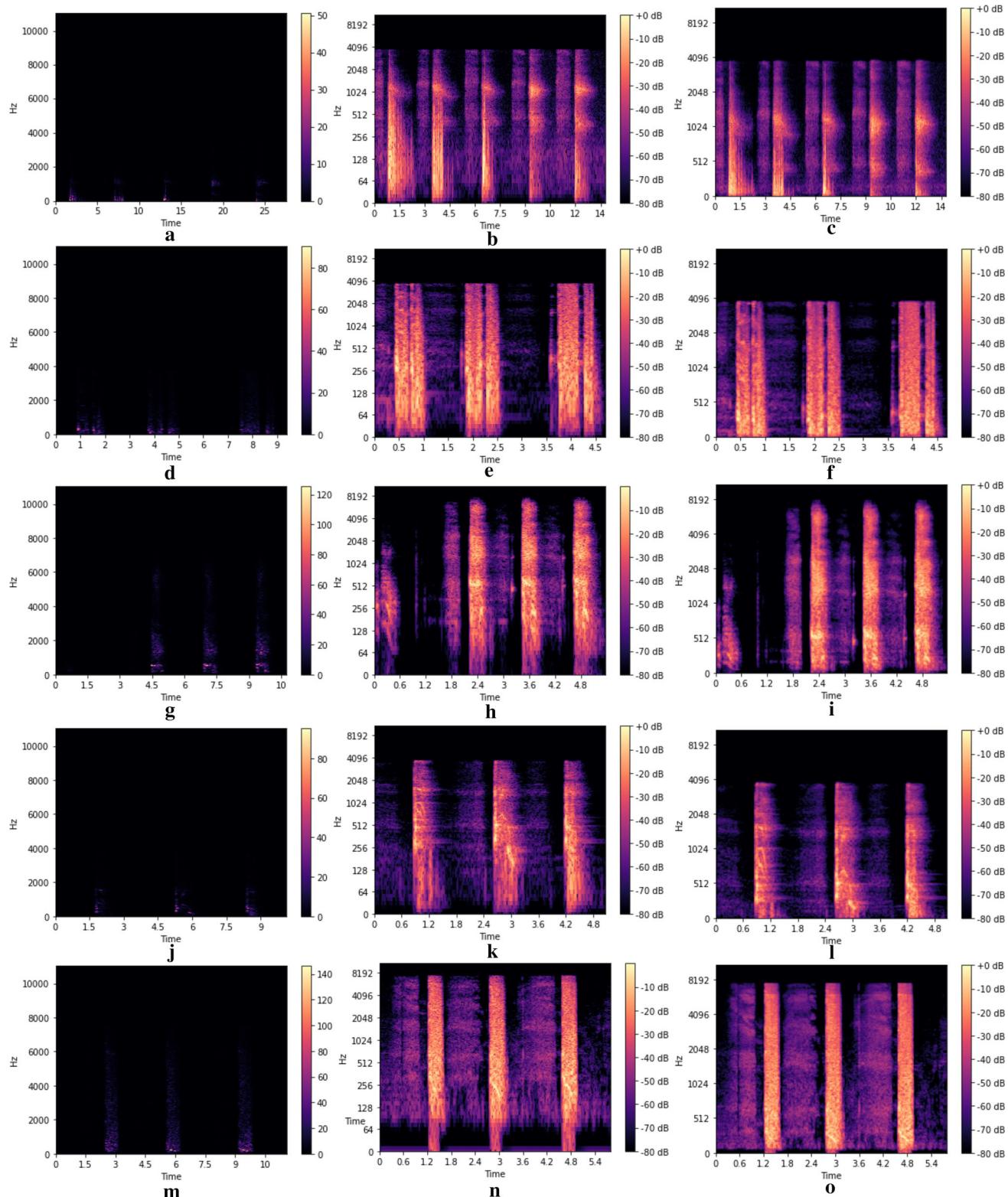
2–3% better accuracy for different respiratory diseases than Model-1 (with max-pooling) performance.

The sample data analysis has done to observe the respiratory sound spectrograms with three different spectrums (1. Soft-Mel spectrum, 2. Log-Mel spectrum, and 3. Modified Mel-frequency spectrum). The analysis shows the variations between the various respiratory diseases and SARS-CoV-2 disease using respiratory sound data. We have applied all three spectrums to analyze respiratory sound disease data, and observed that MMFCC attains better performance than remaining two feature channels to extract the deep respiratory sound features. The feature extraction spectrum analysis for COPD, asthma, pertussis, bronchitis, and COVID-19 diseases is presented in Fig. 9.

Figure 9 (a) shows the Soft-Mell spectrum for COPD disease; (b) represents the Log-Mel spectrogram for COPD disease; (c) presents the MMFCC Spectrogram for COPD disease; (d) depicts the Soft-Mell spectrum for Asthma disease sound; (e) represents the Log-Mel spectrogram for Asthma disease sound; (f) MMFCC Spectrogram for Asthma disease sound; (g) shows the Soft-Mell spectrum for Pertussis disease sound; (h) represents the Log-Mel spectrogram for (i) presents the MMFCC Spectrogram Pertussis disease sound; (j) depicts the Soft-Mell spectrum for Bronchitis disease sound; (k) Log-Mel spectrogram for Bronchitis disease sound; (l) shows the MMFCC Spectrogram for Bronchitis disease sound; (m) presents the Soft-Mell spectrum for COVID-19 disease sound; (n) depicts the Log-Mel



**Fig. 8** The result analysis Model-2 (without max-pooling) for different disease on various datasets



**Fig. 9** The feature extraction spectrum analysis for COPD, Asthma, Pertussis, Bronchitis, and COVID-19 diseases

**Table 5** Comparison of model metrics (precession, recall, accuracy, and  $F_1$  score) on respiratory COVID-19 data with Model-1 (with max-pooling) and Model-2 (without max-pooling)

Model name	Disease class name	Precession (%)	Recall (%)	Accuracy (%)	$F_1$ score (%)
Deep CNN Model-1 (with Max-pooling)	COPD	94	96	91	93
	Asthma	95	97	92	92
	Pertussis	94	96	91	92
	Bronchitis	95	97	92	93
	COVID-19 Variant_1 (Alpha)	96	98	94	95
	COVID-19 Variant_2 (Beta)	91	93	91	92
	COVID-19 Variant_3 (Gamma)	91	93	91	91
	COVID-19 Variant_4 (Delta)	90	92	89	90
	COVID-19 Variant_5 (Omicron)	90	92	88	89
	Healthy symptoms	97	99	96	97
Deep CNN Model-1 (without Max-pooling)	COPD	95	97	92	93
	Asthma	96	98	94	95
	Pertussis	95	97	93	94
	Bronchitis	95	97	95	95
	COVID-19 Variant_1 (Alpha)	97	99	96	97
	COVID-19 Variant_2 (Beta)	94	96	93	93
	COVID-19 Variant_3 (Gamma)	93	95	92	93
	COVID-19 Variant_4 (Delta)	92	94	91	92
	COVID-19 Variant_5 (Omicron)	91	93	90	91
	Healthy symptoms	97	99	97	98

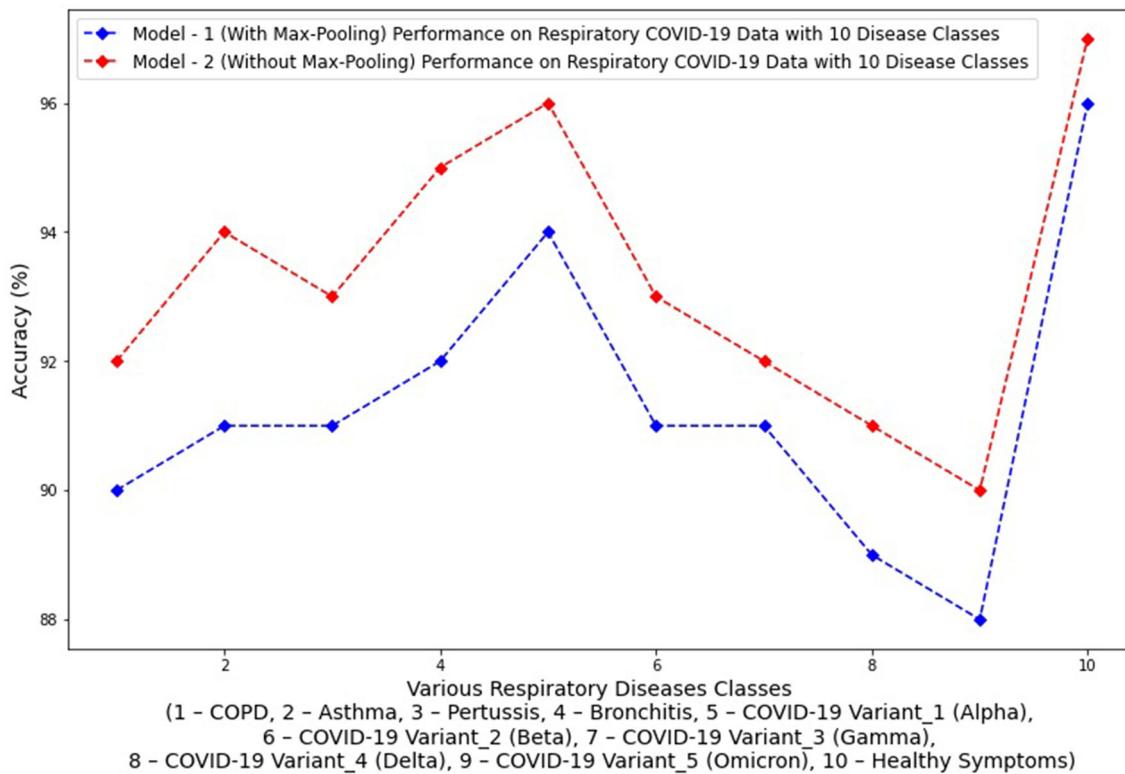
spectrogram for COVID-19 disease sound; (o) represents the MFCC Spectrogram for COVID-19 disease sound.

The performance of model metrics on newly prepared data (Respiratory COVID-19) with Model-1 and Model-2 is represented in Table 5. In this analysis, we have evaluated precession, recall, accuracy, and  $F_1$ -Score for the both the models. The model accuracy

for COPD is 90%, Asthma accuracy is 91%, the accuracy for Pertussis disease class is 91%, Bronchitis accuracy is 92%, accuracy for COVID-19 Variant\_1 (Alpha) is 94%, COVID-19 Variant\_2 (Beta) accuracy is 91%, and the model accuracy for COVID-19 Variant\_3 (Gamma) is 91%, 89% of COVID-19 Variant\_4 (Delta) accuracy, 88% of COVID-19 Variant\_5 (Omicron) disease class accuracy, and 96% of Healthy sounds

**Table 6** Comparison analysis of proposed model and dataset with existing models and datasets

Dataset	Model	Tasks/classes	Accuracy (%)
KDD-Data (COVID-19 sounds data)	SVM [40]	COVID-19 Detection with cough + breath	80
		COVID-19 Detection with cough	82
		COVID-19 vs asthma with breath	80
KDD-Data (COVID-19 sounds data)	VGG-Net [40]	COVID-19 detection with cough	87
		COVID-19 vs asthma with breath	88
ComParE2021-CCS-CSS-Data (COVID-19 sounds data)	COMPARE functionals + SVM CNN + LSTM RNN Fusion of all [47]	COVID-19 detection	72
			69
			74
KDD-Data (COVID-19 sounds data)	1D CNN [35]	COVID-19 detection	90
KDD-Data (COVID-19 sounds data)	Light-Weight CNN [23]	COVID-19 detection	93
COVID-19 Sounds Data (large-scale audio dataset)	Fine-tuned VGG-Net [48]	COVID-19 detection with cough + breath	75
		COVID-19 detection with cough	71
Respiratory COVID-19 data (large scale data + COVID-19 data + augmented data + a few Samples collected from the local hospitals) (this research)	Deep CNN Model-1 (with Max-pooling)	COPD	90
		Asthma	91
		Pertussis	91
		Bronchitis	92
		COVID-19 Variant_1 (Alpha)	94
		COVID-19 Variant_2 (Beta)	91
		COVID-19 Variant_3 (Gamma)	91
		COVID-19 Variant_4 (Delta)	89
		COVID-19 Variant_5 (Omicron)	88
		Healthy Symptoms	96
	Deep CNN Model-1 (without max-pooling)	COPD	92
		Asthma	94
		Pertussis	93
		Bronchitis	95
		COVID-19 Variant_1 (Alpha)	96
		COVID-19 Variant_2 (Beta)	93
		COVID-19 Variant_3 (Gamma)	92
		COVID-19 Variant_4 (Delta)	91
		COVID-19 Variant_5 (Omicron)	90
		Healthy Symptoms	97



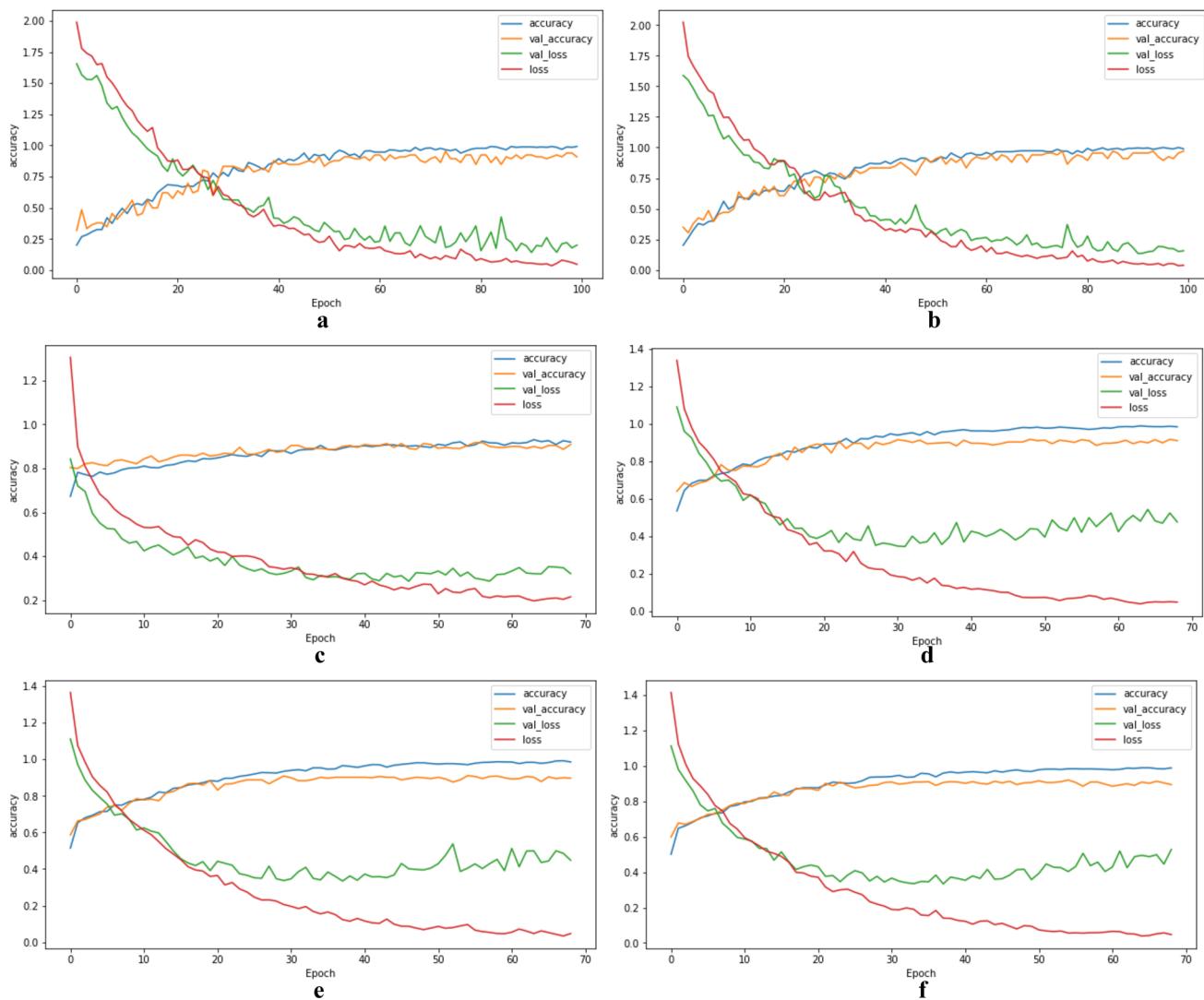
**Fig. 10** The performance of Model-1 and Model-2 on Respiratory COVID-19 Data with 10 different diseases classes

class accuracy with Model-1. The model accuracies with Model-2 (without max-pooling) for different disease classes are COPD—92%, Asthma—94%, Pertussis—93%, Bronchitis—95%, COVID-19 Variant\_1 (Alpha)—96%, COVID-19 Variant\_2 (Beta)—93%, COVID-19 Variant\_3 (Gamma) —92%, COVID-19 Variant\_4 (Delta)—91%, COVID-19 Variant\_5 (Omicron)—90%, and Healthy Symptoms—97%. We observed that the Model-2 is giving 2–3% better results compared with Model-1. The accuracy comparison analysis for Model-1 and Model-2 is depicted in Fig. 10.

Testing accuracies for each model (Model-1 with max-pool and Model-2 without max-pool) with three Mel-spectrogram features channels (MMFCC—Modified Mel-Frequency Cepstral Coefficients, Log-Mel Spectrum, and Soft-Mel Spectrum) is represented in Fig. 11. We have evaluated the model testing accuracy, validation loss, loss, and validation accuracy for both the models. The model test accuracies have been compared with two kernel sizes ( $3 \times 3$ ,  $5 \times 5$ ) for both the models (Model-1 and Model-2). It shows that the  $3 \times 3$  kernel size filter is giving 2–3% better accuracy than  $5 \times 5$  filter size. From this analysis, in fact the accuracy is falling little down for COVID-19 Variant\_2 (Beta), COVID-19 Variant\_3 (Gamma), COVID-19 Variant\_4 (Delta), and COVID-19 Variant\_5 (Omicron) disease classes as a consequence of imbalanced data. We will focus more on data balancing and improving model performance in the future works.

The comparison of proposed model with existing approaches on respiratory disease data is represented in Table 6. The exiting approaches have mostly focused on identification of COVID-19, Pertussis, Bronchitis, and asthma diseases. The proposed approaches are focused COPD disease, Asthma disease, Pertussis, Bronchitis, and COVID-19 with different variants (Alpha, Beta, Gamma, Delta, and Omicron). The regularized deep CNN Model-1 (with max-pooling) attains performances 90%, 91%, 91%, 92%, 94%, 91%, 91%, 89%, and 96% for COPD, Asthma, Pertussis, Bronchitis, COVID-19 Variant\_1 (Alpha), COVID-19 Variant\_2 (Beta), COVID-19 Variant\_3 (Gamma), COVID-19 Variant\_4 (Delta), COVID-19 Variant\_5 (Omicron), and Healthy Symptoms classes.

The Model-2 (without max-pooling) attains good accuracies for COPD, Asthma, Pertussis, Bronchitis, COVID-19 Variant\_1 (Alpha), COVID-19 Variant\_2 (Beta), COVID-19 Variant\_3 (Gamma), COVID-19 Variant\_4 (Delta), COVID-19 Variant\_5 (Omicron), and Healthy Symptoms classes than Model-1. The best accuracies with each model to detect the COVID-19 respiratory disease are depicted in Fig. 12. We have compared the proposed RDCNN models (Model-1 and Model-2) accuracies with existing approaches such as SVM, VGG-Net (Visual Geometry Group Network), LSTM (Long Short-Term Memory), 1D CNN (One Dimensional Convolutional Neural Networks), and Light-Weight CNNs.

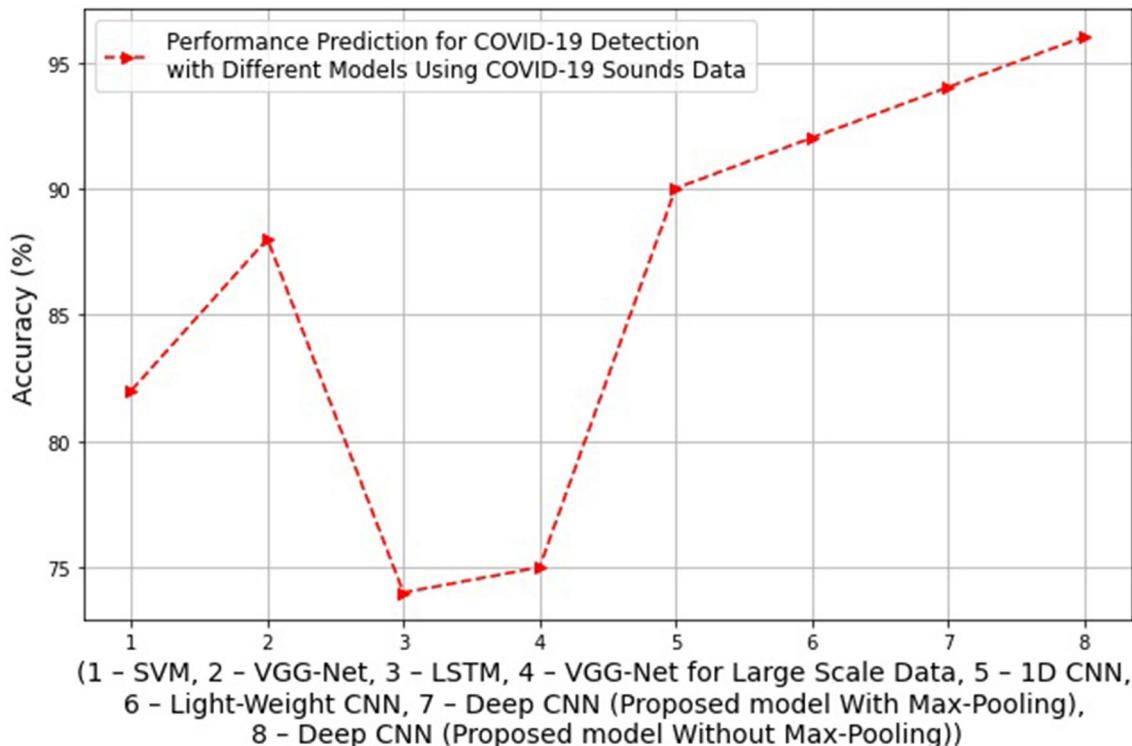


**Fig. 11** Testing accuracies for Model-1 (with max-pooling) and Model-2 (without max-pooling) with three Mel-spectrogram features channels (MMFCC—Modified Mel-Frequency Cepstral Coefficients, Log-Mel Spectrum, Soft-Mel Spectrum). **a** Testing accuracy for Model-1 with MMFCC, **b** testing accuracy for Model-2 with MMFCC, **c** testing accuracy for Model-1 with Log-Mel Spectrum, **d** testing accuracy for Model-2 with Log-Mel Spectrum, **e** testing accuracy for Model-1 with Soft-Mel Spectrum, and **f** testing accuracy for Model-2 with Soft-Mel Spectrum

## 5 Conclusion

The focus of this experimental research is to examine the performance of the RDCNN model on COVID-19 data sets (KDD-data, ComParE2021-CCS-CSS-Data, and Large scale Respiratory COVID-19 data) and to understand the COVID-19 virus dissemination in the five variants. We have examined the COVID-19 sound data by our proposed framework with two RDCNN models (Model-1 and Model-2) that are implemented with and without max-pool operations to identify COVID-19 symptoms. These two models have been observed by applying three different feature extraction methods (Soft-Mel frequency channel, Log-Mel frequency spectrum, and MMFCC spectrum) on actual

data, and we have examined the same models on augmented data. The final results have demonstrated that the RDCNN model with max-pool operation using MMFCC is giving a state-of-the-art performance on three benchmark datasets to identify COVID-19 symptoms. We compared the RDCNN models with two receptive fields ( $1 \times 3$  and  $1 \times 5$ ). The experiments suggested that the  $1 \times 3$  receptive field is giving better performance than the  $1 \times 5$  receptive field. Model-2 with MMFCC achieves 2–3% greater accuracy than that existing DCNN with the MFCC method, and it gives immense performance to detect the respiratory disease with the respiratory sound data. The study also concentrates on the prominence of the utility of Neural Architecture Search (NAS) in determining superior framework classification. We will concentrate to



**Fig. 12** Comparison of best accuracies using different models on COVID-19 sounds data to detect COVID-19 disease

improve the model performance by employing the residual neural network (ResNet) model even on the data imbalance to identify the respiratory diseases.

**Acknowledgements** We would like to thank each and every one who all are trying to stop spreading of SARS-CoV-2 pandemic.

**Data availability statement** This manuscript contains data linked with it that can be found in a data repository. [Authors' comment: The data presented in this work are not available publicly, we have collected COVID-19 large-scale sounds (breath, voice, and cough) data from trusted repositories and cited sources in references [23, 36, 40, 47, 48].] The data are available with the corresponding author and can be given to the researchers and scientists with the reasonable request.

#### Declarations

**Conflict of interest** All the authors have no conflict of interest.

#### References

- E. Saatci, E. Saatci, Determination of respiratory parameters by means of hurst exponents of the respiratory sounds and stochastic processing methods. *IEEE Trans. Biomed. Eng.* **68**(12), 3582–3592 (2021). <https://doi.org/10.1109/TBME.2021.3079160>
- Y. Wang et al., Unobtrusive and automatic classification of multiple people's abnormal respiratory patterns in real time using deep neural network and depth camera. *IEEE Internet Things J.* **7**(9), 8559–8571 (2020). <https://doi.org/10.1109/JIOT.2020.2991456>
- World Health Organization. Coronavirus disease 2019 (covid-19). (2021). <https://covid19.who.int/>. Accessed 12 Mar 2022
- J. Wedel, P. Steinmann, M. Štrakl et al., Can CFD establish a connection to a milder COVID-19 disease in younger people? Aerosol deposition in lungs of different age groups based on Lagrangian particle tracking in turbulent flow. *Comput Mech* **67**, 1497–1513 (2021). <https://doi.org/10.1007/s00466-021-01988-5>
- R. Vaishya, M. Javaid, I.H. Khan, A. Haleem, Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab. Syndrome Clin. Res. Rev.* **14**(4), 337–339 (2020). <https://doi.org/10.1016/j.dsrx.2020.04.012>. (ISSN 1871-4021)
- A. Kumar, P.K. Gupta, A. Srivastava, A review of modern technologies for tackling COVID-19 pandemic. *Diabetes Metab. Syndrome Clin. Res. Rev.* **14**(4), 569–573 (2020). <https://doi.org/10.1016/j.dsrx.2020.05.008>. (ISSN 1871-4021)
- H.J.T. Unwin, S. Mishra, V.C. Bradley et al., State-level tracking of COVID-19 in the United States. *Nat. Commun.* **11**, 6189 (2020). <https://doi.org/10.1038/s41467-020-19652-6>
- D. Easwaramoorthy, A. Gowrisankar, A. Manimaran, S. Nandhini, L. Rondoni, S. Banerjee, An exploration of fractal-based prognostic model and comparative analysis for second wave of COVID-19 diffusion. *Nonlinear Dyn.* **97**(1), 1–12 (2020). <https://doi.org/10.1007/s11047-019-05007-w>

- Dyn. **8**, 1–21 (2021). <https://doi.org/10.1007/s11071-021-06865-7>
9. C. Kavitha, A. Gowrisankar, S. Banerjee, The second and third waves in India: when will the pandemic be culminated? Eur. Phys. J. Plus. **136**(5), 596 (2021). <https://doi.org/10.1140/epjp/s13360-021-01586-7>
10. A. Gowrisankar, L. Rondoni, S. Banerjee, Can India develop herd immunity against COVID-19? Eur. Phys. J. Plus **135**(6), 526 (2020). <https://doi.org/10.1140/epjp/s13360-020-00531-4>
11. S.J. Malla, P.J.A. Alphonse, COVID-19 outbreak: an ensemble pre-trained deep learning model for detecting informative tweets. Appl. Soft Comput. **107**, 107495 (2021). <https://doi.org/10.1016/j.asoc.2021.107495>. (ISSN 1568-4946)
12. B. Korber, W.M. Fischer et al., Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell **182**(4), 812–827.e19 (2020). <https://doi.org/10.1016/j.cell.2020.06.043>. (ISSN 0092-8674)
13. D. Vasireddy et al., Review of COVID-19 variants and COVID-19 vaccine efficacy: what the clinician should know? J. Clin. Med. Res. **13**(6), 317–325 (2021). <https://doi.org/10.14740/jocmr4518>
14. H. Chemaitelly, H.M. Yassine, F.M. Benslimane et al., mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar. Nat. Med. **27**, 1614–1621 (2021). <https://doi.org/10.1038/s41591-021-01446-y>
15. Y. Ai, A. Davis, D. Jones, S. Lemeshow, H. Tu, F. He, P. Ru, X. Pan, Z. Bohrerova, J. Lee, Wastewater-based epidemiology for tracking COVID-19 trend and variants of concern in Ohio, United States, medRxiv 2021.06.08.21258421 (2021). <https://doi.org/10.1101/2021.06.08.21258421>
16. W.T. Gibson, D.M. Evans, J. An, S. J. M. Jones, ACE 2 coding variants: a potential X-linked risk factor for COVID-19 disease. bioRxiv 2020.04.05.026633 (2020). <https://doi.org/10.1101/2020.04.05.026633>
17. I.D. Apostolopoulos, T.A. Mpesiana, COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys. Eng. Sci. Med. **43**, 635–640 (2020). <https://doi.org/10.1007/s13246-020-00865-4>
18. M.K. Gupta et al., Detection and localization for watermarking technique using LSB encryption for DICOM Image. J. Discrete Math. Sci. Cryptogr. (2022). <https://doi.org/10.1080/09720529.2021.2009193>
19. R. Mardani et al., Laboratory parameters in detection of COVID-19 patients with positive RT-PCR; a diagnostic accuracy study. Arch. Acad. Emerg. Med. **8**(1), e43 (2020)
20. Alireza Tahamtan & Abdollah Ardebili, Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev. Mol. Diagn. **20**(5), 453–454 (2020). <https://doi.org/10.1080/14737159.2020.1757437>
21. P.B. van Kasteren, B. van der Veer, S. van den Brink, L. Wijsman, J. de Jonge, A. van den Brandt, R. Molkenkamp, C.B.E.M. Reusken, A. Meijer, Comparison of seven commercial RT-PCR diagnostic kits for COVID-19. J. Clin. Virol. **128**, 104412 (2020). <https://doi.org/10.1016/j.jcv.2020.104412>. (ISSN 1386-6532)
22. S. Malla, P.J.A. Alphonse, Fake or real news about COVID-19? Pretrained transformer model to detect potential misleading news. Eur. Phys. J. Spec. Top. (2022). <https://doi.org/10.1140/epjs/s11734-022-00436-6>
23. L. Kranthi Kumar, P. Alphonse, COVID-19 disease diagnosis with light-weight CNN using modified MFCC and enhanced GFCC from human respiratory sounds. Eur. Phys. J. Spec. Top. (2022). <https://doi.org/10.1140/epjs/s11734-022-00432-w>
24. V. Abreu, A. Oliveira, J.A. Duarte, A. Marques, Computerized respiratory sounds in paediatrics: a systematic review. Respir. Med. X **3**, 100027 (2021). <https://doi.org/10.1016/j.yrmex.2021.100027>. (ISSN 2590-1435)
25. Y. Ma, et al., LungBRN: a smart digital stethoscope for detecting respiratory disease using bi-ResNet deep learning algorithm. In 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2019, pp. 1–4. <https://doi.org/10.1109/BIOCAS.2019.8919021>
26. S.M. Khan, N. Qaiser, S.F. Shaikh, M.M. Hussain, Design analysis and human tests of foil-based wheezing monitoring system for asthma detection. IEEE Trans. Electron Devices **67**(1), 249–257 (2020). <https://doi.org/10.1109/TED.2019.2951580>
27. J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J. Biomed. Health Inform. **22**(1), 173–183 (2018). <https://doi.org/10.1109/JBHI.2017.2655720>
28. L. Brabenec, J. Mekyska, Z. Galaz et al., Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. J. Neural Transm. **124**, 303–334 (2017). <https://doi.org/10.1007/s00702-017-1676-0>
29. R.D. Reid, A.L. Pipe, B. Quinlan, J. Oda, Interactive voice response telephony to promote smoking cessation in patients with heart disease: a pilot study. Patient Educ. Couns. **66**(3), 319–326 (2007). <https://doi.org/10.1016/j.pec.2007.01.005>. (ISSN 0738-3991)
30. E. Maor, J.D. Sara, D.M. Orbello, L.O. Lerman, Y. Levanon, A. Lerman, Voice signal characteristics are independently associated with coronary artery disease. Mayo Clin. Proc. **93**(7), 840–847 (2018). <https://doi.org/10.1016/j.mayocp.2017.12.025>. (ISSN 0025-6196)
31. R. Islam, M. Tarique, E. Abdel-Raheem, A survey on signal processing based pathological voice detection techniques. IEEE Access **8**, 66749–66776 (2020). <https://doi.org/10.1109/ACCESS.2020.2985280>
32. I.M.M. El Emam, M. Fezari, F. Amara, Towards developing a voice pathologies detection system. J. Commun. Technol. Electron. **59**, 1280–1288 (2014). <https://doi.org/10.1134/S1064226914110059>
33. P. Ni, Y. Li, G. Li et al., Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction. Neural Comput. Appl. **32**, 16149–16166 (2020). <https://doi.org/10.1007/s00521-020-04805-x>
34. K.K. Lella, P.J.A. Alphonse, A literature review on COVID-19 disease diagnosis from respiratory sound data. AIMS Bioeng. **8**(2), 140–153 (2021). <https://doi.org/10.3934/bioeng.2021013>

35. K.K. Lella, A. Pja, Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice. *AIMS Public Health.* **8**(2), 240–264 (2021). <https://doi.org/10.3934/publichealth.2021019>
36. K.K. Lella, A. Pja, Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath. *Alex. Eng. J.* **61**(2), 1319–1334 (2022). <https://doi.org/10.1016/j.aej.2021.06.024>. (ISSN 1110-0168)
37. M. S. Jagadeesh, P. J. A. Alphonse, NIT\_COVID-19 at WNUT-2020 task 2: deep learning model RoBERTa for identify informative COVID-19 english tweets. In Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), pp. 450–454 (2020). <https://doi.org/10.18653/v1/2020.wnut-1.66>
38. A. Imran, I. Posokhova, H.N. Qureshi, U. Masood, M.S. Riaz, K. Ali, C.N. John, M.I. Hussain, M. Nabeel, AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked.* **20**, 100378 (2020). <https://doi.org/10.1016/j.imu.2020.100378>
39. C. Aguilera-Astudillo, M. Chavez-Campos, A. Gonzalez-Suarez, J. L. Garcia-Cordero, A low-cost 3-D printed stethoscope connected to a smartphone. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 4365–4368. <https://doi.org/10.1109/EMBC.2016.7591694>
40. C. Brown, J. Chauhan, A. Grammenos, et al., In Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2020). <https://doi.org/10.1145/3394486.3412865>
41. J. Han, K. Qian, M. Song, Z. Yang, Z. Ren, S. Liu, J. Liu, H. Zheng, W. Ji, T. Koike, X. Li, Z. Zhang, Y. Yamamoto, B.W. Schuller, An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety. *Interspeech* (2020). <https://doi.org/10.21437/interspeech.2020-2223>
42. L. Orlandic, T. Teijeiro, D. Atienza, The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Sci Data* **8**(1), 156 (2021). <https://doi.org/10.1038/s41597-021-00937-4>. (Published 2021 Jun 23)
43. S. Pal, I. Ghosh, A mechanistic model for airborne and direct human-to-human transmission of COVID-19: effect of mitigation strategies and immigration of infectious persons. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00433-9>
44. M.B. Alsabek, I. Shahin, A. Hassan, Studying the similarity of COVID-19 sounds based on correlation analysis of MFCC. In 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), 2020, pp. 1–5. <https://doi.org/10.1109/CCCI49893.2020.9256700>
45. J. Laguarta, F. Hueto, B. Subirana, COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open J Eng Med Biol.* **29**(1), 275–281 (2020). <https://doi.org/10.1109/OJEMB.2020.3026928>
46. M. Al Ismail, S. Deshmukh, R. Singh, Detection of COVID-19 through the analysis of vocal fold oscillations. In ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 1035–1039. <https://doi.org/10.1109/ICASSP39728.2021.9414201>
47. B.W. Schuller, et al. The INTERSPEECH 2021 computational paralinguistics challenge: COVID-19 cough, COVID-19 speech, escalation & primates. *CoRR*, abs/2102.13468 (2021). <https://doi.org/10.48550/arXiv.2102.13468>
48. T. Xia, et al. COVID-19 sounds: a large-scale audio dataset for digital respiratory screening. In 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks (2021). <https://openreview.net/forum?id=9KArJb4r5ZQ>. Accessed 23 Nov 2021
49. S. Jayalakshmy, G.F. Sudha, Conditional GAN based augmentation for predictive modeling of respiratory signals. *Comput. Biol. Med.* **138**, 104930 (2021). <https://doi.org/10.1016/j.combiomed.2021.104930>. (ISSN 0010-4825)
50. Z. Mushtaq, S.-F. Su, Environmental sound classification using a regularized deep convolutional neural network with data augmentation. *Appl. Acoust.* **167**, 107389 (2020). <https://doi.org/10.1016/j.apacoust.2020.107389>. (ISSN 0003-682X)
51. B. McFee, et al. Librosa: audio and music signal analysis in python. In Proceedings of the 14th Python Science Conference, no. Scipy, pp. 18–24 (2015). <https://www.semanticscholar.org/paper/librosa%3AAudio-and-Music-Signal-Analysis-in-Python-McFee-Raffel/e5c114afc8c4d4e10ae068ba8e3387cc13e17a6e>. Accessed 18 Nov 2021
52. A. Mesaros, et al. DCASE 2017 challenge setup: tasks datasets and baseline system. In Proceedings of the Detection Classification Acoust. Scenes Events Workshop, 2017. <https://hal.inria.fr/hal-01627981/>. Accessed 13 Nov 2021
53. G. Chaudhari, X. Jiang, A. Fakhry, et al. Virufy: global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough audio samples (2011). <https://doi.org/10.48550/arXiv.2011.13320>
54. N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S.R. Chetupalli, P.K. Ghosh, S. Ganapathy, Coswara—a database of breathing, cough, and voice sounds for COVID-19 diagnosis. *Proc. Interspeech* **2020**, 4811–4815 (2020). <https://doi.org/10.21437/Interspeech.2020-2768>
55. J. Andreu-Perez et al., A generic deep learning based cough analysis system from clinically validated samples for point-of-need COVID-19 test and severity levels. *IEEE Trans. Serv. Comput.* (2021). <https://doi.org/10.1109/TSC.2021.3061402>
56. A. Gowrisankar, T.M.C. Priyanka, S. Banerjee, Omicron: a mysterious variant of concern. *Eur. Phys. J. Plus* **137**, 100 (2022). <https://doi.org/10.1140/epjp/s13360-021-02321-y>
57. R. Gopal, V.K. Chandrasekar, M. Lakshmanan, Analysis of the second wave of COVID-19 in India based on SEIR model. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00426-8>

58. H. Natiq, A. Saha, In search of COVID-19 transmission through an infected prey. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00429-5>
59. P.S. Rana, N. Sharma, The modeling and analysis of the COVID-19 pandemic with vaccination and treatment control: a case study of Maharashtra, Delhi, Uttarakhand, Sikkim, and Russia in the light of pharmaceutical and non-pharmaceutical approaches. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00534-5>
60. D. Ghosh, P.K. Santra, G.S. Mahapatra et al., A discrete-time epidemic model for the analysis of transmission of COVID19 based upon data of epidemiological parameters. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00537-2>
61. R.M. Chen, Analysing deaths and confirmed cases of COVID-19 pandemic by analytical approaches. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00535-4>
62. T.D. Frank, J. Smucker, Characterizing stages of COVID-19 epidemics: a nonlinear physics perspective based on amplitude equations. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00530-9>
63. A. Sioofy Khoojine, M. Mabsuli, M. Shadabfar et al., A proposed fractional dynamic system and Monte Carlo-based back analysis for simulating the spreading profile of COVID-19. *Eur. Phys. J. Spec. Top.* (2022). <https://doi.org/10.1140/epjs/s11734-022-00538-1>

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.