Soft Computing (2022) 26:13405–13429
https://doi.org/10.1007/s00500-022-07499-6

DATA ANALYTICS AND MACHINE LEARNING

Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds

Ali Mohammad Alqudah[1] · Shoroq Qazan[2] · Yusra M. Obeidat[3]

Accepted: 9 September 2022 / Published online: 26 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

In recent years deep learning models improve the diagnosis performance of many diseases especially respiratory diseases. This paper will propose an evaluation for the performance of different deep learning models associated with the raw lung auscultation sounds in detecting respiratory pathologies to help in providing diagnostic of respiratory pathologies in digital recorded respiratory sounds. Also, we will find out the best deep learning model for this task. In this paper, three different deep learning models have been evaluated on non-augmented and augmented datasets, where two different datasets have been utilized to generate four different sub-datasets. The results show that all the proposed deep learning methods were successful and achieved high performance in classifying the raw lung sounds, the methods were applied on different datasets and used either augmentation or non-augmentation. Among all proposed deep learning models, the CNN–LSTM model was the best model in all datasets for both augmentation and non-augmentation cases. The accuracy of CNN–LSTM model using non-augmentation was 99.6%, 99.8%, 82.4%, and 99.4% for datasets 1, 2, 3, and 4, respectively, and using augmentation was 100%, 99.8%, 98.0%, and 99.5% for datasets 1, 2, 3, and 4, respectively. While the augmentation process successfully helps the deep learning models in enhancing their performance on the testing datasets with a notable value. Moreover, the hybrid model that combines both CNN and LSTM techniques performed better than models that are based only on one of these techniques, this mainly refers to the use of CNN for automatic deep features extraction from lung sound while LSTM is used for classification.

Keywords Deep learning · Lung sounds · Pulmonary diseases · Respiratory pathologies · Digital stethoscope

Abbreviations

CNN	Convolutional neural network
LSTM	Long-short term memory
SVM	Support-vectors machine
ANN	Artificial neural network
ReLU	Rectified linear unit
KNN	*K*-nearest neighbor
WHO	World Health Organization
COPD	Chronic obstructive pulmonary disease
KAUH	King Abdullah University hospital
DL	Deep learning
ML	Machine learning
LRTI	Lower respiratory tract infection
TP	True positive
TN	True negative
FP	False positive
FN	False negative
SPO2	Oxygen saturation
ICBHI	International Conference on Biomedical Health Informatics
GMM	Gaussian mixture model
TF	Time–frequency
TS	Timescale
MFCCs	Mel-frequency cepstral coefficients
DTCWT	Dual-tree complex wavelet transform
PSD	Power spectral density
HHT	Hilbert–Huang transform
HOS	Higher order statistics
MCC	Matthews correlation coefficient

✉ Ali Mohammad Alqudah
ali_qudah@hotmail.com

[1] Department of Biomedical Systems and Informatics Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Irbid, Jordan

[2] Department of Computer Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Irbid, Jordan

[3] Department of Electronic Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Irbid, Jordan

1 Introduction

In recent years, lung diseases became the third largest cause of death globally (Lehrer 2018). Based on the World Health Organization (WHO) statistics, the main five lung diseases (Moussavi 2006) are: tuberculosis, lung cancer, chronic obstructive pulmonary disease (COPD), asthma, and acute lower respiratory tract infection (LRTI). These diseases are responsible for the death of more than 3 million people each year worldwide (Chang and Cheng 2008; Chang and Lai 2010). These lung diseases are affecting the overall healthcare system severely while on the other hand, they accordingly affect the general population's lives. Like other disease especially the serious one, prevention is the key for decreasing it affect, diagnosis and treatment in early stages are considered key factors and methods for limiting the negative impact of these deadly diseases. Auscultation of the lung using a stethoscope is the traditional method and one of the widely used diagnostic method which is regularly used by specialists and general practitioners for performing the initial realization of the respiratory system condition.

The sound of the lungs can be normal or abnormal. An irregularity in the auscultated sound typically denotes lung fluid, infection, inflammation, or blockage (Chang and Lai 2010; Sengupta et al. 2016). Many different types of anomalous (adventitious) lung sounds, such as wheezes, stridor, rhonchi, and crackles, superimpose regular sounds. Breathing whistle-like high-pitched continuous waves lasting longer than 80 ms are referred to as wheezes (Naves et al. 2016). These noises are caused by bronchial tube irritation or constriction. Similar to this, stridor noises are high-pitched waves lasting longer than 250 ms and exceeding 500 Hz. They typically develop as a result of tracheal or laryngeal stenosis. Rhonchi are low-pitched, continuous waves of noises with frequencies under 200 Hz that resemble snoring (Naves et al. 2016; Bardou et al. 2018; Palaniappan et al. 2014). They typically start when the bronchial passages get overfilled with liquid or mucus. Crackles are abrupt clicking or rattling noises that can be either fine (short-lived) or coarse (long-lasting) (long duration) (Palaniappan et al. 2014). These noises are a sign of heart failure or pneumonia. Coughing, snoring, and squawking are other respiratory noises. Lung sounds are typically acoustic waves that have frequencies between 100 Hz and 2 kHz. The human ear is only susceptible to waves between 20 Hz and 20 kHz, though (Rocha et al. 2017; Aykanat et al. 2017). Due to the classic manual stethoscope's inability to pick up on the matching respiratory sounds of numerous ailments, these conditions may be misdiagnosed or go unnoticed (Serbes et al. 2013). As a result, crucial information regarding the health of the respiratory organs that were delivered by lower frequency waves is lost throughout the auscultation procedure (Bahoura 2009). Additionally, the quality of the tool, the expertise of the doctor, and the setting can all have an impact on the diagnosis of lung disorders. As a result, electronic stethoscopes have been emerging progressively to take the role of conventional diagnostic equipment (Bahoura 2009; Icer and Gengec 2014). It has the capacity to record lung sounds as signals within a computer, enabling medical professionals to analyze these signals using time–frequency analysis more accurately. Additionally, current developments in artificial intelligence and signal processing help clinicians make decisions when identifying respiratory disorders through lung sounds (Icer and Gengec 2014; Jin et al. 2014).

Moreover, physicians tend to use different strategies like oxygen saturation (SPO2) using plethysmography, spirometry, and arterial blood gas analysis, but lung sound auscultation still vital for physicians due to its simplicity and low cost (Sengupta et al. 2016). The traditional auscultation technique is usually used to collect lung sounds by using a stethoscope (Lehrer 2018; Moussavi 2006; Chang and Cheng 2008). This method is noninvasive, does not need long time for diagnosis, and has no harm effect on patient, but it might lead to wrong diagnosis if the physician is not well trained to use it (Lehrer 2018; Moussavi 2006). Lung sounds are non-stationary which leads to complex analysis and recognition of sounds (Moussavi 2006). Therefore, it is necessary to develop an automatic recognition system to solve the limitations of using the traditional techniques to ensure more efficient clinical diagnosis (Chang and Lai 2010; Sengupta et al. 2016; Naves et al. 2016). In general, there are two types of lung sounds: normal if the lung has no respiratory disorder, or adventitious breathing sound when the lung has a respiratory disorder (Lehrer 2018; Moussavi 2006; Chang and Cheng 2008; Chang and Lai 2010; Sengupta et al. 2016; Naves et al. 2016). Respiratory disorder became a common problem in all sides of the world. Smoking is the most common cause of this disorder (Lehrer 2018; Moussavi 2006), but it can be also caused by genetics and environmental exposure (Chang and Cheng 2008). There are many categories of lung sounds with respiratory disorder, including fine crackle, coarse crackle, polyphonic wheeze, monophonic wheeze, squawk, and stridor, pleural rub, stridor, and squawks (Lehrer 2018; Moussavi 2006; Chang and Cheng 2008; Chang and Lai 2010; Sengupta et al. 2016; Naves et al. 2016; Bardou et al. 2018).

In the last years, large number of different research approaches have been developed and evaluated for automatic detection and classification of lung abnormalities using lung auscultation sounds. Also, the researchers provided many types of feature extraction techniques that have

been used with different types of machine learning (ML) algorithms. ML techniques have been used for lung sounds classifications (Bardou et al. 2018; Palaniappan et al. 2014; Rocha et al. 2017; Aykanat et al. 2017). These techniques mainly applied to create models to find better representations for large-scale unlabeled data (Palaniappan et al. 2014; Rocha et al. 2017; Aykanat et al. 2017; Serbes et al. 2013). Feature-based techniques are commonly used to create automatic systems for classifying lung sounds (Rocha et al. 2017).

The advancement of deep learning (DL) techniques and new developments demonstrating very promising results in different medical applications like diseases detection and classification (Bahoura 2009). DL has many advantages over the ML such as it has an automatic feature extraction, and DL methods are more generic and mitigate the limitations of traditional ML-based methods. Moreover, DL-based methods that are used in recent years for the classification of respiratory abnormalities and pathologies from lung auscultation records have produced very promising results (Bahoura 2009; Icer and Gengec 2014). However, to get the right functionality of DL, DL networks should undergo an extensive training process using a huge training dataset which requires a considerable amount of time and powerful computational resources. As a result, it becomes quite challenging to use DL frameworks in available wearable devices and mobile platforms since it has low computational resources. Many approaches for lung sound classification using CNNs and compared it with features-based approaches. Usually authors used Mel-frequency cepstral coefficient MFCC's statistics extracted from the signals in the first handcrafted features-based approach, and local binary patterns extracted from spectrograms were used in the second approach, while the third approach is based on the design of convolutional neural networks (CNN) (Bardou et al. 2018).

Also, the researchers employed different types of machine learning algorithms with handcrafted features like Mel-frequency cepstral coefficient (MFCC) features in a support-vector machine (SVM) or fed the spectrogram images to the convolutional neural network (CNN). The most common type of classifier with handcrafted features is the SVM algorithm as a classification method for audio and utilized its results to benchmark the CNN algorithm. Moreover, different classification scenarios can be involved for comparison like healthy versus pathological classification; rale, rhonchus, and normal sound classification; singular respiratory sound type classification; and audio type classification with all sound types (Aykanat et al. 2017). Recently, a new method for automatic detection of pulmonary diseases (PDs) from lung sound (LS) signals, where the LS signal modes were evaluated using empirical wavelet transform with fixed boundary points. The time-domain (Shannon entropy) and frequency-domain (peak amplitude and peak frequency) handcrafted features have been extracted from each mode. Then, machine learning classifiers, such as support-vector machine, random forest, extreme gradient boosting, and light gradient boosting machine (LGBM), have been chosen to detect PDs using the features of LS signals automatically. The performance of these features shows a promising result and can be enhanced further for multi-class scenarios like normal versus asthma, normal versus pneumonia, normal versus chronic obstructive pulmonary disease (COPD), and normal versus pneumonia versus asthma versus COPD classification schemes (Tripathy et al. 2022).

In addition, applying different homogeneous ensemble learning methods to perform multi-class classification of respiratory diseases has been raised recently to enhance the performance of the systems. These systems can be applied to a wider range of conditions involved including healthy, asthma, pneumonia, heart failure, bronchiectasis or bronchitis, and chronic obstructive pulmonary disease (Fraiwan et al. 2021b). The other types of ensembles of hybrid methods are combining two different types of deep learning algorithms in one architecture, this can robustly enhance the deep learning performance to recognize pulmonary diseases from electronically recorded lung sounds. But all researchers employ several preprocessing steps that were undertaken to ensure smoother and less noisy signals like wavelet smoothing, displacement artifact removal, and z-score normalization. Usually, the deep learning architectures that consisted of two stages are mainly based on combining convolutional neural networks and bidirectional long short-term memory units (Fraiwan et al. 2021c).

1.1 Our contribution

In this work, we are presenting a study to investigate the ability of three different deep learning models, illustrated by convolutional neural networks, long short-term memory, and a hybrid model between them, in recognizing multiple pulmonary diseases from recorded lung sound signals. The used signals were obtained by merging two publicly available datasets International Conference on Biomedical Health Informatics (ICBHI) 2017 Challenge dataset and King Abdullah University Hospital (KAUH) dataset. The recordings represent signals from patients suffering from normal, asthma, lung fibrosis, BRON, COPD, heart failure, heart failure + COPD, heart failure + lung fibrosis, lung fibrosis, pleural effusion, and pneumonia. A CNN and LSTM network, and hybrid (CNN + LSTM) were designed for the training and testing classification processes to extract information from the temporal domain of the signals in the raw formats without

any preprocessing techniques. The three models have been evaluated on non-augmented and augmented datasets where the two datasets have been utilized to generate four different sub-datasets. Several evaluation metrics were used to evaluate the recognition of diseases using CNN and LSTM networks individually as well as a combination of both networks. To the best of our knowledge, single neural network approaches have often been used in the building of deep learning models for lung sound classification. Therefore, in addition to the suggested hybrid model (CNN + LSTM), the network's capacity to recognize diseases when it was functioning independently as either CNN or LSTM was examined. In addition to feature memorization, the key contribution of this study is the implementation of the standard CNN feature extraction approach and the LSTM network, which increases learning efficiency. Our system presents fundamental differences when compared to the mentioned systems in the literature. Furthermore, the contributions of this work can be summarized as follows:

- The first paper that compares three different types of deep learning models, CNN model, LSTM model, and a hybrid model that combined both CNN and LSTM models.
- The first system is used to classify raw lung auscultation sound without any preprocessing techniques.
- The first system used sound augmentation techniques to enhance the accuracy of the models.
- Using a huge dataset of lung auscultation sounds collected from two different datasets available online.
- A system to detect and classify 11 different classes including ten diseases and healthy.

The remainder of the paper is organized as follows: Sect. 2 provides details about recent related works. Section 3 presents with a detailed explanation of the used dataset and the proposed architectures. Section 4 is the results, including the performance of the proposed different architectures. Section 5 is the discussion of the method results. And finally, Sect. 6 represents the conclusion of the proposed work.

2 Literature review

In recent years, a substantial amount of previous related research that employing both machine learning and deep learning have been proposed. The methods focused on automated raw and processed respiratory sound classification. In this section, the most recent and related works to the paper's topic are discussed. Serbes et al. (2013) have used time–frequency (TF) and timescale (TS) analysis to detect pulmonary crackles. The crackles frequency characteristics were extracted from the non-preprocessed and pre-processed signals using TF and TS analysis. Dual-tree complex wavelet transform (DTCWT) is applied in the pre-processing step to filter out the frequency bands that have no crackle information. They have used *K*-nearest neighbors (KNN), SVM (support-vector machine), and multilayer perceptron to classify crackling and non-crackling sounds and achieved an accuracy of 97.5% from the SVM classifier. Bahoura (2009) have proposed an approach for two classes of lung sounds including normal and wheeze signals, they have used the Mel-frequency cepstral coefficients (MFCCs) for feature extraction and the Gaussian mixture model (GMM) to classify the signals, and they achieved an accuracy of 94.2%.

Icer and Gengec (2014) have used support-vector machine (SVM) for automatic classification. They created features using the frequency ratio of power spectral density (PSD) values and the Hilbert–Huang transform (HHT) to distinguish between three classes of lung sounds including: normal lung sounds, crackles, and rhonchus with an accuracy above 90%. Moreover, Jin et al. (2014) have used SVM to distinguish between different classes of respiratory sounds including normal, wheezing, stridor, and rhonchi and achieved an accuracy between 97.7 and 98.8%. Reyes et al. (2014) have used a technique to obtain the time–frequency (TF) representation of thoracic sounds. The performance of the TF representations for different classes including the heart, adventitious, and normal lung sounds was assessed using TF patterns and they stated that the best performance was achieved from the Hilbert–Huang spectrum (HHS). Higher order statistics (HOS) were used by Naves et al. (2016) for classifying different classes of lung sounds including normal, coarse crackle, fine crackle, monophonic and polyphonic wheezes. They have used genetic algorithms and Fisher's discriminant ratio to reduce dimensionality, and k-NN, and naive Bayes classifiers were used for classification with an accuracy of 98.1% on training data, and 94.6% on validation data.

Orjuela-Canon et al. (2014) used MFCC features along with artificial neural network (ANN) to distinguish between normal sounds, wheezes, and crackles. The achieved performance of classification was 75% for crackles, 100% for wheezes and 80% for normal. Maruf et al. (2015) have used GMM to classify crackles from normal respiratory sounds. They have used a band-pass filter for background noises reduction, and then they extracted three spatial–temporal features, namely pitch, energy, and spectrogram and an accuracy of 97.56% was achieved. A novel attractor recurrent neural networks (ARNN) technique based on the fuzzy functions (FFs-ARNN) for the classification of lung abnormalities was proposed by Bagher et al. and achieved an accuracy of 91% (Khodabakhshi and Moradi 2017). Pinho et al. (2015) used

signal processing methodologies for the detection of crackles in audio files. Their method is based on using fractal dimension and box filtering to extract the window of interest to verify and validate the potential crackle, then extracting the crackle parameters for characterizations.

Islam et al. (2018) used ANN (artificial neural networks) and SVM for classifying lung sounds from 60 subjects 50% of them have asthma. They obtained the best accuracy of 93.3% from SVM scenario. Accuracies of up to 93% and 91.7% were achieved using other configurations of neural network for detecting crackles and wheezes, respectively (Guler et al. 2005). A dataset of seven classes including normal, coarse crackle, fine crackle, monophonic wheeze, polyphonic wheeze, squawk, and stridor were used in another work where different approaches of ANN were used and the results from convolutional neural network (CNN) were the best (Guler et al. 2005; Shuvo et al. 2020; Garcia-Ordas et al. 2020; Tsai et al. 2020; Demir et al. 2020; Kevat et al. 2020; Andrade et al. 2021; Wani et al. 2021).

CNN was also used by Jacome et al. (2019) to deal with respiratory sounds with accuracies of 97% and 87% in detecting inspiration and expiration, respectively. Two types of machine learning algorithms were proposed by Aykanat et al. (2017), Mel frequency cepstral coefficient (MFCC) features in a support-vector machine (SVM) and spectrogram images in the convolutional neural network (CNN). Four data sets were prepared for each CNN and SVM algorithm to classify different respiratory conditions including healthy versus pathological classification; rale, rhonchus, and normal sound classification; singular respiratory sound type classification; and audio type classification with all sound types. They achieved accuracies of CNN 86%, SVM 86%, CNN 76%, SVM 75%, CNN 80%, SVM 80%, and CNN 62%, SVM 62%, respectively. Their results have shown that pre-diagnosis and classifying of respiratory audio can be done accurately using CNN and SVM machine learning algorithms. Bardou et al. (2018) used three approaches, two of them are based on the extraction of a set of handcrafted features trained by three different classifiers including SVM, KNN, and Gaussian mixture models, while they applied CNN for their third approach. The dataset they used consist of seven classes (normal, coarse crackle, fine crackle, monophonic wheeze, polyphonic wheeze, squawk and stridor, the results they achieved show that the CNN outperform the handcrafted feature-based classifiers.

Garcia-Ordas et al. (2020) used a convolutional neural network (CNN) to classify the respiratory sounds into healthy, chronic, and non-chronic disease and achieved 0.993 F-score in the three-label classification. They also have done more challenging classification for different types of pathologies or healthy conditions including URTI, COPD, Bronchiectasis, Pneumonia, and Bronchiolitis and achieved F-score of 0.990 in all classes. Moreover, Fraiwan et al. (2021a) have used an electronic stethoscope to record lung sounds from 112 subjects of healthy and unhealthy conditions (35 healthy and 77 unhealthy) to create new dataset. The dataset contains seven ailments including normal breathing sounds, lung fibrosis, heart failure, asthma, pneumonia, bronchitis, pleural effusion, as well as COPD. This dataset was created for use in machine learning models to distinguish the correct type of lung sounds or detect pulmonary diseases. Fraiwan et al. (2021b) used different ensemble classifiers to perform multi-class classification of respiratory diseases. The dataset they used included a total of 215 subjects with 308 clinically acquired lung sound recordings and 1176 recordings obtained from the ICBHI Challenge database. The recorded data has different conditions including asthma, pneumonia, heart failure, bronchitis, chronic obstructive pulmonary disease, as well as healthy condition. Shannon entropy, logarithmic energy entropy, and spectrogram-based spectral entropy were used for feature representation of the lung sound signals. Bootstrap aggregation and adaptive boosting ensembles were built using decision trees and discriminant classifier. Boosted decision trees achieved the best overall accuracy, sensitivity, specificity, F1-score, and Cohen's kappa coefficient of 98.27%, 95.28%, 98.9%, 93.61%, and 92.28%, respectively. Among the baseline methods, SVM provided an average accuracy of (98.20%), sensitivity (91.5%), and specificity (98.55%) Furthermore, Fraiwan et al. (2021c) have used a deep learning model-based CNNs and bidirectional LSTM (BD-LSTM) for recognizing pulmonary diseases. A dataset of 103 patients recorded at King Abdullah University hospital (KAUH) in Jordan and data from 110 patients were added from the Int. Conf. on Biomedical Health Informatics publicly available challenge database were used. The highest average accuracy achieved in classifying patients based on the pulmonary disease types using CNN + BD-LSTM is 99.62% with a precision of 98.85% and a total agreement of 98.26% between the predictions and original classes within the training scheme.

Nguyen and Pernkopf (2022) proposed a methodology for lung sound classification by employing co-tuning and stochastic normalization to enhance the classification results. They split sound record into 8 s segments then calculating the corrected and normalized spectrogram. After that they co-tuned a Resnet50 based model and applied on three different scenarios 2 classes, 3 three classes, and 4 classes. The results show that their highest performance was obtained using 60–40 training and testing sets for two classes problem using Logmel Spectrogram, ResNet101 and scored specificity of 91.77% and sensitivity

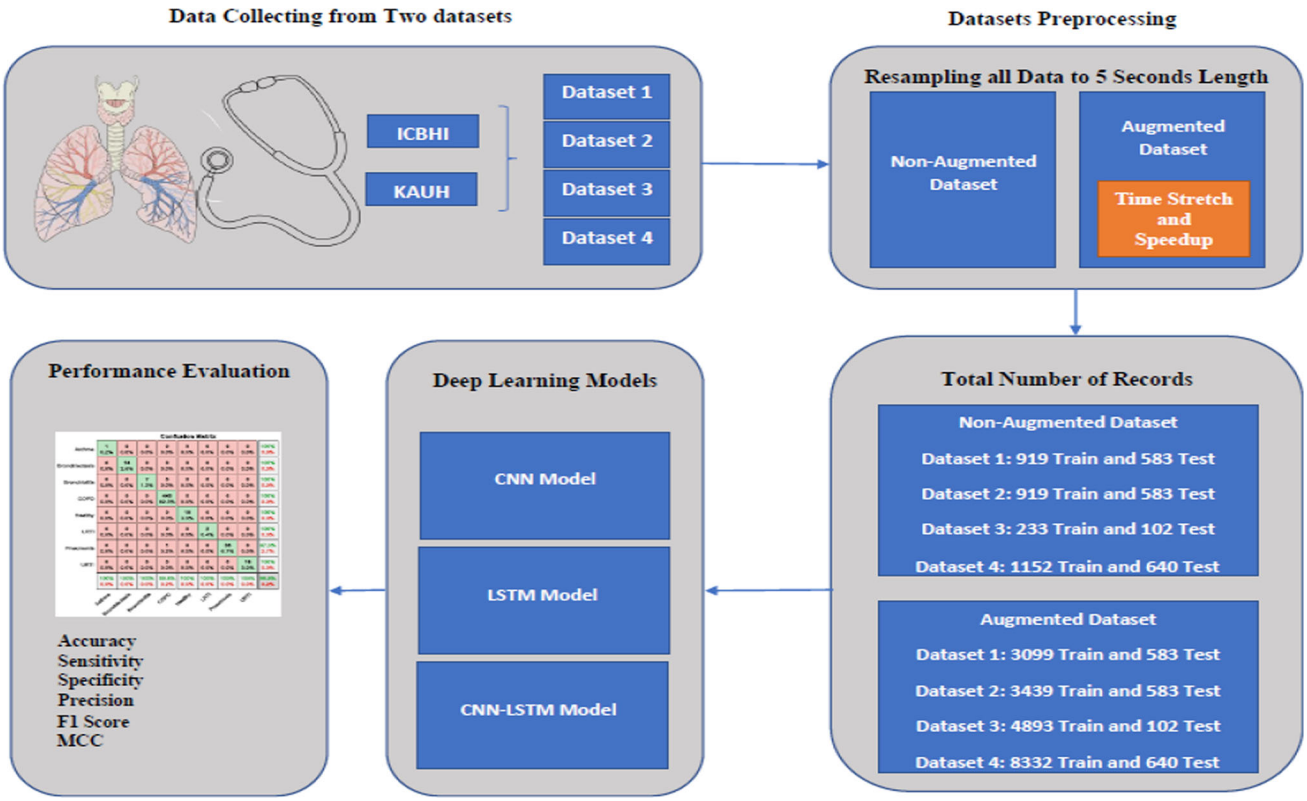

Fig. 1 The complete procedure followed in the proposed study

Table 1 The details of the used datasets

Dataset	Type	Number of classes	Number of records
Dataset 1	Non-augmented	3 (normal, non-chronic, and chronic)	1457 (train: 919; test: 538)
	Augmented		3637 (train: 3099; test: 538)
Dataset 2	Non-augmented	8 (normal, asthma, bronchiectasis, bronchiolitis, COPD, LRTI, pneumonia, and URTI)	1457 (train: 919; test: 538)
	Augmented		3977 (train: 3439; test: 538)
Dataset 3	Non-augmented	8 (normal, asthma, BRON, COPD, heart failure, lung fibrosis, pleural effusion, and pneumonia)	335 (train: 233; test: 102)
	Augmented		4995 (train: 4893; test: 102)
Dataset 4	Non-augmented	11 (normal, asthma, lung fibrosis, BRON, COPD, heart failure, heart failure + COPD, heart failure + lung fibrosis, lung fibrosis, pleural effusion, and pneumonia)	1792 (train: 1152; test: 640)
	Augmented		8972 (train: 8332; test: 640)

of 95.76%. Moreover, Tripathy et al. (2022) proposed a methodology using empirical wavelet transform with fixed boundary points. Where, the time-domain (Shannon entropy) and frequency-domain (peak amplitude and peak frequency) features have been extracted. Then, they employ different classifiers like support-vector machine, random forest, extreme gradient boosting, and light gradient boosting machine have been chosen to detect pulmonary diseases using the extracted features. The best accuracy values were 80.35, 83.27, 99.34, and 77.13% have been obtained using the light gradient boosting machine classifier with fivefolds cross validation for classification systems comparing normal to asthma, normal to pneumonia, normal to COPD, and normal to pneumonia, asthma, and COPD.

Bhatta et al. (2022) proposed a respiratory audio collection to forecast a variety of illnesses, including bronchiectasis, pneumonia, and asthma. To put this study into practice, we used respiratory and disease diagnosis audio datasets. We then extracted characteristics from each audio dataset and

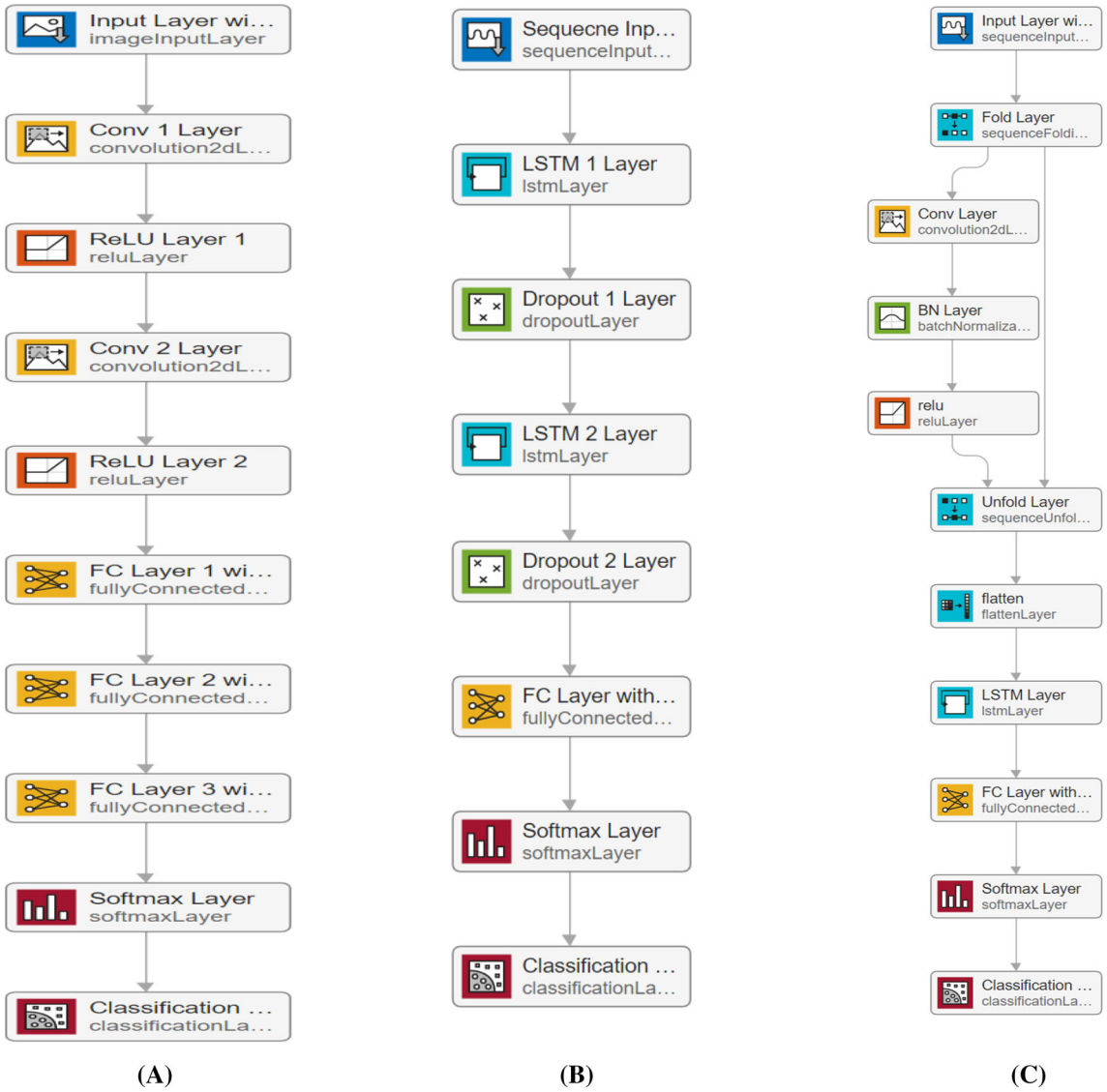

Fig. 2 The used deep learning models

Table 2 The layer details of proposed deep learning models

Layer number	CNN model	LSTM model	CNN–LSTM model
1	Input (size [12000, 1, 1])	Input (size [12000])	Input (size [12000])
2	Conv (filter [3, 1], # filter [16])	LSTM (hidden units 250)	Sequence folding
3	ReLU	Dropout (20%)	Conv (filter [3, 1], # filter [4])
4	Conv (filter [3, 1], # filter [8])	LSTM (hidden units 125)	Batch normalization
5	ReLU	Dropout (20%)	ReLU
6	FC layer (output 250)	FC layer (output # classes)	Sequence unfold
7	FC layer (output 125)	Softmax	Flatten
8	FC layer (output # of classes)	Classification ()	LSTM (hidden units 200)
9	Softmax		FC layer (output # of classes)
10	Classification ()		Softmax
11			Classification ()

trained a convolution neural network (CNN) method, model. We can add any new test data to predict an illness after the training model. The authors report an accuracy and sensitivity values of 86% using their own dataset. Finally, Soni et al. (2022) proposed s method for gapping the problem of heart and lung sounds labeling (diagnosing). They use the ResNet-18 model as a base to generate encodings in the latent space of length *N* for training. They used clinical data, such as age, sex, weight, and sound location, together with the audio file to make use of the common context of the recordings at the patient level. When using patient-specific representations to choose positive and negative pairs, they demonstrate better in downstream tasks for diagnosing heart and lung sounds. The highest performance achieved using linear evaluation and the AUC was 0.752 with 95% confidence interval of 0.715, 0.791.

As we can notice from the scanned literature, most of them are focused on using pre-trained model(s) not their models, or small datasets for evaluation of their methods. Moreover, they only focused on just employing the comparison between the pre-trained model(s) to select the best of them. Also, the literature does not provide a generalized capacity model either because of datasets or not using the augmentation process. Finally, all literature spends most of the research time on the development of signal enhancement or features extraction methods instead of developing new model or systems. Based on that this paper will enrich the literature by combining the two publicly available lung sound datasets to significantly enhance the generalization capacity and model performance for systems that classify adventitious lung sounds and respiratory diseases by employing the augmentation process. Moreover, it will provide a comprehensive comparison between three different types of deep learning models (CNN, LSTM, and CNN–LSTM) based on various datasets by applying them to raw sounds without any enhancement methods or feature extractions except resizing and show the best model for each scenario which open the way for further investigations of designing new models.

3 Materials and methods

As shown in Fig. 1, the adopted methodology consists of the following main phases: data acquisition and preparation, feature extraction, construction and training of the ensemble and baseline classifiers, and finally performance evaluation. These steps are detailed below.

3.1 Lung sounds datasets

In this paper, the data used incorporated two different datasets, both datasets are consisting of stethoscope lung sounds classified with different respiratory diseases. The first dataset was the publicly available International Conference on Biomedical Health Informatics (ICBHI) 2017 Challenge dataset. The second dataset was the new public available King Abdullah University Hospital (KAUH) dataset. Furthermore, both datasets have been merged together to make four different datasets. The four datasets vary in number and type of classes included in each one. Table 1 provides a detailed overview of the used dataset and the four datasets from their merging. Each dataset will be discussed in detail in the next two sections.

3.1.1 KAUH dataset

The KAUH dataset which is a new public available dataset with a total of 70 individuals with various respiratory diseases such as asthma, pneumonia, heart failure, bronchiectasis or bronchitis (BRON diseases), and chronic obstructive pulmonary disease (COPD) (Fraiwan et al. 2021a). A total of 35 healthy controls were also surveyed. The age of the participants in this study was not a variable of interest in order to ensure fair research. The participants ranged in age from youngsters through adults to the elderly. After thoroughly comprehending the parameters of the study and the technique involved, all participants signed a written consent form. The complete dataset comprised of 308 lung sound recordings, each lasting 5 s. Based on the average resting respiration rates for humans (12–20 breaths per minute), this length is sufficient to encompass at least one respiratory cycle and has been used in previous investigations (Fraiwan et al. 2021a). In general, adopting small length data windows alleviates the difficulty of medical data availability while also increasing the model's computing efficiency. Furthermore, training the model on lung sound signals rather than respiratory cycles makes

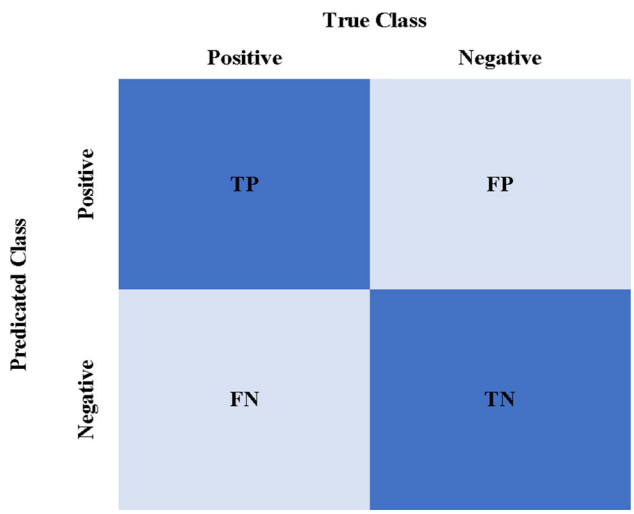

Fig. 3 A confusion matrix example

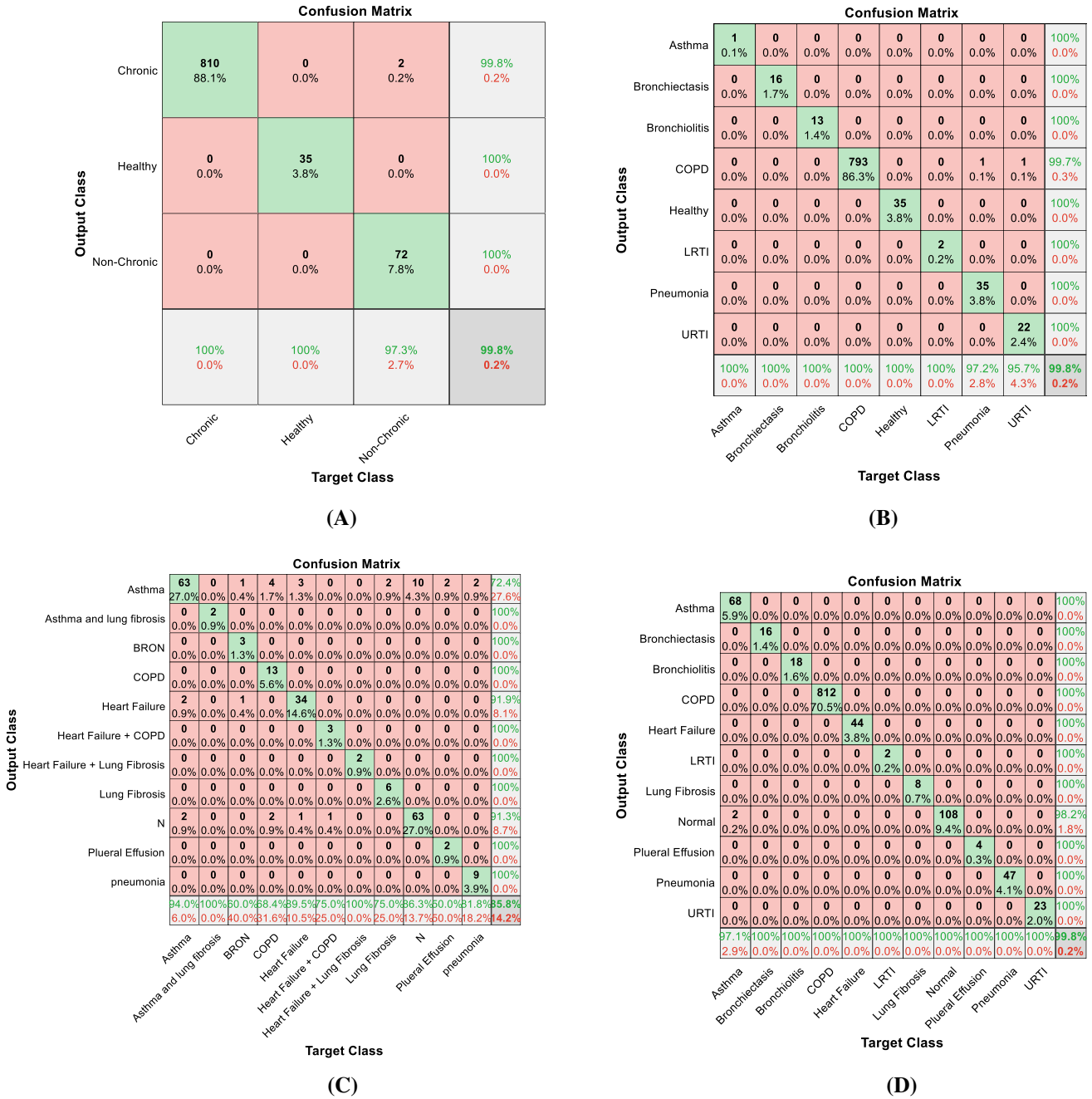

Fig. 4 CNN model training results for non-augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

data curation and labeling much easier. In clinical situations and real-time applications, these distinguished features are usually advantageous (Fraiwan et al. 2021a, b, c).

3.1.2 ICBHI 2017 dataset

The International Conference on Biomedical Health Informatics (ICBHI) 2017 dataset, which is a publicly available benchmark dataset of lung auscultation sounds (Rocha et al. 2017), was used in this research. Two independent research teams from Portugal and Greece have gathered the data (Nuckowska et al. 2019). The dataset contains 5:5 h of audio recordings sampled at multiple frequencies (4 kHz, 10 kHz, and 44 kHz), spanning from 10 to 90 s, in 920 audio samples of 126 participants in various anatomical positions (Chen et al. 2019). The samples are professionally annotated according to two schemes: I according to the corresponding patient's pathological condition, i.e., healthy and seven distinct disease classes, namely Pneumonia, Bronchiectasis,

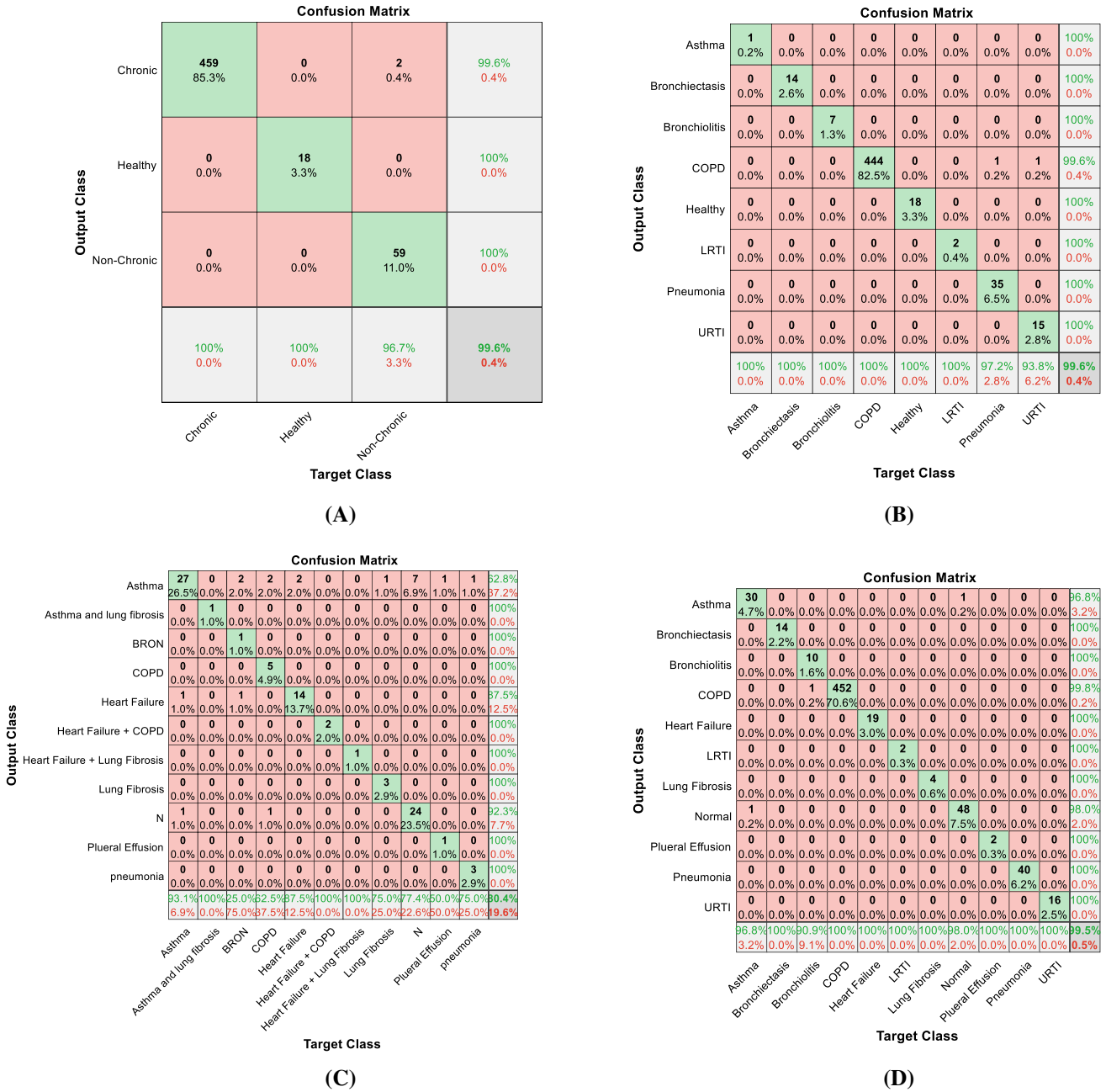

Fig. 5 CNN model testing results for non-augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

COPD, URTI, LRTI, Bronchiolitis, and Asthma; and (ii) according to the presence of respiratory anomalies, i.e., crackles and wheezes in each respiratory cycle. Zhang et al. (2015) contains more information about the dataset and data collecting technique. Table 1 shows the details of the used datasets in our research.

3.2 Augmentation

Data augmentation is a very popular technique that can be used to artificially expand the size of a training dataset in general, this is done by creating modified versions of audios in the dataset (Guler et al. 2005; Shuvo et al. 2020). Training deep learning models on huge datasets can result in more skillful models, and the augmentation techniques can create variations of the audio les that can improve the ability of the fit models to generalize what they have learned to new audios. For instance, with images, we might do things like rotating the image slightly, cropping or scaling it, modifying colors or lighting, or adding some noise to the image (Garcia-Ordas et al. 2020; Tsai et al. 2020; Demir et al.

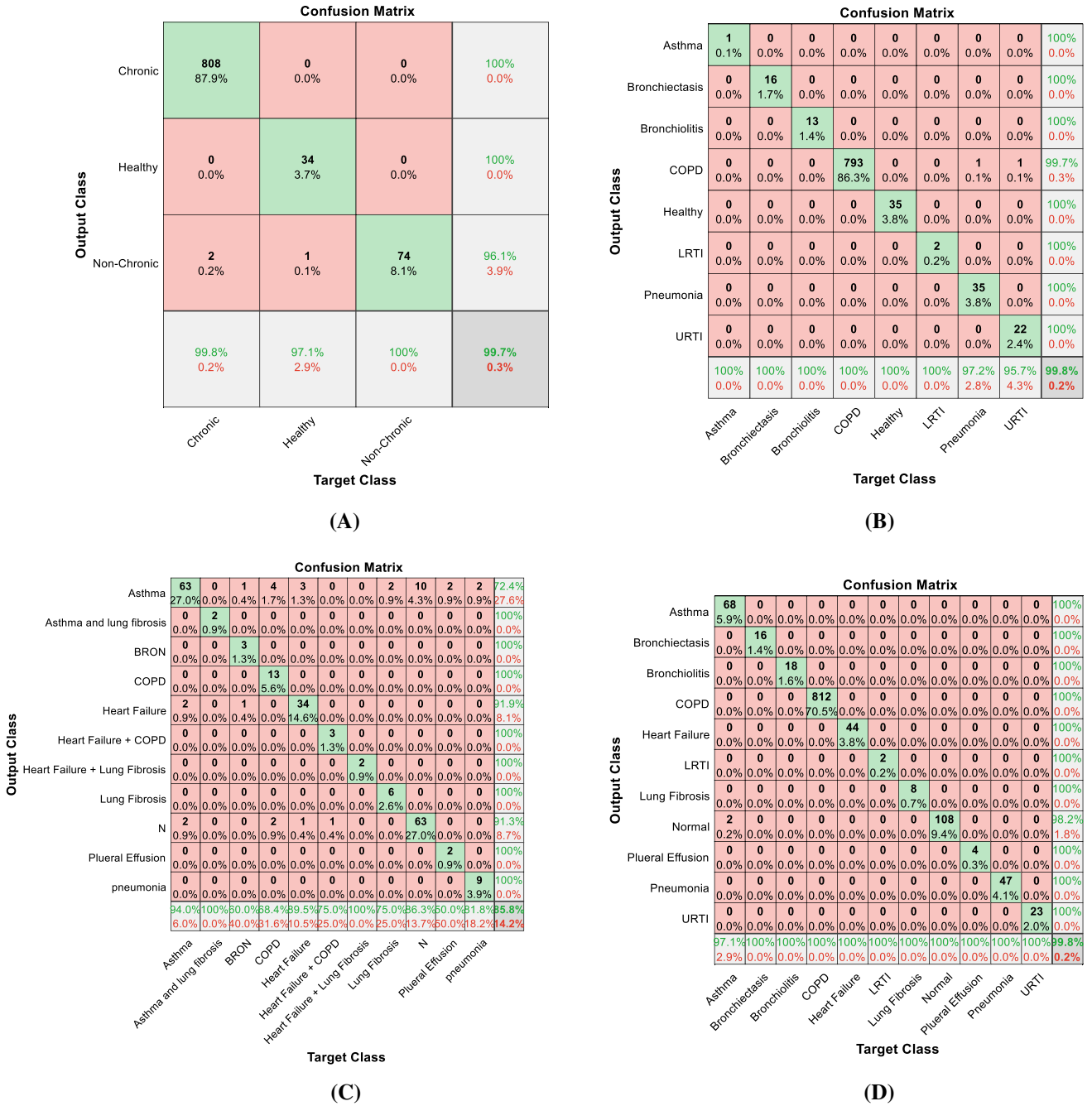

Fig. 6 LSTM model training results for non-augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

2020). Since the semantics of the image have not changed materially, so the same target label from the original sample will still apply to the augmented sample. Just like with images, there are several techniques to augment audio data as well. This augmentation can be done both on the raw audio (Kevat et al. 2020; Andrade et al. 2021; Wani et al. 2021). In this research the following audio augmentation techniques have been applied:

- Time Stretch: randomly slow down or speedup the sound.
- Time Shift: shift audio to the left or the right by a random amount.
- Add Noise: add some random values to the sound.
- Control Volume: randomly increasing or decreasing the volume of the audio.

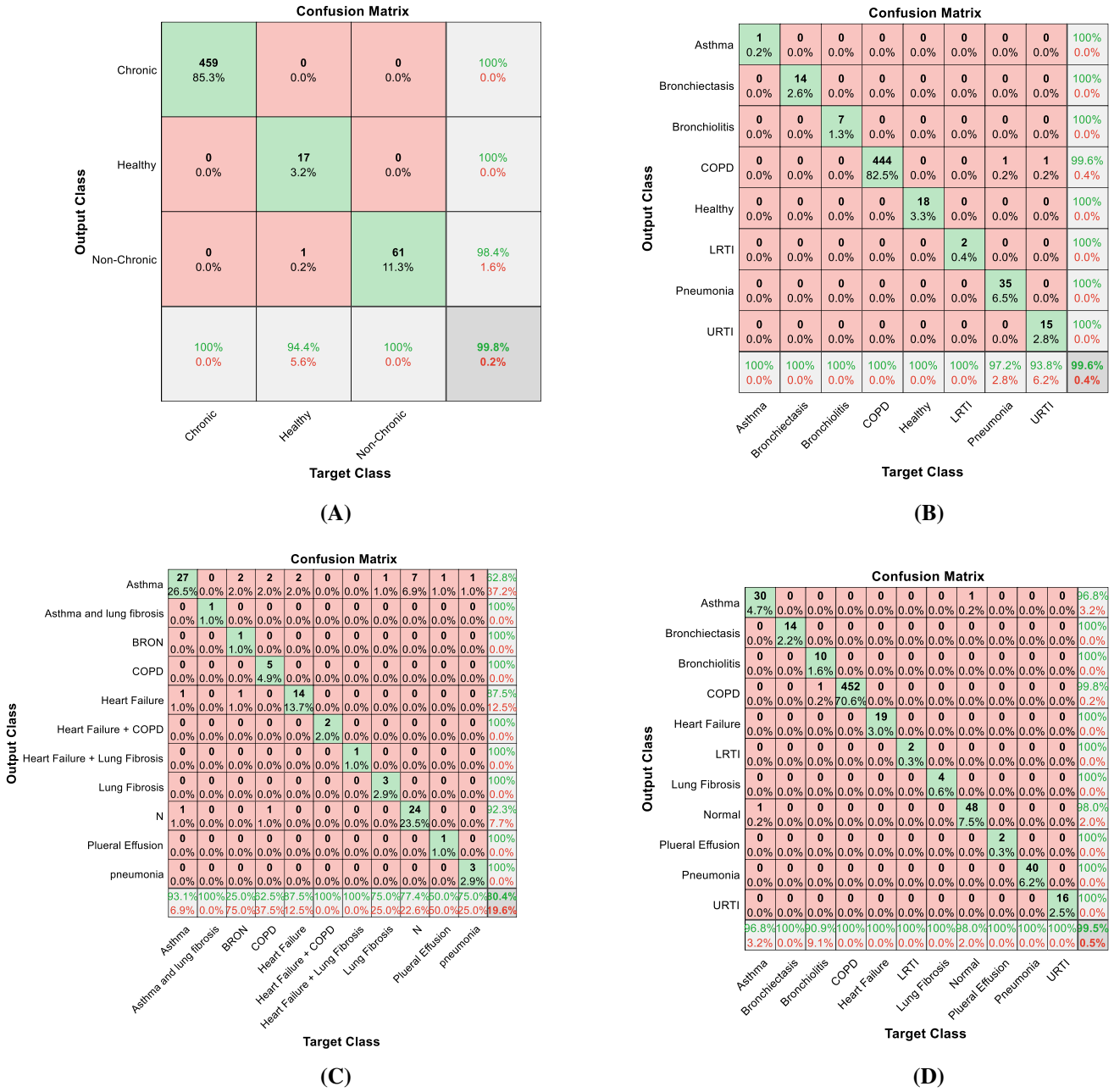

Fig. 7 LSTM model testing results for non-augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

3.3 Deep learning models

Deep learning is one of the newest sorts and state-of-the-art artificial intelligence technologies that has emerged in response to the growing quantity of massive datasets (Alqudah 2020; Esteva et al. 2021; Alqudah et al. 2021a, b; Kanavati et al. 2020). Deep learning is primarily defined and distinguished by the development of a unique architecture made up of many and sequential layers in which successive stages of input processing are carried out (LeCun et al. 1995, 2015). Deep learning is based on and inspired by the human brain's deep structure (LeCun et al. 1995; Alqudah et al. 2021c). The human brain's deep structures have a large number of hidden layers, allowing us to extract and abstract deep features at various levels and from various perspectives. In recent years, a slew of deep learning algorithms has been presented (Alqudah 2020; Esteva et al. 2021; Alqudah et al. 2021a, b, c; Kanavati et al. 2020; LeCun et al. 2015). The most frequently used, powerful, and efficient deep learning algorithms are the

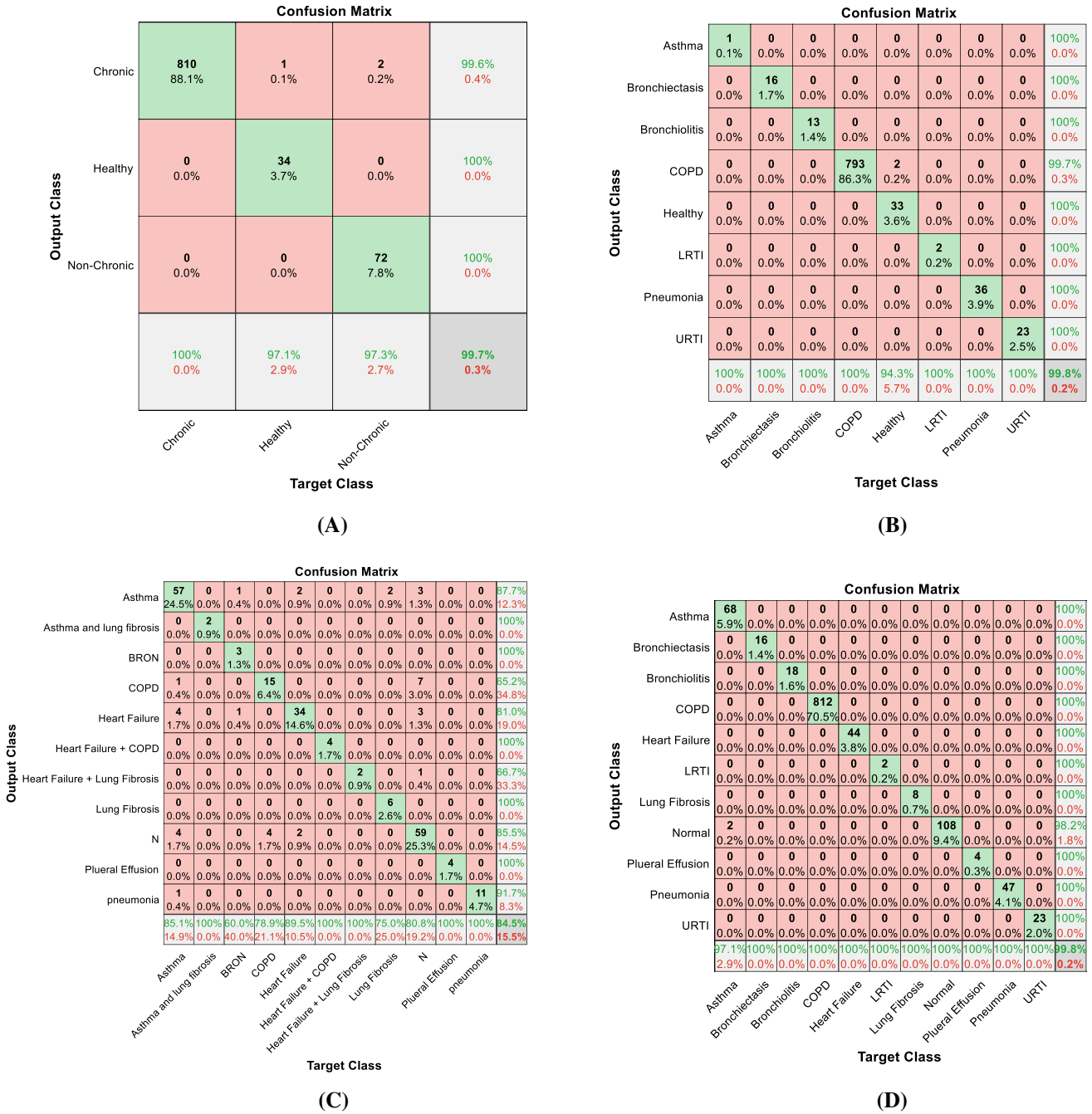

Fig. 8 CNN–LSTM model training results for non-augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

CNN (Alqudah 2020; Alqudah et al. 2021a, c; Alqudah and Alqudah 2022a) and Long Short-Term Memory (LSTM) (Ozturk and Ozkaya 2020; Petmezas et al. 2021; Cinar and Tuncer 2021; Jelodar et al. 2020). In the following subsections we will discuss in detail the developed separated and hybrid CNN and LSTM models. Moreover, Fig. 2 shows the developed models and Table 2 shows the layer details.

3.3.1 CNN model

The CNNs have a large number of hidden layers that use convolution and subsampling techniques to extract deep features from the input data (Alqudah et al. 2021a; LeCun et al. 2015). Input, convolution, RELU (rectified linear unit), fully connected, classification, and output layers are the different types of layers in a CNN. These layers are

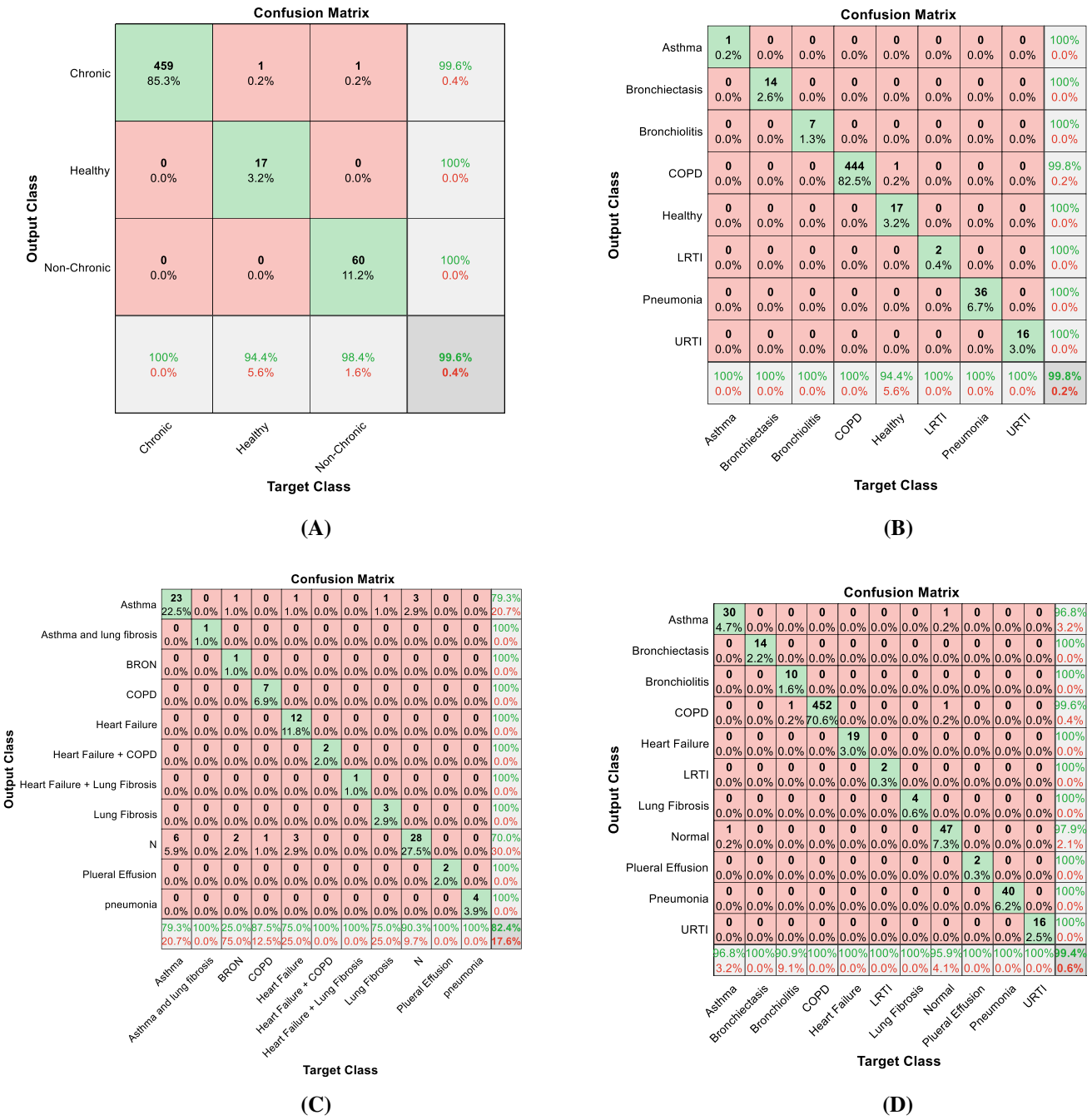

Fig. 9 CNN–LSTM model testing results for non-augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

combined to create a CNN model that can complete the assignment. CNN has excelled in a variety of scientific fields, particularly in the medical field (Chen et al. 2019; Esteva et al. 2021). Deep, representative, and discriminative characteristics are extracted primarily using CNN layers. When utilizing CNN layers, the preceding layers will do downsampling and feature selection, as well as generate categorization of the data. Figure 2A shows the used CNN Model.

3.3.2 LSTM model

Hochreiter and Schmidhuber first proposed the LSTM in 1997 (Shadmand and Mashoufi 2016), and a group led by Felix Gers improved it in 2000 (Gers et al. 2000). Researchers are now introducing many varieties of LSTM, with Zaremba et al. (2014) providing details on LSTM. A memory cell, input gate, output gate, and forget gate make up the most typical LSTM design.

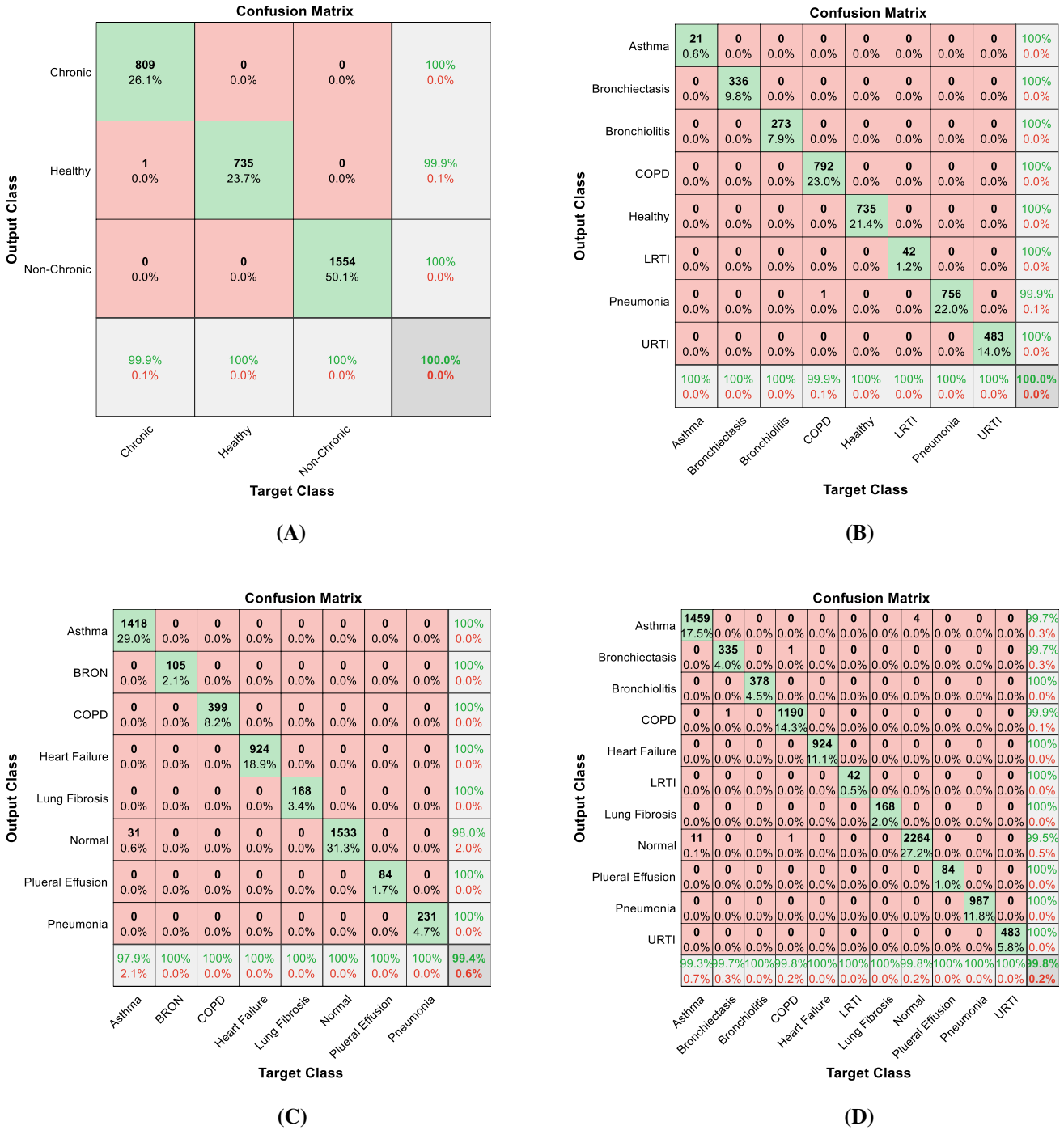

Fig. 10 CNN model training results for augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

Assume that the input, cell, and hidden states at iteration t are x_t, c_t, and h_t, respectively. The cell state $c\ t$ and hidden state h_{t-1} are produced for the current input x_t, the previous cell state c_{t-1}, and its corresponding previous hidden state h_{t-1} (Zaremba et al. 2014). Figure 2B shows the used LSTM model.

3.3.3 CNN–LSTM model

Deep feature extraction and selection from the ECG beat is handled by the CNN blocks, which are the 1D convolutional layer and the max pooling layer in this hybrid model. While the LSTM layer, which is fed these characteristics as time-dependent features, will learn to extract contextual time data (Shahzadi et al. 2018). Deep feature extraction

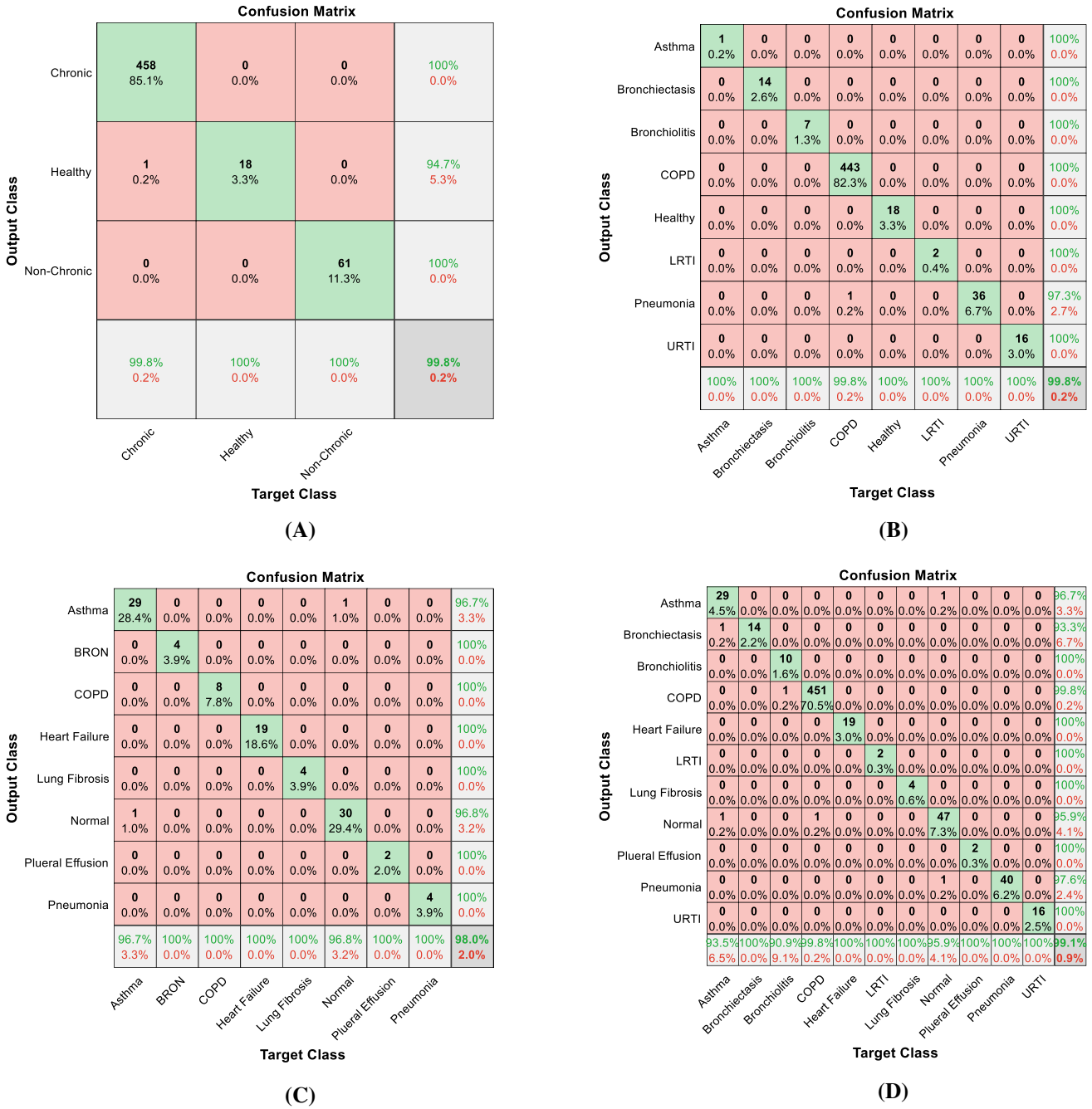

Fig. 11 CNN model testing results for augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset

and classification utilizing a hybrid 1D CNN–LSTM outperforms CNN- or LSTM-based methods, according to our research (She and Zhang 2018). Furthermore, using the LSTM layer allows for considerably shallower models to be built than pure CNN models. Our research shows that adopting a hybrid 1D CNN–LSTM for deep feature extraction and classification outperforms CNN- or LSTM-based methods. Furthermore, employing the LSTM layer allows for a significantly shallower model to be built than pure CNN models, resulting in better performance with fewer parameters. Figure 2C shows the used CNN–LSTM model.

3.4 Performance evaluation

Any artificial intelligence (AI)-based system must have a performance evaluation that corresponds to any new data. The original annotations of the raw lung auscultation sounds were compared to the identical lung sounds annotations predicted by the models to assess the performance

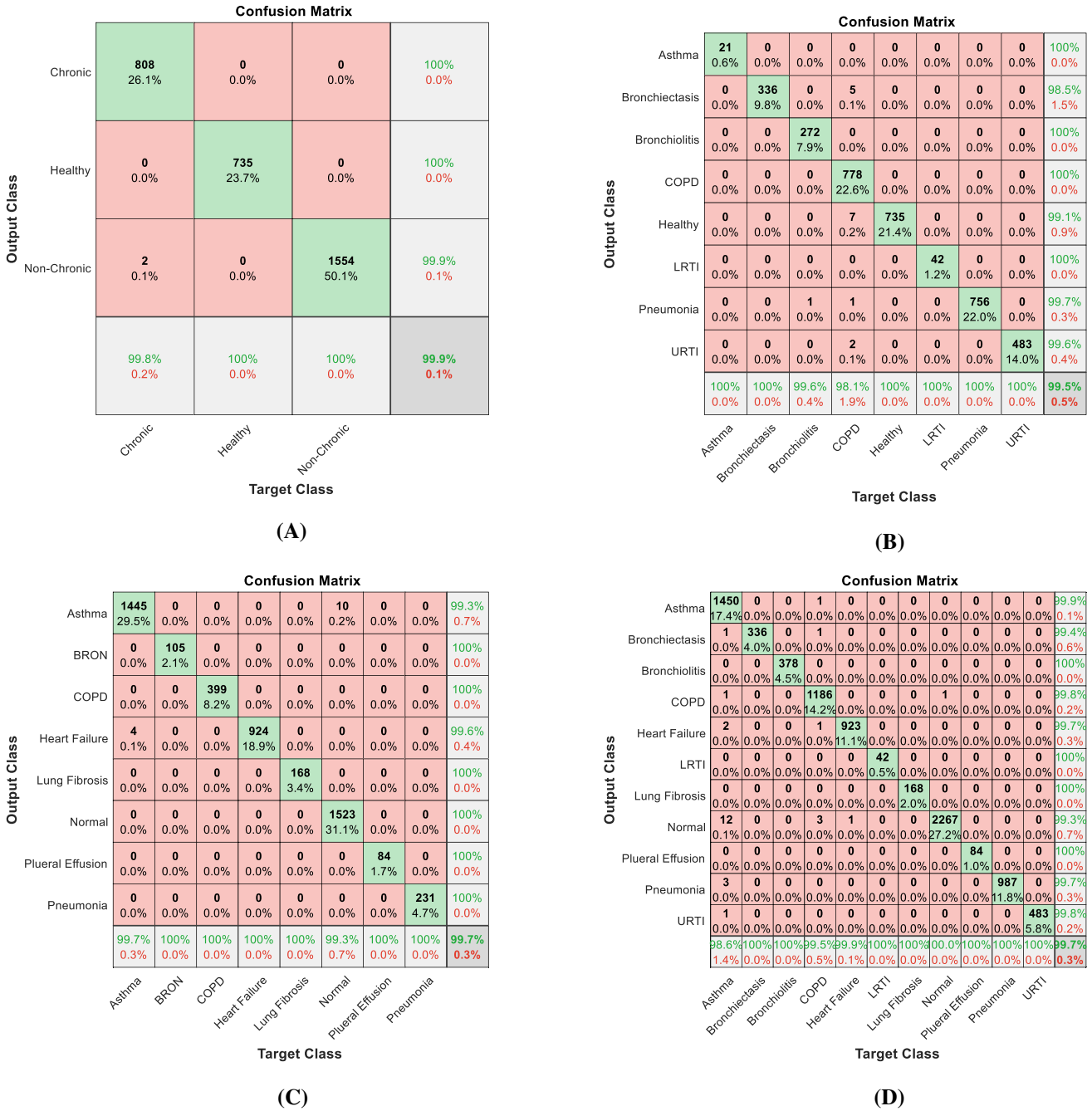

Fig. 12 LSTM model training results for augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

of the developed models. The accuracy, sensitivity, precision, and specificity of the data were then determined using these annotations. These indicators show how exactly lung sounds are diagnosed (Alqudah and Alqudah 2022a). True positive (TP), false positive (FP), false negative (FN), and true negative (TN) are four different sorts of statistical values used to calculate these measures (Kanavati et al. 2020; Alqudah et al. 2021c). All these parameters are extracted from the confusion matrix, the confusion matrix shows four main statistical indices which are used later to calculate performance metric (Obeidat and Alqudah 2021; Alqudah and Alqudah 2022b; Alqudah et al. 2021d), these indices are true positive (TP), false positive (FP), false negative (FN), and true negative (TN) (Alqudah et al. 2020; Al-Issa and Alqudah 2022). Figure 3 shows a simple confusion matrix. Then, the following performance evaluation parameters (accuracy, sensitivity, specificity, precision, F1 score, and MCC have been calculated using these values:

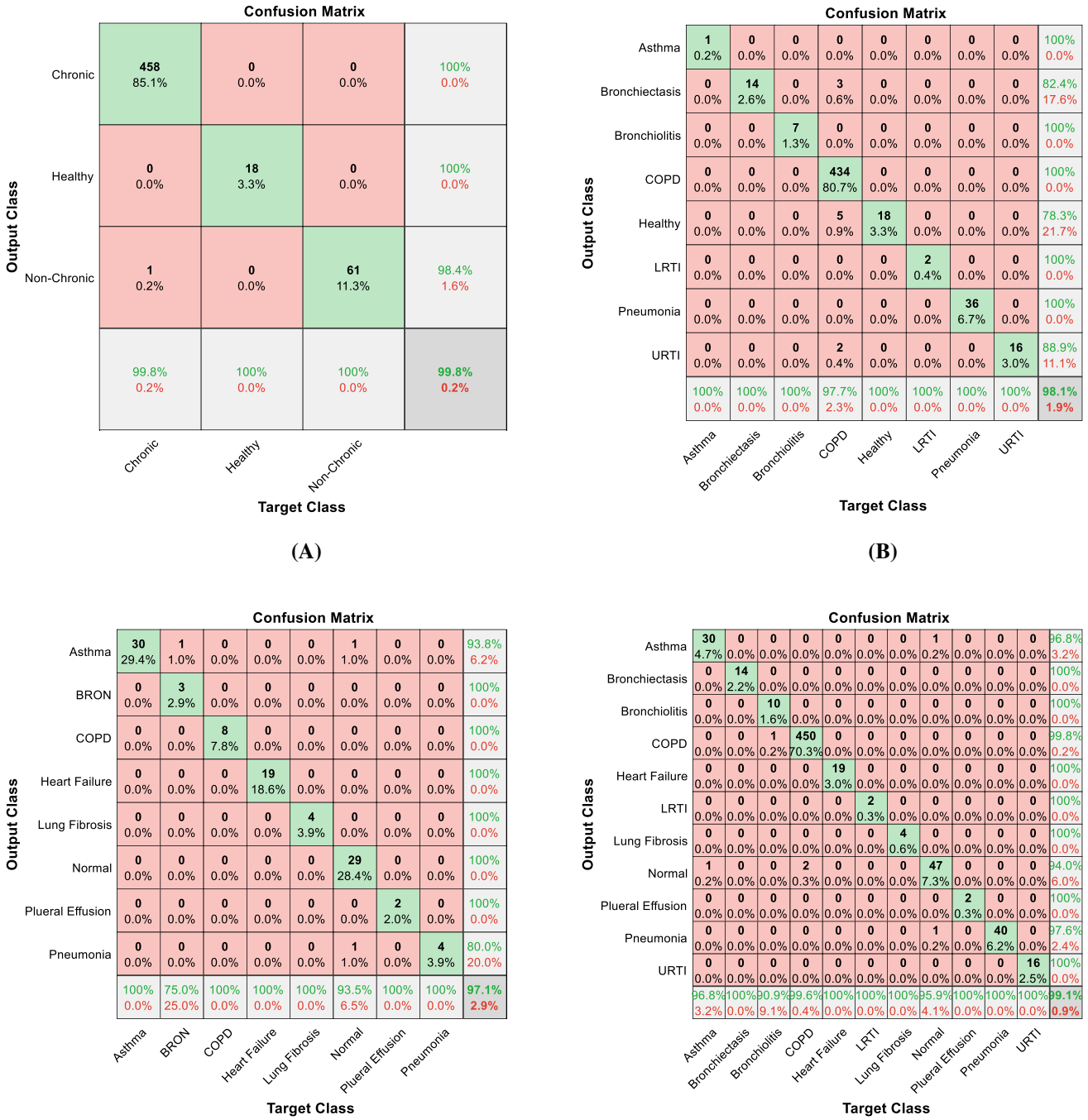

Fig. 13 LSTM model testing results for augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

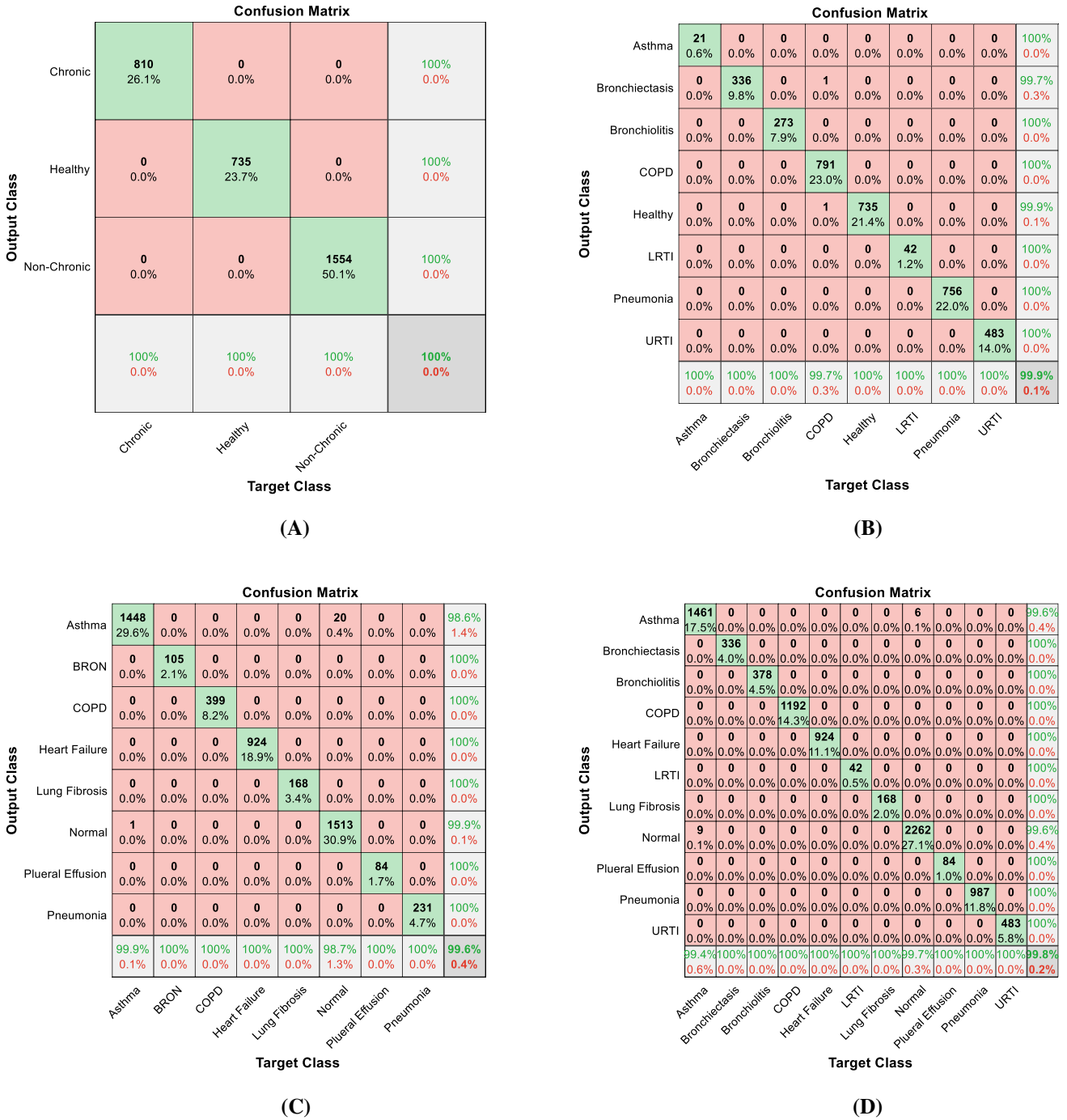

Fig. 14 CNN–LSTM model training results for augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

$$\text{F1 Score} = 2\frac{\text{Precision} * \text{Specificity}}{\text{Precision} + \text{Specificity}} \tag{5}$$

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{6}$$

4 The experimental results

The investigated models were trained tested using a computer with an Intel(R) Core TM i7-6700, a 3.40 GHz CPU, and 16 GB RAM. The training process for each model is taking around 45 min. All models are trained using Adam optimizer, initial learning rate of 0.001, max epochs of 100, mini batch size of 128, and validation frequency of 100.

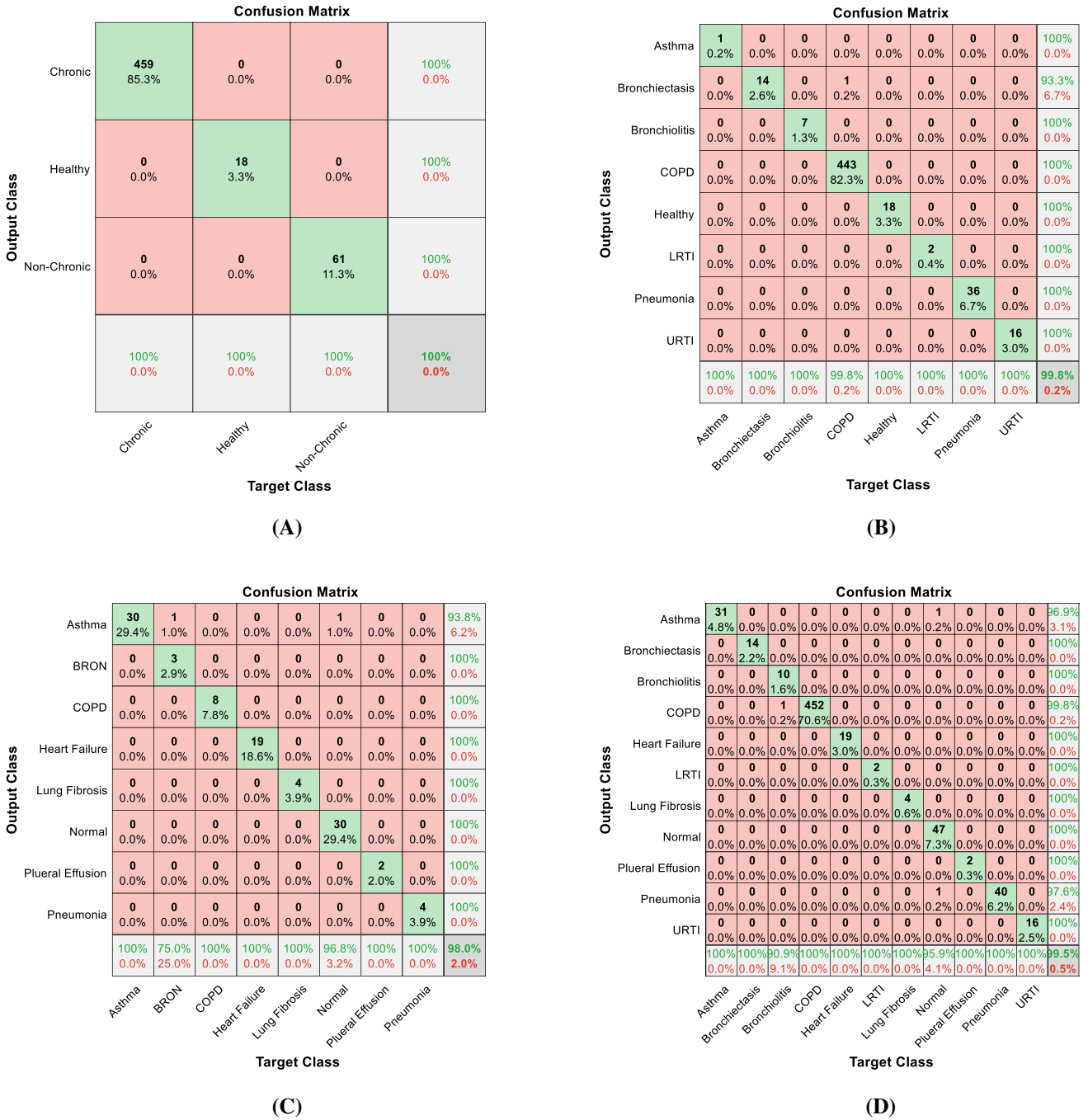

Fig. 15 CNN–LSTM model testing results for augmented datasets, **A** Dataset 1, **B** Dataset 2, **C** Dataset 3, **D** Dataset 4

The performances of all deep learning models are described using the confusion matrix and statistical parameters extracted from it. The confusion matrix represents the results of classification using a certain deep learning model. After calculating the statistical parameters, namely false positive (FP), false negative (FN), true positive (TP), and true negative (TN), the raw lung sound classification effectiveness is compared using the four statistical indices: sensitivity, specificity, precision, and accuracy. The performance evaluation results of all models for different types of raw lung sounds classification using different datasets are shown in the following subsections.

4.1 Non augmented dataset results

The following sections will show the performance of different deep learning models using non-augmented dataset.

Table 3 Comparison between different deep learning models among different datasets using testing sub-datasets

Dataset	Model											
	CNN				LSTM				CNN–LSTM			
	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre
Non-augmented												
1	99.62	98.90	99.15	99.85	99.81	98.14	99.15	99.85	99.62	97.60	99.15	99.85
2	99.62	98.87	99.73	99.94	99.25	98.87	99.73	99.94	99.81	99.30	99.86	99.97
3	80.39	76.86	97.53	94.78	79.41	76.86	97.53	94.78	82.35	84.73	97.71	95.39
4	99.53	98.69	99.92	99.50	99.21	98.69	99.92	99.50	99.37	98.50	99.87	99.47
Augmented												
1	99.81	99.92	99.93	98.24	99.81	99.92	99.93	98.24	100	100	100	100
2	99.81	99.97	99.97	99.66	98.14	99.97	99.97	99.66	99.81	99.97	99.97	99.16
3	98.03	99.18	99.65	99.18	97.05	99.18	99.65	99.18	98.03	96.47	99.65	99.21
4	99.06	98.19	99.87	98.47	99.06	98.19	99.87	98.47	99.53	98.8025	99.92	99.47

4.1.1 CNN model results

In this section we will display the results of CNN model on all of the non-augmented datasets, Fig. 4 shows the training results of all datasets using CNN model. While Fig. 5 shows the testing dataset results of all datasets using CNN model.

4.1.2 LSTM model results

In this section we will display the results of LSTM model on all of the non-augmented datasets, Fig. 6 shows the training results of all datasets using LSTM model. While Fig. 7 shows the testing dataset results of all datasets using LSTM model.

4.1.3 CNN–LSTM model results

In this section we will display the results of CNN–LSTM model on all of the non-augmented datasets, Fig. 8 shows the training results of all datasets using CNN–LSTM model. While Fig. 9 shows the testing dataset results of all datasets using CNN–LSTM model.

4.2 Augmented dataset results

The following sections will show the performance of different deep learning models using augmented dataset.

4.2.1 CNN model results

In this section we will display the results of CNN model on all of the augmented datasets, Fig. 10 shows the training results of all datasets using CNN model. While Fig. 11 shows the testing dataset results of all datasets using CNN model.

4.2.2 LSTM model results

In this section we will display the results of LSTM model on all of the augmented datasets, Fig. 12 shows the training results of all datasets using LSTM model. While Fig. 13 shows the testing dataset results of all datasets using LSTM model.

4.2.3 CNN–LSTM model results

In this section we will display the results of CNN–LSTM model on all of the augmented datasets, Fig. 14 shows the training results of all datasets using CNN–LSTM model. While Fig. 15 shows the testing dataset results of all datasets using CNN–LSTM model.

5 Discussion

In general, computer-aided system for detection of respiratory diseases can expedite diagnostic and treatment decisions and support the study of physiological patterns associated with various respiratory pathologies. In this work, we propose to apply different deep learning models to perform multi-class classification of different types of respiratory diseases. As imperative to all deep learning frameworks, the design stage of the deep learning models targeted toward providing a better classification of the raw lung sounds pattern. Thus, optimized models are key to building effective classification models and enhance the model's predictive

Table 4 Comparison between our proposed methods and literature

References #	Method	Performance
Bahoura (2009)	Features engineering + SVM	Accuracy: 94.2%
Jin et al. (2014)	Features engineering + SVM	Accuracy: 98.8%
Islam et al. (2018)	Features engineering + ANN and SVM	Accuracy-ANN: 91.7%
		Accuracy-SVM: 93.3%
Garcia-Ordas et al. (2020)	CNN	F-score: 0.993
Fraiwan et al. (2021b)	Features engineering + machine learning	Accuracy: 98.27%
		Sensitivity: 95.28%
		Specificity: 98.90%
Fraiwan et al. (2021c)	CNN + BDLSTM	Accuracy: 99.62%
		Sensitivity: 98.43%
		Specificity: 99.69%
Nguyen and Pernkopf (2022)	CNN + corrected and normalized spectrogram	Sensitivity: 95.76%
		Specificity: 91.77%
Tripathy et al. (2022)	Empirical wavelets transform with fixed boundary points and machine learning	Accuracy: 99.34 for normal to COPD Scenario
Bhatta et al. (2022)	CNN	Accuracy: 86%
Soni et al. (2022)	CNN for audio files and clinical data	AUC: 0.752
This study (non-augmented)	CNN (best using Dataset 1)	Accuracy: 99.63%
		Sensitivity: 98.91%
		Specificity: 99.16%
	LSTM (best using Dataset 1)	Accuracy: 99.81%
		Sensitivity: 99.81%
		Specificity: 99.93%
	CNN–LSTM (best using Dataset 2)	Accuracy: 99.81%
		Sensitivity: 99.31%
		Specificity: 99.87%
This study (augmented)	CNN (best using Datasets 2)	Accuracy: 99.81%
		Sensitivity: 99.97%
		Specificity: 99.98%
	LSTM (best using Dataset 1)	Accuracy: 99.81%
		Sensitivity: 99.93%
		Specificity: 99.93%
	CNN–LSTM (best using Dataset 1)	Accuracy: 100%
		Sensitivity: 100%
		Specificity: 100%

accuracy. In general, respiratory sounds are characterized as random and nonlinear signals that are highly complex in nature; Because of the fluctuating lung volume, this is especially true. These characteristics can be seen in both healthy and pathological individuals, but they are more noticeable in pathological lung sounds.

In this study, an investigation was carried out on the use of different deep learning models, as illustrated by the combination of CNN and LSTM neural networks, in identifying pulmonary diseases. The developed models achieved high levels of performance. The highest achieved accuracy/sensitivity/specificity were 100%, 100%, and 100% using the hybrid CNN–LSTM model, which paves the way toward implementing deep learning in clinical settings. Thus, a comparison of the results of the proposed deep learning models using either non-augmented or augmented datasets are shown in Table 3. Using Table 3 we can notice that the hybrid CNNLSTM model outperforms all other models (CNN and LSTM) either augmented or non-augmented datasets.

Clinically, the proposed research ensures the accurate detection and classification of different respiratory diseases from lung sounds. Unlike the traditional stethoscope where diseases are diagnosed manually and based on the practitioner experience, electronic lung sounds combined with a deep learning predictive models reduce the errors in

diseases classification. Therefore, many clinical decisions can be positively affected to prevent any further development of the diseases and any treatment. Furthermore, although manual diagnosis may lead to correct diagnosis in some circumstances, it is highly recommended clinically to build a model(s) that is able of detecting small variations in signals across patients. However, they may be highly affected by the patient-specific information within the same disease. Thus, a deep learning model that can learn from huge number of features which are which could automatically enrich the diagnosis process and act like a supportive decision maker in clinical settings. A comparison between the proposed models and models in the literature is shown in Table 4. Using Table 4 we can notice that our proposed models achieved a better performance than any model in the literature.

6 Conclusion

To sum up, this research paper, proposed a deep learning model based on convolutional neural networks (CNNs), long short-term memory (LSTM), and hybrid of them. These models were utilized for the purpose of raw lung sounds classification using combination of two datasets. The first dataset proposed dataset of lung sounds recorded at King Abdullah University Hospital (KAUH), while the other dataset is the dataset used in the international conference on Biomedical Health Informatics (ICBHI) 2017 Challenge. Different deep learning models like CNN, LSTM, and CNN–LSTM were employed as classification methods and were compared with using the achieved results. The experimental results showed that the hybrid CNN–LSTM classification model generally outperformed the CNN and LSTM methods which are commonly employed in the literature. This research paves the way toward designing and implementing deep learning models in clinical settings to assist clinicians in decision making with high accuracy. Future works will focus on increasing the number of classes of the used dataset to include more records from different subjects and a wider range of diseases such as COVID-19. Such future works will enhance the credibility of the proposed model. Although the current proposed classification deep learning models achieves high performance metrics, it would be improved by making more tuning on the hyperparameters.

For future works, we intend to expand the application of the proposed automatic lung auscultation sounds classification in real-time examination process. Initially, we intend to test the proposed method on more datasets. Then, we plan to develop an embedded system that integrate the developed CNN–LSTM model into the system with a digital stethoscope. Furthermore, we also aim to create an Internet of Things (IoT) system for this approach to be accessible in developing countries, where we find the highest mortality rates. Finally, we intend to analyze the feasibility of the system with the telemedicine system.

7 Limitations

We must acknowledge the limitations of current research. As this research collects data from two datasets there are a few records for some classes, while some classes have a large number of samples. So, the main limitation is the number of samples per class and the required future effort to make records for all diseases. Furthermore, other factors may affect the results like medications used by patients may contribute to the development or deterioration of symptoms; however, the current research did not investigate the effect of these factors.

Acknowledgements The authors would like to thank the authors of the used datasets to make their datasets available online for research use. Also, the authors would like to thank the anonymous reviewers for their valuable comments.

Author's contribution AMQ accomplished the deep learning models development and results generation. SQ and YO accomplished the data and results analysis. All authors accomplished the manuscript writing and revising. All authors read and approved the final manuscript.

Funding This work does not receive any funding or support.

Availability of data and materials The datasets used or analyzed during the current study are available online for free from their corresponding authors.

Declarations

Conflict of interest The authors declare no conflict of interests.

References

Al-Issa Y, Alqudah AM (2022) A lightweight hybrid deep learning system for cardiac valvular disease classification. Sci Rep 12(1):1–20

Alqudah AM (2020) AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. Med Biol Eng Comput 58(1):41–53

Alqudah AM, Alqudah A (2022a) Improving machine learning recognition of colorectal cancer using 3D GLCM applied to different color spaces. Multimed Tools Appl 81(8):10839–10860

Alqudah A, Alqudah AM (2022b) Deep learning for single-lead ECG beat arrhythmia-type detection using novel iris spectrogram representation. Soft Comput 26(3):1123–1139

Alqudah AM, Alquran H, Abu Qasmieh I (2020) Classification of heart sound short records using bispectrum analysis approach

images and deep learning. Netw Model Anal Health Inform Bioinform 9(1):1–16
Alqudah AM, Qazan S, Al-Ebbini L, Alquran H, Qasmieh IA (2021a) ECG heartbeat arrhythmias classification: a comparison study between different types of spectrum representation and convolutional neural networks architectures. J Ambient Intell Hum Comput 13:4877–4907
Alqudah A, Alqudah AM, Alquran H, Al-Zoubi HR, Al-Qodah M, Al-Khassaweneh MA (2021b) Recognition of handwritten arabic and hindi numerals using convolutional neural networks. Appl Sci 11(4):1573
Alqudah AM, Qazan S, Masad IS (2021c) Artificial intelligence framework for efficient detection and classification of pneumonia using chest radiography images. J Med Biol Eng 41:599–609
Alqudah A, Alqudah AM, AlTantawi M (2021d) Artificial intelligence hybrid system for enhancing retinal diseases classification using automated deep features extracted from OCT images. Int J Intell Syst Appl Eng 9(3):91–100
Andrade DS, Ribeiro LM, Lopes AJ, Amaral JL, Melo PL (2021) Machine learning associated with respiratory oscillometry: a computer-aided diagnosis system for the detection of respiratory abnormalities in systemic sclerosis. Biomed Eng Online 20(1):1–18
Aykanat M, Kilic O, Kurt B, Saryal S (2017) Classification of lung sounds using convolutional neural networks. EURASIP J Image Video Process 2017(1):1–9
Bahoura M (2009) Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Comput Biol Med 39(9):824–843
Bardou D, Zhang K, Ahmad SM (2018) Lung sounds classification using convolutional neural networks. Artif Intell Med 88:58–69
Bhatta LN, Bhatta SM, Akshay N (2022) Respiratory analysis—detection of various lung diseases using audio signals. In: 2022 8th International conference on advanced computing and communication systems (ICACCS), vol 1. IEEE
Chang G-C, Cheng Y-P (2008) Investigation of noise effect on lung sound recognition. In: 2008 International conference on machine learning and cybernetics, vol 3. IEEE, pp 1298–1301
Chang G-C, Lai Y-F (2010) Performance evaluation and enhancement of lung sound recognition system in two real noisy environments. Comput Methods Prog Biomed 97(2):141–150
Chen H, Yuan X, Pei Z, Li M, Li J (2019) Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. IEEE Access 7:32845–32852
Cinar A, Tuncer SA (2021) Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN–SVM deep neural networks. Comput Methods Biomech Biomed Eng 24(2):203–214
Demir F, Ismael AM, Sengur A (2020) Classification of lung sounds with CNN model using parallel pooling structure. IEEE Access 8:105376–105383
Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, Liu Y, Topol E, Dean J, Socher R (2021) Deep learning-enabled medical computer vision. NPJ Digit Med 4(1):1–9
Fraiwan M, Fraiwan L, Khassawneh B, Ibnian A (2021a) A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. Data Brief 35:106913
Fraiwan L, Hassanin O, Fraiwan M, Khassawneh B, Ibnian AM, Alkhodari M (2021b) Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. Biocybern Biomed Eng 41(1):1–14
Fraiwan M, Fraiwan L, Alkhodari M, Hassanin O (2021c) Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. J Ambient Intell Hum Comput 13:4759–4771
Garcia-Ordas MT, Benitez-Andrades JA, Garcia-Rodriguez I, Benavides C, Alaiz-Moretfion H (2020) Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. Sensors 20(4):1214
Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. Neural Comput 12(10):2451–2471
Guler I, Polat H, Ergun U (2005) Combining neural network and genetic algorithm for prediction of lung sounds. J Med Syst 29(3):217–231
Icer S, Gengec S (2014) Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. Digit Signal Process 28:18–27
Islam MA, Bandyopadhyaya I, Bhattacharyya P, Saha G (2018) Multichannel lung sound analysis for asthma detection. Comput Methods Prog Biomed 159:111–123
Jacome C, Ravn J, Holsbi E, Aviles-Solis JC, Melbye H, Ailo Bongo L (2019) Convolutional neural network for breathing phase detection in lung sounds. Sensors 19(8):1798
Jelodar H, Wang Y, Orji R, Huang S (2020) Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: NLP using LSTM recurrent neural network approach. IEEE J Biomed Health Inform 24(10):2733–2742
Jin F, Sattar F, Goh DY (2014) New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. Neurocomputing 123:362–371
Kanavati F, Toyokawa G, Momosaki S, Rambeau M, Kozuma Y, Shoji F, Yamazaki K, Takeo S, Iizuka O, Tsuneki M (2020) Weakly-supervised learning for lung carcinoma classification using deep learning. Sci Rep 10(1):1–11
Kevat A, Kalirajah A, Roseby R (2020) Artificial intelligence accuracy in detecting pathological breath sounds in children using digital stethoscopes. Respir Res 21(1):1–6
Khodabakhshi MB, Moradi MH (2017) The attractor recurrent neural network based on fuzzy functions: an effective model for the classification of lung abnormalities. Comput Biol Med 84:124–136
LeCun Y, Bengio Y et al (1995) Convolutional networks for images, speech, and time series. In: The handbook of brain theory and neural networks, vol 3361, No. 10, p 1995
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
Lehrer S (2018) Understanding lung sounds. Steven Lehrer, New York
Maruf SO, Azhar MU, Khawaja SG, Akram MU (2015) Crackle separation and classification from normal respiratory sounds using Gaussian mixture model. In: 2015 IEEE 10th International conference on industrial and information systems (ICIIS). IEEE, pp 267–271
Moussavi Z (2006) Fundamentals of respiratory sounds and analysis. Synth Lect Biomed Eng 1(1):1–68
Naves R, Barbosa BH, Ferreira DD (2016) Classification of lung sounds using higher-order statistics: a divide-and-conquer approach. Comput Methods Prog Biomed 129:12–20
Nguyen T, Pernkopf F (2022) Lung sound classification using co-tuning and stochastic normalization. IEEE Trans Biomed Eng 69:2872–2882
Nuckowska MK, Gruszecki M, Kot J, Wolf J, Guminski W, Frydrychowski AF, Wtorek J, Narkiewicz K, Winklewski PJ (2019) Impact of slow breathing on the blood pressure and subarachnoid space width oscillations in humans. Sci Rep 9(1):1–13
Obeidat Y, Alqudah AM (2021) A hybrid lightweight 1D CNN–LSTM architecture for automated ECG beat-wise classification. Traitement Du Signal 38(5):1281–1291

Orjuela-Canon AD, Gomez-Cajas DF, Jimenez-Moreno R (2014) Artificial neural networks for acoustic lung signals classification. In: Iberoamerican Congress on pattern recognition. Springer, Berlin, pp 214–221

Ozturk S, Ozkaya U (2020) Gastrointestinal tract classification using improved LSTM based CNN. Multimed Tools Appl 79(39):28825–28840

Palaniappan R, Sundaraj K, Sundaraj S (2014) Artificial intelligence techniques used in respiratory sound analysis—a systematic review. Biomed Tech Biomed Eng 59(1):7–18

Petmezas G, Haris K, Stefanopoulos L, Kilintzis V, Tzavelis A, Rogers JA, Katsaggelos AK, Maglaveras N (2021) Automated atrial fibrillation detection using a hybrid CNN–LSTM network on imbalanced ECG datasets. Biomed Signal Process Control 63:102194

Pinho C, Oliveira A, Jacome C, Rodrigues J, Marques A (2015) Automatic crackle detection algorithm based on fractal dimension and box filtering. Procedia Comput Sci 64:705–712

Reyes BA, Charleston-Villalobos S, Gonzalez-Camarena R, Aljama-Corrales T (2014) Assessment of time–frequency representation techniques for thoracic sounds analysis. Comput Methods Prog Biomed 114(3):276–290

Rocha B, Filos D, Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, Natsiavas P, Oliveira A, Jacome C, Marques A et al (2017) A respiratory sound database for the development of automated classification. In: International conference on biomedical and health informatics. Springer, Berlin, pp 33–37

Sengupta N, Sahidullah M, Saha G (2016) Lung sound classification using cepstral-based statistical features. Comput Biol Med 75:118–129

Serbes G, Sakar CO, Kahya YP, Aydin N (2013) Pulmonary crackle detection using time–frequency and time-scale analysis. Digit Signal Process 23(3):1012–1021

Shadmand S, Mashoufi B (2016) A new personalized ECG signal classification algorithm using block-based neural network and particle swarm optimization. Biomed Signal Process Control 25:12–23

Shahzadi I, Tang TB, Meriadeau F, Quyyum A (2018) CNN–LSTM: cascaded framework for brain tumour classification. In: 2018 IEEE-EMBS Conference on biomedical engineering and sciences (IECBES). IEEE, pp 633–637

She X, Zhang D (2018) Text classification based on hybrid CNN–LSTM hybrid model. In: 2018 11th International symposium on computational intelligence and design (ISCID), vol 2. IEEE, pp 185–189

Shuvo SB, Ali SN, Swapnil SI, Hasan T, Bhuiyan MIH (2020) A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram. IEEE J Biomed Health Inform 25:2595–2603

Soni PN et al (2022) Contrastive learning of heart and lung sounds for label-efficient diagnosis. Patterns 3(1):100400

Tripathy RK et al (2022) Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform. IEEE Sens Lett 6(5):1–4

Tsai K-H, Wang W-C, Cheng C-H, Tsai C-Y, Wang J-K, Lin T-H, Fang S-H, Chen L-C, Tsao Y (2020) Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder. IEEE J Biomed Health Inform 24(11):3203–3214

Wani A, Joshi I, Khandve S, Wagh V, Joshi R (2021) Evaluating deep learning approaches for covid19 fake news detection. In: Combating online hostile posts in regional languages during emergency situation: first international workshop, CONSTRAINT 2021, collocated with AAAI 2021, Virtual Event, February 8, 2021, revised selected papers. Springer Nature, p 153

Zaremba W, Sutskever I, Vinyals O (2014) Recurrent neural network regularization. arXiv preprint arXiv:1409.2329

Zhang K, Wang X, Han F, Zhao H (2015) The detection of crackles based on mathematical morphology in spectrogram analysis. Technol Health Care 23(s2):489–494

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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.