A Hybrid Deep Learning Approach for Early Detection of Chronic Obstructive Pulmonary Disease

Lun-Ping Hung[1(✉)], Hsiang-Tsung Yeh[1], Zong-Jie Wu[2], and Chien-Liang Chen[3]

[1] Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei 112, Taiwan
lunping@ntunhs.edu.tw

[2] Department of Industrial Engineering and Management, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan

[3] Department of Innovative Living Design, Overseas Chinese University, Taichung 40721, Taiwan

Abstract. Chronic obstructive pulmonary disease (COPD) is currently the third leading cause of death worldwide. Early detection can help treat the disease and delay its progression. However, chronic diseases are difficult to detect and symptoms often have to develop into severe conditions before they become apparent. Currently, physicians use artificial auscultation as a preliminary means of diagnosing COPD. By detecting the respiratory acoustic phenomena and analyzing the pathology knowledge of patients, physicians can infer and analyze the disease. However, this method still has the possibility of misjudgment or delayed treatment. Therefore, this study uses the ICBHI Respiratory Sound Database Dataset as the basis for analysis data set under the deep learning technology with convolutional neural network models, we classify the features of lung sounds and hope to construct an identification tool that assists in diagnosing COPD. In addition to reducing the time cost of traditional auscultation with this auxiliary tool, after evaluating the model's effectiveness with confusion matrix and accuracy evaluation, we especially estimate its correctness and practical applicability. In the future, it can be recommended for clinical diagnosis and development of an auxiliary diagnosis tool that helps provide early diagnosis of COPD.

Keywords: Deep Learning · Chronic Obstructive Pulmonary Disease · Auxiliary Diagnosis

1 Introduction

Particulate matter (PM) in air pollution is a major factor causing harm to the human body and is also one of the indicators used to assess the level of pollution. When the particle size of suspended matter is equal to or less than 10 μm, it is referred to as PM10. Particles with a size of equal to or less than 2.5 μm are known as fine particulate matter (PM2.5) [1]. Their main components include sulfates, nitrates, ammonia, sodium

D.-J. Deng and J.-C. Chen (Eds.): SGIoT 2023, LNICST 557, pp. 113–121, 2024.
https://doi.org/10.1007/978-3-031-55976-1_11

chloride, black carbon, mineral dust, and water. They also include organic and inorganic complex mixtures suspended in the air [2]. The size of particulate matter determines the final deposition location after inhalation, and PM2.5 can even reach deeper respiratory organs, interfering with gas exchange in the lungs. When the 24-h average concentration of PM2.5 exceeds 35 μg/m^3, it poses a risk to sensitive individuals, such as the elderly, children, and patients with respiratory diseases. If it exceeds 65 μg/m^3, it can harm the health of the general population. According to the World Health Organization, common diseases leading to death due to air pollution include cardiovascular diseases, chronic obstructive pulmonary disease, and cancer [3].

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of death worldwide due to diseases. Early detection of COPD is crucial as it allows for timely intervention and the ability to slow down disease progression. The longer COPD remains undetected and untreated, the more severe the impact on lung function becomes. Once lung function is compromised, it can lead to complications such as cardiovascular diseases and respiratory failure, accelerating the time to death. This is why COPD is often referred to as the "silent killer of the lungs."

Due to the chronic nature of Chronic Obstructive Pulmonary Disease (COPD), it is often challenging to detect symptoms in daily life. The respiratory system, which is essential for human survival, serves as the medium through which COPD develops. This means that anyone can unknowingly develop COPD. By the time noticeable symptoms appear, irreversible damage may have already been done to the body. Dealing with such a large number of patients seeking medical attention for COPD puts an immeasurable burden and pressure on the healthcare system, leading to significant costs and strain on medical resources.

Indeed, auscultation is the primary method for the initial diagnosis of respiratory diseases. However, it is a subjective and highly variable diagnostic approach. It requires sensitive perception and extensive experience from skilled physicians who can combine the identification of specific acoustic phenomena with their medical knowledge to accurately diagnose the disease and determine the appropriate treatment methods. The reliance on the expertise of experienced professionals makes it difficult to standardize the process and can lead to variability in diagnoses among different practitioners. As a result, there is a need for additional tools and technologies to complement auscultation and improve the accuracy and efficiency of diagnosing respiratory conditions such as COPD.

Professional diagnosis of respiratory diseases involves several diagnostic processes, such as spirometry, lung volume measurement, and bronchodilator response testing. Spirometry is used to measure the volume and flow rate of air exhaled from the lungs during maximal inhalation. Combined with the mMRC scale and CAT scale analysis, it helps classify patients into four levels and determines appropriate follow-up procedures for each level. Lung volume measurement is essential for assessing total lung capacity and residual lung volume, enabling the differentiation between obstructive and restrictive diseases. If asthma symptoms are present, bronchodilator response testing is conducted to compare the changes before and after the administration of a bronchodilator spray. Significant improvements in airway obstruction after treatment indicate asthma, while the absence of such improvements may suggest Chronic Obstructive Pulmonary Disease (COPD) [4].

The complex and time-consuming nature of these professional diagnostic processes can take anywhere from 10 to 30 min, and with a high volume of patients seeking medical attention, the time and manpower costs can become overwhelming. This research aims to provide physicians with preliminary assistance in identifying COPD and to assist the general public in developing their own preliminary awareness and judgment of the disease.

2 Literature Review

2.1 Application of Artificial Intelligence Technology in Chronic Obstructive Pulmonary Disease

Artificial intelligence covers the fields of machine learning and deep learning. Some researchers have proposed methods for diagnosing respiratory diseases and many studies related to remote healthcare applications have emerged. Priyanka Choudhury et al. Used supervised machine learning algorithms to distinguish between asthma and chronic obstructive pulmonary disease (COPD) by training models with data from multiple interleukins in blood tests [5]. Dat Tran-Anh et al. Combined the Internet of Things with deep learning to create wearable devices that collect respiratory sounds through the mouth. They used a deep learning model based on the SincNet convolutional neural network to classify deep breathing, heavy breathing (respiratory distress), and normal breathing [6]. Binson V.A. et al. Developed an electronic nose detection system, primarily using integrated machine learning algorithms to analyze volatile organic compounds in exhalations to classify COPD, lung cancer, and health. They validated the results using confusion matrices and performance metrics such as accuracy, sensitivity, and specificity [7].

Due to chronic obstructive pulmonary disease-causing airway obstruction, abnormal respiratory sounds such as wheezing and crackles can be detected through auscultation during the breathing process. Many studies focusing on respiratory sound analysis have emerged, Nishi Shahnaj Haider et al. Classifying asthma, chronic obstructive pulmonary disease, and healthy individuals. They utilize wavelet transformation to extract acoustic features and compare the classification results of four machine learning algorithms using a confusion matrix and accuracy, sensitivity, and specificity metrics [8]. Paweł Stasiakiewicz et al. Use wavelet transformation as acoustic features and employ the support vector machine (SVM) algorithm for classification to differentiate pneumonia, pulmonary fibrosis, heart failure, and chronic obstructive pulmonary disease based on sonorous data, Finally, they evaluate the model results based on accuracy, sensitivity, and specificity indicators [9]. Murat Aykanat et al. Utilized Mel frequency cepstral coefficients as acoustic features and spectrograms as image features to classify different respiratory conditions based on lung sound data. The classification results of Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models were evaluated using accuracy, recall, sensitivity, and specificity [10].

2.2 Deep Learning

Deep Learning (DL), one of the methods of artificial intelligence and a branch of machine learning, was proposed by K. Fukushima in 1980 [11]. Its algorithm is based on neural network models inspired by the human brain. In this model, neurons in the human brain are analogous to nodes in computer calculations, and the connections between neurons are analogous to the weights between nodes. The architecture mimics human sensory input, training logic, and classification results, with input layers, hidden layers, and output layers. The most significant difference from traditional machine learning is its ability to effectively handle unstructured data, such as language sentences, images, sounds, and more. Unstructured data implies diverse content, and deep learning can provide the model with multidimensional input data from various perspectives to improve classification performance [12]. Additionally, in deep learning, transfer learning allows for dynamic updates of classifiers, eliminating the need to incur additional time and cost when incorporating new data.

2.3 Acoustic Features

The extraction of acoustic features is a crucial step in acoustic recognition, involving the analysis of sound amplitude signals. Mel-scale Frequency Cepstral Coefficients (MFCC) is a common method for acoustic feature extraction. Human hearing is less sensitive to high frequencies, and MFCC calculations take advantage of this characteristic by using Mel-scale filter banks that approximate the human ear's equal-loudness frequency response. These filters transform the signal into appropriately sized acoustic features. The process begins with initial audio preprocessing, followed by a fast Fourier transform to convert the time-domain signal into the frequency-domain signal. Subsequently, a series of mathematical formulae compress and transform the data into acoustic features that represent the corresponding audio segment [13]. Mel Frequency Cepstral Coefficients (MFCC) Framework Flowchart, as shown in Fig. 1:

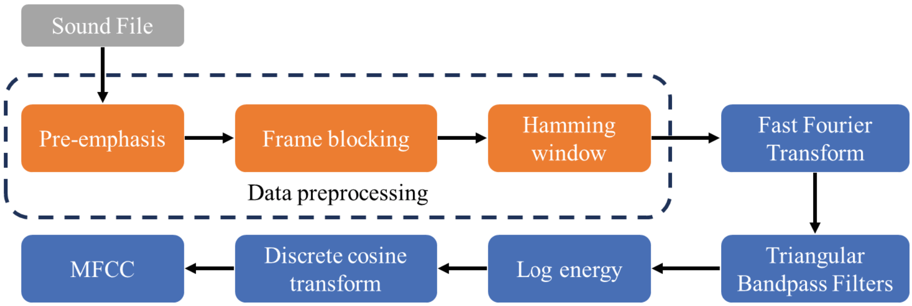

Fig. 1. MFCC Framework Flowchart

Pre-emphasis is the process of passing an audio signal through a high-pass filter. Its purpose is to eliminate noise from vocal cords and lips during the speech production process, primarily emphasizing the resonance peaks in the high-frequency range. Frame blocking involves grouping several sample points into a single observation unit. The Hamming window multiplies each audio frame to enhance the continuity at the left and right ends of the frame. Because it's not easy to observe the characteristics of a sound signal in the time domain, the Fast Fourier Transform (FFT) is used to transform it into the frequency domain to examine the energy distribution. Different energy distributions can represent different sound characteristics. The transformed spectral energy is multiplied by a set of 20 triangular bandpass filters to compute the logarithmic energy output of each filter, which represents the energy of a frame and is an important feature of speech. The logarithmic energy, known as log energy, is the volume of a frame and a crucial feature of speech. The 20 log energy values are then used in the Discrete Cosine Transform (DCT) to calculate the Nth-order Mel Frequency Cepstral Coefficients (MFCCs), with N being a user-selectable parameter. Finally, in practical speech recognition applications, delta cepstral coefficients are often added to show how the MFCCs change over time, resulting in Mel Cepstral coefficients.

The Continuous Wavelet Transform (CWT) is also a tool for acoustic analysis. Its core functionality involves providing different lengths of Hamming window scales based on different frequencies. The calculation is based on translation parameters and scale parameters, and it calculates the continuous wavelet transform parameters according to a user-defined number of frequencies, as show in Fig. 2. This study will apply the two aforementioned acoustic feature extraction methods to the recognition model dataset.

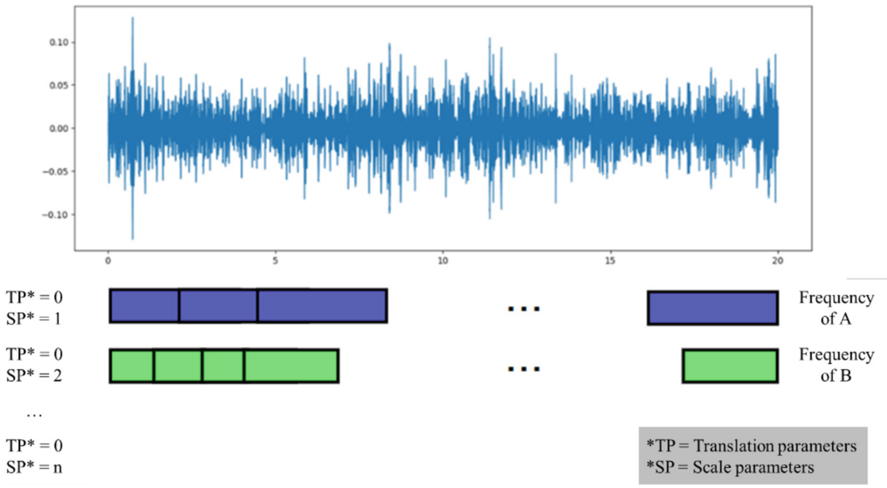

Fig. 2. Continuous Wavelet Transform Operation Visual Chart

3 Method

3.1 Dataset

In this study, the dataset utilizes respiratory sound data provided by the International Conference on Biomedical and Health Informatics (ICBHI). This data was collected by two different research teams from two different countries. The majority of the data was collected in collaboration between the Health Technology Institute of the University of Aveiro and Infante D. Pedro Hospital in Portugal. The remaining portion of the dataset was collected through a collaborative effort involving Aristotle University of Thessaloniki, University of Coimbra, Papanikolaou General Hospital, and Imathia General Hospital in Greece. The dataset encompasses various types of content, including asthma, bronchiectasis, bronchitis, chronic obstructive pulmonary disease, pneumonia, upper respiratory tract infections, lower respiratory tract infections, and healthy cases. Data was collected from seven different anatomical locations, including the front chest on both sides, back on both sides, sides of the body, and the trachea. The data was recorded using a variety of recording equipment, including lapel-style capacitive microphones, traditional stethoscopes, and two electronic stethoscopes, in order to simulate real-world conditions more accurately. The dataset comprises a total of 126 participants and 920 audio files in the WAV format [14].

3.2 Dataset Preprocessing

Data preprocessing is a crucial step before model training. Many unnoticed issues can lead to the model not functioning properly or achieving subpar results, such as incorrect input data or missing values. To ensure the smooth training of the model and even expedite the training process, this stage is divided into three steps: data selection, acoustic feature transformation, and data splitting.In the first step, data selection, given that the research aims to provide recommendations for the diagnosis of chronic obstructive pulmonary disease (COPD), two categories were selected: cases of COPD and healthy cases. There were 64 cases of COPD and 26 healthy cases, totaling 90 individuals and 828 WAV audio files.In the second step, acoustic feature transformation, all audio files were converted into Mel-frequency cepstral coefficients (MFCC) and spectrograms using the librosa audio package in a Python environment. Spectrograms were primarily used as the data for this research's model.Finally, to facilitate the model training, improve model performance, and evaluate model results, the data was split into training, testing, and validation sets. The training set was used for model construction, the testing set for optimization during training, and the validation set to assess the model's effectiveness after construction. The data was initially divided into a 90% training and testing set and a 10% validation set. Within the 90% training and testing set, an 80% training set and a 20% testing set were further split.

3.3 Analysis Model

This study will use the convolutional neural network SincNet mentioned in the literature as the foundation for constructing the training model. We will adjust certain internal parameters based on the aforementioned pre-processed dataset and proceed with the training of the COPD recognition model.

Create a simple one-dimensional CNN model from scratch. By observing that the first data point has 193 parameters, the input layer is set to (193, 1). There are 7 hidden layers, all using the ReLU activation function. The final fully connected layer uses the sigmoid function, suitable for binary classification output. The loss function used is also appropriate for binary classification, binary_crossentropy. Figure 3 shows the model design source code.

```
model = Sequential()
model.add(Conv1D(64, kernel_size=5, activation='relu', input_shape=(193, 1)))

model.add(Conv1D(128, kernel_size=5, activation='relu'))
model.add(MaxPooling1D(2))

model.add(Conv1D(256, kernel_size=5, activation='relu'))

model.add(Dropout(0.3))
model.add(Flatten())

model.add(Dense(512, activation='relu'))
model.add(Dense(2, activation='sigmoid'))

model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

Fig. 3. One-dimensional CNN Model Building Code

3.4 Predictive Assessment

To understand if any unexpected anomalies occur during the model training process, one can observe the changes in model accuracy and loss values to identify issues with the training and test sets. Ideally, as the model's accuracy increases, the loss values decrease. However, it is also possible to encounter situations of overfitting and underfitting. When overfitting occurs, it means that the model performs very well on the training data, resulting in low loss values, but it performs poorly during testing, leading to low accuracy. In the case of underfitting, the model exhibits poorer performance on the training data.

To ensure the feasibility of the diagnostic recommendations provided, this study will utilize a validation set to evaluate the model's performance by calculating metrics in the form of a Confusion Matrix:

- TP (true positive): Actual cases of COPD correctly identified as COPD.
- TN (true negative): Actual cases without COPD correctly identified as non-COPD.
- FP (false positive): Actual cases with COPD incorrectly identified as non-COPD.
- FN (false negative): Actual cases without COPD incorrectly identified as COPD.

Using these metrics from the Confusion Matrix, the following indicator can be calculated:

Accuracy is the ratio of correctly classified samples in the validation dataset to the total number of samples, and it can serve as a comprehensive performance score for the model.

4 Conclusion

Currently, a complete dataset has been obtained and two categories, COPD and healthy, have been selected. Data preprocessing has also been completed, with audio files transformed into Mel-frequency cepstral coefficient (MFCC) features. Successful experiments in building recognition models on this dataset using methods from other domains have been conducted, verifying the feasibility of using this dataset in our research. From Fig. 4, it can be observed that the trained model's recognition results, with the SincNet model and data preprocessing settings, exhibit good recognition performance.

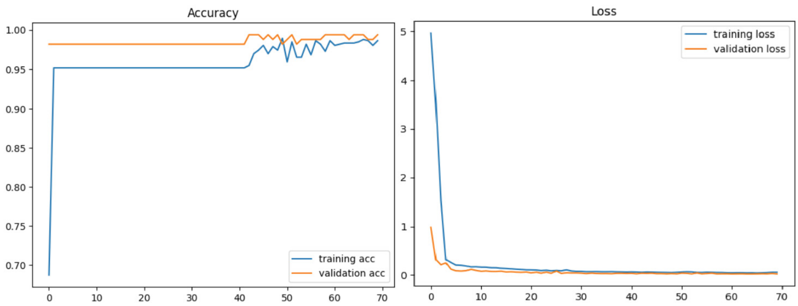

Fig. 4. Training process of a one-dimensional CNN model

5 Discussion

This study utilized the aforementioned publicly available dataset and, through data preprocessing and model development, achieved promising initial results. It is our hope that this research can provide recommendations for predicting the presence of chronic obstructive pulmonary disease (COPD) based on acoustic respiratory features, offering objective diagnostic suggestions for healthcare professionals in the future. We also intend to incorporate other models to assess their effectiveness and feasibility while exploring the relevant features of COPD with input from other researchers and expert physicians. This will enable the construction of a more precise and effective convolutional neural network model for deep learning, which can be used to provide valuable predictions and guidance for COPD.

References

1. Health Promotion Administration-Ministry of Health and Welfare. Air pollution health self-protection area (2022). Available from: https://www.hpa.gov.tw/pages/list.aspx?nodeid=441

2. World Health Organization: Ambient (outdoor) air pollution (2022). Available from: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health
3. World Health Organization: 7 million premature deaths annually linked to air pollution (2014). Available from: https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution
4. Taichung Hospital-Ministry of Health and Welfare: Pulmonary function tests; Available from: https://www.taic.mohw.gov.tw/?aid=52&pid=60&page_name=detail&iid=243
5. Choudhury, P., Biswas, S., Singh, G., Pal, A., Ghosh, N., Ojha, A.K.: Chaudhury, K, Immunological profiling and development of a sensing device for detection of IL-13 in COPD and asthma. Bioelectrochemistry **143**, 107971 (2022)
6. Tran-Anh, D., Vu, N.H., Nguyen-Trong, K., Pham, C.: Multi-task learning neural networks for breath sound detection and classification in pervasive healthcare. Pervasive and Mobile Computing **86**, 101685 (2022)
7. Binson, V.A., Subramoniam, M., Mathew, L.: Detection of COPD and Lung Cancer with electronic nose using ensemble learning methods. Clinical Chimica Acta **523**, 231–238 (2021)
8. Haider, N.S., Behera, A.K.: Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD). Biocybernetics and Biomedical Engineering **42**(1), 42–59 (2022)
9. Stasiakiewicz, P., et al.: Automatic classification of normal and sick patients with crackles using wavelet packet decomposition and support vector machine. Biomed. Signal Process. Control **67**, 102521 (2021)
10. Aykanat, M., Kılıç, Ö., Kurt, B., Saryal, S.: Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing **1**, 65 (2017)
11. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics **36**, 193–202 (1980)
12. Parashar, A., Parashar, A., Ding, W., Shabaz, M., Rida, I.: Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches. Pattern Recogn. Lett. **172**, 65–73 (2023)
13. Nishikawa, K., Akihiro, K., Hirakawa, R., Kawano, H., Nakatoh, Y.: Machine learning model for discrimination of mild dementia patients using acoustic features. Cognitive Robotics **2**, 21–29 (2022)
14. ICBHI Respiratory Sound Database; Available from: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge