ISSN 1870-4069

# Suspicious Lung Disease Prediction from Auscultation Sounds Using Neural Networks

Beatriz Anshel Sánchez-García, Said Polanco-Martagón, Yahir Hernández-Mier, Marco Aurelio Nuño-Maganda, Jorge Arturo Hernández-Almazán

Universidad Politécnica de Victoria, Mexico

{1430220,spolancom,yhernandezm,mnunom, jhernandeza}@upv.edu.mx

**Abstract.** Early detection of any disease is an important factor in the recovering expectation of any patient. Detection of a disease in an advanced stage can lead to serious or even fatal consequences. To diagnose lung diseases or disorders, physicians use auscultation, which consists of listening body sounds through a stethoscope. This technique requires an outstanding experience of the physician to detect different diseases, and currently artificial intelligence methods are being tested to work as a tool to help in lung disease detection. This article proposes a lung sounds classification method to detect lung issues in suspicious and un- suspicious patients, based on neural networks. This research uses a public dataset, which contains two types of sounds: healthy and lung disease patient sounds. This dataset has a notorious lack of proportion in its data, therefore two balanced techniques were implemented, oversampling and SMOTE, to generate several neural networks models. According to the performed experiments, an accuracy greater than 97% was obtained on the tested dataset.

**Keywords:** Neural networks, lung sound, classification, oversampling, SMOTE.

## 1 Introduction

According to [24], respiratory diseases are one of the leading causes of serious diseases worldwide, exceeding 4 million deaths per year. The World Health Organization, in 2017, reported that 10% of the global mortality rate was accounted for by respiratory diseases [21]. Some of these diseases are: Asthma, COPD (Chronic Obstructive Pulmonary Disease), ARI (Upper Respiratory Tract Infection), Bronchiectasis, LRTI (Lower Respiratory Tract Infection), Pneumonia, among others.

Moreover, in 2019 began what would become a global pandemic caused by the virus known as SARS-CoV-2, which is also called COVID-19 [18] and primarily affects the lungs. Hence the importance of detecting any abnormalities in lung sounds.

Physicians commonly use auscultation with a stethoscope as a method to diagnose lung diseases, which is considered to be invasive for the patient. Commonly, the detected sounds are only used during the duration of the examination and are not stored for subsequent analysis.

To be able to use these sounds for research, a medical protocol to generate a dataset would be required, which will result in lots of time and effort, expensive medical equipment and additional expert help in the classification. When acquiring a sound dataset, the stored sound will commonly present noise in the form of digestive sounds, heart sounds, noise due to the stethoscope scraping the skin, among others, hence a cleaning process using electronic or digital filters will be required. These aspects show that the process to generate a dataset of lung sounds is a complicated and expensive process.

In literature, several public datasets can be found [14, 26], which can be down loaded and used freely. However, these datasets have only educational purposes, which implies that in most cases, the data have been extracted and classified for a single purpose and are therefore not proportionally distributed for each category, hindering later data analysis.

Being able to diagnose by auscultation whether or not a patient is suspicious of any pulmonary disease depends heavily on the experience of the physician. In this work, we propose the implementation of a tool to support the diagnosis of lung diagnosis, taking an unbalanced database and processing it to carry out the classification of sounds through a neural network and thus, in a future work, implement it in an embedded system and offer a less invasive solution for the patient.

## 2 Related Work

In recent decades, lung sounds classification has become a topic of great interest, since it yields relevant information about lung conditions. Several works have been carried out focused on this classification. In [2], a study was conducted between three machine learning approaches to perform classification.

Two of these approaches were based on the extraction of a set of features, which were trained with three classifiers: support vector machines, nearest neighbors and Gaussian mixture models. The third approach is based on convolutional neural networks (CNN). They carried out feature extraction using 12 MFCC (Mel Frequency Cepstral Coefficients) from sounds and the local binary pattern extracted from spectrograms.

They used the dataset of the R.A.L.E. [29] (Respiration Acoustics Laboratory Environment) which contained approximately 50 recordings of respiratory sounds. This dataset required a prior request to the author to grant permission to use the data, it is not a public dataset. In this same work [3], the authors carried out a data augmentation using spectrogram clipping techniques and vocal tract length perturbation. Subsequently, they classified sounds into 7 different classes: normal, coarse crackle, fine crackle, monophonic wheeze, polyphonic wheeze, squawks and stridor.

They obtained 91.12% accuracy using support vector machines with the MFCC and 95.96% accuracy with CNN, after performing 1 million iterations. Although CNNs are one of the most widely used models for classification, it is important to note that its performance depends largely on the learning parameters, on whether the dataset is large and on the number of iterations carried out to train it, which can be time consuming and requires significant computational resources. On the other hand, in [17], authors performed a multichannel classification using a recurrent convolutional neural network. Using their proposed device, they obtained lung sounds from 16 healthy people and from 7 people diagnosed with idiopathic pulmonary fibrosis (IPF), obtaining a total of 23 sounds in total.

Subsequently, they extracted features from the spectrogram to later classify as healthy or pathological (binary classification), using an MLP (Multilayer Perceptron), a BiGRNN (Bidirectional Gated Recurrent Neural Network) and a ConvBiGRNN (Convolutional Bidirectional Gated Recurrent Neural Network).

The latter being the one that obtained the best results with F1 = 92.4%. However, the small number of patients with IPF, the large age difference between healthy patients and patients with IPF, the little variety in terms of lung conditions, among others, would cause the models not to have a correct training process.

On the other side, in [7], the authors carried out a multiclass classification: normal, asthma, heart failure, pneumonia, bronchiectasis and bronchitis and chronic obstructive pulmonary disease. They obtained 308 sound recordings and 1176 from the ICBHI (International Conference on Biomedical and Health Informatics Challange) dataset [14], the same dataset that was used in this study, performed this data fusion to solve the unbalance problem that arose.

They obtained 98% accuracy by making use of the entropy features with Powered Decision Trees. Although his technique works and his results are high, they do not exceed those obtained in this study. In other research [13], a telemedicine framework was built to predict respiratory pathology by lung sound examination. They compared all three approaches with machine learning for lung sound detection. The proposed telemedicine framework trained through Bagging and Boosting classifiers, Improved Random Forest, AdaBoost and Gradient Boosting algorithm. Using a set of handdrawn features they obtained 95% of accuracy.

In another article [30], authors used a dataset composed of five types of lung sounds: normal, coarse crackle, fine crackle, monophonic and polyphonic wheezing. They used higher order statistics (HOS) to extract features, second, third and fourth order accumulators, and genetic algorithms (GA), achieving 96.9% of accuracy. However, these last two required a long training time (between 30 and 200 minutes), which can be considered as a disadvantage.

Based on the results found in the literature, some of the methodologies and tools used in the related works can be useful. Table 1 summarizes the information related to this work that was found in the literature.

The main aspects that were used to make this Table are: the used method, the used datasets, the classification method and their results. It can be seen that the previously proposed models have already reached optimal values for classification. However, the results can be considered unreliable, since they may show an optimal value but perform part of the classification in an erroneous way, since they present an imbalance in their data.

## 3 Objective

The objective of this research is to train a multilayer perceptron to predict whether a patient is suspected or not suspected of lung disease through the classification of lung sounds using a public database by applying data balancing techniques.

**Table 1.** Summary background.

| Ref | Dataset | Balanced dataset | Clasification | Multi-channel/ single-channel | Method | Sounds | Metric |
|---|---|---|---|---|---|---|---|
| [17] | Own | Yes | Binary | Multi-channel | ConvBiGRNN | Healthy and idiopathic pulmonary fibrosis | F1 = 92.4% |
| [17] | Own | Yes | Binary | Multi-channel | BiGRNN | Healthy and idiopathic pulmonary fibrosis | F1 = 86.1% |
| [17] | Own | Yes | Binary | Multi-channel | MLP | Healthy and idiopathic pulmonary fibrosis | F1 = 50.2% |
| [7] | Own and public (The same) | Most of the Classes | Multiclass | Single-channel | Boosted Decision Trees | Healthy, asthma heart failure pneumonia, bronchitis and chronic obstructive pulmonary disease | Accuracy = 98.27% |
| [7] | Own and public (The same) | Most of the Classes | Multiclass | Single-channel | Linear Discriminat | Healthy, asthma heart failure pneumonia, bronchitis and chronic obstructive pulmonary disease | Accuracy = 86.41% |
| [7] | Own and public (The same) | Most of the Classes | Multiclass | Single-channel | Support Vector Machine | Healthy, asthma heart failure pneumonia, bronchitis and chronic obstructive pulmonary disease | Accuracy = 98.20% |
| [7] | Own and public (The same) | Most of the Classes | Multiclass | Single-channel | K-Nearest Neighboors | Healthy, asthma heart failure pneumonia, bronchitis and chronic obstructive pulmonary disease | Accuracy = 97.04% |
| [13] | Own and public | Not mentioned | Binary | Single-channel | SVM | Coarse ralescence, crackles and wheezing | Precision = 81.0% |
| [13] | Own and public | Not mentioned | Binary | Single-channel | KNN | Coarse ralescence, crackles and wheezing | Precision = 94.1% |
| [13] | Own and public | Not mentioned | Binary | Single-channel | Naive Bayes | Coarse ralescence, crackles and wheezing | Precision = 81.0% |
| [30] | Public | Not mentioned | Multiclass | Single-channel | Improved-RF | Healthy, coarseness, coarse crackles, monophonic wheezing monophonic wheezing stridor and squawking | Accuracy = 98.76% |
| [30] | Public | Not mentioned | Multiclass | Single-channel | AdaBoost | Healthy, coarseness, coarse crackles, monophonic wheezing monophonic wheezing stridor and squawking | Accuracy = 96.29% |
| [30] | Public | Not mentioned | Multiclass | Single-channel | Gradient Boosting | Healthy, coarseness, coarse crackles, monophonic wheezing monophonic wheezing stridor and squawking | Accuracy = 94.71% |
| [2] | Public | No | Multiclass | Single-channel | Ensembling CNN | Healthy, coarse crackles, fine crackles, monophonic wheezing, poliphonic wheezing stridor and squawking | Accuracy = 95.56% |
| [2] | Public | No | Multiclass | Single-channel | MFC-SVM | Healthy, coarse crackles, fine crackles, monophonic wheezing, poliphonic wheezing, stridor and squawking | Accuracy = 91.12% |
| [2] | Public | No | Multiclass | Single-channel | LBP-SVM | Healthy, coarse crackles, fine crackles, monophonic wheezing, poliphonic wheezing, stridor and squawking | Accuracy = 71.21 |
| [2] | Public | No | Multiclass | Single-channel | MFCC-CNN | Healthy, coarse crackles, fine crackles, monophonic wheezing, poliphonic wheezing, stridor and squawking | Accuracy = 91.67% |
| [2] | Public | No | Multiclass | Single-channel | LBP-SVM | Healthy, coarse crackles, fine crackles, monophonic wheezing, poliphonic wheezing, stridor and squawking | Accuracy = 80.00% |

**Table 2.** Computer specs.

| | Specification |
|---|---|
| Operating system | Windows 11 |
| System | 64 bits |
| Processor | Intel(R) Core(TM) Graphics 3.70 GHz |
| RAM | 8 GB |
| Hard drive | SSD 494.5 GB |

# 4 Methodology

## 4.1 Dataset

In order to carry out the training of a neural network, it is necessary to have a large enough dataset to ensure optimal learning. In this work, a public dataset [14] was used, which was developed by two research teams in Portugal and Greece. It contains 920 recordings (900 of sick people and 20 of healthy people) which were obtained from 126 people and its duration varies between 20 and 90 seconds.

These sounds were obtained with the following devices: AKG C417L Micro- phone (AKGC417L), 3M Littmann Classic II SE Stethoscope (LittC2SE), 3M Litmmann 3200 Electronic Stethoscope (Litt3200), WelchAllyn Meditron Mas- ter Elite Electronic Stethoscope (Meditron), and the age of the patients who took part of the tests where children, adults and elderly.

The dataset is organized as follows:

– 920 .wav sound files.

– 920 .txt files.

– A file that organizes the diagnostic of each patient.

– A file that describes the elements are contained in the name of each sound.

– A file that organizes the 91 sound names from dataset.

– A file that describes demographic information from each patient.

The sounds were taken from 7 different positions on the chest: trachea, an- terior left, anterior right, rear left, rear right, lateral left and right lateral, and they include clean breath sounds as well as noisy sounds. The diseases found in the dataset are Asthma, COPD (Chronic Obstructive Pulmonary Disease), ARI (Upper Respiratory Tract Infection), Bronchiectasis, LRTI (Lower Respiratory Tract Infection) and Pneumonia, which in this work were grouped in the “suspicious” class.

**Used Hardware:** This section gives details about the hardware that was used during the development and implementation of the proposed model. Table II gives details of the computer equipment that was used.

### 4.2 Feature Extraction

In the tests performed, five features were extracted from each sound, the most commonly used in the literature were chosen. Each sound consists of feature vectors of length 24, including the MFCC (20), i.e. the four chosen features plus the 20 MFCCs. These features are described below:

**Zero Crossing Rate:** As shown in Figure 1, zero crossing rate is presented as the average of the number of times the signal switches between positive and negative within the time window. The speed at which these crossings occur is considered as a measure of the frequency content of a signal. It is commonly used in speech recognition and music information restoration [23, 32].

**Spectral Centroid:** It is a characteristic that details the timbre of a sound approaching the perceptive brightness it possesses, in addition, as shown in Figure 2, it is the one in charge of pointing out where the "center of mass of the sound" is located and it is presented as the result of the weighted average of its frequencies. In the same way, it can be considered as the center of gravity spectrum frequency components have. It is worth mentioning that this does not remain fixed but varies and consequently, the features also change [20, 15].

**Spectral Decrease:** It is defined as the measure of the asymmetry that the spectral shape of a signal possesses, as shown in Figure 3, it represents the frequency that is below a punctual percentage of the total spectral energy. Also, it is known as the *N* percentile of the magnitude distribution of the spectrum and its values are usually 85 or 95 percent [4, 31, 8].

**Mel Frequency Cepstral Coefficients (MFCCs):** These coefficients are a set of between 10 and 20 features that help to show the shape of a spectral envelope. Obtaining these coefficients is one of the most efficient and important techniques of voice parameterization, it also uses the Fourier Transform to carry out the obtaining of the frequencies of a signal. Its main purpose is to obtain a consistent, complete and adequate representation in order to achieve a statistical model of the sound with a greater degree of precision [11]. In Figure 4 we can see an example obtained from a lung sound.

**Chroma Features:** They are defined as a type of musical scales, as shown in Figure 5, the spectrum is drawn in 12 containers of sounds or notes, these have intervals that are always at the same distance and are equivalent to 12 semitones (also called chroma) belonging to the musical octave [27, 22].

### 4.3 Neural Networks

Artificial neural networks are machine learning techniques that simulate the learning of biological organisms. The strengths of synaptic connections commonly occur randomly in response to external stimulation, this is how learning occurs in living organisms. This biological mechanism is simulated in neural networks, which contain calculation units called neurons.

The computational units are connected to each other through weights, which fulfill the same function as the strengths of synaptic connections in biological organisms. In a few words, they are like a type of machine learning algorithm, which is based on the behavior of neurons in the brain. In this work, a MLP will be used to carry out the classification of lung sounds.

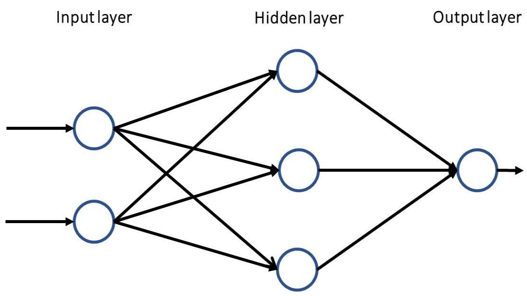


**Fig. 1.** Diagram of the structure of a two input layer MLP.

**Multi-layer Perceptron (MLP):** The MLP [9] is one of the simplest, most common and widely used deep learning models of neural networks. They are also called feedforward since the flow of information through the network is forward, that means, there is no feedback.

The perceptron consists of a simple mathematical function where a set of inputs is entered, some mathematical operations are carried out and a result of said operation is obtained. The equation representing the perceptron is shown in equation 1:

$$y = \Sigma(w_i * x_i + c), \tag{1}$$

where, $w_i$ represents the input weights, $x_i$ the perceptron inputs and c the activation function.

**MLP composition:** MLP architecture is made up from the following elements [28]:

- An input layer, commonly identify as x.
- A number of hidden layers.
- An output layer, commonly identify as y.
- A group of weights and biases for each layer, commonly identify as w y b.
- An activation function for each hidden layer.

Figure 2 shows the architecture of a two-layer MLP (it is important to mention that it is very common for the input layer to be excluded from the total number of layers of the neural network):

Equation 2 shows the computation of the output layer, $y$, of an MLP. This output layer is illustrated in Figure 6:

$$y = \sigma(W_2\sigma(W_1x + b_1) + b_2), \tag{2}$$

where, $y$ represents the output, $W$ the weights and $b$, the biases. We can observe in equation 2, that the only variables that affect the result obtained in the output 'y', are the weights (W) and the biases (b). That is why it is so important that these variables have the correct values, in order to determine the best possible prediction. This process of adjusting the values is known as training a neural network [25].

If we look at the process of each of the interactions that occur in the network layers, it can be found that it is divided into 2 events [16]:

**Table 3.** Activation functions.

| Function | Equation |
|---|---|
| Identity | $f(x) = x$ |
| Sigmoid | $f(x) = 1$ |
| Tanh | $f(x) = tanh(x) = 2 - 1$ $-2x\ 1+e$ |
| ReLU | $f(x) = \{Oforx \leq 0$ |
| | $f(x) = \{Oforx \geq 0$ |

The computation of outputs from the input layer to the output layer, also called Feedforward.

An update of the values of the weights and biases, this process is known as backpropagation.

**Activation Functions:** When designing a neural network model, it is required to define which activation function will be used. This function has an important role, since they are the ones in charge of taking an input value and transforming it to move it to the next layer [10]. The Table 3 shows the activation functions used in this work.

**Data Balancing Techniques:** In this study an unbalanced dataset was used, thus two techniques were used to better distribute the data among the classes.

**Oversampling:** This technique is responsible for balancing the distribution of data by duplicating the elements of the minority class. However, examples containing noise are generated which could lead to problems when training a classification model [3].

**SMOTE:** It is a technique that performs an oversampling of the minority class in order to increase the number of elements and get a more balanced dataset. Unlike OverSampling, SMOTE does not make copies of minority elements, but rather takes features of these elements and their neighbors to subsequently generate new elements with a combination [3].

### 4.4 Results Comparison

For our case of classification, we were used only two possible variables: suspicious or unsuspicious, therefore, we were used a confusion matrix with 2 class labels. Each of the predictions can be one of four outcomes shown in the confusion matrix:

- True Positive (TP): True prediction and actually true.
- True Negative (TN): False prediction and actually false.
- False Positive (FP): True prediction and actually false.
- False Negative (FN): False prediction and actually true.

In addition, four metrics were implemented that were the basis for obtaining the results, which are described below:

**Precision:** It is responsible of measuring the quality of the machine learning model in classification tasks, the lower the dispersion of the data, the greater the precision. It is the

**Table 4.** Neural Network Topologies used in this work.

| Test number | Layers | Neurons |
|---|---|---|
| Test 1 | 2 | 128, 64 respectively |
| Test 2 | 2 | 50, 60 respectively |
| Test 3 | 4 | 60, 70, 80, 90 respectively |
| Test 4 | 4 | 50, 60, 70, 80 respectively |
| Test 5 | 4 | 50, 60, 90, 100 respectively |

result of dividing the true positives by all the positive results (including both the true and fake ones). Briefly, it is the percentage of positive cases obtained by the model [28].

The precision is given by equation 3:

$$Precision = \frac{TP.}{TP+FP}. \tag{3}$$

**Recall:** Also known as the True Positive Rate, it is which informs us of the number of positive cases that were correctly identified by the model [28]. It is calculate as shown in equation 4:

$$Recall = \frac{TP.}{TP+FN}. \tag{4}$$

**F1-Score:** It is the result of the combination of precision and recall in a single value, which makes it practical, since it is much easier to compare the performance of the model [28].

F1 is computed by taking the harmonic mean between precision and recall, as shown in equation 5:

$$f1 = 2 \times \frac{precision \times recall}{precision+recall}. \tag{5}$$

With these metrics, four possible cases can be obtained for each class [1]:

**High precision and low recall:** The model has difficulties detecting the class, but when it does, it is reliable.

**High precision and high recall:** The model has an excellent handling of this class.

**Low precision and high recall:** The model detects the class well, but tends to include samples that do not belong to that class.

**Low precision and low recall:** The model cannot correctly classify the class.

When working with a dataset that presents imbalance, a high precision value is regularly obtained in the majority class and a low value in the minority class. These cases frequently occur in the health area, which is why in this work we used data contrasting techniques.

**Accuracy:** Shows the percentage of cases that the model managed to get right [28].

It is computed through equation 6:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \tag{6}$$

It is the most used measure to evaluate the quality of the models, taking values between 0 and 1. The closer it is to 1, the better.

**Table 5.** Evaluation metrics results (Unbalanced data).

| Test number | Accuracy | Recall | F1 score | Precision |
|---|---|---|---|---|
| Test 1 | 0.95% | 0.98% | 0.97% | 0.97% |
| Test 2 | 0.97% | 1.00% | 0.98% | 0.97% |
| Test 3 | 0.96% | 0.97% | 0.98% | 0.98% |
| Test 4 | 0.95% | 0.98% | 0.97% | 0.96% |
| Test 5 | 0.97% | 0.98% | 0.98% | 0.98% |

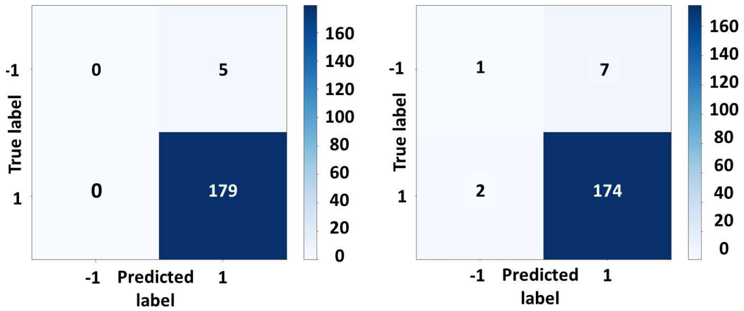


**Fig. 2.** Confusion matrix set one (test 2 vs test 4).

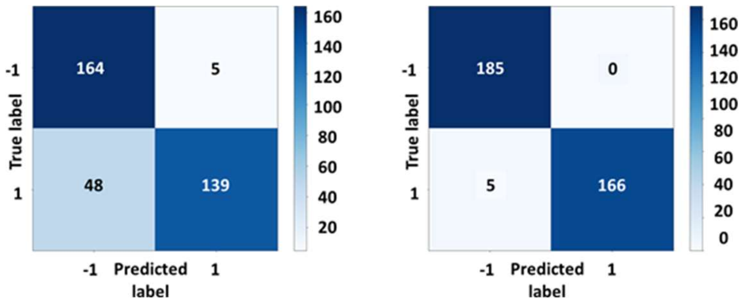


**Fig. 3.** Confusion matrix set two (test 2 vs test 3).

**Table 6.** Evaluation metrics results (OverSampling).

| Test number | Accuracy | Recall | F1 score | Precision |
|---|---|---|---|---|
| Test 1 | 0.96% | 0.93% | 0.96% | 1.00% |
| Test 2 | 0.85% | 0.74% | 0.83% | 0.96% |
| Test 3 | 0.98% | 0.97% | 0.98% | 1.00% |
| Test 4 | 0.97% | 0.95% | 0.97% | 1.00% |
| Test 5 | 0.94% | 0.90% | 0.94% | 1.00% |

## 5 Results

After training the neural network, 3 different sets of tests were developed, with a total of 5 tests each. The model was validated through the use of Cross Validation. As can be seen in Table 4, five different network topologies were used for each test, these were chosen

**Table 7.** Evaluation metrics results (SMOTE).

| Test number | Accuracy | Recall | F1 score | Precision |
|---|---|---|---|---|
| Test 1 | 0.94% | 0.91% | 0.93% | 0.96% |
| Test 2 | 0.89% | 0.78% | 0.87% | 0.99% |
| Test 3 | 0.97% | 0.95% | 0.97% | 1.00% |
| Test 4 | 0.98% | 0.97% | 0.98% | 1.00% |
| Test 5 | 0.94% | 0.88% | 0.93% | 1.00% |

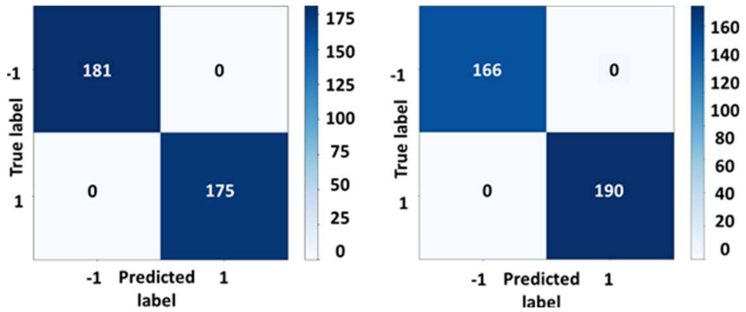


**Fig. 4.** Confusion matrix set one (test 4 vs test 2).

after several tests where it was obtained that the best results were obtained with networks that had 2 to 4 hidden layers and 50 to 100 neurons each.

In the first set of tests, the training and classification proceeded with unbalanced data, as can be seen in Table 5, the results obtained contain percentages that indicates that the classification is being carried out correctly.

However, Figure 2 shows us that despite the metrics indicating that an optimal classification was being generating, this was only being done for the positive class, which means that the model was not learning the negative class, since the dataset was unbalanced, That was the main reason why it was decided to use techniques that would help countervail the data.

To improve the results from the first set of tests, OverSampling technique were used for training and classifying the second set, which was used to balance datasets class.

As in test set 1, the evaluation metrics showed high percentages (a supposed indication of a correct classification). However, it can be seen in Figure 3 (test 3 vs. test 2) that the OverSampling technique was not having the expected results. The reason was that the technique used only duplicated the minority class sounds, therefore the neural network was not optimally trained. And although in test 3, the metrics and the confusion matrix showed good results, the training could not be fully trusted, since it was being carried out with duplicate data. As a solution, in the third set of tests the technique was replaced by SMOTE, which, unlike Oversampling, did not duplicate the sounds, but took one of them and generated a new one applying a minimal change.

The results of applying this technique are shown in Table 7. It can be seen in Figure 4 (test 4 vs test 2) that the classification result into values of 0, which corresponds to the cells of values classified in an erroneous way. This shows that the training of the network had achieved its objective and the model had a greater control over the classes. Based on the evaluation metrics and the confusion matrix, it was found that the test that obtained the best results was test 4 of set 3 (SMOTE technique) with 98%, 97%, 98% and 100% accuracy, exhaustiveness, F-value and precision respectively.

## 6 Conclusion

In this work, a method for lung sound classification based on Neural Networks is presented. It is important to note that if a dataset is not balanced, it will be difficult to generate models to correctly predict an outcome. That is why the contribution of this work is considered to be the use of data balancing techniques, which showed a considerable improvement in the results obtained by the model.

The classes used for this work were divided into 10 versus 90. When the proposed techniques were used, 200% of artificial data were obtained, yielding 50% vs. 50%. The classes that were used for this work were divided into 10% vs 90%. When the proposed techniques were used, 200% of artificial data were obtained. In the performed experiments, the SMOTE technique achieved the best results, getting a 98.97% of accuracy.

As a future work, it is intended to develop a digital stethoscope for the automatic detection of a probable lung disease, that is why this simple classification model is used, since the implementation would be much simpler due to its low processing cost and resources. It should be noted that the proposed model is just a tool to support the detection of pulmonary diseases and does not intend to replace the diagnosis of a health expert.

## References

1. Aggarwal, C. C.: Neural networks and deep learning: A textbook. Springer (2018) 10.1007/978-3-319-94463-0
2. Bardou. D., Zhang, K., Ahmad, S.: Lung sounds classification using convolutional neural networks. Artificial Intelligence in Medicine, vol. 88, pp. 58–69 (2018) doi: 10.1016/j.artmed.2018.04.008
3. Zhang, J., Bloedorn, E., Rosen, L., Venese, D.: Learning rules from highly unbalanced data sets. In: Fourth IEEE International Conference on Data Mining, pp. 571–574 (2004) doi: 10.1109/ICDM.2004.10015
4. Cord M., Cunningham, P.: Machine learning techniques for multimedia: Case Studies on Organization and Retrieval. Springer (2008) doi: 10.1007/978-3-540-75171-7
5. Deng, M., Meng, T., Cao, J., Wang, S., Zhang, J., Fan, H.: Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Networks, vol. 130, pp. 22–32 (2020) doi: 10.1016/j.neunet.2020.06.015
6. Er, M. B.: Heart sounds classification using convolutional neural network with 1D-local binary pattern and 1D-local ternary pattern features. Applied Acoustics, vol. 180, pp. 108152 (2021) doi: 10.1016/j.apacoust.2021.108152
7. Fraiwan, L., Hassanin, O., Fraiwan, M., Khassawneh, B., Ibnian, A. M., Alkhodari, M.: Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. Biocybernetics and Biomedical Engineering, vol. 41, no. 1, pp. 1–14 (2021) doi: 10.1016/j.bbe.2020.11.003
8. Theodoros, G., Pikrakis, A.: Introduction to audio analysis (2014) doi: 10.1016/ C2012-0-03524-7
9. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT Press (2016) www. deeplearningbook.org

10. Gavin, H.: Mastering machine learning with scikit-learn: Apply effective learning algorithms to real-world problems using scikit-learn (2014)
11. Abdul, Z. K., Al-Talabani, A. K.: Mel frequency cepstral coefficient and its applications: A Review. IEEE Access, vol. 10, pp. 122136–122158 (2022) doi: 10.1109/access.2022.3223444
12. Nor-Syazwani, I.: Speech recognition using Mel frequency cepstral coefficient (MFCC). Universiti Tun Hussein Onn Malaysia, 2010
13. Jaber, M. M., Abd, S. K., Shakeel, P. M., Burhanuddin, M. A., Mohammed, M. A., Yussof, S.: A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. Measurement, vol. 162, pp. 107883 (2020) doi: 10.1016/j.measurement.2020.107883
14. Kaggle.: Respiratory Sound Database. English (2019) www.kaggle.com/datasets/vbookshelf/respiratory-sound-database
15. Lokki, T., Müller, M., Serafin, S., Välimäki, V.: Sound and music computing. Multidisciplinary Digital Publishing Institute (2018) doi: 10.3390/books978-3-03842-908-1
16. Loy, J.: Neural network projects with python. Packt Publishing Ltd (2019)
17. Messner, E., Fediuk, M., Swatek, P., Scheidl, S., Smolle-Jüttner, F. M., Olschewski, H., Pernkopf, F.: Multi-channel lung sound classification with convolutional recurrent neural networks. Computers in Biology and Medicine, vol. 122, p. 103831 (2020) doi: 10.1016/j.compbiomed.2020.103831
18. Gobierno de México. COVID-19. Español (2021) coronavirus.gob.mx/covid-19/
19. Millstein, F.: Deep learning: 2 manuscripts-deep learning with Keras and convolutional networks in python (2018)
20. Park, T. H.: Introduction to digital signal processing. World Scientific (2009) doi: 10.1142/6705
21. Perna, D.: Convolutional neural networks learning from respiratory data. In: IEEE International Conference on Bioinformatics and Biomedicine (2018) doi: 10.1109/bibm.2018.8621273
22. Poynton, C.: Digital video and HD: algorithms and interfaces (2007)
23. Lawrence, R. R.: Introducing to digital speech processing. Publishers Inc (2007)
24. Naves, R., Barbosa, B. H. G., Ferreira, D. D.: Classification of lung sounds using higher-order statistics: A divide-and-conquer approach. Computer Methods and Programs in Biomedicine, vol. 129, pp. 12–20 (2016) doi: 10.1016/j.cmpb.20 16.02.013
25. Raschka, S., Liu, Y. H., Mirjalili, V.: Machine learning with PyTorch and Scikit-Learn (2022)
26. E-learning resources: Reference database of respiratory sounds-wheezes (2022) www.ers-education.org/e-learning/reference-database-of-respiratory-sounds/wheezes
27. Taha-Sencar, H., Memon, N.: Digital image forensics: There is More to a Picture than Meets the Eye (2012) doi: 10.1007/978-1-4614-0757-7
28. Siahaan, V., Hasiholan-Sianipar, R.: Step by step tutorials on deep learning using scikit-learn, Keras, and tensorflow with python GUI (2021)
29. The R.A.L.E. repository. English (2008) www.rale.ca/lungsounds.html
30. Ulukaya, S., Serbes, G., Kahya, Y. P.: Resonance based separation and energy based classification of lung sounds using tunable wavelet transform. Computers in Biology and Medicine, vol. 131, pp. 104288 (2021) doi: 10.1016/j.compbiomed. 2021.104288
31. Zelkowitz, M. V.: Advances in computers vol. 44 (2010)
32. Zhang, T., Kuo, C. C. J.: Content-based audio classification and retrieval for audiovisual data parsing. Springer US (2001) doi: 10.1007/978-1-4757-3339-6