Wind Sounds Classification Using Different Audio Feature Extraction Techniques

Wala'a N. Jasim
Department of Pharmacognosy, College of Pharmacy, University of Basra, Iraq
E-mail: Walaa.jasim@uobasrah.edu.iq

Saba Abdual Wahid Saddam and Esra'a J. Harfash
Department of Computer Science, College of Computer Science and Information Technology
University of Basra, Iraq
E-mail: Saba.Saddam@uobasrah.edu.iq, esra.harfash@uobasrah.edu.iq

Keywords: audio signal, audio feature extraction, ZCR, FFT, LPC, PLP, CNN, CNN classification

Received: 10/12/2021

In this research, different audio feature extraction techniques are implemented and classification approaches are presented to classify seven types of wind. We applied features techniques such as Zero Crossing Rate (ZCR), Fast Fourier Transformation (FFT), Linear predictive coding (LPC), and Perceptual Linear Prediction (PLP). We know that some of these methods are good with human voices, but we tried to apply them here to characterize the wind audio content. The CNN classification method is implemented to determine the class of input wind sound signal. Experimental results show that each of these extraction feature methods give different results, but classification accuracy that are obtained by using PLP features return the best results.

Povzetek: V tej raziskavi se izvajajo različne tehnike ekstrakcije zvočnih funkcij in predstavljeni so klasifikacijski pristopi za razvrščanje sedmih vrst vetra. Kjer smo uporabili tehniko funkcij, kot so Zero Crossing Rate (ZCR), Fast Fourier Transformation (FFT), Linear Prediction Coding (LPC), Perceptual Linear Prediction (PLP). Vemo, da nekatere od teh metod dobro vplivajo na človeške glasove, vendar smo jih poskušali uporabiti tukaj za označevanje zvočne vsebine vetra. Za določitev razreda vhodnega zvočnega signala vetra je uporabljena klasifikacijska metoda CNN. Eksperimentalni rezultati kažejo, da je vsaka od teh metod ekstrakcijskih lastnosti dala različne rezultate, vendar se je za klasifikacijo lastnosti PLP izkazalo, da imajo najboljše rezultat.

1 Introduction

Processing of an audio signal generally includes extracting the most important features from it, analyzing, determining the presence of a specific pattern in the signal, and evaluating its behavior pattern, as well as how a particular signal is related to other similar signals. The sound signal has different types such as the speech signal, animal sounds, sounds of specific events in our life, music, and environmental sounds. Therefore, the processing of the audio signal has clearly developed during the past few years, especially with regard to analyzing the audio signals and extracting the most important characteristics from and classifying it [1].

Any signal that represents a sound has a number of parameters such as amplitude, frequency, bandwidth, etc. These qualities can be used in many audio signal processors. Figure 1. shows a representation of any audio signal with its parameters, amplitude and time [2]. Audio processing techniques involve the extraction of the features of a wave signal file, followed by decision-making schemes to detect and classify the inputted sound.

It is critical to order the audio information into different classes like discourse, music, or clamor for quicker and precise access of the information [3]. Then the classification of the audio content is one of the significant and interesting issues. It has 2 main parts, which are: audio feature extraction and [4]. The feature extraction of audio is one important base of present audio signal processing research and evolution. The audio features are an information which can be produced from an audio signal. An information represents contextual information. The features can be divided into groups, that contain definitions of set for the features. In spite of these problems being somewhat different in nature, they heavily lean on groups related to features audio. Low level features are calculated immediately from the audio signal

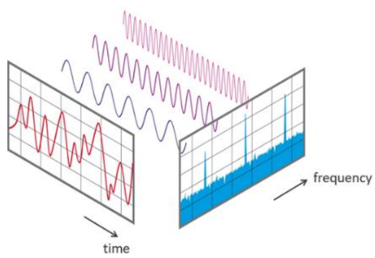

Figure 1: The Time and Frequencies of Sound Signals.

in a frame-by-frame basis oftentimes like zero-crossing rate, and signal energy spectral centroid [5]. The classification of audio is one of the most widespread utilizing cases and includes taking a sound and assigning it to one of various classes. For example, the function could be to identify the kind or sound source. The recently increased attention in deep learning has attracted many scientific and practical applications in different fields of signal processing, oftentimes the processing of traditional signal is outperforming on wide range. In most recent wave, deep learning first produced attraction and interest in image processing, however has then been vastly adopted in environmental sound processing, music and speech processing in addition to a wide range of areas such as chemistry, genomics, quantum, drug discovery, recommendation systems and natural language processing. As a result, previously utilized techniques in processing of audio signal, like Gaussian mixture models(GMM), non-negative matrix factorization and hidden Markov models(HMM) were often bested via DL models, in applications where enough data is obtainable [6]. Many scientific problems and fields have witnessed great developments through the use of deep learning, which has led to its improvement and increase in its achievement rate, for example, computer vision, natural language processing, and also in the field of sound area, like music recommendation and speech recognition [7]-[9].The sound classification systems based on deep neural networks such as CNNs have undergone important improvements in the recognition and classification capability of models. None the less, their complexity of computational and inappropriate exploration of universal dependencies for long sequences restrict perfections in their results of classification [10].

Many researches in the recently years have been achieved in the automatic sound classification and detection area in outdoor environments. Some researchers focused their studying about the environmental sounds such as natural and human produced whilst others have focused and specified the detection and classification of various species of animals [11,15].

However, the objective of this paper is to introduce a wind sound detection and classification system, that is focused on the classification of some classes of wind sounds. According to the information features contained in the signal about the frequency and time space, these features are used to investigate and classify the wind audio signal. The classification of seven classes of wind sounds are applied, these classes are (Soft, Howling, Ghost, Blizzard, Cold, Desert, Strong, Scary) wind. Several extraction features techniques are implemented to extract the most important features in time or frequency of the wind audio signal. These techniques are: ZCR, FFT, LPC, PLP. CNN model is used here to classify the wind sounds.

The rest of the paper is organized as follows; after showing the introduction in section 1, the related work is given in section 2. section 3 explains some feature technique of sounds. Sections 4 explains audio Deep Learning Models, and section 5 shows the database and main steps with all techniques used to complete the system work, and the results of accuracy performance. Finally, the conclusion is given in the last section.

2 Literature review

Nowadays, the classification of sound is a wide field of studying that has great attracted interest from many researchers. With an improvement of Deep CNN and its effective utilize in computer vision(CV), language modeling, recognition of speech, and other regarding fields, it is confirmed that architecture of CNN based out-classes the classical ways in different classification missions. Which is why, they were stratified in the automatic sound event recognition task in recently years. As is the case in this paper presented in which we used CNN to classify wind sounds and predict what will happen after the wind.

Pablo Zinemanas et al. [7] They proposed a new explicable DL model for automatic sound classification, that interpret its foretelling base on likeness of the inputting to a group of learned proto-types in a latent space. Their proposed consist of two main components: an auto-encoder and a classifier. The model of inputting is a representation of time frequency for the audio signal. The aim of the auto-encoder was for representing the inputting into a latent space of beneficial, features which were learned through training step. Then the encoded inputting was utilized via the classifier in order to make a foretelling. Their proposed model realizes results which were similar to that of state-of-art approaches in 3 various tasks of sound classification including music, environmental audio and speech. Two automatic techniques are presented in order to prune their proposed model. Their model was opened source and it was chaperoned via a application web for the editing manual model, that let for a human-in-the-loop debugging method.

Loris Nanni et al. [16] presented work for combining different clustering techniques with a Siamese NN and in order to produce a variation space that is then utilized to train the SVM for classification of animal audio. They used free datasets of animal audio which consist of sounds of birds and cats. They used an SVM for classifying a spectrogram via its variation vector. Their research proposed technique showed based on variation space implement good on both classification tasks with no ad-hoc optimization of clustering approaches. Their results showed that the stand-alone CNNs is worked not better than the combination of CNN-based methods which applied on animal audio classification.

Silvia Liberata Ullo et al. [17] are presented a hybrid model for accurate and automatic of environmental sounds classification. They used Optimal allocation sampling (OAS) in order to extract the samples of informative from any class. The samples that have been acquired via OAS are turned into the spectrogram containing the representation of Time Frequency Amplitude via utilizing a Short-Time Fourier Transform (STFT). They used pre-training networks and classified it by applied multi-classification methods such as Decision Tree (DT) {fine, medium, coarse kernel}, K-Nearest Neighbor(K-NN) {fine, cosine, medium, cubic, coarse and weighted kernel}, SVM, Linear Discriminant Analysis (LDA), Bagged Tree and Softmax classifiers to extract multiple deep features. They used a ESC-10 dataset for the evaluation of the methodology. Their proposed method is proved robust, promising and effective comparing with other techniques that using the same dataset.

Md. Rayhan Ahmed et al [18] are presented system by using the Convolutional Neural Network (CNN) for processing turn into a short sound event audio file to an image of spectrogram and feed that image to (CNN) for processing. The features that are produced from the image are utilized for classification of different environmental sounds events like fire cracking, sea waves, dog barking, raining, lightning, etc. They have utilized the log-mel spectrogram auditory feature to train six-layer stack of CNN model. They are predestined the accuracy of their model to classify the environmental sounds in three datasets and they are carried out an accuracy for the urbansound8k, the ESC-10 and the ESC-50 datasets 92.9%, 91.7% and 65.8% consecutively. Their studying is showed a comparative between Adam Optimizer and RAdam optimizer utilized for training the model to correctly classify the environmental sound from architecture of image recognition.

Diez Gasponet et al [19] presented an automatic system for detecting and classifying sounds, particularly those generated via insects and birds among other sounds that can be heard in an environment of natural. They compared the performance of three various features: mel frequency cepstral coefficients (MFCC), log mel filtered spectrogram (Mel Spectrogram) and log spectrogram (STFT). They generated a sound dataset in order to the development their system. The recording dataset is contained three various Natural Parks, with sounds of many insect species and birds and background noises. Their proposed system is used the neural networks NN to detect and classify sound frames. Their experiments are offered good accuracy in detection and classification of sound frames and with high results compare with other approaches.

Yu Su et al [20] are proposed two combination features to allow a more universal environment representation sounds and CNN is presented with a four-layer to get better the implementation of ESC with suggested grossed features. These features were (Log-mel spectrogram, chroma, spectral contrast and tonnetz). In their proposed system Log-mel spectrogram, chroma, spectral contrast and tonnetz are aggregated to compose the feature sets of LMC, and MFCC is jointed with spectral contrast, chroma and tonnetz to compose the MC feature sets. Then, the CNN trained with various features are fused utilizing the Dempster–Shafer evidence theory to form TSCNN-DS model. The results of their system refer that the features of combination with the four-layer CNN were suitable of the problems for environment sound taxonomic and considerably outperformed other classic techniques. The TSCNN-DS model is achieved an accuracy of classification of 97.2%.

Aditya Kamparia et al [21] are proposed system to classify the sounds of environmental based upon the produced spectrograms of these sounds by using deep learning networks. They used CNN in both stages: feature extraction stage and classification stage. They utilized the spectrogram images of environmental sounds for training the tensor deep stacking network (TDSN) and the convolutional neural network (CNN). They applied two datasets for their experimental work: ESC-50 and ESC-10. Two systems have been trained on their used datasets, and the carried out the accuracy was 49% and 77% in the CNN and 56% in the TDSN trained on the ESC10. From their experimental work, they concluded that their proposed system for classification of sound using the spectrogram sounds images can be effectively used to evolve the sound recognition and classification systems.

Marielle Malfante et al [22] are presented addresses the environmental monitoring issue. specially, their proposed is focused on the using of systems acoustic for monitoring of passiving acoustic of ocean vitality for fish populations. In their study, they used 84 features in feature extraction stage and used a forward selection approach: features are ranked by the importance according to their weight in the RF model in Features selection stage. They built a discriminative model by using Support Vector Machines (SVM) and random forest (RF) which are most important supervise machine learning techniques. Their features proposed to describe the acquisitions came from an inclusive state of the art in different domains that acoustic signals classification is performed, included of music, environmental sounds and speech. In addition, their studying proposed for extracting features from three representations of the data (frequency, time, and cepstral domains). On real fish sounds recorded on different areas their proposed classification scheme is tested and obtained 96.9% correct classification.

Sunit Sivasankaran et al [23] are presented algorithms to classify sounds of environmental in order to provide information of contextual to devices like hearing aids for the best performance. They utilized signal sub-band energy for constructing signal-dependent dictionary and matching pursuit algorithms to get a scattered representation of a signal. They were applied and used the coefficients of sparse vector as weight values for computing the weighted features. These features in the previous step with MFCC have been utilized as feature vectors for classification. The results of their Experimental are showed that their proposed method achieved accuracy 95.6 % whilst classified 14 classes of sound of environmental by utilizing the (GMM).

Siddarth Sigtia et al [24] are presented Automatic Environmental Sound Recognition (AESR) algorithms and developed them with fixed consideration for counting cost. By their experiment, Mel-frequency cepstral coefficient (MFCC) features were extracted from the audio. MFCC features are wide used in environmental sound recognition and in speech recognition. They proved that AESR algorithm could made the most of a limited amount of computing power by compare the performance of sound classification as its computational cost function. Their results offered that DNN produced the best accuracy for classification sound across a computational costs range, whilst GMM yielded a sensible good accuracy with small cost, and SVM stand between both in terms of adjustment between computational cost and accuracy.

3 Features extraction techniques

The Features mean something that values can be quantified and measured numerically using specific techniques available for it. For example, sample rate and sample data are two things that a sound wave is made of primarily. Now several transformations can be performed on the sample rate and sample data to extract important valuable features from it [25-27].

The accuracy of the system relies on the features and classification methods. Extracting efficient features is an important phase in the front-end module of building an sound classification system. For each class of sound there are some features that distinguish it from the rest of the other types of sounds, ut that the sound signal of one class may change with time, and this change may occur on any of the sound variables, such as amplitude or frequency. In the following paragraphs, it explains some of the techniques that are used to extract features from sound file. Some are specialized in extracting features from the time space and others from the frequency space.

3.1 The zero crossing rat

The ZCR is the rate of change of a sign signal (from positive to negative or inverse) along the signal. Speech recognition and music processes topics often use this feature in many of their processing. Its value is high with percussion sounds such as those found in minerals and rocks [28], The ZCR is defined according to the following equation [29]:

$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} \left| sgn[x_i(n)] - sgn[x_i(n-1)] \right|, \quad (1)$$

Where: sgn(·) is the sign function, i.e

$$sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0, \\ -1, & x_i(n) < 0. \end{cases} \quad (2)$$

3.2 Discrete fourier transform

The Spectrum features are important in digital audio processing. A spectrum can be represented Mathematically using Fourier transform of a signal, where the time domain of signal is converted to the frequency domain. This means, a spectrum is the frequency domain representation of the input audio's time-domain signal [30]. Mathematically, the Discrete Fourier Transform (DFT) transforms a limited sequence of samples of equally spaced of a function into a sequence of same-length of equally spaced samples of the discrete-time Fourier transform (DTFT), that is a complex valued function of frequency. The DFT transforms a sequence of *N* complex numbers into another sequence of complex numbers, which is defined by [31].

$$\hat{x}(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi i k n/N} \quad \text{for } k = 0, 1, \ldots, N-1. \quad (3)$$

As DFT deals with a limited data amount, it may be conducted in computer devices via the numeral algorithms or even devoted hardware. Those performances often employ effective Fast Fourier Transform (FFT) algorithms;[3] both "FFT" and "DFT" terms are typically utilized in an interchangeable manned. Prior to its current usage, the "FFT" initialism may have also been utilized for the ambiguous term "Finite Fourier Transform"[32].

3.3 Linear Predictive Coding (LPC)

In audio signal processing and speech processing, the LPC is a method used mostly for representing the spectral envelope of a digital signal of that represent speech in compressed form, using the information of a linear predictive model [33] :

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \quad (4)$$

For the periodic signals with a period N_P, it's evident that $(S(n) \approx S(n-N_p))$. However, that isn't what LP is doing; it estimates (S(n)) from $(P(P<<N_P))$ most recent (S(n)) values through the linear prediction of its value. for LP, the coefficients of the predictor (α_k's) are determined (i.e. estimated) through the minimization of the summation of the squared differences between actual sound samples and linearly predicted sound samples.

3.4 Perceptual Linear Prediction (PLP)

PLP are used to extract features which are used to characterize the audio data. PLP can be defined as approximation of 3 aspects that are associated with the perceptron, which are: the resolution curves of critical band, curve for equal loudness and power law relation of the intensity loudness, the process for the determination of PLP coefficients are describe in the figure [34]:

The LPC and PLP are use a lot in techniques of features extraction in the fields of speech recognition and speaker verification [32], but in our research we used to distinguish the wind signal.

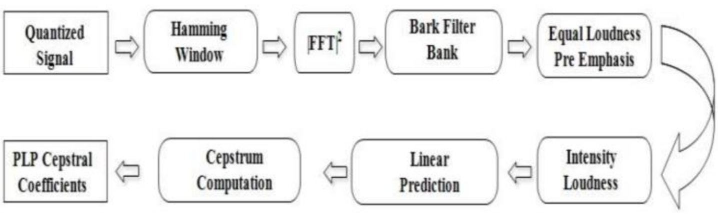

Figure 2: PLP Parameter Computations.

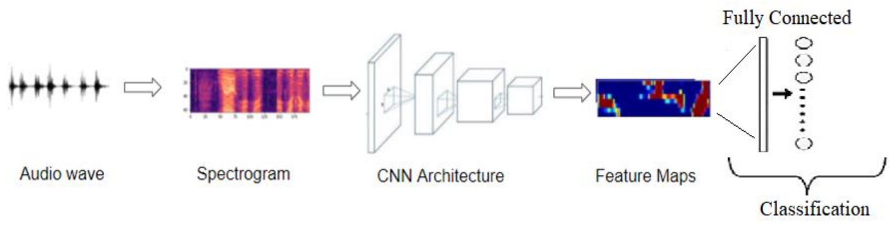

Figure 3: Schematic Diagram of Basic Audio Convolution Neural Network.

4 Audio deep learning models

Now that we understand what a Spectrogram means, we perceived which it is a duplicate compact representation of an audio signal. It is an wonderful way to hold the major features of audio data [2], then we need a deep learning to classify these feature.

Convolutional neural networks ('CNNs') are capable of achieving state-of-the-art results in both image and audio classification. CNNs are often exercised for missions that the input data exhibits the features of locality (where features in data have a local-spatial-support) and translation invariance (where features in data are location independent). he basic structure of a CNN is similar in organization and function to the visual cortex, and it is also designed to mimic the way neurons communicate within the human brain. The architecture of a CNN consists of a stack of discrete layers that convert the input volume to the output volume via a differentiable function. A few distinct types of commonly used layers are [CNN2], as in Figure 3:

1 Convolution layers: to preserve the spatial orientation of the features.

2 Grouping Layers: They are used to reduce (reduce) the input pattern.

3 Fully Connected Layers: The output from the CNN convolutional and aggregate layers is the pattern feature vector. The objective of the fully connected layer is to utilized the vector of these features to classify the input patterns into several categories based on a labeled training data set.

5 The methodology and results

In this part the methods of Working System are introduced. Typically, there are some main stages for building a system for general any sound signal classification, i. e., feature extraction, model training model and classification**.** The Figure 4 gives an briefly an overview of the our system.

The sound signal entered here is the wind sound signal, from this signal the distinctive features are extracted from it in the stage of features extracting, and then these features are submitted to the classification model, The CNN was relied upon as a model for classification, due to the strength of this technique in Separate the different classes, where the final decision.

5.1 The database

In this research, we try to classify among the different sounds that come from the wind and that change according to the change in the environmental condition that the wind sounds predicted.

We have collected data from different websites over the Internet, to suit our needs for specific specifications of the wind sound signal. So we gathered these voices from different locations, and from these sites we got some

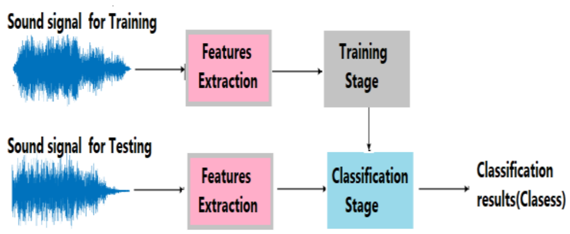

Figure 4: The Structure of Audio Classification System.

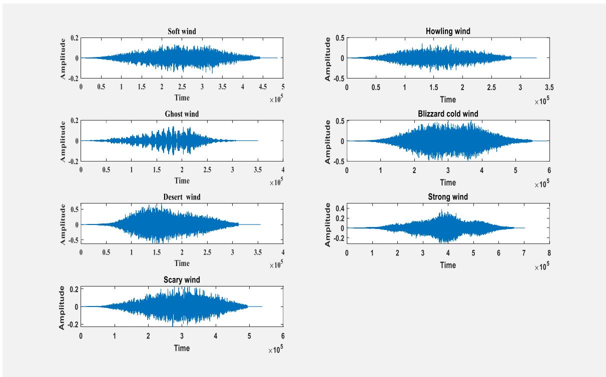

Figure 5: Some Samples Wind wav sounds.

sound signals are:

- mixkit.co/free-sound-effects/wind/
- www.zapsplat.com/sound-effect-category/wind/
- www.soundjay.com/wind-sound-effect.html
- And because we are specialists in processing the audio signal, we already have a set of this data

There are seven different wind sound classes (soft, howling, ghost, blizzard, desert, strong, scary) winds that are adopted in this system, contains 350 sound signals, 210 sounds for training and 140 for testing and we are equal the different length of sounds by Appling solaf function. All these sounds were carrying the same following characteristics:

- The sounds is wav files

- The sampling rate here is 44100
- and all sound files of 16 bits of length and mono.
- Each sound wind signal is treated independently, subdividing it into continuous frames.

The Figure 5 is shown Some samples wind sound that dependent in this research.

5.2 Results and discussion

In this work, before starting to extract the features, the signal are decomposed into blocks of length 256 samples, in order to control the changes that occur to the signal over time. Then, the features of wav files is extracted, that will help to classify wind files into different classes. We will discuss the accuracy results of application different techniques of audio features extraction that are described in Section 3 and recommend the preferred and good feature extraction to classification using CNN model.

5.2.1 ZCR features

Zero-crossing rate is used in this paper, we used the zero-crossing rate of audio Activity to determine the most important features of the different wind signals through this method. Where for the input wind signal X then the Zero-crossing rate is calculated: Y = Zero-crossing (X), where Y Where is the ratio of the intersection of the signal X with the zero axis. Figure 6 show the Plot the signal and the resulted Zero-crossing .

Figure 7 presents a wind signal with the respective ZCR sequence. It shows the differences of values of ZCR with different parts of the signal. Then Depending upon the resulted information of ZCR, we are tried to determine the class of wind based on this type of feature information. The accuracy rate obtained here is (74.43%), and Figure (4) shows a CNN scheme training that gave this final result.

5.2.2 FFT features

The FFT algorithm is used to compute the DFT of input sound vector x. It computes the N-point DFT. The FFT is applied on the input audio signal to extract the most important features of the frequencies that present in the signal. Then the DFT Coefficient is more important to look at the overall frequency content of signal. Figure 8 show the results of DFT coefficients of double side, in (a)without shifting and in (b) applying shifting, by shifting the zero-frequency component to the center. In any case, if the results of the transformation with length N then then half of the data will be taken from 0 to N-1/2.

The DFT always applied with sound processing, because the data of sound data is represent discrete samples, and is good describe the frequencies present in a signal. Then here, we applied the FFT to obtain the most important frequencies in the incoming audio signal and study their efficiency in classification the wind signal.

Figure 9 show the overall results of classification using DFT coefficients as input features to CNN model, and we see that the accuracy result without FFT shifting is 74.43, compare with FFT shifting where the accuracy is 77.19 and is somehow litter better.

5.2.3 LPC features

The LPC is used to calculate spectrum of the signal. In this work, we are presented a Linear Predictive Cepstral Coefficients (LPCC), to extract the important and effective features of sound wind signal to be classification. We want here to know if it is useful in classifying this type of signal. Since we have calculated different coefficients of LPC starting from 8 to 14 coefficients, and the results of accuracy classification are similar here with all coefficients. However, the results may be slightly better with coefficients 13 and above. Where was the CNN is gave accuracy 74.29%. Figures 10 shows the gradient in accuracy according to the number of coefficients and

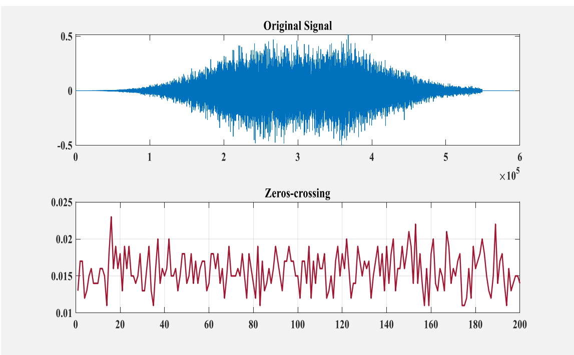

Figure 6: Show the Signal and the Resulted Zero-Crossing rate.

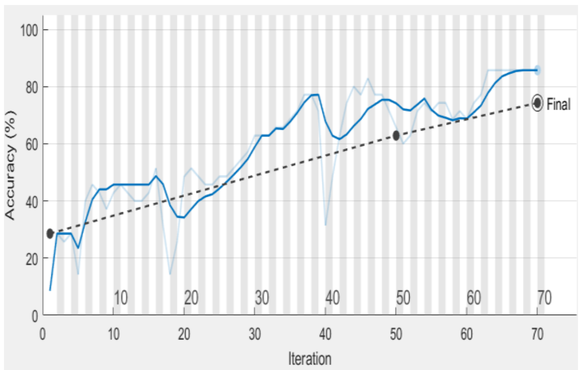

Figure 7: The CNN Training with Final Result (74.43%).

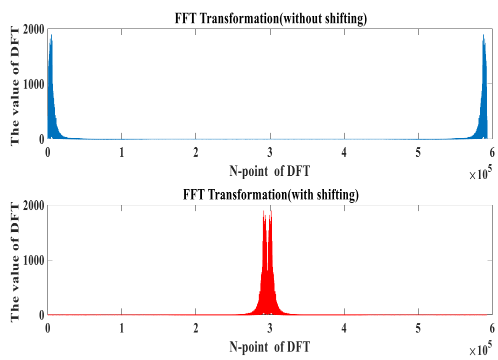

Figure 8: N-point DFT Calculated Using FFT.

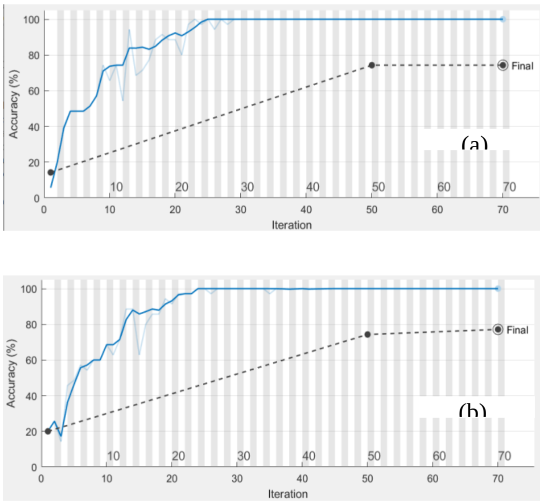

Figure 9: Accuracy of Using DFT Coefficients by Applying FFT.

figure 11 shows the highest accuracy obtained with LPC features. But in general, the results in this LPC feature extraction techniques are moderate.

5.2.4 PLP features

In the present study also the PLP features are obtained from input signal, where the wind wav file is inputted to the feature extraction approaches PLP. PLP with different dimensional feature values will be computed for the given wav file. Figures 12 shows the gradient in accuracy according to the number of coefficients that calculated by PLP method. The best accuracy rat with13 PLP coefficients with CNN model, as shown in figure 13.

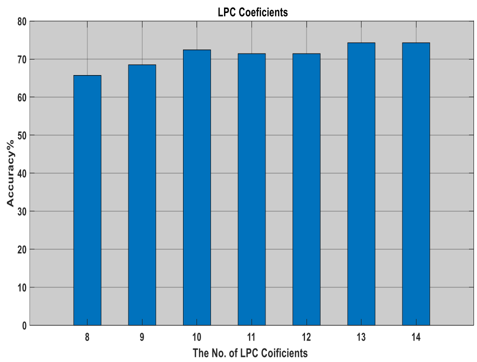

Figure 10: The Accuracy with Different LPC Coefficients.

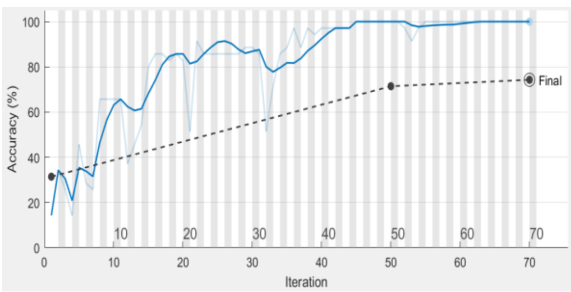

Figure 11: The Highest Accuracy (74.29%) with LPC Features.

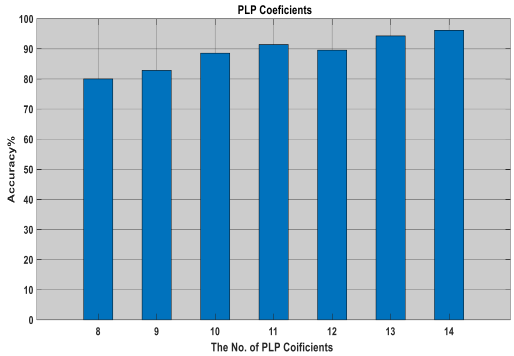

Figure 12: The Accuracy with Different PLP Coefficients.

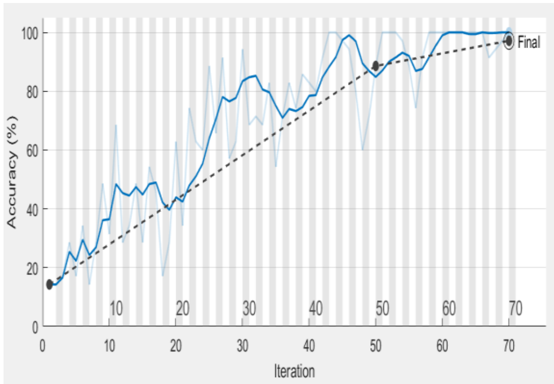

Figure 13: The Highest Accuracy (96.14%) with PLP 14 Coefficients Features.

Table 1 show all the General results obtained from the implementation CNN with each types of features:

Compared to other researches, we are interested in identifying the best features that fit the signal we are treating, in addition to that we are trying to find reasons to raise the accuracy of any method for extracting features with modest results.

Now we are say that all the methods as in table 1 that were followed here were successful in giving distinguished results for the classification of wind sounds, and the following is a set of notes about our observations on this work:

- Performance of ZCR, FFT and LPC is decreased with the increase of the number of classes.
- The size of the database had an effect, on the accuracy of the results, if the database was larger, the results

Feature Audio Technique	Error Accuracy with CNN
ZCR	74.43%
FFT WITH SHIFTING	77.19%
FFT WITHOUT SHIFTING	74.43%
LPC	74.29%
PLP	96.14%

Table 1: The Accuracy precision of each audio features techniques.

with all the four-feature technique would have been better.

- The PLP method is the best among of all, where the accuracy is 96.14%, and represent important result,
- So it can be implemented PLP successfully with this type of sound signals, and with other audio signals.
- ZCR, FFT and LPC methods can be adopted to classify the audio signals unlike human voices, but with the wind signal, you need less number of classes in order to obtain a greater classification performance.

6 Conclusions

Research on voice signals and its treatment is considered to be very important, as it has a great role in solving many problems that can be solved by analyzing and treating those signals. The treatment of the sound of the wind is considered equally important, especially the classification of different wind sound signals. In this research we designed a system that reads the wind sounds and classifies them to one of seven wind classes.

A CNN-Model of audio classification is used, because it is considered as a good performer in classification problems, so we adopted it in this work to measure its efficiency with this type of data. In this research the CNN model is implemented with four different methods of audio features extracting. We tried to deduce which of these types of features extraction techniques is the best for distinguishing wind audio signals, as opposed to distinguishing only than human signals. In this system the ZCR, FFT, LPC and PLP are adopted as audio features for this classification problem. We note that the performance of ZCR, FFT and LPC is decreased with the increase of the number of classes. The accuracy of CNN with ZCR is 74.43% and with FFT shifting, it is 77.19%, while the precision with LPC features is 74.29%. But the PLP method is the best among them all, where the accuracy is 96.14%. PPL can therefore be implemented successfully with this type of sound signals, and with other audio signals as well.

We would like to add that the methods of extracting characteristics for audio signals cannot be limited to specific signals, but yes, they can be very successful with certain signals. Certain changes and many tests can be made to make a specific method fit a specific signal. In the future, we hope to build a strong system capable of processing wide types of wind signals and forecasting the type of wind in order to determine the future state of the atmosphere, by adopting classification models known for their strength such as CNN architectures

References

[1] Garima S, Kartikeyan U and Sridhar K. (2020). Trends in Audio Signal Feature Extraction Methods. Elsevie, vol. 158, pp. 1-21. https://doi.org/10.1016/j.apacoust.2019.107020

[2] Purwins H, Li B and Virtanen T., (2019). Deep Learning for Audio Signal Processing. IEEE Journal of Selected Topics in Signal Processing,vol. 14, pp. 1-14. https://doi.org/10.1109/jstsp.2019.2908700

[3] Andersson T. (2004). Audio Classification and Content Desicription. M Sc. Thesis, Department of Computer Science Electrical Engeerring, University of Techonology.

[4] Liang S and Fan X. (2014). Audio Content Classification Method Research based on Two-Step Strategy. International Journal of Advanced Computer Science and Applications,vol. 5, pp. 57-62. https://doi.org/10.14569/ijacsa.2014.050307

[5] Moffat D, Ronan D and Reis J. (2015). An Evaluation of Audio Feature Extraction Toolboxes. in Proc. of the 18th International Conference on Digital Audio Effects, pp. 1-7.

[6] Krizhevsky A, Sutskever I and Hinton G. (2012) . Imagenet Classification with Deep Convolutional Neural Networks. in the Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 84-90.

[7] Zinemanas P, Rocamora M and Miron M.. (2021). An Interpretable Deep Learning Model for Automatic Sound Classification. Electronics, vol. 10, pp. 1-23. https://doi.org/10.3390/electronics10070850

[8] Ian Goodfellow (2016). Deep Learning. MIT Press: Cambridge, 800 pp, ISBN: 0262035618. https://doi.org/10.16997/ahip.1015

[9] Xie J, Hu K, Zhu M, Yu J and Zhu Q. (2019). Investigation of Different CNN-Dased Models for Improved Bird Sound Classification. IEEE Access, vol. 7, pp. 175353-175361. https://doi.org/10.1109/access.2019.2957572

[10] Yang L, Zhao H. (2021). Sound Classification Based on Multihead Attention and Support," Mathematical Problems in Engineering,., pp. 1-11. https://doi.org/10.1155/2021/9937383

[11] Su Y, Zhang K, Wang J and Madani K. (2019). Environment Sound Classification Using a Two-Stream CNN Based on Decision-Level Fusion. in Sensors, vol. 19, issue 7,pp. 1-15. https://doi.org/10.3390/s19071733

[12] Valero X and Alías F. (2012), Analysis of the Acoustic Signal By Bio-Inspired Cepstral Coefficients and their Application to the Recognition of Soundscapes. VIII Congr. Ibero-americano Acústica, pp. 1-9.

[13] Chang C. and Doran B. (2016), Urban Sound Classification: With Random Forest SVM DNN RNN and CNN Classifiers, in CSCI E-81 Machine Learning and Data Mining Final Project Fall 2016, Harvard University NCambridge.

[14] Acevedo M. (2009), Automated Classification of Bird and Amphibian Calls Using Machine Learning: A Comparison of Methods, Ecological Informatics,vol. 4, pp. 206-214.

[15] Adavanne S, Drossos K, Cakir E and Virtanen T. (2017), Stacked Convolutional and Recurrent Neural Networks for Bird Audio Detection, in 25th European signal processing conference (EUSIPCO), IEEE, pp. 1729-1733.

https://doi.org/10.23919/eusipco.2017.8081505

[16] Nanni L, Rigo A, Lumini A and Brahnam S. (2020), Spectrogram Classification Using Dissimilarity Space, Applied Sciences, vol. 10, pp. 1-17.

[17] Ullo S, Khare S, Bajaj V and Sinha G (2020)., Hybrid Computerized Method for Environmental Sound Classification, IEEE Access, vol. 8, pp. 124055-124065. https://doi.org/10.1109/access.2020.3006082

[18] Ahmed M. (2020), Automatic Environmental Sound Recognition (AESR) Using Convolutional Neural Network, International Journal of Modern Education & Computer Science,vol. 12, pp. 41-54. https://doi.org/10.5815/ijmecs.2020.05.04

[19] Gaspon D. (2019), Deep Learning For Natural Sound Classification, in Inter-Noise and Noise-Con Congress and Conference Proceedings, Institute of Noise Control Engineering, pp. 5683-5692.

[20] Khamparia A. (2019), Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network, IEEE Access, vol. 7, pp. 7717-7727. https://doi.org/10.1109/access.2018.2888882

[21] Malfante M. (2018), Automatic Fish Sounds Classification, The Journal of the Acoustical Society of America, vol. 143, pp. 2834-2846. https://doi.org/10.1121/1.5036628

[22] Sivasankaran S and Prabhu K. (2013), Robust Features for Environmental Sound Classification, in 2013 IEEE International Conference on Electronics, Computing and Communication Technologies, IEEE, pp. 1-6.

[23] S. Sigtia, et al Sigtia S, Stark A, Krstulovic S and Plumbley M. (2016), Automatic Environmental Sound Recognition: Performance Versus Computational Cost, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, pp. 2096-2107. https://doi.org/10.1109/taslp.2016.2592698

[24] Paraskevas I. (2005), Phase as a Feature Extraction Tool for Audio Classification and Signal Localisation, University of Surrey (United Kingdom).

[25] Zhang J. (2021), Music Feature Extraction and Classification Algorithm Based on Deep Learning, Hindawi, Scientific Programming, pp. 1-9. https://doi.org/10.1155/2021/1651560

[26] Wala'a, N. J. & Esra J. H., (2018), Human Activity Recognition System to Benefit Healthcare Field by Using Hog and Harris Techniques with K-NN Model, International Journal of Computer Applications, vol. 180, pp. 7-9. https://www.ijcaonline.org/archives/volume180/number40/29393-2018917045

[27] Wala'a, N. J. & Esra J. H., (2019), Recognition Normal and Abnormal Human Activities by Implementation k-Nearest Neighbor and Decision Tree Models, Journal of Theoretical Applied information Technology, vol. 96, pp 6423-6443. http://www.jatit.org/volumes/Vol96No19/13Vol96No19.pdf

[28] Giannakopoulos T and Pikrakis A. (2014), The Matlab Audio Analysis Library, Academic Press, Book Chapter in Introduction to Audio Analysis, pp. 233-240. https://doi.org/10.1016/b978-0-08-099388-1.00017-0

[29] Müller M. (2015), Musically Informed Audio Decomposition, Springer, Book Chapter in Fundamentals of Music Processing on pp. 415 -480. https://doi.org/10.1007/978-3-319-21945-5_8

[30] Herman R. (2016), An Introduction to Fourier Analysis, Chapman and Hall/CRC, eBook ISBN 9781315367064. https://doi.org/10.1201/9781315367064

[31] Terenzi A, Cecchi S, Sorcioni S and Piazza F. (2019), Features Extraction Applied to the Analysis of the Sounds Emitted by Honey Bees in a Beehive, in 2019 11 International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, pp. 03-08. https://doi.org/10.1109/ispa.2019.8868934

[32] Grama L and Rusu C. (2017), Audio Signal Classification Using Linear Predictive Coding and Random Forests, in 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), IEEE, pp. 1-9. https://doi.org/10.1109/sped.2017.7990431

[33] Dave N (2013), Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition, International journal for Advance Research in Engineering and Technology,vol. 1, pp. 1-4.

[34] Hershey S, Chaudhuri S., et al. (2017), CNN Architectures for Large-Scale Audio Classification, in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/icassp.2017.7952132