IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON TRENDS, PERSPECTIVES AND PROSPECTS OF MACHINE LEARNING APPLIED TO BIOMEDICAL SYSTEMS IN INTERNET OF MEDICAL THINGS

Received October 11, 2018, accepted October 30, 2018, date of publication November 15, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2881199

Binary Neural Networks for Classification of Voice Commands From Throat Microphone

FÁBIO CISNE RIBEIRO[1], RAPHAEL TORRES SANTOS CARVALHO[2], PAULO CÉSAR CORTEZ[1], VICTOR HUGO C. DE ALBUQUERQUE[3], (Member, IEEE), AND PEDRO PEDROSA REBOUÇAS FILHO[4]

[1]Department of Teleinformatics Engineering, Federal University of Ceará, Pici Campus, Fortaleza 60440-970, Brazil
[2]Federal Institute of Education, Science and Technology of Ceará, Jaguaribe 63475-000, Brazil
[3]Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza 60811-905, Brazil
[4]Federal Institute of Education, Science and Technology of Ceará, Maracanaú 61939-140, Brazil

Corresponding author: Fábio Cisne Ribeiro (fabiocr@gmail.com)

ABSTRACT Multi-class pattern classification has many applications including speech recognition, and it is not easy to extend from two-class neural networks (NNs). This paper presents a study about using binary classifiers with NNs together with a perceptual linear prediction (PLP) method for feature extraction to increase the classification rate of voice commands captured using a throat microphone, comparing this method with a single NN. Because there is no other data set with voice commands captured using a throat microphone in the Brazilian Portuguese language in researched literature, we created a data set with isolated voice commands with utterances captured from 150 people (men and women). All the voice samples are captured in Brazilian Portuguese, and they are the digits "0" through "9" and the words "Ok" and "Cancel". The results show that the throat microphone is robust in noise environment, achieving 95.4% of hit rate in our speech recognition system with multiple NNs using the one-against-all approach, better performance than a simple NN that reach 91.88%. This result is very representative, since both classifiers obtained high hit rates. But, it requires 535% more time for training the multiple NNs compared with simple NN. The best configuration on PLP extraction order is 9 or 10 for voice samples captured by the throat microphone, which was observed that poor stressed vowel and fricative-like words "3" and "7" in Portuguese confuses the classifier.

INDEX TERMS Multi-class pattern recognition, speech recognition, neural networks, binary classifiers.

I. INTRODUCTION

The voice is one of the principals means of human communication and as an acoustic signal it carries significant information about some individual characteristics [1]–[3]. Speech is the most complex signal to classify since it depends of the physiological systems of human vocal tract, and can be influenced by transformations due to semantics, linguistics and acoustic. Furthermore, the physical speech production also changes from one person to another and, consequently, each person has a different utterance to same word [4]. If we consider only captured samples of a single person, the utterances of a word are different in time and in frequency.

There are many problems that make the speech recognition system complex. Factors such as linguistics, dialects, speakers, sex, age, region, even these factors change during life. In addition, there are factors arising from the speech capture system itself, such as sensors, wires, and the environment, alter the captured signal [5], [6].

Thus, appropriate choice of each component of the voice recognition system is one of the most difficult tasks. A speech recognition system can be composed of three basic components: acquisition, feature extraction and classification. Sometimes it takes a few other components such as preprocessing or filtering to be held after the acquisition. There are many studies in the literature about each of these components. In acquisition step, we can highlight studies comparing the use of traditional microphones and throat microphones and the use in combination of different transducers [7]–[10]. There are various methods of extracting features of voice signals, such as: Perceptual Linear Predictive (PLP) [11]–[13], Linear Prediction Coding (LPC) [14], [15], Mel-Frequency Cepstrum Coefficients (MFCC) [6], [16], Discrete Wavelet Transform (DWT) [17] and Relative Spectra Filtering of log domain coefficients (RASTA-PLP) [2], [18]. In addition to these methods with their combinations, there are various other methods that we can found in literature [19], [20].

Finally, for the last step we present several pattern classification methods that has been created for two-class classification problems. There are theoretical studies concentrated exclusively on researching binary functions including the methods using artificial neural network such as the perceptron and the error backpropagation (BP) algorithm [21]–[25]. Expand a two-class classification system to a multi-class it is complex and can cause performance degradation. Because of this, binarization techniques have arisen to treat multi-class problems by decomposing the original problem into multiple two-class classification problems [26]–[30]. Have several different methods decomposing a K-class pattern problem in two-class [31]–[37].

This paper presents a comprehensive and competitive study in multi-class neural network classification using supervised learning to classify voice commands in Brazilian Portuguese captured using a throat microphone. Our study main contribution is a comparison between the performance of two major system architectures using PLP features extracted of the speech samples: a simple neural network and a system of multiple neural networks modeled using the one-against-all approach. A summary is shown below in bullet form:

- Throat microphone chosen as acquisition method to increase signal–to–noise ratio in noise environment.
- Definition of PLP as feature extraction.
- Creation a data set with isolated voice commands in the Brazilian Portuguese.
- Evaluation the PLP order in our acquisition scenario.
- Evaluation the results using a single neural network in a speech recognition system.
- Evaluation the results using a multiple neural network using one-against-all in a speech recognition system.
- Comparison of the results and training processing time between both neural network methods.

The remainder of this work proceed with the following organization. Section II presents a brief description about voice acquisition using throat microphone. The Section III presents a description about PLP feature extraction method. Section IV explains multi-class pattern classifications using single and multiple neural networks. In Section V, are showed the results and performances of the neural network systems with different modeling approaches for our dataset, and the voice samples used in the study are described in greater detail. The Section VI concludes the study.

II. ACQUISITION VOICE SAMPLES USING THROAT MICROPHONE

Before we develop a speech recognition system, we must define an effective method audio capture, especially considering the high level of expected noise for trouble. For this, it is essential specify a robust method of acquiring signal in a noisy environment, since speech recognition systems have better performance when noiseless voice signals are used. In places with high levels of background noise, such as factories or streets, the speech recorded using a close–talking microphone is degraded and hence the recognition accuracy of a speech recognizer decreases as the signal–to–noise ratio (SNR) of its input is reduced [38]–[41]. Thus, the close–talking microphone does not seem the best option to use in noisy environments, it is necessary to choose another type of transducer for the acquisition of voice signals.

(a)

(b)

FIGURE 1. Throat microphone: (a) throat position; (b) transducer model used.

The throat microphone, placed in contact with the skin close to the Adam's apple (see Figure 1(a)), i.e. surrounding the larynx, can be seen as a transducer of the vibrations of the body tissues (skin, bone, cartilage, ...) that captures the muscles movements produced by the human speech apparatus [8], [19], [42]. Since it does not capture ambient sound, it makes it robust in noisy environments, but the voice signal captured by throat microphone has a limited frequency bandwidth because bones and tissue feature like a low-pass filter [43]. The Figure 1(b) shows the throat transducer model used in this study.

Due to this spectrum reduction the speech throat signal, they suffer a reduction of intelligibility compared to the voice recorded by close–talking microphone [19]. The main difference occurs by the lack of fricatives that are generated by the passage of the sound signal by the lips and mouth. This causes

TABLE 1. Comparison of speech recognition accuracy rate of studies in clean and noise environment using close–talking (CT) and throat microphone (TM).

Study	Clean Environment		Noise Environment	
	CT	TM	CT	TM
Heracleous *et al.* [38]	88.5%	72.4%	41.1%	42.3%
Yegnanarayana *et al.* [40]	88.6%	94.3%	25.0%	93.3%
Dupont *et al.* [42]	98.9%	95.7%	65.4%	81.6%
Radha *et al.* [44]	35%	56%	-	-

some high frequency components been lost in the signal obtained using throat microphone. But in noisy environment the close–speaking microphone capture much noise that can reduce the SNR affecting the voice becoming unintelligible, making the throat microphone a better approach. As shown in Dupont *et al.* [42], the noise level is almost constant for the throat microphone up to a noise level of 75 dBA, that is similar to a very bustling street. This result means that, up to this level, background noise is practically not captured by the throat microphone [40], [44], [45].

Typically, the signal of throat microphone contains low frequency, amplitude or energy when compared to the close–speaking microphone signal, but it is interesting to note that the throat speech is a high quality one [40], [45].

In Heracleous *et al.* [38], Yegnanarayana *et al.* [40] and Dupont *et al.* [42], they compare the recognition rate of a speech recognition system using both, throat microphone and close–talking microphone, in clean and noise environment that are showed in Table 1. It also show the recognition rate of the study of Radha *et al.* [44], but it only measure recognition rate in clean environment. Those comparison reveal the robustness of throat microphone. Therefore, the throat microphone has been choose to be used in this work to capture voice commands.

An important information about those results is that they did not use the same methods for feature extraction, pattern recognition, language or database, so the direct comparison between those studies is unreliable, not being possible to evaluate the quality of a work comparing with the others due to the difference of context. The relevant information observed is the accuracy rate of close–talking microphone and throat microphone in the same study.

III. FEATURES EXTRACTION

For speech processing, the signal need to be represented in a parametric form [14]. This process is called feature extraction wherein the data space is converted into a feature space having the same dimension as the original data space, but it is represented by a reduced number of effective characteristics. The methods PLP, LPC and MFCC have been used by the most of speech processing systems [2], [11], [14], [46]. Since PLP method is more suitable to human hearing [13], we selected PLP parametrization technique to use in this work.

The PLP method is based on the conception of the human hearing physiology trying to represent the human speech.

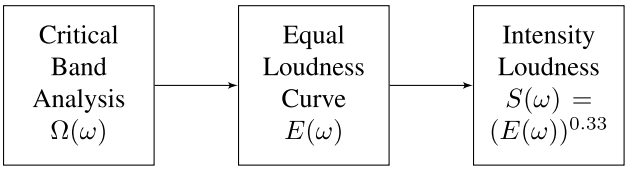


FIGURE 2. Block Diagram of PLP features extraction process [13], [47].

The PLP method is similar to the LPC excepting by the spectral information that PLP try to satisfy the nature of the human auditory model [47]. The steps of PLP extraction are based on our perceptual auditory characteristics and are composed of three phases called cubic–root, which are: Critical Band Analysis, Equal Loudness Curve and Intensity Loudness Conversion. This PLP processing steps are shown in Figure 2.

The first stage, Critical Band Analysis, consists of applying a conversion from hertz frequency to bark frequency, that act more appropriate like human hearing in spectral information. Here, power spectrum is wrapped along the axis frequency into bark frequency [47]. Then, a pre–emphasis equal loudness curve $E(\omega)$ is applied to the filter–bank coefficients to reproduce the feeling of hearing human audition. The equalized vector is converted using Stevens power law, $S(\omega) = (E(\omega))^{0.33}$.

The distinct phases of PLP features computation is shown in Figure 3. These steps are common in almost all the feature extraction process. Firstly, the audio frame is passed through a Hamming Window, then it is applied a Fast Fourier Transform (FFT) and it is computed the power spectrum of the speech signal as $P(\omega) = Re(S(\omega))^2 + Im(S(\omega))^2$ [13]. Then, PLP extraction is applied as described above. After, the follow auditory distortion line spectrum is calculated with a linear prediction (LP), where the predictive coefficients are processed as a spectral power signal. Finishing calculating the ceptral coefficients taking the predictive coefficients with a recursion that are similar of the logarithm of the model spectrum proceeded by an inverse Fourier transform [19].

IV. MULTI-CLASS PATTERN CLASSIFICATIONS USING NEURAL NETWORKS

Pattern recognition (PR) is a part of learning machine whose goal is to classify information (patterns) using a priori knowledge or/and statistical information extracted from those patterns [3]. Those techniques are typically used to classify data into groups, making the system understand how to distinguish such groups, allowing the classification of new data within this set of groups [48], [49]. Neural Networks (NN) are one of the most classifiers known in the literature and, therefore, it was chosen to classify voice commands in this work. The NN models are described by their network topology, neuron features, learning rules and training methods. The topology term refers to the entire network structure, specifying how the inputs, outputs and hidden layers are interconnected [29].

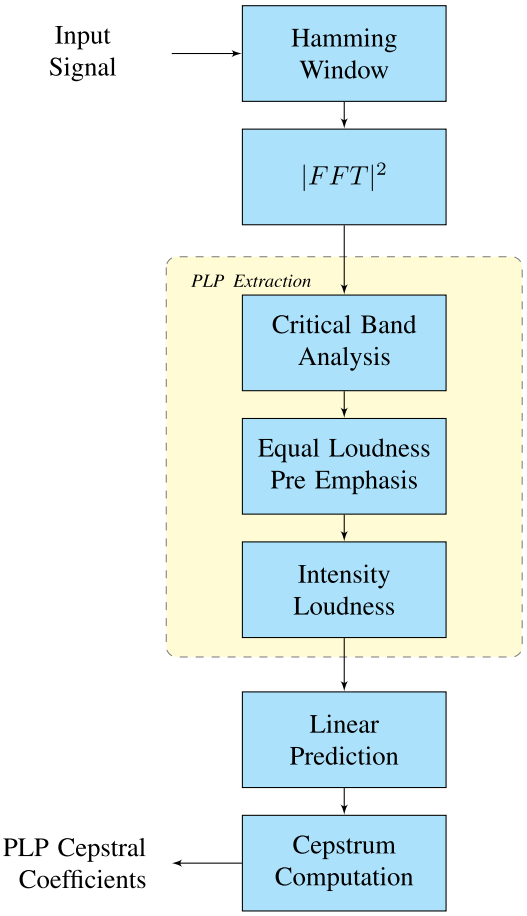


FIGURE 3. PLP Features Computation [13], [47].

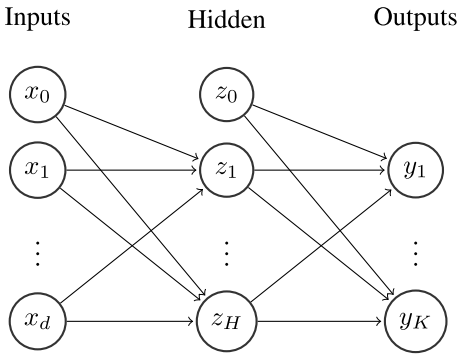


FIGURE 4. Single Neural Network architeture for implementing *K*-class pattern classification.

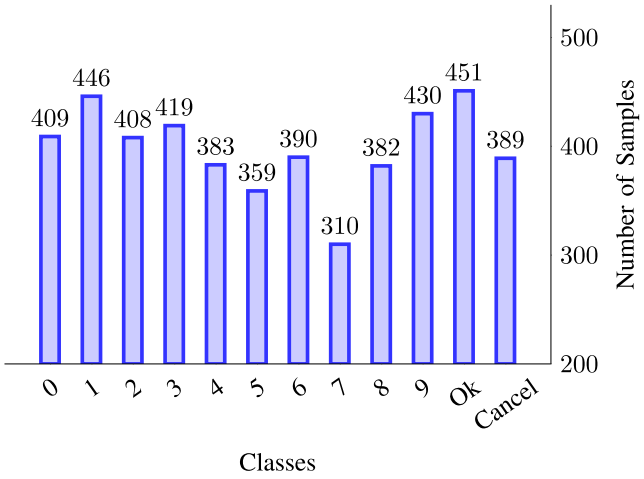


FIGURE 5. Histogram of the amount of samples per class.

A Multi-class pattern recognition system accurately trace an set of input feature to a set of output space with more than two classes [27]. Usually, build a classifier to differ only two classes is more easy than considering more than two classes, because it simplify the definitions of decision boundaries [50]. Basically has two process used in researched literature to deal with multi-class problems using binary classification techniques: adaptation of the internal operations of the classifier training algorithm and decomposition of the multi-class problem on a group of two-class classification process. The learning algorithm is the primary issue when extends to a multi-class pattern, showing that is a complex task, and in some cases, even impractical [49].

A K-class pattern recognition system can be developed using two different architectures, a single neural network with K outputs or a system of multiple neural networks [51]. Decomposing the original problem into multiple two-class classification problems reducing the multi-class classification in to K binary problems. The two more popular methods are: one-against-all (OAA) [31], [33] and one-against-one (OAO) [34], [52]. In the OAA approach, each K classes are trained against all other classes, i.e. the training data set is adapted to train each classifier using the positive data with the classes belong to then, and making all others classes been negative [35], [51]. In the OAO approach, each K classes are trained against every other classes, i.e. the training data set is adapted to train each acceptable pair of classes discarding all others classes. The OAO approach can be developed only in a system of $K(K-1)/2$ binary neural networks. Although these two methods are the most obvious, Allwein *et al.* [53] shown that have many other ways that a multi-class problem may be decomposed on to a number of binary classification problems [35].

A. K-CLASS PATTERN CLASSIFICATION USING A SYSTEM OF MULTIPLE NEURAL NETWORKS

A K-class pattern recognition system can be developed using $M > 1$ neural networks. Each neural network have their own training using a adapted part of the training data set. To integrate all M neural networks a decision rule usually is used to obtain the final output. This methodology including value of M, training method, decision rule determines the modeling scheme. As we have seen, multiple neural network has some different schema, OAA and OAO, making them a powerful tool.

1) THE ONE-AGAINST-ALL APPROACH

The one-against-all (OAA) approach use a system of K binary neural networks, NN_i, $i = 1, ..., K$, and each neural network,

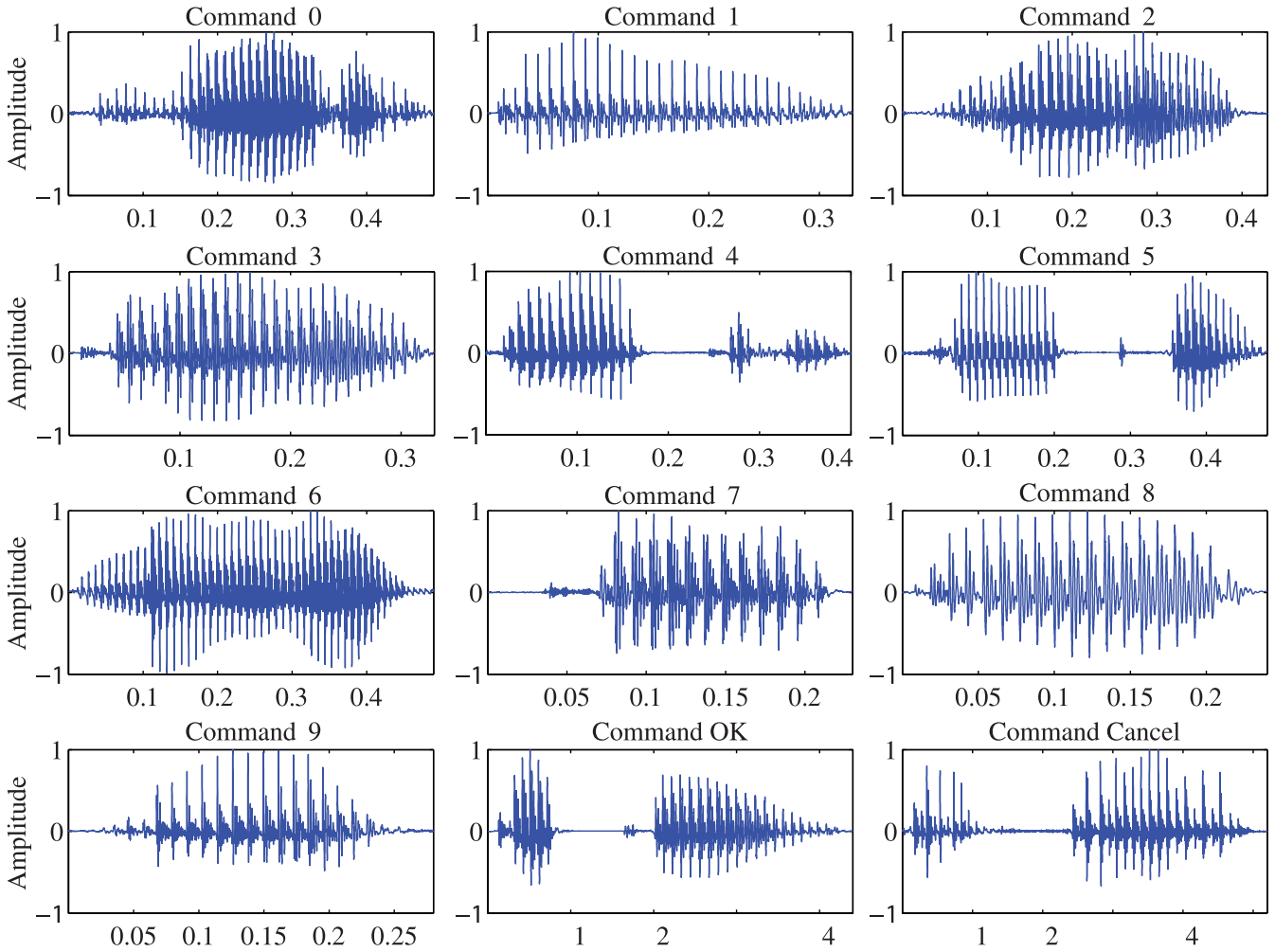


FIGURE 6. Examples of all voice commands in Brazilian Portuguese captured using a throat microphone.

NN_i has one output node O_i with output function f_i being modulated based on output y_i to $f_i(\bar{x}) = 1$ or $f_i(\bar{x}) = 0$ to represent whether the input pattern $\bar{x}$ belongs to class i or output $f_i(\bar{x}) = 0$ when does NOT belong to class i. Each neural network use the same data set in training process but use different labels in classes. Because of this feature, this system of K binary neural networks has a number of advantages [51]:

- Each neural network has they own feature space and, therefore, they can select the features that best fit each neural network;
- Each neural network has they own architecture such as the number of hidden layers, the amount of hidden nodes, the calculation of the activation functions, etc.;
- The training process of each K binary neural networks is independent, so it can be executed simultaneously in different machines to speed up training time.

Though these advantages, Ou and Murphey [51] describes that the OAA approach has two major drawbacks: it may have difficulty in learning minority classes if the training set is imbalanced, and the decision boundaries obtained by the system can overlap or not in the space of features. This problem of the data being highly imbalanced set for each individual neural network occurs when the number of training sample in each class is approximately equal. So, when K is large, the training set for neural network i is highly imbalanced. This may result in totally ignoring the minority classes. In the case of backpropagation algorithm, the neural network can be biased toward the majority class, or to the "other classes" [35], [51].

2) THE ONE-AGAINST-ONE APPROACH

The concept of the K-class pattern classification arises separating the $K(K-1)/2$ class into two-class classification problems applying the OAO modeling approach [51], also recognize as pair-wise method [32]. In this approach, each problem is solved by a neural network, denoted as $NN_m(i, j)$, $m = 1, 2, \cdots, K(K-1)/2$, which is a neural network trained to discriminate class i from class j, for $1 \leq i < j \leq K$ [51].

The improvement of OAO method arise in the training independent of $K(K-1)/2$ binary neural networks, this

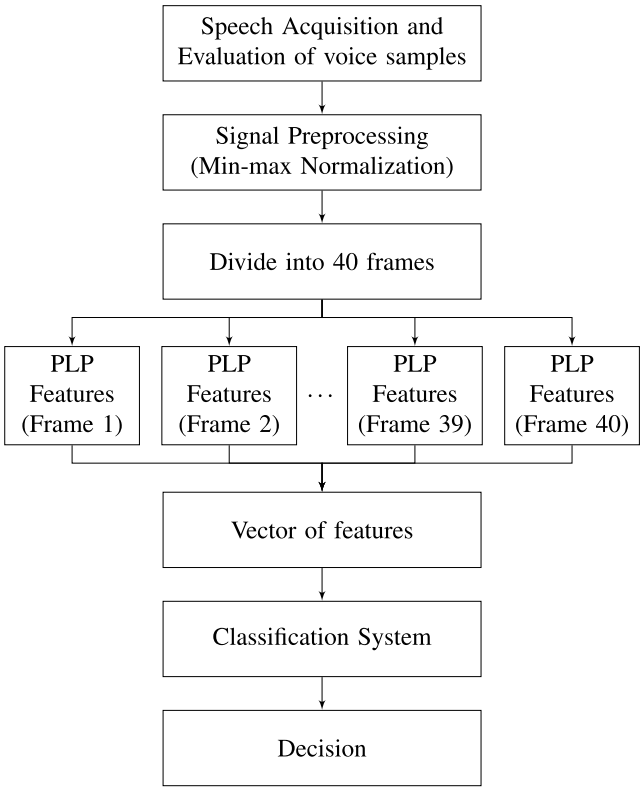


FIGURE 7. Summary of the algorithm used for analyzing the voice signal.

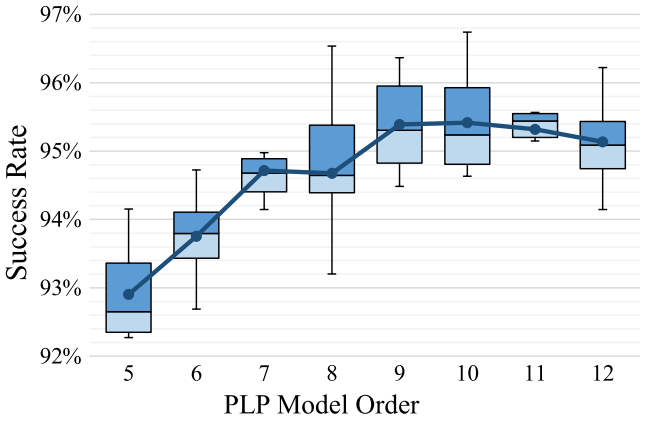


FIGURE 8. Box plot of binary neural networks.

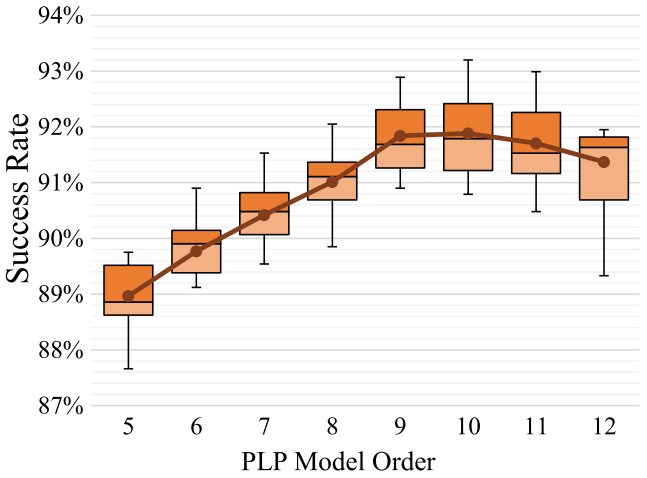


FIGURE 9. Box plot of single neural network.

grant a level of redundancy in classification of pattern classes, enhancing the generalization of the classifier. Requiring training a amount $K - 1$ distinct neural networks. Furthermore, a neural network system modeled with OAO has good flexibility in terms on generation of features in independent spaces, independent neural network architectures, and simultaneous training of multiple neural networks on different computers, as well as the OAA approach. Due to each neural network is trained using only two classes, OAO modeling method has no problems of imbalance of the data set and the feature space is less likely to have uncovered space as determined by modeling of OAA method [51].

A disadvantage of this approach is that for large values of K, several neural networks are required, which can increase the computational cost in training and also for sorting. Furthermore, the decision function is much more complex than in the OAA modeling method.

3) AGGREGATION SCHEMES FOR BINARIZATION TECHNIQUES

During the testing phase, after the data has been pre-classified by each binary neural network, occurs that appears more than one good probabilistic result, so a decision rule must be taken, normally it is used the probabilistic confidence of each classifier together to obtain the final output.

There are different methods to combine its outputs. In recent year, some strategies have been developed, for instance, new methods of probability estimates, binary-tree based strategies or dynamic classification schemes, to complement some recognized approach such as Weighted Voting, Pair-wise Coupling or Max-Wins rule. A detailed study about the major aggregation strategies for binarization techniques is described in Galar *et al.* [50]. For the one-against-all approach, the most commonly used aggregation strategy is the maximum confidence strategy (MAX), where the result output class is defined by the class with the highest positive value.

B. K-CLASS PATTERN CLASSIFICATION USING A SINGLE NEURAL NETWORK

A K-class pattern classification might be implemented by a single neural network in an architecture with d input nodes and K output nodes, as we can see in Figure 4, where d is the dimension of an input feature vector and K is the amount of output nodes in the neural network system. For a training data example $\bar{x}$, the expected output of the neural network at the output node y_i is set to 1 if and only if $\bar{x}$ belongs to class i, otherwise it is set to 0, for $i = 1, \ldots, K$. Since only one neural network it is used to represent multiple classes, the structure

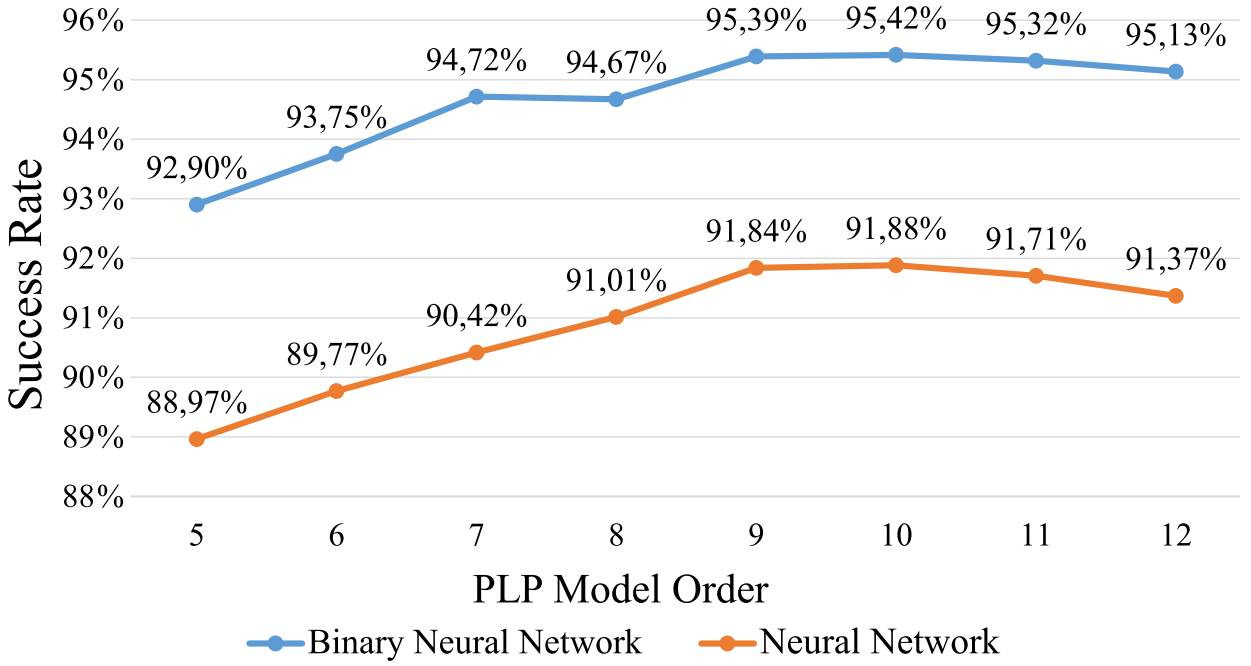


FIGURE 10. Comparing average hit rates of single neural network and binary neural networks.

of the neural network must be more complex than others used in the binary neural networks [51]. This is usually because all classes use the same feature space during the training phase. Thus, the higher the dimension of the feature space, the higher the complexity of the neural network architecture, which is already high in a single neural network system with K output nodes.

In general, the training step for a single neural network is simple to deal with, considering only one training data set and neural network to train. However, the training time will be very long when the training data are many and the number of pattern classes is large, which makes the fine tuning of the neural network structure and learning parameters difficult.

Ou and Murphey [51] identified that this single neural network system has many features in common with a system of multiple binary neural networks modeled by OAA, for example both ''ignores the minority classes when the training data are imbalanced since the output node corresponding to a minority class was set to 1 in much less time than a node corresponding to a majority class''. However, they also pointed out that there are differences between the two systems. One difference is that, during the single neural network learning step, the training data from all classes is presented to all neurons, allowing the neural network to set an optimal boundary decision. In addition to this difference, the single neural network design share features in hidden layer to multiple classes, and a part of the neurons can specialized for a few classes, that allow be ignored in other classes, reducing the weights associated with them. Therefore, if the neural network is properly trained, it can minimize the feature space of regions uncovered and overlapped.

V. RESULTS AND DISCUSSION

The effectiveness of the proposed approach to extract content descriptive metadata has been evaluated and the details are discussed in this section. First, we outline the setup of the experiment and the datasets on which the experiments are executed.

A. EXPERIMENTAL PARADIGM

To evaluate the classification systems studied in this work, we created a data set with isolated voice commands in the Brazilian Portuguese language, with utterances captured from 150 people (men and women), pronouncing each command three times, using a throat microphone. The captured commands consisted of the digits 0 through 9 and the words Ok and Cancel, i.e. the data set has 12 words classes.

All voices samples were quantized in amplitude with 16 bits, recorded in mono-channel WAV format with 48,000 Hz of sampling frequency to preserve the fidelity of the signal. The voice samples were recorded in an open environment with presence ambient noise from cars, wind, animals and people, in which volunteers were asked to produce each word in about 2 seconds. This means that there were no strict criteria for registration of the words. Thus, utterances of the same word may differ considerably in length and consequently in number of samples. A total of 5400 utterances were captured, of which we have selected 4776 samples after analysis by removing very noisy or corrupted samples. The number of samples per class is shown in Figure 5, as we can see, the database generated is relatively balanced with little variation between classes.

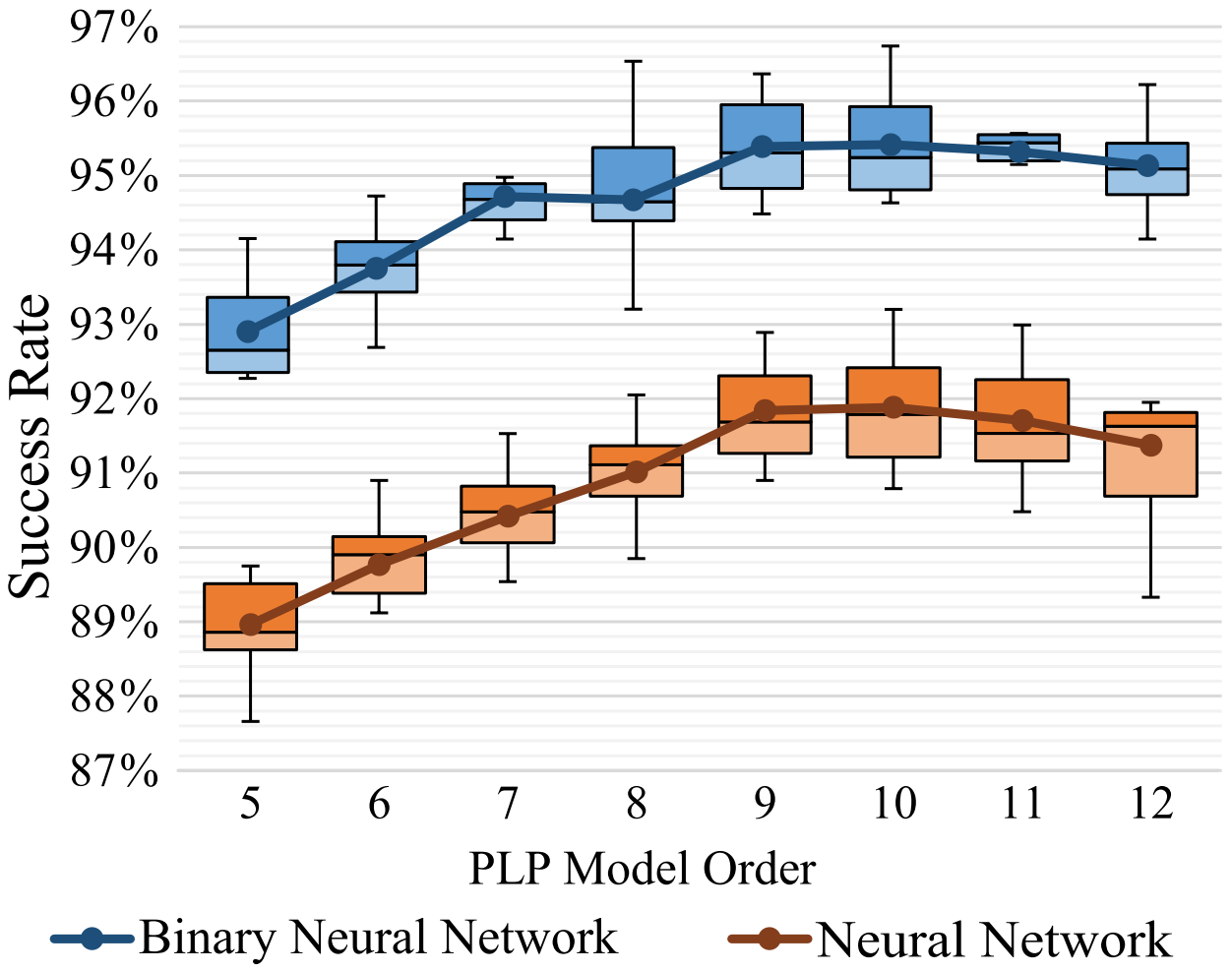


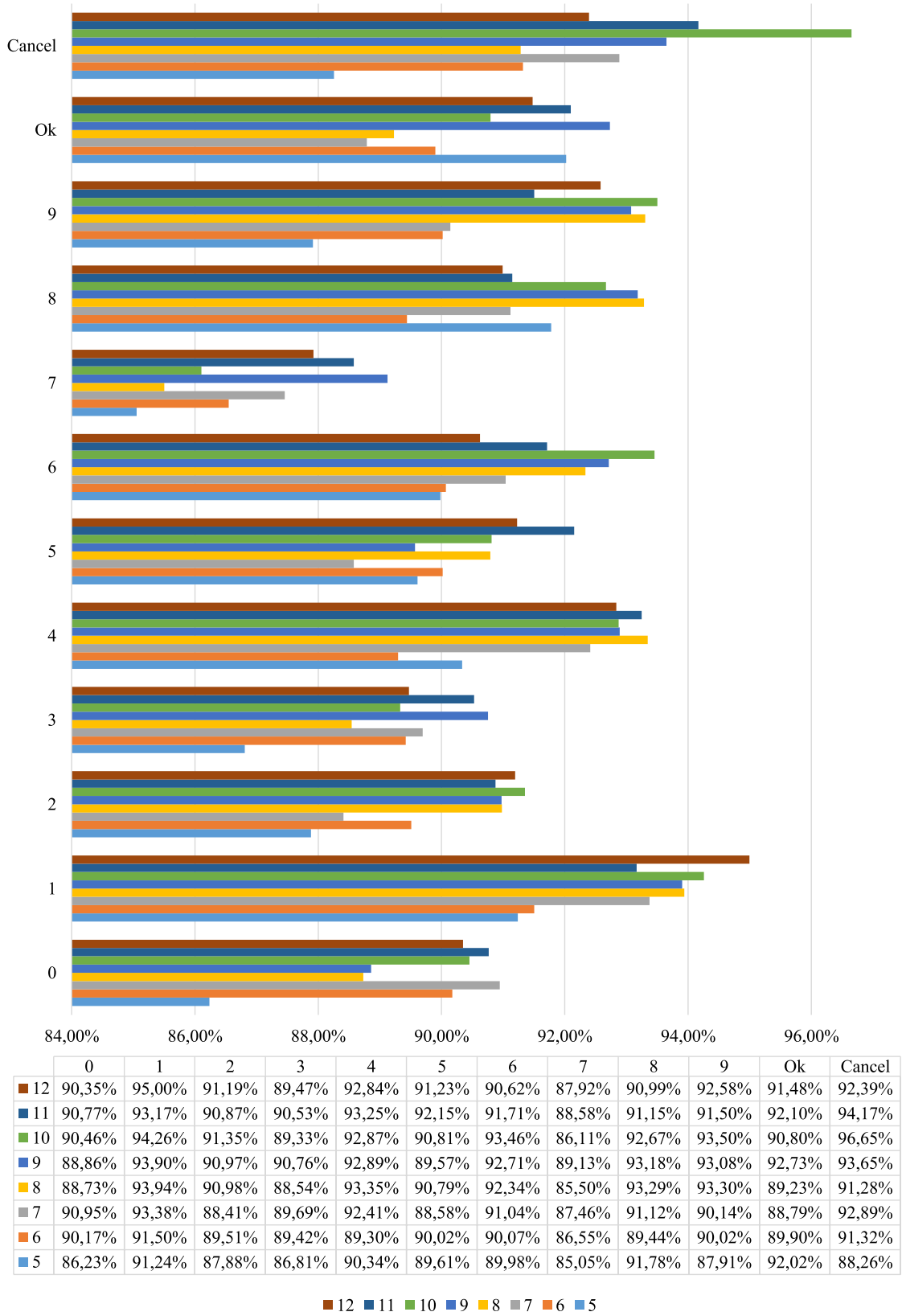
FIGURE 11. Box Plot comparing single neural network and binary neural networks.

Then we filter some noise present in the signals and remove the silence inserted at the beginning and end of each sample. Thus, the samples were of different sizes, but lasting no more than one second. After removal of silence, the signals were processed in the min-max amplitude normalization to result in signals with amplitudes between -1 and 1. The Figure 6 shows examples of each command captured using throat microphone after this pre-processing step.

Features are extracted of the signals in order that the classification system can distinguish each class. As the classification system studied in this paper is based on supervised neural networks, which depend on a fixed size of the input vector, there is a big problem for the classification of voice commands spoken spontaneously because, as we have seen, the length of each word spoken by different people can vary considerably in time. One solution to this problem is to perform a normalization of the speech, such that all the utterances have the same size. The most basic ways to make this normalization is limiting the number of frames to be used for the classifier. In this paper, we realize the normalization of each sample, fragmenting it into 40 frames of 25 milliseconds with or without overlap between adjacent frames. These values were chosen because, despite of voice signal is non-stationary, it can be considered stationary for periods of time between 10 and 30 milliseconds and using 40 frames we can break signals up to one second in duration [8], [54]. The number of frames used was defined empirically after many simulated tests.

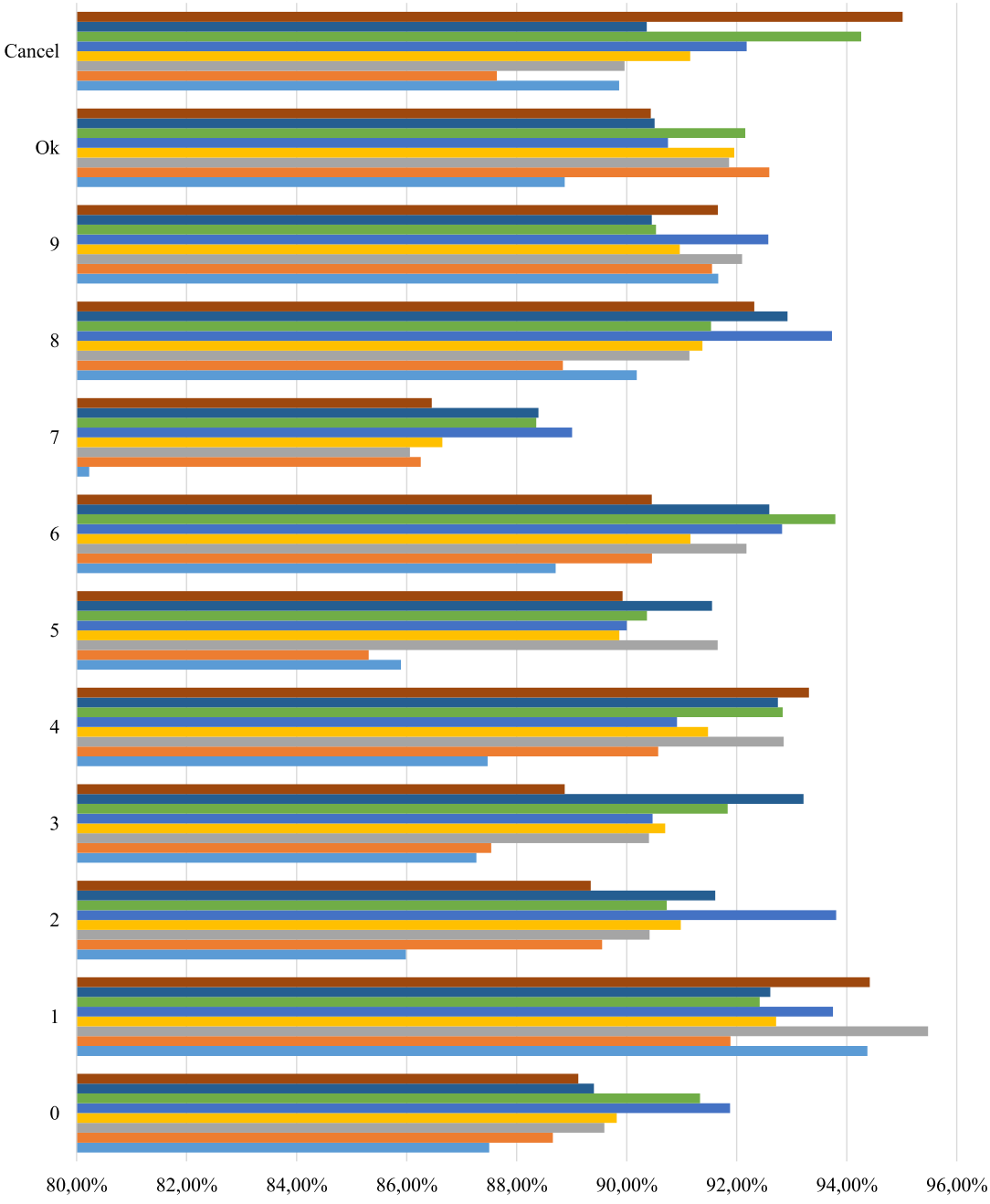
The PLP feature extraction method is applied to each frame, varying the order of the autoregressive PLP model between 5 and 12, which are the values with the best average performance for isolated words recognition applications, as demonstrated by Hermansky [13]. An order N of the autoregressive PLP model generates $N + 1$ features extracted for each frame. Thus, for 40 frames are extracted a total of $40(N + 1)$ features of each voice sample, which will make up the classification system input vector. Once extracted the features, the problem is to obtain a discriminant function to separate the different classes in feature space [3]. The Figure 7 presents a summary of our voice commands recognition system.

In this work, we compare two classification system: a single neural network and a system with 12 binary neural networks using the OAA modeling approach. For both we used



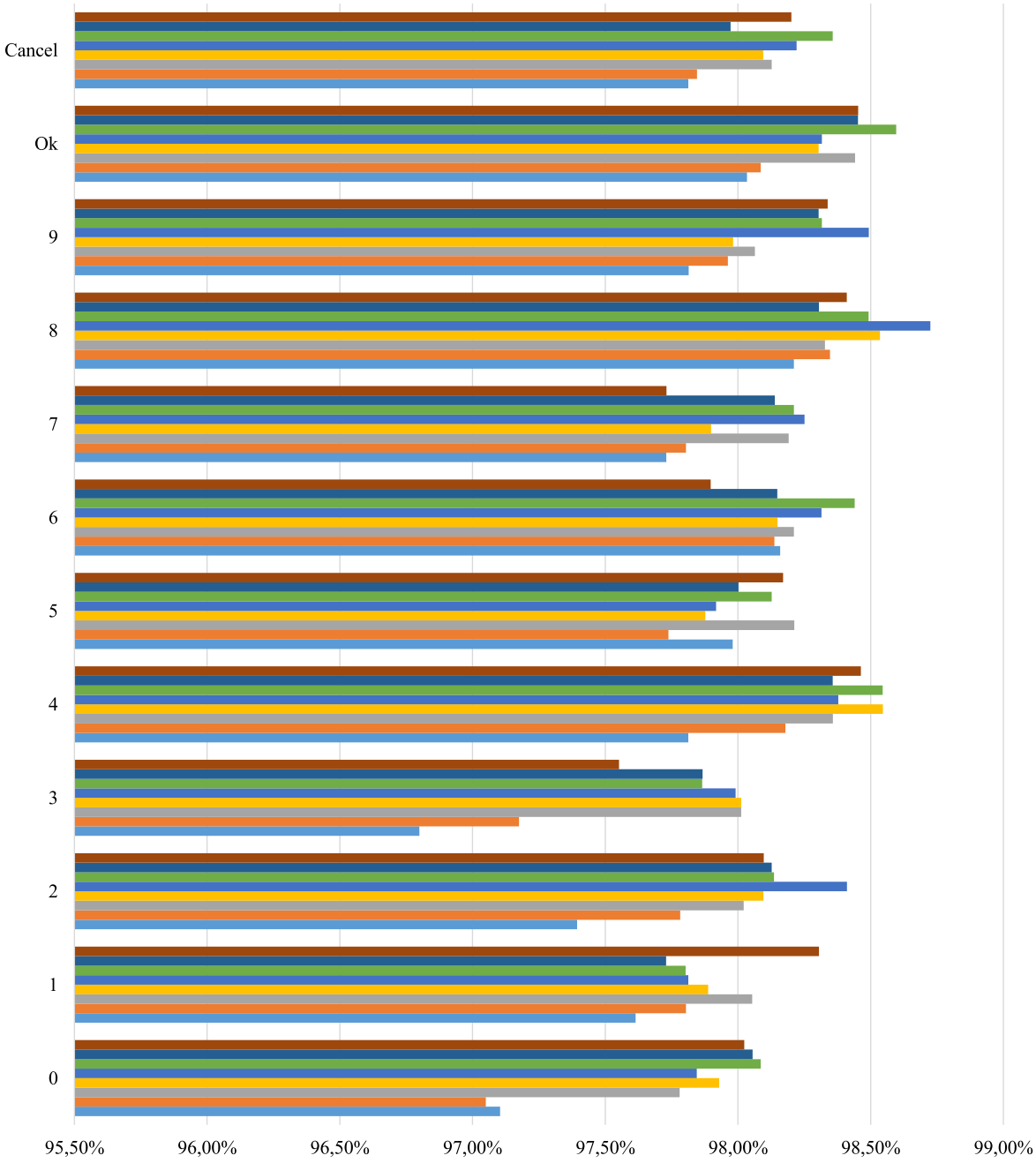
	0	1	2	3	4	5	6	7	8	9	Ok	Cancel
12	90,35%	95,00%	91,19%	89,47%	92,84%	91,23%	90,62%	87,92%	90,99%	92,58%	91,48%	92,39%
11	90,77%	93,17%	90,87%	90,53%	93,25%	92,15%	91,71%	88,58%	91,15%	91,50%	92,10%	94,17%
10	90,46%	94,26%	91,35%	89,33%	92,87%	90,81%	93,46%	86,11%	92,67%	93,50%	90,80%	96,65%
9	88,86%	93,90%	90,97%	90,76%	92,89%	89,57%	92,71%	89,13%	93,18%	93,08%	92,73%	93,65%
8	88,73%	93,94%	90,98%	88,54%	93,35%	90,79%	92,34%	85,50%	93,29%	93,30%	89,23%	91,28%
7	90,95%	93,38%	88,41%	89,69%	92,41%	88,58%	91,04%	87,46%	91,12%	90,14%	88,79%	92,89%
6	90,17%	91,50%	89,51%	89,42%	89,30%	90,02%	90,07%	86,55%	89,44%	90,02%	89,90%	91,32%
5	86,23%	91,24%	87,88%	86,81%	90,34%	89,61%	89,98%	85,05%	91,78%	87,91%	92,02%	88,26%

FIGURE 12. Average hit rates per class of single neural network.



	0	1	2	3	4	5	6	7	8	9	Ok	Cancel
12	89,12%	94,42%	89,35%	88,87%	93,31%	89,92%	90,46%	86,46%	92,32%	91,66%	90,44%	95,02%
11	89,41%	92,61%	91,61%	93,21%	92,75%	91,55%	92,59%	88,40%	92,92%	90,45%	90,51%	90,37%
10	91,33%	92,42%	90,73%	91,84%	92,84%	90,37%	93,79%	88,36%	91,53%	90,53%	92,16%	94,26%
9	91,88%	93,75%	93,81%	90,47%	90,91%	90,00%	92,83%	89,01%	93,73%	92,57%	90,75%	92,18%
8	89,82%	92,72%	90,98%	90,70%	91,48%	89,87%	91,16%	86,65%	91,38%	90,97%	91,95%	91,15%
7	89,59%	95,48%	90,42%	90,40%	92,86%	91,65%	92,18%	86,06%	91,14%	92,10%	91,86%	89,96%
6	88,66%	91,89%	89,55%	87,54%	90,57%	85,31%	90,46%	86,26%	88,84%	91,55%	92,59%	87,64%
5	87,50%	94,37%	85,99%	87,27%	87,47%	85,90%	88,71%	80,23%	90,18%	91,66%	88,87%	89,86%

FIGURE 13. Average hit rates per class of binary neural networks.



	0	1	2	3	4	5	6	7	8	9	Ok	Cancel
12	98,02%	98,31%	98,10%	97,55%	98,46%	98,17%	97,90%	97,73%	98,41%	98,34%	98,45%	98,20%
11	98,06%	97,73%	98,13%	97,87%	98,36%	98,00%	98,15%	98,14%	98,31%	98,30%	98,45%	97,97%
10	98,09%	97,80%	98,14%	97,87%	98,55%	98,13%	98,44%	98,21%	98,49%	98,32%	98,60%	98,36%
9	97,85%	97,81%	98,41%	97,99%	98,38%	97,92%	98,32%	98,25%	98,73%	98,49%	98,32%	98,22%
8	97,93%	97,89%	98,10%	98,01%	98,55%	97,88%	98,15%	97,90%	98,54%	97,98%	98,30%	98,10%
7	97,78%	98,05%	98,02%	98,01%	98,36%	98,21%	98,21%	98,19%	98,33%	98,06%	98,44%	98,13%
6	97,05%	97,80%	97,78%	97,18%	98,18%	97,74%	98,14%	97,80%	98,35%	97,96%	98,09%	97,85%
5	97,10%	97,61%	97,39%	96,80%	97,81%	97,98%	98,16%	97,73%	98,21%	97,81%	98,03%	97,81%

FIGURE 14. Average hit rates per classifier in the binary neural networks.

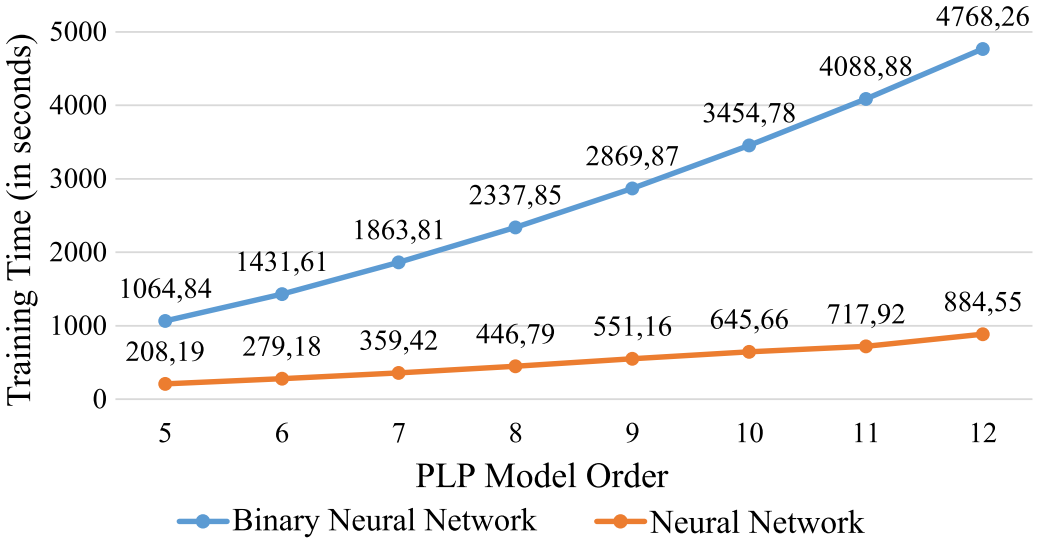


FIGURE 15. Comparison between the time for training.

the topology of the Multi-Layer Perceptron (MLP), as one of the most Artificial Neural Network (ANN) commonly used to separate data is not linearly separable like as voice word samples. We use an MLP neural network with two layers, being one hidden layer, whose number of neurons q has been determined using the KolmogorovâĂŹs rule, $q = 2n + 1$, where n is the number of attributes used in classification. The parameters of neural network are 300 training epochs, MSE desired 10^{-4} and learning factor of 0.01, were empirically chosen after several simulations to find the values with better accuracy.

B. EVALUATION

To evaluate the performance of each classification system, the data set is divided into two sets: one of training and another of test. Samples are randomly shuffled and them 80% is selected to the training, while 20% is used in the test. The evaluation is based on hit rates average, maximum, minimum and standard deviation. The results are drawn from 10 independent simulations.

The Figure 8 shows the result of the binary neural networks using the OAA modeling approach varying the order of the autoregressive PLP model between 5 and 12. As it is possible to analyze even increasing the order of the PLP extraction, i.e., increasing the amount of embedded features in the classifier, the larger average hit rate, 95.38% and 95.42%, occurred with the order of extraction between 9 and 10 respectively. In the Figure 9, we can see the same behavior for a single neural network, where the highest average hit rates, 91.84% and 91.88%, occurred with the order of PLP extraction between 9 and 10 respectively.

Comparing the two approaches, we find that for all values of the order of PLP extraction, single neural network always had lower performance than the binary neural networks, as shown in the Figure 10. The average hit rate of the difference for each PLP extraction order between the two neural networks is on average 3.79%, that is, the real gain observed in tests by using a set of binary neural networks is approximately 4%. A relevant data is that this 4% presents to be almost constant in all PLP extraction orders comparing between two neural network, that can be seen in Figure 10. Proving that, even with some independence of the presented data, the binary neural networks it behaves better at a level almost fixed than the single neural networks.

Comparing the boxplots of network neural single and neural networks binary, shown in Figure 9 and Figure 8 respectively, in a single boxplot, as shown in Figure 11, we find that even considering the best performance of neural single network in the best possible configuration, this still got underperformed the worst performance of the set of binary neural network. This result allows us to verify that schema binary multiple neural networks are more efficient than for a single neural network.

To verify that the performance gain by using a set of binary neural networks was real, we found the average performance of each class for different values of PLP extraction order. The Figure 12 shows the individual performance of each class to the single neural network in which it is found that increasing the PLP extraction order, the average hit rate of each class also increased following a similar pattern to the previous described in this work and also demonstrated by Hermansky [13], in that the best performance was generally obtained from the PLP extraction order of between 9 and 11. Moreover, we verify that among all classes, the "7" class had the worst performance, with less than 90% rates, while most classes achieved superior performance to 92%. For the classifier composed of binary neural networks, hit rates per class obtained similar results to those obtained by single neural network, as shown in Figure 13.

Despite of performance by class have been very similar between the two approaches, the binary neural network could have higher average performance due to the comparison scheme output of each classifier before the final classification. As shown in Figure 14, it appears that most of the individual classifiers received performance between 97.5% and 98.5%. This result confirms that there are few cases where the samples are misclassified, which considerably increases the average performance of the classifier multiple binary neural networks.

Comparing the mean hit rate of binary neural network that was evaluated in a noise environment that reached 95.42% of success rate, show that our approach has better results that all studies compared on Table 1 in a noise environment that are: 42.3%, 81.6% and 93.3%.

Finally, the comparison between the time for training the two classification approaches, using a CPU Intel Core i7-4810MQ @ 2.8 GHz and 8 GB RAM, is shown in Figure 15. This result confirms that the binary neural networks need more time for training that the single neural network, approximately 535% more in PLP extraction order 10, since each binary neural network is trained independently and this time exponentially grows as it increases the number of extracted features with increase of PLP extraction order. However, this value can be improved by using parallelism with multiple processors to optimize the training of each binary network.

VI. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Our paper intends to increase research interest on the use of binary classifiers with neural networks for voice commands classification. Although most of the current research are to find new algorithms for pattern classification and extraction of new features of voice commands, we claim that it is still worth the study on existing methods in the literature, only grouped and used in such a way which allows achieving much higher hit rates than through conventional techniques. Thus, we compared the results obtained by a classifier based on a single neural network and a group of binary neural networks, using the one-against-all approach, recognizing voice commands captured using a throat microphone and whose characteristics were extracted using the Perceptual Linear Predictive method (PLP).

The results showed that the use of multiple binary neural network reach 95.42% of hit rate, increasing the recognition performance for this problem in approximately 4% compared to the use of a simple neural network. This comparison also presents multiple binary neural to be always better, almost constant in all PLP extraction orders. We found that the best configuration of PLP extraction order is 9 or 10, with a order less than 8 it did not represent all of data information, and more than 11 increase data that confuses the classifier. Beside there was an increase in time for classifier training, that occurs because each binary classifier is responsible for determining an area decision from a given class and all other, increasing thereby the surface of separation between the classes and, consequently, the performance of the classifier. These 4% increase in hit rate is very representative, since both classifiers obtained hit rates higher than 90%.

Another important result of this work is that some classes had worse results due to the pronounced features of these words in Portuguese like "3" and "7". These words are characterized by having poor stressed vowel and fricative that can decrease the sound intensity and, therefore, they carry a smaller amount information that can be extracted by the PLP technique.

The data set with samples of voice commands of 150 people in Brazilian Portuguese captured using throat microphone and the application of these methods previously discussed in this data set is another important contribution of our work. Since no other work was found in the literature with these characteristics.

Summary of results obtained:

- Throat microphone is robust in noise environment.
- Production a data set of voice commands of 150 people in Brazilian Portuguese.
- The best configuration was reached with PLP extraction order of 9 or 10.
- Binary neural network mean hit rate obtained is 95.42%.
- Single neural network mean hit rate obtained is 91.88%.
- Best hit rate found in studies evaluated in a noise environment comparing binary neural network that reach 95.42% with others showed in Table 1 (42.3%, 81.6% and 93.3%).
- Observation that poor stressed vowel and fricative confuses the classifier like words "3" and "7" in Portuguese.
- Increase of 535% in time for training comparing binary neural networks with single neural network.

In the future, we will explore other techniques binarization classification to improve the performance of the classifier based on neural networks, we intend to extend this technique to other types of classifiers, and also compare with others classifiers like Support Vector Machine (SVM), Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM).

FUNDING

This research was funded by Financiadora de Estudos e Projetos (FINEP/Brazil) grant number 01.14.0068.03.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support by PolibrásNet, Computer Systems Engineering Laboratory (LESC), Department of Teleinformatics Engineering (DETI), of Federal University of Ceará (UFC), and Federal Institute of Education, Science and Technology of Ceará (IFCE), Brazil, by supporting this research. They also thank all the other researchers in LESC and the students who helped in contributing to generation of the dataset.

REFERENCES

[1] V. Zwass, "Speech recognition," in *Encyclopedia Britannica*. Britannica Academic, Feb. 2016. Accessed: Nov. 15, 2018. [Online]. Available: http://academic.eb.com/levels/collegiate/article/speech-recognition/126497

[2] A. Mansour, F. Chenchah, and Z. Lachiri, "Emotional speaker recognition in real life conditions using multiple descriptors and i-vector speaker modeling technique," in *Multimedia Tools and Applications*. New York, NY, USA: Springer, Jul. 2018, pp. 1–18, doi: 10.1007/s11042-018-6256-2.

[3] R. T. S. Carvalho, C. C. Cavalcante, and P. C. Cortez, "Wavelet transform and artificial neural networks applied to voice disorders identification," in *Proc. 3rd World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Oct. 2011, pp. 371–376.

[4] N. N. Diep and A. A. Zhdanov, "Neuron-like approach to speech recognition," *Program. Comput. Softw.*, vol. 44, no. 3, pp. 170–180, May 2018, doi: 10.1134/S0361768818030088.

[5] M. Ravanelli and M. Omologo, "Automatic context window composition for distant speech recognition," *Speech Commun.*, vol. 101, pp. 34–44, Jul. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639318300128

[6] M. Alsulaiman, A. Mahmood, and G. Muhammad, "Speaker recognition based on arabic phonemes," *Speech Commun.*, vol. 86, pp. 42–51, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639315300649

[7] K. N. Khadar, "A review on multimodal speaker recognition," *Asian J. Pharmaceutical Clin. Res.*, vol. 10, no. 13, pp. 382–384, Apr. 2017. [Online]. Available: https://innovareacademics.in/journals/index.php/ajpcr/article/view/1976

[8] N. Radha, A. Shahina, P. Prabha, S. B. T. Preethi, and K. A. Nayeemulla, "An analysis of the effect of combining standard and alternate sensor signals on recognition of syllabic units for multimodal speech recognition," *Pattern Recognit. Lett.*, pp. 1–17, Oct. 2017, doi: 10.1016/j.patrec.2017.10.011.

[9] A. Vijayan, B. M. Mathai, K. Valsalan, R. R. Johnson, L. R. Mathew, and K. Gopakumar, "Throat microphone speech recognition using MFCC," in *Proc. Int. Conf. Netw. Adv. Comput. Technol. (NetACT)*, Jul. 2017, pp. 392–395.

[10] L. R. Mathew and K. Gopakumar, "Voice analysis using acoustic and throat microphones for speech therapy," in *Proc. Interspeech*, Sep. 2018, pp. 173–174.

[11] K. Feroze and A. R. Maud, "Sound event detection in real life audio using perceptual linear predictive feature with neural network," in *Proc. 15th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2018, pp. 377–382.

[12] T. A. Mesallam *et al.*, "Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms," *J. Healthcare Eng.*, vol. 2017, Oct. 2017, Art. no. 8783751. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5672151/

[13] H. Hermansky, "Perceptual linear predictive PLP analysis of speech," *J. Acoust. Soc. Amer.*, vol. 87, no. 4, pp. 1738–1752, Apr. 1990.

[14] S. Agrawal and D. K. Mishra, "Speaker verification using mel-frequency cepstrum coefficient and linear prediction coding," in *Proc. Int. Conf. Recent Innov. Signal Process. Embedded Syst. (RISE)*, pp. 543–548, Oct. 2017.

[15] F. Jelinek, *Statistical Methods for Speech Recognition*. Cambridge, MA, USA: MIT Press, 1997.

[16] B. B. Ali, W. Wójcik, O. Mamyrbayev, M. Turdalyuly, and N. Mekebayev, "Speech recognizer-based non-uniform spectral compression for robust MFCC feature extraction," *Przeglad Elektrotechniczny*, vol. 94, no. 6, pp. 90–93, Jun. 2018.

[17] R. Marinescu, A. G. Rusu, C. Burileanu, and D. Bica, "Simultaneous speech detection based on MFCC-DTW with two-stage normalization," in *Proc. 41st Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2018, pp. 1–5.

[18] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 578–589, Oct. 1994.

[19] R. Visalakshi and P. Dhanalakshmi, "Performance of speaker identification using CSM and TM," *Int. J. Speech Technol.*, vol. 19, no. 3, pp. 457–465, Sep. 2016.

[20] D. Nath and S. K. Kalita, "Composite feature selection method based on spoken word and speaker recognition," *Int. J. Comput. Appl.*, vol. 121, no. 8, pp. 18–23, Jan. 2015.

[21] D. K. Ghose and S. Samantaray, "Modelling sediment concentration using back propagation neural network and regression coupled with genetic algorithm," *Procedia Comput. Sci.*, vol. 125, pp. 85–92, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050917327771

[22] O. Krestinskaya, K. N. Salama, and A. P. James, "Analog backpropagation learning circuits for memristive crossbar neural networks," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[23] V. Ranganathan and S. Natarajan, "A new backpropagation algorithm without gradient descent," *CoRR*, pp. 1–15, Jan. 2018. [Online]. Available: https://arxiv.org/abs/1802.00027

[24] J. Dou, H. Yamagishi, Z. Zhu, A. P. Yunus, and C. W. Chen, "A comparative study of the binary logistic regression BLR and artificial neural network ANN models for GIS-based spatial predicting landslides at a regional scale," in *Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools: Fundamentals, Mapping and Monitoring*, vol. 1. Cham, Switzerland: Springer, 2018, pp. 139–151.

[25] I. Hubara, E. Hoffer, and D. Soudry, "Quantized back-propagation: Training binarized neural networks with quantized gradients," in *Proc. ICLR Workshop*, Feb. 2018, pp. 1–4. [Online]. Available: https://openreview.net/forum?id=Bye10KkwG

[26] M. D. Gregorio and M. Giordano, "An experimental evaluation of weightless neural networks for multi-class classification," *Appl. Soft Comput.*, vol. 72, pp. 338–354, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S156849461830440X

[27] R. Jayakrishnan, G. N. Gopal, and M. S. Santhikrishna, "Multi-class emotion detection and annotation in malayalam novels," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2018, pp. 1–5.

[28] J. Xu, X. Liu, Z. Huo, C. Deng, F. Nie, and H. Huang, "Multi-class support vector machine via maximizing multi-class margins," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2017, pp. 3154–3160.

[29] Z.-L. Zhang, X.-G. Luo, and S. García, and F. Herrera, "Cost-sensitive back-propagation neural networks with binarization techniques in addressing multi-class problems and non-competent classifiers," *Appl. Soft Comput.*, vol. 56, pp. 357–367, Jul. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494617301400

[30] A. Fernández, V. López, M. Galar, M. J. D. Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowl.-Based Syst.*, vol. 42, pp. 97–110, Apr. 2013.

[31] R. Anand, K. Mehrotra, C. K. Mohan, and S. Ranka, "Efficient classification for multiclass problems using modular neural networks," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 117–124, Jan. 1995.

[32] D. Price, S. Knerr, L. Personnaz, and G. Dreyfus, "Pairwise neural network classifiers with probabilistic outputs," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, G. Tesauro, D. S. Touretzky, and T. K. Leen, Eds. Cambridge, MA, USA: MIT Press, Jan. 1994, pp. 1109–1116.

[33] C. K. Aridas, S.-A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Random resampling in the one-versus-all strategy for handling multi-class problems," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, G. Boracchi, L. Iliadis, C. Jayne, and A. Likas, Eds. Cham, Switzerland: Springer, 2017, pp. 111–121.

[34] Z.-L. Zhang, X.-G. Luo, S. García, J.-F. Tang, and F. Herrera, "Exploring the effectiveness of dynamic ensemble selection in the one-versus-one scheme," *Knowledge-Based Syst.*, vol. 125, pp. 53–63, Jun. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705117301557

[35] H. Liu *et al.*, "Action understanding based on a combination of one-versus-rest and one-versus-one multi-classification methods," in *Proc. 10th Int. Congr. Image Signal Process. Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–5.

[36] B. Krawczyk, M. Galar, M. Woźniak, H. Bustince, and F. Herrera, "Dynamic ensemble selection for multi-class classification with one-class classifiers," *Pattern Recognit.*, vol. 83, pp. 34–51, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320318301857

[37] A. Rocha and S. K. Goldenstein, "Multiclass from binary: Expanding one-versus-all, one-versus-one and ECOC-based approaches," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 289–302, Feb. 2014.

[38] P. Heracleous, J. Even, F. Sugaya, M. Hashimoto, and A. Yoneyama, "Exploiting alternative acoustic sensors for improved noise robustness in speech communication," *Pattern Recognit. Lett.*, vol. 112, pp. 191–197, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865518303040

[39] M. E. Ayadi, A.-K. S. Hassan, A. Abdel-Naby, and O. A. Elgendy, "Text-independent speaker identification using robust statistics estimation," *Speech Commun.*, vol. 92, pp. 52–63, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016763931630231X
[40] B. Yegnanarayana, A. Shahina, and M. R. Kesheorey, "Throat microphone signal for speaker recognition," in *Proc. INTERSPEECH, 8th Int. Conf. Spoken Lang. Process. (ICSLP)*, Jeju Island, South Korea, Oct. 2004, pp. 1–4.
[41] S. Roucos, V. Viswanathan, C. Henry, and R. Schwartz, "Word recognition using multisensor speech input in high ambient noise," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, Apr. 1986, pp. 737–740.
[42] S. Dupont, C. Ris, and D. Bachelart, "Combined use of close-talk and throat microphones for improved speech recognition under non-stationary background noise," in *Proc. Workshop (ITRW) Robustness Issues Conversational Interact.*, Norwich, U.K., Oct. 2004, pp. 1–4.
[43] M. A. T. Turan and E. Erzin, "Source and filter estimation for throat-microphone speech enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 2, pp. 265–275, Feb. 2016.
[44] N. Radha, A. Shahina, G. Vinoth, and A. N. Khan, "Improving recognition of syllabic units of Hindi languagae using combined features of throat microphone and normal microphone speech," in *Proc. IEEE Int. Conf. Control, Instrum. Commun. Comput. Technol. (ICCICCT)*, Jul. 2014, pp. 1343–1348.
[45] M. Sahidullah *et al.*, "Robust voice liveness detection and speaker verification using throat microphones," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 1, pp. 44–56, Jan. 2018.
[46] S. S. Upadhya, A. N. Cheeran, and J. H. Nirmal, "Thomson multitaper MFCC and PLP voice features for early detection of Parkinson disease," *Biomed. Signal Process. Control*, vol. 46, pp. 293–301, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809418301897
[47] N. Dave, "Feature extraction methods LPC, PLP and MFCC in speech recognition," *Int. J. Adv. Res. Eng. Technol.*, vol. 1, no. 6, pp. 1–4, Jul. 2013.
[48] R. Kumari and S. K. Srivastava, "Machine learning: A review on binary classification," *Int. J. Comput. Appl.*, vol. 160, no. 7, pp. 11–15, 2017.
[49] A. C. Lorena, A. C. P. L. F. D. Carvalho, and J. M. P. Gama, "A review on the combination of binary classifiers in multiclass problems," *Artif. Intell. Rev.*, vol. 30, nos. 1–4, pp. 19–37, Dec. 2008.
[50] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, Aug. 2011.
[51] G. Ou and Y. L. Murphey, "Multi-class pattern classification using neural networks," *Pattern Recognit.*, vol. 40, no. 1, pp. 4–18, Jan. 2007.
[52] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: A stepwise procedure for building and training a neural network," in *Neurocomputing, Algorithms, Architectures Applications* (NATO ASI Series), F. F. Soulié and J. Hérault, Eds. New York, NY, USA: Springer-Verlag, 1990, pp. 41–50.
[53] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, Sep. 2001.
[54] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, vol. 118. Piscataway, NJ, USA: IEEE Press, 2000.

FÁBIO CISNE RIBEIRO received the degree in electronic engineering from the University of Fortaleza, Brazil, in 2004, and the M.Sc. degree in teleinformatics engineering, in the field of computer vision, from the Federal University of Ceará, in 2008, where he is currently pursuing the Ph.D. degree. He has experience in the research fields of audio processing, speech recognition, embedded systems, pattern recognition, artificial intelligence, intelligent systems, computer vision, image processing, and industrial automation.

RAPHAEL TORRES SANTOS CARVALHO received the degree in teleinformatics enginerring from the Federal University of Ceará (UFC) in 2009, the master's degree in teleinformatics engineering, in the field of biomedical engineering, from UFC in 2012, and the M.B.A. degree in project management from the Christus University Center (Unichristus) in 2017. He is currently a Professor with the Federal Institute of Education, Science and Technology of Ceará. He has experience in the research fields of audio processing, speech recognition, signal processing, artificial intelligence, intelligent systems, pattern recognition, biomedical signal processing, embedded systems, and computer networks.

PAULO CÉSAR CORTEZ received the B.Sc. degree in electrical engineering from the Federal University of Ceará, Brazil, in 1982, and the M.Sc. and Ph.D. degrees in electrical engineering from the Federal University of Paraíba, in 1992 and 1996, respectively. He is currently a Full Professor at the Department of Teleinformatics Engineering, Federal University of Ceará. His fields of interest include image and signal analysis, computer vision, biomedical signal processing, and biomedical systems.

VICTOR HUGO C. DE ALBUQUERQUE received the degree in mechatronics technology from the Federal Center of Technological Education of Ceará in 2006, the M.Sc. degree in teleinformatics engineering from the Federal University of Ceará in 2007, and the Ph.D. degree in mechanical engineering with emphasis on materials from the Federal University of Paraíba in 2010. He is currently an Assistant VI Professor of the Graduate Program in Applied Informatics, University of Fortaleza. He has experience in computer systems, mainly in the research fields of applied computing, intelligent systems, visualization and interaction, with specific interest in pattern recognition, artificial intelligence, image processing and analysis, and automation with respect to biological signal/image processing, image segmentation, biomedical circuits, and human/brain–machine interaction, including augmented and virtual reality simulation modeling for animals and humans. Additionally, he has research at the microstructural characterization field through the combination of non-destructive techniques with signal/image processing and analysis, and pattern recognition.

PEDRO PEDROSA REBOUÇAS FILHO received the degree in mechatronics engineering from the Federal Institute of Ceará, Brazil, in 2008, the M.Sc. degree in teleinformatics engineering, in the field of biomedical engineering, and the Ph.D. degree in teleinformatics engineering from the Federal University of Ceará in 2010 and 2013, respectively. He has been an Assistant Professor with the Federal Institute of Ceará since 2008. From 2015 to 2016, he was a Post-Doctoral Researcher at the University of Porto, Portugal. He has co-authored of over 100 articles in national and international journals and conferences. Also, he has been involved in several research projects in the field of computer vision, medical image, and embedded systems.