Kernel SVM Classifiers based on Fractal Analysis for Estimation of Hearing Loss

Mohamed Djemai and Mhania Guerti

Abstract—Hearing screening consists of analyzing the hearing capacity of an individual, regardless of age. It identifies serious hearing problems, degree, type and cause of the hearing loss and the needs of the person to propose a solution. Auditory evoked potentials (AEPs) which are detected on the EEG auditory cortex area are very small signals in response to a sound stimulus (or electric) from the inner ear to the primary auditory areas of the brain. AEPs are noninvasive methods used to detect hearing disorders and to estimate hearing thresholds level. In this paper, due to the nonlinear characteristics of EEG, Detrended Fluctuation Analysis (DFA) is used to characterize the irregularity or complexity of EEG signals by calculating the Fractal Dimension (FD) from the recorded AEP signals of the impaired hearing and the normal subjects. This is to estimate their hearing threshold. In order to classify both groups, hearing-impaired and normal persons, support vector machine (SVM) is used. For comparably evaluating the performance of SVM classifier, three kernel functions: linear, radial basis function (RBF) and polynomial are employed to distinguish normal and the abnormal hearing subjects. Grid search technique is selected to estimate the optimal kernel parameters. Our results indicate that the RBF kernel SVM classifier is promising; it is able to obtain a high training as well as testing classification accuracy.

Keywords—Auditory evoked potentials, Hearing Thresholds, Detrended Fluctuation Analysis, Grid search, Support Vector Machine.

I. INTRODUCTION

Hearing loss is a pathological state of hearing characterized by a partial or total loss or even early or late loss of sound perception. The consequences of hearing impairment include miscommunication and psychological problems [1]. In the absence of corrective measures, there could be 630 million people with hearing loss by 2030 and almost 900 million by 2050 [2].

Hearing screening consists of analyzing the hearing capacity of an individual, regardless of age. It identifies serious hearing problems, degree, type and cause of the hearing loss and the needs of the person to propose a solution. AEP signal is the screening technique used to distinguish between pathological and healthy cases. AEP signals are a recording of a subject's electroencephalogram (EEG) from the auditory pathways leading sound from the inner ear to the primary auditory areas of the brain in response to a short auditory stimulus. AEPs are an objective tool for assessing hearing function, used to identify potential problems in the auditory neural pathway and to estimate hearing thresholds [3].

Fractals are mathematical objects used to describe natural phenomena such as clouds, branches of trees, rocky coasts, leaves, the bronchi of our lungs that present a certain irregularity or roughness, in the face of which Euclidean geometry does not allow correctly describe this irregularity. Fractals have a character called self-similarity or scale invariance. Several natural phenomena have longer-term time dependencies: the correlations in the series remain in a durable way. These properties indicate the presence of a fractal structure. It is thanks to the introduction of the theory of fractals (from the Latin fractus: irregular, interrupted) by Benoît Mandelbrot in the 1970s that a new description of these complex objects could be established [4]. The fractal dimension (FD) is a non-integer number; it can be a fraction, an irrational number or a whole number which measures the degree of irregularity of an object and also makes it possible to quantify the notion of self-similarity or the degree of fluctuation of time series.

DFA is prominent method to quantify the fractal-scaling index of time series. In this study, DFA method is employed to estimate the FD of hearing impaired and normal subjects.

I. RELATED WORK

A. Detrended Fluctuation Analysis (DFA)

DFA is a method specially designed for the analysis of signals possessing the property of self-similarity [5] and detection of long-term correlations in non-stationary time series. It is the most frequently used method in different areas due to its simple construction and to its excellent results [6] such as: DNA sequencing, Study of heart rate variability, long-time weather records, structure clouds, geology, ethnology, economic time series and solid state physics [7]. Ivan Seleznov et al. applied DFA method to identify activation changes in brain dynamics...
during mental computations, which reveals a permanent information communication during brain activity [8]. Jing et al. [9] used the DFA to assess the temporal correlation properties of the EEG in drug dependence. The results obtained confirmed the effect of the drug-related stimulus on the EEG scaling behavior.

We compute the integrated series from each data point:
\[ y(k) = \sum_{k=1}^{K} (x(k) - \bar{x}), \]
where \( \bar{x} \) is the average of the global signal.

The integrated series \( y(k) \) is then divided into equal-sized, non-overlapping boxes of length \( n \). In each box, the linear fit is calculated by using least squares, the resulting vector, \( y_n(k) \), is then subtracted to \( y(k) \) as follows:
\[ F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y(k) - y_n(k))^2}, \]
(2)
The above computation will be repeated for segment sizes \( n \) (different scales) to give a relationship between \( F(n) \) and \( n \). Typically, \( F(n) \) will increase linearly with segment size \( n \). The slope of the line \( F(n) \) determines the scaling \( \alpha \) exponent [10]. FD = 3 - \( \alpha \).

B. Support Vector Machine

SVMs are a class of supervised learning algorithms designed to solve discrimination and regression or anomaly detection problems. It seeks to find among an infinity of linear classifiers (hyperplanes), the optimal hyperplane which separates the data into two different classes by following the criterion of maximum margin.

The margin is the distance between the separation boundary and the observations closest to the hyperplane (support vectors) [11].

SVM is the most widely used technique due to its better generalization performance and capability to work well in higher dimensional space [12]. It has better learning ability and smaller test errors than other methods for different datasets [13].

B.1. Separable Data

For two-classes, separable training data sets, there is an equation hyperplane \( w \cdot x + b = 0 \) such as:
\[
\begin{align*}
& \{w \cdot x_i + b \geq 1 \text{ for } y_i = +1 \\
& \{w \cdot x_i + b \leq -1 \text{ for } y_i = -1 
\end{align*}
\] (3)
The last two constraints can be combined into:
\[ y_i (w \cdot x_i + b) \geq +1. \]
(4)
The optimal separating hyper plane can be obtained by solving the following optimization problem:
\[
\text{min} \frac{1}{2} w^T w \text{ subject to } y_i (w \cdot x_i + b) \geq +1.
\]
(5)
The problem is a quadratic optimization problem; the Lagrangian associated to the problem becomes:
\[
L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{m} \alpha_i (y_i (w \cdot x_i + b) - 1),
\]
(6)
Where \( \alpha_i \geq 0 \) are the Lagrange multipliers.

By differentiating \( L \) with respect to \( w \) and \( b \), \( L \) is converted into a dual Lagrangian \( L_D(\alpha) \)
\[
\max_{\alpha} L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i \cdot x_j,
\]
(7)
s.t \( : i = 1, ..., m \alpha_i \geq 0 \) and \( \sum_{i=1}^{m} \alpha_i y_i = 0 \),

\( L_D(\alpha) \) should be maximized with respect to \( \alpha_i \) to obtain the best hyperplane. The \( \alpha_i \) vector can be calculated also as a quadratic optimization problem.

B.2. Non-separable Data

In the nonlinearly separable case, SVM is modified by introducing slack variables \( \xi_i \) for measuring classification errors with \( \xi_i \geq 0 \) for \( i = 1, ..., m \).

Therefore, we can write the optimization problem as:
\[
\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i,
\]
(8)
s.t. \( y_i (w \cdot x_i + b) + \xi_i \geq +1, \quad \xi_i \geq 0. \)

The parameter \( C > 0 \) controls the tradeoff between increasing the margin and reducing the errors. The Lagrange multipliers \( \alpha_i \) should be employed to solve the Optimization problem Eq. (8) that transforms it to dual form.

In the linearly separable case, a non-linear vector mapping function \( \phi \) should be used to transform the data to a higher-dimensional feature space. This process is done through the kernel function, presumably making the separation easier in that space [14].

The optimal hyperplane can be obtained by solving:
\[
\max_{\alpha} L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j),
\]
(9)
s.t \( 0 \leq \alpha_i \leq C, i = 1, ..., m \) and \( \sum_{i=1}^{m} \alpha_i y_i = 0 \),

where \( \alpha = (\alpha_1, \alpha_2, ..., \alpha_m) \) is the vector of non negative Lagrange multipliers associated with the constraints \( \sum_{i=1}^{m} \alpha_i y_i = 0 \) and \( 0 \leq \alpha_i \leq C, i = 1, ..., m \). Three commonly used kernel functions are listed below:

The Gaussian radial basis function (RBF): \( K(x_i, x_j) = \exp -\frac{|x_i - x_j|^2}{2\sigma^2} \).

The polynomial kernel: \( K(x_i, x_j) = (x_i, x_j + 1)^d \);

The linear kernel: \( K(x_i, x_j) = x_i, x_j \).

Here, \( \sigma \) and \( d \) are kernel parameters. \( \sigma \) is the spread of the Gaussian function.

The order of the polynomial kernel \( d \) controls the flexibility of SVM model and affects its accuracy [15].

C. Grid Search

There are many optimization Kernel parameters methods for SVM such as: particle swarm optimization algorithm, genetic algorithm and Grid Search.

Huang et al. [16], used wavelet features and traditional spectral features as input features to construct fusarium head blight
detection models in combination with the particle swarm optimization support vector machines (PSO-SVM) approach. PSO is applied to simultaneously optimize both the penalty parameter \( c \) and the radial basis function parameter \( \gamma \). The results show that the PSO-SVM detection method yielded a higher overall accuracy compared to the back propagation neural network (BPN) detection method.

Djemai et al [17] developed a hybrid approach in which they combined the genetic algorithm and SVM model to classify normal and abnormal hearing subjects using FD features that were extracted from the AEP responses of the subjects applying DFA algorithm. This combination leads to achieve considerably higher performance compared to standard SVM.

In this study, Grid Search is chosen because it straightforward hyperparameter tuning method [18]. It is easier to implement and it is that it cannot get stuck in local maximum. The Grid Search method is a technique whose goal is to methodically determine the best combination of hyperparameters over the designated range that can facilitate the construction of a model in a given set. It consists of exploring all the possible combinations of values on a set of models that differ from each other in their parameter values, which are on a grid seeking the best combination that has achieved the best performance score on the test data. X. Wang et al. developed a hybrid algorithm based on SVM and the search grid method to make better use of soil salinization information [19]. F. Budiman applied the search grid method to analyze and test the optimization range of SVM-RBF kernel parameter values to recognize the image possessing geometric decorative motifs [20].

D. K-fold Cross Validation (KCV)

Cross-validation (CV) is an approach employed to evaluate the efficiency of machine learning models. It is used in modeling based on data that has properties such as: complexity, distribution, correlation between variables, etc [21].

The KCV randomly splits the training data \( T \) into \( k \) equal groups. One fold is going to be used as the validation set, and the rest \( T_1 \ldots T_k, T_{k-1} \) are for the training set. This procedure should be done \( k \) times but using a different fold for the validation set.

II. MATERIALS AND METHODS

A. Dataset

Twenty participants, ages 15 to 29 (15 men, 5 women), took part in the experiment, divided into two groups: the normal hearing group consisting of ten participants between the ages of 21 and 28; The hearing-impaired group consists of five females and five males, ages 15 to 29. Participants in both groups are in good health and are not taking any medicines. First, the hearing threshold \( t_h \) is measured for all participants using screening pure-tone audiometry [10].

- Normal hearing was defined as having \( t_h \leq 20 \text{ dB} \).
- Hearing impaired was defined as having \( t_h > 20 \text{ dB} \).

B. Data analysis

EEG signals were recorded using 10-20 electrode positioning system with 19 electrodes [10]. Using the headphones, the participants were prepared to perceive click sound at 4000 Hz, 2000 Hz, 1000 Hz and 500 Hz with an intensity of 20 dBHL in the right and the left ear. We record the AEP signals with a sampling frequency of 256 Hz. This procedure should be done for five trials and the participants are given a one-minute rest period between trials.

For all participant, FD was estimate from AEP signals obtained from 19 channels by DFA method [10]. We obtained the database that was used in this experiment by acoustic research lab, University Malaysia Perlis. For comparably evaluating the efficiency of linear kernel SVM classifier, polynomial kernel SVM classifier and RBF kernel SVM classifier, we perform experiments to distinguish between the normal hearing subjects and the hearing impaired subjects using FD vectors obtained from the subject's recorded AEP.

In this work, 5-fold CV technique was selected to estimate the competence of the SVM classifiers. The results obtained will be optimized by grid search method. Grid search technique is selected to estimate the optimal parameters for three kernel functions: linear, radial basis function (RBF) and polynomial, in which the values of the parameters are changed over the selected parameter ranges with fixed step sizes, and the efficiency of all groups of parameters is measured and compared. The range of \( C, \sigma \) and \( d \) values tested is as follows:

For the linear and RBF kernel SVM classifiers, The range of \( C \) is from \( 2^{-5} \) to \( 2^{10} \), increasing in steps of the power of two; The Gaussian width \( \sigma \) for the RBF kernel ranged from \( 2^{-1} \) to \( 2^{5} \), increasing in steps of the power of two; The order of the polynomial kernel \( d \) ranged from 1 to 10 by integers.

The model with the optimized parameters has the highest classification rate.

Fig. 2 shows the flowchart of SVM classifier model using grid search.

![Flowchart of SVM classifier model using grid search](image-url)
III. RESULTS AND DISCUSSION

The linear kernel SVM classifier yields accuracy equal to 82.86% with parameter C worth $2^5$. As shown in Fig. 2.

Fig. 2: Linear kernel parameter optimization using grid search

Figures 3 shows the parameters optimization results of polynomial kernel SVM classifier, using grid search method. The best classification accuracy of 85.71% achieves at $C=2^{-2}$ and $d=2$.

Fig. 3: Polynomial kernel parameters optimization using grid search

As shown in Fig. 4, Using grid search method, find out the optimal values of $\sigma$ and margin parameter $C$. Which gives $\sigma = 2^{-1}$ and $C=2^2$ with the classification accuracy 95.82%.

Fig. 4: RBF kernel parameters optimization using grid search

The classification rates of linear kernel SVM classifier, RBF kernel SVM classifier and polynomial kernel SVM classifier and are given in Table I.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Ear</th>
<th>Linear SVM</th>
<th>Polynomial SVM</th>
<th>RBF-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>R</td>
<td>84.16</td>
<td>81.73</td>
<td>96.40</td>
</tr>
<tr>
<td>1000</td>
<td>R</td>
<td>76.23</td>
<td>72.13</td>
<td>89.43</td>
</tr>
<tr>
<td>2000</td>
<td>R</td>
<td>78.54</td>
<td>74.33</td>
<td>90.70</td>
</tr>
<tr>
<td>4000</td>
<td>R</td>
<td>81.09</td>
<td>77.43</td>
<td>89.04</td>
</tr>
<tr>
<td>500</td>
<td>L</td>
<td>82.30</td>
<td>78.57</td>
<td>92.46</td>
</tr>
<tr>
<td>1000</td>
<td>L</td>
<td>90.24</td>
<td>88.20</td>
<td>92.66</td>
</tr>
<tr>
<td>2000</td>
<td>L</td>
<td>83.06</td>
<td>80.13</td>
<td>93.30</td>
</tr>
<tr>
<td>4000</td>
<td>L</td>
<td>84.59</td>
<td>83.70</td>
<td>93.49</td>
</tr>
</tbody>
</table>

For a frequency of 500 Hz for the right ear, the classification accuracy obtained from the comparison result indicates that grid search method could to choose the most appropriate $\sigma$ and $C$ parameters among a big set of parameters, which allowed RBF kernel SVM algorithm to obtain the highest training and testing classification accuracy of 96.40% and 94.13% respectively. For a frequency of 4000 Hz for the left ear, RBF kernel SVM classifier has the maximum training and testing classification accuracy of 93.49 % and 91.30% respectively.

For a frequency of 500 Hz for the right ear, the classification accuracy obtained from the comparison result indicates that grid search method could to choose the most appropriate $\sigma$ and $C$ parameters among a big set of parameters, which allowed RBF kernel SVM algorithm to obtain the highest training and testing classification accuracy of 96.40% and 94.13% respectively. For a frequency of 4000 Hz for the left ear, RBF kernel SVM classifier has the maximum training and testing classification accuracy of 93.49 % and 91.30% respectively.
In this paper, we have developed three SVM classifier models including RBF kernel, polynomial and linear kernels to distinguish the normal hearing group and the abnormal hearing group using FD features obtained from the subject's recorded AEP signals applying DFA method. The AEP signal is stimulated at four distinct frequency in the right and left ear at a fixed sound intensity level. Grid search is selected to optimize the kernel parameters.

Through the results obtained, it can be considered that the FD extracted from the AEP signals is an appropriate parameter for estimating the threshold of perception for hearing. It is also considered that DFA method more efficient and faster to find this parameter. From the results obtained in this experiment, for hearing impaired persons, the values of the FD are relatively high compared to normal hearing ones due to the longer response time to the stimulus.

Our results indicate that the grid search was able to find near optimal parameter combination within given ranges. RBF kernel SVM classifier is promising; it is able to obtain a high training as well as testing classification accuracy and to achieve considerably higher performance compared to linear SVM classifier and polynomial SVM classifier.

IV. CONCLUSION

In this paper, we have developed three SVM classifier models including RBF kernel, polynomial and linear kernels to distinguish the normal hearing group and the abnormal hearing group using FD features obtained from the subject's recorded AEP signals applying DFA method. The AEP signal is stimulated at four distinct frequency in the right and left ear at a fixed sound intensity level. Grid search is selected to optimize the kernel parameters.

Through the results obtained, it can be considered that the FD extracted from the AEP signals is an appropriate parameter for estimating the threshold of perception for hearing. It is also considered that DFA method more efficient and faster to find this parameter. From the results obtained in this experiment, for hearing impaired persons, the values of the FD are relatively high compared to normal hearing ones due to the longer response time to the stimulus.

Our results indicate that the grid search was able to find near optimal parameter combination within given ranges. RBF kernel SVM classifier is promising; it is able to obtain a high training as well as testing classification accuracy and to achieve considerably higher performance compared to linear SVM classifier and polynomial SVM classifier.

REFERENCES

Mohamed Djemai received the Engineer Diploma degree from Ziane Achour University, Djelfa, Algeria in 2006 and the Magister degree in Electronic engineering from Ecole Nationale Polytechnique, Algiers, Algeria, in 2011. He is currently a teacher at Ziane Achor university, Djelfa, Algeria. His research fields are speech processing, automatic classification and feature selection.

Mhania Guerti received her MSc in Language and linguistic communication sciences from the ILP Algiers, Algeria in collaboration with the CNET Lannion (France), in 1984, and Ph.D. from ICPINPG, Grenoble, France, in 1993. She is an Ex-Full Professor in the Department of Electronics at Ecole Nationale Polytechnique. Her research fields are speech processing, acoustics and audiovisual systems.