Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Improved Methodology of SVM to Classify Acoustic Signal by Spectral Centroid

S. Kavitha[1], J. Manikandan[2]

[1]Professor, Department of ECE, Hindusthan Institute of Technology, Coimbatore, Tamil Nadu, India

[2]Professor Department of Mechanical Engineering, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

E-mail: [1]kavithamani2003@gmail.com, [2]manikandan.hcet@gmail.com

Abstract

Acoustic signal classification issues are addressed in this work using spectral examination, channel extracting the features from the input and machine learning algorithm. This brief article examines the effect of various settings on feature extraction. This feature-level channel combination's accuracy increase is then observed. To categorise things, pattern recognition utilises a variety of classification schemes. "Pattern" refers to the measures that must be categorised with accurate feature extracted. Images and audio signals are among the most common kinds of measurements. The proposed Support Vector Machine (SVM) is used for the necessity of an effective categorization of acoustic signals driven by the continual improvements in multimedia technology. This study uses two machine learning algorithms to enhance audio classification and categorization. The proposed SVM achieves superior performance than the other ML algorithm by spectral features.

Keywords: Machine Learning (ML), spectral analysis, SVM, audio classification, Spectral Centroid feature

1. Introduction

Comparatively, research in audio classification and retrieval is very recent compared to other closely related disciplines like speech recognition and speaker identification. In spite of this, audio categorization is a vital component in the present field of audio processing and content analysis. As a rule of thumb, audio categorization is a problem of pattern recognition. Classifying items into several categories or classes is the purpose of pattern recognition, which is a branch of science. In certain cases, these items might be photographs, signal waveforms, or any other measures that need organisation [1-5].

DOI: https://doi.org/10.36548/jtcsst.2021.4.005
Received: 17.03.2022, received in revised form: 16.04.2022, accepted: 01.05.2022, published: 10.05.2022

Data may be mined for information using machine learning. The autonomous identification of the any group category or pattern is based on the given dataset by machine learning research. Design and development of algorithmic systems that enable computers to learn and adapt depending on data are at the heart of this scientific field's focus.

1.1 Classification of Audio Signals

To categorise an audio signal, characteristics are extracted from the sound and applied in order to determine its class. Pattern recognition utilises a variety of classification schemes to categorise. Analyzing and identifying patterns in data is the goal of pattern recognition. Techniques like character recognition, voice analysis, picture analysis and medical diagnostics are all crucial to the development and use of these technologies. Thus, categorization is the primary goal of pattern recognition. A pattern-recognition system can be considered as a two-step gadget. The first step is the extraction of features, and the next is the categorization of those features [6-9].

1.1.1 Supervised Learning

When training instances are labelled with the proper outcome, students undergo supervised learning, which provides immediate feedback on their progress in learning. In this instance, the classes to which the training samples belong are already known.

1.1.2 Unsupervised learning

Unsupervised learning produces no error signals since it does not provide a desired outcome. It refers to the challenge of discovering hidden patterns in unlabelled data. When learning a new algorithm, similar-type input vectors are clustered together.

Multimedia technology improvements necessitate the need for effective categorization of audio signals in order to improve the accuracy and accessibility of content-based retrieval from large databases. In many cases, audio information plays a crucial role in determining the meaning of multimedia. Increasing amounts of data need the employment of automated systems for filtering, processing, and storing it. For video indexing and content analysis, audio data may be a crucial source of information. An audio signal's weak or noisy sound makes it difficult for the listener to discriminate between various kinds of sound. This issue, on the other hand, has been very difficult to solve with computers. Automatic categorization of audio signals is a major difficulty in this discipline. Media services, search engines, and

intelligent human-computer systems are all interested in improving their ability to classify audio signals [10 - 13]. Audio categorization is necessary for the following reasons:

- There should be a range of processing options for various kinds of audio.
- A single subclass is narrowed down during the retrieval procedure after classification.

Each bit of audio should be processed and indexed separately, so that it may be instantly compared and discovered.

2. Literature Survey

An assigning object to a certain class or category is an important data mining function. Individual data objects are grouped into distinct categories based on previous knowledge in classification, which is an example of a supervised learning technique. The classifier's performance is heavily influenced by the data's attributes [14]. The researchers in the fields of data mining and machine learning usually focus on classification as a subject of particular interest. The properties of an item are used to construct a classification function or model. The training set consists of a predetermined number of items. In order to build a classification function or model, the link between the properties of the training set and the classes, is analysed. This model or function may then be used to categorise subsequent items. This enables us to better comprehend the database's rank structure [15].

3. Methodology

The feature extraction and classification are the two main components of an audio signal classification system. To which set of classes a sound is most likely to fit is identified by extracting important information from the sound. It's important to understand the exact sector in which the classifier will be used before designing features. Classification algorithms, on the other hand, treat input data as if it were a collection of random integers.

3.1 Infrared and Visible Spectra

This infrared part of magnetic spectral density has wider wavelength with medium frequency than visible spectra that covers one. There are several methods covered, but absorption spectroscopy is the main one [16]. Molecules absorb certain frequencies according to their structure, which is why infrared spectroscopy takes use of this. In these absorptions,

the frequency of the absorbed radiation matches that of the bond or group that is vibrating, and hence called resonant. There are several limitations to current infrared spectroscopy methods that demand the use of infrared-optimized optical equipment. In this article, how to produce infrared spectral measurements is demonstrated using spectral characteristics that can be seen through power spectrum [17 – 19].

3.1.1 Amount of Time it takes to reach zero

The measures of frequency content of a signal in an easy-to-understand way. An average zero crossing rate may be used to assess the frequency content of a narrowband signal's movement. Speech, on the other hand, is substantially less precise when transmitted across a wide area network. A short-time average zero-crossing rate may be used to derive the spectral characteristics. In order to calculate the average across N samples, each pair of samples is first verified for zero crossings.

3.1.2 Short-Time Energy

In a variety of audio classification difficulties, Short-Time Energy (STE) has been used. When it comes to identifying voiced and unvoiced speech, STE gives a framework. Second, STE may be used to identify high-quality speech from quiet.

3.1.3 Aspects of spectral centrifugal Force

In digital signal processing, the spectral centroid is used to describe a spectrum. It serves as a visual cue to the "centre of mass" of the frequency spectrum and is linked to the perceived brightness of the sound. A Fourier transform may also be used to find out how many frequencies there are in the signal, so the weighted average can be calculated.

In other words, it measures the sound's "centre of mass" or "brightness." For example, many musical settings have distinct centres of gravity. So, for every 2 second window, it computes a 1-dimensional spectral centroid feature. If many windows are utilised, a single feature vector for each recording is created by averaging the mean and standard deviations of the features taken from each window [20]. The raised classification accuracy is observed i.e., classification with spectral centroid improved the accuracy of classification around 70%.

3.1.4 Aspect Ratios of Flux

The spectral Flux is the average fluctuation in the spectrum value between neighbouring frames in a particular audio clip. Music, on the other hand, does not have a

syllable-rate structure like speech, but it does have a syllable-rate structure like speech. Environmental sounds have a greater range of spectral flux fluctuation than speech or music [21]. To identify environmental noises with significant periodicity, the acoustic property that distinguishes these sounds is the frequency modulation. It is also able to distinguish between music, speech, and environmental noises.

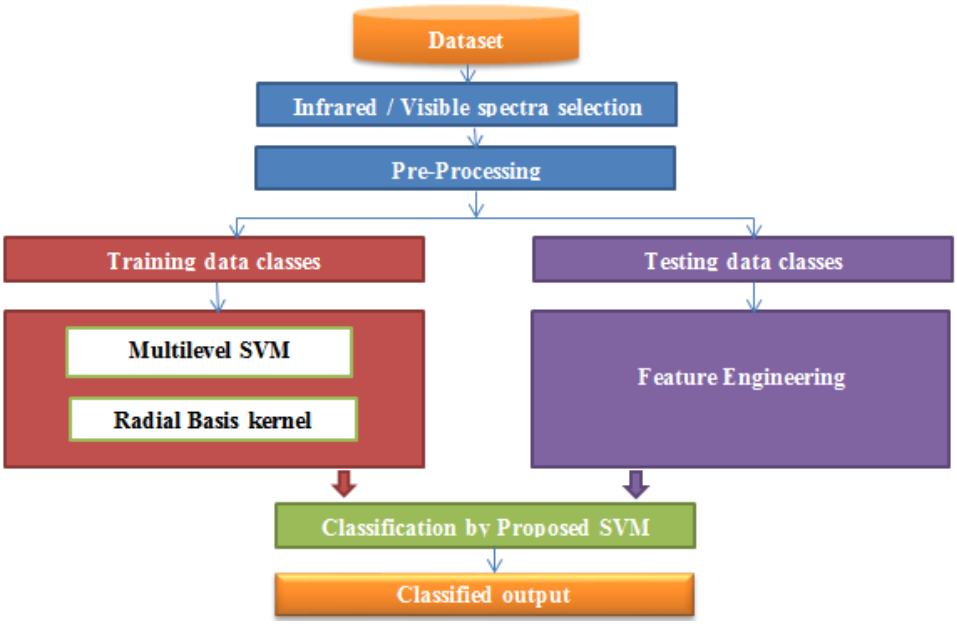

Figure 1. Block diagram of proposed SVM architecture

3.2 Proposed SVM Method

On a speaker verification job, this study tests the effectiveness of the support vector machine. It seems as though support vector machines, which can only make binary judgments, would be a good fit for speaker verification. Classifiers trained using many benchmark databases, which contains recordings of high-quality speech from cooperating speakers, performed comparable to those trained using an older method of normalising the polynomial kernel. The radial base function kernel's attributes are imposed on the polynomial kernel via the normalisation procedure. It was because to this strategy that good results are acquired using SVMs on this database.

3.3 Radial basis kernel with SVM

There are three degrees of kernels used to classify from the input signal as follows;

1. Quadratic
2. Cubic

3. Polynomial equation

These are used to do computations in any d-dimensional space; where d > 0.99. An endless polynomial equation may be generated by expanding an exponential signal. This kernel can be utilised to make the classification line indefinitely strong by employing exponents.

There are significant differences in the way acoustic sceneries and speech are structured. Even the characteristics of an acoustic environment vary from one kind to the next. Some settings include just low, middle, and high frequencies, whereas others have a wide range of frequencies.

3.4 Training / Testing of Dataset

The correct categorization of training instances is traded for a larger decision function margin with training parameter. An increase in the significant margin that results from using a training value, is less than one result in a simpler decision function at the expense of training accuracy. To put it in another way, training acts as an SVM regularisation parameter.

Both channels of the dataset's binaural recordings are extracted individually to take use of the extra cues they include. Then, for each recording, a single feature vector is obtained by feature-level channel combination. The system's detection accuracy can be improved by the use of components from both media. In order to classify a scene, the system uses a typical SVM-based technique with a radial base function kernel. Final feature vectors are modelled using SVM using the LIBSVM library [22].

4. Results and Discussion

The transformed training patterns serve as the support vectors, and both are located near to the hyperplane of separation during the sampling period. To determine the optimum hyperplane, the support vectors are used as training examples. Figure 2 shows the obtained results from the proposed SVM classifier.

The feature extraction approaches are provided a wave file to work with samples of original and determined. Once this is done, the wave file's spectral feature values may be determined accurately for further classification. Finally, the procedure outlined above is repeated until a total of 100 wave files have been generated for accurate classification. The power spectrum graph has been shown in the figure 3. The performance of the proposed

model is measured by many metrics such as accuracy, precision and sensitivity as tabulated in table 1.

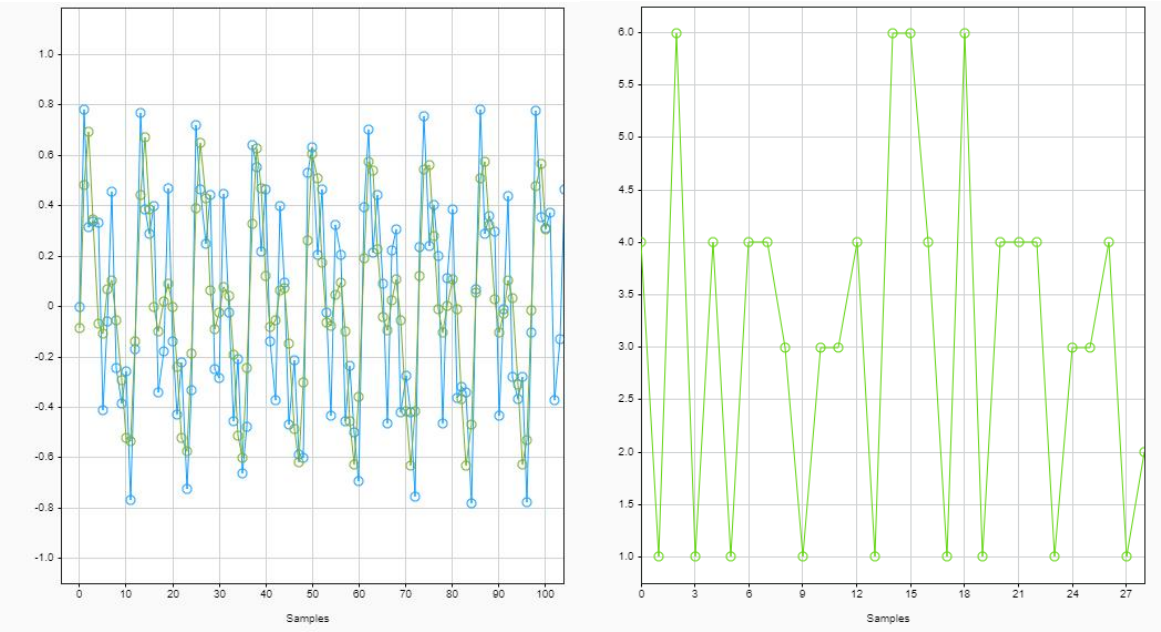

Figure 2 a) Obtained results from the proposed SVM classifier (Original vs Obtained)
b) Accurate predicted samples in the signal

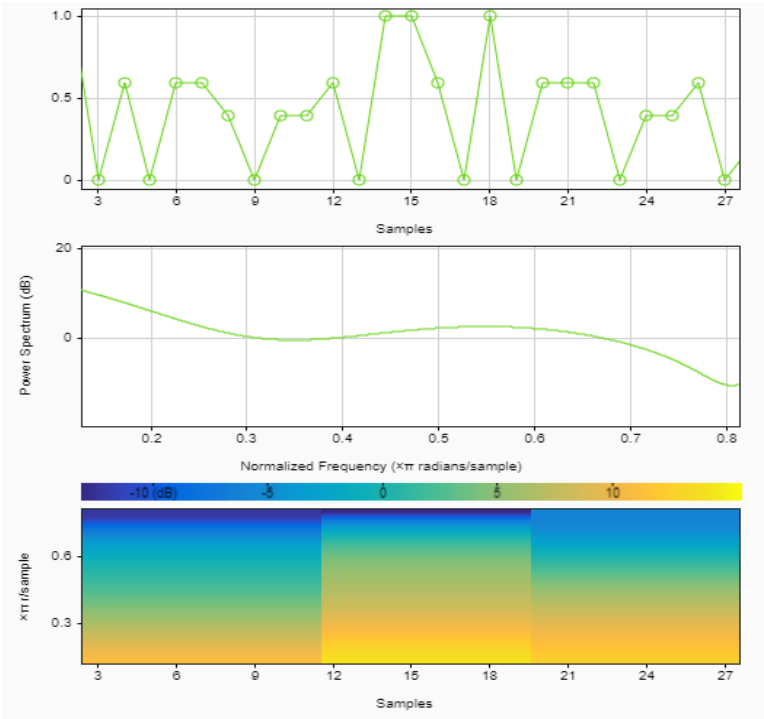

Figure 3. Power spectrums from obtained results

The proposed SVM has the greatest accuracy rate for jobs involving audio signal categorization. SVM's finest results may be described to the algorithm's fundamental nature.

For example, the SVM is a classification system with a goal of minimising the upper limit on its anticipated error.

Table 1. Obtained results from proposed SVM classifier

S.No	Model	Precision	Sensitivity	Identification Accuracy	Computation speed	Prediction error
1	Random Forest	83.64%	84.41%	85.18%	High	0.175
2	Naïve Bayes	81.56%	82.8%	84.04%	High	0.567
3	k-NN	88.67%	89.33%	89.99%	Low	0.073
4	Proposed SVM Model	93.78%	95.26%	96.67%	Moderate	0.013

Many hyper-planes may be possible, but only the one with the highest margin of error is considered the best. A larger margin means a smaller generalisation error to minimize it. From table 1, probability based classifiers fail in this research work than other machine learning algorithms.

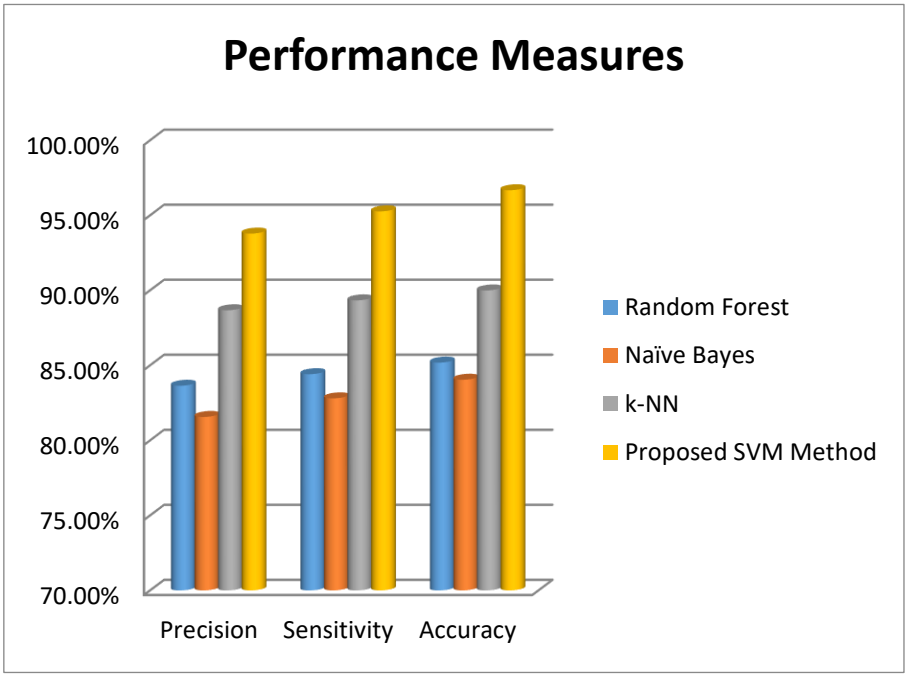

Figure 4. Overall performance measures between Machine learning algorithms

The proposed SVM looks for the hyperplane that best separates distinct types of data so that it may be applied to new data in the future. There is a hyperplane that maximises distance to the nearest point from each class, dubbed as the "maximum margin hyperplane".

5. Conclusion

In comparison to the existing ML techniques, the suggested SVM with radial basis function kernel produced good results with substantially less training data. If the classifier can recognise a scene properly after just a few frames, then future work will concentrate on early detection. When separate feature sets offer uncorrelated information, they may be integrated to make a joint conclusion. This is the inspiration for future work in audio signal categorization. Furthermore, the ML method may be taught and evaluated on acoustically imperfect data. For effective coding and signal amplification, this is a need because of the enormous dispersal of telecommunication networks. It is interesting to discover which functions are the most reliable. Achieving this aim, requires the exploration of other audio components. Therefore, for the sake of enhancing the accuracy and reliability of the audio categorization, additional audio classes are to be added to the classification scheme.

References

[1] Lim and Chang, “Enhancing Support Vector Machine-Based Speech/Music Classification using Conditional Maximum a Posteriori Criterion,” Signal Processing, IET, vol. 6, no. 4, pp. 335-340, 2012.

[2] Md. Al Mehedi Hasan and Shamim Ahmad. predSucc-Site: Lysine Succinylation Sites Prediction in Proteins by using Support Vector Machine and Resolving Data Imbalance Issue. International Journal of Computer Applications 182(15):8-13, September 2018.

[3] Gerazov, B.; Ivanovski, Z. Kernel Power Flow Orientation Coefficients for Noise Robust Speech Recognition. IEEE/ACM Trans. Audio Speech Lang. Process. 2015, 23, 407–419.

[4] Venkitaraman, A.; Adiga, A.; Seelamantula, C.S. Auditory-motivated Gammatone wavelet transform. Signal Process. 2014, 94, 608–619.

[5] Gerazov, B.; Ivanovski, Z. Gaussian Power flow Orientation Coefficients for noise-robust speech recognition. In Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014; pp. 1467–1471.

[6] Hend Ab. ELLaban, A A Ewees and Elsaeed E AbdElrazek. A Real-Time System for Facial Expression Recognition using Support Vector Machines and k-Nearest Neighbor

Classifier. International Journal of Computer Applications 159(8):23-29, February 2017.

[7] Kruspe, D. Zapf, and H. Lukashevich, "Automatic speech/music discrimination for broadcast signals," in INFORMATIK 2017, M. Eibl and M. Gaedke, Eds. Gesellschaft f ¨ur Informatik, Bonn, 2017, pp. 151–162.

[8] Pikrakis and S. Theodoridis, "Speech-music discrimination: A deep learning perspective," in 2014 22nd European Signal Processing Conference (EUSIPCO), Sept 2014, pp. 616–620.

[9] K. Khonglah and S. R. M. Prasanna, "Low frequency region of vocal tract information for speech / music classification," in 2016 IEEE Region 10 Conference (TENCON), Nov 2016, pp. 2593–2597.

[10] Lim and J. h. Chang, "Enhancing support vector machine-based speech/music classification using conditional maximum a posteriori criterion," IET Signal Processing, vol. 6, no. 4, pp. 335–340, June 2012.

[11] K. Khonglah and S. R. M. Prasanna, "Speech / music classification using vocal tract constriction aspect of speech," in 2015 Annual IEEE India Conference (INDICON), Dec 2015, pp. 1–6.

[12] Wang, J.C.; Lin, C.H.; Chen, B.W.; Tsai, M.K. Gabor-based nonuniform scale-frequency map for environmental sound classification in home automation. IEEE Trans. Autom. Sci. Eng. 2014, 11, 607–613.

[13] Palo, H.K.; Mohanty, M.N.; Chandra, M. Novel feature extraction technique for child emotion recognition. In Proceedings of the 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), Visakhapatnam, India, 24–25 January 2015; pp. 1–5.

[14] Zão, L.; Coelho, R.; Flandrin, P. Speech Enhancement with EMD and Hurst-Based Mode Selection. IEEE/ACM Trans. Audio Speech Lang. Process. 2014, 22, 899–911.

[15] Dranka, E.; Coelho, R.F. Robust Maximum Likelihood Acoustic Energy Based Source Localization in Correlated Noisy Sensing Environments. J. Sel. Top. Signal Process. 2015, 9, 259–267.

[16] Valero, X.; Alías, F. Gammatone Wavelet features for sound classification in surveillance applications. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012; pp. 1658–1662.

[17] Chungsoo Lim Mokpo, Yeon-Woo Lee, and Joon-Hyuk Chang, "New Techniques for Improving the practicality of a SVM-Based Speech/Music Classifier," IEEE

International Conference on Acoustics, Speech and Signal Processing, pp. 1657-1660, 2012.

[18] Theodorou, T., I. Mporas and N. Fakotakis, 2012. Automatic sound classification of radio broadcast news. Int. J. Signal Process. Image Process. Patt. Recogn., 5: 37-48.

[19] Md. Al Mehedi Hasan and Shamim Ahmad. predSucc-Site: Lysine Succinylation Sites Prediction in Proteins by using Support Vector Machine and Resolving Data Imbalance Issue. International Journal of Computer Applications 182(15):8-13, September 2018.

[20] Poonam Sharma and Anjali Garg. Feature Extraction and Recognition of Hindi Spoken Words using Neural Networks. International Journal of Computer Applications 142(7):12-17, May 2016.

[21] Hend Ab. ELLaban, A A Ewees and Elsaeed E AbdElrazek. A Real-Time System for Facial Expression Recognition using Support Vector Machines and k-Nearest Neighbor Classifier. International Journal of Computer Applications 159(8):23-29, February 2017.

[22] Serwach, M., & Stasiak, B. GA-based parameterization and feature selection for automatic music genre recognition. In Proceedings of 2016 17th International Conference Computational Problems of Electrical Engineering, CPEE 2016.