

Classification of EEG Signals using Nonlinear Features and Preprocessing Techniques



Saneesh Cleatus T, Thungamani M

Abstract: In this paper we study the effect of nonlinear preprocessing techniques in the classification of electroencephalogram (EEG) signals. These methods are used for classifying the EEG signals captured from epileptic seizure activity and brain tumor category. For the first category, preprocessing is carried out using elliptical filters, and statistical features such as Shannon entropy, mean, standard deviation, skewness and band power. K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) were used for the classification. For the brain tumor EEG signals, empirical mode decomposition is used as a pre-processing technique along with standard statistical features for the classification of normal and abnormal EEG signals. For epileptic signals we have achieved an average accuracy of 94% for a three-class classification and for brain tumor signals we have achieved a classification accuracy of 98% considering it as a two class problem.

Keywords: EEG, EMD, Epilepsy, Brain Tumor

I. INTRODUCTION

Electroencephalography (EEG) is a noninvasive technique used to record the electrical signal generated due to neural interactions in the brain. Because of the noninvasive nature, EEG plays a prominent role in the diagnostics of various brain related disorders. EEG from scalp gives excellent temporal resolution which is a depiction of brain's condition and its function. However the electrical signal obtained from the scalp undergoes heavy noise effects due to artifacts such as eye movement, muscle movement etc. in order to overcome the effects of these artifacts we perform pre processing of the signals. Several preprocessing methods are available to improve the signal to noise ratio of EEG signals [1]. However since a complex biological system like the brain, the underlying dynamics is nonlinear. The traditional filtering methods may not be enough to process and model as EEG signals are neither band-limited nor stationary [2],[3]. The basic goal of EEG analysis is to extract useful information from recordings. One cannot resort to a particular method of analysis owing to the complexity of the

system under study. The standard protocol is to choose a method from a class of methods used for EEG analysis, depending upon the objective of the study. There is always a demand for improving the reliability and accuracy of methods, and it is an active area of research. We have conducted a survey on the existing methods for modelling EEG signals. The methods we were interested were,

1. Time and frequency domain approaches.
2. Statistical models.

In the time and frequency domain approaches, traditional time domain methods like co-variances and correlations are generally be used. The frequency domain approaches are mostly used to characterize EEG signals to find the dominant rhythm in the signal. The main drawback of FFT based methods is that it is not well suited for non-stationary signals. For the processing or pre-processing of EEG signals, standard linear transformation techniques like Singular Value Decomposition (SVD) and its related variations Principal Component Analysis (PCA), and Independent Component Analysis (ICA) are widely used [4].

Statistical models are widely used if a signal is said to nonlinear in nature as well as not stationary. Nonlinear extensions of Auto Regressive (AR) and Auto Regressive Moving Average (ARMA) models are also popular for the modelling of nonlinear signals. These models are used to get estimates of the best regression nonlinear polynomial fit. For modelling non-stationary time series, ARMA is generalized to ARIMA (Auto Regressive Integrated Moving Average) models. ARIMA processes reduce to ARMA when the data is fitted to the difference data [5].

II. METHODOLOGY

Objective of this work is to study the effect of different nonlinear techniques which is used in the preprocessing of EEG signals. Also, the different statistical parameters are used as features to check the classification accuracy. The proposed methods are verified on two categories of EEG data set. These data set corresponds to abnormalities such as brain tumor and epilepsy.

A. Analysis of epileptic EEG signals

In this section the detection of epileptic seizures from EEG signals using nonlinear methods are described. Figure 1 shows the methodology used in the processing of epileptic signals. It consists of three stages

1. Pre-processing of the raw EEG signal
2. Feature extraction.
3. Classification.

Manuscript received on June 08, 2021.

Revised Manuscript received on June 15, 2021.

Manuscript published on June 30, 2021.

* Correspondence Author

Saneesh Cleatus T*, Department of Electronics and Communication Engineering, BMS Institute of Technology and Management Affiliated to Visvesvaraya Technological, Bangalore(Karnataka), India. E-mail: saneesh@bmsit.in

Dr. Thungamani M, Department of Computer Science, College of Horticulture, University of Horticultural Sciences, Bangalore(Karnataka), India. E-mail: thungamani_k@rediffmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The data has been obtained from the University of Bonn. It consists of five sets namely Z, O, N, F and S. Each set contains 100 files of single-channel EEG segments and each file has a duration of 23.6 seconds.

The scalp EEGs from normal subjects with eyes open is set Z and eyes closed is set O. Sets N, F, and S were recorded using intracranial EEGs from epileptic patients in which sets N and F were recorded when there were no seizures whereas set S was recorded when seizures occurred. The sampling rate used to record the data was 173.61 Hz [6].

Block diagram of the methodology

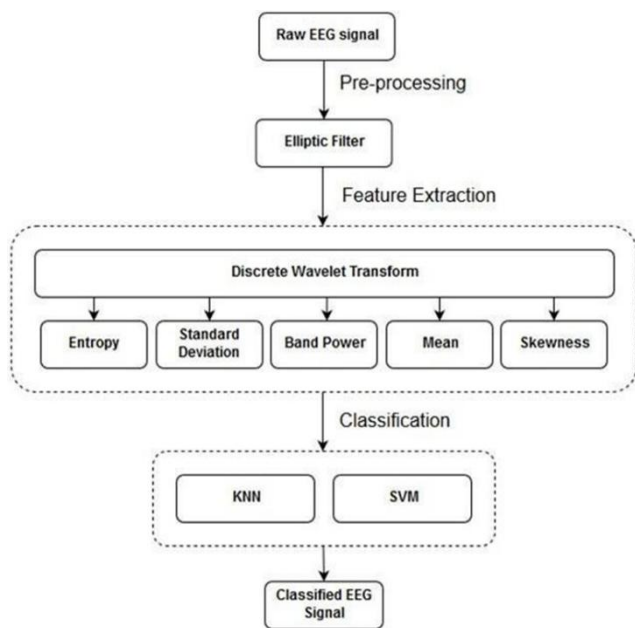


Figure 1. Overall methodology adopted to classify the epileptic EEG signal

As indicated in the figure 1, the first stage is the collection of the EEG data. This data which was obtained from BONN university database is passed through elliptical filter because of the reason that elliptical filters has got sharp cutoff frequencies. The preprocessed signal was given as input to the feature extraction stage. At first the discrete wavelet transform of the preprocessed EEG signal obtained. The resultant signal is evaluated for the below mentioned features.

Shannon Entropy

Shannon’s Entropy is the minimum number of bits needed to encode the signal. The usage of Shannon entropy makes the computational process fast with comparable accuracy. This also ensures dimensionality of the features to the minimum. This reduction in dimensionality and computational complexity contributes to the improvement of the speed of classifier.

Shannon Entropy (SE) is defined in equation (1).

$$SE = -\sum_{i=1}^k p(i) \log_2 p(i) \quad (1)$$

where $p(i)$ is the probability or normalized frequency of unique values [7], [8].

Mean, Standard Deviation, Skewness and Band Power are the other features which were used [9]. Further these features are given to two classification algorithms such as K Nearest

Neighborhood (KNN) classifier and Support Vector Machine (SVM) classifier [10].

In KNN, first training data set and a test set were provided. The training data set consists of extracted features from the sub-bands of an epileptic and a non-epileptic person. The test data set consists features of either epileptic and non-epileptic or both. The training data set is then fed to the classifier which plots the values in an n-dimensional space. A value from the test set is taken and plotted in the same n-dimensional space and then the specified number of nearest neighbours to it is calculated. If the majority of the neighbours belong to the epileptic class, it is classified as epileptic data. It works similarly for the non-epileptic data as well. This is done for each and every value of the test set and then the accuracy is calculated.

In SVM, a train and test set is created in a similar fashion as in KNN. Each feature value of the train set is plotted in an n-dimensional space. Then classification is performed on the test set by finding the hyper-plane that best segregates the two classes and accuracy is calculated.

B. Analysis of EEG signals of brain tumor data.

For the classification of brain tumor data, we have chosen the EEG Data available from Temple university TUH EEG corpus. The data have been handpicked from the descriptions available in the clinicians report along with each edf file. The edf file is then converted into MATLAB readable text file due to size limitations of edf file. Many of this edf file contained 31 channels with each channel of sample length more than fifty thousand.

The text converted edf data is then used for empirical mode decomposition (EMD) analysis. EMD decomposes the EEG signal into number of constituent signals called Intrinsic Mode Function (IMF). Further statistical features on the IMFs such as mean, variance, standard deviation, temporal skew, Higuchi fractal dimension were computed. Apart from the temporal features, we have evaluated the spectral features such as spectral skewness and power spectral density [11]-[14]. The power spectral density of a sequence is computed as in equation 2

$$P(w) = \sum_{-\infty}^{\infty} r_y[n] e^{-jwn} \quad (2)$$

Where $r_y[n]$ represents the autocorrelation of $y[n]$, defined as:

$$r_y[n] = E(y[m]y^*[m])$$

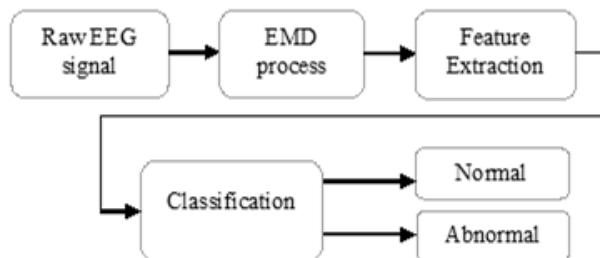


Figure 2. Methodology adopted to classify the brain tumor EEG signal.



Visual analysis of result obtained after performing PSD on IMFs shows that the statics of PSD can be one of the relevant features for feature extraction. Variation coefficient is another feature which is extracted and is defined as the ratio of standard deviation to mean. The spectral features such as spectral skew(β_s) is defined in equation 3. Another feature spectral kurtosis is basically a 4th order statistical tool which signifies the occurrence of transients and their locations in the frequency domain by decomposing the kurtosis of a signal as a function of frequency.

$$\beta_s = \frac{\sum_w \left(\frac{w - \mu_s}{\sigma_s}\right)^3 p(w)}{\sum_w p(w)} \quad (3)$$

After extracting these features the vector is created by concatenating the temporal and spectral features of each IMF, which is shown below

$$F = [\mu_t, \sigma_t, V_s, \beta_t, H_t, \sigma_s, \beta_s, K_s]$$

where μ_t is the mean, σ_t is the standard deviation, V_s is the variance, β_t is the temporal skew, H_t is the Higuchi fractal dimension, σ_s is the variation coefficient of power spectral density, β_s is the spectral skewness and K_s is the spectral kurtosis.

The values obtained from feature extraction process are tabulated which becomes input to the classifier. In the classifier, the best relevant features combination subset is selected and the classifier chosen makes use of the feature combination to classify the EEG signals as normal and abnormal signals [15],[16].

III. RESULT AND DISCUSSION

A. Results of proposed method for epileptic EEG signals

At every second we have calculated the features for all the five bands in the EEG namely Gamma, Beta, Alpha, Theta and Delta. The results of all the features are tabulated in table I.

Table- I: Feature values obtained

| Feature | Class | Minimum | Maximum |
|--------------------|-------------|-----------------------|-----------------------|
| Shannon Entropy | Seizure EEG | -6.2×10 ⁷ | -9.2×10 ⁵ |
| | Normal EEG | -3.9 ×10 ⁶ | -9.5×10 ⁹ |
| Mean | Seizure EEG | -1.96 | 1.41 |
| | Normal EEG | -0.39 | 0.56 |
| Standard deviation | Seizure EEG | 12.72 | 419.22 |
| | Normal EEG | 6.03 | 70.05 |
| Skewness | Seizure EEG | -3.02 | 5.57 |
| | Normal EEG | -0.59 | 1.19 |
| Bandpower | Seizure EEG | 161.91 | 17.57×10 ⁴ |
| | Normal EEG | 36.37 | 4.096×10 ³ |

The output of the three data sets from the five bands were used as inputs to the trained classifier[17],[18]. Table 1 provides the classification accuracy of every band as training data against selected band as testing data. The accuracy ranges from 90% to 100% with an average accuracy of 94 %

Table- II: Classification accuracy against each band.

| Train | Test | Classifier | Accuracy |
|-------|-------|------------|----------|
| Gamma | Gamma | SVM | 90% |
| | Beta | | 92% |
| | Alpha | | 92% |
| Beta | Gamma | SVM | 92.67% |
| | Beta | | 99% |
| Alpha | Gamma | SVM | 90.67% |
| | Beta | | 95.33% |
| | Alpha | | 98.67% |

| | | | |
|-------|-------|-----|--------|
| | Theta | | 90.67% |
| Theta | Beta | KNN | 94.67% |
| | Alpha | | 95.33% |
| | Theta | | 94.67% |
| Delta | Beta | SVM | 94% |
| | Alpha | | 100% |
| | Theta | | 94% |

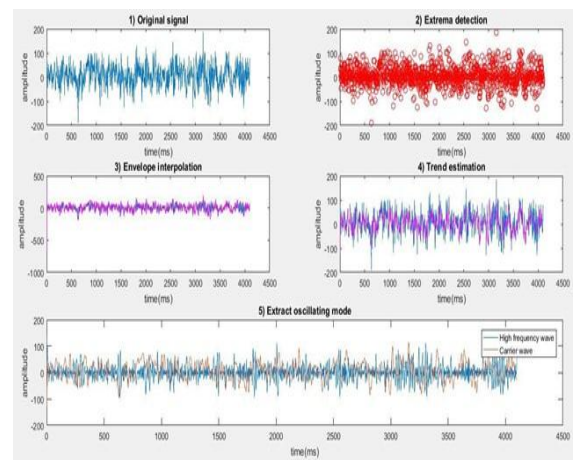
B. Results of proposed method for tumor affected brain EEG signals.

The proposed EMD based brain tumor detection method was implemented using MATLAB. The publicly available clinical data consisting of multichannel tumor EEG signals that are sampled at 250Hz using 16 bits per sample were taken for tumor detection. The EDF files obtained involves both EEG specific channels along with the supplementary channels like detected bursts, EKG, EMG, and photonic stimuli. This data was monitored during various stages such as sleep, awake, drowsiness. The data from set A (normal) and set B (abnormal) sampled at 173.61Hz, each containing 100 txt files were taken as normal EEG data. Both normal and tumor dataset were used to determine IMF's, Hilbert transforms and PSD, to achieve extraction of temporal and spectral features. Below figures show the extracted characteristics from both Normal and Brain tumor EEG. Highest order IMF obtained after EMD process from normal EEG and tumour EEG is shown in figure 3(a) and 3(b).

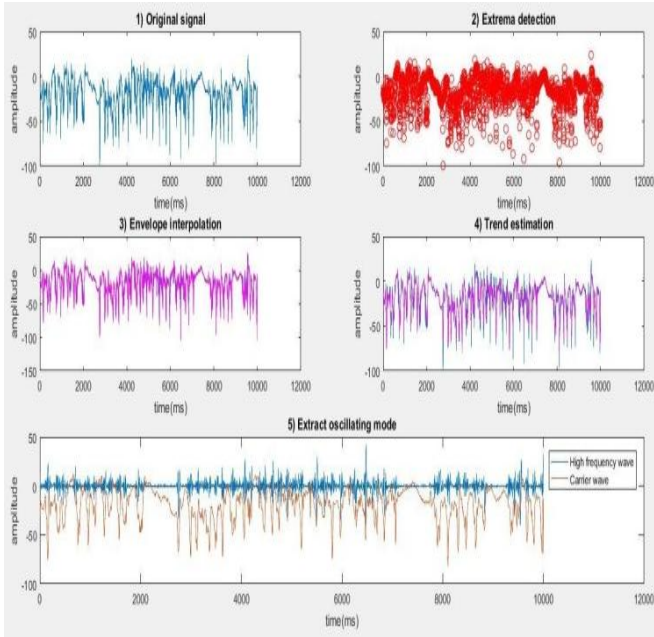
The results of the temporal feature values and spectral features values for normal EEG signal and tumor EEG signal is calculated and represented in table III.

Table- III: Classification accuracy against each band.

| Features | Normal Min | Normal Max | Tumor Min | Tumor Max |
|-----------------------|------------|------------|-----------|-----------|
| Temporal skew | -0.08 | 0.043 | -0.225 | 0.1636 |
| Spectral skew | 0.2073 | 1.45 | -0.95 | 0.86 |
| Mean | -0.89 | 0.556 | -0.233 | 0.014 |
| Variance | 234 | 7933 | 3.68 | 184 |
| Standard deviation | 15.29 | 89 | 1.919 | 13.569 |
| Variation Coefficient | -450 | -63.22 | -110.2 | -32.5 |
| Spectral kurtosis | 2.04 | 5.3942 | 1.795 | 4.295 |

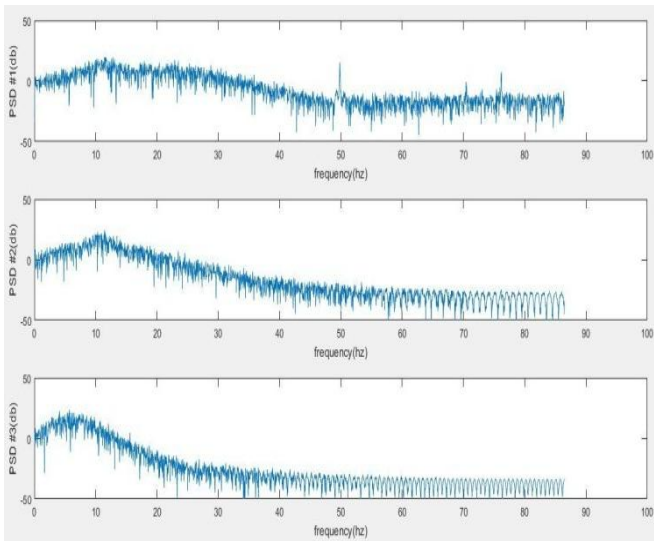


3 (a)

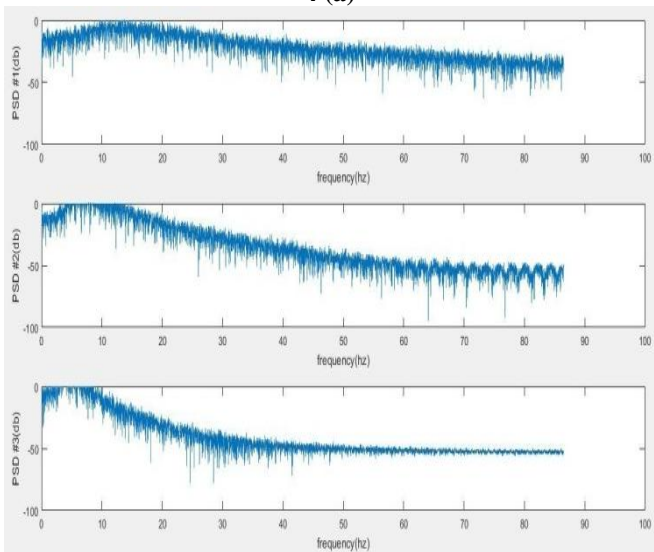


3 (b)

Figure 3 (a). EMD Decomposition of normal EEG.
3 (b). EMD Decomposition of tumor EEG.



4 (a)



4 (b)

Figure 4(a) . PSD plot of IMFs from normal EEG
4(a) . PSD plot of IMFs from tumor EEG.

It was observed that, when 100 normal samples and 100 tumor samples were taken for training the appropriate classifier, 85% accuracy was obtained. As the number of samples increased to 200 for normal and 200 abnormal in the training data, 94% accuracy was reached. Similarly, when 300 normal and 300 abnormal samples were used, finally 98% accuracy was achieved.

IV. CONCLUSION

This paper proposes a technique to distinguish EEGs of two category of abnormalities such as epilepsy and brain tumor. We have successfully demonstrated the extraction of various nonlinear features and classification. We have obtained promising results for BONN university data base. For complex signals obtained from data sources such as THU EEG corpus needs improvement in our techniques which will be the scope for future work.

REFERENCES

1. S. Krishnan and Y. Athavale, "Trends in biomedical signal feature extraction," Biomed. Signal Process. Control, vol. 43, 2018, pp. 41–63.
2. N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," Physical Review Letters, vol. 45, no. 9, pp. 712–716, 1980.
3. F. Takens, "Detecting strange attractors in turbulence," Springer, pp. 366–381, 1981.
4. S. Sanei and J. A. Chambers, EEG Signal Processing, vol. 1. 2007.
5. J. Pardey, S. Roberts, and L. Tarassenko, "A review of parametric modelling techniques for EEG analysis," Med. Eng. Phys., vol. 18, no. 1, pp. 2–11, 1996.
6. Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E, 64, 061907
7. N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," Comput. Methods Programs Biomed., 2005.
8. D. Phung, D. Tran, W. Ma, P. Nguyen, and T. Pham, "Using Shannon Entropy as EEG Signal Feature for Fast Person Identification," 22nd Eur. Symp. Artif. Neural Networks, Comput. Intell. Mach. Learn. (ESANN 2014), 2012.
9. R. Shiavi, Introduction to applied statistical signal analysis: Guide to biomedical and electrical engineering applications. Elsevier, 2010.
10. F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," Journal of Neural Engineering. 2007.
11. Y. Zhang, S. Zhang, and X. Ji, "EEG-based classification of emotions using empirical mode decomposition and autoregressive model," Multimed. Tools Appl., pp. 1–14, 2018.
12. G. Rodríguez-Bermúdez and P. J. García-Laencina, "Analysis of EEG Signals using Nonlinear Dynamics and Chaos : A review," Appl. Math. Inf. Sci., vol. 9, no. 5, pp. 2309–2321, 2015.
13. J. Roschke and J. B. Aldenhoff, "A nonlinear approach to brain function: Deterministic chaos and sleep EEG," Sleep, 1992.
14. J. Fell, J. Röschke, and P. Beckmann, "Deterministic chaos and the first positive Lyapunov exponent: a nonlinear analysis of the human electroencephalogram during sleep," Biol. Cybern., 1993.
15. Murugesan, M. and Sukanesh, R., 2009. Towards detection of a brain tumor in electroencephalogram signals usingsupport vector machines. International Journal of Computer Theory and Engineering, 1(5), p.622.
16. Murugesan, M. and Sukanesh, R., 2009, December. Automated detection of brain tumor in EEG signals usingartificial neural networks. Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT'09. International Conference on (pp. 284-288). IEEE.

AUTHORS PROFILE



Saneesh Cleatus T, is working as Associate Professor in the Department of Electronics and Communication Engineering at BMS Institute of Technology and Management, Bangalore. Saneesh is currently pursuing his PhD in the applications of machine learning techniques in the biomedical signal processing from Visvesvaraya Technological University, Karnataka. He has completed his master's degree from Cochin University of Science and Technology, Kerala. He has got 12 international publications and his research interests includes biomedical signal processing, machine learning etc. Prior to academic career he worked in the industry for over eight years in the field of embedded systems and particularly in building automation product development. He is a chartered engineer and a Member of IEEE, Institution of Engineers(India), ISTE and Computer Society of India.



Dr. Thungamani M, obtained BE degree in Computer Science & Engineer from Bangalore University, M.Tech in Computer Science & Engineering from Dr. MGR University Chennai and awarded Doctor of philosophy in Computer Science in 2014, from RV College of Engineering as a research centre of Tumkur University, Tumkur. She is having 12 years of teaching experience out of it 7 years in Engineering colleges, CIT Gubbi, Tumkur, KIT Tiptur and as a Associate Professor in BMS Institute of Technology & Management Bengaluru, Visvesvaraya Technological University, Belgaum and in 2015 joined and presently working as Assistant Professor of Computer Science, College of Horticulture (UHS) campus, GKVK, Bengaluru. Her area of research is Pattern Recognition, Image processing, and Big Data analysis. She had published 16 papers in International Journals and presented 10 papers in National conference. At present she is guiding four Ph.D scholars from the Visvesvaraya Technological University. She has been nominated by the university (UHS) as a master trainer and technical team member of application software, "Academic Management Software (AMS) system" for effective implementation of the software across the institutions of University of Horticultural Science, Bagalkot. As a host and technical coordinator she is involved organizing and conducting international conference and webinar which are organized by the department of Biotechnology and crop improvement, college of Horticulture, Bengaluru. She is a life member of Indian Society for Technical Education(ISTE) and Institution of Electronics and Telecommunication Engineers (IETE)