

Modelling EEG Dataset for Stress State Recognition using Decision Tree Approach

Mamata S Kalas¹, B. F. Momin²

¹Research Scholar, Walchand College of Engg, Sangli

²Associate Professor, Department of Computer science and Engg, WCE, Sangli

Abstract -Electroencephalography (EEG) is a predominant tool for learning the stress behavior. This work concentrates towards stress detection by means of eye states. This work proposes a framework which would be supportive in identifying human stress level and as an outcome, distinguishes a normal or stressed person. In this work, we used decision trees, carried out the performance analysis and found that it gives good performance in recognizing the stress states. This analysis is performed with reference to eye state: whether eyes are closed indicating rest, open eyes with blinks

Keywords- Electroencephalography (EEG), Decision trees, Stress detection

I. INTRODUCTION

There are several human activities like vehicle driving, heavy equipment operation, hazardous materials manipulation, where the awareness and conscious control is a very important factor. The principal reason for measuring stress is to quantify the mental cost of performing tasks in order to predict operator and system performance [3]. We have to characterize mental states of operator performance, by finding patterns in timely changing physiological, measures like EEG (Electro Encephalography), with eye blinks. The said research work has significance in the general field of safety of individual working in monotonous environment like driving vehicle, landing an aircraft, operating machines and technology implications of a cognitive task analysis, in air force research, army research, and transport organizations concentrating on technology-mediated attention enhancing safety [3]. The relevant information from each measurement is extracted via evaluation of a reduced set of selected features. These features are primarily obtained from filtered and processed versions of the raw time measurements with calculations of certain statistical and descriptive parameters. Selection of the reduced set of features was performed using DWT algorithm, thus constraining the computational cost of the real-time implementation. Different classification approaches have been studied, but classifier algorithms chosen for this investigation because they represent a good tradeoff between the intelligence of the solution and computational complexity [4]. This research paper deals with determining performance characteristics of EEG dataset using classifier models like Decision tree, SVM classifier & linear regression model. Comparative analysis is carried out to determine classification accuracy, time taken

for processing, Confusion matrix and predictive analysis of actual v/s observed values.

II. PROPOSED WORK

The aim of proposed technique is to design and develop a method that analyze the stress level effectively and also tackle the intersubject difference of human stress and provide a techniques for reducing the stress among individuals for improving their performance in work.

There are few researches on stress reduction techniques, therefore to control the health related issues generate from stress, stress reduction technique is important in relation to technology. So it motivates to do research on stress reduction. Consequently, this work is concerned with estimation of stress and analysis of human stress level by using the cluster-based analysis method and also developed a system will be helpful to reduce the human stress.

System Design

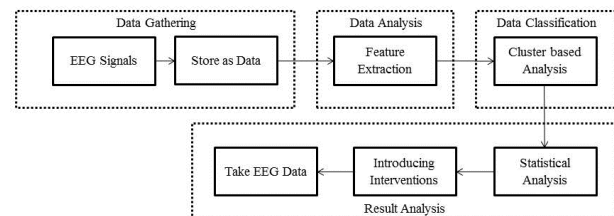


Figure 1. Architecture of Proposed System

In the proposed approach, the goal is to reduce the human stress after detecting the stress by using the EEG signals. The objective of this research is to accurately estimate the human stress and diagnose the human stress level. The estimation of stress can be done by analyzing the EEG features and the human stress level i.e. stress or relaxed

mode is shown by using clustering process [1][25]. The k-means clustering will be used for dividing the subjects into subgroups for predicting the human stress level [1][25]. The aim is to reduce stress by introducing the interventions into the system if stress level of human is high and do the statistical analysis for checking that stress level is reduced or not. In the proposed approach, a method will be implemented to reduce the human stress, so that they can efficiently improve his/her performance in the work[25].

Stress reduction is an objective of proposed approach. To achieve this objective steps used in this study are describe below [25].

1. Get the test data for checking the stress.
2. Do the data preprocessing on collected data.
3. Classifying data using clustered based analysis method.
4. If stress is high then introducing the intervention into the system.
5. Perform statistical analysis on data for getting the results.
6. Analyze the obtained results by performing comparative analysis of classifier models.

First phase is to collect standard EEG Datasets, second phase is pre-processing, third phase is feature extraction, fourth phase is classification and final phase is emotion/stress detection. **Inputs are taken from EEG Readings of RMS MAXIMUS Data obtained from Matrix Radiotherapy from Kolhapur.** After collecting standard dataset, signals need to be preprocessed. Pre-processing aims to simplify subsequent processing operations, improve signal quality without losing significant information. In this step, the recorded signals are processed to clean and remove noisy artifacts such as eye blinks in order to get the relevant information embedded in the signals. The next phase is feature extraction. After getting noise-free signals from the preprocessing phase, important features from the brain activities signals are extracted. Representative features obtained from the previous stage are classified using classification approach.

III. COMPUTATIONAL RESULTS

We conducted experiments on participants who have no history of psychiatric problems or neuro disorders. The purpose was clearly explained to them and consent was obtained from them. A wireless EEG device, Emotiv Epoch head set was placed according to international 10-20 system. The electrodes were attached to the scalp at position AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 as shown in Figure 1[5].

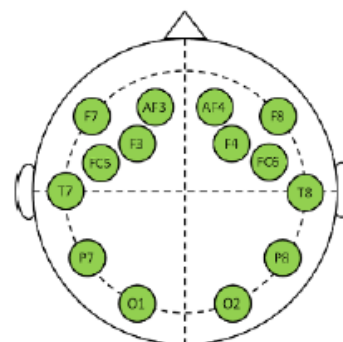


Figure 2. Electrode placement on the scalp.

The duration of the measurement was 117 seconds. The eye state was detected via a camera during the EEG measurement and added later manually to the file after analyzing the video frames. '0' indicates the eye-closed and '1' the eye-open state. All values are in chronological order with the first measured value at the top of the data.

EEG data was recorded. We validate the method using a participant study engaged in reading activity[8]. Using a support vector machine (SVM) classifier and other classifiers we perform comparative analysis of classification of EEG data and detect the stress of an individual by analyzing eye state. For this classification procedure, we have formed the decision tree rules, textual representation of decision tree, Complexity Table of decision tree, confusion matrix(Error Matrix). This can be done by averaging the results of ten 10-fold cross-validation runs. In each of one 10-fold cross-validation run the data is split into 70% training, 15% for testing, 15% for validation partitions/folds[9]

1. Decision Tree:

Decision tree methodology is a commonly used data mining method for establishing classification systems based on multiple covariates or for developing prediction algorithms for a target variable. This method classifies a population into branch-like segments that construct an inverted tree with a root node, internal nodes, and leaf nodes. The algorithm is non-parametric and can efficiently deal with large, complicated datasets without imposing a complicated parametric structure. When the sample size is large enough, study data can be divided into training and validation datasets. Using the training dataset to build a decision tree model and a validation dataset to decide on the appropriate tree size needed to achieve the optimal final model [22]. The decision tree method is a powerful statistical tool for classification, prediction, interpretation, and data manipulation that has several potential applications in medical research.

Using decision tree models to describe research findings has the following advantages:

- Simplifies complex relationships between input variables and target variables by dividing original input variables into significant subgroups.
- Easy to understand and interpret.
- Non-parametric approach without distributional assumptions[22].

Input: node n , partition D , split selection method S

Output: decision tree for D rooted at node n

Top-Down Decision Tree Induction Schema:[26]

BuildTree(Node n , data partition D , split selection method S)

- (1) Apply S to D to find the splitting criterion
- (2) **if** (a good splitting criterion is found)
- (3) Create two children nodes $n1$ and $n2$ of n
- (4) Partition D into $D1$ and $D2$
- (5) **BuildTree**($n1$, $D1$, S)
- (6) **BuildTree**($n2$, $D2$, S)

endif

Decision tree algorithm for classification

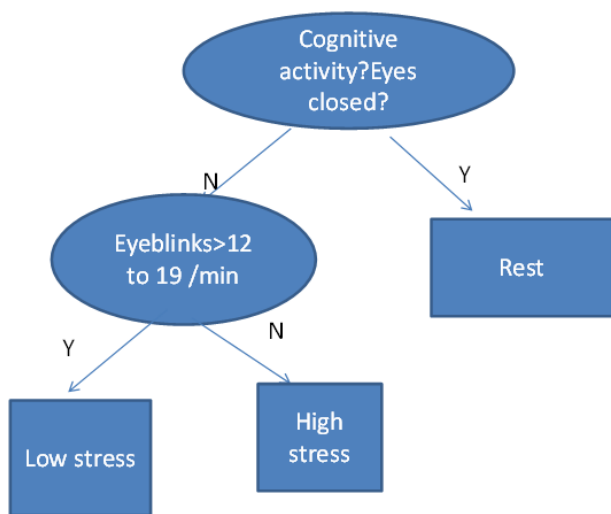


Fig 3. Decision tree model for classification of stress

Decision tree model is one of the most common data mining models can be used for predictive analytics. The reported investigation depicts optimum decision tree architecture achieved by tuning parameters such as Min split, Min bucket, Max depth and Complexity. DT model, thus derived is easy to understand and entails recursive partitioning approach implemented in the “rpart” package. Moreover the performance of the model is evaluated with reference Mean Square Error (MSE) estimate of error rate. DT is a nonparametric supervised learning method used for classification .DT creates a series of binary decisions on the

features which best distinguishes classes. Min Split Min Bucket, max Depth, Complexity [9][24].

The quality assessment is carried out using test data set and eventually evaluated in terms of mean squared error (MSE) and correlation coefficient (r). Mean squared error is given by equation (3). The Y_i represents the observed value, where, $i=1, 2, \dots, n$ denote the values of the class variable of the i^{th} observation and \hat{Y}_i denote the predicted value of the i^{th} observation. The difference $(Y_i - \hat{Y}_i)$ is termed as an error. Then mean square error is defined as,

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \text{----- (1)}$$

The parameters described below are irrespective of tool. It is important to understand the role of parameters used in tree modeling. These parameters are available in R & Python [23].

1. Minimum samples for a node split

Defines the minimum number of samples (or observations) which are required in a node to be considered for splitting.

Used to control over-fitting. Higher values prevent a model from learning relations which might

be highly specific to the particular sample selected for a tree.

Too high values can lead to under-fitting hence; it should be tuned using CV.

2. Minimum samples for a terminal node (leaf)

Defines the minimum samples (or observations) required in a terminal node or leaf.

Used to control over-fitting similar to min_samples_split.

Generally lower values should be chosen for imbalanced class problems because the regions in

which the minority class will be in majority will be very small.

3. Maximum depth of tree (vertical depth)

The maximum depth of a tree.

Used to control over-fitting as higher depth will allow model to learn relations very specific to a particular sample.

Should be tuned using CV.

4. Maximum number of terminal nodes

The maximum number of terminal nodes or leaves in a tree.

Can be defined in place of max_depth. Since binary trees are created, a depth of ‘n’ would

produce a maximum of 2^n leaves.

5. Maximum features to consider for split

The number of features to consider while searching for a best split. These will be randomly selected.

As a thumb-rule, square root of the total number of features works great but we should check upto

30-40% of the total number of features [23].

Tuning parameter	Description	Value of DT Model
Min split	Minimum number of observations that must exist in a node resulting from a split before a split will be performed	20
Min Bucket	This is the minimum number of observations allowed in any leaf node of the decision tree	7
Max depth	This is the maximum depth of any node of the final tree	30
Complexity	This parameter is used to control the size of the decision tree and to select optimal tree size.	0.01

Table 1: Details of tuning parameters varied in Rattle to obtain optimized decision tree model for stress detection analysis.

Fig.3 shows decision tree derived in the present investigation represents supervised learning model for stress detection using eye blinks. Table 1 gives details of tuning parameters varied in Rattle to obtain optimized decision tree model for stress detection analysis. R is a popular statistical programming language with a number of extensions that support data processing and machine learning tasks [12]. However, interactive data analysis in R is usually limited as the run time is single-threaded and can only process data sets that fit in a single machine’s memory[24].

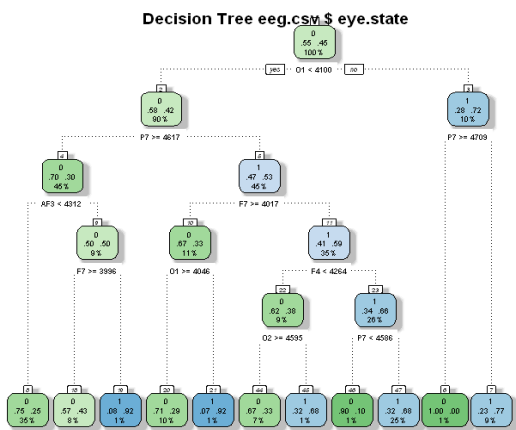


Figure 4: Decision Tree

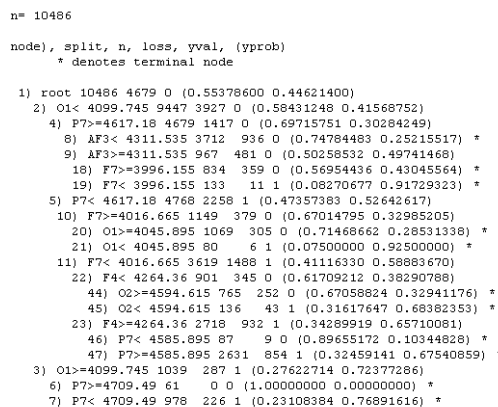


Figure 5: Textual representation of decision tree

Table 2 shows Performance evaluation of the model. The table depicts the number of iterations carried out and accuracy in results in terms of the cross-validated error, which is the *xerror* column of the table. The CP (complexity parameter) value reveals that as the tree splits into more nodes, the complexity parameter is reduced. But we also note that the cross validation error starts to increase as we further split the decision tree. This tells the algorithm to stop partitioning, as the error rate is not improving [9][24].

Table 2: Complexity table for DT model

level	CP	nsplit	relerror	xerror	xstd
1	0.099380	0	1.00000	1.00000	0.0108791
2	0.068711	1	0.90062	0.90212	0.0107327
3	0.045095	3	0.76320	0.78628	0.0104444
4	0.014747		0.71810	0.72558	0.0102404
5	0.014533	5	0.70336	0.71532	0.0102021
6	0.013037	6	0.68882	0.70229	0.0101518
7	0.011862	7	0.67579	0.68391	0.0100777
8	0.010686	9	0.65206	0.66638	0.0100036
9	0.010000	10	0.64138	0.66296	0.0099887

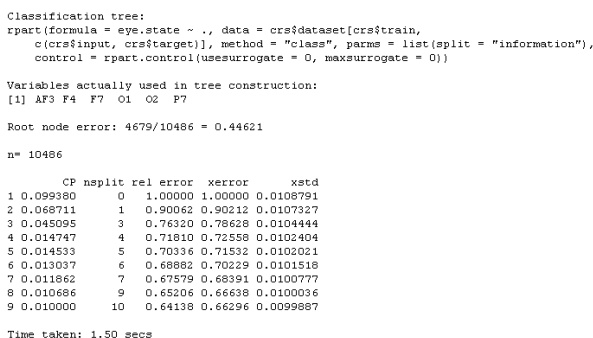


Figure 6: Decision tree model & its performance

For each of these classification procedures, an estimator for the misclassification error, a confusion matrix, and the corresponding receiver-operating characteristic (ROC) is computed by averaging the results of ten 10-fold cross-validation runs. In each of one 10-fold cross-validation run the data is split into 70% training, 15% for testing, 15% for validation partitions/folds[9]

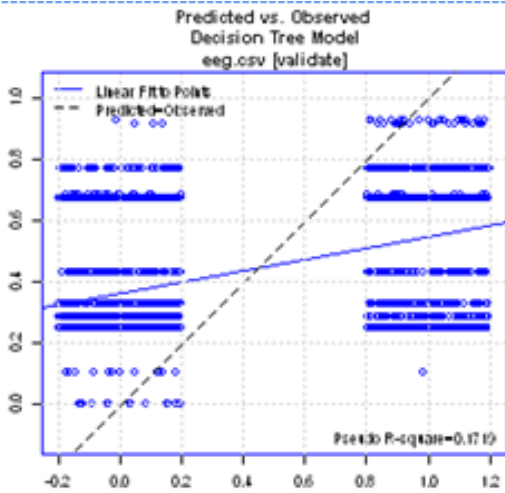


Figure 7: Comparative Performance of Classifiers with Validation Data Set

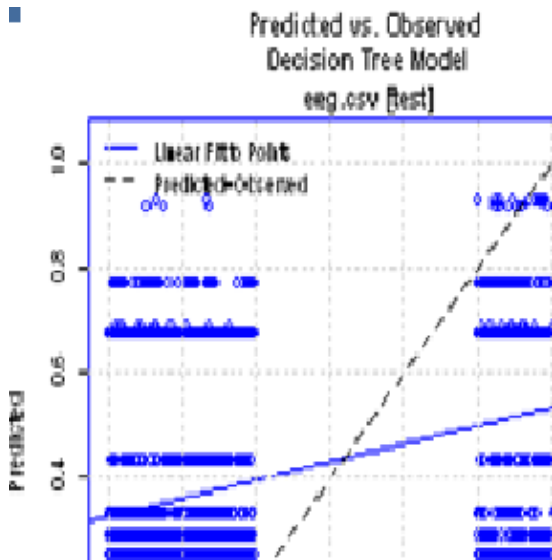


Figure 8: Comparative Performance of Classifiers with Test Data Set

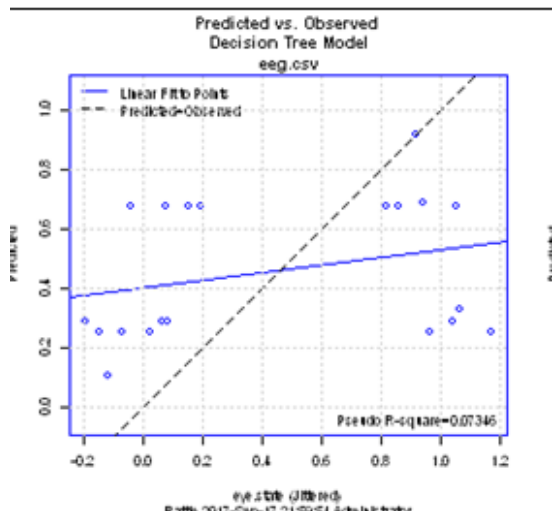
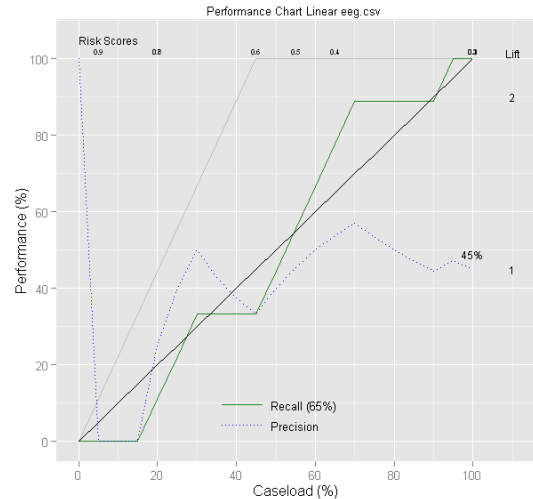


Figure 9: Comparative Performance of Classifiers with New Data Set

observed values. Two lines are also plotted, one being a linear fit to the actual points, and the other being the perfect fit, if the predicted values were the same as the actual observations. The Pseudo R-Squared is a measure that tries to mimic the R-Squared. It is calculated as the square of the correlation between the predicted and observed values. The closer to 1, the better[12]



Risk charts are particularly suited to binary classification tasks, which are common in data mining. A risk chart is particularly useful in the context of the audit dataset, and for risk analysis tasks in general. The audit dataset has a two class target variable as well as a so called risk variable, which is a measure of the size of the risk associated with each observation. Observations that have no adjustment following an audit (i.e., clients who have supplied the correct information) will of course have a risk of zero associated with them. Observations that do have an adjustment will usually have a risk associated with them, and for convenience we simply identify the value of the adjustment as the magnitude of the risk[12]. Rattle uses the idea of a risk chart to evaluate the performance of a model in the context of risk analysis [12].

Caseload is the percentage of the entities in the dataset covered by the model at a particular probability cutoff, so that with a cut off of 0, all (100%) of the entities are covered by the model. With a cutoff of 1 (0%) no entities are covered by the model. A diagonal line is drawn to represent a baseline random performance. Then the percentage of positive cases (the recall) covered for a particular caseload is plotted, and optionally a measure of the percentage of the total risk that is also covered for a particular caseload may be plotted. Such a chart allows a user to select an appropriate trade-off between caseload and performance

The Predicted vs. Observed plot is relevant for regression models (predicting a continuous value rather than a discrete value). It will display the predicted values against the

Model	Area under recall curve
Decision tree	77%

Confusion Matrix: A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known.

Actual Class	Predicted class		
		Yes	No
Yes		TP	FN
No		FP	TN

Error Matrix

Error matrix for the Decision Tree model on eeg.csv (counts):

```

Predicted
Actual 0 1
0 7 4
1 4 5
    
```

Error matrix for the Decision Tree model on eeg.csv (proportions):

```

Predicted
Actual 0 1 Error
0 0.35 0.20 0.36
1 0.20 0.25 0.44
    
```

Overall error: 0.4, Averaged class error: 0.4040404

IV CONCLUSION AND FUTURE WORK

In the present paper we have reported modeling of Eye blink patterns by analyzing Electro Encephalography (EEG), a diagnostic method which is widely used in stress detection analysis. The reported investigation depicts optimum decision tree architecture achieved by tuning the number of trees and choice of variables for partitioning the dataset. The results showcases prediction of the patterns based on the Characterization of mental states of operator performance by EEG with eye blinks data by using the Decision tree modeling. In terms of errors, for decision tree classifier, the recall accuracy is 77%, overall error is 0.4, averaged class error is 0.44.

Future Work

We need to model the multitask performance, which is one of the major challenges of cognitive modeling, since performing multiple tasks at the same time is common in daily life. Accidents can occur during driving. Due to high stress, driver has diverted attention while driving. The decision-making power and slower reaction are consequences of accumulated stress, some thoughtless decisions can easily be made. To prevent this, it is necessary to make a software suite that can deliver an alert message to the driver to calm down or to be careful after recognizing the stress state.

REFERENCES

- [1] Qianli Xu, Member, IEEE, Tin Lay Nwe, Member, IEEE, and Cuntai Guan, Senior Member, IEEE, 'Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals', IEEE Journal of Biomedical And Health Informatics, Vol. 19, No. 1, January 2015.
- [2] Teresa A. Branscome and Jock O. Grynovicki, 'An Investigation of Factors Affecting Multi-task Performance in an Immersive Environment', Army Research Laboratory, Aberdeen Proving Ground, MD 21005-5425, December 2007
- [3] Mrs.Mamata S.Kalas, Dr. B. F. Momin, 'Blink detection using Electrooculography for stress detection ', National Research Symposium on Computing - RSC 2016 ISBN: 978-81-931456-1-8, Dec. 19-20, 2016.
- [4] Mohino-Herranz I, et al.' Assessment of Mental, Emotional and Physical Stress through Analysis of Physiological Signals Using Smartphones' Sensors (Basel). 2015.
- [5] Vanitha V*, Krishnan P, 'Real time stress detection system based on EEG signals', Biomedical Research, An International Journal of Medical Sciences, ISSN: 0970-938X(Print) — 0976-1683 (Electronic)
- [6] [https://en.wikipedia.org/wiki/Stress\(biology\)](https://en.wikipedia.org/wiki/Stress(biology))
- [7] <https://en.wikipedia.org/wiki/Electroencephalography> (accessed on 23 November 2017)
- [8] Andreas Bulling, Student Member, IEEE, Jamie A. Ward, Hans Gellersen, and Gerhard Troster, Senior Member, IEEE, 'Eye Movement Analysis for Activity Recognition' Using Electrooculography', IEEE Transactions on pattern analysis and machine intelligence, Vol 33, No 4, April 2011.
- [9] Kamath RS* and Kamat RK,' Modeling Mice Down Syndrome through Protein Expression:A Decision Tree based Approach', Research Journal of Pharmaceutical, Biological and Chemical Sciences, September October 2016 ISSN: 0975-8585.
- [10] Christoph Lehmann^aThomas Koenig, Vesna Jelic, Leslie Pritchep, Roy E. John, Lars Yadolah Dodge, Thomas Dierks,' Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)' Journal of Neuroscience Methods, Volume 161, Issue 2, 15 April 2007, Pages 342-350
- [11] Prashant S Lahane*, and Mythili Thirugnanam,' Literature review of Human Stress analysis using EEG Signal', Research Journal of Pharmaceutical, Biological and Chemical Sciences. ISSN: 0975-8585
- [12] Graham W. 'Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery'. Springer, UK, 2011, pp. 245-268.
- [13] Joachims, T. 'Text categorization with support vector machines: Learning with many relevant features'. In Proceedings of the European Conference on Machine Learning, Chemnitz, Germany, 21–23 April 1998; pp. 137–142
- [14] Radhika Deshmukh, Manjusha Deshmukh, 'Mental Stress Level Classification: A Review', International Journal of Computer Applications (0975 – 8887) International

Conference on Advances in Science and Technology (ICAST-2014)

- [15] Seyyed Abed Hosseini and Mohammad Bagher Naghibi-Sistani, 'Classification of Emotional Stress Using Brain Activity, Applied Biomedical Engineering.
- [16] C. Ranjith and Dr.M. Mohanapriya, 'Real-time Stress Detection and Classification Using Relevance Vector Machine', JARDCS, 2017, Pages: 458-470
- [17] J]VernonL whernW. DavidHairstonbKalebMcDowellbMarissaWesterfieldcKayRobbins,'Detection and classification of subject-generated artifacts in EEG signals using autoregressive models', Journal of Neuroscience Methods, Volume 208, Issue 2, 15 July 2012, Pages 181-189
- [18] SaidatulArdeenawatie*, Paul Murugesu Pandiyan , Sazali Yaacob, 'Mental Stress Level Classification Using Eigenvector Features and Principal Component Analysis', May, 2013, Vol. 3 Iss. 5, PP. 254-261
- [19] NORIZAM SULAIMAN et al,' Novel Methods for Stress Features Identification using EEG Signals', DOI 10.5013/IJSSST.a.12.01.04, ISSN: 1473-804x 27 online, 1473-8031 print
- [20] Jing Zhai and Armando Barreto,' Stress Recognition Using Non-invasive Technology', Biomedical sciences in instrumentation · February 2006.
- [21] Nargess Nourbakhsh, Yang Wang and Fang Chen,'GSR and Blink Features for Cognitive Load Classification', Citeseerx.ist.psu.edu.
- [22] Yan-yan SONG,* and Ying LU,'Decision tree methods:applications for classification and prediction', Shanghai Arch Psychiatry. 2015 Apr 25; 27(2): 130–135
- [23] <https://www.analyticsvidhya.com> > Machine Learning, A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python).
- [24] M.S.Kalas*, B.F.Momin'Modelling 'Characteristics of Eye Movement Analysis for stress detection – Performance Analysis using Decision tree approach', ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 7, Issue 4, October 2017
- [25] Mamata S Kalas, B.F.Momin,'Stress detection and reduction using EEG Signals',published in IEEEExplore ICEEOT'2016 Proceedings of the International Conference on Electrical Electronics and Optimization Techniques. March 3-5 2016 Included in IEEE on 24th November 2016.
- [26] Raghu Ramakrishnan,'Database Management Systems', 2nd edition. [27] Dimigen, O., Kliegl, R., & Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: a study with fixationrelated brain potentials. Neuroimage, 62 (1), 381-393

Authors Profile



Mrs. Mamata S. Kalas, having been graduate from University Of Mysore, in 1993, From B.I.E.T,Davangere, started her professional carrier there itself. Since 1995, she has worked as lecturer at D.Y.Patil's college of Engg., Bharati Vidyapeeth's College Of Engg.,Kolhapur.She is M.Tech(CST) Graduate and her dissertation work is based on image segmentation using parametric distributional clustering. She has been awarded with M.TECH (CST) from Shivaji University, Kolhapur of Maharashtra in June 2009. She is persuing Ph. D in computer science and Engineering at walchand College of Engineering, Reseach center, Shivaji University, under the guidance of Dr.B.F.Momin. She is currently working as an Associate Professor at KIT'S College of Engg, Kolhapur.She is in her credit, 19 Years of teaching experience, eight papers presented for international conferences,6 papers presented for n national conferences, seven papers published in international journals. Her areas of interest are pattern recognition and artificial intelligence, Computer Architecture, System programming.