IJEIT

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 7, Issue 4, October 2017

Modelling Characteristics of Eye Movement Analysis for stress detection – Performance Analysis using Decision tree approach

M.S.Kalas[1*], B.F.Momin[2]
[1]Research scholor, Walchand College of Engg, Sangli
[2]Asso. Prof, Dept of CSE, Walchand College of Engg, Sangli

Abstract: ElectroOcculography (EOG) data is used to develop and adjust the method of stress detection in operators and academicians who are engaged in continuous reading and learning operations. By monitoring the eyes, and EOG analysis it is believed that the symptoms of operator's stress can be detected early enough to avoid mistakes. There are several human activities where the awareness and conscious control is a very important factor: vehicle driving, heavy equipment operation, hazardous materials manipulation. The principal reason for measuring stress is to quantify the mental cost of performing tasks in order to predict operator and system performance. We have to characterize mental states of operator performance, by finding patterns in timely changing physiological, measures like EOG (Electro Occulography), with eye blinks. Various computational approaches based on EOG signals have been developed for analyzing and detecting stress of an individual. We report decision tree (DT) modeling of EOG data characteristics. We have done the comparative analysis of eye movement analysis for stress detection. We validate the method using an eight participant study in an office environment using an example set of two activities, Physical activity related to operating heavy machines, ,Cognitive activities like working with computers: copying a text, reading a printed paper, taking handwritten notes, watching a video, and browsing the Web. We also include periods with no specific activity (the NULL class).It is found that Decision tree model gives good performance with least error.

Keywords: **ElectroOcculography, Decision Tree, Efficiency, Rattle.**

I. INTRODUCTION

Stress detection is an ongoing research topic among both psychologists and engineers. Various technologies are developed on human stress detection using wearable sensors and bio signal processing. Stress can be detect from human bio-signals such as Electrooculogram (EOG) data ,Electroencephalography (EEG), Electromyography (EMG), Electrocardiography (ECG), Galvanic Skin Response (GSR), Blood Volume Pulses (BVP), Blood Pressure (BP), Skin Temperature (ST) and Respiration. Also human physiological features are used to measure the stress level using physiological signals. There is difference between individuals when he/she response to stress i.e. how a person's physiological feature changes in response to stressful events. In proposed system EEG (Electroencephalography) stress detection technique will be used.

In recent years, there have been some studies of collecting and processing electrooculography (EOG) signals. An EOG signal contains only two channels, a horizontal channel and a vertical channel, and there are fewer artifacts in EOG signals than in EEG signals. Therefore, EOG signals are easy to analyze and we can build the classifier models and carryout the comparative analysis of performance in terms of time taken for processing, Error rate, and Correlation coefficient for getting high accuracy only in processing clean EOG signals.

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.

II. RELATED WORK

A. Classification algorithms

Decision Tree: Theoretical considerations:

A decision tree is a set of conditions arranged in a hierarchical structure [1][23]. This is a classification/predictive model in which a data item is categorized by following the path of fulfilled conditions from the root of the tree till reaching a leaf. The leaf corresponds to a class label. A set of classification rules can be easily derived from decision tree. The basic algorithm for decision tree is the greedy algorithm that constructs decision trees in a top down recursive divide-and-conquer manner [22]. It explores the structure of a set of data, while developing easy to visualize decision rules for a classification tree. In next section Algorithm is given which explains construction of decision tree for classification.

Step 1: A subset of data taken as input and evaluate all possible splits.
Step 2: The single variable is found which best splits the data in to two groups. The best split decision. i.e. the split

with the highest information gain is chosen to partition the data in two subsets.
Step 3: The data is separated and then this process is applied separately to each subgroup.
Step 4: The method (step1 to 3) is called recursively until the subgroups either reach a minimum size or until no improvement can be made.

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[19][20]. 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[24].

III. MATERIALS AND METHODS

This section explores details of experiment conducted for the classification of eye blink patterns with classifier model, Decision trees classifiers. For sake of simplicity we compared eight participant study in an office environment using an example set of two activities, Physical activity related to operating heavy machines, ,Cognitive activities like working with computers: copying a text, reading a printed paper, taking handwritten notes, watching a video, and browsing the Web. We also include periods with no specific activity (the NULL class) in separate analysis. The following classification algorithms are applied and compared 1. For each of these classification procedures, an estimator for the misclassification error, a confusion matrix

We report Decision tree modeling of stress detection by analyzing EOG data. Present study exhibits performance estimation of various random forest configurations and compares the classification accuracy. 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 EOG with eye blinks data by using the Decision tree modeling.

Column1: timestamp [sec]
Column 2: timestamp [usec]
Column 3: horizontal EOG [Mobi amplitude units]
Column 4: vertical EOG [Mobi amplitude units]
Column 5: ground truth annotation

Ground truth annotation format:

Null: 1
Read: 2
Browse: 3
Write: 4
Video: 5
Copy: 6
disspeak: 7 // distraction speak
disphone: 8 // distraction phone

Participants (number gender comment)
----

(1) Male
(2) Male
(3) Male
(4) Male
(5) Male
(6) Female
(7) Female
(8) Male

Experimental Runs (number type comment)
(0) Run 1, performing a random sequence of office activities
(1) Run 2, performing a random sequence of office activities
The number in brackets corresponds to the number appended to each filename.

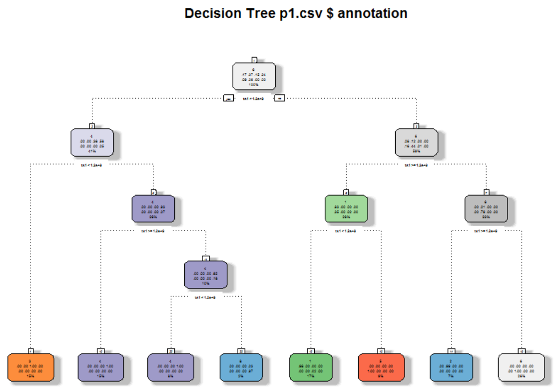

Fig 1: Decision Tree

The decision tree is a very convenient and efficient representation of knowledge [16]. To model the characteristics of eye blinks for stress patterns. We have employed decision tree approach. It starts with a single root node that splits into multiple branches, leading to further nodes, each of which may further split or else terminate as a leaf node. Associated with each nonleaf node will be a test or question that determines which branch to follow [15]. The leaf nodes contain the decisions.

DT is a nonparametric supervised learning method used for classification. DT creates a series of binary decisions on the features which best distinguishes classes. Following are the set of tuning parameters varied in Rattle to obtain optimized decision tree model for eye blink patterns for stress detection using EOG data:

The quality assessment is carried out using test data set and eventually evaluated in terms of mean squared error

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 7, Issue 4, October 2017

level	CP	nsplit	relerror	xerror	xstd
1	0.319928	0	1.000000	1.0000000	0.0022512
2	0.224145	1	0.4800717	0.4800717	0.00257428
3	0.207886	2	0.4559269	0.4559072	0.00243645
4	0.123381	3	0.2480414	0.2480217	0.00199575
5	0.092575	4	0.1246604	0.1246408	0.00149211
6	0.012106	5	0.0322854	0.0322428	0.00078707
7	0.010000	7	0.0078737	0.0080312	0.0039642

(MSE) and correlation coefficient (r). Mean squared error is given by equation (1). The Y_i represents the observed value, where, i=1, 2 ….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,

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

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

Tuning parameter	Description	Value of DT Model
Min split	Minimum number of observations thatmust exist in a node resulting from a split before a spilt 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

IV. COMPUTATIONAL DETAILS, DISCUSSION AND RESULTS

Derived optimized decision tree entails values for tuning parameters such as Min split; Min bucket, Max depth and complexity are 20,7,30 and 0.01 respectively.

Performance evaluation of the model is summarized in Table 2. This complexity table explains iterations and associated change in the accuracy of the model as new levels are added to the tree. We are most likely interested in 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.

Table 2: Complexity table for DT model

```
Summary of the Decision Tree model for Classification (built using 'rpart'):

n= 68419

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 68419 50802 6 (0.17 0.069 0.15 0.24 0.092 0.26 0.004 0.019)
   2) ts1< 1.237451e+09 28099 11846 4 (0 0 0.38 0.58 0 0 0 0.045)
     4) ts1< 1.237451e+09 10597    18 3 (0 0 1 0.0017 0 0 0 0) *
     5) ts1>=1.237451e+09 17502  1267 4 (0 0 0 0.93 0 0 0 0.072)
      10) ts1>=1.237451e+09 10495     0 4 (0 0 0 1 0 0 0 0) *
      11) ts1< 1.237451e+09 7007  1267 4 (0 0 0 0.82 0 0 0 0.18)
        22) ts1< 1.237451e+09 5711     4 4 (0 0 0 1 0 0 0 0.0007) *
        23) ts1>=1.237451e+09 1296    33 8 (0 0 0 0.025 0 0 0 0.97) *
   3) ts1>=1.237451e+09 40320 22703 6 (0.28 0.12 0 0.0005 0.16 0.44 0.0069 0)
     6) ts1>=1.237452e+09 17949  6562 1 (0.63 0 0 0 0.35 0 0.015 0)
      12) ts1< 1.237452e+09 11647   277 1 (0.98 0 0 0 0 0 0.024 0) *
      13) ts1>=1.237452e+09 6302    17 5 (0.0027 0 0 0 1 0 0 0) *
     7) ts1< 1.237452e+09 22371  4754 6 (0.00036 0.21 0 0.00089 0 0.79 0 0)
      14) ts1>=1.237452e+09 4757    31 2 (0.0017 0.99 0 0 0 0.0048 0 0) *
      15) ts1< 1.237452e+09 17614    20 6 (0 0 0 0.0011 0 1 0 0) *

Classification tree:
rpart(formula = annotation ~ ., data = crs$dataset[crs$train,
    c(crs$input, crs$target)], method = "class", parms = list(split = "information"),
    control = rpart.control(usesurrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:
[1] ts1

Root node error: 50802/68419 = 0.74251

n= 68419

        CP nsplit rel error    xerror       xstd
1 0.319928      0 1.0000000 1.0000000 0.00225132
2 0.224145      1 0.6800717 0.6800717 0.00257428
3 0.207886      2 0.4559269 0.4559072 0.00243645
4 0.123381      3 0.2480414 0.2480217 0.00199575
5 0.092575      4 0.1246604 0.1246408 0.00149211
6 0.012106      5 0.0320854 0.0322428 0.00078707
7 0.010000      7 0.0078737 0.0080312 0.00039642

Time taken: 5.00 secs
```

Fig 2: Textual representation of Decision Tree

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 7, Issue 4, October 2017

```
Tree as rules:

 Rule number: 10 [annotation=4 cover=10495 (15%) prob=10495.00]
   ts1< 1.237e+09
   ts1>=1.237e+09
   ts1>=1.237e+09

 Rule number: 22 [annotation=4 cover=5711 (8%) prob=5707.00]
   ts1< 1.237e+09
   ts1>=1.237e+09
   ts1< 1.237e+09
   ts1< 1.237e+09

 Rule number: 23 [annotation=8 cover=1296 (2%) prob=33.00]
   ts1< 1.237e+09
   ts1>=1.237e+09
   ts1< 1.237e+09
   ts1>=1.237e+09

 Rule number: 15 [annotation=6 cover=17614 (26%) prob=20.00]
   ts1>=1.237e+09
   ts1< 1.237e+09
   ts1< 1.237e+09

 Rule number: 4 [annotation=3 cover=10597 (15%) prob=18.00]
   ts1< 1.237e+09
   ts1< 1.237e+09

 Rule number: 14 [annotation=2 cover=4757 (7%) prob=0.00]
   ts1>=1.237e+09
   ts1< 1.237e+09
   ts1>=1.237e+09
```

Fig 3: Decision Tree as rules (partial)

```
Error matrix for the Decision Tree model on p1.csv [validate] (counts):

      Predicted
Actual    1    2    3    4    5    6    7    8
     1 2440    3    0    0    3    0    0    0
     2    0 1012    0    0    0    0    0    0
     3    0    0 2231    0    0    0    0    0
     4    0    0    1 3549    0    4    0    7
     5    0    0    0    0 1375    0    0    0
     6    0    3    0    0    0 3715    0    0
     7   71    0    0    0    0    0    0    0
     8    0    0    0    0    0    0    0  247

Error matrix for the Decision Tree model on p1.csv [validate] (proportions):

      Predicted
Actual    1    2    3    4    5    6 7    8 Error
     1 0.17 0.00 0.00 0.00 0.00 0.00 0 0.00     0
     2 0.00 0.07 0.00 0.00 0.00 0.00 0 0.00     0
     3 0.00 0.00 0.15 0.00 0.00 0.00 0 0.00     0
     4 0.00 0.00 0.00 0.24 0.00 0.00 0 0.00     0
     5 0.00 0.00 0.00 0.00 0.09 0.00 0 0.00     0
     6 0.00 0.00 0.00 0.00 0.00 0.25 0 0.00     0
     7 0.00 0.00 0.00 0.00 0.00 0.00 0 0.00     0
     8 0.00 0.00 0.00 0.00 0.00 0.00 0 0.02     0
```

Fig 4: Error Matrix for Decision Tree Model

```
Error matrix for the Decision Tree model on p1.csv (counts):

      Predicted
Actual 1 2 3 4 5 6 7 8
     1 3 0 0 0 0 0 0 0
     2 0 3 0 0 0 0 0 0
     3 0 0 2 0 0 0 0 0
     4 0 0 0 2 0 0 0 0
     5 0 0 0 0 4 0 0 0
     6 0 0 0 0 0 3 0 0
     8 0 0 0 0 0 0 0 3

Error matrix for the Decision Tree model on p1.csv (proportions):

      Predicted
Actual    1    2   3   4   5    6 7    8 Error
     1 0.15 0.00 0.0 0.0 0.0 0.00 0 0.00     0
     2 0.00 0.15 0.0 0.0 0.0 0.00 0 0.00     0
     3 0.00 0.00 0.1 0.0 0.0 0.00 0 0.00     0
     4 0.00 0.00 0.0 0.1 0.0 0.00 0 0.00     0
     5 0.00 0.00 0.0 0.0 0.2 0.00 0 0.00     0
     6 0.00 0.00 0.0 0.0 0.0 0.15 0 0.00     0
     8 0.00 0.00 0.0 0.0 0.0 0.00 0 0.15     0
```

Fig 5: Error Matrix for decision tree on test dataset

In the present investigation we have employed Decision tree modeling for classifying EOG data for eye blinks for stress detection in to ten classes of patterns.

V. PERFORMANCE METRICS

For developing a new detector and estimating its potential application performance, it is very important to examine properly the detection quality. The leave-one-out (LOO) cross-validation approach is used to assess the performance of the system for stress detection. The total average accuracy based on some feature and the classifier is the average of the accuracy of all single channels based on the same feature and same classifier.

To provide an easier-to-understand method to measure the detection quality, the well-known performance indicators, including accuracy (Acc), sensitivity (Sn), and specificity (Sp), are described as follows: where TP (true positive) denotes the number of the data inputs that refer to fatigue state correctly classified as fatigue. FP (false positive) is the number of data inputs that refer to normal state classified as stress state. TN (true negative) is number of the data inputs that refer to normal state correctly classified as normal state. FN (false negative) is the data inputs that refer to stress state classified as normal state.

VI. CONCLUSION AND FUTURE WORK

In the present paper we have reported modeling of Eye blink patterns by analyzing Electro Occulography (EOG), a diagnostic method which is widely used in stress detection analysis. A decision tree model is one of the most common data mining models. It is popular because the resulting model is easy to understand. 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. Result concludes that DT prediction is a suitable approach since the resulting analysis is much more accurate and precise. Consistent with our earlier reported investigations [4], the modelling of the Eye blink patterns by analyzing Electro Occulography (EOG) demonstrates strong correlation of efficiency with ts1 and ts2.In future we will work with modeling of eye blink patterns for an operator engaged in multiple task.

REFERENCES

[1] Sundar C, Chitradevi M, Geetharamani G. International Journal of Computer Applications, 2012; 47(14): 19-25.

[2] Tomas P, Krohova J, Dohnalek P, Gajdos P. 36th International Conference elecommunications and Signal Processing 2013; 620 – 923.

[3] Karabulut EM, Ibrikci T. Journal of Computer and Communications 2014; 2: 32-37.

[4] Arif M. Biomaterials and Biomechanics in Bioengineering 2015; 2(3): 173-183.

[5] Sundar C, Chitradevi M, Geetharamani G. Journal of Computer Science 2013; 9(2): 198-206.

[6] Magenes G, Signorini MG, Arduini D. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks 2000; 3: 637 – 641.

[7] Sahin H, Subasi A. Applied Soft Computing 2015; 33: 231-238.

[8] Lichman M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California 2013; School of Information and Computer Science.

[9] Alfirevic Z, Devane D, Gyte, Gillian ML. In Alfirevic, Zarko. Cochrane Database of Systematic Reviews 2006; doi:10.1002/14651858.CD006066.

[10] Huang M, Hsu Y. Journal of Biomedical Science and Engineering 2012; 5: 526-533.

[11] Ocak H. J. Med. Syst. 2013; 37(2): 1-9.

[12] Menai MEB, Mohder FJ, Al-mutairi F. J. Med. Bioeng. 2013; 2(1).

[13] Karabulut EM, Ibrikci T. J. Comput. Commun. 2014; 2(09): 32.

[14] Czabanski R, Jezewski M, Wrobel J, Horoba K, Jezewski J. In 14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics 2008; Springer Berlin Heidelberg.

[15] Martinez AM, Kak AC. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001; 23(2):228-233.

[16] Barghout, Lauren. "Spatial-Taxon Information Granules as Used in Iterative Fuzzy-Decision-Making for Image Segmentation." Granular Computing and Decision-Making. Springer International Publishing, 2015.

[17] Grivell RM, Alfirevic Z, Gyte GM, Devane D. Cochrane Database Syst. Rev. 2010; 1.

[18] Breiman L. Machine Learning 2001; 45(1): 5-32.

[19] Kamath RS, Dongale TD, Pawar P, Kamat RK. Research journal of Pharmaceutical, Biological and Chemical Sciences 2016; 7(4): 830-836.

[20] Kamath R, Kamat R. Educational Data Mining with R and Rattle, River Publishers, Netherland, 2016, pp. 65-67.

[21] Graham W. Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery. Springer, UK, 2011, pp. 245-268.

[22] Tomas P, Krohova J, Dohnalek P, Gajdos P. 36th International Conference Telecommunications and Signal Processing 2013; 620 – 923.

[23] J. Jobson, Applied multivariate data analysis: volume II: Categorical and Multivariate Methods. Springer Science & Business Media, 2012.

[24] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," 1990.

AUTHOR BIOGRAPHY

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 pursuing Ph. D in computer science and Engineering at Walchand College of Engineering, Research 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.

Dr. Bashirahamad F. Momin is working as Associate Professor & Head, Dept. of Computer Science & Engineering, Walchand College of Engineering Sangli, Maharashtra State India. He received the B.E. and M.E. degree in Computer Science & Engineering from Shivaji University, Kolhapur, India in 1990 and 2001 respectively. In February 2008, he had completed his Ph.D. in Computer Science and Engineering from Jadavpur University, Kolkata. He is recognized Ph.D. guide in Computer Science & Engineering at Shivaji University, Kolhapur. His research interest includes pattern recognition & its applications, data mining and soft computing techniques, systems implementation on the state of art technology. He was a Principal Investigator of R & D Project titled "Data Mining for Very Large Databases" funded under RPS, AICTE, New Delhi, India. He had delivered a invited talk in Korea University, Korea and Younsea University, Korea. He had worked as "Sabbatical Professor" at Infosys Technologies Ltd., Pune. He is a Life

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 7, Issue 4, October 2017
Member of "Advanced Computing and Communication Society", Bangalore INDIA. He was a student member of IEEE, IEEE Computer Society. He was a member of International Unit of Pattern Recognition