Galvanik Skin Response signal based Cognitive Load classification using Machine Learning classifier

Abstract— **Cognitive load (CL) classification is an important research issue in the human-computer interaction paradigm. It is evident from recent research that Galvanic Skin Response (GSR) can be used to sense cognitive load. CL analysis is important for understanding the mental growth of the child and the psychology of patients going through the different traumatic situation. Inspired by such novel application, a novel model is designed. In this work, a technique has been demonstrated to measure and evaluate the level of human cognitive load for different tasks by collecting GSR from 40 student participants. The students are asked to sit for a test to solve three different tasks — reading comprehension, solving mathematics, and cracking Sudoku. Time domain features have been extracted from participant's GSR signals while these tasks are being analyzed. Some parameters such as Correlation Dimension (CD), Lempel-Ziv Complexity (LZC), Hurst Exponent (HE) and Shannon Entropy (SE) are used to analyze CL and these are used as features while accomplishing classification. Level of stress or cognitive load is strongly observed with the one-way ANOVA test and box-whisker plots. Next several machine learning algorithms are used to classify the various level of cognitive load. Using Naïve Bayes algorithm 91.5% accuracy is obtained which is better than existing method.**

Keywords*— *Cognitive load, GSR signal, features, ANOVA test, Classification.

I. INTRODUCTION

Cognitive Load (CL) is referred to as the effort being used in human working memory. Cognitive load classification make a great impact on the various sector of human life. High cognitive load is harmful to human health as well as working performance. It is important to observe the cognitive load of psychological patient condition at a real-time. Besides, it will help in intelligence exam, measuring working capability, criminal detection program etc. The theory of cognitive load was first developed in the study of problem-solving for distinguishing experts and novices [1]. There were two types of problem strategies - 'means-end strategy' and 'non-specific goal strategy' for various problems such as chess, algebra, maze etc. Cognitive load measurement has potential impact in our lives in many ways, such as safety in driving [2], aviation [3] etc. CL is needed to be kept at tolerable levels so that the people will not lose their working ability. There are various physiological signals which can be used to measure the CL. Research works have been reported using EEG [4-5] for measuring CL. Among all signals, the GSR signal is now drawing the attention of researchers because the GSR device is simple and easy to operate to collect data and the cost is comparatively low. It is interesting to consider how GSR devices work. Human bodies show some level of electrical conductivity depending on sweating glands. Thus, the GSR signal varies with the state of the sweat glands in the skin. Human sweating is controlled by the sympathetic nervous system, and skin conductance is an indication of psychological or physiological stimulation. GSR signals can be used to develop personal health care systems. Researchers [6] used GSR signals to discriminate between stressful events; solving the arithmetic problem under timed conditions and evaluate social threats considering mental and psychological conditions. Human emotion could be recognized with the cognitive load using the GSR signals [7]. Different types of arithmetic and reading tasks were also used to assess cognitive load using GSR [8]. Wernaart et al. [9] measured two types of cognitive loads (intrinsic and extraneous) during solving Sudoku and puzzles. Mundell et al. [10] evaluated human working performances in high stress from the data of low-stress work. Saitis et al. [11] used EDA and EEG both for measuring Cognitive load blind people in an indoor and outdoor challenging task. This kind of experiment will help to design intelligent assistive navigation aids for the visually impaired. The non-linear parameter e.g. CD, HE, LZC etc. are widely used to analyze the signal of EEG, ECG etc. These parameters are used to evaluate different mental stress [12] under different condition and discriminated epileptic seizure, alcoholic, normal condition, and diagnosis [13-14]. In this research, we want to analyze cognitive load in three different tasks — reading comprehension, solving math and cracking Sudoku. A wearable device, called E4 Wristband, is used to collect data. Then, time domain features are extracted from GSR to assess CL. Finally, the box-whisker plot and one-way ANOVA test are used to find the variance among these three tasks. CD, LZC, HE and SE are used to observe non-linear characteristics for GSR signal. 15 features have extracted from the GSR signal to classify the cognitive load. Random forest, KNN, Naïve Bayes classifier are used. Naïve Bayes accuracy is better than others.

The remaining part of the paper is organized as follows. In Section II, the brief description of data collection has been described. The methodology is presented in section III which includes data pre-processing, CL analysis, time domain feature, feature extraction and feature selection and classifier. The experimental results are evaluated and discussed in section IV. In Section V, a conclusion is given.

II. DATA COLLECTION

2.1 Task Description

We collected data from three common cognitive tasks (shown in Table 1), reading comprehension, solving math and cracking Sudoku. Tasks are the same for all participants.

Table 1. Task description

Task	Number of problems	Time(min) allowed
Comprehension	2	6
Math	15	10
Sudoku	2	20

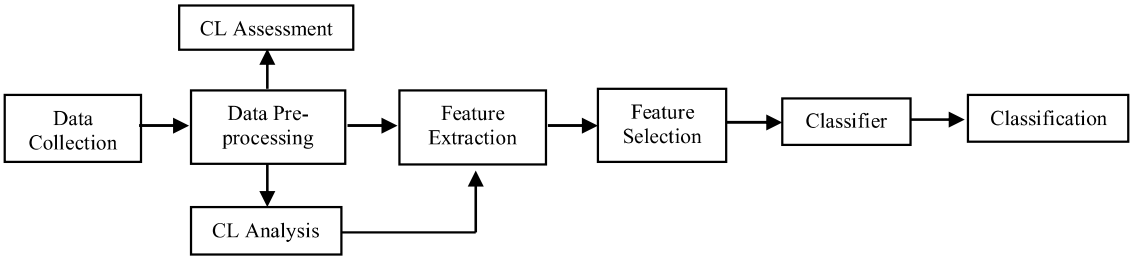

Fig 1. A Simple Diagram of Proposed Method

Five MCQ questions are provided for each participant for each task. In sequence, the first problem is easier than the second one. The second problem is easier than the third problem. Participants are required to solve all the problems following the time schedule as mentioned in Table 1. Data is recorded continuously with a 1-minute break between each task. A question paper was given for comprehension, math and Sudoku problems for each participant.

2.2 Apparatus

We have collected data from 40 right-handed participants (35 males and 5 females) with a mean age of 23±1.3. All participants are healthy and sign a consent form before their participation. They are all undergraduate University students. We have collected GSR signals using the E4 wristband [15] with its standard sampling frequency of 4 Hz for GSR. Participants have worn this band on the left hand during the task, and the right hand is used for writing the answer to the questions supplied.

III. METHOD

3.1 Data pre-processing

GSR signal varies from person to person. Therefore, we need to normalize the data for each person across the entire experiment. The normalization was done by dividing each data point of a particular subject by the sum of all data points of that particular subject [8]. Before normalization, we filter the GSR signals for smoothing the data. In Fig. 1 normalize data are plotted.

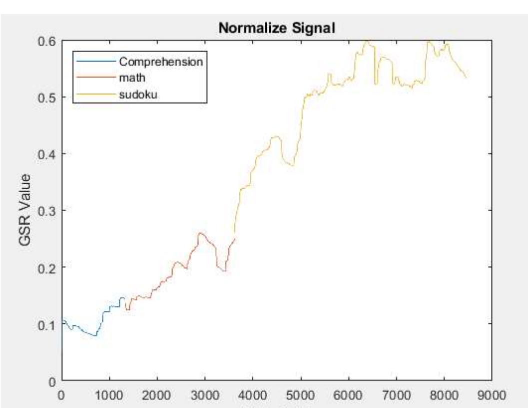

Fig 2. Normalize Signal of sample 20

3.2 Time Domain Feature

Time domain feature is calculated from GSR signal. Time domain features means here the mean value of each task of each subject. This value is used to assessment the CL level.

3.3 Cognitive Load Analysis

Non-linear parameters are widely used to analysis physical signal. CD [16], LZC [17], SE and HE [18] are used here to measure the chaotic property of signal. From Fig.2. It is clear that this signal has no periodic order. These nonlinear parameters are used to check the state of the GSR signal at different working stage.

3.4 Feature Extraction

For classify, various features have been extracted. They are mean, standard deviation, mean of first difference of filter signal, mean of second difference of filter signal, mean of first difference of normalize signal, root mean square, My pulse Percentage Rate (MYOP), entropy, log detector and Power Spectrum Density (PSD). Besides to measure nonlinear parameter value i.e. CD, LZC, HE and SE value for each subject are used as features.

3.5 Feature Selection and Classifier

It is known that extra number of irrelevant features are harmful for classification result, it may case for time consuming in training period and over fitting in result. To improve result, Chi-square feature selection method is used to find out the best features from 15 features. Three types of classifier are used to classify cognitive load, they are Random forest, KNN and Naïve Bayes.

IV. RESULT & DISCUSSION

4.1 Cognitive load Assessment

Firstly, we visualize the numerical values of the three tasks in a boxplot as shown in Fig. 3. From the box plot, we understand that the mean value of Sudoku is higher than others. It is interesting to observe that the comprehension task for all subjects are easier than mathematical tasks, and the mathematical task is easier than Sudoku. It is generally known that the GSR value increases with the increase in stress level. So, the value represents the discrimination among the tasks.

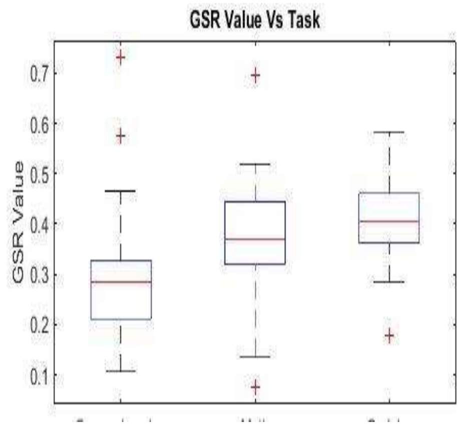

Fig. 3 : Box-whisker plot considering three tasks

Table 2. ANOVA test results for time domain features

All tasks	Comprehension and math	Math and Sudoku	Sudoku and comprehension
F=13.14 p=7.09x10^{-5}	F=9.66 p=0.0026	F=3.11 p=0.0816	F=25.94 p=2.31x10^{-6}

We also apply ANOVA for all tasks as shown in Table 2. According to the ANOVA test, the *p*-value is less than 0.05 for all tasks, but if we try to find the variance between two tasks only Sudoku and math's *p*-value is not less than 0.05. This means the level of stress for these two tasks are not statistically different. Though we believe the Sudoku is harder than mathematics, this is not significant.

4.2 Cognitive Load Analysis

Now the 40 subjects signal are considered as a one time series for the three different tasks. It is added one after another subject to observe the chaotic properties of a signal.

Table 3. Variation of parameters values for different tasks

Parameters / Task	CD	LZC	SE	HE
Comprehension	0.903	33	0.0848	0.47
Math	0.925	52	0.2133	0.44
Sudoku	0.974	99	0.5045	0.42

Table 3. Presents the CD, LZC, SE and HE for three different tasks. Here CD value is 0.903, 0.925 and 0.974 for comprehension, math and Sudoku respectively. Times series uniformity is measured using CD. When the task is easy, complexity is less and complexity is high for the difficult task. If the work is easy to sweat diffusion is less due to less stress which provides little CD value, which represents less stress in the brain and greater CD value for high stress. The degree of randomness and predictability of a time series is measured using Shannon Entropy. If the randomness is high, then entropy is high and if predictability is high then lower entropy. For high cognitive load, randomness is high (Sudoku-0.5045) and it decreases according to cognitive load (Math-0.2133, comprehension-0.0848). The LZC nature can be utilized to quantify the uniformity of binary sequence. Sequences with specific normality don't have a too huge complexity, but the complexity develops with developing long sequence and abnormality. The value of LZC is 0.33, 0.52 and 0.9 for comprehension, math, and Sudoku respectively. It is also suggested that Sudoku gives more random data than others. HE measures the long term memory of time series data. If the HE value (H) is 0<H<.5 then the time series has no correlation. From table 3, it is observed that all tasks H are in this range. H decrease or increase depends on the lag between two pair of data points. When the randomness increases the lag between pairs also increases so H decrease. It is known that, H=E+1-D, here D is correlation dimension and E is Euclidian distance. It is clear that if H increase D is decrease and vice versa. The values of H's are 0.47, 0.44 and 0.42 for comprehension, math and Sudoku respectively

4.3 Classification

This analysis is done with Inter (R) Core (TM) i5 – 6200 CPU, 2.4 GHz. 4 GB RAM, 64 bit OS Windows 10. And this analysis and classification using MATLAB 2018a and Python respectively.

Three different classifiers are used to classify the load according to tasks. It is a 3 label classification. It is already mentioned that, classifier are KNN, Random Forest and Naive Bayes. The performance parameter: n_estimator =7, criterion = 'entropy' and min_samples_leaf =2 for Random forest, 5 nearest neighbors for KNN and for Navies Bayes default parameter has been used. For KNN, K is an important key for accuracy. From 1 to 10 K value has been changed to see how accuracy change. K = 5 gives the highest accuracy for KNN. For random forest number of the tree in the forest has an impact on accuracy. A number of tree is changed from 1 to 15, highest accuracy is gotten when tree's number is 7. 25% and 75% data are splitted for test and train. Here the correlation matrix is presented.

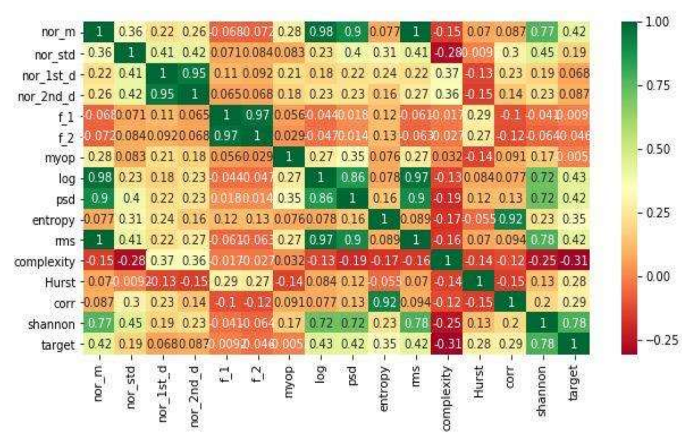

Fig. 4. Correlation matrix with heat map among the features

Correlation matrix with heat map present the correlation between the features and target values with features. From this map it is easily understood the highly correlated feature with the target value. Automatically 10 best features are selected from 15 features using Chi-square method. The accuracy of KNN, Naïve Bayes, and Random forest without nonlinear features 75%, 91.5% and 87.5% respectively. The proposed approach is compared with Nourbakhsh [19], where authors have used SVM, Naive Bayes for binary and 4 class classification. It has been seen that [19] the accuracy have crossed 80% in some case of binary classification but for 4 class classification, their value is always lower than 65%. Hence it can be claimed that the proposed approach is more efficient and applicable for it is better accuracy.

V. CONCLUSION

GSR signal can be used as an efficient indicator of human stress. Ever since the birth of 'nonlinear science' chaotic property of physiology, biomedical engineering, and theoretical biology are searching for meaningful chaotic parameters in physiological processes. In this research, Cognitive Load (CL) is measured using GSR signal for three task reading comprehension, solving math and cracking Sudoku. CL is proportional to complex and hard work. If the problem is hard CL will increase. Next, various parameters are applied such as Correlation Dimension (CD), Shannon entropy (SE), Lempel-Ziv Complexity (LZC) and Hurst exponent (HE) to observe the nonlinear property of time series. CD, SE, and LZC give the value of the same pattern that is when the problem more complex and hard then those parameters value also increased. CD and SE are used to measure the variability, randomness, and LZC is used to measure the uniformity in a binary sequence. But HE measures the long term memory of time series and its' result is reversed. HE results represent that the data are not correlated because they are $0 < HE < 0.5$ range. In the next section to classify cognitive load 3 well-known algorithm are used that are Random forest, Navies Bayes and KNN. After using Chi-square feature selection method Naïve Bayes gives 91.5% accuracy. Random Forest and KNN accuracy are 87.5% and 75% respectively. Feature selection method helps to improve classification accuracy which represent a better result than the present method. In future this research will implement in a real time application.

REFERENCES

[1] J. Sweller, Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science. 12 (1988): 257-285.

[2] J. Engstrom, E.Johansson, J. Ostlund. Effects of Visual and Cognitive Load in Real and Simulated Motorway Driving. Transportation Research Part F, 8, 2 (2005), 97-120.

[3] K. Huttunen, H. Keränen, E. Väyrynen, R.Pääkkönen, T. Leion. Effect of cognitive load on Speech Prosody in Aviation: Evidence from Military Simulator Flights. Applied Ergonomics, 42(2011), 348-357.

[4] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd,G. A.Preston&C. T. Silva. A User Study of Visualization Effectiveness Using EEG and Cognitive Load. IEEE Symposium on Visualization, 30 (2011), 791-800

[5] P. Antonenko, F. Paas, R.Grabner, T. Gog. Using Electroencephalography to Measure Cognitive Load. EducPsychol Rev (2010) 22:425-438.

[6] C. Setz, B. Arnrich, J.Schumm, R. La Marca, G.Tröster. Discriminating Stress from cognitive Load using a Wearable EDA Device. Technology, 14, 2 (2010), 410-417.

[7] C. L. Lisetti, F. Nasoz. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. EURASIP Journal on Applied Signal Processing 2004:11, 1672-1687.

[8] N. Nourbaksh, Y. Wang, F. Chen, R. A.Calvo. Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and

[9] Reading Tasks. OZCHI' 12, November 26-30, 2012, Melbourne, Victoria, Australia

[10] G.Wernaart. Cognitive load measurement: different instruments for different types load? Master Thesis. Utrecht: Faculty of social and BehaviouralSciences (2013).

[11] C. Mundell, J. P.Vielmaand T. Zaman. Predicting Performance under stressful conditions using Galvanic Skin Response, 2016.

[12] C. Satis, M. Z. Parvez, K. Kalimeri. Cognitive Load assessment from EEG and peripheral Biosignals for the Design of Visually Impaired Mobility Aids. Hindawi, Wireless Communications and Mobile Computing Volume 2018, Article ID 8971206, 9 pages.

[13] K. Natarajan, R. Acharya U, F. Alias, T. Tiboleng and S. K. Puthusserypady. Nonlinear analysis of EEG signals at different mental states. BioMed Central.

[14] S. Geng, W. Zhou, Q. Yuan, D. Cai, Y. Zeng. EEG non-linear feature extraction using correlation dimension and Hurst exponent. Neurological Research 2011 VOL. 33 NO. 9.

[15] https://www.empatica.com/get-started-e4.

[16] P. Grassberger, I. Procassia, Measuring the strangeness of strange attractors, Physica D 9 (1983) 189—208.

[17] A. Lempel, J. Ziv, On the Complexity of Finite Sequences , IEEE Trans. on Information Theory, January 1976, p. 75–81, vol. 22.

[18] H. E. Hurst, (1951). "Long-term storage capacity of reservoirs". *Transactions of American Society of Civil Engineers*. 116: 770.

[19] N. Nourbakhsh, F. Chen, Y. Wang, and R. A. Calvo. 2017. Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference. ACM Trans. Interact. Intell. Syst. 7, 3, Article 12 (September 2017), 20 pages.