A non-EEG Biosignals Dataset for Assessment and Visualization of Neurological Status

Javad Birjandtalab, Diana Cogan, Maziyar Baran Pouyan, Mehrdad Nourani
Quality of Life Technology Laboratory
The University of Texas at Dallas, Richardson, TX 75080
{birjandtalab, diana.cogan, mxb112230, nourani}@utdallas.edu

***Abstract*—Neurological assessment can be used to monitor a person's neurological status. In this paper, we report collection and analysis of a multimodal dataset of Non-EEG physiological signals available in the public domain. We have found this signal set useful for inferring the neurological status of individuals. The data was collected using non-invasive wrist worn biosensors and consists of electrodermal activity (EDA), temperature, acceleration, heart rate (HR), and arterial oxygen level (SpO2). We applied an efficient non-linear dimension reduction technique to visualize the biosignals in a low dimension feature space. We could cluster the four neurological statuses using an unsupervised Gaussian Mixture Model. The experimental results show that our unsupervised method can accurately separate different neurological statuses with an accuracy of greater than 84%.**

***Index Terms*—biosensor signal processing, data visualization, unsupervised machine learning, neurological status assessment.**

I. Introduction

A. Background

Recent developments in wearable sensors allows us to monitor different human physiological activators in an efficient, non-invasive, and inexpensive manner. Monitoring and assessing of neurological activity is traditionally done using surface or implanted electrodes to collect EEG signals. However, EEG based monitoring is neither comfortable nor practical during normal daily activities. We investigated the feasibility of monitoring an individual's neurological status using extracerebral biosignals collected with a comfortable and effective wrist-worn platform.

Wrist worn devices are easy to use and portable, which makes them suitable health monitoring tools for daily life. Heart rate, skin temperature and wrist movement can all be screened by state of the art smart watches [1]. Moreover, changes in skin conductance, which indicate changes in the sympathetic nervous system, can be monitored at the wrist [2]. The amount of oxygenated hemoglobin can be measured by a wrist worn device using a finger cuff [3]. Several companies are working to develop devices capable of reading SpO2 accurately at the wrist using reflectance oximetry. These companies include Empatica with their E4 wristband, and Samsung with their Gear Smartwatch [4]; as of this writing, there is no device on the market with that capability.

B. Proior work

Recently various ranges of physiological signals are introduced to monitor neurological activity which can be categorized as cerebral and extra-cerebral biosignals. Authors used a combination of functional near-infra-red spectroscopy (fNIRS) and electroencephalogram (EEG) to monitor neural activities by measuring oxygenated and deoxygenated hemoglobin concentrations and electrical activity [5]. Although Such cerebral biosignals are accurate and reliable, they cannot be used in wearable devices. On the other hand, extra-cerebral biosignals are potentially suitable to be implanted for health monitoring in daily life and are shown to be informative indicators for neurological status monitoring. Researchers used electrodermal activity, ECG, heart rate, and respiratory rate for emotional reactivity monitoring in individuals with suicidal behaviors [6]. In [7], the authors analyzed a set of physiological signals like electrodermal activity (EDA), temperature (Temp), accelerometer (Acc), heart rate (HR), and arterial oxygen saturation (SpO2) that are monitored comfortably by a wrist worn device for supervised neurological status monitoring.

C. Key Contribution

To the best of our knowledge there is no publicly available dataset of multi-modal wrist worn based biosignals for neurological status assessments. In this paper, we introduce our dataset by explaining our experiment design and data collection set up in detail. Our intent is to encourage other researchers to extract new knowledge out of these prerecorded biosignals from different neurological statuses.

Data visualization can provide a better understanding of the effect of neurological statuses on different biosignals. So, we visualized all the biosignals collected from both wrist worn devices in a visible feature space. Moreover, labeled data is not always available in real life. Hence, we clustered different neurological statuses using GMM clustering without providing any prior knowledge. The excellent data separation we obtained using unsupervised learning indicates that our set of biosignals is sufficient for distinguishing among neurological statuses.

II. Data collection set-up

A. Design of Experiment

The experimental procedures involving human subjects described in this work were approved under UTD IRB # 12-29 by the Institutional Review Board at the University of Texas at Dallas, Richardson, Texas, USA. We collected data from 20 college students (14 males, 6 females). Our goal was to distinguish responses to the different types of stresses shown in Figure 1. We designed the following experiment and asked our volunteers to perform the following tasks in order.

1) *First Relaxation*: five minutes.

2374-7390/16 $31.00 © 2016 IEEE
DOI 10.1109/SiPS.2016.27

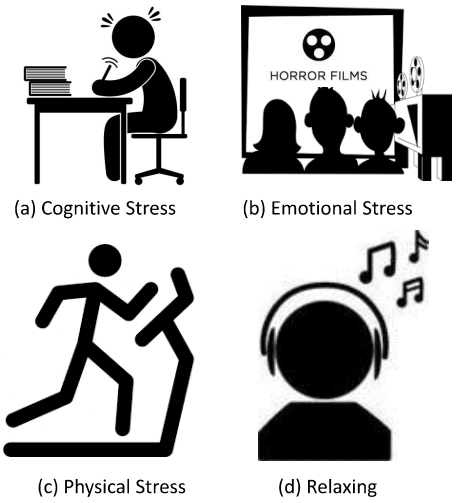


Fig. 1. Different Neurological statuses.

2) *Physical Stress*: Stand for one minute, walk on a treadmill at one mile per hour for two minutes, then walk/jog on the treadmill at three miles per hour for two minutes.
3) *Second Relaxation*: five minutes.
4) *Cognitive Stress*: Count backwards by sevens, beginning with 2485, for three minutes. The Stroop test consisted of reading the names of colors written in a different color ink, then saying what color the ink was. In both tests, the volunteer was alerted to errors by a buzzer.
5) *Third Relaxation*: five minutes.
6) *Emotional Stress*: The volunteer was told he/she would be shown a five minute clip from a horror movie in one minute. After the minute of anticipation, a clip from a zombie apocalypse movie, *The Horde* was shown.
7) *Forth Relaxation*: five minutes.

During the four relaxation sessions the subject was asked to sit quietly and listen to a portion of *Binaural*, which was composed for use in mediation, and hence is quiet, soothing music. The objective of the relaxation sessions was to establish a baseline for the physiological metrics we were measuring. That way we could see how each metric changed during the three tasks and what the sensitivity and specificity of each metric was. The data from subject 1 is shown in Figure 2.

B. Data from Affectiva

Figure 3 (a) presents an Affectiva Q Curve [8], which records EDA, Temp, and 3D Acc; EDA represents electrical changes measured at the surface of the skin that are caused by changes in the sympathetic nervous system. Changes in the sympathetic nervous system are caused by events such as emotional arousal, cognitive workload, or physical effort. Sweating caused by these events may not be sensed by an individual but the electrical conductance of the skin increases. Skin temperature measurements are required for proper interpretation of the EDA measurement [8].

TABLE I
SUBJECTS INFORMATION.

Subjects ID	Age	Gender	Height [cm]	Weight [kg]
1	30	M	177	94
2	28	M	172	68
3	28	M	177	91
4	22	M	167	58
5	30	M	182	82
6	30	F	167	58
7	33	F	157	90
8	27	M	182	64
9	25	M	177	68
10	23	M	180	64
11	26	M	170	71
12	32	F	162	53
13	20	F	167	64
14	19	F	160	50
15	23	M	165	64
16	24	M	180	54
17	23	M	167	57
18	23	M	177	64
19	22	M	167	64
20	24	F	160	44

C. Data from Nonin

Figure 3 (b) presents a Nonin 3150 Wireless WristOx2 Oximeter, which records HR and SpO2 using pulse oximetry [3][9]. SpO2 is an estimate of the amount of oxygen in the blood calculated by oxygen saturation. The oxygen saturation is the ratio of oxygenated hemoglobin (HbO_2) to deoxygenated hemoglobin (Hb) in the blood. Oxygenated and deoxygenated hemoglobin have bright and dark red colors, respectively. The pulse oximeter integrated into the Nonin WristOx2 finger cuff emits red and infrared light and measures the ratio of infrared and red light transmitted through the finger to estimate oxygen saturation [10]. Normal values of SpO2 range between 95% and 100%.

The Nonin WristOx2 returns a flag when it does not receive valid data. These drop outs may be caused by a loose or displaced finger cuff. We observed that drop outs were rare and brief (less than five samples at a time). Because both HR and SpO2 changes were slow relative to the lengths of dropouts, we replaced missing (invalid) HR and SpO2 records with the last valid reading.

The Affectiva and Nonin devices were time synced to each other by synching each of them to the same laptop. The laptop ran software designed to interface with the two wrist worn devices. The nVision package interfaces to the Nonin WristOx2 and the Q package interfaces to the Affectiva Q curve. Each device was time synced to the laptop daily using the appropriate function of its respective software package [8][11].

The raw data from both devices and read-me files, all in Comma Separated Values (CSV) format, are now available at www.utdallas.edu/∼nourani/bioinformatics /biosensor_data for public academic use. Table I lists demographic information of all twenty subjects including age, gender, weight, and height. There are two data files for each volunteer because the Q and Nonin data were collected at

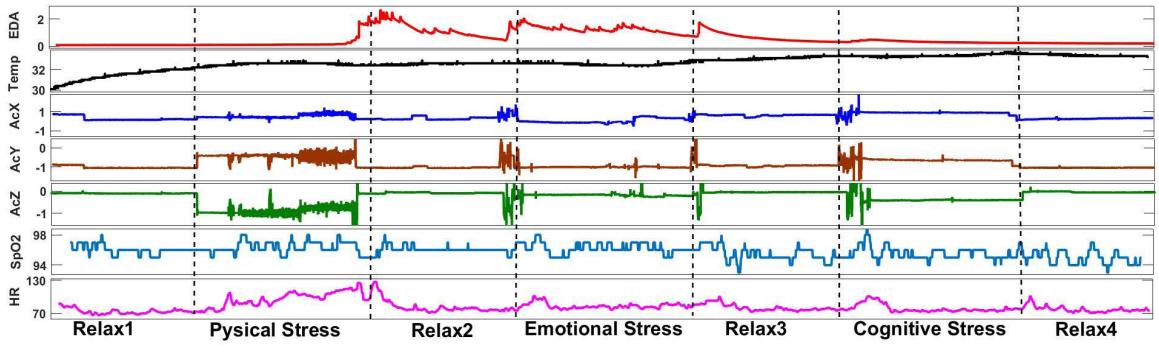


Fig. 2. Raw data from multi-modal biosignals for subject 1.

(a) (b)

Fig. 3. (a) The wrist worn Affectiva collects EDA, Acc, and (b) the Nonin 3150 Wireless WristOx2 collects HR and SpO2.

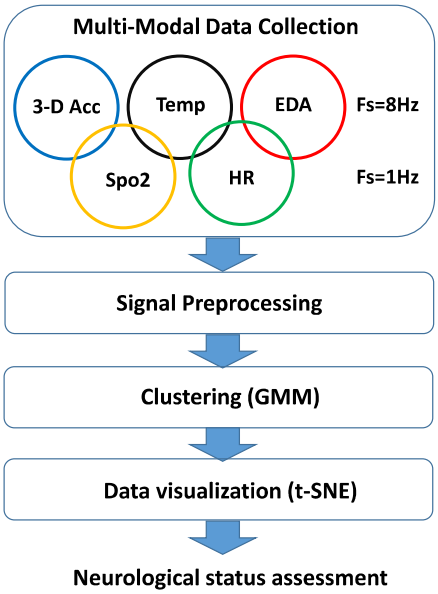


Fig. 4. Overall view of data collection process.

different frequencies and so could not be readily combined. Note that, the experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board at the University of Texas at Dallas. We used the same data collection platform at a hospital epilepsy monitoring unit to successfully detect the seizures of 6 out of 11 patients in another study [12].

III. Neurological Status Assessment

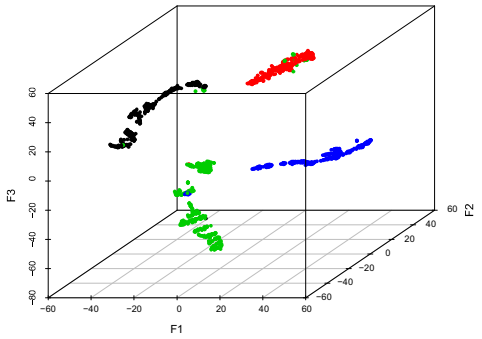
Figure 4 shows the overall view of our proposed approach. First, we collected Non-EEG biosignals using two wrist worn devices including electrodermal activity (EDA), temperature, acceleration, heart rate (HR), and arterial oxygen level (SpO2). Then, we performed the data preprocessing by removing the invalid data points. Next, we grouped the four neurological statuses using Gaussian Mixture Model (GMM) clustering for each subject. Finally, we applied an state-of-the-art dimension reduction technique to visualize all the biosignals in 2D/3D feature space.

A. Clustering Method

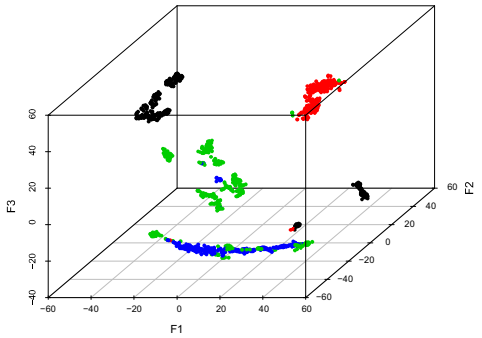
We downsampled the signals collected by the Affectiva eight times so that all signals have the same time resolution of 1 Hz (i.e. each data point corresponds to one second of data). Since biosignals in this dataset have different modalities and the ranges of the features are different, we standardized the features by dividing the value of each feature by its variance. The ”Relaxation” class has more data than do the other three classes, so we use data from only the ”First Relaxation” phase in order to balance our dataset. So, we analyzed approximately 1400 frames (1 sec windows) of all neurological states per subject. In order to cluster the neurological statuses in an unsupervised manner, we applied a GMM (Gaussian Mixture Model) algorithm [13]. GMM finds the maximum likelihood of statistical model and find different clusters in an unsupervised manner. Note that GMM is a fast algorithm for learning mixture models [14]. Like any other human biosignals, non-extracerebral signals response to neurological status varies from subject to subject. That is the key reason that we applied the neurological statuses clustering in a subject-specific manner. We used labels only for evaluation purposes; clustering is done without prior knowledge of the neurological statuses.

B. Data Visualization

Since it is not possible to visualize more than three dimension of space, we mapped the data form original space into 2D and 3D space which both have their own pros and cons. It is intrinsically easier to track data points in 2D feature space compared with 3D. However, 3D visualization make it possible to investigate the data points from different



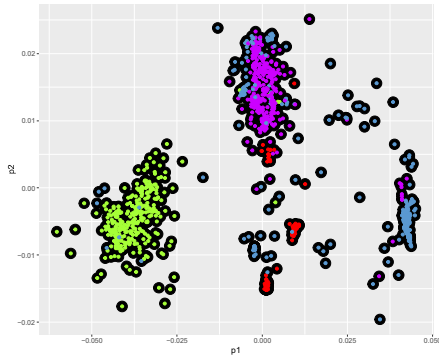
(a) Sub10, 3D t-SNE



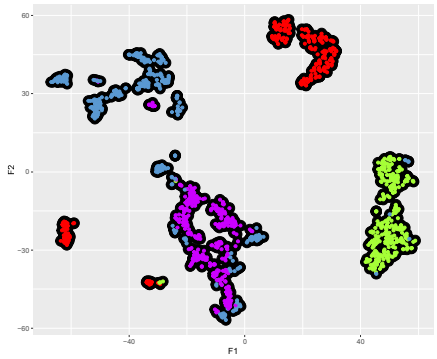
(b)Sub14, 3D t-SNE

Fig. 5. 3D Visualization of Non-EEG biosignals in different Neurological Status: relaxation(red), emotional stress (blue), cognitive stress(purple), physical stress (green).

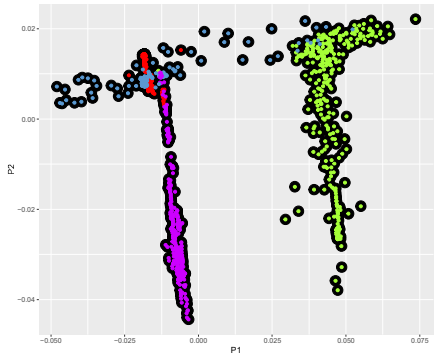
angles. Representation of multimodal signals (those with more than 3 features) in a 2D or 3D feature space allows us to visualize how individual features vary among different neurological statuses. One possible solution is dimension reduction, which maps high dimension signal information into a lower dimension feature space. Principal Component Analysis (PCA), a popular dimension reduction technique, has been widely used for this purpose. However, it works based on finding principal directions with the maximum variance in the data [15] and is not able to preserve neighbor distance between data points. In this work, we propose a state-of-the-art nonlinear dimension reduction technique known as t-distributed Stochastic Neighbor Embedding (t-SNE) [16]. The t-SNE calculates the pairwise similarities of the data points in high and low dimension and minimizes their joint distribution. By minimizing the joint distribution, the relationship among neighboring data points is preserved. The t-SNE estimates the corresponding two dimensions such that data points which are similar/different, are mapped near/far in the lower dimension space [17].



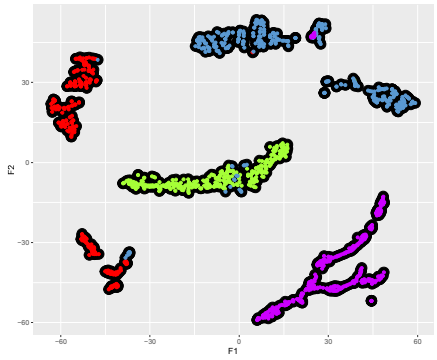
(a) Sub10, 2D PCA



(b) Sub10, 2D t-SNE



(c) Sub14, 2D PCA



(d) Sub14, 2D t-SNE

Fig. 6. 2D Visualization of Extracerebral biosignals in different Neurological Status: relaxation(red), emotional stress (blue), cognitive stress (purple), physical stress (green).

TABLE II
CONFUSION MATRIX AND STATISTICAL METRICS (IN%) FOR DIFFERENT NEUROLOGICAL STATUSES AVERAGED OVER ALL 20 SUBJECTS.

	Proposed Method				[7]			
Class	Relaxation	Physical	Emotional	Cognitive	Relaxation	Physical	Emotional	Cognitive
Relaxation	94.2	0.1	5.4	0.3	98.7	0	0.8	0.9
Physical stress	1.8	90.9	3.8	3.5	0	99.1	0.8	0
Emotional stress	13.4	5	65.8	15.8	1	0.9	98.4	0
Cognitive stress	1	1.4	10.2	87.3	0.3	0	0	99.1
Sensitivity	85.4	93.3	77.2	86.4	98.7	99.1	98.4	99.1
Specificity	93.6	91.1	70.2	86.4	98.9	98.7	99	98.7
Precision	94.2	90.8	65.8	87.4	99.5	99.1	96.2	99.1
Accuracy	84.6				98.8			

IV. EXPERIMENTAL RESULTS

We applied 2D and 3D t-SNE to the 7 biosignal features (Accelerometer has 3 features since it is recorded in 3D). Figure 5 shows the biosignals mapped into a 3D feature space. These three dimensions (F1, F2, and F3) are not any of the original dimensions. Instead, they are three new dimensions, which are extracted by similar/dissimilar data point in high-dimension space. In order to compare the performance of t-SNE with conventional PCA, we ran both techniques on two subjects. Figure 6 shows that t-SNE outperforms PCA in terms of grouping the different neurological statuses since it is able to preserve the distances between neighbors in the dimension reduction process. Table II shows the effectiveness of our proposed method in terms of a confusion matrix, of sensitivity, specificity, and precision. We compared our unsupervised clustering approach with a supervised neural network classifier [7]. Although we did not use any label information in this work, our results show that our technique is experimentally comparable with the supervised method. Our results indicate that the biosignals were appropriately selected and may be used to cluster different neurological status even without prior knowledge. We use labels only to evaluate the effectiveness of our clustering methodology. Note that, we ran dimension reduction only for purpose of data visualization; all the clustering analyses were performed in the original feature space.

V. CONCLUSION

The dataset we present here was collected by monitoring a set of biometrics (HR, SpO2, EDA, Temp and Acc) that provide information on changes in neurological statuses. The data was collected using two wrist worn devices and was analyzed using a state-of-the-art non-linear dimension reduction technique to visualize the biosignals in a low dimension feature space. Unsupervised GMM clustering was used to confirm that our set of biosignals is able to distinguish among 4 neurological states. We have made our data in www.utdallas.edu/∼nourani/Bioinformatics/Biosensor_Data for public use by researchers in this field.

REFERENCES

[1] Empatica, inc., e4 wristband. [Online]. Available: www.empatica.com, AccessedSeptember27,2015

[2] M.-Z. Poh, "Continuous assessment of epileptic seizures with wrist-worn biosensors," Ph.D. dissertation, Massachusetts Institute of Technology, 2011.

[3] Nonin puresat pulse oximetry technology. [Online]. Available: www.nonin.com/puresat,AccessedFebruary10,2016

[4] Samsung inc., samsung gear s. [Online]. Available: www.samsung.com/global/microsite/gears/gearsfeatures,AccessedSeptember15,2015

[5] H. O. Keles, R. L. Barbour, H. Aghajani, and A. Omurtag, "Multi-modality mapping approach for evolving functional brain connectivity patterns: A fnirs-eeg study," in *Biomedical Optics*. Optical Society of America, 2014, pp. BT5B–2.

[6] A. M. Amiri, M. Abtahi, A. Rabasco, M. Armey, and K. Mankodiya, "Emotional reactivity monitoring using electrodermal activity analysis in individuals with suicidal behaviors," in *2016 10th International Symposium on Medical Information and Communication Technology (ISMICT)*, March 2016, pp. 1–5.

[7] D. Cogan, M. B. Pouyan, M. Nourani, and J. Harvey, "A wrist-worn biosensor system for assessment of neurological status," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, pp. 5748–5751, 2014.

[8] Affectiva inc. [Online]. Available: www.affectiva.com/q-sensor,2013

[9] Turner medical inc. [Online]. Available: www.turnermedical.com

[10] K. Minolta, "Basic understanding of the pulse oximeter. how to read spo2," *Konica Minolta Sensing, Inc. Ramsey*, 2006.

[11] Nonin medical inc., "wristox2 3150 wireless oem pulse oximeter. [Online]. Available: www.nonin.com/OEMSolutions/WristOx23150-OEM, AccessedJune9,2013

[12] D. Cogan, J. Birjandtalab, M. Nourani, J. Harvey, and V. Nagaraddi, "Multi-biosignal analysis for epileptic seizure monitoring," *International Journal of Neural Systems*, 2016.

[13] G. Xuan, W. Zhang, and P. Chai, "Em algorithms of gaussian mixture model and hidden markov model," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 1. IEEE, pp. 145–148, 2001.

[14] J. Birjandtalab, M. B. Pouyan, and M. Nourani, "Unsupervised eeg analysis for automated epileptic seizure detection," in *First International Workshop on Pattern Recognition*. International Society for Optics and Photonics, 2016, pp. 100 110M–100 110M.

[15] N. Kambhatla and T. K. Leen, "Dimension reduction by local pca," *Neural Computation*, vol. 9, no. 7, pp. 1493–1516, 1997.

[16] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[17] J. Birjandtalab, M. B. Pouyan, and M. Nourani, "Nonlinear dimension reduction for eeg-based epileptic seizure detection," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 595–598, 2016.