INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

Human Dysthymia Detection Through Biological Processes Using Machine Learning.

Ms. M.S.Namose[1], Ms.Pooja Bhatia[2], Ms.Devyani Ubhale[3], Mr.Yash Kalwani[4], Mr.Prajyot Bhadange[5]

[1]Professor, [2,3,4&5] Students

Department of computer science

JSPM Narhe Technical Campus Pune, Maharashtra. India

ABSTRACT

Over a third of patients suffering from epilepsy continue to live with recurrent disabling seizures and would greatly benefit from personalized seizure forecasting. While electroencephalography (EEG) remains most popular for studying subject-specific epileptic precursors, dysfunctions of the autonomic nervous system, notably cardiac activity measured in heart rate variability (HRV), have also been associated with epileptic seizures. This work proposes an unsupervised clustering technique which aims to automatically identify preictal HRV changes in 9 patients who underwent simultaneous electrocardiography (ECG) and intracranial EEG presurgical monitoring at the University of Montreal Hospital Center. A 2-class k-means clustering combined with a quantitative preictal HRV change detection technique were adopted in a subject- and seizure-specific manner. Results indicate inter and intra-patient variability in preictal HRV changes (between 3.5 and 6.5 min before seizure onset) and a statistically significant negative correlation between the time of change in HRV state and the duration of seizures (p<0.05). The presented findings show promise for new avenues of research regarding multimodal seizure prediction and unsupervised preictal time assessment.

Keywords: HRV, machine Learning, stress detections, K-means clustering, Data acquisition, svm.

I. INTRODUCTION

Stress is the difficulty of an organism to maintain its homeostasis, often induced by external stimuli that cause mental or physical imbalance. When anyone is exposed to a stress condition, the autonomic nervous (ANS) system is triggered which results in the suppression of the parasympathetic nervous system and it further activates the sympathetic nervous system. This reaction known as the fight-or- flight response and it involve physiological appearances like: vasoconstriction of blood vessels, increased blood pressure, increased muscle tension and a variation in heart rate (HR) and heart rate variability

The fight-or- flight response can involve physiological manifestations such as: vasoconstriction of blood vessels, increased blood pressure, increased muscle tension and a change in heart rate (HR) and heart rate variability (HRV). Between these, HRV has become a regular metric for the valuation of the state of body and mind, with many signs derived from HRV. Being regularly used for recognizing mental stress or absence thereof Heart rate variability or HRV is the physiological phenomenon of the variation in the time interval between consecutive heartbeats in milliseconds. Known as the RR interval (RRI). A best HRV points to healthy physiological function, adjust- capability and flexibility. More or bigger HRV (beyond normal) point to a disease or irregularity. Less HRV on the other side points to a reduced regulatory capability and is seen as a sign of stress, anxiety, and a number of other health difficulties.

Identifying stress has been the focus of much research as an increasing body of evidence suggests a rising prevalence of stress-related health conditions associated with the stressful contemporary lifestyle. A significant portion of contemporary stress is due to occupational stress. Professional stress can lead to prolonged health conditions such as heart disease as well as can tend to immediate disastrous effects such as accidents, injury and death too. People are susceptible to acute stress due to the sensitive nature of their work. It is imperative that mental stress in some people is monitored to prevent in- jury to personnel or the public. In our work, our objective is to influence HRV statistics and unsupervised machine learning systems, to notice mental stress in different people. With the growth of modern machine learning, deep Learning methods, these methods have been utilized in the analysis and study of heart rate variability. Machine learning and deep learning methods have earlier been used with HRV and electrocardiography (ECG) data for several applications like: weakness and stress detection, student stress prediction, heart failure or catastrophe detection, cardiac arrhythmia classification. The majority of former arts, though, are supervised.

As compared to EEG recordings, electrocardiography (ECG) signals are non-invasive, less visible/stigmatizing (compared to an EEG headset) and can be measured continuously with very little physical interference with patients' normal daily activity. For these reasons, HRV has-been studied in patients with epilepsy for various applications including seizure prediction and detection [7,12]. The aforementioned studies have extracted well established time and frequency-domain features from HRV signals which were incorporated into supervised algorithms for machine learning classification. However, the main challenge encountered while applying supervised algorithms is the requirement of predetermined labeling of samples (interictal, preictal, ictal, postictal). While the ictal and postictal (post- seizure) states can be visually confirmed by expert neurologists, no standard preictal time (period preceding a seizure) has been yet established. Thus, previous studies employing supervised epileptic state classification based on HRV have conventionally assumed preictal periods ranging between 10 and 15 minutes. Since preictal time can vary in a subject-specific manner, a generic preictal time could not reflect the physiological subject-specific preictal time and may account for the poor classification performances reported for certain patients [14]. Furthermore, recent reviews have agreed upon the importance of considering subject- and seizure-specific changes in both ECG and EEG while predicting and detecting seizures.

II. LITERATURE REVIEW

Researchers have discovered the uses of machine learning algorithms for stress prediction by forming the patterns of questionnaires dataset and another dataset. U. Reddy et al. [1] Stress issue is a typical issue for the individuals who are working in IT proficient industry. Individuals working in that industry face issues like change in their way of life and work societies, which has overstated the danger of stress. As per their paper they utilizing OSMI (open sourcing Mental Illness) Survey dataset 2017 from tech industry.[1] Different Machine learning strategies utilized (boosting, packing, Decision trees, Random timberland) and various kinds of characteristics they utilizing like sexual orientation, age, family ancestry e- produce medical advantages and so on. As per study they found that 75% individuals working in tech organization were marginally at the danger of creating pressure

G. Harrison et al. [2]
In the realm of globalization, innovation is expanding quickly and we as a whole are encompassed by advancements which are valuable in our everyday life and one of the most fundamental innovations is the cell phone which has become a significant piece of everyone's life. As indicated by Occom's 2017 records, 94% of grown- ups in the UK have a cell phone; and more than three fourth of those are advanced mobile phones .Deloitte's portable purchaser study (2016) proposes 33% of Smartphone clients don't really make voice calls by any stretch of the imagination. Rather, our telephones are utilized as portable PCs, for browsing email, shopping internet, getting to news, downloading music and recordings, taking part in web-based social networking, requesting nourishment, taking a gander at maps… the rundown goes on. We rely upon the web for each solution to our inquiries. We are reliant on innovation or simply dependent on this. We feel truly awkward and fragmented when we overlook our cell phone. Truth be told, as indicated by inquire about has demonstrated that a couple of individuals experience basic weight and uneasiness when they are disengaged from their phones.

Samrat Qaraqe et al. [3] As indicated by this paper, these sensor signs can distinguish feelings of anxiety whereas physiological detecting gadgets are utilized to gather these signs. Signal pre-processing must be executed to select valuable highlights from accumulated signs. To choose valuable highlights from the accumulated sign, signal pre-handling must be actualized.
When the highlights are fearless, the AI calculations can be applied to assemble the grouping model. The accelerometer information can identify development and furthermore be considered as a measure to see physical exercises. Pulse varieties, galvanic skin reaction and skin temperature can reflect self-sufficient sensory system movement so their highlights are extremely helpful for anticipating the feeling of anxiety of a person. Bolster vector machines, choice trees and Random woods are instances of powerful order calculations for mental pressure identification. Besides warm imaging innovation just as wearable sensors can be utilized for discovery. One of the promising procedures to foresee a singular feeling of anxiety is profound learning.

Huijie Lin et al. [4]. As indicated by the report youngsters somewhere in the range of 8 and 11 years, will confront different psychosocial stressors, they revealed that understudies with lower monetary status, who get cash from their companions, cause clashes among relatives and furthermore lead a worry toward understudies. Stress was found to massively influence understudies' wellbeing and physical movement including lack of sleep, weakness, and uneasiness which cause low fixation and lead to sorrow. Sandhi and Asrabadi (1994). In an investigation in the midst of college understudies announced achiness to visit the family, racial segregation, language hindrances, as the primary driver of worry among understudies. Understudies are likewise discovered hard to settle in another workplace uniquely who are away from home and they observe nature as a factor of pressure. Due to the non-accessibility of darlings and absence of understudies embraces the outrageous degree of stress. As the understudies originate from different social foundations so they feel isolated from their way of life and feeling have them a culture stun. They stay tangled between odd estimations of their way of life and new standards where they are living, which gets clashes and makes them. Additionally, verbal and nonverbal correspondence, methods of conduct contrast from culture to culture likewise turn into the explanation of worry in understudies. Subsequently, correspondence issue in new condition was likewise detailed in their investigation which lead to worry among the understudies

III. METHEDOLOGY

A. Data Description

1) Continuous ECG recordings from 9 patients, implanted with intracranial EEG (iEEG) macro-electrodes, who underwent invasive presurgical monitoring at the University of Montreal Hospital Center (CHUM) were retrospectively

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

analyzed in this study. All patients were diagnosed with focal epilepsy and underwent continuous intracranial video-EEG, and simultaneous ECG monitoring (~ 2 weeks). Patients who had at least 6 lead electro-clinical seizures were selected. Lead seizures were identified as seizures occurring at least one hour after a preceding seizure. The project's protocol was approved by the University of Montreal Hospital Research Center's ethics committee.

2) Data

A total of 100 lead seizures were collected, annotated by an expert epileptologist, and segmented for further processing. Only seizures which lasted at least 10 seconds, and which were accompanied by clinical manifestations were included. Electrical onset and termination were visually annotated using intracranial video-EEG recordings. iEEG is currently the gold standard for seizure annotation since it has less noise, allows for the measure of high frequencies and displays the onset earlier since the seizure begins in the brain and then propagates to the scalp. The ECG recordings were sampled at 2000 Hz and segmented to samples beginning 10 minutes before seizure onset (in line with previous investigations) [8- 11] and ending 10 seconds after seizure onset.

B. Preprocessing and HRV feature extraction

1) Data preprocessing

Non-overlapping raw ECG segments of duration equal to610 seconds (10 minutes before and 10 seconds after electrical iEEG onset) were first visually inspected to identify possible loss of signal and large artifacts. Recordings including such events were excluded from the study. The remaining recordings were subsequently preprocessed using a threefold technique proposed in [15] which removes baseline wandering, power line interferences and impulse artifacts. Backward and forward band-pass (5-30 Hz) filtering was applied to remove baseline wandering and power line interference. Median filtering with a non-overlapping 60 sec moving window was then applied to remove impulse artifacts.

2) R peak detection and HRV correction

HRV was calculated by first detecting all QRS complexes (heart beats) in the ECG signal and then measuring the time delay between adjacent heart beats. The most commonly used R peak detection algorithm [16-17] was adopted to detect QRS complexes. Briefly put, the Pan & Tompkins algorithm includes preprocessing steps to enhance R peaks and relies on an adaptive threshold in line with a regularization technique to accurately identify the locations of R peaks. The accuracy of the algorithm was visually verified for each recording. Based on detected R peaks' locations, the HRV was calculated by computing the difference (in ms) between adjacent R peaks. This difference is also referred to as the RR interval (RRI).

a. K-means Clustering

The traditional K-means clustering algorithm was utilized to cluster the data or records transformed into 18 engineered features. We employed the K-means clustering algorithm implemented in the Scikit-Learn Python library with k =2 and all other parameters set as default. However, common distance metrics, such as the Euclidean distance used in K- means, are not useful in finding similarity in high-dimensional data. As a result, we explored auto encoders (AEs) as a way to compress the raw samples into a lower dimensional latent space (2D in our case), and then search for Patterns or clusters within the compressed (or encoded) data. We discuss auto encoders in detail in the following subsection.

b. Auto Encoders

Auto encoders are neural nets that ingest a sample, *x*, and attempt to reconstruct the sample at the output. When the auto encoder involves a hidden layer, *h*, that is of lower dimension than *x*, it is called an under complete auto encoder. The idea is to encode the data into a lower dimensional, *h*, which contains the maximum noticeable structures of the data. The learning process of an auto encoder involves minimizing a loss function, *J*:

$$J = L(g(f(x)) \text{ where } h = f(x)$$

Where f is the encoder, g is the decoder and *L* is a loss function that penalizes g(*h*) for being dissimilar to *x*. We explored both mean squared error (MSE) and mean absolute error (MAE) as the loss function, *L*, and found mean absolute error to offer better convergence and lower reconstruction error compared with MSE.

The architecture of the LSTM auto-encoder (LAE) is shown in figure 3B. The encoder consists of a LSTM layer with hidden dimension of 20 followed by a linear transformation to the 2D bottleneck. The decoder consists of a dense transformation of the 2D bottleneck to 20 dimensions followed by vector repetitions (30 times) and a LSTM layer followed by a dense layer that reconstructs the input sample. DBSCAN Clustering

We employed the DBSCAN clustering algorithm to identify clusters in the latent representation of HRV data given by the AEs. DBSCAN is a density-based algorithm that clusters densely-packed samples together while disregarding samples in low-density areas as outliers.

c. Training and Evaluation

The AEs were implemented using tensor flow 1.12.0 (tf.keras) deep learning library. Adam optimizer with a learning rate of 1e-4 and a batch size of 64 was utilized to train AEs. a 5-fold cross validation scheme was utilized to train the models and tune the hyper parameters. Both CAE and LAE were trained for 300 epochs. CAE loss plateaued after nearly 150 epochs while LAE plateaued much later at about 290 epochs. Both models were trained using a virtual machine with a 12-core CPU and 24 GB of RAM.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

IV. RESULTS & DISCUSSION

a. Traditional Feature Engineering and K- means

As an initial try at unsupervised classification of HRV data, we hired K-means clustering on 18 engineered structures. The K-means clustering outcomes for k = 2 are shown in figure 2, plot shown in 2D for mean heart rate (Mean HR) and root mean square of successive differences (RMSSD). As shown in the figure, the recognized clusters seem synthetic without a clear separation in the data. This is in part due to the high dimensionality of the data and the essential inclination of K-means clustering algorithm to cluster samples, irrespective of true parting inside the data.

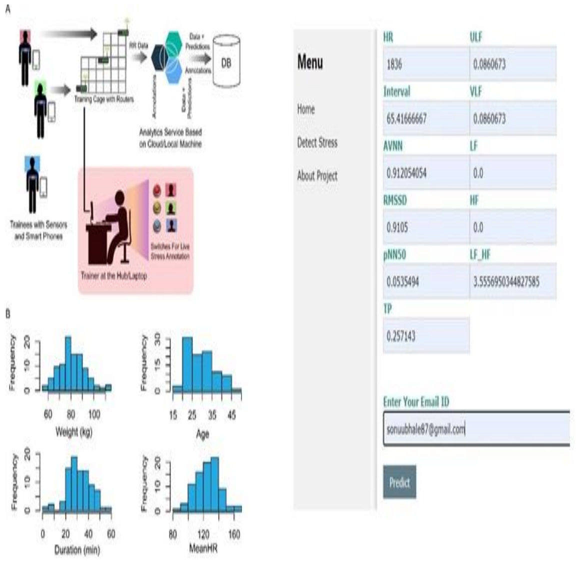
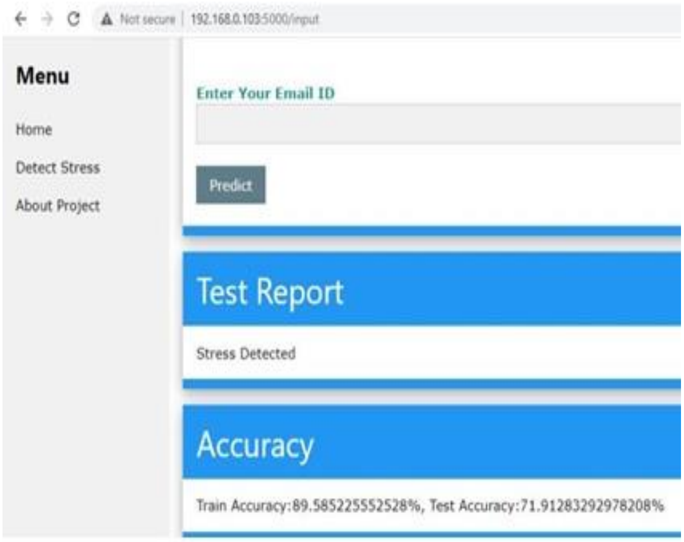


Fig. 1. Data collection and HRV data to be analyzed, predicted or labeled. B) Subject trainee statistics: Weight in kilo- grams, Age, HRV data time period is collected, MHR in beats / minute.

Fig. 3. Project windows and real time results



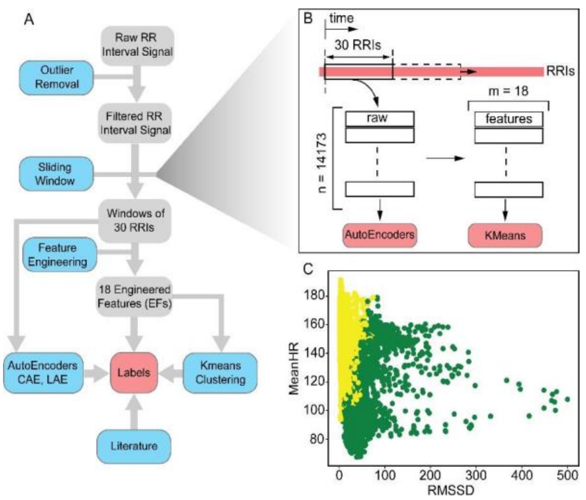


Fig. 2. The workflow for unsupervised classification of HRV data in this Work B) Sliding window transformation of the RRI into non-over-lapping Windows of 30 time steps. C) The results of K-means clustering with $k = 2$ on the samples Using the 18 engineered features. As seen in the plot K-means does not produce distinct well-separated clusters within the data.

b. Discussion

We studied various methods for unsupervised classification of HRV data collected from 100 trainees. The motive was to identify mental stress using the HRV data. We had begun with time and frequency domain HRV features / structures joint with old K-means clustering. The K- means clusters, however, appeared arbitrary and did not produce well-separated clusters within the data. To tackle this problem and bring out any existing hidden patterns within the data, we explored auto encoders (AEs). We incorporated two different neural net architectures namely the convolutional and LSTM neural nets and built two different auto encoders that were trained and evaluated separately A K-Nearest-Neighbor classifier was fit to the validation data. The resultant KNN models were used to make forecasts for hidden samples. Having observed two separate clusters in the data for each model, we set out to determine two points:

Clusters produced by which AE are the true meaningful

ISSN : 2454-9924

clusters about mental stress recognition and 2) which cluster (or label) corresponds to mental strain. So to find an answer to this, we used the benefit of deep-rooted HRV markers of mental stress within the literature. We selected and determined four HRV indicators of mental stress reported in the literature namely: RMSSD, Max-HR, Mean- RR and LF-HF Ratio. We observed that, for the CAE-en- coded clusters, the values of these four markers showed a significant discrepancy across the clusters. In addition, the marker values for one cluster (cluster "0") were predominantly associated with mental stress according to the literature.

V. CONCLUSION

We presented a new approach for unsupervised finding of mental stress from raw HRV data using auto encoders. We demonstrated that classical K-Clustering combined with time, frequency domain structures was not suitable to detect mental stress. We then explored two different architectures of auto encoders to encode the data and find underlying patterns that may enable us to find out mental stress in an unsupervised manner. We trained convolutional and LSTM auto encoders and demonstrated that despite being more powerful and producing lower reconstruction error, LSTM auto encoders failed to identify useful patterns within the data. On the other hand, the convolutional auto encoders with their much fewer trainable weights, produced clusters that were verifiably distinct and pointed to different levels of mental stress according to the reported markers of mental stress. As per the results given by the convolutional auto encoder, 90% plus samples collected from trainees during a practice were mentally stressed at the same time less than 10% had normal HRV. While our proposed approach offers promising preliminary results toward unsupervised detection of mental stress, we recognize a number of shortcomings that must be addressed with additional experiments and data. For instance, our training dataset was relatively small and additional data, including new modalities (e.g. motion, voice, etc.), would improve the accuracy of our trained models. Moreover, our method didn’t consider intrinsic differences in HRV of different individuals which could be investigated with further experiments and data. In addition, it is imperative that the observed results in this work are thoroughly validated via new experiments.

VI. REFERENCES

[1] T. Mohana Priya, Dr. M. Punithavalli & Dr. R. Rajesh Kanna, Machine Learning Algorithm for Development of Enhanced Support Vector Machine Technique to Predict Stress, Global Journal of Computer Science and Technology: C Software & Data Engineering, Volume 20, Issue 2, No. 2020, pp 12-20

[2] Ganesh Kumar and P. Vasanth Sena, "Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit," International Journal of Computer Science and Network Security, Vol. 15, issue 9, Sep. 2015, pp. 222-234

[3] Gyusoo Kim and Seulgi Lee, "2014 Payment Research", Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[4] Chengwei Liu, Yixiang Chan, Syed Hasnain Alam Kazmi, Hao Fu, "Financial Fraud Detection Model: Based on Random Forest," International Journal of Economics and Finance, Vol. 7, Issue. 7, pp. 178-188, 2015

[5] Hitesh D. Bambhava, Prof. Jayeshkumar Pitroda, Prof. Jaydev J. Bhavsar (2013), "A Comparative Study on Bamboo Scaffolding And Metal Scaffolding in Construction Industry Using Statistical Methods", International Journal of Engineering Trends and Technology (IJETT) – Volume 4, Issue 6, June 2013, Pg.2330-2337.

[6] P. Ganesh Prabhu, D. Ambika, "Study on Behaviour of Workers in Construction Industry to Improve Production Efficiency", International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development (IJCSEIERD), Vol. 3, Issue 1, Mar 2013, 59-66[1] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: A meta- analysis and review of the literature," Psychiatry in- vestigation, vol. 15, no. 3, p. 235, 2018.

[7] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," presented at the 4th European conference of the international federation for medi- cal and biological engineering, 2009, pp. 1366–1369.

[8] Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," Front Public Health, vol. 5, pp. 258–258, Sep. 2017.

[9] U. Kraus et al., "Individual daytime noise exposure during routine activities and heart rate variability adults: a repeated measures study," Environmental health perspectives, vol. 121, no. 5, pp. 607–612, 2013.

[10] W. von Rosenberg, T. Chanwimalueang, T. Adjei, U. Jaffer, V. Goverdovsky, and D. P. Mandic, "Resolv- ing Ambiguities in the LF/HF Ratio: LF-HF Scatter Plots for the Categorization of Mental and Physical Stress from HRV," Front Physiol, vol. 8, pp. 360–360, Jun. 2017.

[11] M. Weippert, M. Kumar, S. Kreuzfeld, D. Arndt, A. Rieger, and R. Stoll, "Comparison of three mobile de- vices for measuring R–R intervals and heart rate var- iability: Polar S810i, Suunto t6 and an ambulatory ECG system," European journal of applied physiology, vol. 109, no. 4, pp. 779–786, 2010.

[12] R. Gevirtz, P. Lehrer, and M. Schwartz, "Cardi- orespiratory biofeedback," Biofeedback: A practi- tioner's guide, pp. 196–213, 2016.

[13] R. McCraty and F. Shaffer, "Heart rate variability: new perspectives on physiological mechanisms, as- sessment of self- regulatory capacity, and health risk," Global advances in health and medicine, vol. 4, no. 1, pp. 46–61, 2015.

[14] S. Järvelin-Pasanen, S. Sinikallio, and M. P. Tar- vainen, "Heart rate variability and occupational stress-systematic review," Ind Health, vol. 56, no. 6, pp. 500–511, Nov. 2018.

[15] M. Kivimäki, P. Leino-Arjas, R. Luukkonen, H. Riihimäki, J. Vahtera, and J. Kirjonen, "Work stress and risk of cardiovascular mortality: prospective cohort study of industrial employees," Bmj, vol. 325, no. 7369, p. 857, 2002.]

[16] S. Salminen, M. Kivimäki, M. Elovainio, and J. Vahtera, "Stress factors predicting injuries of hospital personnel,"

ISSN : 2454-9924

American journal of industrial medicine, vol. 44, no. 1, pp. 32–36, 2003.H. Soori,

[17] M. Rahimi, and H. Mohseni, "Occupational stress and work-related unintentional injuries among Iranian car manufacturing workers," 2008.

[18] R. Bhardwaj, P. Natrajan, and V. Balasubramanian, "Study to Determine the Effectiveness of Deep Learning Classifiers for ECG Based Driver Fatigue Classification," presented at the 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), 2018, pp. 98–102.

[19] A. Saeed, T. Ozcelebi, J. Lukkien, J. van Erp, and S. Trajanovski, "Model adaptation and personalization for physiological stress detection," presented at the 2018 IEEE 5th International Conference on Data Sci- ence and Advanced Analytics (DSAA), 2018, pp. 209–216.

[20] L. Wang and X. Zhou, "Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals," Sensors, vol. 19, no. 7, p. 1502, 2019.

[21] S.-H. Song and D. K. Kim, "Development of a Stress Classification Model Using Deep Belief Networks for Stress Monitoring," Healthcare informatics research, vol. 23, no. 4, pp. 285–292, 2017.

[22] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, and B.-T. Zhang, "Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals," TELE- MEDICINE and eHEALTH, vol. 24, no. 10, pp. 753– 772, 2018.

[23] [18] Y. Smirnov, A. Popov, and V. Kharytonov, "Heart rate variability analysis for epileptic seizures onset detection," in Proc. IEEE 2019 SPSympo, Krakow, 2019, 10.1109/SPS.2019.8881986.

[24] [19] J. Jeppesen, A. Fuglsang-Frederiksen, P. Johansen, et al., "Seizure detection based on heart rate variability using a wearable electrocardiography device," Epilepsia, vol. 60, pp. 2105–2113, 2019.

[25] [20] L. Gagliano, E. Bou Assi, D. K. Nguyen, and M. Sawan, "Bispectrum and recurrent neural networks: Improved classification of interictal and preictal states," Sci Rep, vol. 9, no. 15649, 2019.

[26] [21] L. Gagliano, E. Bou Assi, D. H. Toffa, D. K. Nguyen, and M. Sawan, Fellow, IEEE Unsupervised Clustering of HRV Features Reveals Preictal Changes in Human Epileps