Teaching Machines to Know Your Depressive State: On Physical Activity in Health and Major Depressive Disorder

Kun Qian[1], *Member, IEEE*, Hiroyuki Kuromiya[1],
Zixing Zhang[2], *Member, IEEE*, Jinhyuk Kim[3], Toru Nakamura[4], Kazuhiro Yoshiuchi[5],
Björn W. Schuller[2], *Fellow, IEEE*, and Yoshiharu Yamamoto[1], *Member, IEEE*

Abstract— A less-invasive method for the diagnosis of the major depressive disorder can be useful for both the psychiatrists and the patients. We propose a machine learning framework for automatically discriminating patients suffering from the major depressive disorder ($n = 14$) and healthy subjects ($n = 17$). To this end, spontaneous physical activity data were recorded via a watch-type computer device equipped by the participants in their daily lives. Two machine learning models are investigated and compared, i.e., support vector machines, and deep recurrent neural networks. Experimental results show that, both of the two methods, i.e., the static model fed with human hand-crafted features, and the sequential model fed with raw data can reach a promising performance with an unweighted average recall at 76.0 % and 56.3 %, respectively.

I. INTRODUCTION

As a serious, chronic, and recurring psychiatric disorder, major depressive disorder (MDD) has a high lifetime prevalence (approximately 15 %) in the U. S. population [1]. Previous studies indicated that, patients suffering from MDD may have high risks for cardiovascular morbidity and mortality [2]. Besides, MDD is an independent risk factor for coronary artery disease, myocardial infarction, and sudden cardiac death [3], [4]. More seriously, suicide risk might be greatly increased in depression [5], [6]. Therefore, an objective evaluation of depressive mood can be significant and essential, both for the diagnosis, and the treatment for MDD patients.

With the help of signal processing and machine learning, physiological signals collected from MDD patients can be used for developing automated methods on diagnosing the depressive disorder. For example, electroencephalography (EEG) data were used in a method based on mixture of factor analysis to classify three types of psychiatric illness, i.e., MDD, schizophrenia (SCZ), and bipolar disorder (BD) [7]. Mumtaz *et al.* proposed a method based on logistic regression to discriminate MDD patients and healthy controls via EEG data [8]. Among EEG signals, other physiological signals, e.g., arterial pulse signals [9], heart rate variability [10], images [11], and audio signals [12] were used for extracting representations for building MDD machine learning models. The experimental results of these limited studies are encouraging and promising. However, the methods mentioned above are expensive, and might be time-consuming for practical usage. It is hence extremely desirable both for doctors and patients to find an easy, cheap, and less-invasive method to screen the MDD.

In this study, we propose a machine learning framework based on analysis of the spontaneous physical activity (SPA) data [13], as recorded via a watch-type computer device participants are equipped with. In our study, we provide two models investigated and compared, i.e., support vector machines (SVMs) fed with human hand-crafted features, and deep recurrent neural networks (DRNNs) fed with raw sensor data. The main contributions in this work are: Firstly, to the best of our knowledge, it is the first time to use SPA data for diagnosis of MDD via a machine learning framework, and the experimental results demonstrate its promising potential; secondly, it furthers the study based on the statistical analysis of SPA data in [14]; thirdly, we find that, even fed with the raw data, the presented deep learning model can be useful to extract features itself, which can be a good direction for future studies. The rest of this paper will be organised as follows: Firstly, the database and methods used in this study will be briefly introduced in Section II. Then, experimental results and discussion will be given in Section III and Section IV, respectively. Finally, Section V concludes the work and the findings in this paper.

II. MATERIALS AND METHODS

A. Database

This study is approved by the research ethic committees of the Teikyo University, Japan, and the University of Tokyo, Japan. There are 14 patients with MDD (12 males (M), 2 females (F); age: 34.0±5.7 years, age range: 22-42 years), and

[1]Kun Qian, Hiroyuki Kuromiya, and Yoshiharu Yamamoto are with the Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. {qian, kuromiya, yamamoto}@p.u-tokyo.ac.jp

[2]Zixing Zhang, and Björn W. Schuller are with GLAM – the Group on Language, Audio & Music, Imperial College London, 180 Queens Gate, Huxley Bldg., London SW7 2AZ, UK. {zixing.zhang, bjoern.schuller}@imperial.ac.uk

[3]Jinhyuk Kim is with the Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka 432-8011, Japan. kimj@inf.shizuoka.ac.jp

[4]Toru Nakamura is with the Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. t-nakamura@sangaku.es.osaka-u.ac.jp

[5]Kazuhiro Yoshiuchi is with the Department of Stress Sciences and Psychosomatic Medicine, Gaduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. kyoshiuc-tky@umin.ac.jp
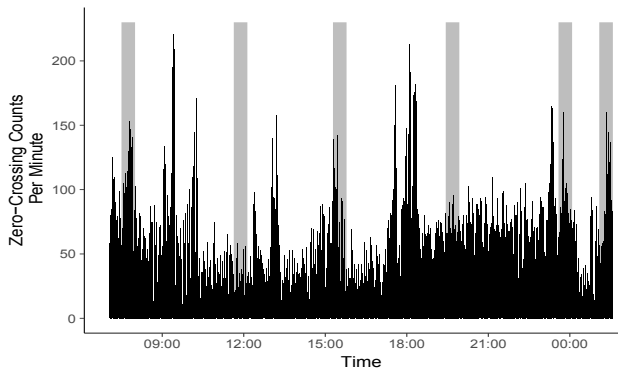
Fig. 1. An example of the SPA data recorded by a participant. The 30 min SPA data before a self-reported symptom was made by the participants (highlighted by grey colour background), are selected to be the instances.

TABLE I
THE NUMBER [#] OF INSTANCES IN EACH DATA SET.

	Train	Dev	Test	Σ
MDD	425	137	135	697
Healthy	436	147	147	730
Σ	861	284	282	1 427

17 healthy control subjects (17 M, 0 F; age: 43.2±11.8 years, age range: 24-58 years) involved. The MDD patients are from the Mizonokuchi Hospital of Teikyo University, Kanagawa, Japan. The healthy subjects are full-time office workers from the University of Tokyo, Japan. Detailed information of the MDD patients and the healthy subjects can be found in [14].

The SPA data were recorded via a watch-type device equipped with an activity monitor (Ambulatory Monitors Inc., Ardsley, NY, USA). The sensor is capable to detect small changes in bodily acceleration ($\geqslant 0.01$ G/rad/s). In this study, the *zero-crossing counts* accumulated for every one minute are used as the SPA data. Based on the previous study [14], we select 30 minute (min) SPA data before a self-reported symptom was quasi-regularly made by the participants as the *instance* for further machine learning (see Fig. 1).

Considering the case of *subject-independency*, we split the whole database into a *train*, a *development* (dev), and a *test* set, respectively. In each data set, both, the number participants, and the number of instances, occupies approximately 60 %, 20 %, and 20 % of the overall database, respectively. The information on data partition can be found in Table I.

B. Feature Extraction

The SPA data recorded from the participants are the *low-level descriptors* (LLDs) of their behaviours, which can be used to extract features for a further machine learning phase. As outlined, we investigate two methods, namely, functionals based on these LLDs, and raw data with deep learning. For both of these two methods, the input is standardised to eliminate outliers before feeding into the classification model.

1) Functionals: The changes of the LLDs over a given period of time can carry important information for further building of a machine learning model. As mentioned in [15], the *supra-segmental* features can summarise the information of a time unit. The principle of functionals is to map the time series to a scalar value per applied functional, which can make the features suitable for static models, e. g., SVMs. In this study, 9 statistical functionals (*cf.* [16]), i.e., *maximum*, *minimum*, *mean*, *range*, *standard deviation*, *skewness*, *kurtosis*, *slope*, and *bias* of linear regression approximation, are calculated from 30 min LLDs from each instance.

2) Deep Learning: With the help of the recent development of *deep learning* [17], features can be learnt by the model itself if the data size can be considerably large. Inspired by the success achieved in speech emotion recognition [18], we introduce a sequential deep learning model to the area of MDD diagnosis via the SPA data. Compared with the classical machine learning methods mentioned above, there are no human hand-crafted features needed when using the proposed deep learning paradigm. As a kind of time series, the 30 min SPA data will be firstly framed into segments of 10 min window (with 50 % overlap), and sorted by time steps as the sequential input for the DRNNs.

C. Machine Learning Models

1) Support Vector Machines: The diagnosis of MDD is a kind of binary classification problem. We use SVMs [19], which were popular and standard machine learning models in the past two decades. The aim of an SVM classifier is to find the best hyperplane that maximises the separation between classes. When performing classification, the instances will be mapped to a multi-dimensional space, and the predictions will be given based on which side of the gap they fall onto. Formally, in a binary classification problem (e. g., $y_i \in \{-1, +1\}$, y_i is the label of the *i*-th instance), SVMs intend to find the optimal margin separating hyperplane by solving the optimisation problem [19]:

$$minimise : \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{n} \alpha_i,$$
$$subject \ to : \sum_{i=1}^{n} \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C, \ i = 1, \dots, n, \quad (1)$$

where α_i is the *Lagrange* multiplier of the training sample ($\boldsymbol{x}_i$, y_i), and C is a pre-defined parameter. $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, i.e., *kernel function* [19], which can make SVMs analyse also nonlinearly separable problems. There are several commonly used kernel functions, e. g., *linear*, *polynomial*, and *radial basis function* (RBF).

2) Recurrent Neural Networks: As the SPA data may contain contextual information, we intend to evaluate the time series learning capacity of the model via RNNs [20]. Different from the traditional feed-forward neural networks (FNNs) [21], RNNs can learn to model sequential information by incorporating the outputs of a previous time step as the additional inputs for the current time step. However, basic RNN models face the 'vanishing gradient' challenge (cf. [22]) during the training process via *backpropagation through time* (BPTT) [23]. Therefore, in this study, we
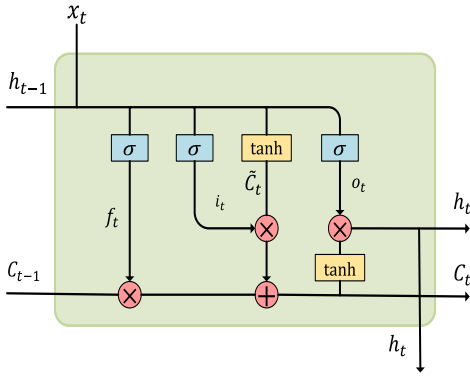
Fig. 2. The structure of an LSTM cell.

Fig. 3. The UARs of the proposed models on the development, and test sets.

selected long short-term memory (LSTM) cells [24] to overcome the vanishing gradient problem. The behaviour of an LSTM cells (see Fig. 2) can be defined as follows:

$$i_t = \sigma(W_{x_i}x_t + W_{h_i}h_{t-1} + W_{c_i}c_{t-1} + b_i), \qquad (2a)$$

$$o_t = \sigma(W_{x_o}x_t + W_{h_o}h_{t-1} + W_{c_o}c_t + b_o), \qquad (2b)$$

$$f_t = \sigma(W_{x_f}x_t + W_{h_f}h_{t-1} + W_{c_f}c_{t-1} + b_f), \qquad (2c)$$

$$\tilde{c}_t = \tanh(W_{h_c}h_{t-1} + W_{x_c}x_t + b_c), \qquad (2d)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \qquad (2e)$$

$$h_t = o_t \odot \sigma_h(c_t), \qquad (2f)$$

where σ is the logistic sigmoid function, i, o, f, and c represent the *input gate*, the *output gate*, the *forget gate*, and the *cell state*, respectively. In addition, h represents the hidden layer, W is the weight matrix, b is the bias matrix, and $\odot$ denotes the element-wise multiplication.

D. Evaluation Metrics

Considering the imbalanced characteristic of the SPA data distribution (*cf.* Table I), we use the unweighted average recall (UAR) [25] as the general evaluation metric. In addition, the commonly used *accuracy*, *sensitivity*, *specificity*, *precision*, *recall*, *F1-measure*, and *the area under the curve* (AUC) of the receiver operating characteristic are provided. When reporting significance testing, a one-tailed *z*-test is used.

III. EXPERIMENTAL RESULTS

A. Experimental Setup

Experiments were basically done in an environment of Matlab R2018b by MathWorks. The SVM was implemented by the popular open source toolkit LIBSVM [26]. All the hyper-parameters of the models were optimised on the development set, and applied to the test set. For SVMs, kernels were selected from *linear*, *polynomial*, *sigmoid*, and *RBF*. The *C*-value for SVMs was optimised using a grid-search strategy from $\{10^{-5}, 10^{-4}, \ldots, 10^4, 10^5\}$. For LSTM RNN, the sequence length was selected to be 10, and the batch size was set to 25. The optimisation algorithm was chosen to be 'Adam', and the drop out rate was set to 0.1. The structure of the DRNN model was two-layers with LSTM cells. The
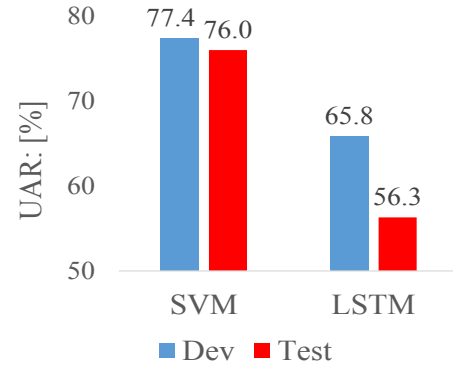
TABLE II
CONFUSION MATRICES (NORMALISED: [%]) OF THE PROPOSED TWO
MODELS ON THE TEST SET.

(a) SVM				(b) LSTM		
Pred ->	MDD	Healthy		*Pred* ->	MDD	Healthy
MDD	59.4	40.6		MDD	83.7	16.3
Healthy	7.5	92.5		Healthy	71.1	28.9

number of neurons was optimised from the candidates as 64, 128, and 512.

B. Results

We can find from Fig. 3 the best UARs achieved by the proposed two models by the dev set, and the corresponding results using the optimised hyper-parameters on the test set. In general, SVM outperformed LSTM in this study, both for the dev set (77.4 % vs 65.8 %, $p < .002$), and the test set (76.0 % vs 56.3 %, $p < .001$). From Table II, we can see that, SVMs have a higher recall for finding healthy subjects than LSTM, while the latter is superior when recognising MDD patients. Table III shows the other evaluation metrics for the two proposed models by the test set.

IV. DISCUSSION

We can find that, the 'classical' pattern recognition approach, i. e., functionals extracted from LLDs fed into an SVM, can reach an excellent performance on UAR both on the development (77.4 %) and test (76.0 %) sets. However, when looking at the performance of sensitivity (see Table II(a) and Table III), there is room from improvement in future work. We may consider that, currently, functionals used in this study might be insufficient to describe the complicated information inherited in SPA data. One future direction could be finding some more advanced human hand-crafted features fed into the static models such as SVM. When considering the performance of LSTM RNNs, there is a big gap between its UAR on the development, and the test sets (65.8 % vs 56.3 %). We may consider that, limited to the size of data set analysed, the DRNNs could be overfitting during the training process. Encouragingly, the deep learning system (based on LSTM model) works well in detection of

	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-measure	AUC
SVM	76.8	59.4	92.5	87.8	59.4	70.9	87.0
LSTM	57.4	83.7	28.9	56.2	83.7	67.2	37.8
p-value	< .001	< .001	< .001	< .001	< .001	-	< .001

MDD (see Table II(b) and Table III), in which, a higher sensitivity can be achieved by LSTM when comparing with SVM. Therefore, to improve the generalisation and capacity of DRNNs like LSTM, more available SPA data will be needed.

V. CONCLUSION

In this study, we proposed a machine learning framework for automatic diagnosis of MDD via the SPA data collected from participants' daily lives. The SVM and LSTM models performed successfully as classifiers for the data set used in this work. In future work, we will explore bag-of-words approaches for the activity data to extract higher representations from LLDs [27], which might describe the behaviour of the MDD patients in a better way.

ACKNOWLEDGMENT

REFERENCES

[1] D. S. Hasin, R. D. Goodwin, F. S. Stinson, and B. F. Grant, "Epidemiology of major depressive disorder: Results from the National Epidemiologic Survey on Alcoholism and Related Conditions," *Archives of General Psychiatry*, vol. 62, no. 10, pp. 1097–1106, 2005.

[2] D. L. Hare, S. R. Toukhsati, P. Johansson, and T. Jaarsma, "Depression and cardiovascular disease: A clinical review," *European Heart Journal*, vol. 35, no. 21, pp. 1365–1372, 2013.

[3] A. Nicholson, H. Kuper, and H. Hemingway, "Depression as an aetiologic and prognostic factor in coronary heart disease: A meta-analysis of 6362 events among 146 538 participants in 54 observational studies," *European Heart Journal*, vol. 27, no. 23, pp. 2763–2774, 2006.

[4] G. A. Alvares, D. S. Quintana, I. B. Hickie, and A. J. Guastella, "Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: A systematic review and meta-analysis," *Journal of Psychiatry & Neuroscience*, vol. 41, no. 2, pp. 89–104, 2016.

[5] S.-C. Park, M.-S. Lee, S. W. Hahn, T.-M. Si, S. Kanba, M.-Y. Chong, C. K. Yoon, P. Udomratn, A. Tripathi, N. Sartorius, N. Shinfuku, M. M. Maramis, and Y. C. Park, "Suicidal thoughts/acts and clinical correlates in patients with depressive disorders in Asians: Results from the reap-ad study," *Acta Neuropsychiatrica*, vol. 28, no. 6, pp. 337–345, 2016.

[6] X. Fang, C. Zhang, Z. Wu, D. Peng, W. Xia, J. Xu, C. Wang, L. Cui, J. Huang, and Y. Fang, "Prevalence, risk factors and clinical characteristics of suicidal ideation in Chinese patients with depression," *Journal of Affective Disorders*, vol. 235, pp. 135–141, 2018.

[7] A. Khodayari-Rostamabad, J. P. Reilly, G. Hasey, H. MacCrimmon *et al.*, "Diagnosis of psychiatric disorders using EEG data and employing a statistical decision model," in *Proc. EMBC*. Buenos Aires, Argentina: IEEE, 2010, pp. 4006–4009.

[8] W. Mumtaz, A. S. Malik, S. S. A. Ali, M. A. M. Yasin, and H. Amin, "Detrended fluctuation analysis for major depressive disorder," in *Proc. EMBC*. Milan, Italy: IEEE, 2015, pp. 4162–4165.

[9] A. H. Khandoker, V. Luthra, Y. Abouallaban, S. Saha, K. I. Ahmed, R. Mostafa, N. Chowdhury, and H. F. Jelinek, "Identifying depressed patients with and without suicidal ideation by finger photoplethysmography," in *Proc. EMBC*. Orlando, FL, USA: IEEE, 2016, pp. 1842–1845.

[10] M. Kobayashi, G. Sun, T. Shinba, T. Matsui, and T. Kirimoto, "Simple and objective screening of major depressive disorder by heart rate variability analysis during paced respiration and mental task conditions," in *Proc. EMBC*. Seogwipo, South Korea: IEEE, 2017, pp. 1316–1319.

[11] S.-H. Chu, C. Lenglet, M. W. Schreiner, B. Klimes-Dougan, K. Cullen, and K. K. Parhi, "Anatomical biomarkers for adolescent major depressive disorder from diffusion weighted imaging using SVM classifier," in *Proc. EMBC*. Honolulu, HI, USA: IEEE, 2018, pp. 2740–2743.

[12] S. Harati, A. Crowell, H. Mayberg, and S. Nemati, "Depression severity classification from speech emotion," in *Proc. EMBC*. Honolulu, HI, USA: IEEE, 2018, pp. 5763–5766.

[13] T. Nakamura, K. Kiyono, H. Wendt, P. Abry, and Y. Yamamoto, "Multiscale analysis of intensive longitudinal biomedical signals and its clinical applications," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 242–261, 2016.

[14] J. Kim, T. Nakamura, H. Kikuchi, K. Yoshiuchi, T. Sasaki, and Y. Yamamoto, "Covariation of depressive mood and spontaneous physical activity in major depressive disorder: Toward continuous monitoring of depressive mood," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1347–1355, 2015.

[15] F. Eyben, *Real-time Speech and Music Classification by Large Audio Feature Space Extraction*. Cham, Switzerland: Springer International Publishing, 2015, Doctoral Thesis.

[16] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Classification of the excitation location of snore sounds in the upper airway by acoustic multi-feature analysis," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1731–1741, 2017.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M. A. Nicolaou, B. Schuller, and S. Zafeiriou, "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network," in *Proc. ICASSP*. Shanghai, P. R. China: IEEE, 2016, pp. 5200–5204.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.

[21] C. M. Bishop, *Pattern recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[22] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber *et al.*, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Neural Networks*. Piscataway, NJ, USA: IEEE Press, 2001, pp. 237–244, J. F. Kolen, and S. C. Kremer, Ed.

[23] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] B. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," in *Proc. INTERSPEECH*, Brighton, UK, 2009, pp. 312–315.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[27] K. Qian, M. Schmitt, C. Janott, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "A bag of wavelet features for snore sound classification," *Annals of Biomedical Engineering*, vol. 47, no. 4, pp. 1000–1011, 2019.