Pattern recognition of mental stress levels from differential RRI time series using LSTM networks

Nan Bu
Dep. of Control and Information Systems Engineering
National Institute of Technology, Kumamoto College
Kumamoto, 861–1102 JAPAN
Email: bu@ieee.org

Masanori Fukami and Osamu Fukuda
Dep. of Information Science, Graduate School of
Science and Engineering, Saga University
Saga, 840–8502 JAPAN
Email: fukudao@cc.saga-u.ac.jp

2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech) | 978-1-6654-1875-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/LIFETECH52111.2021.9391853

***Abstract*—This paper reports some experimental results of a preliminary study on development of a mental stress recognition method using a deep learning-based model. This model acquires time-varying characteristics in heart rate variability (HRV) data with a long short-term memory (LSTM) deep network. The method proposed in this paper attempts to directly input differential R-R intervals (RRI) time series into the LSTM network. Training and classification performance have been discussed with stress level recognition experiments, in which HRV data were obtained with mental stress levels induced in a virtual reality environment. Test configurations have examined with different number of classes, segmentation lengths, and datasets. Comparing with the experimental results of the authors' previous studies, it is found that using differential RRI can largely ease the convergence difficulty in LSTM training process.**

***Index Terms*—Stress recognition, mental stress, differential R-R intervals, deep learning, long short-term memory**

I. Introduction

Recognition of stress levels, especially those related to emotional and affective states, is of great importance in order to deal with depression and suicide risk, and to help people overcome stress related disorders. A variety of physiological signals and a wide spectrum of machine leaning techniques have been discussed in this field [1]. Features and evaluation indices extracted from electrocardiography (ECG) are widely accepted since ECG is non-invasive and inexpensive for wearable applications. Heart rate variability (HRV) has been intensively investigated to bring new methods and techniques to clinical and healthcare applications [2]. For example, a multilevel stress detection method has been developed based on well-selected feature sets of HRV indices [3].

Recently, with the emergence of deep learning (DL) improved performance is expected through novel proposals on incorporation of DL network architectures into stress recognition schemes. Masood and Alghamdi have developed a DL framework for mental stress classification, where features of bio-signals, including ECG, are extracted and fed into a convolutional neural network (CNN) architecture [4]. On the other hand, challenges of end-to-end stress recognition have been made by feeding HRV data direct into a CNN in order to construct a data-driven model for mental stress [5].

This work was supported by JSPS KAKENHI Grant Number JP19H04506.

Temporal characteristics of HRV have also been discussed for the purpose of stress analysis. In the field of machine learning, recurrent neural network (RNN) is considered good at dealing with time-varying features. A long short-term memory (LSTM) network has been applied to time sequences of HRV features, and it was found that exploiting time variation in these features had improved the classification accuracy of healthy subjects and patients with heart diseases [6].

The authors have carried out research works using a deep LSTM network in order to acquire time-varying and discriminative features between mental stress levels *directly* from R-R intervals (RRI) time series [7]. By skipping the feature extraction stage, this method attempts to develop an end-to-end stress recognition method. Relative high classification rates were obtained in some parts of the experimental results, however, not all training trials converged well and overfitting of the training process made the recognition results inclined to some particular stress class(s). Although this study validated possibility of stress recognition based on raw RRI data and deep LSTM network, detailed discussion and systematic examination are further required in order to improve the performance of this DL-based model.

On the other hand, it has been reported that use of differential RRIs may improve discrimination performance for normal subjects and patients with congestive heart failure [8]. Differential RRI, which is the increment between two consecutive RRI, is also an important aspect or basic value for derivation of HRV evaluation indices [9]. With this consideration, raw RRI time series can be replaced by its differential version.

This paper aims to adopt differential RRI in the deep RNN-based stress recognition method for better convergence and discrimination performance. In order to compare with results of the previous study [7], same experimental configurations are used in this paper.

II. Method

A schematic diagram of the proposed method is shown in Fig. 1. Differential RRI data are directly fed into the LSTM network step by step.

978-1-6654-1875-1/21/$31.00 ©2021 IEEE

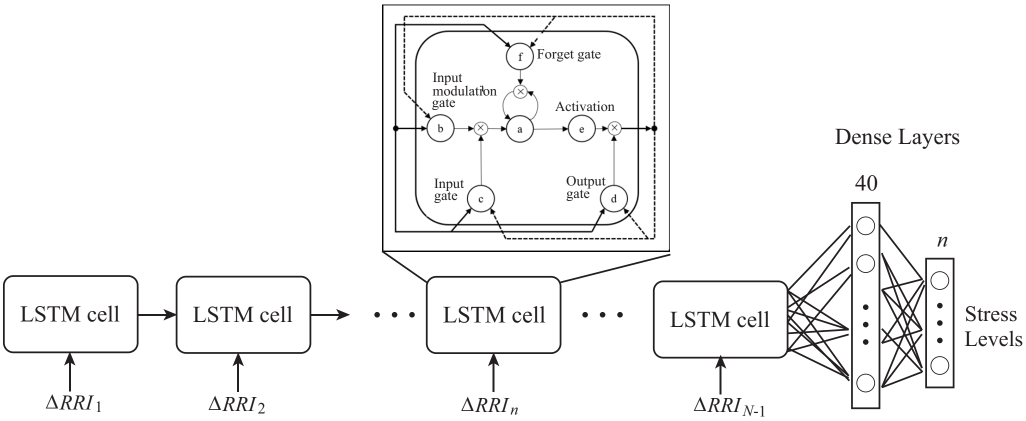

Fig. 1. The proposed method with differential RRI data and LSTM networks for stress level recognition.

A. Differential RRI

The incremental series of a long-range correlated time series is considered informative on the dynamic properties. For RRI time series, RRI_n $(n = 1, \cdots, N)$, its increment series (ΔRRI) is obtained as follows,

$$\Delta RRI_n = RRI_{n+1} - RRI_n \tag{1}$$

As the original RRI time series, ΔRRI possesses inherent nature of uneven sampling. In order to regulate the data length of input for the LSTM cells, cubic spline interpolation is applied to the differential RRI time series and ΔRRI is then evenly sampled with a resampling frequency of 2 Hz. Furthermore, normalization is achieved for each differential RRI time series. Short and ultra-short data length of input is considered as the target of this research project, and in this paper differential RRI time series is segmented with data length L ($L \leq 30$ s).

B. Stress Recognition using Differential RRI and LSTM

Differential RRI data is trained with an LSTM-based deep neural networks (DNN) to identify mental stress levels. As shown in Fig. 1, a many-to-one structure is used to form the LSTM cells, and the dimension of LSTM output is set as 40. The number of LSTM layers, m, is set corresponding to the input data length. The LSTM layers are followed with two dense layers that classify the output of LSTM into classes of stress levels. The number of units in the first dense layer, D_1, is prepared according to dimension of LSTM's output. Here D_1 is 40. Units in the second dense layer are outputs for mental levels n.

III. Stress Recognition Experiments

RRI data was collected in a mental stress induction experiment [10], where subjects were provided with visual stimulus in a virtual reality (VR) environment. Three VR contents are considered as three different mental stress levels. RRI data of eight subjects are used in the experimental study [7]. Figure 2 shows examples of raw RRI data and the normalized ΔRRI data of subject A for three stress levels.

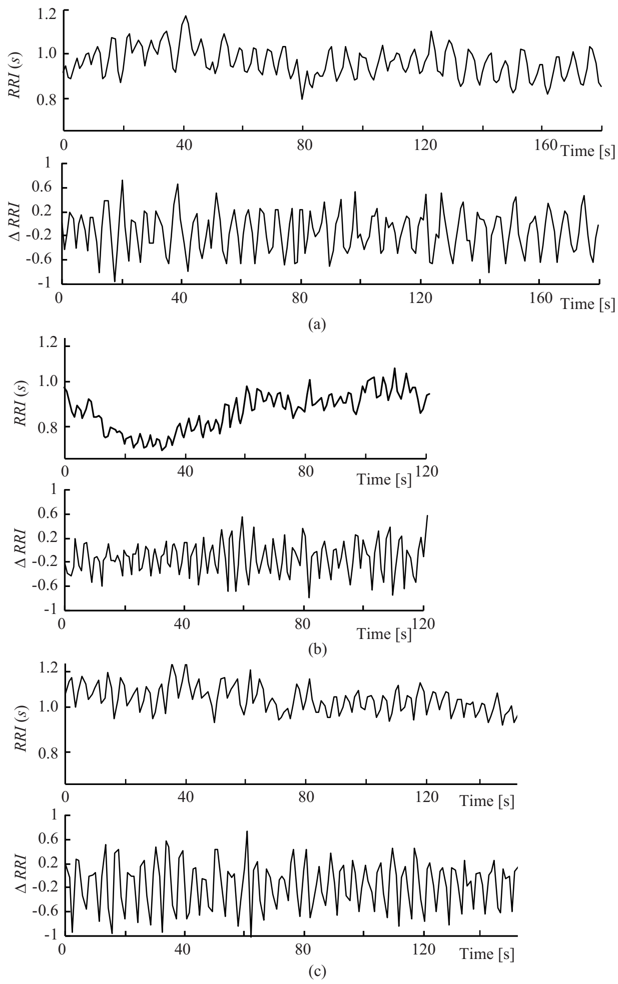

Fig. 2. Examples of RRI and normalized differential RRI data for VR stress sessions of VR(a) Woods scene (180 s), VR(b) Roller coaster (120 s), and VR(c) Journey through the milky way (150 s).

Two-level stress recognition experiments were conducted with data of VR contents of woods scene, VR(a), and roller coaster riding, VR(b), $n = 2$. In addition, stress recognition tests were achieved with data of three VR contents, $n = 3$.

TABLE I
CONFIGURATION OF THREE TEST DATASETS.

	Training data	Validation data
Dataset 1	B-{0∼3}, C-{0∼3}	A-{0∼3}
Dataset 2	A-{0∼3}, C-{0∼3}	B-{0∼3}
Dataset 3	A-{0∼3}, B-{0∼3}	C-{0∼3}

Data of each stress session was divided into three equal parts (A, B, and C), and leave one out cross validation is used to construct three datasets. With the same configuration, the datasets were constructed as depicted in Table I. The labels (A, B, and C) stand for three divided parts and the labels of 1 to 3 correspond to three stress levels. In addition, the number label 0 stands for a pre-test rest session of 180 s in [10], with which a four-class recognition ($n = 4$) is also examined in this paper. In the experiments with different class numbers, corresponding differential RRI data of the classes were used for training and validation. For each training and validation segments, the datasets were generated with length L ($L = 10$, 20, and 30 s) and the shift between segments is 0.5 s.

IV. EXPERIMENTAL RESULTS

Recognition accuracies are calculated based on the validation data in each dataset with the best recognition results during training processes. Comparison of the recognition accuracies between the proposed method and those reported in [7] are listed in Tables II to VI, under test configurations of different combinations (n, L). In these tables, mark "–" denotes that the training processes do not converge well within a predefined number of epochs.

TABLE II
COMPARISON OF RECOGNITION ACCURACIES BETWEEN METHODS USING *RRI* AND Δ*RRI* FOR $n = 2$ AND $L = 10$ S [%].

		A	B	C	D	E	F	G	H
Dataset 1	*RRI*	–	94.2	–	95.5	–	–	–	–
	Δ*RRI*	83.3	90.0	–	94.9	–	59.3	79.5	50.0
Dataset 2	*RRI*	100	82.3	–	80.3	–	68.2	100	–
	Δ*RRI*	93.6	82.8	66.7	99.2	–	59.9	90.5	–
Dataset 3	*RRI*	95.8	65.2	50.0	83.7	–	–	83.5	–
	Δ*RRI*	95.5	100	80.2	93.3	–	75.2	–	–

TABLE III
COMPARISON OF RECOGNITION ACCURACIES ($n = 2$ AND $L = 20$ S) [%].

		A	B	C	D	E	F	G	H
Dataset 1	*RRI*	–	–	–	–	–	–	–	–
	Δ*RRI*	100	96.5	–	100	–	58.8	–	–
Dataset 2	*RRI*	–	–	–	82.9	–	–	–	–
	Δ*RRI*	97.5	97.6	79.9	77.9	–	98.1	97.6	–
Dataset 3	*RRI*	100	51.3	–	90.7	–	–	88.8	–
	Δ*RRI*	94.0	82.1	–	86.9	–	61.6	89.4	–

TABLE IV
COMPARISON OF RECOGNITION ACCURACIES ($n = 2$ AND $L = 30$ S) [%].

		A	B	C	D	E	F	G	H
Dataset 1	*RRI*	100	–	–	–	83.3	–	–	–
	Δ*RRI*	–	77.4	93.8	–	–	72.7	66.1	–
Dataset 2	*RRI*	–	100	–	–	83.3	–	–	–
	Δ*RRI*	99.2	100	60.9	86.4	–	93.6	100	82.0
Dataset 3	*RRI*	100	57.5	–	–	100	–	100	–
	Δ*RRI*	50.0	100	–	90.1	–	70.5	–	–

TABLE V
COMPARISON OF RECOGNITION ACCURACIES ($n = 3$ AND $L = 30$ S) [%].

		A	B	C	D	E	F	G	H
Dataset 1	*RRI*	–	–	–	–	–	66.7	–	–
	Δ*RRI*	–	80.6	65.0	70.7	–	52.8	–	53.6
Dataset 2	*RRI*	–	–	–	–	–	–	76.7	88.4
	Δ*RRI*	–	96.1	68.3	54.5	–	53.9	84.5	41.7
Dataset 3	*RRI*	66.7	–	–	–	–	–	–	39.2
	Δ*RRI*	78.6	66.3	78.1	–	–	51.9	64.4	–

From these tables, convergence performance varies largely between test configurations and datasets. Such difference can also be recognized between subjects. In two-class tests (Tables II and III), training processes did not converge for most data of subjects E and H. Although the proposed method gave a converged trial for dataset 1 with $L = 10$ s (subject H), the recognition result indicates an overfitting occurred for this test trial. On the other hand, when data length increased to 20 s increase of accuracies can be found in some test conditions, however, some other test trials failed in convergence, for example, dataset 3 of subject C and dataset 1 of subject G. With a further change of L to 30 s (see Table IV), similar trends can be recognized while decrease of accuracies can be found like the case of dataset 3 (subject A).

Difference of convergence performance between datasets is obvious for both methods. This demonstrates the time-varying characteristics in RRI time series and its differential version with respect to the construct process of the datasets. The proposed method is able to handle such characteristics, while the previously developed method fails for a large part of the datasets. With different data length, the situations changes greatly so that a simple trend between different L cannot be concluded easily.

Table VII summaries the recognition accuracies of two methods for different combinations (n, L). Statistical data of

TABLE VI
COMPARISON OF RECOGNITION ACCURACIES ($n = 4$ AND $L = 20$ S) [%].

		A	B	C	D	E	F	G	H
Dataset 1	*RRI*	–	50.0	–	–	–	–	–	–
	Δ*RRI*	50.4	70.7	50.3	56.4	59.4	42.2	42.9	47.1
Dataset 2	*RRI*	–	–	–	–	–	–	–	–
	Δ*RRI*	65.5	61.8	61.5	74.0	–	–	51.8	31.1
Dataset 3	*RRI*	59.3	66.1	–	–	–	43.6	34.9	66.0
	Δ*RRI*	49.2	84.5	78.1	61.1	32.2	–	59.6	60.4

TABLE VII
SUMMARY OF RECOGNITION ACCURACIES FOR TWO METHODS WITH DIFFERENT TEST CONFIGURATIONS [%].

		LSTM with *RRI*					LSTM with Δ*RRI*				
n	L	Mean	S.D.	Max	Min	Converged Trails	Mean	S.D.	Max	Min	Converged Trails
2	10 s	83.2	14.9	100	50.0	12	82.0	14.7	100	50.0	17
	20 s	82.7	16.7	100	51.3	5	87.9	13.0	100	58.8	15
	30 s	90.5	14.3	100	57.5	8	82.8	15.4	100	50.0	16
3	10 s	71.3	16.1	95.1	33.3	23	55.9	12.7	84.1	39.9	24
	20 s	74.0	19.4	100	33.3	12	62.9	15.7	100	38.3	22
	30 s	67.5	16.3	88.4	39.2	5	66.3	14.1	96.1	41.7	16
4	10 s	52.3	10.3	70.8	32.0	24	47.1	9.8	67.2	31.7	24
	20 s	53.3	11.6	66.1	34.9	6	56.7	13.5	84.5	31.1	21
	30 s	57.8	8.2	74.5	50.5	6	58.8	14.8	82.0	36.2	18

TABLE VIII
CONFUSION MATRIX OF RESULT FOR SUBJ. B WITH DATASET 3 ($n = 4$ AND $L = 30$ S).

		Predicted Labels			
		Rest	VR(a)	VR(b)	VR(c)
True Lables	Rest	56	0	0	4
	VR(a)	6	53	0	1
	VR(b)	8	9	0	5
	VR(c)	0	28	12	0

mean accuracies and standard deviations (S.D.) are obtained from all 24 test trials (three datasets $\times$ eight subjects) under each condition. In addition, the numbers of converged trials, among 24 test trials, are shown in this table as well. The mean and S.D. results do not show large difference between two methods. However, the method based on differential RRI showed improved convergence performance with more trials among all 24 tests. Generally, the proposed method goes with more converged trials. When data length increases, the LSTM with *RRI* deteriorates more since the original RRI data varies with different data range which cannot be easily handled by LSTM network.

Results of other two test configurations are listed in Tables V and VI. Convergence performance does not degrade much, however, deterioration in the accuracies is evident in these tables. The recognition problems turn harder for $n = 3$ and 4, and most the incorrect classification results are ascribed to overfitting between parts of the classes. Table VIII shows the confusion matrix of results for subject B (Dataset 3, four classes, and $L = 30$ s), and the recognition accuracy for this configuration is 45.4%. Data of classes rest and VR(a) were relatively well recognized, however, the other two classes were mis-classified even after the training process converged.

V. CONCLUSION

Differential R-R interval (RRI) time series is utilized in this paper in order to improve training and classification performance of an LSTM-based stress level recognition method. From the experimental results shown in Section IV, training performance has been improved largely by replacing the RRI data with its differential version. Changes of recognition accuracies do not show a strong trend between data length and datasets. When the difficulty of recognition problems increases for three and four classes, deterioration in the accuracies is clear. Even for these conditions the convergence keeps in a relative high level.

Although good recognition accuracies are not obtained in this paper, validity of the proposed method and its outstanding performance comparing to the method developed in [7] have been confirmed. For our future studies, more experiments are required for detailed investigation of the proposed method. Comparison of the accuracies requires more careful and rigorous analysis in order to realize more robust and stronger DL-based end-to-end modeling for stress recognition.

REFERENCES

[1] G. Giannakakis, D. Grigoriadis, K. Giannakaki, *et al.*, “Review on psychological stress detection using biosignals,” IEEE Trans. on Affective Computing, doi: 10.1109/TAFFC.2019.2927337. in press

[2] R. Hartmann, F. M. Schmidt, C. Sander, and U. Hegerl, “Heart rate variability as indicator of clinical state in depression,” Frontiers in Psychiatry, vol. 9, Article–735, 2019.

[3] M. Zubair and C. Yoon, “Multilevel mental stress detection using ultra-short pulse rate variability series,” Biomed. Signal Process. Control, vol. 57, Article-101736, 2020.

[4] K. Masood and M. A. Alghamdi, “Modeling mental stress using a deep learning framework,” IEEE Access, vol. 7, pp. 68446–68454, 2019.

[5] G. Giannakakis, E. Trivizakis, M. Tsiknakis, and K. Marias, “A novel multi-kernel 1D convolutional neural network for stress recognition from ECG,” Proc. 2019 8th Int. Conf. on Affective Computing and Intelligent Interaction Workshops and Demos, 2019, pp. 1–4.

[6] A. Leite, M. E. Silva, and A. P. Rocha, “Classification of HRV using long short-term memory networks,” Proc. 2020 11th Conf. of the European Study Group on Cardiovascular Oscillations, 2020, pp. 1–2.

[7] N. Bu, M. Fukami, and O. Fukuda, “Mental stress level recognition using heart rate variability (HRV) and a deep recurrent neural network,” Proc. 26th Int. Symp. Artificial Life and Robotics, 2021, pp. 44–49.

[8] C. Liu and R. Gao, “Multiscale entropy analysis of the differential RR interval time series signal and its application in detecting congestive heart failure,” Entropy, vol. 19, no. 6, 251, 2017.

[9] N. Bu, “Poincaré analysis based on short-term heart rate variability data for stress evaluation,” Proc. 40th Ann. Int. Conf. IEEE Eng. in Medicine and Biology Society, 2018, pp. 2973–2976.

[10] N. Bu, ”Mental stress tests with autonomic responses induced in virtual reality environment,” Proc. IEEE 2nd Global Conf. Life Sciences and Technologies, 2020, pp. 117–120.

 978-1-6654-1875-1/21/$31.00 ©2021 IEEE