

Mood Prediction in Consideration of Certainty Factor Using Multilayer Deep Neural Network and Storage-Type Prediction Models

Yusuke Kajiwara,* Haruhiko Kimura,¹ and Takashi Oyabu²

College of Information Science and Engineering, Ritsumeikan University,
1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

¹Graduate School of Natural Science and Technology, Kanazawa University,
Kakuma, Kanazawa, Ishikawa 920-1192, Japan

²Kokusai Business Gakuin College, Minami-cyo 6-12, Kanazawa, Ishikawa 920-0919, Japan

(Received August 6, 2015; accepted February 9, 2016)

Keywords: deep neural network, threshold optimization, depression, biological information, weather information

Depression has become a social problem in Japan. To prevent depression, people need to recognize their mental health in daily life. Previous research supports mental health care by predicting tomorrow's mood with 73% accuracy using weather information and biological information. However, the mood after 2 d or later could not be predicted using an existing system. In this paper, we propose multilayer-deep neural network (M-DNN) and storage-type prediction models (STPMs) to predict mood two weeks in advance with high accuracy. The M-DNN outputs predictions as well as unpredictable data using a deep neural network and threshold optimization in each prediction layer. The threshold optimization determines the threshold that maximizes a certainty factor. The certainty factor is calculated from the predictive accuracy of M-DNN and the amount of unpredictable data. The STPMs interpolates the unpredictable data by accumulating the predictions output from M-DNN. The amount of unpredictable data output from the M-DNN is decreased by STPMs. Experiments show that M-DNN and STPMs can predict mood two weeks in advance with 70% accuracy. The predictive accuracy in M-DNN+STPMs is 11% higher than that in DNN. Hence, M-DNN+STPMs is an effective method for mood prediction.

1. Introduction

Depression has become a social problem in Japan. Depressed patients are difficult to cure because these patients think only of bad things and fall into a vicious cycle. Therefore, depression needs to be prevented. To prevent depression, people need to recognize their mental health in daily life. However, they cannot recognize correctly their own mental health.

Previous research⁽¹⁻⁴⁾ supports mental health care by estimating current moods and predicting future moods. C. T. Eagle⁽⁵⁾ defines a mood as an emotional state. Moods have positive valence and negative valence for long times. Moods are changed by body condition and environment. Valenza *et al.*⁽¹⁾ estimated bipolar disorder with 95% accuracy using the heart rate of patients. Yoshino and Matsuoka⁽²⁾ and Kinase and Venture⁽³⁾ estimated the current mood in daily life from

*Corresponding author: e-mail: kajiwara@de.is.ritsumei.ac.jp

a heart rate and muscle potential during walking. On the other hand, Kajiwara *et al.*⁽⁴⁾ predicted tomorrow's mood with 73% accuracy using yesterday's mood, weather information, and biological information. However, the moods 2 d later and beyond could not be predicted using this system. Thus, the prediction system needs to improve predictive accuracy and extend the prediction period for practical use in daily life.

In this study, we proposed multilayer-deep neural network (M-DNN) and storage-type prediction models (STPMs) to predict mood over a long period with high accuracy using biological information and weather information. The prediction error is composed of bias, variance, and irreducible error. Bias is error between target outputs and the expected value of the estimate. Variance is error from fluctuations in the training set. The irreducible error is composed of observation noise and unseen information. The machine learning algorithm has the ability to reduce bias and variance. However, the irreducible error cannot be reduced by a machine learning algorithm because machine learning cannot improve accuracy of observations and cannot obtain unseen information.

Mood is decided by a person's subjective evaluation. Hence, the mood includes much observation noise. An enforced prediction decreases the predictive accuracy. M-DNN does not necessarily predict the future mood from noisy data. M-DNN is composed of prediction layers that are the deep neural network⁽⁶⁾ and threshold optimization. M-DNN outputs a prediction and unpredictable data in each prediction layer. The deep neural network output a class probability, which is an index of prediction difficulty. Hence, if the class probability is less than the threshold, M-DNN outputs unpredictable data. Otherwise, M-DNN outputs the prediction. The threshold optimization determines the threshold that maximizes the certainty factor. The certainty factor is calculated from the predictive accuracy and the amount of unpredictable data. The amount of unpredictable data output from the M-DNN is decreased by STPMs. The STPMs interpolates the unpredictable data by accumulating the predictions output from the M-DNN.

A person who has symptoms such as depressive mood, anorexia, and lack of sleep is likely to be depressed for a long time. Therefore, M-DNN and STPMs predict the mood two weeks later or beyond. Biological information and behavioral information are obtained with a body composition monitor, a blood pressure monitor, and a pedometer. Weather information is obtained online for free.

We explain M-DNN and STPMs in § 2. In § 3, the M-DNN and STPMs are compared with existing machine learning algorithms such as the deep neural network, the random forest, the linear kernel support vector machine, and the radial basis function kernel support vector machine. In § 4, conclusions and suggestions for future work are provided.

2. Mood Prediction in Consideration of the Certainty Factor by M-DNN and STPMs

Figure 1 shows the processing flowchart of a mood prediction system. First, a user uploads the biological information, current mood, and environmental information. Environmental information is the residence of the user. Second, the user assesses the current mental state in the subject on the basis of the biological information and current mood. Third, a prediction system such as M-DNN obtains biological information, current mood, and weather information. Fourth, M-DNN and STPMs output the prediction and unpredictable data from the biological information, current mood, and weather information. Fifth, the user takes precautionary measures against depression on the basis of the predicted future mood. The user can preserve a healthy mental state by taking

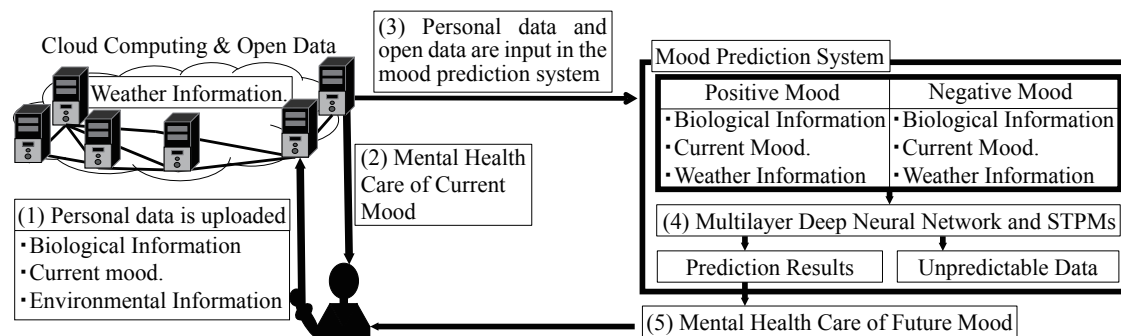


Fig. 1. Processing flowchart of mood prediction system.

precautionary measures and the cure. The mood prediction system is a personal adaptive system as in previous research,⁽¹⁻⁴⁾ and it cooperates with the mood estimation system.⁽¹⁻³⁾

Figure 2 shows a block diagram of M-DNN and STPMs. The M-DNN and STPMs predict the future mood and output unpredictable data. The mood types are positive mood and negative mood. First, the deep neural network in the first prediction layer output the class probability of the future mood by inputting biological information, current mood, and weather information. Deep neural networks in i th layer are defined as $DNN(i)$. Second, the threshold optimization determines the threshold for the maximizing certainty factor in $DNN(i)$. The certainty factor in $DNN(i)$ is calculated from the predictive accuracy of $DNN(i)$ and the amount of unpredictable data. The threshold in the i layer is defined as $T_c(i)$. Third, if the class probability in the future mood is less than $T_c(i)$, M-DNN outputs unpredictable data. Otherwise, M-DNN outputs the prediction. Fourth, $DNN(i)$ outputs the class probability in the future mood by inputting unpredictable data. M-DNN repeats steps 2 to 4 until the amount of unpredictable data is less than 10 or the predictive accuracy in $DNN(i)$ is less than random probability.

The unpredictable data is interpolated using the STPMs. First, the accuracy of the probability is calculated from the precision of M-DNN. The accuracy of the probability is the reliability of accumulating predictions. Second, the threshold optimization determines the threshold for maximizing the certainty factor in STPMs. The certainty factor in STPMs is calculated from predictive accuracy of STPMs and the amount of unpredictable data. The threshold is defined as T_a . Third, if the accuracy of the probability of the future mood is less than T_a , STPMs outputs the unpredictable data. Otherwise, STPMs outputs the prediction.

2.1 M-DNN

M-DNN is composed of prediction layers that are the deep neural network and the threshold optimization. Deep neural networks output a class probability. The class probability is an index of the difficulty of the prediction. The deep neural networks consist of hidden layers. The hidden layers have some units in each layer.

First, the deep neural networks learn explanation variables by unsupervised learning in pre-training as follows. The first hidden layer learns explanation variables using an autoencoder.⁽⁷⁾ The z ($z = 2, 3, \dots, Z$) hidden layer learns output results in the $z-1$ hidden layer with the autoencoder. The Z layer outputs a feature vector. Second, the networks of hidden layers are adjusted by fine turning.

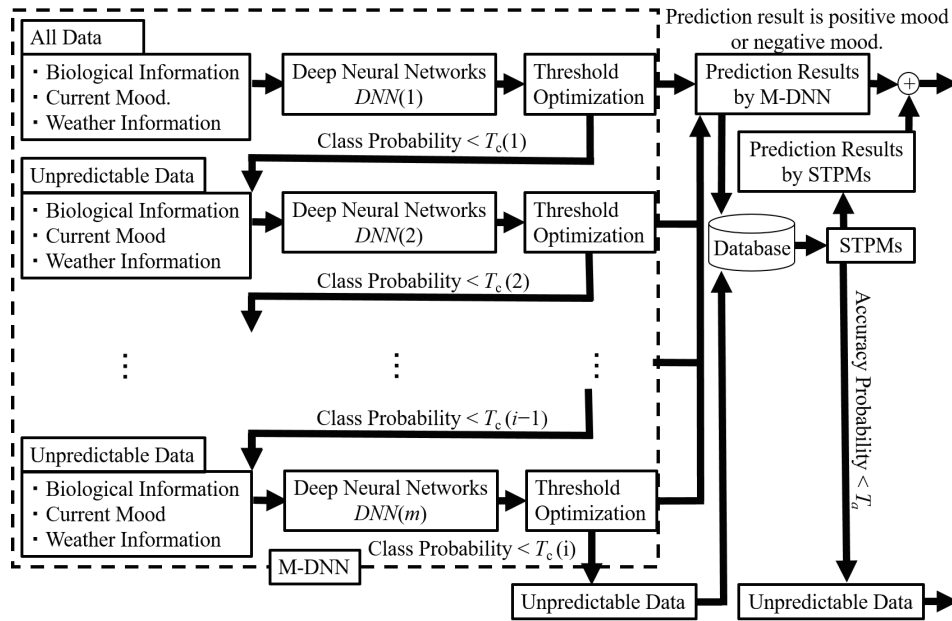


Fig. 2. Outline block diagram of M-DNN and STPMs.

The fine turning learns feature vectors output from the Z layer and recognizes the objective variable by supervised learning. The deep neural network outputs the class probability of each objective variable as

$$\sum_{y \in Y} c(y, b, q) = 1, \tag{1}$$

where Y is an objective value set, y is an objective value, b is the current day, q is a future day, and $c(y, b, q)$ is the class probability of each objective variable. If $c_{\max}(b, q)$ is less than $T_c(i)$, the deep neural network in the i layer outputs unpredictable data. Otherwise, the deep neural network in the i layer outputs $y_{\text{est}}(b, q)$ as

$$y_{\text{est}}(b, q) = \operatorname{argmax}_{y \in Y} c(y, b, q), \tag{2}$$

$$c_{\max}(b, q) = \max_{y \in Y} c(y, b, q), \tag{3}$$

where $c_{\max}(b, q)$ is the maximum class probability in $c(y, b, q)$, $T_c(i)$ is the class probability threshold, and $y_{\text{est}}(b, q)$ is the result of predicting the mood after day q . $T_c(i)$ is determined from maximum certainty factor in $DNN(i)$. The certainty factor maximizes the accuracy of the prediction of $DNN(i)$ and minimizes the amount of unpredictable data as

$$T_c(i) = \operatorname{argmax}_{t_c \in [0,1]} E_c(i, t_c), \tag{4}$$

$$E_c(i, t_c) = \frac{f_c(i, t_c)}{\max_{t_c(i) \in [0,1]} f_c(i, t_c) - f_c(i, 0)} + \frac{1 - u_c(i, t_c)}{(1 - u_c(i, 0)) - \min_{t_c(i) \in [0,1]} (1 - u_c(i, t_c))}, \tag{5}$$

where $E_c(i, t_c)$ is a certainty factor in $DNN(i)$, and t_c is the threshold of the class probability. The term $f_c(i, t_c)$ is an f-measure in applying t_c as

$$f_c(i, t_c) = \frac{2p_c(i, t_c)r_c(i, t_c)}{p_c(i, t_c) + r_c(i, t_c)}, \tag{6}$$

where $p_c(i, t_c)$ is the prediction in applying t_c , and $r_c(i, t_c)$ is recalled in applying t_c . The term $u_c(i, t_c)$ is the percentage of $d_c(i)$ to $d_u(i, t_c)$ as

$$u_c(i, t_c) = \frac{d_u(i, t_c)}{d_c(i)}, \tag{7}$$

where $d_c(i)$ is the amount of input data in $DNN(i)$, and $d_u(i, t_c)$ is the amount of unpredictable data in applying t_c .

2.2 STPMs

Figure 3 shows the outline of STPMs. The term b is a current date; q is the day being predicted. In Fig. 3, a mood in (b, q) equals a mood in $(b-\beta, q+\beta)$ $\{\beta = 1, 2, \dots, Q-q\}$. Hence, STPMs interpolates the unpredictable data in (b, q) from $y_{est}(b-\beta, q+\beta)$. If $a_{max}(b, q)$ for a future mood is less than T_a , STPMs outputs the unpredictable data. Otherwise, STPMs outputs the prediction as

$$y_{est}(b, q) = \operatorname{argmax}_{y \in Y, 1 \leq \beta \leq Q-q} p_m(M(b-\beta, q+\beta), y), \tag{8}$$

$$a_{max}(b, q) = \max_{y \in Y, 1 \leq \beta \leq Q-q} p_m(M(b-\beta, q+\beta), y), \tag{9}$$

where $a_{max}(b, q)$ is the maximum accuracy probability, Q is a future day set, $M(b, q)$ is M-DNN for predicting the mood after q d at b date, and $p_m(M(q), y)$ is the prediction of $M(q)$ in each objective

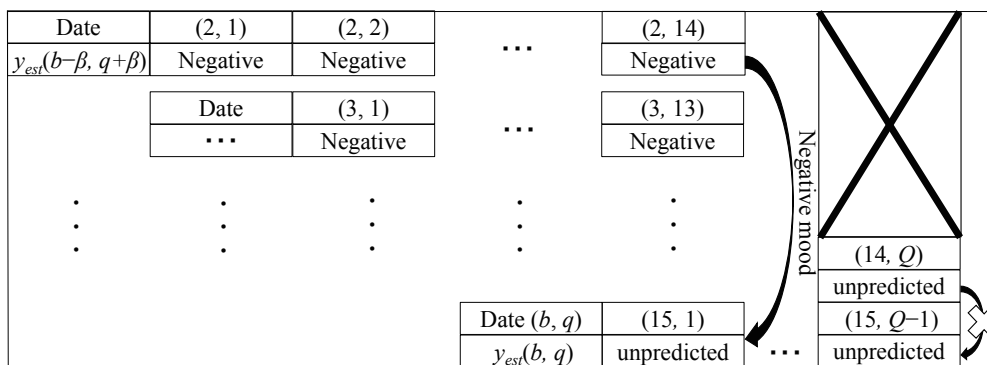


Fig. 3. STPMs interpolates unpredictable data.

value. The term T_a is determined from the maximum certainty factor in STPMs. The certainty factor maximizes the predictive accuracy and minimizes the amount of unpredictable data as

$$T_a = \operatorname{argmax}_{t_a \in [0,1]} E_a(t_a), \quad (10)$$

$$E_a(t_a) = \frac{f_a(t_a)}{\max_{t_a \in [0,1]} f_a(t_a) - f(0)} + \frac{1 - u_a(t_a)}{(1 - u_a(0)) - \min_{t_a \in [0,1]} (1 - u_a(t_a))}, \quad (11)$$

where $E_a(t_a)$ is a certainty factor in STPMs, and t_a is the threshold of the accuracy probability. The term $f_a(t_a)$ is the f-measure in applying t_a as

$$f_a(t_a) = \frac{2p_a(t_a)r_a(t_a)}{p_a(t_a) + r_a(t_a)}, \quad (12)$$

where $p_a(t_a)$ is the prediction for applying t_a , and $r_a(t_a)$ is recalled in applying t_a . The term $u_a(t_a)$ is the percentage of d_a to d_u as

$$u_a(t_a) = \frac{d_u(t_a)}{d_a(i)}, \quad (13)$$

where d_a is the amount of input data in STPMs, and $d_u(t_a)$ is the amount of unpredictable data in applying t_a .

3. Evaluation of M-DNN and STPMs

3.1 Experimental environment

The subject was a 64-year-old healthy man who is an office worker. The measurement period was between 1 July 2013 and 30 June 2015. Table 1 shows the sample size of each mood. The location was Kanazawa City. The subject's biological information was measured with a body composition monitor (Tanita Company, BC-503), a blood pressure monitor (Tanita Company, BP-301), and a pedometer (Tanita Company, FB-723). BC-503 measured the subject's weight (WT), body fat (BF), body mass index (BMI), and muscle mass (MM). BP-301 measured the subject's systolic blood pressure (SBP), diastolic blood pressure (DBP), and pulse (PS). FB-723 measured the number of subject's walking steps (WS). Weather information was obtained for free from the Japan Meteorological Agency website. We obtained maximum temperature (MXT), minimum temperature (MNT), precipitation (PN), snowfall (SF), snow depth (SD), sunshine hours (SH), maximum wind speed (MW), relative humidity (RH), mean of cloudiness (MC), air pressure (AP), and vapor pressure (VP). The mood (MD) was determined on a scale of one to five using questionnaires. Answers of 1 or 2 on the questionnaire are defined as a negative mood. Answers of 3 on the questionnaire are defined as neutral. Answers of 4 or 5 on the questionnaire are defined as a positive mood. The mood and the biological information were measured once a day between 15:00 and 18:00.

Table 1
Sample size of moods on each prediction date.

Prediction date	Sample size of moods		
	Negative	Positive	Neutral
1	79	98	552
2	79	98	551
3	79	98	550
4	79	98	549
5	78	98	549
6	77	98	549
7	77	98	548
8	76	98	548
9	76	97	548
10	76	96	548
11	76	95	548
12	76	95	547
13	76	94	547
14	76	93	547

3.2 Experimental methods

We compared M-DNN and STPMs with each of the classifiers. The classifiers were the deep neural network (DNN), naïve bayes (NB),⁽⁸⁾ Random Forest (RF),⁽⁹⁾ a linear kernel support vector machine (SL),⁽¹⁰⁾ and a radial basis function kernel support vector machine (SR) in the R library for statistical computing. The DNN parameter was configured as the number of units, the number of layers, and activation function. The number of units was sixteen. The number of hidden layers was two. The activation function was a hyperbolic tangent function. Parameters for the other classifiers were configured as defaults in the R library. The explanation variables are the biological information today, the mood today, and the weather information until q d later. The objective variable was the mood after q days. Positive mood and negative mood were predicted up to 14 d later. Evaluation was conducted by a leave-one-out cross-validation (LOOCV)⁽¹¹⁾ in which the f-measure was calculated.

3.3 Experimental results and discussion

Figure 4 shows the predictive accuracy of M-DNN+STPMs, M-DNN, and the other classifiers. The number of layers in M-DNN was nine. From the figure, the average of the f-measure in M-DNN+STPMs was 0.69. M-DNN+STPMs has the highest predictive accuracy in the machine learning for comparison. The average of the f-measure in M-DNN+STPMs was 0.11 higher than that of the DNN. Figure 5 shows the f-measure of M-DNN+STPMs and M-DNN when the number of layers was changed. From the figure, M-DNN with nine layers had a predictive accuracy equal to that of M-DNN with one layer. The difference in the f-measure between M-DNN and M-DNN+STPMs decreased as the number of layers increased. Figure 6 shows the percentage of unpredictable data in M-DNN+STPMs and M-DNN when the number of layers changed. From the figure, the percentage of unpredictable data in M-DNN with nine layers is 0.49 lower than that in M-DNN with one layer. M-DNN focused on prediction of the unpredictable data as prediction

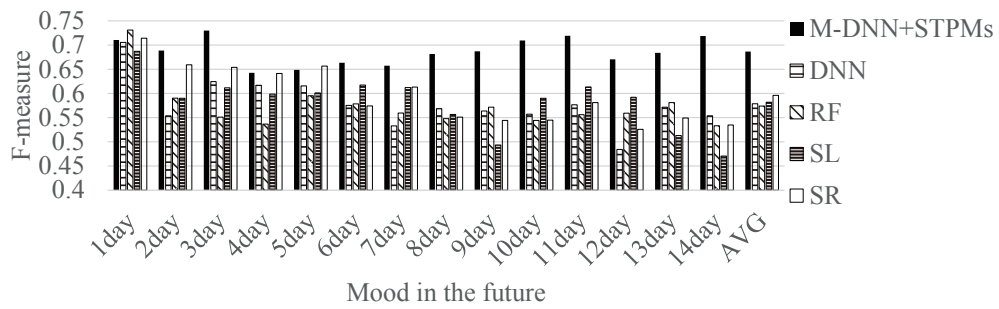


Fig. 4. Prediction accuracy of machine learning.

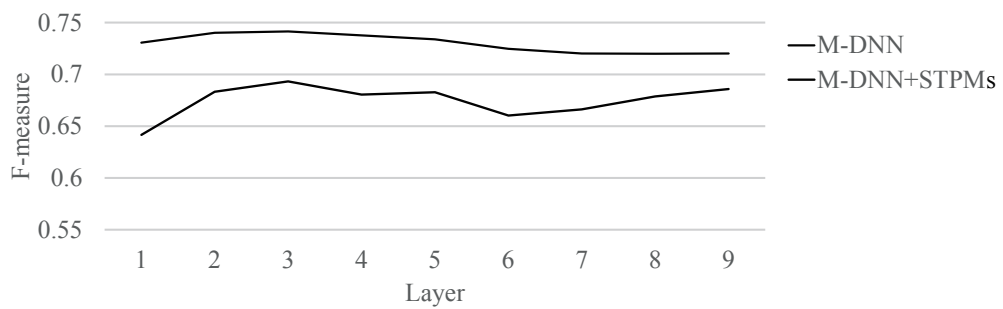


Fig. 5. The f-measure of M-DNN+STPMs and M-DNN when the number of layers changed.

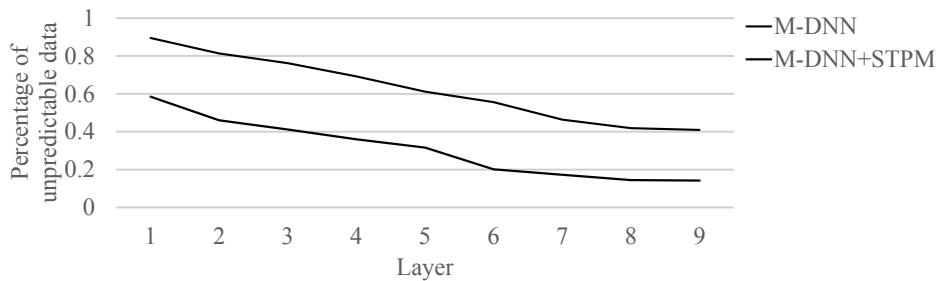


Fig. 6. Percentage of unpredictable data in M-DNN+STPMs and M-DNN when the number of layers changed.

layers increased. $DNN(1)$ learns common patterns in input data from the majority of all input data to maximize predictive accuracy. If any input data have a majority pattern, $DNN(1)$ outputs high class probability, and the first prediction layer outputs the prediction result. On the other hand, if any input data have a minority pattern, $DNN(1)$ outputs low class probability, and the first prediction layer outputs the unpredictable data. $DNN(2)$ learns common patterns in unpredictable data from the majority of all unpredictable data output from $DNN(1)$ to maximize the predictive accuracy. Thus, $DNN(i)$ can focus more on prediction of the unpredictable data than can $DNN(i-1)$. Hence, M-DNN maintains the predictive accuracy and decreases the percentage of unpredictable data by increasing the number of layers. Figure 7 shows the f-measure of M-DNN+STPMs and M-DNN with nine layers. The average of the f-measure in M-DNN+STPMs is 0.03 less than that in M-DNN. Figure 8 shows the percentage of unpredictable data of M-DNN+STPMs and M-DNN

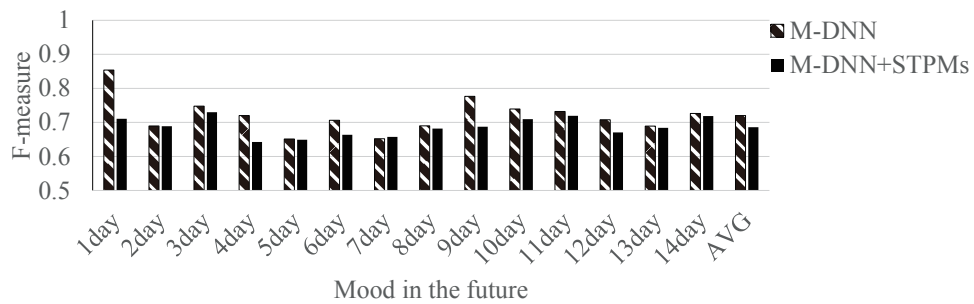


Fig. 7. The f-measure of M-DNN+STPMs and M-DNN up to 14 d later.

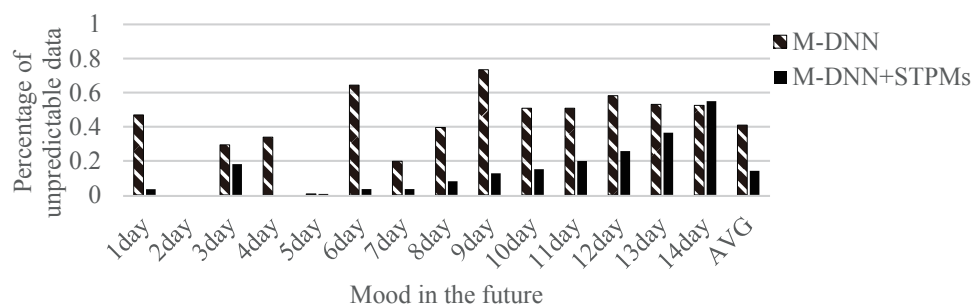


Fig. 8. Percentage of unpredictable data of M-DNN+STPMs and M-DNN up to 14 d later.

with nine layers. From the figure, the percentage of unpredictable data in M-DNN+STPMs is 0.27 less than that in M-DNN. The percentage of unpredictable data in M-DNN+STPMs increased as prediction periods q increased. The number of interpolations of unpredictable data was increased as the number of the predicted results by M-DNN increased. For example, STPMs can interpolate tomorrow's mood by accumulating the predictions up to $Q-q$ d.

If the f-measure of a prediction system decreases, the reliability of the prediction system decreases. In addition to that, if a prediction system outputs a lot of unpredictable data, the practicability of prediction system decreases. Thus, the prediction system should be evaluated by the f-measure and the percentage of unpredictable data. From the F7, the difference in the f-measure between M-DNN+STPMs and M-DNN is -0.03 . From Fig. 8, the difference in the percentage of unpredictable data between M-DNN+STPMs and M-DNN is 0.27. Thus, the sum of the difference values is 0.24. As a result, M-DNN+STPMs is better than M-DNN. Moreover, the experimental results have shown that M-DNN+STPMs can predict mood over a long period better than existing system.⁽⁴⁾

Table 2 shows the important variable of random forest. From the table, today's SBP is an important variable in predicting the mood up to 2 d later. Today's SH and yesterday's SH are important variables in predicting the mood after 3 d and beyond. The average of yesterday's SH in a positive mood is 8.1 ± 0.4 h. The average of today's SH in a positive mood is 8.8 ± 0.9 h. The average of yesterday's SH in a negative mood is 7.4 ± 0.2 h. The average of today's SH in a negative mood is 7.7 ± 0.2 h. Hence, the mood is positive as today's SH, and yesterday's SH increases.

Table 2
Variables importance in predicting mood until 14 d later.

Prediction date q	Variables importance (variable in q d)				
	First	Second	Third	Fourth	Fifth
1	MD(today)	IS(today)	IS(1day)	BX(today)	BF(today)
2	IS(1day)	IS(2day)	AF(1day)	IS(today)	BX(today)
3	IS(2day)	IS(3day)	AF(2day)	IS(today)	AF(3day)
4	IS(3day)	IS(4day)	AF(3day)	AF(4day)	IS(2day)
5	IS(4day)	IS(5day)	AF(4day)	BX(today)	AF(5day)
6	IS(5day)	IS(6day)	VP(today)	AF(5day)	TX(today)
7	IS(6day)	IS(7day)	VP(1day)	AF(6day)	TN(today)
8	IS(7day)	IS(8day)	PS(today)	AF(7day)	TX(today)
9	IS(8day)	IS(9day)	AF(8day)	VP(3day)	TX(2day)
10	IS(9day)	IS(10day)	AF(9day)	VP(4day)	TX(3day)
11	IS(10day)	IS(11day)	DN(today)	AF(10day)	IS(today)
12	DN(today)	IS(11day)	IS(12day)	DN(1day)	AF(11day)
13	DN(1day)	IS(12day)	DN(today)	IS(13day)	IS(today)
14	DN(2day)	IS(13day)	DN(today)	IS(14day)	IS(1day)

4. Conclusion and Future Works

We proposed a predictive method combining M-DNN and STPMs. Experiments show that M-DNN and STPMs can predict mood up to two weeks with 70% accuracy. The predictive accuracy in M-DNN+STPMs is 11% higher than that in DNN. Hence, the M-DNN+STPMs is effective for personalized mood prediction. However, the number of subjects in this study was small. Thus, we will apply the M-DNN+STPMs to many subjects, and we will verify the efficacy of M-DNN+STPMs. On the other hand, the mood prediction system needs to function without biological information to predict mood over a long period. The prediction system used weather information but did not use a weather forecast. In future work, we will predict future moods using a weather forecast. Moreover, we would like to improve predictive accuracy.

References

- 1 G. Valenza, M. Nardelli, A. Lanata, C. Gentili, G. Bertschy, R. Paradiso, and E. P. Scilingo: IEEE J. Biomed. Health Inf. **18** (2014) 1625.
- 2 K. Yoshino and K. Matsuoka: International Conference on Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS) (2012) 1009.
- 3 Y. Kinase and G. Venture: 35th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC) (2013) 4981.
- 4 Y. Kajiwaru, M. Nakamura, H. Kimura, and T. Oyabu: IEEJ Trans. Electron. **135** (2015) 570 (in Japanese).
- 5 C. T. Eagle: Bull. Counc. Res. Music Educ. **32** (1973) 55.
- 6 Y. Bengio: Found. Trends Mach. Learn. **2** (2009) 1.
- 7 G. E. Hinton and R. R. Salakhutdinov: Science **313** (2006) 504.
- 8 D. Pedro and M. Pazzani: Mach. Learn. **29** (1997) 103.
- 9 L. Breiman: Mach. Learn. **45** (2001) 5.
- 10 C. Cortes and V. Vapnik: Mach. Learn. **20** (1995) 273.
- 11 R. Kohavi: Proc. 14th Int. Joint Conf. Artificial Intelligence **2** (1995) p. 1137.