Sensors & Transducers

Published by IFSA Publishing, S. L., 2020

http://www.sensorsportal.com

Classification Accuracy of Support Vector Machine, Decision Tree and Random Forest Modules when Applied to a Health Monitoring with Flexible Sensors

[1, *] Yasutaka Uchida, [2] Tomoko Funayama and [3] Yoshiaki Kogure

[1] Dept. of Life Science, Teikyo Univ. of Science, 2-2-1 Senjusakuragi, Adachi-ku, Tokyo, 120-0045, Japan

[2] Dept. of Occupational therapy, Teikyo Univ. of Science, 2525 Yatsusawa, Uenohara-shi, 409-0193, Yamanashi, Japan

[3] Professor Emeritus, Teikyo Univ. of Science, 2-2-1 Senjusakuragi, Adachi-ku, Tokyo, 120-0045, Japan

[1] Tel.: +81-3-6910-1010, fax: +81+3+6910-3800

E-mail: uchida@ntu.ac.jp

Received: 20 August 2020 /Accepted: 24 September 2020 /Published: 30 October 2020

Abstract: In this study, distinct machine learning-based methods of analysis were used to evaluate the most effective feature combination for a previously proposed system which detects changes in the physical conditions of human beings through eight sensors placed on various locations of a bed and floor near the bed. First, the proposed system was applied to evaluate the presence or absence of restrictions in the movement of the knee joint in two subjects of different age. A set of measurements were obtained from four classes, under distinct scenarios, with 10 case in each class. Then, 30 % of the measured feature values were randomly extracted for testing. Scikit-learn was used as the machine learning library, and the k-nearest neighbor method, logistic regression, perceptron, and multilayer perceptron were used as the four learning models in the first step. The k-nearest neighbor method assigned the highest accuracies to all three feature value combinations that were investigated in this study. The accuracy and feature value of the distribution, which could be high or low, were investigated. It was found that the analysis methods corresponding to the feature value at specific locations were effective. The decision tree method, random forest and support vector machine methods were found to meet our proposed system. The accuracies ranged from 0.6 to over 0.8 when t2-t7 and t4-t7 feature values were combined.

Keywords: Pressure sensor, Healthcare monitoring system, Machine learning, Scikit-learn.

1. Introduction

It is important for elderly people and persons with disabilities to be aware of any health issues as early as possible, particularly if they are living alone. Therefore, it is a beneficial challenge to develop an automatic system that can monitor the health conditions of elderly people on a daily basis and issue warnings before their health conditions deviate from normal standards. Various type of health and welfare equipment, which are primarily reliant on the Internet of Things, have been developed over the past years. [1-7] In particular, human motion monitoring and analysis can be used to detect unusual conditions in people with disabilities. Thus, based on this principle, we proposed a system to detect changes in the physical conditions of human beings using an eight-sensor-system and reported the results of the machine learning approaches implemented in the study. [8-13].

In this study, three of the measurements were carried out, where presence or absence of restrictions in the movement of knee or shoulder joint movements were investigated in two subjects of different ages. Four types of machine learning-based methods were used to determine the accuracies of the three types of measurements.

Based on the results, several methods in the sikit-learn module were compared using the default conditions and algorithms suitable for the measurement system were investigated. The accuracy and features of the distribution, which could be high or low, were investigated. It was found that the methods corresponding to the features in specific locations were effective. The decision tree, random forest and support vector machine methods were found to meet the condition. The applications to the methods was discussed in this paper.

This research was conducted with approval of the Teikyo University of Science's Ethics Committee.

2. Experiment

2.1. Healthcare Monitoring System and Data

As depicted in Fig. 1, the system proposed in our previous report consisted of eight flexible force-sensing resistors, a data acquisition device, and a personal computer. One sensor (s1) was placed on a pillow, three sensors (s2–s4) were placed apart from each other at the edge of a bed, and four more sensors (s5–s8) are placed on the floor near the bed. The output signals were selected from four of the eight sensors.

Feature values t2 to t7, corresponding to sensors s2 to s7, were defined based on the respective time differences between the arrival of the output signals from s1, which specified the initiation of the measurement, and the corresponding sensors, as depicted in Fig. 2. For waveforms, in which the rise time of the signal could not be detected, the final 10 s was used as the rise time of the sensor. This problem can be remedied by modifying the arrangement of the sensors. Existing studies have established the best candidate combinations of feature values to be t2-t7, t3-t9 and t4-t7 [7-11]. In this study, scikit-learn was used as a machine learning library and the k-nearest neighbor method, logistic regression, perceptron, and multilayer perceptron were used as classification modules in the first step.

In the k-nearest neighbor method, the default distance function was the Minkowski distance, the number of neighbors was 5, the weighting parameter was uniform, and the default algorithm was auto. In the logistic regression method, the default value was 1 to prevent over-learning, penalty was set to L2 to prevent over-learning, and the solver was set to liblinear to prevent over-learning. For the perceptron method, the number of iterations was set to 100 and the other values were set to the default values. For the multi-layered perceptron method, the number of iterations was set to 700, the activation function is set to relu, and other values are set to the default values. In the support vector machine method, we used the Gaussian kernel, rbf, which was reported in a previous report; however, we did not compare it with the other methods because of over-training.

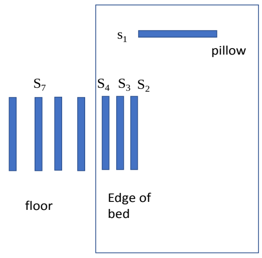


Fig. 1. Position of flexible force-sensing resistors.

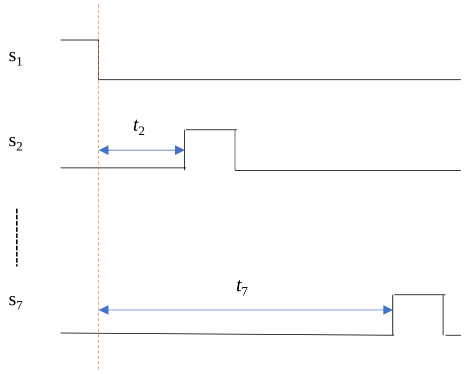


Fig. 2. The definition of the time difference as future values.

2.2. Survey Method

Measurements data were collected by applying the proposed system to two men in their 60 s and 70 s, under two separate conditions: normal range of motion and simulated limited right knee motion. In all instance, each set of learning and test data was created by randomly sampling 70 % and 30 % of the measurement data as learning data and test data, respectively.

3. Results

3.1. Dependence of Analytical Modes

Fig. 3 depicts the accuracies of the t2-t7 feature value combinations under the default settings, which were achieved using various machine learning-based methods of analysis. In each method, 10 sets of learning and test data were constructed by randomly sampling 70 % and 30 % of the measurement data as the learning and test data, respectively, on 10 separate instances. Subsequently, each set of learning and test data, producing 10 accuracy values for each method of analysis. In the figure, KN, Logi, Perc, and MLP denote the k-nearest neighbor method, logistic regression, perceptron and multiple perceptron, respectively. The k-nearest neighbor method assigned the highest accuracies for all predicted values, with 0.54.

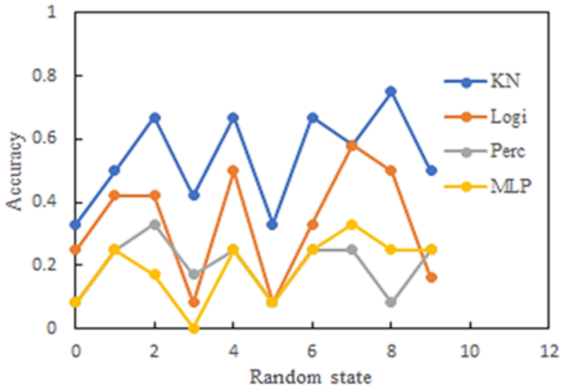


Fig. 3. Difference by analytical model.

3.2. Accuracies of the Three Feature Combinations

Fig. 4 shows the accuracies of the three feature value combinations, t2-t7, t3-t7, and t4-t7, which were evaluated using the k-nearest neighbor method with the randomly extracted datasets. The mean accuracies of feature combinations t2-t7, t3-t7, and t4-t7 over the sample datasets were 0.54, 0.51 and 0.49, respectively. Thus, based on the k-nearest neighbor method, the t2-t7 feature combination was the most accurate.

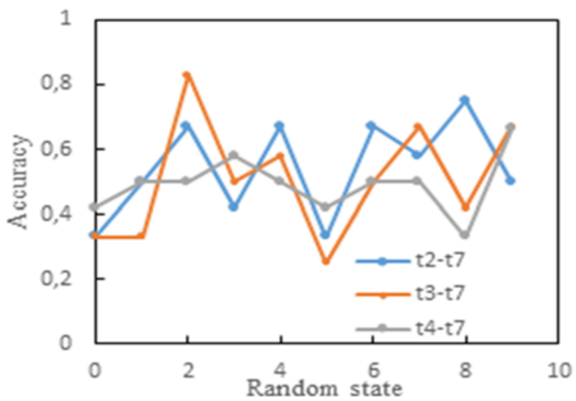


Fig. 4. Difference in accuracy due to differences in feature values.

3.3. Analysis of Accuracy Distribution

As the three types of feature value combinations demonstrated a significant variation in accuracy for the test data, factors affecting the accuracy were investigated. At this stage, t2-t7 was considered.

First, we examined the data extraction rate used for the tests. Although a value of 30 % is commonly used, we adopted a value of 20 % to increase the number of learning data considering the dataset, and a value of 40 % considering the substantial variability of the data.

Fig. 5 illustrates the three extraction rate accuracies evaluated by the k-nearest neighbor method using the randomly extracted datasets. The various extraction rates showed almost similar trend.

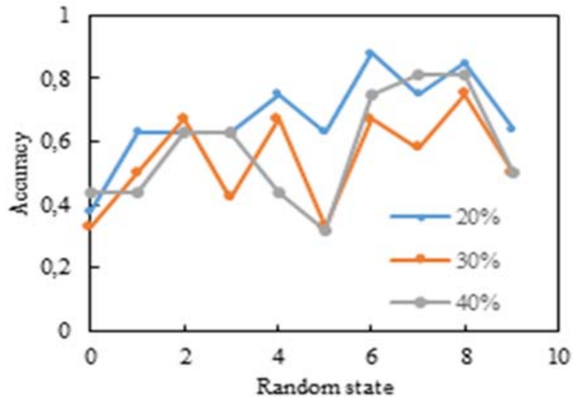


Fig. 5. Three extraction rate accuracy.

The influence of the percentage of the test data on accuracy, including standard errors, is shown in Fig. 6. The mean accuracy values for the test data selection percentages of 20 %, 30 %, and 40 % are 0.68, 0.54, and 0.58, respectively. In the 20 % extraction rate, there were cases in which the extracted test data did not include data points corresponding to the subject's conditions. It was found that 30 % was optimum extraction rate considering both accuracy and standard error.

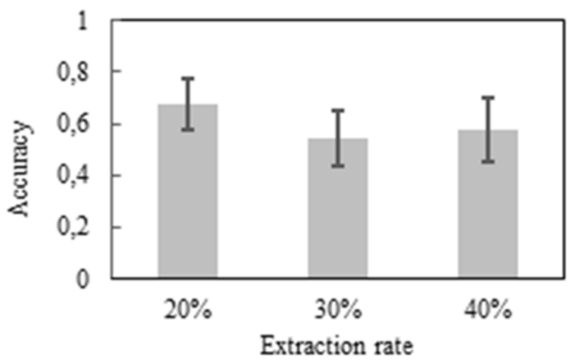
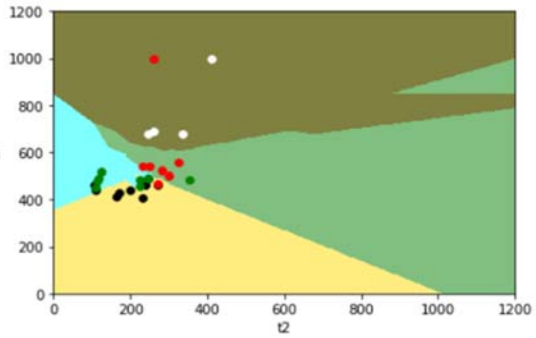


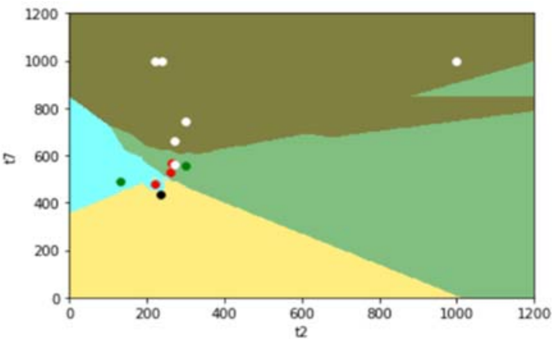
Fig. 6. The influence of extraction rate.

3.4. Analysis of Accuracy Using 2-D Distribution

For feature value combination t2-t7, we investigated the main causes for high and low accuracy using a two-dimensional distribution diagram. First, the distribution of points corresponding to each exercise limitation condition was examined for the high-accuracy case. Fig. 7 depicts the case where a high accuracy of 0.75 was achieved. In the learning data, subject A, represented in the figure by the black circle for normal range motion and green circle for simulated limited right knee motion, demonstrated no specific variation in the presence or absence of motion limitation. For subject B, represented by the red circle for normal range motion and white circle for simulated right knee motion demonstrated no specific variation.



(a) Learning data



(b) Test data

Fig. 7. High accuracy case.

Fig. 8 shows the learning data for low accuracy for features t2-t7, t3-t7 and t4-t7. We found that the shape of the segmented area differed significantly depending on the feature values. Furthermore, in this case, we believe that the difference was largely dependent on the variation of the data for the unrestricted red and white circles of subject B. We observed that the segmented area differed considerably due to the difference in the feature amount and was influenced significantly by the accuracy of the data.

As many of the feature values used a value of 10 s, which was the time the sensor was unresponsive, we believe that this could have been the cause of the common distribution of low-accuracy conditions.

To deal with the data distributed in the upper and rightmost parts of the distribution area, we changed the algorithm to one that uses linear boundaries. We studied three types of distributions: the decision tree, random forest, and support vector machine methods with a kernel as linearly distributed data. In the decision tree method, the default value of gini was set as the growth criterion, the maximum depth of branching was set to 3, and the number of samples required for new branching was set to 2. For the random forest method, the number of trees to be ensembled was set to the default value of 10, and the number of dimensions to be used for randomly selecting features was set to 10. For the support vector machine method, the default parameter for the soft margin penalty was 1 to reduce the risk of overlearning.

3.5. Decision Tree Analysis

Fig. 9 depicts the accuracies of the t2-t7, t3-t7 and t4-t7 feature value combinations evaluated using a decision tree classifier with randomly extracted datasets. In the figure, nb denotes the combination of the feature values from the normal range and simulated limited right knee motions, ns denote that from the normal range and simulated right shoulder motions, bs denote that from the simulated right knee and shoulder motions. The following figures use the same abbreviations.

For the simulated right knee and shoulder motions, the accuracy of random sampling was higher when t2-t7 and t4-t7 were used as feature values. The highest average accuracy of normal range and simulated limited shoulder motions was 073. However, the combination of normal range and simulated knee motions resulted in a lower accuracy of less than 0.54 for all feature combinations.

In most cases that had low accuracy, where there was presence or absence of restrictions, it was observed that no signal from sensor 7 was obtained. This is due to the high dependence of accuracy on the data of a specific subject, suggesting that accuracy can be improved by improving the arrangement of the sensors appropriate for the subject.

3.6. Random Forest Analysis

Fig. 10 depicts the accuracies of the t2-t7, t3-t7 and t4-t7 feature value combinations evaluated using a random forest classifier with the randomly extracted datasets. Differences were found between combinations of motion-restricted conditions; however, relatively high accuracies were obtained for the different motion restricted conditions. The accuracy was lower for the combination of normal and simulated limited knee motions.

This method of analysis also showed that accuracy was greatly influenced by the condition of the knee. Therefore, optimizing sensor placement and feature extraction to match the knee condition may be useful.

3.7. Support Vector Machine Analysis

Fig. 11 depicts the accuracies of the t2-t7, t3-t7 and t4-t7 feature value combinations evaluated using support vector machine classifier, corresponding to the randomly extracted datasets.

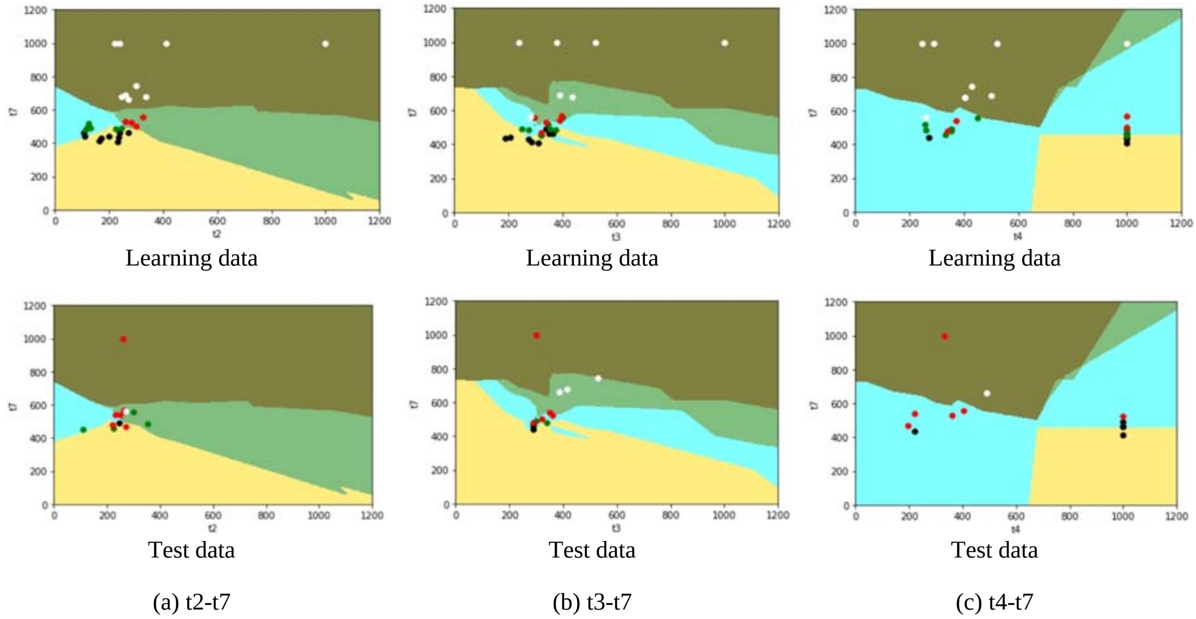


Fig. 8. Low accuracy for learning date of three feature values.

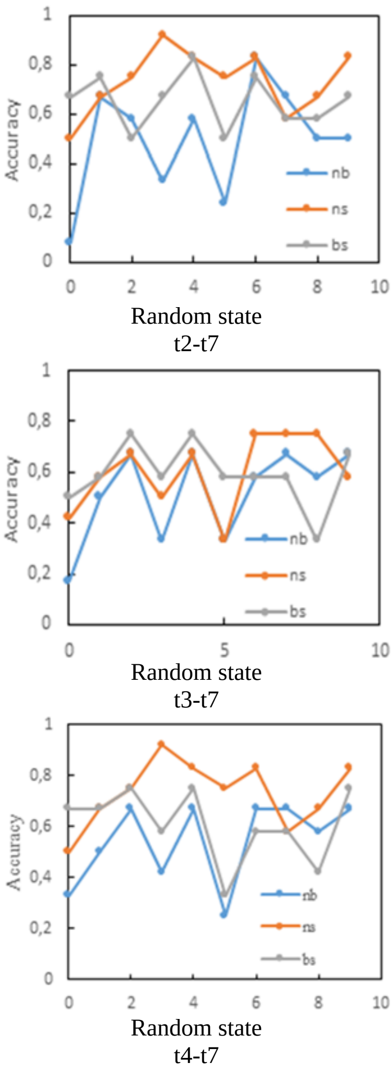


Fig. 9. Accuracies of the t2-t7, t3-t7 and t4-t7 feature value combinations evaluated using the decision tree method.

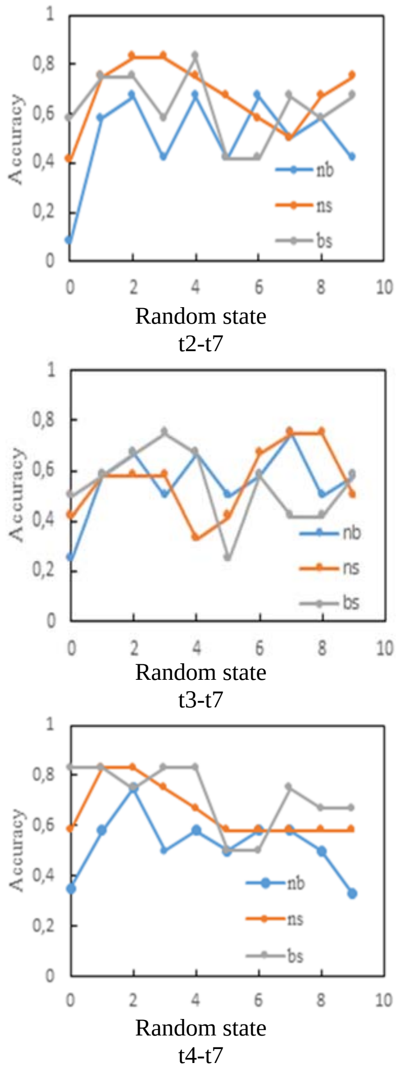


Fig. 10. Accuracies of the t2-t7, t3-t7 and t4-t7 feature value combination evaluated using the random forest method.

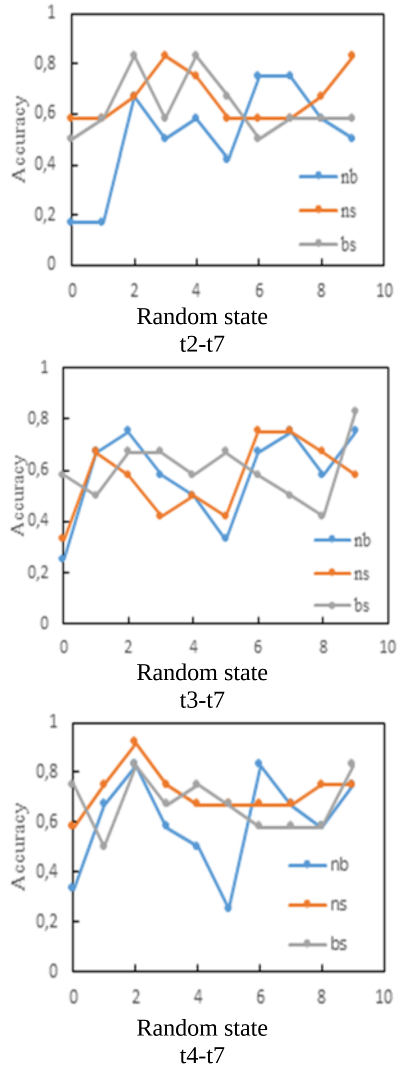


Fig. 11. Accuracies of the t2-t7, t3-t7 and t4-t7 feature value combinations evaluated using a support vector machine.

Large differences were found for different simulated limited motions. When normal range motion and simulated limited knee motion were combined, the accuracy of the three combinations of feature values varied greatly. For the combined normal range and simulated limited motions and combined simulated limited knee and shoulder motions, the accuracies range from 0.3 to over 0.8 when t2-t7 and t4-t7 feature values were adopted.

3.8. Comparison of Accuracy for the Different Simulated Limited Motion

Fig. 12 shows the accuracies for the different simulated limited motions from all feature value combination and analysis methods. The error range is also shown in the figure. As the data were averaged, the accuracy value was relatively low at approximately 0.6. In the previous sections, we found that there was significant difference in the accuracy between simulated limited knee motion and the normal range motion. It was found that the accuracy of the study can be improved by analyzing the walking movement of the subjects with simulated limited knee motion and by installing sensors based on the results of this analysis.

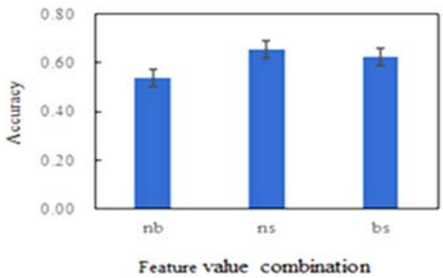
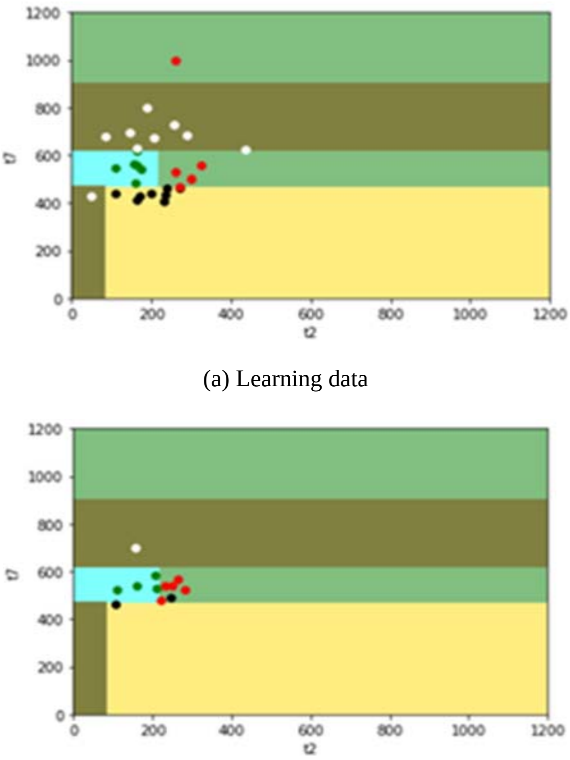


Fig. 12. Accuracies for the different simulated limited motions.

Fig. 13 shows the learning and the test data, which resulted in the highest accuracy of 0.92 using the decision tree method with the random state=3. High accuracy was achieved as there were no case that had no output signal from the sensor, which could reduce accuracy.

4. Conclusions

We evaluated the performance of a system to detect changes in human health conditions using four distinct machine-learning-based methods of analysis. First, the aforementioned system was used to monitor two subjects of different ages under two separate scenarios: no restrict motion and restricted motion for the right knee. Then a set of measurements was obtained. Based on multiple learning and test datasets, which were constructed using the recorded measurements, it was found that the k-nearest neighbor method resulted in highest mean accuracy of 0.54 for the t2-t7 feature combination. The accuracies of all 10 sets of training and test data exhibited considerable variability, ranging between 0.33 and 0.7, owing to an insufficient quantity of training data and the lack of optimization via other methods.



(a) Learning data

(b) Test data

Fig. 13. Distribution data showing the highest accuracy, which was analyzed using the decision tree.

Based on these results, the accuracy and feature value of the distribution, which may be high or low, were investigated, and we found that the analysis methods corresponding to the feature values in specific conditions were effective. It was concluded that decision tree method, random forest method and support vector machine method were suitable for our healthcare system. The accuracies ranged from 0.6 to over 0.8 when t2-t7 and t4-t7 feature values were combined as feature values.

Further analysis on low accuracy demonstrated the significant impact of specific variability on accuracy. It cannot be applied to actual patients due to COVID19 issue. The effectiveness of the system can be demonstrated in the future by integrating the pressure sensor system and the health condition detection system using machine learning.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP17k01590.

References

[1]. M. M. Alam, E. B. Hamida, Surveying Wearable Human Assistive Technology for Life and Safety Critical Applications: Standards, Challenges and Opportunities, *Sensors*, Vol. 14, 2014, pp. 9153-9209.

[2]. A. Arcelus, C. L. Herry, R. A. Goubran, F. Knoefel, H. Sveistrup, M. Bilodeau, Determination on Sit-to-Stand Transfer Duration Using Bed and Floor Pressure Sequences, *IEEE Tran. on Biomedical Engineering*, Vol. 56, Issue 10, 2009, pp. 2485-2492.

[3]. V. G. Sanchez, C. F. Pfeiffer, Nils-Olva Skeie, A Review of Smart House Analysis Methods for Assisting Older People Living Alone, *Sensor and Actuator Networks*, Vol. 6, Issue 3, 2017.

[4]. J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: a survey, *Pattern Recognition Letters*, Vol. 119, 2019, pp. 3-11.

[5]. G. Yang, J. Deng, G. Pang, H. Zhang, J. Li, B. Deng, Z. Pang, J. Xu, M. Jiang, P. Liljeberg, H. Xie, H. Yang, An IoT-Enable Stroke Rehabilitation system Based on Smart Wearable Armband and Machine Learning, *IEEE J. Translation Engineering in Health and Medicine*, Vol. 6, 2018, 2100510.

[6]. Y. Liu, H. Wang, W. Zhao, M. Zhang, H. Qin, Y. Xie, Flezible, Stretchable Sensors for Wearable Health Monitoring: Sensing Mechanisms, Materials, Fabrication Strategies and Features, *Sensors*, Vol. 18, Issue 2, 2018, 645.

[7]. D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G Yang, Deep Learning for Health Informatics, *IEEE J. Biomedical and Health Informations*, Vol. 21, Issue 1, 2017, pp. 4-21.

[8]. Y. Uchida, T. Funayama, Y. Asano, Y. Kogure, Proposal for a simple care system using pressure-sensitive resistors at nursing site, in *Proceedings of the Joint International Conference of Service Science and Innovation and Serviceology of ICSSI' 2018 and ICServ2018*, Taichung, Taiwan, 13-15 November 2018, p. 41.

[9]. Y. Uchida, T. Funayama, Y. Asano, Y. Kogure, Possibility for application of the machine learning to simple care system, *Human Interface Society SIG-ACI-22*, Vol. 20, Issue 11, 2018, pp. 59-62 (in Japanese).

[10]. Y. Uchida, Y. Kogure, T. Funayama, Y. Asano, A study on data classification method from healthcare monitoring system using pressure senor, *Human Interface Society SIG-ACI-23*, Vol. 21, Issue 2, 2019, pp. 15-18 (in Japanese).

[11]. Y. Uchida, T. Funayama, Y. Kogure, SVM classification of data obtained from a health condition monitoring system using flexible force sensing resistors, in *Proceedings of the 5th Int. Conference on Sensors Engineering and Electronics Instrumentation Advances (SEIA'2019)*, Canary Island, Spain, 25-27 September 2019, pp. 117-120.

[12]. Y. Uchida, T. Funayama, Y. Kogure, Support Vector Machine Analysis to Detect Deviation in a Health Condition Monitoring System, *Sensors & Transducers*, Vol. 237, Issue 9-10, 2019, pp. 103-108.

[13]. Y. Uchida, T. Funayama, Y. Kogure, Determination of best future combination for healthcare monitoring system based on evaluation via four distinct machine learning-based methods of analysis, in *Proceedings of the 6th Int. Conference on Sensors Engineering and Electronics Instrumentation Advances (SEIA'20)*, Porto, Portugal, 23-25 September 2020, pp. 43-46.