

Received July 7, 2020, accepted July 15, 2020, date of publication July 27, 2020, date of current version August 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3012167

Detection of Health Abnormality Considering Latent Factors Inducing a Disease

KIICHI TAGO¹, (Member, IEEE), KOSUKE TAKAGI¹, AND QUN JIN², (Senior Member, IEEE)

¹Graduate School of Human Sciences, Waseda University, Tokorozawa 359-1192, Japan

²Faculty of Human Sciences, Waseda University, Tokorozawa 359-1192, Japan

Corresponding author: Qun Jin (jin@waseda.jp)

The work was partially supported by 2016–2018 Masaru Ibuka Foundation Research Project on Oriental Medicine, and Waseda University Grants for Special Research Project Nos. 2019C-729 and 2020C-393.

ABSTRACT Underlying latent factors may cause a person to feel unwell. As the influence of the latent factors increases, the person will become sick. It is difficult to directly measure the influence of latent factors on risk degrees. However, early symptoms of a disease may affect vital signs such as body temperature and blood pressure, which may be a result of the influence of the latent factors. Deep learning is often used to predict the onset of a disease owing to its high accuracy. However, the reliability of this method is limited because of its characteristics of a black-box model. In this study, we propose a new approach to detect health abnormality. We regard the degree of influence of latent factors as the risk of disease and detect health abnormality before the onset of the disease. In our approach, we used a combination of Structural Equation Modeling (SEM) and Hidden Markov Model (HMM). First, in SEM, a domain model was created, and the factor score was estimated based on the relationship between latent factors and the explicit variables influenced by the factors. Thereafter, risk degrees were quantified with HMM using the estimated factor scores, and abnormality was identified in terms of risk degree. Finally, our proposed method was compared with three baselines: PCA (principal component analysis)-based approach, deep learning, and no-degree estimation methods. The average recall of our method was 98.75%, almost the same as the baselines, and false positive rate (FPR) was 0.186%, lower than the baselines. In the five-fold cross-validation comparing with no-degree estimation method, the average accuracy and recall of our method were 99.7% and 98.3% respectively, and FPR was 0.045%, all much better than the baseline. Moreover, our approach can make the process of obtaining the result visible and help detect abnormality sensitively by setting the threshold according to the risk degree, which can contribute to early detection of a disease and improve the reliability of abnormality detection as well.

INDEX TERMS Abnormality detection, personal health data, structural equation modeling (SEM), hidden Markov model (HMM).

I. INTRODUCTION

Health is important to lead a high quality of life. In recent years, with the widespread development of wearable devices, one can easily measure vital signs on their own, without visiting hospitals and attending medical checkups. However, vital signs obtained by a wearable device are mainly used for visualization or summarization. If these personal health data are subsequently analyzed, we can detect health abnormality earlier.

When a person becomes sick, there may be underlying latent factors inducing a disease. These factors influence

The associate editor coordinating the review of this manuscript and approving it for publication was Kezhi Li¹.

and exacerbate one's health. For example, when a person's immunity decreases, the person catches a cold. The influence of these factors cannot be directly observed, and therefore it is difficult to quantify this influence. By identifying such symptoms, diseases can be detected early.

In this study, factors related to or inducing a disease are defined in four dimensions: direct or indirect, and explicit or implicit. Factors other than the direct and explicit are regarded as latent factors. We further quantify disease risk by introducing a risk degree, based on the latent factor scores to effectively detect health abnormality.

Many studies on detecting abnormality have been carried out. For example, abnormal electroencephalogram (EEG) signals are one of the main themes of abnormality

detection [1], [2]. However, these studies used special devices because it was difficult to record EEG signals in daily life. Moreover, these studies focused on when the disease occurred and did not consider other factors, although even small changes in vital signs are important. Thus, more detailed abnormality detection considering latent factors is needed.

In studies that detect or predict disease from health data, deep learning is mainly used because of its high accuracy. However, the reliability of this method is limited because it is a black-box model and therefore cannot visualize how the results were obtained [3]. Thus, there is a need for a white-box method that can visualize how the results are obtained.

In this study, we propose a new approach to detect health abnormality by considering latent factors inducing a disease. The characteristics and contributions of this study are as follows.

- We defined the latent factors and proposed a new approach considering the latent factors that induce a disease for health abnormality detection.
- The latent factor scores were estimated based on the relationship between the factor and explicit variables. Thereafter, based on the scores, risk degrees were quantified, and abnormality was detected according to the risk degree.
- Our proposed approach is comparable to three baselines in terms of accuracy, recall, and false positive rate (FPR) of abnormality detection and can improve the reliability of the result.

The rest of this paper is as follows. Section II describes related research and the position of this study. Section III gives an overview of our approach and defines latent factors. In addition, we describe a method for estimating latent factor scores and risk degrees. Section IV introduces an experiment on abnormality detection and discusses the results. Section V summarizes this study and describes our future work.

II. RELATED WORK

A. METHODS TO DETECT ABNORMALITIES

Approaches to detect abnormalities are mainly divided into three categories based on distance, machine learning, and prediction [4].

The first approach uses a distance. In this approach, the distance between the record and a certain value is calculated. A criterion is set as a threshold, and abnormal data are detected when the distance is over the threshold. PCA (principal component analysis) and Hotelling's control chart are the conventionally used methods for this approach [5], [6]. Hotelling's control chart is the classic method of calculating the degree of abnormality from the mean and variance of a single variable, and it is used when there are many normal records. Ma and Lee [5] proposed a monitoring system for long-term patients, and they adopted Hotelling's control chart in their system. Amor *et al.* [6] proposed a method of conducting multivariate analysis using PCA. In their method, at first, anomalies are calculated from multivariate analysis. If abnormalities are over a threshold,

abnormal vital signs are investigated at the same time. Balasubramaniyam *et al.* [7] proposed using T-shirts as wearable devices to detect abnormalities in respiratory rate.

The second approach is machine learning. In this approach, the k-nearest neighborhood (KNN) method is often used. Asniar and Surendro [8] utilized the KNN method to detect students who have a greater risk of dropping out. Yan *et al.* [9] adopted this method for large sets of health data. They improved KNN so that distributed processing could be performed. Using their method enables high-dimensional data to be analyzed.

The third approach is predicting. Many studies use stochastic models, such as a Bayesian network. Kawanishi *et al.* [10] regarded features of dementia as anomalies. They detected the abnormal data from the patients' motions of drawing pictures. They adopted a method based on variational Bayesian methods and conducted abnormality detection by unsupervised learning. Hela *et al.* [11] proposed detecting abnormalities in daily environmental factors and used a Markov chain network to predict abnormalities.

In addition to these approaches, the method of deep learning has recently been used. Caroprese *et al.* [12] investigated tools for electronic health data processing with deep learning. Hsiao *et al.* [13] used deep learning to analyze the risk of cardiovascular disease and the relationship between living environment and disease risk. Li and Li [14] used deep learning and collaborative filtering to predict readmission for diabetic patients. However, deep learning cannot explain the process of how the results were obtained. Thus, many studies have been conducted to improve the interpretability of the results [3], [15].

B. ANALYSIS FOR LATENT FACTOR

Research using the concept of latent factors reveal factors that can be directly estimated from the obtained data. This concept has been widely adopted in many fields, such as behavioral science, psychology, marketing, and education.

Chen *et al.* [16] suggested that by examining common dependencies between multiple explicit variables, latent factors could be estimated. For example, in the education field, the explicit variable is the response to a test item, and the latent factors estimated can be cognitive skills. Latent factors are estimated mainly using Structural Equation Modeling (SEM) and factor analysis. In studies using SEM, researchers assume the latent factors, create a model, and verify the adaptability of the model to the dataset. Using factor analysis, Gagnon *et al.* [17] investigated whether factors such as organizational characteristics could affect recruitment when introducing electronic health data. Using SEM, May *et al.* [18] analyzed the factors that could influence people's vehicle automation acceptance and demonstrated the relationships among the latent factors.

C. POSITION OF THIS STUDY

Studies in recent times are based on deep learning methods mainly because they show better accuracy than conventional

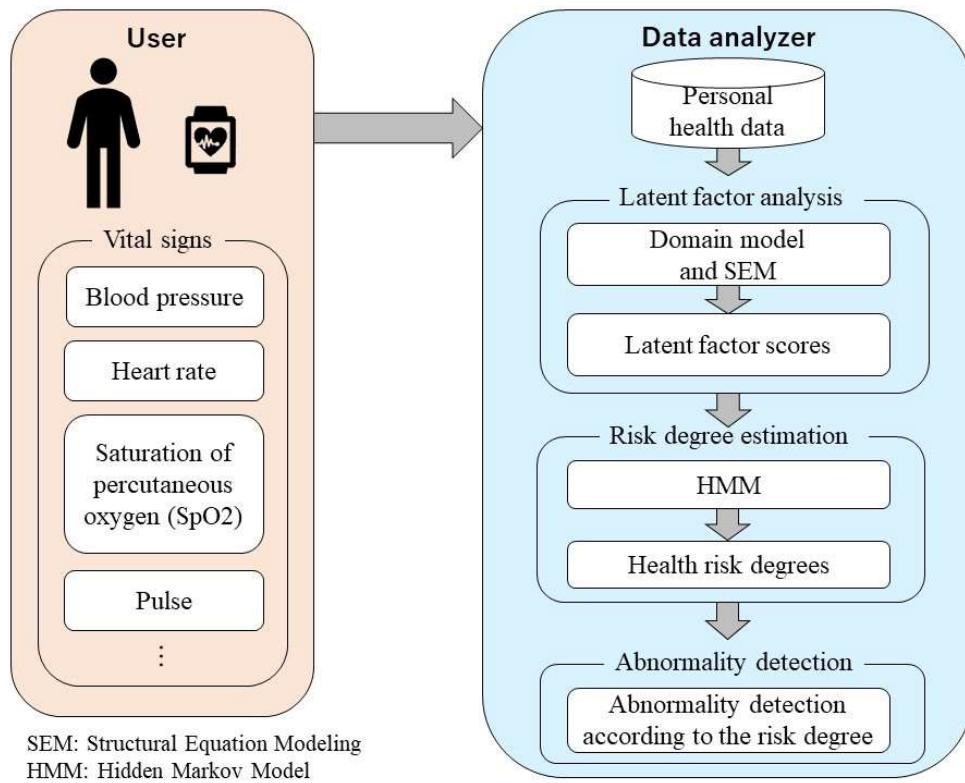


FIGURE 1. Abnormality detection considering latent factors.

methods. However, since deep learning has the characteristics of a black-box model, it is difficult to explain the process by which the results are obtained. Some studies have tried to create a new model that can visualize the process of how the result was obtained as explainable artificial intelligence [3], [19], but many issues persist. Therefore, there is a need for a white-box approach that has almost the same accuracy as deep learning.

Studies with anomaly detection often categorize participants by age or sex [7], but factors inducing a disease have not been considered. Many studies have focused on detecting abnormalities when a person is healthy but not on the onset of sickness or the serious stage of the disease. Ismail *et al.* [20] analyzed factors associated with specific diseases and presented a model to reveal positive and negative correlations with factors and health conditions. In their approach, health condition refers to attributes such as vital signs and BMI, and factor refers to behaviors such as fluid intake. However, the health condition was not quantified and used for abnormality detection. To solve these problems, we proposed a new approach to detect health abnormalities according to risk degrees.

In our previous studies, we proposed approaches to estimate latent factors [21] and applied them for abnormality detection [22]. However, risk degree estimation was not performed in these studies. In this study, we extend the previous studies by quantifying risk degree using a combination of SEM and Hidden Markov Model (HMM).

III. ABNORMALITY DETECTION CONSIDERING LATENT FACTORS

An outline of our approach is shown in Figure 1. In this approach, abnormality detection is performed using the following three steps.

The first step is data collection. In this step, personal health data are collected from a user over a certain period. The data are collected via wearable devices or sensors and stored in the data analyzer.

The second step is the analysis of latent factors. In data analysis, the relationship between vital signs and latent factors is analyzed, and the factor score is calculated using SEM. A domain model is created expressing the relationship between variables and latent factors. The factor scores are estimated for each record, which is considered the disease risk.

The third step is risk degree estimation and abnormality detection. In this step, HMM is used to estimate the risk degree, and health abnormalities are detected according to the risk degree.

A. THE DEFINITION OF LATENT FACTOR

Many studies use the term “factor,” but with inconsistent meaning. We define a factor based on four categories: direct, indirect, explicit, and implicit, as shown in Table 1. In this table, cells with the gray background, (2), (3), and (4), indicate latent factors. Each category is defined as follows.

TABLE 1. Types of factors.

	Direct	Indirect
Explicit	(1) Direct, Explicit	(2) Indirect, Explicit
Implicit	(3) Direct, Implicit	(4) Indirect, Implicit

- Direct factor is one that produces the result without any interim connection.
- Indirect factor is one that arrives at the result through two or more connections.
- Explicit factor is one that leads to the identification of the result that is obvious from the data.
- Implicit factor is one that can lead to the result only by applying meta-knowledge related to tacit knowledge.

As per this definition, the result includes phenomena, facts, and findings obtained by analyzing the data. By applying the definition mentioned, factors inducing a disease can be identified as follows, corresponding to the cell numbers in Table 1.

(1) Falling sick can be attributed to one direct and explicit factor. For example, a person ate food that was past its expiry date and thus fell sick.

(2) Falling sick can be attributed to two or more chain-reaction indirect and explicit factors. For example, a person not vaccinated against the flu contracted the infection by coming into contact with infected persons.

(3) Based on meta-knowledge and the combination of observed data, a factor can predict whether a person will contract a disease before the person actually falls sick. For example, when a smoker tends to be diagnosed with lung cancer, the factor inducing cancer can be estimated from the frequency of smoking and the period of smoking before the person is diagnosed with lung cancer.

(4) Based on meta-knowledge and combining it with the observed data, two or more chain-reaction indirect factors inducing a disease can be estimated before a person actually falls sick. For example, lifestyle-related illnesses can be predicted in a busy person from the person's working hours and duration of sleep.

Many studies analyzed the observable factors that induce a disease, such as toxic substances and viruses. However, there are factors that are related implicitly or indirectly to a disease. As defined, such factors are only revealed by analyzing the relationship between the factors or by using meta-knowledge. These factors are defined as Indirect and Explicit (2), Direct and Implicit (3), and Indirect and Implicit (4). In this study, we consider these as latent factors inducing a disease.

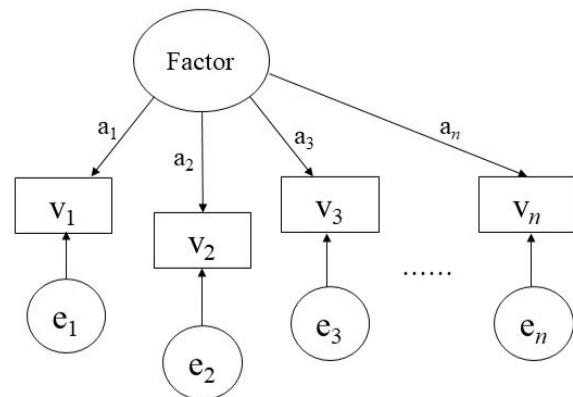
B. CALCULATING LATENT FACTOR SCORES BASED ON EXPLICIT VARIABLES

When analyzing latent factors such as factors inducing a disease, we adopt SEM, which is a statistical method. Using SEM, latent factor scores can be calculated from the relationship between the factors and the explicit variables.

To use SEM, we need a domain model that expresses the relationship between explicit variables and latent factors. In

our previous study, we defined an activity factor as a latent factor and verified the effectiveness of the domain model on personal health data [21]. In that study, motivation to perform activities was considered a latent factor, and we analyzed how behaviors such as step count were changed by motivation.

SEM requires a domain model as shown in Figure 2. In the model, a combination of explicit variables influenced by a latent factor is assumed. Based on the assumed relationship, the factor score is calculated. For example, when heart rate and respiratory rate are high, the risk of illness may increase, and the risk degree becomes high. Thus, factor scores can be considered a measure of disease risk. The scores can be calculated based on their correlations with explicit variables.

**FIGURE 2.** Example of a domain model.

SEM includes factor analysis and path analysis. By using SEM, latent factor scores can be calculated based on the relationship between the factor and explicit variables. The correlation can be calculated from the covariance. When two datasets X, Y as $(X, Y) = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ are obtained, the covariance of X and Y can be expressed as in Eq. (1) [23].

$$\text{Cov}(X, Y) = E[x_i y_i] = E[(X - \bar{X})(Y - \bar{Y})] \quad (1)$$

where $E(z)$ is the expected value of z , and \bar{X}, \bar{Y} is the average of X and Y .

For example, in Figure 2, when it is assumed that a latent factor affects three explicit variables (v_1, v_2, v_3), the relationships between v_1, v_2, v_3 and the latent factor scores $F = \{f_1, f_2, \dots, f_n\}$ can be expressed as follows [23].

$$\begin{aligned} v_1 &= a_1 f_i + e_1 \\ v_2 &= a_2 f_i + e_2 \\ v_3 &= a_3 f_i + e_3 \end{aligned}$$

where f_i represents the latent factor score. By definition [24], the variance of f_i is assumed to be 1, and the average of f_i is assumed to be 0. a_i is the regression coefficient, which indicates the degree of correlation. e_i is an independent constant value, which is regarded as an error. As a prerequisite of SEM, e_i is not affected by other variables. In addition, e_i has no

covariance with other variables. The variance of e_i is assumed to be 1, and the average is assumed to be 0.

When one considers two variables (v_1, v_2), the covariance of v_1 and v_2 can be calculated as follows [24].

$$\begin{aligned} \text{Cov}(v_1, v_2) &= E[(a_1 f_i + e_1)(a_2 f_i + e_2)] = a_1 a_2 E[f_i^2] \\ &\quad + a_1 E[f_i e_2] + a_2 E[f_i e_1] + E[e_1 e_2] = a_1 a_2 \end{aligned} \quad (2)$$

where $E[f_i^2]$ is equal to 1, since $E[f_i^2]$ can be expressed as $E[(f_i - E[f_i])^2]$ according to Eq. (1), which is the same equation with the variance of f_i [24]. $E[f_i e_1]$, $E[f_i e_2]$, and $E[e_1 e_2]$ are equal to 0, since the covariance of f_i and e_1 , f_i and e_2 , and e_1 and e_2 are equal to 0, and they are not correlated.

$\text{Cov}(v_2, v_3)$ and $\text{Cov}(v_3, v_1)$ can be calculated the same way with $\text{Cov}(v_1, v_2)$. Therefore, one can obtain values of $a_1 a_2$, $a_2 a_3$, and $a_3 a_1$. v_1 , v_2 and v_3 are explicit variables (e.g., step counts, heart rate), therefore, values of $\text{Cov}(v_1, v_2)$, $\text{Cov}(v_2, v_3)$, and $\text{Cov}(v_3, v_1)$ are calculated by Eq. (1). Consequently, simultaneous equations are obtained as follows [24].

$$\begin{aligned} a_1 a_2 &= C_1 \\ a_2 a_3 &= C_2 \\ a_3 a_1 &= C_3 \end{aligned}$$

where C_1 , C_2 , and C_3 represent values of $\text{Cov}(v_1, v_2)$, $\text{Cov}(v_2, v_3)$, and $\text{Cov}(v_3, v_1)$, respectively. By solving the equations, one can obtain values of a_1 , a_2 , and a_3 . Using these values, one can further perform factor analysis.

In the factor analysis, multiple regression analysis is conducted to estimate the value of f_i [25]. As described in [25], F can be estimated by Eq. (3) [25].

$$\hat{F} = X R^{-1} A f \quad (3)$$

where X is a mean deviation matrix of explanatory variables and $A f$ is the factor loading matrix. For the example mentioned, $A f$ can be expressed as follows [25].

$$A f = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad (4)$$

R represents a correlation matrix between explanatory variables. The correlation matrix for v_1 , v_2 , and v_3 can be expressed as follows [25].

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix} \quad (5)$$

where r_{12} , r_{23} , and r_{13} represent correlation coefficients of (v_1, v_2) , (v_2, v_3) , and (v_3, v_1) , respectively. X is the mean deviation matrix of the explicit variables, and one can estimate F as factor scores. This method is known as the regression estimation. If the latent factor scores F are high, explicit variables such as heart rate will be affected and will change. We regard the calculated latent factor scores as the disease risk and use them to estimate risk degrees.

C. ESTIMATING RISK DEGREES AND DETECTING ABNORMALITIES

The factor scores estimated by SEM do not consider the effects of previous records. In the real world, risk degrees change continuously over time and are usually affected by the previous degree. Therefore, the estimated factor scores should be taken as time-series data, and risk degree estimation should be performed considering the influence of the previous record.

To resolve this problem, we adopt HMM, which is a stochastic model, for risk degree estimation. In [26], to estimate the hidden parameters for people having different attributes and from varied populations, SEM was combined with HMM. In the study, time series data, which have many random effects, can be modeled. We propose a combined approach using both HMM and SEM, in which HMM was used to estimate not only hidden states but also the probability of state transitions. In addition, it is possible to calculate the probability of a transition from the previous state. By applying this method, it is possible to obtain the probability of how the risk degree changes. In HMM, it is assumed that the data is obtained based on the current state, and the next state is stochastically determined by the current state. This connection is called a “Markov chain,” and it is possible to estimate the next state based on the current state as time-series data.

In our approach, the estimated factor scores are used as input for HMM. Figure 3 shows the relationship between SEM and HMM. In this figure, f_n represents the factor score, and the state S represents the estimated risk degree. The estimated factor scores are used as time-series data, and the current state can be estimated considering the influence of the previous record.

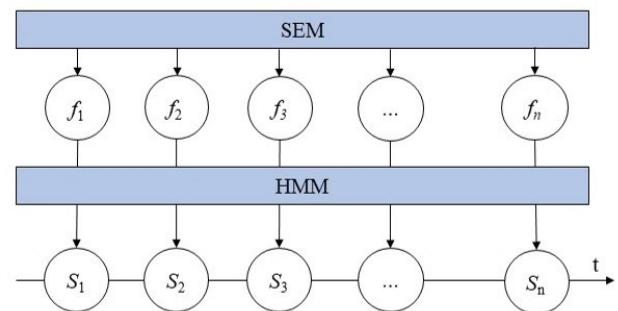


FIGURE 3. Relationship of SEM and HMM.

The estimated states are regarded as disease risk degrees in this study. After estimating the risk degrees using HMM, abnormalities are detected according to the degree. To detect an abnormality, in this study, we adopted a detection method proposed by Ide et al. [27]. In their method, the weights of variables and the parameters of Gaussian Markov Random Fields Mixture are estimated using variational Bayes. After estimating these parameters, the abnormality score is calculated based on its negative conditional log likelihood [27].

If the abnormality score is higher than a threshold, it is regarded as abnormal. The abnormality score for each variable could be calculated, therefore, it is possible to identify which vital sign shows abnormality.

IV. EXPERIMENT AND DISCUSSION

A. OVERVIEW OF DATASET

In our experiments, open data recorded by monitoring patients and released as the MIMIC database¹ were used. In the MIMIC database, there are multiple datasets. Amor *et al.* [6] used the dataset of No. 221. To compare our result with the previous study, we used the same dataset.

Dataset No. 221 contains the vital signs of a female patient with a brain injury for approximately 24 h. The recorded vital signs are as follows: average blood pressure (ABP), systolic blood pressure (SBP), diastolic blood pressure (DBP), saturation of percutaneous oxygen (SpO₂), respiratory rate, heart rate (HR), and pulse rate. The sampling rate was approximately one second. The number of records was 83,363, and Amor *et al.* [6] used the first 20,000 records. Therefore, the first 20,000 records were also used in this study. In the records, there were some missing values, which were removed, and finally 19,951 records were obtained. We divided the data into training data and test data. Of the 19,951, we treated the first 951 records as normal health statuses and used them as training data. The remaining 19,000 records were used as test data. The test data was also used for calculating latent factor scores and estimating risk degrees.

B. ESTIMATING RISK DEGREES BY SEM AND HMM

Using the test data, a domain model was created to apply SEM. The SEM implementation used the lavaan library of the statistical processing software R.

In the case of brain injury, blood pressure and heart rate are known to change when the disease state changes [28]. In addition, brain injury guidelines indicate that blood pressure and heart rate need to be monitored [29]. In medical environments, SpO₂ is also monitored [30]. Based on medical knowledge, a domain model was created assuming that latent factors affect ABP, HR, and SpO₂. Figure 4 shows the quantified latent factor scores, which are averaged every minute and plotted as a line graph. When the scores change, the vital signs are affected, and the patient's risk degree will change. Figure 5 shows the weights for explicit variables. It is considered that there is a positive correlation between the latent factor scores and blood pressure. Additionally, HR has a weak negative correlation with the latent factor scores. The latent factor scores shown in Figure 4 can be regarded as the disease risk. When the risk is far from zero, it means that the vital sign changed greatly. According to Figure 5, it can be seen that blood pressure increases when the risk increases. It also shows that heart rate is higher with low risk, but not as variable as blood pressure. In this patient's case, the risk may mainly relate to blood pressure.

¹<https://physionet.org/physiobank/database/mimicdb/>

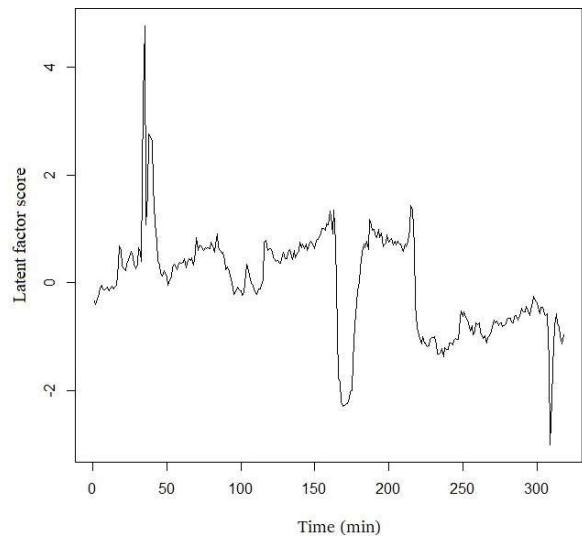


FIGURE 4. Changes in the estimated factor score.

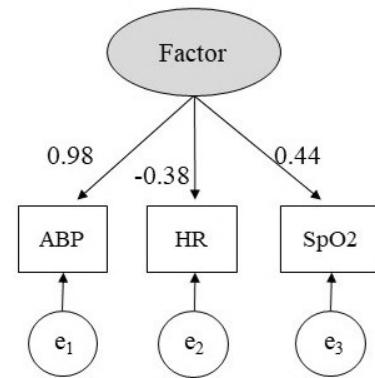


FIGURE 5. Weights from the latent factor to each vital sign.

The validity of the domain model can be verified by examining how well the data fit into the model. Several indicators of goodness of fit have been proposed, and we used common indicators (p-value [31], goodness of fit index (GFI) [32], cumulative fit index (CFI) [33], standardized root mean square residual (SRMR) [34], and root mean square error of approximation (RMSEA) [35]). The calculated indicators are shown in Table 2. The results revealed that the p-value and RMSEA did not meet the criteria. As described in [36], it is known that the p-value tends to decrease as the number of data points increases. In [25], it is mentioned that most models do not fit if the data have over 500 records, but we used 19,000 records in this experiment. Therefore, we judged that our domain model is acceptable though the p-value is smaller than 0.05. RMSEA was 0.11, which did not meet the criteria. However, the goodness to fit was close to the acceptable threshold, and several other indicators met the goodness of fit. Based on these results, the model was adopted in this study.

After estimating latent factor scores on the severity of brain injury by SEM, risk degrees were estimated using HMM.

TABLE 2. Adaptability of our model.

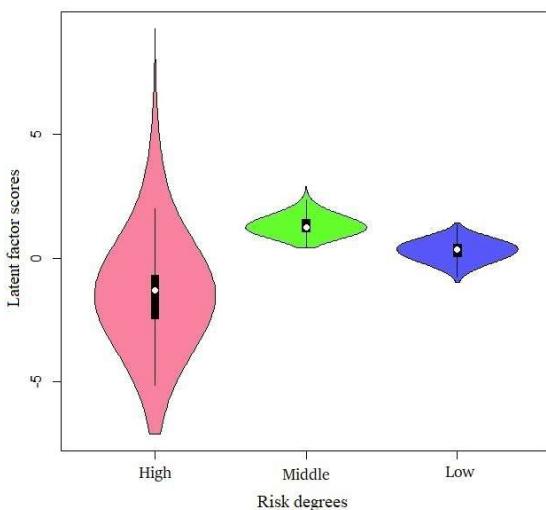
Indicators	Our model	Acceptance threshold
p-value	0.00	less than 0.05
GFI	0.99	more than 0.9
CFI	0.96	more than 0.9
SRMR	0.03	less than 0.1
RMSEA	0.11	less than 0.1

As HMM can perform state (risk degree in this study) estimation without training data, the model was created using only test data. We assumed three risk degrees for the states, i.e., high, middle, and low, and the states (risk degrees) change stochastically. To use HMM, we adopted R and RHMM packages.

Abnormality detection was performed for each risk degree estimated by HMM. The transition probabilities for each risk degree are shown in Table 3. In this table, the probabilities of changing from high to low and from low to high were smaller than the change from high or low to middle. From the middle risk degree, it changes to high and low with almost the same probability. We consider this result rational because the transition of risk degree from low to high and vice versa would be rare than from middle. Figure 6 shows the distribution of latent factor scores for each estimated risk degree. In this figure, the left graph is considered to be an unstable and high-risk degree because the variance is large. As the middle graph has a higher average value than the right graph, the middle one was regarded as the middle-risk degree and the right one as the low-risk degree. For the implementation of the abnormality detection, R and sGMRFmix packages were used.

TABLE 3. Transition probabilities of each risk degree.

	High	Middle	Low
High	0.998	0.001	0.0003
Middle	0.001	0.997	0.002
Low	0.001	0.010	0.987

**FIGURE 6.** Distributions of latent factor scores for each risk degree.

C. PREPARING DATA AND DETECTING ABNORMALITIES

The purpose of this study was to detect health abnormality, i.e., to judge whether the health status of an individual is normal or abnormal, which is a problem of binary classification and has been widely investigated in related studies, such as [1], [2], [6]. Furthermore, the disease risk degree was introduced in this study, and was estimated using the latent factor scores as the input to HMM.

To evaluate the approach proposed in this study and to compare it with the method presented in [6], we used the same data set they used, and inserted abnormalities in the same way as they did. Their study injected random synthetic anomalies at different time instants, which has not been elaborated. In this study, we inserted abnormalities to the test data clustered by the risk degrees, as shown in Table 4. We inserted four sets, each 100 records, to each clustered data as abnormalities to verify the detection performance.

To detect abnormalities, we calculated the abnormality score of each vital sign using the same method presented in [27]. After calculating the abnormality score, the threshold was set, and abnormalities were detected. At first, we set a threshold for all vital signs. Next, we adjusted the threshold for each vital sign to raise the F-measure value [37] higher. Finally, we compared the F-measures and selected the vital sign threshold that had the highest F-measure value.

D. EVALUATION METRICS

The performance of abnormality detection was evaluated in terms of accuracy, recall, and FPR, which are representative indicators [37].

Accuracy is defined in Eq. (6).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

where TP stands for “True Positive”, TN for “True Negative”, FP for “False Positive”, and FN for “False Negative”. In this study, “True” means correct detection (abnormality), and “False” means wrong detection (non-abnormality). “Positive” stands for abnormal, and “Negative” for normal. Therefore, TP indicates the number of correctly detected abnormalities, and TN represents the number of correctly classified non-abnormalities. FP is the number of wrongly judged abnormalities (actually, normal), and FN is the number of wrongly classified non-abnormalities (i.e., abnormalities that could not be detected). Accuracy is a measure on how many abnormalities are correctly detected in this study.

Recall is defined in Eq. (7).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

Recall is an indicator on how many true abnormalities are detected.

FPR is defined in Eq. (8).

$$\text{FPR} = \frac{FP}{FP + TN} \quad (8)$$

TABLE 4. Inserted abnormal data.

Abnormality number	Record number	Changed vital signs	Inserted abnormal data
No. 1	1000-1099	HR	Change each vital sign to minimum value minus twenty in the degree
No. 2	2000-2099	ABP, SBP, DBP	Change each vital sign to maximum value in the degree
No. 3	3000-3099	SpO2	Change each vital sign to minimum value minus ten in the degree
No. 4	4000-4099	SBP, DBP, HR, PULSE	Change each vital sign to maximum value in the degree

TABLE 5. Summary of the approaches.

Approach	Risk degree estimation	Abnormality detection method
PCA-based	-	Based on the Squared Prediction Error [6]
Deep learning	-	Auto-encoder [38]
No-degree estimation	-	sGMRFmix algorithm [27]
This study	✓	sGMRFmix algorithm [27]

FPR is a measure on how many non-abnormalities are wrongly classified to be abnormal.

Using the test data, which are used to estimated risk degrees based on SEM and HMM, an evaluation of abnormality detection was performed. The test data was divided according to the risk degree and abnormal health data was inserted for each degree. We verified whether the inserted abnormalities in the test data could be detected.

To verify the effectiveness of our approach, we set three approaches as the baseline: PCA-based approach, deep learning, and no-degree estimation. Table 5 shows a summary of the approaches. In our approach, risk degrees were estimated though other studies do not conduct risk degree estimation. The abnormality detection method for PCA-based approach is based on the Squared Prediction Error [6]. Deep learning approach uses auto-encoder [38] for abnormality detection. The abnormality detection for no-degree estimation and our approach uses the algorithm based on the sGMRFmix algorithm [27]. Since only recall and FPR were used for the evaluation in [6], to compare all of their results, we adopted the same indicators, i.e., recall and FPR. The details of the baseline approaches are described as follows.

(1) PCA-based approach [6]: In this approach, the previous study performed abnormality detection based on the same dataset as ours. They inserted abnormal health data and used PCA to calculate abnormality. The detection method was different from ours.

(2) Deep learning: In this approach, all records were regarded as one degree, and 400 abnormal records were inserted in the test data as shown in Table 4. For the deep learning model, an auto-encoder was used. With this model, it is possible to calculate the degree of an abnormality even if the training data does not include abnormalities. The model was trained using the training data, and the abnormality score for the test data was calculated. The hyper-parameters of the auto-encoder are as follows. Binary cross-entropy is a loss function, and number of epochs is 100. We set the threshold for abnormality when the F-measure was the highest.

(3) No-degree estimation: In this approach, all records were regarded as one degree, and 400 abnormal records were inserted in the test data as shown in Table 4. The method

used for abnormality detection was the same one as our approach, which was presented in [27]. We set the threshold for abnormality when F-measure becomes the highest.

In addition, we performed a five-fold cross-validation for no-degree estimation and our approach. In this validation, we adopted accuracy as an indicator in addition to recall and FPR. The threshold for abnormality detection was set to be the same as that of Table 6.

E. RESULTS

The comparison of results with three baseline methods is shown in Table 6, where the conditions for the adjusted thresholds of vital signs represent corresponding abnormality scores obtained by the method used in [27]. The abnormality score is defined as the degree of distance from criteria obtained by training with the data of vital signs at normal times. When the score exceeds a threshold, it is judged as abnormal. NA (Not Available) means that the abnormality score could not be obtained. From Table 6, we can see that our approach showed slightly lower recall (98.75%) than deep learning and no-degree estimation methods (both 100%), but slightly higher than PCA-based approach (98.02%). On the other hand, for FPR, our method was 0.186%, much lower than three baselines, 0.7%, 4.5%, and 1.25% respectively. By estimating the risk degrees, the detection could be conducted with high performance. In addition, since the process on how the results were obtained was visible and clear, the reliability of the results was improved.

The result of five-fold cross-validation comparing with no-degree estimation is shown in Table 7. The average accuracy and recall of our method were 99.7% and 98.3% respectively, and FPR was 0.045%, all much better than the baseline.

F. DISCUSSION

In our approach, based on the relationship between vital signs and latent factors, risk degrees were estimated, and different thresholds were set. By setting different thresholds for each estimated risk degree, the performance of abnormality detection was improved. In our experiment for our approach, as shown in Table 6, in abnormality No. 2, as the risk degree

TABLE 6. Results of abnormality detection.

Approach	Abnormality number	Recall (%)	FPR (%)	Adjusted threshold for detecting abnormalities
PCA-based [6]	Average	98.02	0.7	Algorithm based on Jackson and Mudholkar approximation
Deep learning	No. 1	100	18.0	Abnormality > 0.3
	No. 2	100	0.001	Abnormality > 0.8
	No. 3	100	0	Abnormality > 0.3
	No. 4	100	0	Abnormality > 1.1
	Average	100	4.5	-
No-degree estimation	No. 1	100	0.8	HR: NA
	No. 2	100	0.3	DBP > 200
	No. 3	100	0	SpO2: NA
	No. 4	100	0.3	SBP > 20
	Average	100	1.25	-
This study (Case of high risk)	No. 1	100	0	HR: NA
	No. 2	100	0.04	DBP > 100
	No. 3	100	0	SpO2: NA
	No. 4	100	0	SBP: NA
	Average	100	0.01	-
This study (Case of middle risk)	No. 1	100	0	HR: NA
	No. 2	90	0.1	DBP > 20
	No. 3	100	0	SpO2: NA
	No. 4	100	0.1	DBP > 20
	Average	97.5	0.05	-
This study (Case of low risk)	No. 1	100	0	HR > 100
	No. 2	95	0.1	DBP > 20
	No. 3	100	0	SpO2: NA
	No. 4	100	0.1	DBP > 20
	Average	98.75	0.05	-
Average of this study	Average	98.75	0.186	-

TABLE 7. Results of five-fold cross validation.

Approach	Abnormality number	Accuracy (%)	Recall (%)	FPR (%)	Adjusted threshold for detecting abnormalities
No-degree estimation	No. 1	98.3	96.5	0	HR: NA
	No. 2	75.2	65.0	0.3	DBP > 200
	No. 3	100	100	0	SpO2: NA
	No. 4	68.7	61.0	0.3	SBP > 20
	Average	85.6	80.6	0.1	-
This study (Case of high risk)	No. 1	100	100	0	HR: NA
	No. 2	99.5	96.0	0.4	DBP > 100
	No. 3	100	100	0	SpO2: NA
	No. 4	100	100	0	SBP: NA
	Average	99.9	99.0	0.1	-
This study (Case of middle risk)	No. 1	100	100	0	HR: NA
	No. 2	99.7	99.0	0.02	DBP > 20
	No. 3	100	100	0	SpO2: NA
	No. 4	100	100	0	DBP > 20
	Average	99.9	99.7	0.005	-
This study (Case of low risk)	No. 1	100	100	0	HR > 100
	No. 2	98.0	89.3	0.1	DBP > 20
	No. 3	100	100	0	SpO2: NA
	No. 4	99.1	95.0	0	DBP > 20
	Average	99.3	96.1	0.03	-
Average of this study	Average	99.7	98.3	0.045	-

increased, the threshold for DBP became larger. By setting a larger threshold for high risk degree, smaller FPR could be obtained. In addition, the performance was also improved in the five-fold cross verification. By considering latent factors, abnormalities can be detected more effectively.

The indicators to be measured and monitored have been widely investigated, which are generally listed and detailed in medical guidelines [29]. To detect abnormalities, the studies [1], [2], [6] on the indicators to be measured directly judged whether the indicators such as blood pressure and EEG were abnormal or not. By considering the latent factors inducing

a disease and quantifying the degree of disease risk, our findings showed that sensitive abnormality detection could be performed.

The major advantages of our method were that it could detect abnormalities in a visible way, and it is independent to a data set. This study focused on abnormality detection based on personal health data, and the data set from one subject was used to evaluate our proposed approach and compare with a baseline [6], which used the same data set. The result showed that our approach was effective to detect health abnormality. By estimating risk degrees for the quantified disease risk, it is

possible to set an abnormality detection threshold according to the degree. In addition, real-time abnormality detection can be implemented by constructing a domain model and executing the data analyzer on a server or cloud. By setting a threshold for abnormality according to risk degrees, early detection for a disease can be expected. For example, in a healthy state, a strict check of vital signs can help detect abnormalities as early as possible. By paying attention to these abnormalities, early detection and even prevention before the onset of the disease are possible.

Limitations of our approach and improvements that could be made are summarized as follows. Our approach was not conceived to detect multiple diseases at the same time, since the domain model was designed to be applied for one disease only. In addition, the domain model, as the name suggests, is domain-dependent, and it was based on the investigation and medical insights on the relationship between a disease and its vital signs. Therefore, creating such a domain model would require investigations or expertise suitable for the disease. When creating the domain model, it is necessary to carefully consider how to determine the explicit variables (i.e., vital signs influenced by latent factors) that will be included in the model, in addition to its adaptability and rationality. We will improve and verify our approach to creating the domain model by referring to the method to build the health knowledge model proposed in [20]. Moreover, in the comparison experiment, for simplicity, the threshold was set for one vital sign per abnormality detection, although it is possible to set the threshold to detect one abnormality by a combination of multiple vital signs. And as to the threshold setting for abnormality score at high risk degree, for example, the threshold for abnormality score of DBP was set so as to decrease FPR. However, for prompt detection, it is necessary to set a stricter threshold as the risk degree is higher. We will further investigate and explore a more sensitive detection to solve this issue.

V. CONCLUSION

In this study, we proposed an abnormality detection approach, and clarified how the result was obtained. First, latent factors were estimated based on the combination of explicit variables in Structural Equation Modeling (SEM), and latent factor scores were calculated as the disease risk. Second, the risk degrees are estimated using the factor score as the input to Hidden Markov Model (HMM). Thereafter, the health abnormalities were detected according to risk degrees. In our experiment, the performance of abnormality detection was compared with three baselines: PCA-based approach, deep learning, and no-degree estimation methods. The experimental result showed that the average recall of our method was slightly higher than PCA-based approach, but lower than deep learning and no-degree estimation. However, FPR was much lower than all three baselines. In the five-fold cross-validation comparing with no-degree estimation, the average accuracy, recall, and FPR of our method were all much better than the baseline. Furthermore, how the result was obtained in our

approach was clarified in a visible way, and early detection and prevention of disease could be achieved.

We will consider how to create a domain model in a systematic way, and we will try to extend it for multiple diseases in future studies. In addition, we plan to set the threshold to detect one abnormality by a combination of multiple vital signs to further improve the detection accuracy.

REFERENCES

- [1] M. Hammad, A. Maher, K. Wang, F. Jiang, and M. Amrani, "Detection of abnormal heart conditions based on characteristics of ECG signals," *Measurement*, vol. 125, pp. 634–644, Sep. 2018.
- [2] M. Sharma, S. Patel, and U. R. Acharya, "Automated detection of abnormal EEG signals using localized wavelet filter banks," *Pattern Recognit. Lett.*, vol. 133, pp. 188–194, May 2020.
- [3] G. R. Vásquez-Morales, S. M. Martínez-Monterrubio, P. Moreno-Ger, and J. A. Recio-García, "Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning," *IEEE Access*, vol. 7, pp. 152900–152910, 2019.
- [4] H. Ren, Z. Ye, and Z. Li, "Anomaly detection based on a dynamic Markov model," *Inf. Sci.*, vol. 411, pp. 52–65, Oct. 2017.
- [5] F. T. Ma and C. E. Lee, "Integrated control chart for vital signs early warning of long-term care patients," in *Proc. 7th Int. Conf. Ubi-Media Comput. Workshops*, Ulaanbaatar, Mongolia, Jul. 2014, pp. 313–318.
- [6] L. B. Amor, I. Lahyani, and M. Jmaiel, "Data accuracy aware mobile healthcare applications," *Comput. Ind.*, vol. 97, pp. 54–66, May 2018.
- [7] H. Balasubramaniyam, M. S. Vignesh, E. S. Florence, A. Abhirami, and A. Abanah, "Design and development of a IoT based flexible and wearable T-shirt for monitoring breathing rate," in *Proc. 3rd Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Mar. 2019, pp. 376–379.
- [8] Asniar and K. Surendro, "Using data science for detecting outliers with k nearest neighbors graph," in *Proc. Int. Conf. ICT Smart Soc. (ICISS)*, Bandung, Indonesia, Sep. 2014, pp. 1–5.
- [9] K. Yan, X. You, X. Ji, G. Yin, and F. Yang, "A hybrid outlier detection method for health care big data," in *Proc. IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Social Comput. Netw. (SocialCom), Sustain. Comput. Commun. (SustainCom) (BDCloud-SocialCom-SustainCom)*, Atlanta, GA, USA, Oct. 2016, pp. 157–162.
- [10] K. Kawanishi, H. Kawanaka, H. Takase, and S. Tsuruoka, "A study on dementia detection method with stroke data using anomaly detection," in *Proc. 6th Int. Conf. Informat., Electron. Vis., 7th Int. Symp. Comput. Med. Health Technol. (ICIEV-ISC MHT)*, Sep. 2017, pp. 1–4.
- [11] S. Hela, B. Amel, and R. Badran, "Early anomaly detection in smart home: A causal association rule-based approach," *Artif. Intell. Med.*, vol. 91, pp. 57–71, Sep. 2018.
- [12] L. Caroprese, P. Veltri, E. Vocaturo, and E. Zumpano, "Deep learning techniques for electronic health record analysis," in *Proc. 9th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, Jul. 2018, pp. 1–4.
- [13] H. C. W. Hsiao, S. H. F. Chen, and J. J. P. Tsai, "Deep learning for risk analysis of specific cardiovascular diseases using environmental data and outpatient records," in *Proc. IEEE 16th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2016, pp. 369–372.
- [14] X. Li and J. Li, "Health risk prediction using big medical data—A collaborative filtering-enhanced deep learning approach," in *Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2018, pp. 1–7.
- [15] K. Amarasinghe, K. Kenney, and M. Manic, "Toward explainable deep neural network based anomaly detection," in *Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI)*, Jul. 2018, pp. 311–317.
- [16] Y. Chen, X. Li, and S. Zhang, "Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications," *J. Amer. Stat. Assoc.*, to be published, doi: [10.1080/01621459.2019.1635485](https://doi.org/10.1080/01621459.2019.1635485).
- [17] M.-P. Gagnon, D. Simonyan, E. K. Ghandour, G. Godin, M. Labrecque, M. Ouimet, and M. Rousseau, "Factors influencing electronic health record adoption by physicians: A multilevel analysis," *Int. J. Inf. Manage.*, vol. 36, no. 3, pp. 258–270, Jun. 2016.
- [18] K. R. May, B. E. Noah, and B. N. Walker, "Driving acceptance: Applying structural equation modeling to in-vehicle automation acceptance," in *Proc. 9th Int. Conf. Automot. User Interfaces Interact. Veh. Appl. Adjunct (AutomotiveUI)*, New York, NY, USA: Association for Computing Machinery, 2017, pp. 190–194.

- [19] B. M. Keneni, D. Kaur, A. Al Bataineh, V. K. Devabhaktuni, A. Y. Javaid, J. D. Zaintz, and R. P. Marinier, "Evolving rule-based explainable artificial intelligence for unmanned aerial vehicles," *IEEE Access*, vol. 7, pp. 17001–17016, 2019.
- [20] W. N. Ismail, M. M. Hassan, H. A. Alsalamah, and G. Fortino, "CNN-based health model for regular health factors analysis in Internet-of-medical things environment," *IEEE Access*, vol. 8, pp. 52541–52549, 2020.
- [21] K. Tago, K. Takagi, K. Ito, and Q. Jin, "Specifying latent factors with a domain model for personal data analysis," in *Proc. 3rd IEEE Cyber Sci. Technol. Congr.*, Athens, Greece, Aug. 2018, pp. 286–291.
- [22] K. Tago and Q. Jin, "Detection of anomaly health data by specifying latent factors with SEM and estimating hidden states with HMM," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Hangzhou, China, Oct. 2018, pp. 137–141.
- [23] R. O. Mueller, *Basic Principles of Structural Equation Modeling*. Piscataway, NJ, USA: Springer, 1996.
- [24] H. Toyoda, *Covariance Structure Analysis*. Tokyo, Japan: Asakura Shoten, 1998.
- [25] H. Asano, T. Suzuki, and T. Kojima, *Getting Started in Covariance Structure Analysis*. Tokyo, Japan: Kodansha, 2005.
- [26] T. Asparouhov, E. L. Hamaker, and B. Muthén, "Dynamic latent class analysis," *Struct. Equation Model., Multidisciplinary J.*, vol. 24, no. 2, pp. 257–269, 2017.
- [27] T. Ide, A. Khandelwal, and J. Kalagnanam, "Sparse Gaussian Markov random field mixtures for anomaly detection," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 955–960.
- [28] M. Kuroda, M. Kohno, K. Go, T. Sugino, and T. Ukai, "Hormonal study of blood pressure, heart rate, and survival in head-injured patients," *Jpn. Assoc. Acute Med.*, vol. 13, no. 7, pp. 361–368, 2002.
- [29] N. Carney, A. M. Totten, C. O'Reilly, J. S. Ullman, G. W. J. Hawryluk, M. J. Bell, S. L. Bratton, R. Chesnut, O. A. Harris, N. Kissoon, A. M. Rubiano, L. Shutter, R. C. Tasker, J. Vavilala, M. S. A. Wilberger, D. W. Wright, and J. Ghajar, "Guidelines for the management of severe traumatic brain injury, fourth edition," *Neurosurgery*, vol. 80, no. 1, pp. 6–15, 2016.
- [30] T. Miyagi, M. Maeda, Y. Inoue, D. Kondou, F. Yoshimura, A. Ookura, and M. Morioka, "Acute therapy for traumatic head injury," *Jpn. J. Rehabil. Med.*, vol. 50, no. 7, pp. 557–569, 2013.
- [31] J. H. Steiger, "Statistically based tests for the number of common factors," in *Proc. Annu. Meeting Psychometric Soc.*, Iowa City, IA, USA, 1980, pp. 424–453.
- [32] K. G. Jöreskog and D. Sörbom, "Recent developments in structural equation modeling," *J. Marketing Res.*, vol. 19, no. 4, pp. 404–416, Nov. 1982.
- [33] P. M. Bentler, "Comparative fit indexes in structural models," *Psychol. Bull.*, vol. 107, no. 2, p. 238, 1990.
- [34] P. M. Bentler and D. G. Bonett, "Significance tests and goodness of fit in the analysis of covariance structures," *Psychol. Bull.*, vol. 88, no. 3, pp. 588–606, 1980.
- [35] M. Browne and R. Cudeck, "Alternative ways of assessing model fit," in *Testing Structural Equation Models*. Thousand Oaks, CA, USA: SAGE Publications, 1993, pp. 137–162.
- [36] T. Hoshino, K. Okada, and T. Maeda, "Fit indices and model modification in structural equation modeling: A review and new findings," *Jpn. J. Behaviormetrics*, vol. 32, no. 2, pp. 209–235, 2005.
- [37] F. Salfner, M. Lenk, and M. Malek, "A survey of online failure prediction methods," *ACM Comput. Surv.*, vol. 42, no. 3, pp. 1–42, Mar. 2010.
- [38] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.



KIICHI TAGO (Member, IEEE) received the bachelor's degree in arts and the master's degree in human sciences from Waseda University, Japan, in 2013 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the Graduate School of Human Sciences. His research interests include big data, health data analysis, social networking service, natural language processing, cognitive psychology, and behavioral analysis.



KOSUKE TAKAGI received the bachelor's degree in human sciences from Waseda University, Japan, in 2018, where he is currently pursuing the master's degree with the Graduate School of Human Sciences. His research interests include social networking service, natural language processing, abnormality detection, and machine learning.



QUN JIN (Senior Member, IEEE) is currently a Professor with the Networked Information Systems Laboratory, Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Japan. He has been extensively engaged in research works in computer science, information systems, and social and human informatics. His recent research interests include human-centric ubiquitous computing, behavior and cognitive informatics, big data, personal analytics, individual modeling, intelligence computing, blockchain, cyber security, cyber-enabled applications in healthcare, and computing for well-being.

• • •