IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 2, 2019, accepted January 13, 2020, date of publication January 21, 2020, date of current version January 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968393

Context Deep Neural Network Model for Predicting Depression Risk Using Multiple Regression

JI-WON BAEK[1] AND KYUNGYONG CHUNG[2]

[1]Department of Computer Science, Kyonggi University, Suwon-si 16227, South Korea
[2]Division of Computer Science and Engineering, Kyonggi University, Suwon-si 16227, South Korea

Corresponding author: Kyungyong Chung (dragonhci@gmail.com)

This work was supported by Kyonggi University Research Grant 2019.

ABSTRACT Depression is a mental illness influenced by various factors, including stress in everyday life, physical activities, and physical diseases. It accompanies such symptoms as continuous depression, sleep disorder, and suicide attempts. In the healthcare, it is necessary to predict diverse situations accurately. Accordingly, in order to care for mental health, it is necessary to recognize individuals' situations and continue to manage them. In the area of mental diseases and treatment, research has been conducted to find a patient's state with the use of big data and to monitor the worst situation. Mental illnesses typically have depression. Research on Mental healthcare using artificial intelligence do conduct on prediction based on patients' voice, word choice, and conversation length. However, there is not much research on situation prediction in order to prevent depression. Therefore, this study proposes the context-DNN model for predicting depression risk using multiple-regression. The context of the proposed context-DNN consists of the information to predict situations and environments influencing depression in consideration of context information. Each context information related to predictor variables of depression becomes an input of DNN, and variable for depression prediction becomes an output of DNN. For DNN connection, the regression analysis to predict the risk of depression is used so as to predict the potential context influencing the risk of depression. According to the performance evaluation, the proposed model was evaluated to have the best performance in regression analysis and comparative analysis with DNN.

INDEX TERMS Deep neural network, context, depression risk, mental health, multiple regression, healthcare, deep learning, context information.

I. INTRODUCTION

Today, people experience severe physical disorders and psychological stress due to a variety of internal and external factors. People change their emotions depending on the situation of places like school, work, and meeting, and on the situation of weather and time. In particular, those who often feel a sense of depression, lack of enthusiasm, and dysthymia are vulnerable to depression. Although depression is mainly found in people in their 30s and 40s mainly experience depression, it is often detected in juveniles due to academic stress and interpersonal relationship and in elderly persons. As such, depression is a mental illness found in various age groups [1]. Since there is a socially negative view on persons who suffer a mental illness, such patients often hide their illness. In addition, such symptoms as a sense of depression, lack of enthusiasm, and dysthymia are found in most people so that they are easily ignored. According to the statistics about diseases of national interest in healthcare big data [43] provided by Health Insurance Review and Assessment Service [42], the number of depression patients and medical costs continue to increase. Patients visiting medical institutions to treat their depression are mostly in their 50s and 60s. It means that young adults fail to make prevention, care, and treatment properly in their busy life. In this circumstance, more demands for fast treatment and continuous care through early diagnosis are on the rise. If found early, depression has a high cure rate but is likely to recur. To overcome, care for,

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

and prevent depression, it is necessary to predict and prepare for the situation of the risk of depression. Therefore, need is the model that is capable of collecting people's diet, illnesses, stress, and other context information, analyzing correlations between depression and factors on the basis of the collected information, and predicting the context to prevent the disorder.

With the advancement of artificial intelligence (AI), developed have been models to predict results from multimodal data like numbers, images, and videos with the use of reinforcement learning or machine learning [2]. These models utilize a computer system to execute the tasks that need human intelligence and the ability of a machine that imitates human intelligent behavior [3]. Machine learning obtains new knowledge and information through learning with the structured data which are pre-processed from big data [4]. In terms of healthcare, needed is the technology of classifying a user's health state, context-aware, and predicting the potential health risk through time-series prediction. Accordingly, this study proposes the context-DNN model to predict the risk of depression using multiple regression. The proposed model predicts depression risk context in the combination with DNN and context information. It is used to predict the potential context influencing the risk of depression. The context information related to predictor variables of depression are used as inputs of DNN model. The output of DNN model consists of predictor variables of depression. The regression analysis for risk prediction is applied for the DNN connection. In this way, the risk of depression is predicted.

The rest of this paper is organized as follows. Section 2 provides a discovery for relation using regression analysis. In addition, classification and prediction using DNN. Section 3 proposes the context-DNN model for predicting the depression risk using multiple regression. Section 4 provides an evaluation. Finally, conclusions are given in Section 5.

II. RELATED WORK

A. DISCOVERY FOR RELATION USING REGRESSION ANALYSIS

Regression analysis, which is often used in statistics, mathematically estimates linear correlations between data with the use of regression [5]. As variables, independent variables (cause) and dependent variables (result) are used [6]. Regression analysis is distinguished into simple regression analysis and multiple regression analysis according to the number of independent variables. In simple regression analysis, there are one independent variable and one dependent variable, and the cause-and-effect relationship between the two variables is analyzed. In multiple regression analyses, there are one dependent variable and at least two independent variables. The analysis method is used to discovery how at least two independent variables influence one dependent variable [7], [8]. In addition, the regression analysis technique is used to analyze the cause-and-effect relationship between dependent and independent variables. The equation (1) presents multiple regression analysis. In the equation, Y means a dependent variable, and X means an independent variable. β_0 represents Y-intercept, and e means residual. $\beta_1, \beta_2, \cdots, \beta_k (k = 1, 2, \cdots, n)$ represent regression coefficients. Dependent and independent variables should have logical validity. The fundamental goal of regression analysis is to calculate β_0 and the regression coefficients $\beta_1, \beta_2, \cdots, \beta_k (k = 1, 2, \cdots, n)$. In regression analysis, an independent variable influencing a dependent variable is found, and the change of the dependent variable is predicted.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e \tag{1}$$

If a dependent variable changes significantly according to an independent variable on the basis of the results of logical validity and regression analysis in the equation (1), it is possible to estimate that two variables have a cause-and-effect relationship.

The coefficient to judge the relation between variables in a regression equation is the coefficient of determination (R^2) [9]. The coefficient makes it possible to judge goodness-of-fit of a regression equation and goodness-of-fit of an independent variable for a dependent variable. Therefore, the relation between dependent and independent variables is judged with the coefficient of determination in diverse analysis conditions.

M. Gong *et al.* [10] developed the regression analysis method of the relevant factors for personal information protection. In a framework, a differential personal regression analysis model is applied for prediction. A target function is converted to a multinomial expression. The coefficient of a multinomial equation is changed according to the size of the relation between an input variable feature and an output value of a model. Personal information leakage is applied to regression analysis on the basis of deep learning data in order for effective prediction. H. Jiang *et al.* [11] developed the multi-purpose evolutionary approach for fuzzy regression so as to address the issue of multi-purpose optimization. The factors that require subjective judgment are learned and predicted by a genetic algorithm.

B. CLASSIFICATION AND PREDICTION USING DNN

DNN (Deep Neural Network) has the structure of multiple hidden layers between the input layer and the output layer [12]. In this case, the number of nodes in the input layer is determined by the feature of sample data. In the hidden layer, it is possible to learn relations of a variety of nonlinear data. Also, complex nonlinear problems can be solved, and a variety of information can be extracted from high-dimensional data. In the hidden layer connected with the input layer, it is possible to combine with a value of input variable, give a weight, extract a new value, and deliver it to the output layer. Based on the calculated features in the hidden layer, the output layer makes it possible to perform classification and prediction with the use of FNN (Feed-Forward Network) and backpropagation. FFN [13], [14] is a neural network in which information goes forward in one direction, such as from the input layer to the hidden layer, or from the hidden layer to the output layer. In FFN, it is impossible to adjust weight for an error. In backpropagation [15], [16], to reduce a difference between an actual value and predictive value, it is possible to return to the previous node and adjust the weight for each node and its next node. Through the repetition of such a process, it is possible to extract the optimal result. As methods of adjusting a weight, there are Gradient Descent, Stochastic Gradient Descent, and Adam. Gradient descent [17] is a method of adjusting the weight to minimize an error and improve accuracy. Stochastic Gradient Descent [18] is a method of extracting data randomly and then improving speed and updating a weight. Adam [19], [20] is the most used method that makes it possible to improve an update speed of learning and accuracy. DNN model uses an activation function to determine whether input data is delivered to the next layer. ReLu solved the problem of gradient loss. It features very simple calculations and very fast learning. If an input value is negative, '0' is displayed; if positive, an input value is displayed as it is. DNN aims at classification and numerical prediction. SVM (Support Vector Machine) is a representative algorithm for classification. SVM [21] is a method of learning data and classifying new data. It measures the distance between a group and data so as to calculate its center, and then finds the optimal hyperplane and classifies data. Fig. 1 show the DNN based classification and prediction process. After the collected data is pre-processed, it is split into a training set and a test set. For the extraction of data features, data is entered in DNN, and then data features are extracted and displayed in the hidden layer. Data output is classified. To evaluate the model's goodness-of-fit, the model is compared with other models.

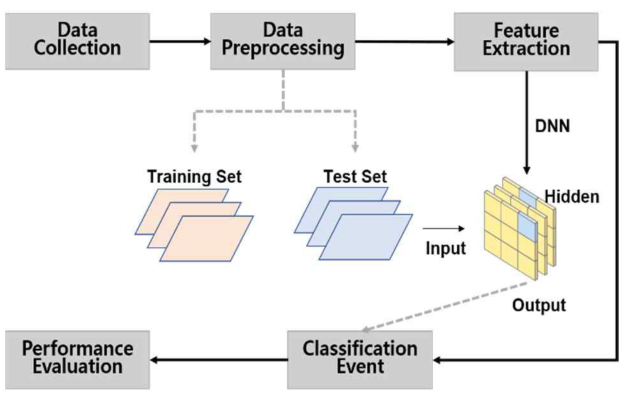

FIGURE 1. DNN based classification and prediction process.

Li *et al.* [22] developed the data hierarchical-classification method using DNN and weight support vector machine. In order to reduce the dimensions and optimal depth of multi-mode data, the method initializes the initial weight and offset of the hierarchical multiple neutral networks and extracts features of the multiple-layer neural networks. To classify high-dimensional data, it changes and monitors parameters with the use of Gradient Descent.

III. CONTEXT DEEP NEURAL NETWORK MODEL FOR PREDICTING DEPRESSION RISK USING MULTIPLE REGRESSION

The prediction model for preventing the risk of depression consists of three steps. The first step is data organization and pre-processing. For the data organization, the raw data offered by the Korea National Health and Nutrition Examination Survey is used. Based on the raw data [24] of Korea National Health and Nutrition Examination Survey [23], calculated is the statistical data for finding a level of the public's health and nutrition and evaluating if health policies and projects are effectively established. As a result, with the medical big data, the greatly influential and threatening medical issues are defined. Accordingly, in the analysis of medical big data, it is possible to predict disease relationships, spread direction, and risk. In addition, by calculating the effective index of the policies on health diseases, it is possible to provide the national health indexes required by WHO or OECD. Korea National Health and Nutrition Examination Survey have about 600 variables of examination, health questions, and nutrition. Examination survey includes the index of obesity, hypertension, diabetes, dyslipidemia, lung disease, liver disease, kidney disease, anemia, oral disease, eyesight, grip, and heavy metals. Health survey includes questions about smoking, drinking, physical activity, obesity, weight control, mental health, safety consciousness, disease susceptibility, medical use, limited activity, quality of life, damage, health examination, education, economic activity, and household survey. Nutrition survey includes food, details of food intake, dietary behavior, use of nutrient description, nutrition education, dietary supplement, stability of food products, energy, and frequency of the food supplying nutrients. Among these variables, context variables of depression are used. In addition, transactions with missing data, NULL data, no idea data/no response data are removed. Through the normalization process, context variables of depression are organized finally.

The second step is variable selection. In multiple regression analysis, context variables most related to depression are selected. The target variable for selecting the context variable is depression, and an independent variable is the one for prediction. A context variable that meets the significance level of 0.05 is selected, and a regression equation is also drawn.

The last step is the prediction of depression risk in context-DNN. Based on the selected context variables, the context information with high relation is extracted, and then is used as an input of context-DNN for learning. For the connection of the learned DNN, the regression equation is applied. Each output of context-DNN is put in the regression equation to predict a degree of depression risk. Fig. 2 show the context-DNN based regression prediction process for the risk of depression.

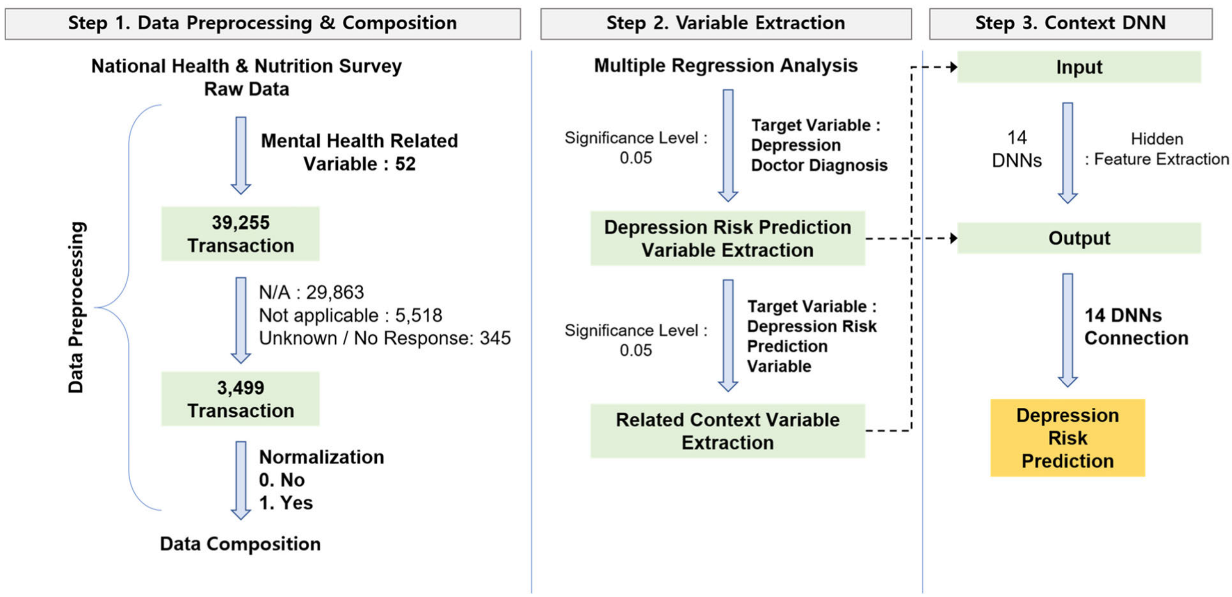

FIGURE 2. Context DNN process for predicting depression risk using multiple regression.

A. DATA PREPROCESSING AND COMPOSITION FOR DEPRESSION PREDICTION

Data is used to predict a user's risk of depression according to physical illnesses or external conditions. This study utilizes the context variables related to depression among the raw data of the Korea National Health and Nutrition Examination Survey at the Centers for Disease Control and Prevention Korea.

In order to predict the context of depression, unnecessary variables are removed so as to improve analysis accuracy, and data is pre-processed to fit the model. The pre-processing method is used to integrate, reduce, convert, refine, and process incomplete data with low quality [15], [25], [26]. Pre-processing makes it possible to support accurate analysis and improve analysis quality. The data means 39,225 persons' data consisting of 262 context variables, including age, whether to have dyslipidemia, bran stroke, myocardial infarction, angina, arthritis, tuberculosis, asthma, diabetes, thyroid disease, stomach cancer, liver cancer, colon cancer, breast cancer, cervical cancer, lung cancer, thyroid cancer, depression, hepatitis B, or hepatitis C, whether to be diagnosed by medical doctor, the first diagnosis time, and whether to treat such a disease. Among 262 context variables, the context variables closely related to depression are extracted. In addition, from the extracted variables, removed are the transactions that have NULL data, missing data, and no idea/no response based category-type data (9, 99 and 999). The structured data consists of categorical data and continuous data. Categorical data can be classified into particular categories (nominal data and ordinal data). Continuous data has an infinite value and is presented numerically. Table 1 shows the data organization for depression prediction.

For the prevention of overfitting, pre-processed data is split into a training set (70%), a validation set (10%), and a test set (20%). A training set is used to find the optimal model. A validation set is used to judge a model's goodness-of-fit. A test set is used to evaluate performance. In order to solve the problem of overfitting that occurs at the time of adapting to a training set excessively and thereby losing the ability of generalization, the weight decay technique is applied.

B. VARIABLE SELECTION USING MULTIPLE REGRESSION ANALYSIS

Depression can incur in diverse context so that independent variables are multiple. In multiple regression analyses, the variables for predicting the risk of depression are extracted. At this time, a significance level is applied. In addition, a regression equation is extracted. The risk of depression is set as a target variable, and a context variable for predicting depression is set as an independent variable. A significance level represents the stochastic precision of judgment in the hypothesis test [27]. A significance level is presented with (1-confidence). If a significance level is 0.1, it means that confidence is 0.9. In order to predict depression in highly reliable analysis, this study sets a significance level to 0.05 in regression analysis. When there are two variables, a regression equation is used to predict one variable from the other variable [10]. Table 2 shows the result of the regression analysis at the significance level of 0.05. Estimate means an estimated value. Std. Error represents a standard error. t-value is a t-test statistical value, meaning statistics of the influence of variables. Fourteen predictor variables of depression risk are extracted.

TABLE 1. The data organization for depression prediction.

No	Attributes	Name	Value
1	D_1_1	Subjective health state	1 : Very good 2 : Good 3 : Normal 4 : Bad 5 : Very bad
2	BD1_11	Yearly drinking frequency	1 : Not drinking at all 2 : Less than once a month 3 : About once a month 4 : 2 or 4 times a month 5 : 2 or 3 times a week. 6 : More than 4 times a week
3	BD2_1	A drinking amount per time	1 : 1-2 cups 2 : 3-4 cups 3 : 5-6 cups 4 : 7-9 cups 5 : More than 10 cups
4	Total_slp_wk	Average daily sleep hours on weekdays	0 : Not at all 1 : 1 day 2 : 2 days 3 : 3 days 4 : 4 days 5 : 5 days 6 : 6 days 7 : 7 days (every day)
5	Total_slp_wd	Average daily sleep hours on weekends	
6	BE3_31	The number of days of walk in a week	
...	...	...	...
49	BE3_71	Daily whether to have physical activity with high intensity	0 : No 1 : Yes
50	BE3_81	Daily whether to have highly intensive physical activity with middle intensity	0 : No 1 : Yes
51	BE5_1	The number of days of muscle exercise in a week	0 : Not at all 1 : 1 day 2 : 2 days 3 : 3 days 4 : 4 days 5 : More than 5 days
52	pa_aerobic	Practice rate of aerobic physical activity	0 : No 1 : Yes

Equation (2) is used to predict the depression predicted as the result of regression analysis. In terms of Depression Risk Prediction of y-intercept, the probability of information on depression risk is generated as an output. Each one of the extracted predictor variables is multiplied by estimated value, and then the multiplied results are all added up in order to draw the probability information on depression risk.

TABLE 2. The result of the regression analysis.

Variable	Description	Estimate	Std.Error	t-value
D_1_1	Subjective health state	0.01844	0.004418	4.173
DI6_dg	Whether to be diagnosed with angina by medical doctor	0.045029	0.025016	1.800
DM1_dg	Whether to be diagnosed with arthritis by medical doctor	0.046192	0.011481	4.023
DJ2_dg	Whether to be diagnosed with tuberculosis by medical doctor	0.045188	0.018096	2.497
DJ4_dg	Whether to be diagnosed with asthma by medical doctor	0.080489	0.02209	3.644
DE2_dg	Whether to be diagnosed with thyroid disease by medical doctor	0.030937	0.017001	1.820
DC2_dg	Whether to be diagnosed with liver cancer by medical doctor	0.171391	0.099607	1.721
EC1_1	Economic activity state	-0.03217	0.007072	-4.550
BD1_11	Yearly drinking frequency	-0.00706	0.002206	-3.199
BE8_1	Daily sitting hours (time)	-0.00228	0.000979	-2.325
mh_PHQ_S	The sum of points for nine items in depression screening scale	0.010489	0.001021	10.273
mh_stress	Stress recognition rate	0.027921	0.008387	3.329
L_LN	Whether to go without lunch a day ago	0.023797	0.012959	1.836
L_BR_FQ	Weekly breakfast frequency in the latest one year	-0.00763	0.004112	-1.855

$$\begin{aligned} DRP = {} & 0.016958 + (0.01844 \times D\_1\_1) \\ & + (0.045029 \times DI6\_dg) + (0.046192 \times DM1\_dg) \\ & + (0.045188 \times DJ2\_dg) + (0.080489 \times DJ4\_dg) \\ & + (0.030937 \times DE2\_dg) + (0.171391 \times DC2\_dg) \\ & + (-0.032174 \times EC1\_1) + (-0.007057 \times BD1\_11) \\ & + (-0.002287 \times BE8\_1) + (0.010489 \times mh\_PHQ\_S) \\ & + (0.027921 \times mh\_stress) + (0.023797 \times L\_LN) \\ & + (-0.00763 \times L\_BR\_FQ) \end{aligned} \tag{2}$$

To find the relation of relevant variables, there are methods using an ontology, coefficient of correlation, association rule, and regression analysis. This study utilizes regression analysis as a method of finding a clear relation and cause-and-effect relationship, in consideration of data features. In order to find the potential context that has a high relation with depression predictor variables, 14 regression variables are used as a dependent variable and the relevant context is extracted. For the extraction of context variables, 14 classes are organized according to the raw data guidelines [24] of

TABLE 3. The context information that have highly relations with regression variables of depression risk.

Class	Depression Risk Prediction Variable	Related Context Information
C1	D_1_1	[D_2_wk], [D_2_1], [age] [HE_sbp], [HE_dbp], [HE_wt], [HE_BMI], [HE_glu]
C2	DI6_dg	[DI6_pr], [BS3_1], [BS3_2], [HE_HP], [HE_DM]
C3	DM1_dg	[DM1_pr], [DM2_dg], [DM2_pr], [DM2_ag], [DM2_pt], [DM3_dg], [DM3_pr]
C4	DJ2_dg	[DJ2_pr], [DJ2_ag], [DJ2_pt]
C5	DJ4_dg	[DJ4_pr], [DJ4_pt], [DJ4_3]
C6	DE2_dg	[DE2_pr], [DE2_ag], [DE2_pt], [DC7_dg], [DC7_pr], [HE_tsh], [HE_ft4], [HE_tpoab], [HE_uiod]
C7	DC2_dg	[DC2_pr], [DC1_pr], [DC4_pr]
C8	EC1_1	[EC_occp], [EC_stt_1], [EC_stt_2], [EC_wh], [EC_wht_0], [EC_wht_23], [EC_wht_5], [EC_lgw_4], [EC_lgw_5], [EC_pedu_1], [EC_pedu_2]
C9	BD1_11	[BD1], [BD2], [BD2_1], [BD2_31], [BD2_32]
C10	BE8_1	[BE8_2], [BE3_32], [pa_aerobic], [LQ4_03]
C11	mh_PHQ_S	[BP_PHQ_1] ~ [BP_PHQ_9], [BP16_11], [BP16_12], [BP16_13], [BP16_14], [Total_slp_wk], [Total_slp_wd]
C12	mh_stress	[BP1], [BP7]
C13	L_LN	[L_LN_FQ], [L_LN_TO], [L_LN_WHO]
C14	L_BR_FQ	[L_BR], [L_BR_TO], [L_BR_WHO]

the Korea National Health and Nutrition Examination Survey and the data of the National Health Information Portal [28]. Each class has the context variables related to 14 depression predictor variables. Table 3 shows the context variables that have highly relations with regression variables of depression risk.

C. CONTEXT PREDICTION USING CONTEXT DEEP NEURAL NETWORK MODEL

This study proposes context Deep Neural Network (context-DNN) to predict the probability information on context as a model in the combination with context information and DNN. Context is information on a situation expanded to a surrounding circumstance, time, and an environment, beyond the simple meaning of the text in a dictionary. To composed context information, context variables are defined and used depending on outer context, inner context, service context [1], [29]. Data is classified into structured data, unstructured data, and semi-structured data. The data area is extracted in order to show its meaning well depending on data type [30]–[32]. Structured data has a given data format and uses the relational database since it supports operations. Semi-structured data such as XML and HTML have a data format but are not operable. Unstructured data such as video and image has no data format and is not operable. Context variable data is structured and numerical so that it is operable. Context variables represent context information collected in everyday life, such as sleep hours, an activity amount, an amount of food intake, frequency, and whether to have an illness. In addition, context information is used as an input of the context-DNN model.

Fig. 3 shows the architecture of the context-DNN model. A DNN model has the context information related to depression predictor variables as inputs. As outputs of the DNN model, depression predictor variables are displayed. For depression prediction, the weight of a regression equation and an output of DNN are combined together. DNNs for 14 context are individually learned, and SUM operation is executed to make one model. At this time, the weight of the regression equation means an estimated value of a depression predictor variable in the equation. An output of context-DNN model represents a degree of depression risk.

IV. RESULT AND PERFORMANCE EVALUATION

A. CONTEXT PREDICTION FOR DEPRESSION RISK

To predict the risk of depression depending on changing context, deep learning is applied. A DNN is a sort of machine learning to make the hidden pattern of large data or documents modeled as a network with complex multiple layers. In multiple hidden layers, patterns of various input data are learned through classification and clustering [12], [33], [45], [46]. The proposed context-DNN model is the method of making dynamic context of depression as context information for prediction. As a neural network model to predict context information of depression, the model has N context variables related to predictor variables of depression risk as inputs, and predictor variables of depression risk as outputs. There are 14 predictor variables of depression risk, each of which is used for a DNN. A total of 14 DNNs are individually learned. With an increase in data, many DNN layers are piled up. Individual learning can solve the problem and shorten the time needed for learning. For the connection of the learned DNNs, the regression equation (2) is applied. An output of DNN is put in the regression equation to predict a degree of depression. Fig. 4 show the context prediction process for the risk of depression.

In Fig. 4, the input data of the DNN model in the context of relational context variables. Hidden layers become dense. The model of fourteen DNNs uses Adam as an optimization function, relu as an activation function, and Mean Absolute Error as a loss function in hidden layers. In the output layer, a variable fitting the input context is extracted. The variables extracted in DNNs are put in the regression equation to predict the risk of depression. The result of risk prediction is a value between 0 and 1. The predicted risk of depression is categorized into four stages-good, not bad, risky, and very

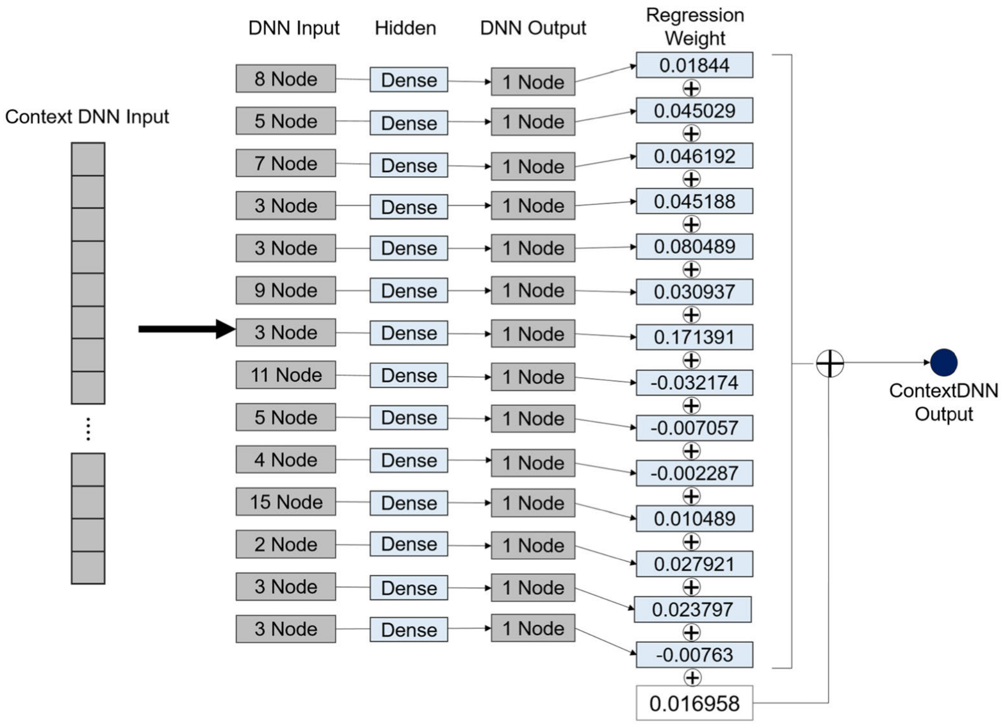

FIGURE 3. The architecture of the context-DNN model.

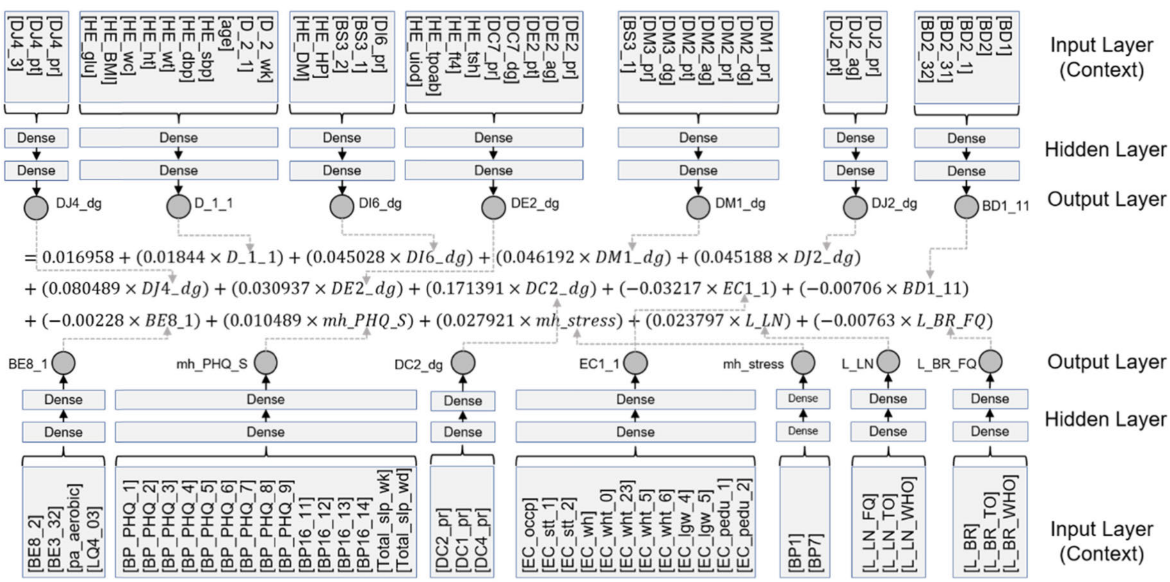

FIGURE 4. Context prediction process for the risk of depression.

risky-on the basis of the depression scale 'CES-D' of National Mental Health Center [23], [44], [48].

If a value of depression is between 0.0 and 0.55, it means 'good'; if between 0.56 and 0.60, it means 'not bad'; if between 0.61 and 0.64, it means 'risky'; if between 0.65 and 1.0, it means 'very risky'. For example, if the values of classes C1 to C14 as DNN outputs are 2, 1, 1, 1, 1, 1, 1, 6, 20, 30, 1, 1, and 3, each one of them is put in the regression equation to execute an operation. As a result, 0.696 is calculated. The value means 'very risky' stage. Accordingly, it is possible to predict the context for depression risk and to find one's depression stage. Fig. 5 shows the prediction system for depression risk according to context.

In Fig. 5, the prediction system for the risk of depression consists of user information, context information, history of illnesses, and analysis of depression risk. User information

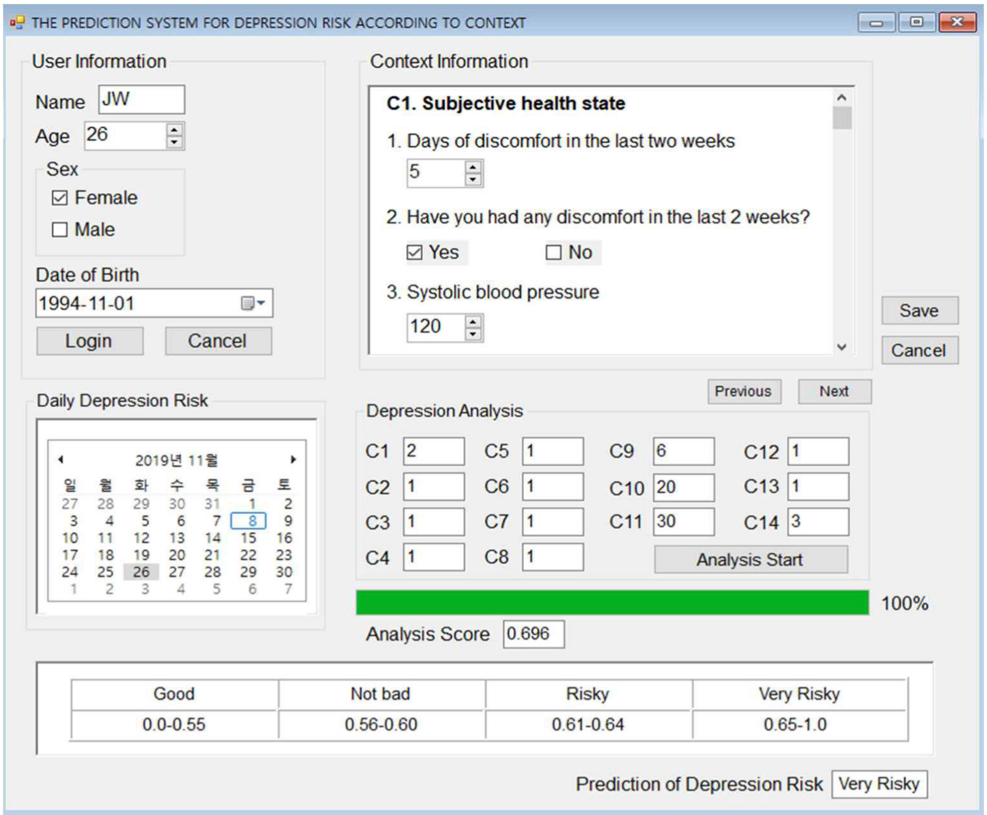

FIGURE 5. The prediction system for depression risk according to context.

can be accessed with a user's name, age, sex, and date of birth. When a user has access, it is possible to find the monthly records of depression [47]. In terms of context information, a user answers questions, each of which has a value recorded. The values are used in the regression equation for predicting the risk of depression, and the calculated result is analyzed. As a result, a stage of depression risk (good, not bad, risky, and very risky) is given to the user. In this way, the user is able to find a degree of depression risk.

B. PERFORMANCE EVALUATION

The operating system and hardware specifications of the context-DNN proposed in this study are Window10 Pro, Intel(R) Core(TM) i5-4690 3.50GHz, and 16GB RAM. As software, R Studio 1.1.463 and R deep learning library *karas* were used for design. The performance of the proposed model is evaluated in terms of goodness-of-fit, cross-validation, accuracy, and recall.

As for performance, firstly, the goodness-of-fit of the proposed context-DNN is evaluated. For comparative evaluation, multiple regression analysis, general DNN, and the proposed context-DNN model are applied to predict the risk of depression risk. Table 4 shows the results of the goodness-of-fit evaluation. Results of Acc (accuracy) and Loss was used training data for weights and bias learning. In addition, results of val_Acc (validation accuracy) and val_Loss was used validation data for the evaluation performance of hyper parameters.

TABLE 4. Results of goodness-of-fit evaluation.

	Depression Risk Prediction	
Multiple Regression	RSE	0.1992
	Multiple R^2	0.1149
	R^2 adj	0.1018
DNN	Loss	4.541
	Acc	0.8546
	val_Loss	5.429
	val_Acc	0.8457
Context-DNN	loss	2.541
	Acc	0.9546
	val_Loss	3.429
	val_Acc	0.9457

In the table 4, RSE represents Residual standard error [34], which is the difference between a predicted value in the regression model and real data. Multiple R^2 represents the variance rate of dependent variables which can be explained by the regression model [35]. If the number of variables increases, a coefficient of determination becomes large.

A rise in variables only leads to high explanatory power. To overcome the problem, R^2 adj (Adjusted R-squared) is used [36]. It represents the variance rate of a dependent variable in consideration of the number of variables. A loss value is used to evaluate the difference between the data obtained by learning and real data [37]. The larger the value is, the more the learned data is inconsistent with real data. According to the performance evaluation, the context-DNN model showed the best performance. In the regression analysis to predict the risk of depression, a residual value was measured to be about 0.2. Although the difference between the predicted value and real value is low, independent variables to predict depression do not have enough explanatory power. In addition, the context-DNN model had a lower value of loss and higher accuracy than general DNN. It means that the context-DNN model has better goodness-of-fit than multiple regression analysis and general DNN in terms of the prediction of depression risk.

Secondly, to evaluate context-DNN model, a cross-validation test is conducted with the use of K-fold. The cross-validation technique can use all data sets as evaluation data and prevents overfitting that arises in a particular evaluation data set [44], [45]. All data sets can be used as training data. Therefore, it is possible to improve accuracy and prevent under fitting that occurs due to a lack of data [38], [39]. K-fold cross-validation splits data into k data sets and then evaluates a model k time. Of k data sets, one set is used as test data, and k-1 sets are used as training data. Learning occurs through k times of repetition [40], [41]. This study utilizes 10-fold cross-validation to learn from data sets. When experimental data is put in, the probability information on a degree of context information is generated as an output. If new data comes in, a degree of model's prediction is analyzed. Table 5 shows an error of the mean according to 10-fold cross-validation. If new information comes in with an error of the mean (about 0.001 to 0.02), the proposed model is capable of making an accurate prediction.

TABLE 5. An error of the mean according to 10-fold cross-validation.

Cross-Validation	Average Error
1-Fold	0.014914
2-Fold	0.014686
3-Fold	0.019643
4-Fold	0.013444
5-Fold	0.019033
6-Fold	0.012358
7-Fold	0.017855
8-Fold	0.012406
9-Fold	0.016997
10-Fold	0.012895

Thirdly, with the use of the pre-processed data sets, the accuracy and recall of context-DNN and DNN are evaluated according to their epoch. Fig. 6 show the comparative result of accuracy between context-DNN and DNN. According to the evaluation of accuracy, context-DNN improved accuracy by learning DNNs individually. The proposed model was evaluated to have high accuracy since it accomplished accurate learning according to a learning rate. On the contrary, DNN with one model failed to learn from data so that it had low accuracy.

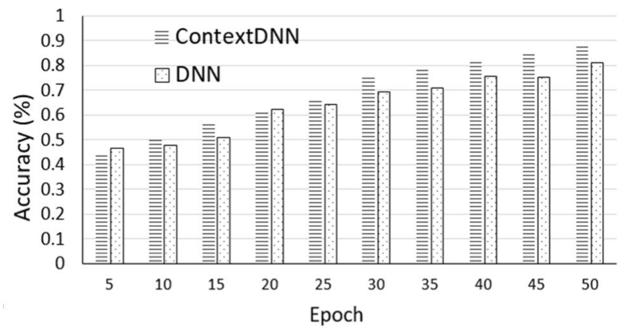

FIGURE 6. Comparative result of accuracy between context-DNN and DNN.

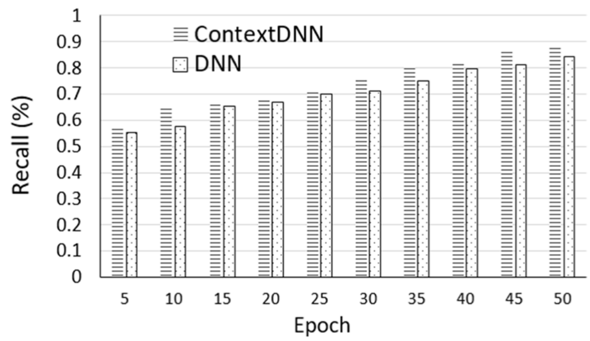

FIGURE 7. Comparative result of recall between context-DNN and DNN.

Fig. 7 shows the comparative result of recall between context-DNN and DNN. As shown in Fig. 7, the proposed context-DNN model had excellent recall according to the epoch. It predicts the risk of depression in consideration of changing context. For this reason, the model's recall was evaluated to be high when tit was tested with new data. On the contrary, general DNN supports general prediction only.

Since the DNN had many hidden layers, it failed to learn data patterns. Therefore, it had low recall according to a user's changing context. Given the results, context-DNN is good to predict a flexible context and care for mental health.

V. CONCLUSION

This study proposed the Context-DNN model for predicting depression risk using multiple-regression. The context-DNN model was developed in the way of making the context information influencing depression as context and then combining context and DNNs. In consideration of context information, the method of predicting the risk of depression was developed. The context-DNN model to prediction method pre-processes the context variables of the data in the Korea

National Health and Nutrition Examination Survey and then generates data. In regression analysis, variables related to depression are extracted. In this case, variables of depression are set as a dependent variable, and context variables that have meaningful associations are extracted. The context variables related to depression are used to generate context information. Finally, the context-DNN model is applied to predict the risk of depression. Context is context information related to depression. Context is used as the input data of each DNN model. The context information is learned by each DNN, and then 14 models are created. To connect all of these models together, the regression equation to predict depression is applied. In the equation, it is possible to predict a degree of depression risk. The predicted value is a value between 0 and 1. The risk of depression has four stages, which are 'good', 'not bad', 'risky', and 'very risky'.

The proposed Context-DNN conducted four performance evaluations. First, as a result of the conformity assessment, the multiple regression analysis evaluated the variable's explanatory power as unfit, and the DNN was evaluated more unfit due to the lower loss function and accuracy than the Context-DNN. Second, cross-validation was conducted to evaluate the model of Context-DNN. Third, accuracy and recall were compared with general DNN. As a result, the performance of Context-DNN was better than that of general DNN. The proposed model learned DNNs individually so as to improve accuracy and utilized predictor variables through regression analysis. Therefore, the proposed context-DNN model makes it possible to judge one's context for the risk of depression accurately, to continue to care for mental health, and to prevent depression. In the future, we will implement a model that available personalized health management by classifying contexts according to individual situations into internal and external context information and discovering external and internal influence factors that affect each.

REFERENCES

[1] J.-W. Baek, H. Jung, and K. Chung, "Context mining based mental health model for lifecare platform," *Medico-Legal Update*, vol. 19, no. 1, pp. 674–679, Apr. 2019.

[2] J. C. Kim and K. Chung, "Prediction model of user physical activity using data characteristics-based long short-term memory recurrent neural networks," *KSII Trans. Internet Inf. Syst.*, vol. 13, no. 4, pp. 2060–2077, Apr. 2019.

[3] J. Guan, "Artificial intelligence in healthcare and medicine: Promises, ethical challenges and governance," *Chin. Med. Sci. J.*, vol. 34, no. 2, pp. 76–83, Jun. 2019.

[4] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, and V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease," *Future Gener. Comput. Syst.*, vol. 83, pp. 366–373, Jun. 2018.

[5] G. Olmschenk, Z. Zhu, and H. Tang, "Generalizing semi-supervised generative adversarial networks to regression using feature contrasting," *Comput. Vis. Image Understand.*, vol. 186, pp. 1–12, Sep. 2019.

[6] M. Sato-Ilic, "Knowledge-based comparable predicted values in regression analysis," *Procedia Comput. Sci.*, vol. 114, pp. 216–223, Jan. 2017.

[7] J. Chantana, Y. Kawano, A. Kamei, and T. Minemoto, "Description of degradation of output performance for photovoltaic modules by multiple regression analysis based on environmental factors," *Optik*, vol. 179, pp. 1063–1070, Feb. 2019.

[8] J. Zhang and L. C. Thomas, "Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD," *Int. J. Forecasting*, vol. 28, no. 1, pp. 204–215, Jan./Mar. 2012.

[9] Ö. Polat, "A robust regression based classifier with determination of optimal feature set," *J. Appl. Res. Technol.*, vol. 13, no. 4, pp. 443–446, Aug. 2015.

[10] M. Gong, K. Pan, and Y. Xie, "Differential privacy preservation in regression analysis based on relevance," *Knowl.-Based Syst.*, vol. 173, pp. 140–149, Jun. 2019.

[11] H. Jiang, C. Kwong, C. Chan, and K. Yung, "A multi-objective evolutionary approach for fuzzy regression analysis," *Expert Syst. Appl.*, vol. 130, pp. 225–235, Sep. 2019.

[12] N. M. Joy, M. K. Baskar, and S. Umesh, "DNNs for unsupervised extraction of pseudo speaker-normalized features without explicit adaptation data," *Speech Commun.*, vol. 92, pp. 64–76, Sep. 2017.

[13] R. Tang, S. Fong, S. Deb, A. V. Vasilakos, and R. C. Millham, "Dynamic group optimisation algorithm for training feed-forward neural networks," *Neurocomputing*, vol. 314, pp. 1–19, Nov. 2018.

[14] M. Khashei and M. Bijari, "Hybridization of the probabilistic neural networks with feed-forward neural networks for forecasting," *Eng. Appl. Artif. Intell.*, vol. 25, no. 6, pp. 1277–1288, Sep. 2012.

[15] H. H. Örkcü and H. Bal, "Comparing performances of backpropagation and genetic algorithms in the data classification," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3703–3709, Apr. 2011.

[16] X. Wang, Z. Tang, H. Tamura, and M. Ishii, "A modified error function for the backpropagation algorithm," *Neurocomputing*, vol. 57, pp. 477–484, Mar. 2004.

[17] E. Moulay, V. Léchappé, and F. Plestan, "Properties of the sign gradient descent algorithms," *Inf. Sci.*, vol. 492, pp. 29–39, Aug. 2019.

[18] A. Sharma, "Guided stochastic gradient descent algorithm for inconsistent datasets," *Appl. Soft Comput.*, vol. 73, pp. 1068–1080, Dec. 2018.

[19] Z. Chang, Y. Zhang, and W. Chen, "Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform," *Energy*, vol. 187, Nov. 2019, Art. no. 115804.

[20] J. Shi, J. Song, B. Song, and W. F. Lu, "Multi-objective optimization design through machine learning for drop-on-demand bioprinting," *Engineering*, vol. 5, no. 3, pp. 586–593, Jun. 2019.

[21] B. Mei and Y. Xu, "Multi-task least squares twin support vector machine for classification," *Neurocomputing*, vol. 338, pp. 26–33, Apr. 2019.

[22] K. Li, Y. Wu, Y. Nan, P. Li, and Y. Li, "Hierarchical multi-class classification in multimodal spacecraft data using DNN and weighted support vector machine," *Neurocomputing*, vol. 259, pp. 55–65, Oct. 2017.

[23] *Korea Centers for Disease Control and Prevention & National Nutrition Survey*. Accessed: Nov. 10, 2019. [Online]. Available: https://knhanes.cdc.go.kr/

[24] *The Sixth Korea National Health and Nutrition Examination Survey*, Korea Centers for Disease Control and Prevention, Cheongju-si, South Korea, 2015.

[25] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, p. 107, Jan. 2008.

[26] A. Idri, H. Benhar, J. Fernández-Alemán, and I. Kadi, "A systematic map of medical data preprocessing in knowledge discovery," *Comput. Methods Programs Biomed.*, vol. 162, pp. 69–85, Aug. 2018.

[27] S. Hart and B. Weiss, "Significance levels for multiple tests," *Statist. Probab. Lett.*, vol. 35, no. 1, pp. 43–48, Aug. 1997.

[28] *Korea Centers for Disease Control and Prevention*. Accessed: Nov. 10, 2019. [Online]. Available: http://health.cdc.go.kr/

[29] J. W. Baek, K. Chung, J. Kim, and H. Jung, "Cloud-based ontology context mining using deep learning in healthcare," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, no. 8, pp. 296–300, 2019.

[30] D. Bacciu and D. Castellana, "Bayesian mixtures of hidden tree Markov models for structured data clustering," *Neurocomputing*, vol. 342, pp. 49–59, May 2019.

[31] Y. Lin, Z. Jun, M. Hongyan, Z. Zhongwei, and F. Zhanfang, "A method of extracting the semi-structured data implication rules," *Procedia Comput. Sci.*, vol. 131, pp. 706–716, Jan. 2018.

[32] J. Gardner and L. Xiong, "An integrated framework for de-identifying unstructured medical data," *Data Knowl. Eng.*, vol. 68, no. 12, pp. 1441–1451, Dec. 2009.

[33] M. J. Shafiee, B. Chwyl, F. Li, R. Chen, M. Karg, C. Scharfenberger, and A. Wong, "StressedNets: Efficient feature representations via stress-induced evolutionary synthesis of deep neural networks," *Neurocomputing*, vol. 352, pp. 93–105, Aug. 2019.

[34] Y. Chen, J. Jiang, and A. Narayan, "A robust error estimator and a residual-free error indicator for reduced basis methods," *Comput. Math. Appl.*, vol. 77, no. 7, pp. 1963–1979, Apr. 2019.
[35] C.-L. Cheng, Shalabh, and G. Garg, "Coefficient of determination for multiple measurement error models," *J. Multivariate Anal.*, vol. 126, pp. 137–152, Apr. 2014.
[36] L. Ricci, "Adjusted -squared type measure for exponential dispersion models," *Statist. Probab. Lett.*, vol. 80, nos. 17–18, pp. 1365–1368, Sep. 2010.
[37] W. Zhang, Y. Du, T. Yoshida, and Y. Yang, "DeepRec: A deep neural network approach to recommendation with item embedding and weighted loss function," *Inf. Sci.*, vol. 470, pp. 121–140, Jan. 2019.
[38] L. Xu, O. Hu, Y. Guo, M. Zhang, D. Lu, C.-B. Cai, S. Xie, M. Goodarzi, H.-Y. Fu, and Y.-B. She, "Representative splitting cross validation," *Chemometrics Intell. Lab. Syst.*, vol. 183, pp. 29–35, Dec. 2018.
[39] J. Li, K. Perrine, L. Wu, and C. M. Walton, "Cross-validating traffic speed measurements from probe and stationary sensors through state reconstruction," *Int. J. Transp. Sci. Technol.*, vol. 8, no. 3, pp. 290–303, Sep. 2019.
[40] P. Jiang and J. Chen, "Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation," *Neurocomputing*, vol. 198, pp. 40–47, Jul. 2016.
[41] G. Jiang and W. Wang, "Error estimation based on variance analysis of K -fold cross-validation," *Pattern Recognit.*, vol. 69, pp. 94–106, Sep. 2017.
[42] *Health Insurance Review and Assessment Service (HIRA)*. Accessed: Nov. 15, 2019. [Online]. Available: http://www.hira.or.kr/eng/
[43] *Healthcare Bigdata Hub of Health Insurance Review and Assessment Service*. Accessed: Nov. 15, 2019. [Online]. Available: http://opendata.hira.or.kr/
[44] K. Chung and J. Kim, "Activity based nutrition management model for healthcare using similar group analysis," *Technol. Health Care*, vol. 27, no. 5, pp. 473–485, Sep. 2019.
[45] J.-W. Baek, J.-C. Kim, J. Chun, and K. Chung, "Hybrid clustering based health decision-making for improving dietary habits," *Technol. Health Care*, vol. 27, no. 5, pp. 459–472, Sep. 2019.
[46] J.-C. Kim and K. Chung, "Associative feature information extraction using text mining from health big data," *Wireless Pers. Commun.*, vol. 105, no. 2, pp. 691–707, Mar. 2019.
[47] K. Chung and R. C. Park, "Cloud based u-healthcare network with QoS guarantee for mobile health service," *Cluster Comput.*, vol. 22, no. S1, pp. 2001–2015, Jan. 2019.
[48] C.-W. Song, H. Jung, and K. Chung, "Development of a medical big-data mining process using topic modeling," *Cluster Comput.*, vol. 22, no. S1, pp. 1949–1958, Jan. 2019.

JI-WON BAEK received the B.S. degree from the School of Computer Information Engineering, Sangji University, South Korea, in 2017. She is currently pursuing the master's degree with the Department of Computer Science, Kyonggi University, South Korea. She has worked with the Data Management Department, Infiniq Company, Ltd. She has been a Researcher with the Data Mining Laboratory, Kyonggi University. Her research interests include data mining, data management, knowledge systems, automotive testing, deep learning, medical data mining, healthcare, and recommendation.

KYUNGYONG CHUNG received the B.S., M.S., and Ph.D. degrees from the Department of Computer Information Engineering, Inha University, South Korea, in 2000, 2002, and 2005, respectively. He has worked with the Software Technology Leading Department, South Korea IT Industry Promotion Agency (KIPA). From 2006 to 2016, he was a Professor with the School of Computer Information Engineering, Sangji University, South Korea. Since 2017, he has been a Professor with the Division of Computer Science and Engineering, Kyonggi University, South Korea. His research interests include data mining, artificial intelligence, healthcare, knowledge systems, ambient intelligence, and recommendation.