IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 22, 2020, accepted October 2, 2020, date of publication October 16, 2020, date of current version November 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031762

Hybrid Multi-Modal Deep Learning using Collaborative Concat Layer in Health Bigdata

JOO-CHANG KIM[1] AND KYUNGYONG CHUNG[2]
[1]Department of Computer Science, Kyonggi University, Suwon 16227, South Korea
[2]Division of Computer Science and Engineering, Kyonggi University, Suwon 16227, South Korea

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

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 20CTAP-C157011-01).

ABSTRACT A health model based on data has various missing values depending on the user situation, and the accuracy of a health model requiring variables that the user cannot collect appears low. A deep learning health model is fitted by learning weights to increase accuracy. In the process of applying a deep-learning-based health model to the user situation, accuracy may be degraded if learning is omitted. In this paper, we propose hybrid multimodal deep learning using a collaborative concat layer in health big data. The proposed method uses a machine learning technique to alleviate the issue caused by the change in the data observation range according to a change in the user situation, and occurring in multimodal health deep learning. It is a layer composed of the connection, input, and output of the model of the collaborative node (CN). A CN is a node that predicts absent variables through filtering using the similarity of input values. With CN, a collaborative concat layer (CCL) that handles missing values from the input of the health model can be configured, and the issue related to missing values occurring in the health model can be resolved. With the proposed CCL, it is possible to reuse existing models or construct new models through the concatenation of several single-modal deep learning models. By evaluating the effect on the input and output of the model according to the structural position of the CCL, various networks can be configured, and the performance of the single-modal model can be maintained. In particular, the accuracy of a deep learning model is more stable when the CCL is used, suggesting the experiment progress based on the assumption that a specific variable is absent depending on the user situation.

INDEX TERMS Health bigdata, data imputation, multi-modal, model concatenate, hybrid learning.

I. INTRODUCTION

Based on the development of 4th industrial technologies such as communication, information, and sensors, various convergence industries are growing rapidly. This is causing great changes not only in industrial fields from smart farms to smart homes, but also in the daily lives of people [1]. In particular, with the supply of health-related electronic medical records, personal health records, personal health equipment, etc., the data-based health industry is accumulating large amounts of data mainly in various countries, companies, and research institutes [2]. In order to utilize the accumulated data, support projects for machine learning and deep learning are in progress, such as building and disclosing data related to the public interest at the national level. Various health-related data generated from various devices are collected and used [3]. Personal health networks are being formed around IoT devices and smartphones, and large amounts of data generated from the networks are being distributed [4]. With the development of information and communication, smart phones, personal health devices, IoT, etc., data is also widely collected mainly by consumers [5]. Activity, location, heart rate, stress, blood pressure, weight, body fat, slee and others are continuously collected from various types of devices in everyday life through smart watches, smart bands, and smart phone APPs [6]. When data began to be collected, data mining and machine learning, which can be used with a small amount of data, attracted attention, but

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

deep learning is currently attracting attention as an environment to collect and distribute large amounts of data is created [7]–[9]. However, the model created through deep learning requires precise learning, so it has a problem that it is difficult to flexibly cope with the rapidly changing user's surroundings, health condition, and activity. The health status of a person is rapidly changed by various situation variables, and each user has a different device, so the composition of the situation variables is diverse, placing a limit to the use of the generalized model [10]–[12]. Health requires a deep learning model that can flexibly cope with the different variables available depending on the user's situation. Rather than a large deep learning model, therefore, it is necessary to divide it into smaller ones and apply a model suitable for the user's situation [13]. It is necessary to study hybrid modeling that complements the disadvantages arising from each other and maximizes the advantages by integrating data mining or machine learning techniques with a general deep learning model. In this article, we propose hybrid multi-modal deep learning using Collaborative Concat Layer in health big data. This is a method of flexibly connecting several single-modal models that can be used according to the user's situation with the data mining technique.

This article is organized as follows. Chapter 2 finds out smart health and multi-modal deep learning, while chapter 3 describes hybrid multi-modal deep learning using Collaborative Concat Layer in the health big data. Chapter 4 describes performance experiments of the proposed method, and Chapter 5 draws conclusions.

II. RELATED STUDIES

A. TRENDS IN SMART HEALTHCARE

In the medical industry, smart health is attracting attention as the available data increase, including personal health information, electronic medical records, national health information, and genome information. In particular, a variety of data are collected from wearable devices on a mobile health platform without limitation in place and time. Smart watches, smart bands, personal health devices and smart phone applications are accumulating user's movements, vital signs, lifelog, location, weather and others over time. These data are directly related to the user's health, and various studies are underway to utilize them. With the development of the industry, the range of data that can be collected by wearable devices is increasing, including body temperature, humidity, illuminance, temperature, air pressure, ultraviolet light, heart rate, activity, and sleep. Health models using data mining, machine learning, and deep learning are being developed to utilize such data. The National Health and Nutrition Examination Survey [14] data are collected through health, examination, and nutrition surveys.

Personal health habits, family history, disease records, and medical records are collected through health surveys, and pulse, blood pressure, weight, height, and blood sugar are collected through regular checkups. The nutrition survey collects the individuals' frequency of meals, the amount of food, the amount of water consumed, and whether to take nutrients through a survey. Although far from normal PHR, it has a high potential value because it contains many variables related to health. It consists of about 600 variables that can be created and managed by individuals. By analyzing this with data mining, it is possible to infer the relationships or rules of variables used in health. Figure 1 shows the smart healthcare networks.

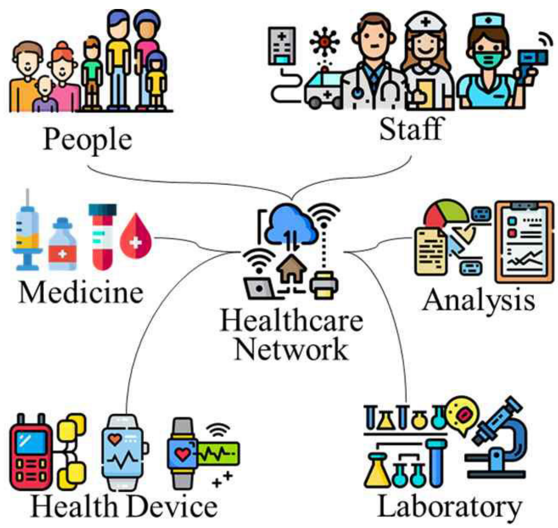

FIGURE 1. Smart healthcare networks.

B. DEEP LEARNING-BASED SMART HEALTHCARE MODEL

According to the 4th Industrial Revolution, data is extensively collected, and the smart health industry using it is continuously growing. Smart Health has been using a large amount of data due to the spread of various devices and information and communication engineering, and various deep learning models have been developed to utilize it [15], [34]. To this end, integration of data collected through various paths is required. In the case of a model that occurs in everyday life, values with the same time-step can be collected for one variable. In addition, there is a problem that it is difficult to process with one input when the observation time point is different depending on the variable. In data mining and deep learning, various models are created depending on variable composition, training data, and goals. In smart health, time-series, non-time-series, and image type data are dominant. Time-series data are data whose collected cycle is constant and fast, such as activity, heart rate, weight, and sleep time [16]. Non-time-series data is data with inconsistent cycles, such as doctor's diagnosis, prevalence, family history, or operation status. Image data represents CT, MRI, X-ray, ultrasound, etc. [17]. As shown above, various variables affect human health, and integration or connection is difficult because each feature is different. In addition, a large amount of learning data is required for learning of a deep learning model. The frequency of occurrence of health data is significantly different. If learning data is integrated and configured at a certain point in time, data with fast cycles causes loss of information, and data with slow cycles is filled with the same value in most transactions, causing a problem of low impact on the model results [18].

Time series data is data with temporal continuity and is modeled using RNNs such as LSTM and GRU. This can collect a large amount of data according to the user and configure learning data according to the user. The spread of IoT and personal health devices made the collection of personal health-related data easier, which became the driving force for the movement of health data producers from institutions to individuals. Currently, health-related data is being collected from various devices such as health bands, sleep meters, and personal blood glucose meters. Therefore, smart health services using personal health data have been developed [19].

Typically, a deep learning model suitable for time series data is a method of predicting current or future values from the user's existing records. It predicts changes in activity, heart rate, blood pressure and others to monitor user health and derive variables that negatively affect it. It models non-time-series data using the general aata mining technique. This means that there is usually one transaction for the user, so it is necessary to collect data from many users. Variables appearing in EMR such as health checkups and National Health and Nutrition Examination Survey [14] are used as non-time series data [20]. A large amount of data is analyzed with various data mining techniques and the results are used in universal services. Decision trees, association rules, and deep neural networks are mainly used. This enables the prediction or inference of the health status of actual users [21]. Image data utilizes data from many users, such as non-time series data. It predicts the user's health status by learning images generated in medical processes such as CT, MRI, and X-ray. Neural networks using various techniques are used based on CNN, which has high efficiency in image data [22].

When data is integrated in smart health and used for learning or analysis, problems such as feature extraction and absence of variables arise [23]. To overcome this problem, health data representing different characteristics is configured as each single-modal. Depending on the variables included in each modal, there are variations in learning efficiency or modeling results. The single-modal model according to the data characteristics can be configured as multi-modal to compensate for the shortcomings of each modal [24]. Different single-modal models are required according to the user's situation, and a general multi-modal model incorporating them has a problem that it is difficult to satisfy the diversity of users. Therefore, there is a need for a flexible way to connect single-modal models. Figure 2 shows the deep learning method according to health data.

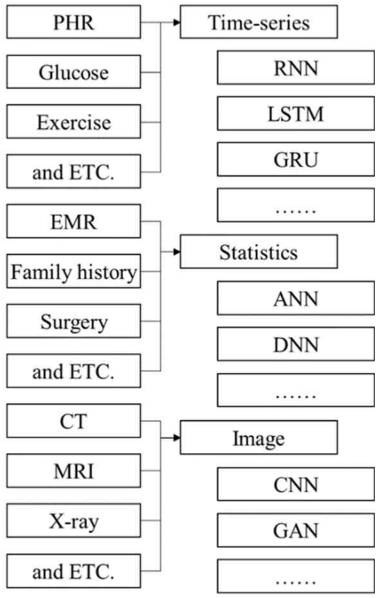

FIGURE 2. Deep learning method according to health data.

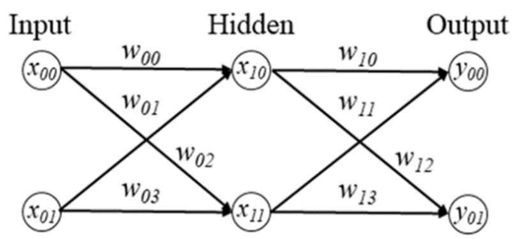

FIGURE 3. Basic structure of a neural network.

III. HYBRID MULTI-MODAL DEEP LEARNING USING COLLABORATIVE CONCAT LAYER IN HEALTH BIGDATA

Collected through various paths depending on the device or observation target, health data has different occurrence cycles, units, and shapes, so it is difficult to integrate it into one data set [25]. In addition, data is distributed across many institutions and personal devices. The main observation targets of health data are people, and it is difficult to integrate them due to various variable configurations or privacy issues. Various data in the forms of time series, non-time series, and images are collected according to the collection path. Since human health is changed by various factors, it is predicted using various types of data, and different models are required depending on the characteristics of the data to be predicted. Thus, the health model has been developed in various forms and variable configurations. Health data has a multi-modalitic feature collected through multiple channels. A multi-modal deep learning health prediction model is attracting attention, which extracts features with a single-modal model according to each data feature and combines them. The health model using deep learning shows higher performance than the existing machine learning or data mining. For this purpose, a neural network learning process using a large amount of data is required. The deep learning model is a neural network composed of multiple layers. Connections between nodes configured in the layer are made using the basic weight of deep learning, and the result is transmitted from the node to the next node through the calculation of the input value

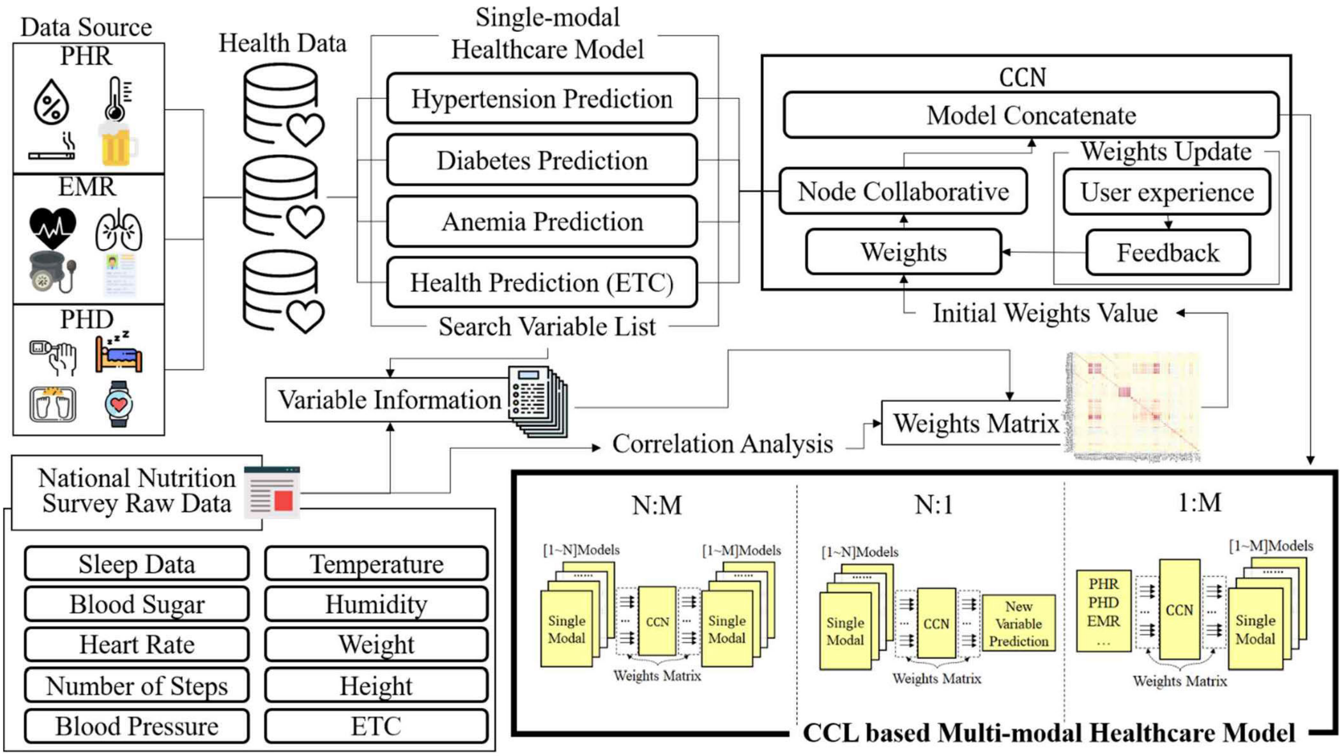

FIGURE 4. Architecture of hybrid multi-modal deep learning in the proposed health big data.

and the weight. The model update is repeated to output the value closest to the correct answer for the input value to learn the characteristics of the data [26]. Figure 3 shows the basic structure of a neural network.

Figure 3 is a simple neural network of $2 \times 2 \times 2$ type, which consists of 2 inputs and 2 outputs. Equation (1) shows the calculation method of the hidden node x_{10} in Figure 3, and equation (2) shows the calculation method of the output node y_{00}. In equations (1) and (2), indicates bias. Deep learning repeats weight correction through learning to find the optimal weight value. The initial value of the weight affects the learning efficiency or results. The initial value of the weight is generally 0, but a random value is used or an Autoencoder is used to perform pre-training [27].

$$x_{10} = x_{00} \times w_{00} + x_{01} \times w_{01} + b \quad (1)$$

$$y_{00} = x_{10} \times w_{10} + x_{11} \times w_{11} + b \quad (2)$$

This structure has a problem that accuracy is significantly lowered in the case of the absence of an input value. If an input value is absent for a short time, it can be replaced by an average value or median value. However, this method has a problem that in the case of long-term absence of input value due to the user's situation in smart health temporarily, it is impossible to reuse the already learned neural network model. For example, in a neural network model that predicts chronic diseases using context information such as height, weight, age, and blood sugar, information on blood sugar is collected through personal health devices. In general, long-term absence occurs because blood glucose meters do not have a device if a person is not a diabetic. In this case, the neural network model can be used even in an environment without a blood glucose meter by predicting blood sugar in which absence occurs by using a relationship between variables derived through correlation mining. In addition, it can be extended to a multi-modal neural network model through the connection between neural network models. By linking data mining and deep learning, it is possible to create hybrid models that overcome the limitations of common models. Collaborative Concat Layer (CCL) is a method of linking based on correlation mining. It uses the Weight Matrix to predict the absent input value through the relationship between variables. Figure 4 shows the architecture of hybrid multi-modal deep learning in the proposed health big data.

A. WEIGHT MATRIX USING VECTORIZATION BY HEALTH VARIABLE

The Weight Matrix is a matrix for storing the initial weight of the Collaborative Concat Node. This consists of the result of data mining of raw data for the National Health and Nutrition Survey in 2016 and 2018. In 2016 and 2018, additional information on PHQ is collected and the two-year dataset is used. The National Health and Nutrition Examination Survey [14] data is statistical data for various people and contains a large amount of missing values depending on the situation of people. Missing values directly affect the results of data mining and deep learning, thereby requiring pre-processing.

TABLE 1. Variable information.

Variable Name	Code	Average	Notice
age	age	47.26	User's age
subjective health	D_1_1	2.81	User's own health status
hypertension diagnosis	DI1_dg	6.50	Hypertension diagnosis result by doctor
dyslipidemia	DI2_dg	6.92	Dyslipidemia diagnosis result by doctor
stroke	DI3_dg	7.85	Stroke diagnosis result by doctor
myocardial infarction	DI5_dg	7.93	Myocardial infarction diagnosis result by doctor
weight	HE_wt	62.32	User recent weight
kidney	HE_ht	162.28	User recent height
waist circumference	HE_wc	80.63	User recent waist circumference
body mass index	HE_BMI	23.55	User's latest body mass index
......	......	......	......

In this article, 196 variables including current disease, doctor's diagnosis, smoking, drinking, etc., which are directly related to health services, are used in the raw data of the National Health and Nutrition Examination Survey. Variable names may vary depending on the device or provider of health data. The National Health and Nutrition Examination Survey uses separate variable codes. Therefore, a matrix containing the description and code is added to find the same variable. Also, the average value of the variables is added for the calculation of CF. This is composed using the Guidelines for using the raw data of the National Health and Nutrition Examination Survey [28]. Table 1 shows the variable information.

In the entire transaction, 10,488 cases selected through data refinement and processing are used. Of these, 7,342 cases, which is 70%, are randomly used for the analysis and the remaining 3,146 are used for experiment and evaluation. Table 2 shows the raw data of the National Health and Nutrition Examination Survey.

Weight Matrix is constructed through the correlation analysis between variables of pre-processed data. The correlation analysis is a method of quantifying the direction and magnitude of correlation, dependency, and similarity inherent between two variables with a correlation coefficient. Equation (3) shows how to quantify the correlation coefficient between two variables. In equation (3), x_n is the n th value of the variable x and $\overline{x}$ is the average value of the variable.

$$\mathrm{Cor}\,(\mathrm{x},\mathrm{y}) = \frac{\sum_{n=1}(x_n-\overline{x})\times(y_n-\overline{y})}{\sqrt{\sum_{n=1}(x_n-\overline{x})^2\times\sqrt{\sum_{n=1}(y_n-\overline{y})^2}}} \tag{3}$$

The correlation coefficient is expressed in the range of -1 to $+1$, indicating that when the correlation coefficient is -1, the direction of the two variables is different and the degree is 1. It indicates that when the correlation coefficient is $+1$, the directions of the two variables are the same and the degree is 1. The closer to 0, the smaller the correlation, dependency, and similarity between the two variables. The result of the correlation analysis on 196 variables showed that 19,503 pairs of variables and weights are stored in the Weight Matrix. Table 3 shows the weight matrix. It is a matrix which is symmetrical based on a diagonal. The matrix uses key1 as the x-axis and key2 as the y-axis to search for weights. When the database is updated, the weight value may change. The result of the correlation analysis showed that 1 is shown when key1 and key2 are the same. For convenience in the operation of the node proposed in this article, it is stored as 0 if key1 and key2 are the same.

TABLE 2. Raw data of national health and nutrition examination survey.

No.	age	D_1_1	DI1_dg	DI2_dg	......	HE_wt	HE_ht	HE_wc	......
1	76	2	1	0	......	59.7	166.7	81.4	......
2	39	2	0	0	......	62.7	171.7	75.7	......
3	35	2	0	0	......	58.3	169.5	77.6	......
4	71	3	0	0	......	53.1	169	68.1	......
5	68	3	1	1	......	61.3	158.1	89.8	......
6	28	2	0	0	......	53.9	156.7	75.9	......
7	45	3	0	0	......	48.8	160.7	66.9	......
8	70	3	1	0	......	53.7	145.3	86	......
9	37	4	0	0	......	55	159.8	71.7	......
10	67	2	0	1	......	73.9	165.7	86.9	......
11	57	3	0	0	......	61.8	158	89.1	......
12	46	3	0	0	......	68.3	175.8	86.3	......
13	33	2	0	0	......	59.3	168.5	73	......
14	41	3	1	0	......	80.1	173.1	89.5	......
15	23	2	0	0	......	64.5	162	74.7	......
16	73	4	1	1	......	52.8	151.3	83	......
17	78	2	1	1	......	56.3	155	77.6	......
......	......	......	......	......	......	......	......	......	......

TABLE 3. Weight matrix.

key1 / key2	age	D_1_1	DI1_pr	DI2_pr	DI3_pr	...	N_DIET	N_DIET_WHY	N_WAT_C
age	0.00	0.30	-0.48	-0.32	-0.15	...	-0.06	-0.08	-0.06
D_1_1	0.30	0.00	-0.23	-0.20	-0.13	...	-0.05	-0.07	-0.04
DI1_pr	-0.48	-0.23	0.00	0.35	0.17	...	0.06	0.08	0.02
DI2_pr	-0.32	-0.20	0.35	0.00	0.10	...	0.10	0.12	-0.00
DI3_pr	-0.15	-0.13	0.17	0.10	0.00	...	0.02	0.03	0.01
DI5_pr	-0.10	-0.08	0.09	0.07	0.08	...	0.02	0.03	-0.01
DI6_pr	-0.15	-0.11	0.13	0.12	0.06	...	0.02	0.03	-0.00
...	...	...	...	...	...	...	...	...	...
N_DIET	-0.06	-0.05	0.06	0.10	0.02	...	0.00	0.91	-0.06
N_DIET_WHY	-0.08	-0.07	0.08	0.12	0.03	...	0.91	0.00	-0.07
N_WAT_C	-0.06	-0.04	0.02	-0.00	0.01	...	-0.06	-0.07	0.00

B. STRUCTURE OF COLLABORATIVE NODE

The Collaborative Node(CN) uses the correlation coefficient as the weight initial value. The correlation coefficient is not a weight learned through a neural network, but it is a weight widely used in data mining and machine learning, and various predictions using it are possible [29], [3], [4]. In the neural network, the operation of a node outputs the result by applying the input value and weight to the activation function. Figure 5 shows node (n_{00}) in collaborative node. Node(n_{00}) outputs $x_{00}\wedge$ using the output values from *00*th to *ij*th of the deep learning model connected to CN.

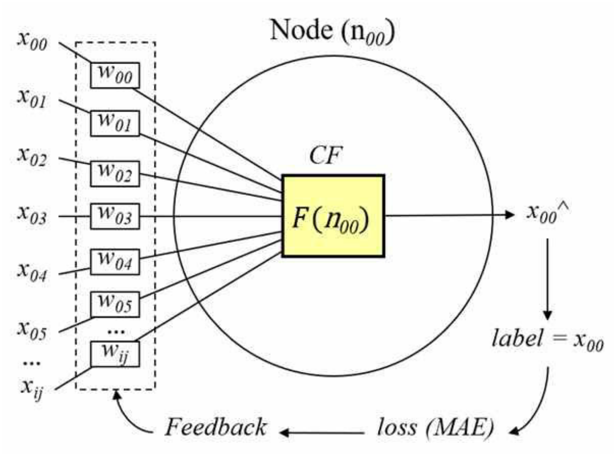

FIGURE 5. Node () in collaborative node.

In Node (n_{00}), w_{01} represents the correlation coefficient of the input variables x_{00} and x_{01}, and is obtained from the weight matrix using the key value searched in the information matrix. The key1 value is determined according to the output of the CN node, and the key2 value is determined by the output of the connected deep learning model. In Node (n_{00}), key1 is $x_{00}\wedge$ and key2 is x_{ij}.

The weight w_{00} represents the correlation coefficient of the input variables x_{00} and x_{00}, and if key1 and key2 are the same, the weight is set to 0 with the same variable. This is to exclude x_{00} from the prediction of $x_{00}\wedge$ because x_{00} and x_{00}^are variables with the same meaning when the output of Node (n_{00}) is x_{00}^. In the case of node (n_{01}), w_{00} indicates the correlation coefficient of x_{01} and x_{00} in node (n_{01}), w_{01} is 0 as the correlation coefficient of x_{01} and x_{01}.

Equation (4) shows the activation function of the proposed node (n_{00}). $x_{00}\wedge$ indicates the output of the node (n_{00}), and w is the weight. b indicates bias. $\overline{x_{ij}}$ is the average value of the variable obtained from the variable information. As correlation coefficients and input values are required for the calculation, prediction is possible even in an initial state in which learning has not progressed. w_{ij} and $\overline{x_{ij}}$ are variables that can be learned in the node calculation. This may have a different value depending on the actual user, and if it is learned with a value close to the actual user value, the accuracy of the node may increase. Therefore, a deep learning model connected by a collaborative node can exhibit higher accuracy when the weight and the variable information matrix are updated to suit actual users.

$$\mathrm{f}\,(n_{00}) = x_{00}^{\wedge} = \overline{x_{00}} + \frac{\sum\limits_{i,j=0} (x_{i,j} - \overline{x_{i,j}}) \times w_{i,j}}{(\sum\limits_{i,j=0} |w_{i,j}|)} \tag{4}$$

Fig. 5 is a node that complements the problem that occurs in general neural networks with machine learning. Since the output between neural network models is concated by applying a cooperative filtering technique with a machine learning algorithm, it is possible to connect models without additional learning.

C. HYBRID MULTI-MODAL DEEP LEARNING USING COLLABORATIVE CONCAT LAYER

Smart Health requires the provision of appropriate services according to the user's situation. The range of available data varies depending on the user's device or surrounding environment. In addition, if there is a change in the user's situation, the health model cannot operate normally.

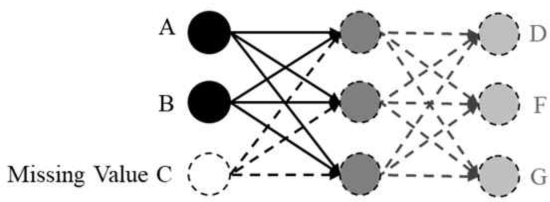

FIGURE 6. Error of the health model according to user situation.

Figure 6 shows the error of the health model according to the change of user situation. When the absence of the input variable C occurs, this affects all nodes under it, thus lowering the accuracy of the model. In order to compensate for this problem, the health model must be configured to be complementary. The proposed Collaborative Node can be used to replace the absent variable. To this end, the output variable (key2) and input variable (key1) of different models are compared and the parameters of the node are brought. The value of an absent variable is calculated through the operation of a node. Also, it should work as it is when the variable is not absent. Algorithm 1 represents the Hudo code that generates the CCL.

Multiple CNs are used to construct a layer that copes with missing values caused by various errors in health data.

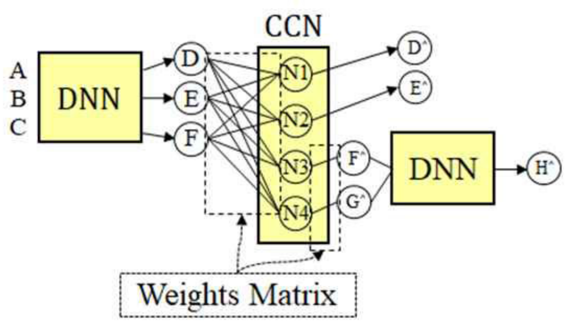

FIGURE 7. Example of mulit-modal model.

Figure 7 shows an example of a multi-modal model using CCL. This is an example of connecting the DNN model predicting the output D, E, and F through the input values A, B, and C and the DNN model predicting H through the

Algorithm 1 Collaborative Concat Layer Generation

Previous Model n, Post Model m
$x'_{n,i}$: Model n's Output i
$y'_{m,i}$: Model m's Input j
k: number of variables
CNi: Collaborative Concat Layer's Node i

Search Output Variables in Model()
Search Input Variables in Model()
Create CNi
CNi input = for(i, 1:Kn) for(j, 1:Km)
initial weight $w_k(x'_{n,i}, y'_{m,i}) = \text{similarity}(x_{n,i}, y_{m,i})$ // from Weights Matrix
parameter $P_{n,i} = \text{average } x_{n,i}$// from variable information
CNi output = for(i, 1:K_n) for(j, 1:K_m)
if($x'_{n,i} == y'_{m,i}$)
    if($x'_{n,i}! = Null)y'_{m,i} = x'_{n,i}$
  else $y'_{m,i} = Prediction$

input values F and G. The CCL searches the output variable and input variable of the DNN connected during the initial creation in the variable matrix, and retrieves the initial value of the weight from the weight matrix. In CCL, the output of CN1 is D', N2 is E', N3 is F', N4 is G', and the output of the rear DNN is H'. G' is a variable that is absent in the output of the potential DNN, and is searched in the weight matrix by using the input variable of the rear DNN connected from N4 as key1. This is a structure in which the value substituted through CN enters the input value of the rear DNN model when there is no value of the variable F or G. The proposed CCL can be used for flexible configuration to connect multiple models.

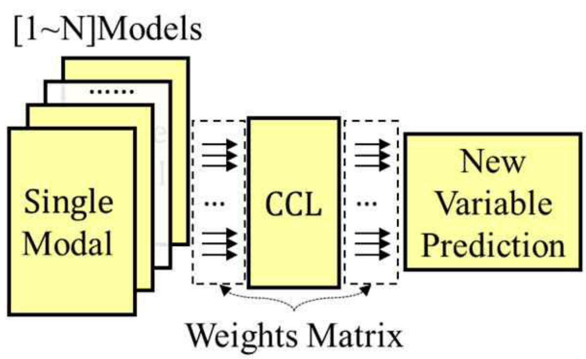

FIGURE 8. N-CCL multi-modal model.

Figure 8 shows the N-CCL-M multi-modal model. This is a structure in which N front single-modal models and m rear single-modal models are connected using CN. The output value of the model on the front and the weights matrix are used to connect in the form of replacing the variable in which the absence occurred among input variables of the model at the rear.

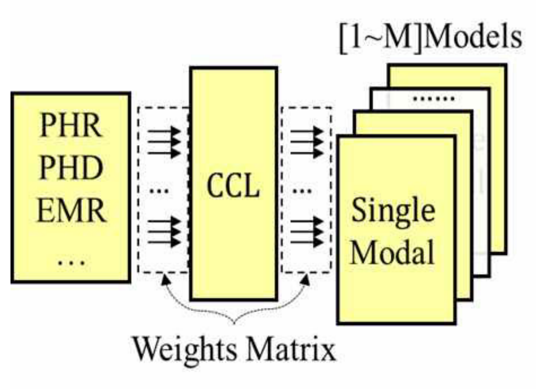

FIGURE 9. CCL-M multi-modal model.

Figure 9 shows the N-CCL Multi-modal Model. In the N-CCL model, the CCL can predict new variables using the weights matrix and output values of a single-modal model on the front. This is possible only for variables present in the weights matrix, but more variables can be predicted if the weights matrix is updated in the future.

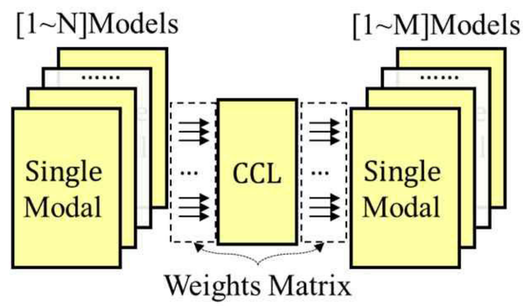

FIGURE 10. N-CCL-M multi-modal model.

Figure 10 shows the CCL-M Multi-modal Model. This is a configuration in which a single-modal model is located at the rear, and if a specific variable is absent depending on the user's situation, it is replaced with the value predicted in the CCL. This allows the use of a single-modal model at the back. Through the proposed N-CCL-M, N-CCL, and CCL-M Multi-modal Model, the characteristics learned in each model can be changed and used according to the changing user's situation. Therefore, a flexible health service is possible through the construction of a flexible model according to the user's situation.

IV. EXPERIMENT ENVIRONMENT

A. DEEP LEARNING MODEL FOR EXPERIMENT

In this article, we propose a method to connect the previously learned deep learning-based health model through the CCL, which is a method that analyzes the characteristics of variables through machine learning technique and uses the relationship appearing in the results of the analysis. As an experiment to evaluate the proposed method, we connect

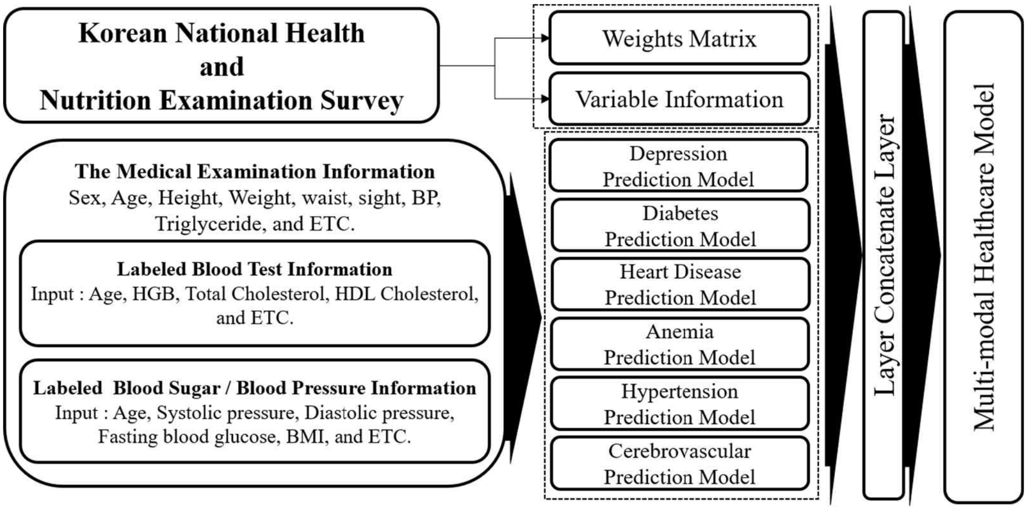

FIGURE 11. Deep learning health model for experiment.

the pre-learned single-modal health model to the CCL to construct a multimodal model and evaluate its accuracy. The experiment concatenates the previously learned deep learning model using a machine learning technique. This is to verify whether meaningful results appear, and we confirm the usefulness of the hybrid multimodal deep learning model that combines deep learning and machine learning techniques. When the constructed single-modal model is connected to the CCL, the experiment is conducted focusing on the degree to which the performance of the existing model is maintained. For the experiment, the DNN model learns to predict disease using health data, and we observe accuracy when applying the proposed method. Figure 11 shows the deep learning health model for the experiment.

For the data for the experiment, the National Health and Nutrition Examination Survey [14], health examination information, national health examination blood pressure blood glucose data [30], and national health examination blood test data are used [31]. Most of the variables are data that can be collected depending on the user's interests or device retention. Health examination information is periodically collected from members of the National Health Insurance. It is a data set that includes variables such as height, weight, vision, blood pressure, blood sugar, and cholesterol [32], [33], [35]. The National Health Examination Blood Pressure Blood Sugar Data is a data set that includes age, blood pressure, pre-meal blood sugar, diabetes and hypertension prevalence. The National Health Examination Blood Test Data is a data set that includes blood test data such as age, hemoglobin, cholesterol, and triglycerides, and variables such as anemia, ischemic heart disease, and prevalence of cerebrovascular diseases.

The blood test data consists of 1 million rows, of which 600,000 are used for model learning, and about 120,000, 20% of them, are used as validation data. The rest is used as test data. The blood test data is used to create three DNN models: a heart disease prediction model, a cerebrovascular prediction model, and an anemia prediction model. The inputs of the three models are the same as gender, age, hemoglobin, cholesterol, triglycerides, and HDL cholesterol.

Heart Disease prediction model consists of Input(6) * dense(64, relu) * dense(64, relu) * dense(64, relu) * Output(1,Sigmoid). The epoch is 10 and the accuracy is 0.725.

Cerebrovascular prediction model is composed of Input(6) * dense(4, relu) * dense(6, relu) * dense(8, relu) * Output(4, sigmoid). When the epoch is 5, the model's accuracy is 0.733.

The anemia prediction model consists of Input(6) * dense(16, relu) * dense(16, relu) * dense(16, relu) * Output(1,Sigmoid). The epoch is 20 and the accuracy is 0.762.

Blood pressure blood glucose data is composed of 1 million rows, of which 600,000 are used for model training, and about 120,000, 20% of them, are used as validation data. Blood pressure blood glucose data are used to create two DNN models: a Diabetes prediction model and a Hypertension prediction model. The inputs for both models are the same: gender, age, systolic blood pressure, diastolic blood pressure, pre-meal blood sugar, and BMI.

The Diabetes prediction model consists of Input(6) * dense(64, relu) * dense(64, relu) * Output(1,Sigmoid). The epoch is 5 and the accuracy is 0.960.

The hypertension prediction model consists of Input(6) * dense(64, relu) * dense(64, relu) * dense(64, relu) * Output(1,Sigmoid). The epoch is 10 and the accuracy is 0.971.

The National Health and Nutrition Examination Survey Data uses 10,488 cases selected through refinement and processing. Of these, 7,342, which is 70%, are randomly used for analysis and learning, and the remaining 3,146 are used for evaluation. This is used to create a depression prediction model. The depression prediction model is a DNN model that predicts depression by entering hypertension, diabetes, smoking, stress diagnosis, and degree of stress awareness. It consists of Input(14) * dense(64, relu) * dense(64, relu) * dense(64, relu) * Output(1,Sigmoid). The epoch is 20 and the accuracy is 0.956.

B. EXPERIMENT RESULTS

The hardware environment for the experiment consists of CPU i5 8400, DDR4 32GB, and GTX1660 Super. The software environment consists of windows10, Python 3.6, and keras 2.3.0. The experiment evaluates the accuracy of the Concatenate method with CCL, which proposes a single-modal health model, and Concatenate method with Fully Connected [34,35]. The DNN model uses CCL and FC to evaluate the performance of the multi-modal health model created by Concatenate in the form of N-CCL-M, N-CCL, CCL-M.

The evaluation is conducted in three types, and Experiment 1 assumes the absence of n input variables and evaluates the change in accuracy when the model is connected in the form of N-CCL-M. Experiment 2 assumes the absence of n input variables and evaluates the change in accuracy when the model is connected in the form of N-CCL. Experiment 3 evaluates the accuracy of the new variable n when the model is connected in the form of CCL-M. For the experimental data, we use 3,146 cases divided by National Health and Nutrition Examination Survey data into evaluation data. In addition, FC requires learning, and only the FC layer of the multi-modal model connected using 7,342 cases used as analysis data is learned separately. Each experiment was repeated 10 times to evaluate the average accuracy.

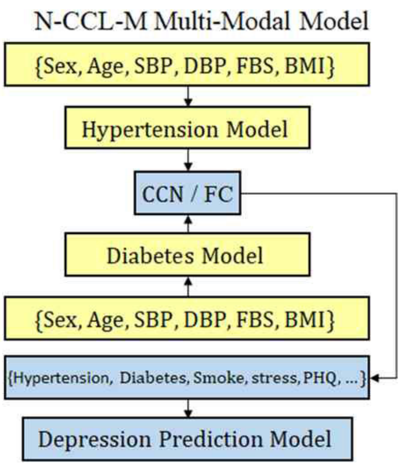

FIGURE 12. N-CCL-M multi-modal health model.

Figure 12 shows a model in the form of N-CCL-M. It is a multimodal health model that predicts depression using the output of the hypertension prediction model and diabetes prediction model, PHQ, and stress. In Figure 12, SBP represents systolic blood pressure, DBP represents diastolic blood pressure, FBS represents fasting blood sugar, and BMI represents body mass index. Experiment 1 assumes the degree of stress awareness and whether smoking among the input variables as missing values, and predicts the missing value using the CCL or FC to compare the accuracy of the result of entering them into the depression prediction model, and the result of entering the output of the heart disease prediction model and the diabetes prediction model into the depression prediction model as they are.

TABLE 4. Result of Experiment 1.

Models	Accuracy			
Variables	All	-stress	-smoke	-stress, -smoke
Diabetes	0.956	0.656	0.711	0.574
MM N-CCL-M	0.933	0.874	0.891	0.859
MM FC0	0.651	0.434	0.472	0.394
MM FC30	0.860	0.518	0.536	0.579
MM FC50	0.942	0.665	0.681	0.596

Table 4 shows the results of Experiment 1. In the variables in Table 4, all indicates that all variables are available, and -stress and -smoke indicate that the variables are absent. MM represents a multi-modal model, while FC0 represents a model with 0 epoch. FC50 represents a model with 50 epochs.

According to the result of Experiment 1, the proposed method maintained the performance of the existing model by more than 89%. In particular, in the absence of both variables, diabetes can be predicted with more stable accuracy. The method of using FC appears similar to the existing model according to the amount of learning. However, in the evaluation using test data, if the input variable is set to NULL by assuming the absence of a variable, the performance of the existing model is significantly lowered. However, by learning features from other variables in the form of multi-modal, it shows slightly higher accuracy than single-modal. The method using FC shows a problem that requires a lot of time and data for optimization depending on the number of nodes or the initial weight value. The proposed method using CCL can be replaced with a meaningful value in a relatively short time.

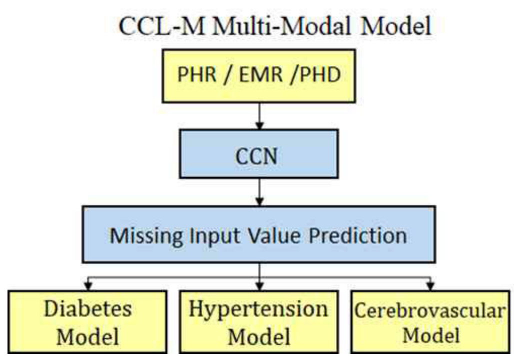

FIGURE 13. CCL-M multi-modal health model.

Figure 13 is a multi-modal model in the form of CCL-M, and it is a multi-modal health model that predicts and replaces

TABLE 5. Result of Experiment 2.

Models	Accuracy					
	Diabetes		Hypertension		Cerebrovascular	
Variables	All	-FBS	All	-BP	All	-BP
Single-modal	0.960	0.711	0.971	0.680	0.733	0.571
MM N-CCL-M	0.960	0.918	0.971	0.908	0.733	0.674

TABLE 6. Result of Experiment 3.

Models	Accuracy				
Variables	Arthritis	Tuberculosis	Asthma	Dyslipidemia	Avg.
MM N-CCL-M	0.914	0.894	0.876	0.921	0.901
MM FC0	0.684	0.657	0.778	0.818	0.734
MM FC30	0.860	0.889	0.852	0.892	0.873
MM FC50	0.936	0.911	0.878	0.957	0.920

missing values occurring in each model at the rear by using the user's context information. Experiment 2 evaluates the accuracy of the model when the predicted value from the user's context information is input as the input values of the Diabetes, Hypertension, and Cerebrovascular models. Table 5 shows the results of Experiment 2. In the variables in Table 5, -FBS represents the absence of pre-meal blood sugar, and -BP represents the absence of systolic and diastolic blood pressure.

In Experiment 3, the single-modal model, like Experiment 1, has a problem that performance is significantly lowered due to the absence of input variables. The proposed method was found to maintain an accuracy of at least 91% by replacing the absent variables. When all variables are available in the CCL, the performance of the existing model can be maintained. Therefore, by adding the proposed CCL, it is possible to increase the resource with the increased parameters of the model, but it can be used to deal with exceptions.

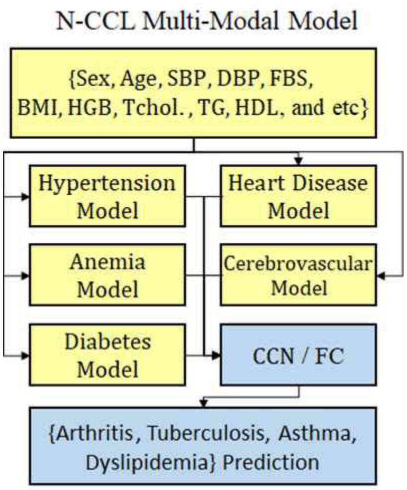

FIGURE 14. N-CCL multi-modal health model.

Figure 14 shows a multimodal model in the form of N-CCL. It is a multimodal health model that predicts arthritis, tuberculosis, asthma, and dyslipidemia using the output of models for predicting hypertension, heart disease, anemia, cerebrovascular disease, and diabetes. In Figure 14, HGB represents hemoglobin, Tchol represents total cholesterol, and HDL represents cholesterol contained in high-density lipoproteins. Other health-related variables that can be collected from health examination information are included. Experiment 3 compares the performance of the method of predicting arthritis, tuberculosis, asthma, and dyslipidemia by the CCL using the output of N models with that of the DNN model predicting arthritis, tuberculosis, asthma, and dyslipidemia using the output of N models. Table 6 shows the results.

In Experiment 3, the prediction for the new variable shows an average accuracy of 0.02 higher than the proposed method using the FC. However, in order to predict a new variable, additional learning is required until a new model is suitable, and there is still a problem that it is difficult to use an existing learned model when an external new variable occurs. The prediction of new variables using the proposed method has the advantages of relatively simple implementation and no need for additional learning.

The result of the experiment showed that the method using CCL proposed by the user-centered health model that changes from time to time can concatenate between models more simply, and is not much different from the method of using FC in terms of performance. Therefore, it is found that reuse and concatenate have the advantages for various health prediction models through the proposed the CCL. Thus, the hybrid multi-modal deep learning that connects the existing deep learning model with the machine learning technique can improve the model to suit the actual user situation or the surrounding situation.

Although the CCL does not exhibit very high accuracy, it has the advantage of being ready to use immediately without further learning or work. In addition, performance can be improved by optimizing the correlation coefficient or the average value between the key parameter variables. The CCL can improve accuracy by updating the parameter weight matrix and variable information. In addition, when a new variable is collected and a correlation coefficient between variables is added, more scalability can be obtained. The average value of each variable in the current variable information is the average by universally collected data, which may vary depending on the actual user. In the future, we plan to conduct research on CCL parameter optimization for users.

V. CONCLUSION

In deep learning, learning efficiency or accuracy varies depending on the variable, learning data, and neural network configuration. Processing all variables in a neural network model has the advantages of simplicity and easy implementation. In particular, deep learning has the advantage that the model extracts and learns features by itself. However, if several input and output variables are included in a single neural network, computational complexity may increase, resulting in reduced efficiency or accuracy in actual use. In addition, in the case of variables indicating a clear relationship, efficient deep learning is generally possible. However, if this is not the case, there is a concern that learning will proceed in a direction different from the intended. Therefore, we propose hybrid multimodal deep learning using a collaborative concat layer in health big data. The proposed method efficiently utilizes overflowing health data. This is a method to flexibly manage the observation range of increasing health data and the deep learning model being developed variously. The CCL is a data-mining-based Concat that uses the initial value of the weight as a correlation coefficient and the activation function as collaborative filtering in deep learning. By setting the correlation coefficient as the initial value, it is possible to perform calculations without additional learning on the model. In addition, by using cooperative filtering, it is possible to handle missing values due to internal or external errors. In the event of the absence of an input variable according to a user situation in a pre-learned health deep learning model, a correlation coefficient-based weight matrix and the CCL can be used. A performance experiment involving the proposed method demonstrated that the loss of accuracy in the general situation is in an allowable range, which can be considered approximately 10%, indicating higher accuracy than in the case of the general model in the absence of specific user variables. Therefore, in the concatenation of a deep learning model using CCL by applying a machine learning technique, more types of variables are used, and a more diverse and flexible health model can be constructed. In addition, in the case of a healthcare model, the accuracy of the model may differ depending on the user situation, and research aimed at increasing the accuracy of the CCL by fitting the average value and similarity weight of each variable, the main parameters of the CCL proposed in the future, to the actual user will be conducted.

REFERENCES

[1] J.-C. Kim and K. Chung, "Depression index service using knowledge based crowdsourcing in smart health," *Wireless Pers. Commun.*, vol. 93, no. 1, pp. 255–268, Mar. 2017.

[2] J.-C. Kim and K. Chung, "Mining health-risk factors using PHR similarity in a hybrid P2P network," *Peer-to-Peer Netw. Appl.*, vol. 11, no. 6, pp. 1278–1287, Nov. 2018.

[3] K. Chung, H. Yoo, and D.-E. Choe, "Ambient context-based modeling for health risk assessment using deep neural network," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 4, pp. 1387–1395, Sep. 2018.

[4] S. Park, B. Cha, K. Chung, and J. Kim, "Mobile IoT device summarizer using P2P Web search engine and inherent characteristic of contents," *Peer-to-Peer Netw. Appl.*, vol. 13, no. 2, pp. 684–693, Jul. 2019.

[5] J.-C. Kim and K. Chung, "Knowledge-based hybrid decision model using neural network for nutrition management," *Inf. Technol. Manage.*, vol. 21, no. 1, pp. 29–39, May 2019.

[6] H. Yoo and K. Chung, "Mining-based lifecare recommendation using peer-to-peer dataset and adaptive decision feedback," *Peer-to-Peer Netw. Appl.*, vol. 11, no. 6, pp. 1309–1320, Nov. 2018.

[7] S. Durga, R. Nag, and E. Daniel, "Survey on machine learning and deep learning algorithms used in Internet of Things (IoT) healthcare," in *Proc. 3rd Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Erode, India, Mar. 2019, pp. 1018–1022.

[8] J.-C. Kim and K. Chung, "Neural-network based adaptive context prediction model for ambient intelligence," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 4, pp. 1451–1458, Aug. 2018.

[9] K. Chung and H. Yoo, "Edge computing health model using P2P-based deep neural networks," *Peer-to-Peer Netw. Appl.*, vol. 13, no. 2, pp. 694–703, Apr. 2019.

[10] J.-C. Kim and K. Chung, "Multi-modal stacked denoising autoencoder for handling missing data in healthcare big data," *IEEE Access*, vol. 8, pp. 104933–104943, May 2020.

[11] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[12] H. Yoo and K. Chung, "Heart rate variability based stress index service model using bio-sensor," *Cluster Comput.*, vol. 21, no. 1, pp. 1139–1149, Mar. 2018.

[13] A. Gumaei, M. M. Hassan, A. Alelaiwi, and H. Alsalman, "A hybrid deep learning model for human activity recognition using multimodal body sensing data," *IEEE Access*, vol. 7, pp. 99152–99160, Aug. 2019.

[14] Korean National Health and Nutrition Examination Survey. *Korea Centers for Disease Control and Prevention*. Accessed: Jan. 3, 2020. [Online]. Available: https://knhanes.cdc.go.kr/

[15] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[16] M. I. Pramanik, R. Y. K. Lau, H. Demirkan, and M. A. K. Azad, "Smart health: Big data enabled health paradigm within smart cities," *Expert Syst. Appl.*, vol. 87, pp. 370–383, Nov. 2017.

[17] K. Chung and H. Jung, "Knowledge-based dynamic cluster model for healthcare management using a convolutional neural network," *Inf. Technol. Manage.*, vol. 21, no. 1, pp. 41–50, Aug. 2019.

[18] J.-C. Kim and K. Chung, "Mining based time-series sleeping pattern analysis for life big-data," *Wireless Pers. Commun.*, vol. 105, no. 2, pp. 475–489, Mar. 2019.

[19] 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.

[20] J.-W. Baek and K. Chung, "Context deep neural network model for predicting depression risk using multiple regression," *IEEE Access*, vol. 8, pp. 18171–18181, Jan. 2020.

[21] S. Patel and H. Patel, "Survey of data mining techniques used in healthcare domain," *Int. J. Inf. Sci. Techn.*, vol. 6, nos. 1–2, pp. 53–60, Mar. 2016.

[22] C. M. Kim, E. J. Hong, K. Chung, and R. C. Park, "Driver facial expression analysis using LFA-CRNN-based feature extraction for health-risk decisions," *Appl. Sci.*, vol. 10, no. 8, pp. 2956–2976, Apr. 2020.

[23] 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.

[24] D. Nie, J. Lu, H. Zhang, E. Adeli, J. Wang, Z. Yu, L. Liu, Q. Wang, J. Wu, and D. Shen, "Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages," *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, Jan. 2019.

[25] L. Li, B. Du, Y. Wang, L. Qin, and H. Tan, "Estimation of missing values in heterogeneous traffic data: Application of multimodal deep learning model," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105592.

[26] W. Wang, B. C. Ooi, X. Yang, D. Zhang, and Y. Zhuang, "Effective multi-modal retrieval based on stacked auto-encoders," *Proc. VLDB Endowment*, vol. 7, no. 8, pp. 649–660, Apr. 2014.

[27] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2391–2402, Sep. 2017.

[28] S. Kweon, Y. Kim, M.-J. Jang, Y. Kim, K. Kim, S. Choi, C. Chun, Y.-H. Khang, and K. Oh, "Data resource profile: The Korea national health and nutrition examination survey (KNHANES)," *Int. J. Epidemiol.*, vol. 43, no. 1, pp. 69–77, Feb. 2014.

[29] K.-Y. Chung, D. Lee, and K. J. Kim, "Categorization for grouping associative items using data mining in item-based collaborative filtering," *Multimedia Tools Appl.*, vol. 71, no. 2, pp. 889–904, Jul. 2014.

[30] Blood Pressure and Blood Glucose Data. *National Health Insurance Sharing Service*. Accessed: Jan. 3, 2020. [Online]. Available: https://nhiss.nhis.or.kr/

[31] Blood Test Data. *National Health Insurance Sharing Service*. Accessed: Jan. 3, 2020. [Online]. Available: https://nhiss.nhis.or.kr/

[32] J.-S. Kang, J.-W. Baek, and K. Chung, "PrefixSpan based pattern mining using time sliding weight from streaming data," *IEEE Access*, vol. 8, pp. 124833–124844, Jul. 2020.

[33] W. Kim and J. Chun, "An improved approach for 3D hand pose estimation based on a single depth image and Haar random forest," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 8, Aug. 2015.

[34] H. Kim, "A SCORM-based e-learning process control model and its modeling system," *KSII Trans. Internet Inf. Syst.*, vol. 5, no. 11, 2011.

[35] S.-E. Ryu, D.-H. Shin, and K. Chung, "Prediction model of dementia risk based on XGBoost using derived variable extraction and hyper parameter optimization," *IEEE Access*, vol. 8, pp. 177708–177720, Sep. 2020.

JOO-CHANG KIM received the B.S. and M.S. degrees from the School of Computer Information Engineering, Sangji University, South Korea, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Kyonggi University, South Korea. He has been a Researcher with the Data Mining Laboratory, Kyonggi University. Since 2020, he has been a Lecturer with the Department of Software Convergence Engineering, Inha University, South Korea. His research interests include data mining, data management, knowledge systems, machine learning, deep learning, big data, healthcare, and recommendation systems.

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 for the Software Technology Leading Department, 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. He was named the 2017 Highly Cited Researcher by Clarivate Analytics. His research interests include data mining, artificial intelligence, healthcare, knowledge systems, HCI, and recommendation systems.