



# Ambient context-based modeling for health risk assessment using deep neural network

Kyungyong Chung<sup>1</sup> · Hyun Yoo<sup>2</sup> · Do-Eun Choe<sup>3</sup> 

Received: 19 May 2018 / Accepted: 3 September 2018 / Published online: 14 September 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Context computing is a branch of ambient intelligence (AmI) research, which has been rapidly emerging in the support of intelligent smart health platform solution. To develop reliable ambient computing using the hybrid peer-to-peer network and Internet of Things, machine learning, deep learning, artificial intelligence, and context awareness have been applied. This study proposes an ambient context-based modeling for a health risk assessment using deep neural network. In the proposed method, we collected medical information from chronic disease patients such as EMR, PHR, and medical histories, as well as environmental data from a health platform. Subsequently, heterogeneous data are integrated through selecting, cleaning, modeling, and evaluating the collected raw data and then the context is created. The structured input data such as a sensor data are normalized by transforming the time domain data to the frequency domain information. Using a deep neural network, the normalized data are applied to create an ambient context. A deep neural network is composed of the following three layers: input layers with treated and untreated data; hidden layers where connection strength is trained as a weight; and output layers of trained results. In the deep neural network layers, the control of the weight of training data enables repeated learning to create an ambient context pattern. Using an ontology inference engine, unstructured/structured data, including individual health data and environmental information, and their context is presented as ontology metadata. In the knowledge base, hidden association relationships are discovered through mining. To inform the individual health conditions exposed to the individual environmental contexts, a health risk assessment model is developed with a set of the ambient context pattern learned with metadata and a deep neural network. The Minkowski distance formula, which defines a normalized geometrical distance between two nodes, is used to measure the similarity between the patients with chronic disease and the individual user based on the context. In the proposed model, the risk is represented as a similarity-based index. The risk assessment model can be implemented into the individual risk alert/prevention system. The model may significantly impact the healthcare industry as well as ambient intelligence research, thus contributing to improve the quality of human life of the future society.

**Keywords** Ambient context · Data mining · Big data · Deep learning · Deep neural network · Healthcare

✉ Do-Eun Choe  
dochoe@pvamu.edu

Kyungyong Chung  
kyungyong.chung@gmail.com

Hyun Yoo  
rhpa0916@gmail.com

<sup>1</sup> Division of Computer Science and Engineering, Kyonggi University, 154-42, Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16227, South Korea

<sup>2</sup> Department of Computer Information Engineering, Sangji University, 83, Sangjidae-gil, Wonju-si, Gangwon-do 26339, South Korea

<sup>3</sup> Department of Civil and Environmental Engineering, Prairie View A&M University, Mail Stop 2510, P.O. Box 519, Prairie View, TX 77446, USA

## 1 Introduction

Ambient intelligence that enables healthcare at any time and any place is critical in a healthcare system. Recently, a smart healthcare platform using a variety of advanced IoT technologies is developed. A customized context awareness system is highly applicable in terms of healthcare and emergent response, which is a base model of caring for chronic disease patients conveniently in a health platform. To ordinary people, chronic diseases mean various illnesses that progress slowly and require a long time for treatment and healing. Medically, chronic diseases mean the illnesses with symptoms that from last 6 months to more than 1 year. Examples include chronic respiratory disease (CRD) caused by

industrialization or tuberculosis as a communicable disease. Currently, with the increasing ageing population, the number of elderly persons with chronic diseases has increased. Accordingly, of the total medical costs, the medical cost of the elderly is increasing. Cancer is also a chronic disease with a high death rate. This chronic disease causes a long-term illness, of which it is often difficult to make an accurate and timely prognosis. In particular, high blood pressure and diabetes have very long illness duration and cause various complications depending on care. Therefore, when a disease incurs, not only does treatment need to be provided in a primary medical institution with high accessibility but also the continued care for the patient afflicted (Kim and Chung 2014). A chronic disease patient's self-care significantly affects the disease progression. Healthcare services using ambient intelligence is a critical medical practice to overcome a chronic disease and lower the medical cost for chronic disease patients. With the increasing demands from chronic disease patients, an advanced context awareness service with high safety and usefulness is highly required.

The data for healthcare service are comprised of context-awareness-based basic data, such as the medical big data offered by hospitals, public institutions, and public health centers, as well as bio big data, and lifelog big data. In the case of lifelog information, personal life experience information can be saved by collecting sensor information and obtaining the recorded data. Regardless of the place and time, daily life information can be unconsciously collected, recorded, and then saved (Chaib et al. 2018; Mashal et al 2016). Chen et al. (2016) proposed a life extension method using ambient intelligence in a wireless sensor network. Mashal et al. (2016) developed an efficient recommendation system for IoT and a graphics-based recommendation algorithm. As such, through a variety of ambient sensing, daily life information, including position, movement, time, place, and bio signal, are used (Chung and Lee 2004). A user's interest, preference, lifestyle, and others are applied to an inference engine and are automatically saved (Adomavicius and Tuzhilin 2015). With the advancement of digital systems in large hospitals and the development of IoT, a massive amount of lifelog big data is collected through wearable devices. In addition, the information offered by the Korean Meteorological Administration and National Health Service becomes more diversified and specified. Therefore, the technology of integrating and processing these heterogeneous big data is a significant factor to create the healthcare service for chronic disease patients. Recently, such a technology has primarily focused on the integration of electronic medical record (EMR) and personal health record (PHR), and personal information offering services based on public health data (PHD) is provided. These heterogeneous big data are to be integrated and processed based on AI and data mining to establish the health data (Rho et al. 2015). Large hospitals

attempt to connect and integrate EMR, PHR, and PHD. In the heterogeneous big data, it is necessary to integrate structured and unstructured data to extract a meaningful knowledge base. With the use of the context awareness system, it is necessary to integrate personally acquired IoT based on bio-log big data and a variety of health records; further, intelligent algorithm-based analysis and customized information need to be conducted (Kim and Chung 2014). For information integration, the collaborative filtering algorithm of mining technology, deep-learning-based deep neural network algorithm, and other algorithms should be combined efficiently based on their advantages. Using the combined hybrid algorithm, the learning model can be improved and an optimized health risk assessment model with a high prediction rate can be provided.

The composition of this study is as follows. In Sect. 2, we describes the big-data based data mining healthcare model, In Sect. 3, we propose an ambient context-based modelling for health risk assessment using a deep neural network. In Sect. 4, we describes the performance evaluation, and Sect. 5 provides a conclusion.

## 2 Big-data based data mining healthcare model

The bio information that can be obtained from a user for health services includes pulse rate, height, weight, ECG, body temperature, body mass index, eyesight, hearing, blood pressure, blood sugar, EMG, and EEG. Bio signal data can be divided into numerical bio signal and time-series bio signal depending on their characteristics. Numerical bio signal data that include height, weight, body temperature, and blood sugar is a discrete value obtained at a particular point (Jung and Chung 2016a, b; Jung et al. 2016; Kim and Chung 2017). For the numerical data, whether one is normal or abnormal in terms of health is judged based on the accurate criteria depending on age, sex, chronic diseases, and others. As for health conditions, the Apriori algorithm of data mining is used to obtain the association rules. The Apriori algorithm creates the association rules in the transactions presented as a set of items. Improved and advanced algorithms include the Apriori TID, Apriori Hybrid, and DHP. Using the decision tree of C4.0 algorithm, potential health risks are analyzed. The logical model of a decision tree contains the "if-then" conditional form. A logical product is conducted from a root node along with the input variables, and depending on the result, a classification is performed. The attributes of the bio signal data are converted to finite discrete values such that the potential health risk can be easily determined. The time-series bio data, which include blood pressure, pulse rate, ECG, and EMG, are continuously obtained values from collecting and recording time-series

data. Whether time-series bio signals are normal are determined by graph waves in each zone, or by the sequential pattern search method using the AprioriAll and AprioriSome algorithms of data mining (Agrawal and Srikant 1994). In terms of the sequential pattern search, sequential association is used for a meaningful relationship to measure a relationship sequence. To create a sequential pattern, it is necessary to measure the time-series bio signals in the time sequence of their occurrence. If the time-series bio signal data are analyzed and meaningful rules are obtained, the potential health risk can be predicted (Chung and Park 2018). Figure 1 illustrates the big-data based data mining process.

The healthcare big data open system of the Health Insurance Review and Assessment Service (HIRA 2018) provides a variety of medical big data. In medical statistics by disease and action, the big data about the diseases whose information are highly required or pertain to social issues are received and preprocessed. Such information can be offered within the range of the health insurance data. The preprocessed medical big data are sent to an NAS file server in a health platform and uses the interaction technology in a P2P networking environment. Data are extracted in the XML format, and information is exchanged with the health platform consistent with its format. In a tree form, objects can communicate with each other through a parser. Regarding the medical big data, interesting disease statistics, statistics of examination and operations, multi-frequency disease statistics, statistics by medical treatment action and type, and drug statistics by disease are applied for the topic modeling of data mining. The result of the topic modeling shows that diabetes had a meaningful association relationship with hyperlipidemia (Song et al. 2017). Regarding the number of topics, the disease classification criteria in medical statistics are applied for the specification. The result shows that the number of topics is limited to 50. The more independent

words are allocated to a topic, the weaker is their clustering. A topic with small allocated words has stronger clustering. Topic modeling uses the similarity in topic distribution to create associated words.

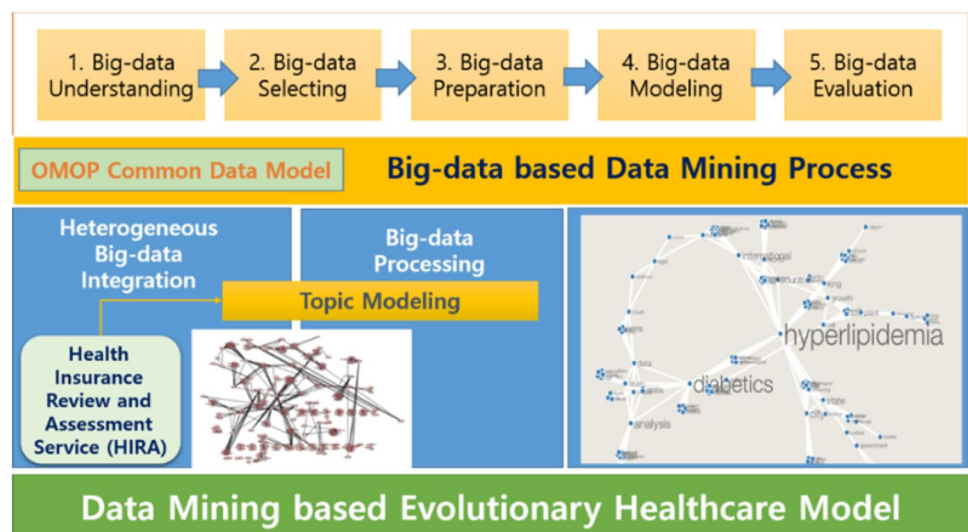
To advance the healthcare industry, structured and unstructured data need to be used to establish a distributed-file-process-based common data model (OMOP CDM) using data mining, text mining, reality mining, and social mining (OHDSI 2018). Further, in the expansion of the multi-omics technology and disease research, they can be used as a CDM-based decision-making support system. Using the clinical database, decision-making can be developed for experts and consumers, and be used at hospitals. Accordingly, a variety of big data, such as medical data, lifelog, and context information, are integrated to develop an evolutionary healthcare model.

### 3 Ambient context-based modeling for health risk assessment using deep neural network

#### 3.1 Context Information collection and preprocessing

To collect a chronic disease patient's health context information, an AMI system should collect each patient's physical data including EMR and PHR, and their environmental information including PHD and open API data (OAD). EMR, which stands for electronic medical record, is digitalized hospital data related to a patient's health conditions and medical history. The EMR also includes a patient's records about hospitalization, discharge, and examination. An individual's health condition and medical history data are the result value of a risk situation. Accordingly, the data can

**Fig. 1** Big-data based data mining process



become a comparative criterion of risk factors. EMRs can be different depending on the medical institutions. Therefore, it is necessary to process the basic data and create a common module for integration. Table 1 presents the EMRs after the data preprocessing. For practical uses, the basic data processing in hospitals accompanies a manual process depending on the database type. This process involves data cleansing, data preprocessing, and data merging, in that order (Chung et al. 2016; Jung and Chung 2016a, b). In data cleansing, the data required for a common data model is extracted from the initial data. It involves the processing of basic abnormal numbers and missing values. For missing values, the default value of the attribute belonging to the same class is used for the accuracy of the prediction. As each data factor has a different meaning, data processing is performed such that the meaning of each parameter value is equal. Finally, data merging is performed to integrate the database of each medical institution. In a health platform, the view table function of the database functions as a multi-step use method (Kim and Chung 2016).

PHR data, as a personal health record, include blood pressure, blood sugar, and weight (Yoo and Chung 2018).

Recently, CCR/CCD-based personal health record (PHR) measured by IoT devices such as a smart band and a smart scale is used. To classify the contexts depending on a user's environment, the PHR analysis algorithm is required. PHR data vary depending on whether a personal measuring device obtains data and on the environment. Although the same measuring device is used, various problems, such as resolution, data omission, and noise, can arise. Hence, a variety of resolutions are applied and an algorithm strongly resistant of missing values is used. The data measured from an ambient sensor can have a frequency form. As a universal algorithm for data normalization with a frequency form, the Fourier transform algorithm is applied. The Fourier transform algorithm can analyze a changing point in the frequency waves in real time such that it the data generated in rapidly changing circumstances can be easily analyzed. If short-time Fourier transform that is improved for the analysis of time–frequency relationship is used, the time-series data measured from an ambient sensor can be analyzed simultaneously. The short-time Fourier transform using a window is presented in formula (1). The start of a window is  $WS$ , and its ending is  $WE$ . The size of width is the whole. In formula (1),  $t$  means

**Table 1** EMRs after the data preprocessing

Item	Style	Attribute
SEX	Category	Gender (male 1/female 2)
AGE	Continuous	Age
HEIGHT	Continuous	Height
WEIGHT	Continuous	Weight
:	:	:
TOT_CHOL	Continuous	Total cholesterol levels
DBP	Continuous	Blood pressure (diastolic)
SBP	Continuous	Blood pressure (systolic)
GGT	Continuous	r-GTP levels
:	:	:
Q_HTN	CHAR	Past hypertension diagnosis/drug prescription 0: Past history(X) Drug(X) 1: Past history(O) Drug(X) 2: Past history(O) Drug(O)
:	:	:
SMK_YN	Category	0: Do not smoke 1: Smoking in the past 2: Smoking now
:	:	:
Q_PA_NEW	Category	Athletic activity Low activity group(1): –600 Met min Medium activity group(2): 600–3000 Met min High activity group(3): 3000 Met min–
Q_FHX_HTDZ	Category	Family history of heart disease 0: No 1: Yes
Q_FHX_STK	Category	Family history of stroke 0: No 1: Yes

time;  $f$  is the frequency;  $i$  is an imaginary number. The width of  $WS$  and  $WE$  is a periodic value.

$$G(f) = \int_{WS}^{WE} f(t)e^{-i2\pi ft} dt \tag{1}$$

For the PHD, the raw data of the Korean National Health and Nutrition Examination Survey generated by the Korea Centers for Disease Control and Prevention (KCDCP 2018) are used. The raw database of the Korean National Health and Nutrition Examination Survey (KCDCP 2015) is used, and the raw data are composed of examination, health, and nutrition surveys. The examination and health survey conducted in a type of expert questionnaire includes a basic survey and an examination of family history, and a separate health questionnaire survey is conducted. A nutrition survey includes the dietary life, food intake and frequency, and food stability. The examination items using a variety of equipment are the thyroid disease test, pulmonary function test, and tuberculosis test (chest X-ray). Oral tests and eye tests are conducted in medical practice. Additionally, color sense tests, ear, nose, and throat tests, bone density/osteoarthritis tests, and muscular tests are conducted (Rho et al. 2016). Because the extracted data include missing and abnormal data, preprocessing is applied to remove such data before use.

The OAD is a user’s ambient information. For the data, the public data offered by the Korean Meteorological Administration are used (KMA 2018). For the OAD, the National Information Society Agency integrates the data of the Korean Meteorological Administration and provides them in the type of open API and dataset (Jung and Chung 2016). Examples include the health weather index and living weather index (Kim and Chung 2016). The health weather index offered includes the asthmatic probability index, lung disease probability index, brain stroke probability index, and et al. More details are presented in Table 2. The risk index of the collected data is presented in 3–5 steps. Therefore, it is used after a 15-step conversion.

### 3.2 Deep neural network model for ambient context awareness

To use ambient context efficiently, the integrated health platform of the preprocessed personal physical data (EMR and PHR), as well as the PHD and OAD of the ambient context information are developed. The first preprocessed basic data include the unstructured data. Accordingly, the already developed unstructured data processing methodology for decision making is used dataset (Jung and Chung 2016). Examples include the health weather index and living weather index (Kim and Chung 2016). In particular, the

**Table 2** Details of health weather index

No.	Response name	Explanation	Use	Info
1	AirpollutionList	Air pollution diffusion index	X	
2	AsthmaLungList	Asthmatic probability index	O	Combined
3		Lung disease probability index	O	
4	BrainList	Brain stroke probability index	O	
:	:	:	:	
16	FsnList	Food-poisoning index	O	
17	InflList	Flu-probability index	O	
18	SkinList	Skin disease probability index	O	
19	UltrvList	UV index	O	
:	:	:	:	

EMR data contain medical keywords of the texts entered by medical staff. The data are relatively limited and explicit, compared to the general-language-based data. Therefore, the ontology establishment technique of an expert system is applied.

The ontology establishment technique is composed of the keyword processing technology for collecting text data, and the search and analysis mining work for extracting the data meaning. The keyword processing technology is used to establish the ontology based on medical keywords with the text data. Mining work uses an inference engine to infer an association relationship from the ontology knowledge base and draws a conclusion (Jung and Chung 2015). The typical association relationship analysis algorithms for extracting association relationships among medical words, diseases, and patterns include the Apriori algorithm and FP-growth algorithm. The Apriori algorithm extracts more fundamental and typical results in the pattern mining research. Even though the FP-growth algorithm contains low operations and a high efficiency, the result is similar. Accordingly, to extract the standard data, the Apriori algorithm is used as an ontology inference engine. In the extracted information, the meaning and association of the attributes are established. From the established information, inference engine rules are created using semantic tags. Based on the created rules, a knowledge base is created and saved. The knowledge base has association relationships between the chronic disease patient data and complications. With its use, the contexts of chronic diseases such as diabetes and cardiovascular disease are designed. The semantic inference engine creates a hidden influence index and adds the reference rules. To update the existing inference rules, an entropy-based feedback is applied to the added data. Hence, the individual risk occurrence can be predicted.

Regarding the PHR, the patterns of the bio data created by an ambient sensor in a user’s wearable device or smart home automation system are analyzed. The distributed heterogeneous data are integrated with a context awareness computing technique. Ambient context awareness computing means the system of recognizing a user’s context by providing customized information service related to a user’s environment. A general context awareness computing system conducts context classification in consideration of the contexts. PHR data have a variety of items and are massively collected as big data. Additionally, the importance of each item is different depending on the chronic disease. Accordingly, the algorithm that evaluates a weight by item and makes a selection is used. A deep neural network is an evolutionary algorithm favorable in processing massive data. Hence, it is necessary to create a learning model with distributed file framework and training data. The created learning model adjusts a weight along with data collection and performs repeated learning. A general deep neural network model is composed of input layers with the data input function, hidden layers of learning connection strength as a weight, and output layers for the data output. Learning can be categorized into supervised learning, e.g., classification or regression, and unsupervised learning, e.g., clustering or association, which are selected depending on a given

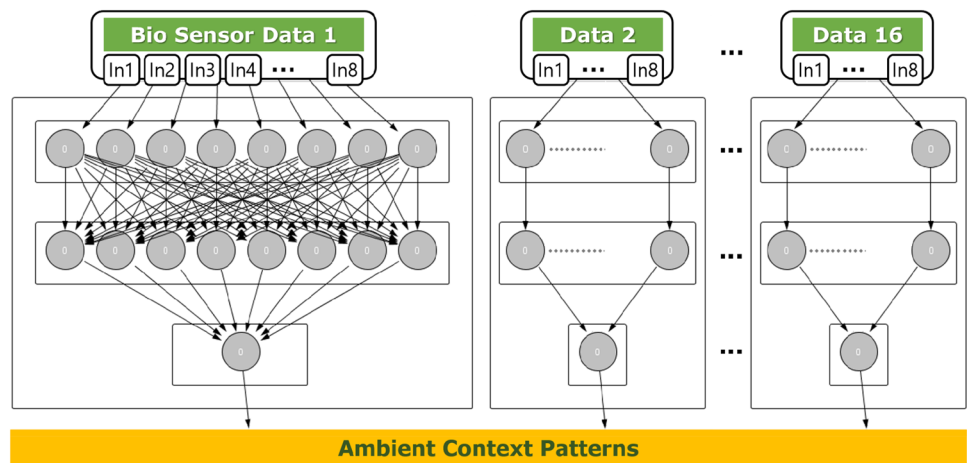
situation whether the corresponding output data exist or not. In terms of the repeated feedback evaluation, reinforcement learning is used. Hence, a deep neural network enables the modeling of nonlinear relationships (Kim and Chung 2018). In this study, based on the frequency input, eight nodes were generated. A hidden layer has hidden nodes with the same form. Figure 2 illustrates the deep neural network model for ambient context awareness.

An output layer has 16 nodes for the created data. The result value of each node is presented in 256 steps. Therefore, the result value ranges from hexadecimal 00 to FF. When the node result values are drawn in order, an ambient context pattern is created.

### 3.3 Data integration and risk factor extraction for health risk prediction

Table 3 shows the integrated data after preprocessing for the analysis of individual context awareness. These data are used as the basic data for context awareness. The processed data have both the metadata created by an ontology inference engine and the result of a deep neural network model. A risk context awareness model should evaluate the health risk from the mixed data in real time. The EMR data received by an expert group become the criterion of the risk context

**Fig. 2** Deep neural network model for ambient context awareness



**Table 3** Integrated data after preprocessing

Ambient context	Detailed context	Input
External context	Temperature, humidity, illumination, noise, position, date, time	OAD
Internal context		
Static body information context	Sex, age, race, chronic disease	EMR PHD
Dynamic body information context	Pulse, blood pressure, body temperature, movement	PHR
Etc	Available resources (battery, display, network) Availability situation (facilities, resources, equipment) Access status (user, permission information, accessibility)	

data. In the course of evaluating the data similarity between a patient and an ordinary person, each data weight is determined. Hence, it is necessary to apply a similarity weight algorithm and the data for presenting an individual’s ambient context. To create the ambient context, the PHR context pattern extracted from the PHR, OAD, and PHD extraction data are combined together. Among the data of an individual’s ambient context, the EMR data in which a chronic disease patient’s disease become a dependent variable becomes a risk reference value. The closer to the common reference value of the chronic disease patients, the higher the risk. To calculate the similarity weight of an ordinary person and a chronic disease patient, the Minkowski distance is applied (Song et al. 2017). The Minkowski distance can be considered as a generalization from the analytical geometry to the Euclidean geometry. For a p-dimension in the Euclidean geometry, the norm *m* is added to the distance *d(x,y)* between the points *x* and *y*. This norm “*m*” means the norm in linear algebra. The Minkowski distance is written as shown in formula (2).

$$d(x, y) = \left( \sum_{i=1}^{norm} |x_i - y_i|^m \right)^{\frac{1}{m}} \tag{2}$$

*x* represents an individual’s ambient context data, and *y* means the already saved ambient context data. The result value *d(x,y)* means the difference in discrete values. The larger the difference is, the more the value increases. Consequently, in the condition of 0 without similarity, the result 1 means the same environment is extracted. If it is converted to a probability, in the condition of 0%, the result of 100% is presented. Based on the selected data, a user can predict its own risk context.

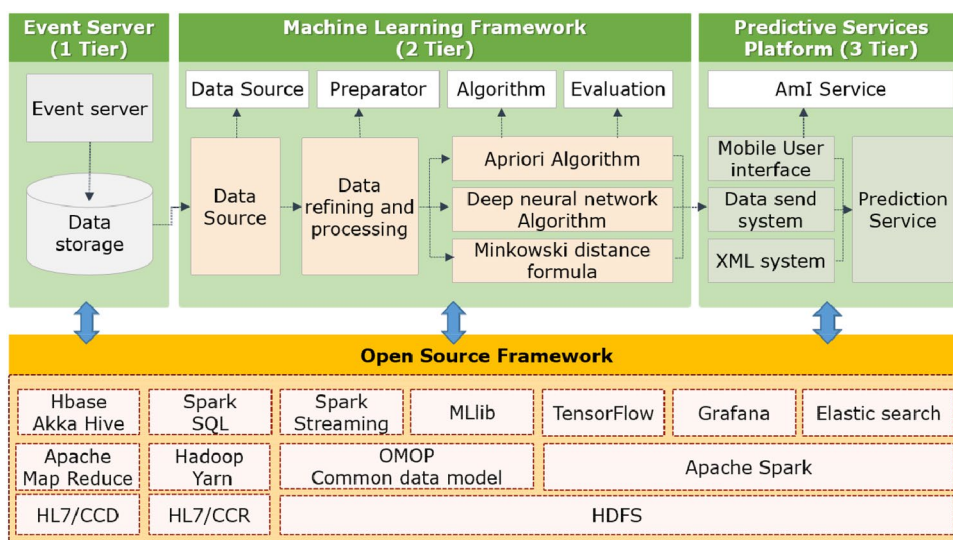
### 4 Performance evaluation

The EMR data are typically composed of a personal health information management module and a clinical information management module. Universally, it uses HL7/CCD and HL7/CCR protocols for transmission outside (HL7 2018). The PHR data use the universal ZigBee such that IEEE 101073 is used as a detailed protocol for connection. Each protocol encrypts data such that they are favorable to security and complement the latest sensitive privacy problem. To collect and save data with safety, a data collection system is required. The collection system contains the extended big data gateway system based on the common data model (CDM) with distributed file processing. In addition, for the data mining of structured and unstructured data, a data warehouse is created. A gateway system collects and processes the distributed data warehouse information with the standardized XML protocol. The system architecture is illustrated in Fig. 3.

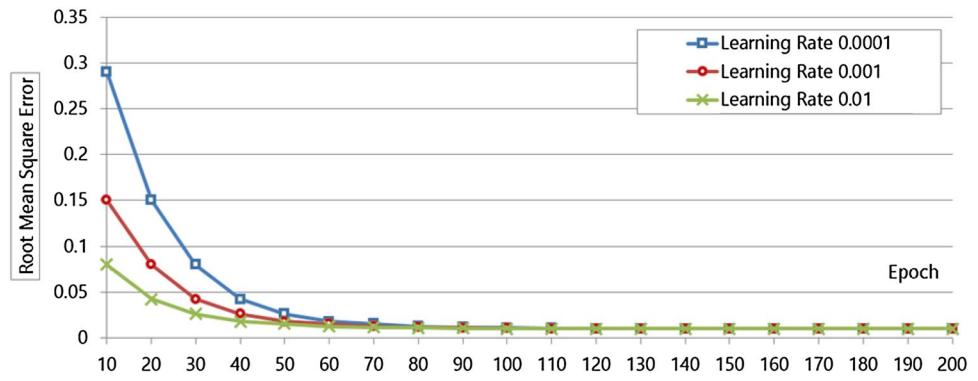
The deep neural network, the key to ambient context pattern operations, has different performances depending on the learning rate. Therefore, before the final system is established, the different performances of the deep neural network depending on the learning rate are evaluated. Figure 4 shows the result of the evaluation depending on the learning rates.

In the condition where the learning rate was 0.01, the results show that learning more than 60 times is effective. In addition, for the comparative evaluation of performance between using and not using a deep neural network model, the root mean square error (RMSE) of the patient data and new data are used. The RMSE is a general method used to evaluate an actual value and a predicted value. Using RMSE, the accuracy of the implemented model can be evaluated. The RMSE can be presented as shown in formula (3), where

Fig. 3 System architecture



**Fig. 4** Result of evaluation depending on learning rates



**Table 4** Result of performance evaluation

Depth neural network	RMSE result (%)
Apply	48.31
None	52.65

$i$  is each of the extraction step,  $n$  is the total extraction count,  $O_i$  is the actual result, and  $P_i$  is the algorithm result value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \tag{3}$$

As the basic data, 4659 records of the raw data offered by the Korean National Health and Nutrition Examination Survey are used (KCDPCP 2015). Among them, the records without answers or errors are excluded through preprocessing. Consequently, 10% of the 3365 records, or 337 data, are modeled. In the case of a new user for the RMSE, the data used in the model is excluded, and 5% of the remaining data, or 168 data are used as random irregular data. The result of the performance evaluation is shown in Table 4. According to the experiment, the health-risk assessment model using the deep neural network improved the accuracy. The predicted risk rate drawn from the extracted ambient context pattern and similarity weight operations are presented in Table 5. With the final offered results (predicted health-risk rate, similarity of a risk group, and an ambient context pattern), a user can actively prevent his/her potential health risk.

### 5 Conclusions

This study proposed an ambient context-based modeling for health risk assessment using a deep neural network. The proposed AmI system is a healthcare-based technology that can effectively provide personalized services to a chronic disease patient. AmI for chronic disease patients can integrate heterogeneous big data, such as medical data (health

**Table 5** Ambient context patterns and predicted health-risk rate

No.	User	Ambient context pattern	Similarity of risk group	Health-risk rate (%)
1	***	ff 0a 27 1a 01 00 0a 0000 71 01 65...	0.23	23
2	***	ff 0a 27 1a 01 00 0a 0000 71 01 65...	0.40	40
3	***	ff 0a 27 1a 01 00 0a 0000 71 01 65...	0.41	41
4	***	ff 0a 27 1a 01 00 0a 0000 71 01 65...	0.03	3
:	:	:	:	:
168	***	ff 0a 27 1a 01 00 0a 0000 71 01 65...	0.11	11

data of hospitals and public institutions), personal health records, and lifelogs, and integrate the personal context such as nutrition, environment, and weather data in the distributed-file-processing-based CDM. Integrating and processing heterogeneous structured and unstructured data from various sources were necessary to extract new meaningful data and potentially use it as the knowledge base for AmI. Therefore, we preprocessed the structured data of time domain data convert to frequency domain data using a modified Fourier transform algorithm. For the unstructured data, an ontology inference engine and the Apriori algorithm of data mining were applied to analyze the association of contexts and infer the hidden context related to health risk. For context awareness, a deep neural network algorithm was used to learn the context. An output layer had 16 nodes, and the result value of each node was presented with 256 steps. In comparison with the EMR disease risk patients, the health-risk-factors-based individual risk ambient contexts were extracted. The client system is realized by mobile apps, which may easily collect useful ambient information from the environments and provide health information with appropriate alerts directly to the users. To evaluate the performance of the developed health risk assessment model, a performance



variation was analyzed upon the learning rate of the deep neural network. Effective application of the algorithm for the integrated data to maximize the advantage of each algorithm would further improve the performance of the suggested system. Hence, this study revealed a great potential of the health risk alert service for chronic disease patients. Although the number of data sets used in this research may not be sufficient to provide a reliable information, the reliability of the prediction is to be achieved as the system developed in this research is used in the field since the number of data will be rapidly growing. Data integration from various sources and development of efficient algorithms can be one of future study needed to improve the current research. In addition, the proposed model can be used as a clinical decision support system for chronic disease prevention and early disease detection for patients undergoing surgical treatment.

**Acknowledgements** This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-0-01405) supervised by the IITP (Institute for Information & communications Technology Promotion).

## References

- Adomavicius G, Tuzhilin A (2015) Context-aware recommender systems, recommender systems handbook. Springer, New York, pp 191–226
- Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of 20th international conference on very large data base, USA, pp 487–499
- Chaib A, Boussebough I, Chaoui A (2018) Adaptive service composition in an ambient environment with a multi-agent system. *J Ambient Intell Hum Comput* 9(2):367–380
- Chen RC, Hsieh CF, Chang WL (2016) Using ambient intelligence to extend network lifetime in wireless sensor networks. *J Ambient Intell Hum Comput* 7(6):777–788
- Chung KY, Lee JH (2004) User preference mining through hybrid collaborative filtering and content-based filtering in recommendation system. *IEICE Trans Infor Syst* E87-D(12):2781–2790
- Chung K, Park RC (2018) Chatbot-based healthcare service with a knowledge base for cloud computing. *Cluster Comput*. <https://doi.org/10.1007/s10586-018-2334-5>
- Chung K, Kim JC, Park RC (2016) Knowledge-based health service considering user convenience using hybrid wi-fi P2P. *Inf Technol Manag* 17(1):67–80
- HIRA (2018) Health insurance review and assessment service (online). <http://opendata.hira.or.kr>. Accessed 16 Apr 2018
- HL7 (2018) Health level seven international (online). <http://www.hl7.org>. Accessed 16 Apr 2018
- Jung H, Chung K (2015) Sequential pattern profiling based bio-detection for smart health service. *Cluster Comput* 18(1):209–219
- Jung H, Chung K (2016a) Knowledge-based dietary nutrition recommendation for obese management. *Inf Technol Manag* 17(1):29–42
- Jung H, Chung K (2016b) Life style improvement mobile service for high risk chronic disease based on PHR platform. *Cluster Comput* 19(2):967–977
- Jung H, Yoo H, Chung K (2016) Associative context mining for ontology-driven hidden knowledge discovery. *Cluster Comput* 19(4):2261–2271
- KCDCP (2015) 6th Korean National Health and Nutrition Examinations Survey (KNHANES VI-1). Korea Centers for Disease Control and Prevention, Cheongju
- KCDCP (2018) Korea Centers for Disease Control & Prevention (online). <http://www.cdc.go.kr/CDC/eng>. Accessed 16 Apr 2018
- Kim JH, Chung K (2014) Ontology-based healthcare context information model to implement ubiquitous environment. *Multimed Tools Appl* 71(2):873–888
- Kim SH, Chung K (2016) Emergency situation monitoring service using context motion tracking of chronic disease patients. *Cluster Comput* 18(2):747–759
- Kim JC, Chung K (2017) Emerging risk forecast system using associative index mining analysis. *Cluster Comput* 20(1):547–558
- Kim JC, Chung K (2018) Mining health-risk factors using PHR similarity in a hybrid P2P network. *Peer-to-Peer Netw Appl* 11(6):1278–1287
- KMA (2018) Korea Meteorological Administration (online). <http://web.kma.go.kr/eng>. Accessed 16 Apr 2018
- Mashal I, Alsaryrah O, Chung TY (2016) Testing and evaluating recommendation algorithms in internet of things. *J Ambient Intell Hum Comput* 7(6):889–900
- OHDSI (2018) Observational Health Data Sciences and Informatics (online). <https://www.ohdsi.org>. Accessed 16 Apr 2018
- Rho MJ, Jan KS, Chung KY, Choi IY (2015) Comparison of knowledge, attitudes, and trust for the use of personal health information in clinical research. *Multimedia Tools Appl* 74(7):2391–2404
- Rho MJ, Kim HS, Chung K, Choi IY (2016) Factors influencing the acceptance of telemedicine for diabetes management. *Cluster Comput* 18(1):321–331
- Song CW, Jung H, Chung K (2017) Development of a medical big-data mining process using topic modeling. *Cluster Comput*. <https://doi.org/10.1007/s10586-017-0942-0>
- Yoo H, Chung K (2018) Mining-based lifecare recommendation using peer-to-peer dataset and adaptive decision feedback. *Peer-to-Peer Netw Appl* 11(6):1309–1320

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.