Improvement of the Personalized Mobile U-Health Service System

Byung-Won Min

Department of Information Communication Engineering,
Mokwon University, Doan-dong 800, Seo-gu, Daejon, 302-729, Korea
minfam@mokwon.ac.kr

Abstract

There are many problems with applying the machine learning technique, which is widely used in the conventional healthcare field, during the mobile u-health service analysis step. First, research on the mobile u-health service is just beginning, and there are very few cases where the existing techniques have been applied in the mobile u-health service environment. Second, since the machine learning technique requires a long learning period, it is not suitable for application in the mobile u-health service environment, which requires real-time disease management. Third, the various machine learning techniques that have been proposed until now do not include a way to assign the weight factors to the disease-related variables, and thus its use as a personalized disease prediction system is somewhat limited.

This paper proposes PCADP, which is an ontology-based personalized disease prediction method, to solve such problems and to interpret the bio data analysis of the mobile u-health service system as a process. Moreover, the mobile u-health service ontology framework was modeled as a semantics type in order to meaningfully express the mobile u-health data and service statement based on PCADP.

To validate the performance and efficiency of the PCADP technique proposed in this paper, the 5-cross validation method was used to measure the accuracy of the prediction. The validation of PCADP using a virtual disease group verified that the technique proposed in this paper shows much greater accuracy compared to existing methods. Moreover, the PCADP prediction method improved the flexibility and real-time attributes, which are the essential elements of any diagnosis technique in the mobile u-health environment, and showed efficiency in the continuous improvement of the monitoring and system of the diagnosis process.

Keywords: *Mobile U-Health Service System; Personalized; Ontology; PCADP; 5-cross validation*

1. Introduction

As health has become the key value of a society, the demand for personalized medical services based on specialization and diversification is increasing. Moving beyond the face-to-face oriented service, new services such as preventive medicine and post-treatment management as well as customized medical services are in greater demand. Furthermore, with the rising interest in health resulting from the increase in diseases caused by changes in eating habits and stress, patients are demanding medical services from the position of consumers, and the hospital information system is transforming the medical service paradigm from a costly medical agency-centric one to a routine home health and specialized medical service-centric one. Therefore, the demand for u-health services is expected to greatly increase in the

future as there will be a strong demand for medical checkups and medical services anywhere and anytime without having to visit a medical agency [1-6].

The service generally consists of four stages: the sensing stage, in which the health customer's bio signals are collected using sensors; the monitoring stage, in which the bio signals undergo preliminarily processing; the analysis stage, in which a new health index is generated from machine learning, such as pattern recognition, and data mining of the obtained data; and the feedback stage, in which the customer is notified of his/her health status. In particular, the analysis stage, whereby a new health index is calculated from the accumulated bio signals, may be considered as the key to and the (essential) platform of, the mobile u-health service [7-9].

Many people believe that applying the conventional machine learning method at the analysis stage will yield effective results. First, research on the mobile u-health service is just beginning, and there have been very cases of applying existing techniques to the mobile u-health service environment. However, the fact that it requires a long learning time and that it is difficult to back-track the diagnosis result makes it difficult to be applied in the analysis stage as the mobile u-health service must be able to monitor the disease anywhere and anytime. Although there are many machine learning methods, there is not much difference in their ability to predict and diagnose the health condition of individuals using the accumulated data. As such, the key interest is reported to be how to improve learning speed through the selection of numerous variables [10-15].

This paper proposes a PCADP (Personalized Computer Aided Diagnosis Probability) method suitable for the personalized mobile u-health service environment. This statistical disease diagnosis method features a flexible structure, real-time processing, continuous enhancement, and monitoring of the diagnosis process, which are the main criteria of a diagnosis method in the mobile u-health service environment. The technique uses the IA (Interrelation Appearance) matrix and CI (combination interrelation) matrix to develop an extended diagnosis system that ensures the reliability of the data and application service, which has been identified as a source of the problems with existing systems, and which can diagnose even general diseases such as stress, obesity, and diabetes.

This paper also proposes a mobile u-health service platform that is equipped with the key functions and tools needed by mobile u-health service developers, and which supports the systematic development of mobile u-health services. The proposed service platform is significant in that the new mobile u-health services are developed and advanced in the platform. The proposed platform enables the registration and reuse of the developed common features and mobile u-health services. The main items of u-health services supported by the platform include the acquisition of bio data, management, analysis, knowledge deduction, and decision-making support. The key components of the proposed platform architecture include mobile u-health progress management, user management, mobile message handling, huge temporary database management, security management, data mining, and pattern recognition modules.

This paper is organized as follows: Chapter 2 describes the personalized disease prediction technique for calculating a new health index using the user's accumulated bio signal data; Chapter 3 describes the bio data acquisition, management, analysis, knowledge deduction, and decision-making process in the mobile u-health service platform; Chapter 4 describes the validation test to present the potential of the proposed disease prediction method in detail, and uses the test result to compare it with conventional disease diagnosis methods and derive the implications; and, lastly, Chapter 5 presents the conclusion and a future study plan.

2. Mobile U-health Service Personalized Disease Prediction Method

2.1. Disease Diagnosis Algorithm Architecture

The personalized disease prediction algorithm stores the interrelationship pattern among the bio signals and symptoms of the normal group and the patient group, and then derives the probability equation using the bio signal and symptoms of the subject to diagnose the presence or not of a disease in the subject. Figure 1 shows the overall architecture of the disease diagnosis algorithm.

In the learning stage, the interrelation pattern among the bio signals and symptoms of the patient group and the normal group are saved. In the decision tree stage, the <bio signals set, symptoms set> pairs are categorized and combined according to the disease type in order to create the root tree connected by the node and edge, using the data accumulated in the learning stage. Then, the important decision tree (DT) value is obtained from the data using the allelic association and the Hardy-Weinberg equilibrium. In the prediction stage, the probability equation - using the bio signal and symptoms of the subject - is devised to diagnose the extent of disease of the subject.

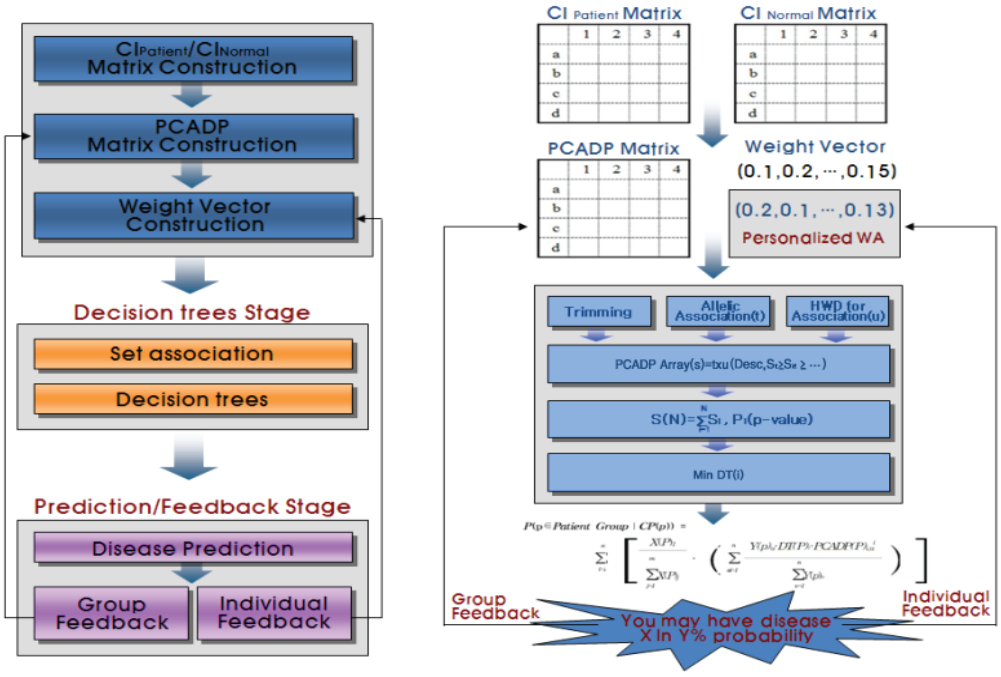


Figure 1. Structure of Personalized Diagnosis Method

The learning stage consists of three detailed processes. In the first process, a CI matrix to record the appearance frequency of <bio signals set, symptoms set> of the patient group and the normal group is generated. The CI matrix handles not only the bio signals and symptoms but also the bio signals sets and symptoms sets. In the second stage, the probability of the <bio signals set, symptoms set> pairs appearing in the patient group is calculated to generate the PCADP matrix. In the third process, the WA (Weight Assignment) matrix is generated for each user in order to provide a personalized service specific to each individual. In the decision tree stage, the DT value is obtained using the HWD statistics, which constitute the χ^2-

statistical value representing the level distribution of the <bio signals set, symptoms set> pair with the AA statistics, which constitute the χ^2-statistical value measuring the relationship to the disease, and the Hardy-Weinberg equilibrium as the null hypothesis. In the prediction stage, the disease status of the subject is diagnosed with a certain <bio signals set, symptoms set> pair based on the PCADP matrix and the weight vector generated in the learning stage. Lastly, in the feedback stage, the system is compensated using the user's response to the result upon completion of the disease diagnosis. It improves the accuracy of the disease diagnosis framework and enables a precise personalized diagnosis by separating the individual user feedback mechanism from the group user mechanism.

2.2. Learning Stage

The learning stage is the step in which the PCADP matrix and the weight vector to be used in the prediction stage are generated. In this paper, the weight vector is assumed to be determined with the help of specialists in a particular disease. This section describes the process of generating the PCADP matrix, which contains the probability data of the <bio signals set, symptoms set> pair appearing in a group having a certain disease. To obtain a PCADP matrix, the $CI_{Patient}$ and CI_{Normal} matrices containing the appearance frequency of the <bio signals set, symptoms set> of the patient group and the normal group, respectively, must first be generated.

2.2.1. IA Matrix

To facilitate understanding, the IA matrix, which contains the appearance frequency of the <bio signal, symptom> pair, will be introduced before describing the CI matrix used for the appearance frequency of the <bio signals set, symptoms set> pair in a certain group. Each row of an IA matrix represents a bio signal, and each column of an IA matrix represents a symptom. Therefore, each element of an IA matrix represents the appearance frequency of the <bio signal, symptom> pair.

The CI matrix expands the IA matrix to handle not only the bio signals and symptoms but also the bio signals sets and symptoms sets. Therefore, the CI matrix records the appearance frequency of the <bio signals set, symptoms set> pair rather than the <bio signal, symptom> pair.

2.2.2. CI Matrix

When the CI $CI_{Patient}$ and CI_{Normal} matrices, representing the CI matrix of the patient group and the normal group, respectively, are obtained, the probability of the <bio signals set, symptoms set> pairs appearing in the patient group is calculated, and the PCADP matrix having the same structure as the CI matrix can be generated. In this paper the normalization function is applied in the equation to determine the element values to compensate for the learned data of the patient group and the normal group. Assuming $Total_{Patient}$ and $Total_{Nomal}$ to represent the total number of <bio signals set, symptoms set> pairs accumulated in the $CI_{Patient}$ matrix and the CI_{Normal} matrix, respectively, each element of the PCADP matrix can be expressed by Equation (2.1).

$$PCADP_{ij} = \frac{Total_{Nomal} \cdot (CI_{Patient})_{ij}}{Total_{Nomal} \cdot (CI_{Patient})_{ij} + Total_{Nomal} (CI_{Patient})_{ij}} \quad (2.1)$$

After obtaining the $PCADP_{ij}$ probability values, the probability of a certain <bio signals set, symptoms set> pair appearing in a patient group can be calculated using the PCADP matrix. However, diagnosing the disease using only the PCADP matrix will mean that all bio signals and symptoms must have the same weight factors. Since the PCADP matrix itself does not have any weight factors, it cannot reflect the fact that some bio signals and symptoms have a greater impact than others on the disease. Knowing that certain bio signals or symptoms generally have a relatively bigger impact than others on a certain disease, it is not appropriate to diagnose the disease only with the PCADP matrix. Therefore, the third stage of learning introduces the concept of a weight vector relating to the relative impact of the bio signals and symptoms on a certain disease to compensate for the limitations of the PCADP matrix.

2.2.3. Weight Vector

As explained above, some bio signals and symptoms may have a higher weight factor than others for prediction. Since the PCADP matrix does not support the weight factor assignment, this paper introduces the weight vector to reflect the relative impact of the bio signals and symptoms when predicting a disease. The bio signal weight vector and the symptom weight vector are assumed to be {wb1, wb2, … , wbn} and {ws1, ws2, …, wsm}, respectively. The initial values of the weight vector will be determined with help from experts on the disease. In the future, the optimum weight vector may be obtained through additional experiments. This paper does not exclusively describe the weight factors of the bio signals sets and symptoms sets. The weight factors of the bio signals sets and symptoms sets can be easily obtained through average calculation and normalization using the bio signals weight vector and symptoms weight vector.

2.3. Decision Tree Stage

The personalized disease prediction method used in this study identifies a certain pattern in complex data and makes a prediction statistically. However, if there is a large volume of data input, it takes a long time to calculate and hence becomes more difficult to find the pattern. To solve this problem, set association and the decision tree are used to reduce the amount of data input into the disease prediction system so that it can handle the large data sets and increase the accuracy of prediction.

First, the PCADP matrix values are selected using set association and the decision tree from those generated in the learning stage, and they are given as the input values of the disease prediction probability equation. The time required for prediction is reduced even for a large volume of data to improve the accuracy and efficiency of disease prediction. Set association is a method of finding the important decision tree value using the allelic association and Hardy-Weinberg equilibrium data. To that end, HWD statistics, which constitute the χ^2-statistical value representing the level distribution of the <bio signal set, symptom set> pair with the AA statistics, which constitute the χ^2-statistical value measuring the relationship to the disease, and the Hardy-Weinberg equilibrium as the null hypothesis, are needed. It means that the DT values with high HWD deviation in the patient group are particularly related to a certain disease. On the other hand, the DT values having high HWD deviation in the normal group means an error has occurred. Therefore, a trimming process for removing or zeroing those DT values is needed.

2.4. Prediction Stage

The disease can be assessed after generating the PCADP matrix and weight vector in the learning stage. The prediction stage uses a probability equation to assess the disease of a certain <bio signals set, symptoms set> pair based on the PCADP matrix, weight value and DT values generated in the leaning and decision tree stages. The value calculated by the devised probability equation represents the probability of a subject being in the disease group. In this paper, the probability value is assumed to be the probability of a subject having the disease in question.

To explain the disease probability equation, it is assumed that the PCADP matrix, the bio signals weight vector X, and the symptoms weight vector Y have already been generated. And CP(p) is defined as a set of <bio signals set, symptoms set> pairs of a subject p in whom the presence of a certain disease has not yet been determined. The PCADP(p) matrix is a rearrangement of CP(p) in PCADP matrix form. The bio signal weight vector and symptom weight vector in the PCADP(p) matrix are expressed as X(p) an Y(p), respectively. Assuming the finally obtained matrix to be m X PCADP(p), the m-length bio signal weight vector to be X(p), and the n-length symptom weight vector to be Y(p), the disease diagnosis of a subject having CP(p) is calculated with Equation (2.2).

$$P(\mathbf{p} \in Patient\ Group \mid CP(p)) = \sum_{i=1}^{m} \left[\frac{X(P)_i}{\sum_{j=1}^{m} X(P)_j} \cdot \left(\sum_{u=1}^{n} \frac{Y(p)_u \cdot DT(p)_i \cdot PCADP(p)_{i,u}^{i}}{\sum_{v=1}^{n} \mathbf{Y(p)}_v} \right) \right] \quad (2.2)$$

2.5 Feedback Stage

It cannot be said for certain that the diagnosis of diseases using the above procedure will always be correct. Indeed, it is difficult to assure its reliability, particularly given that the result is predicted with the limited bio signals and symptoms in the ubiquitous environment. This section describes the user feedback mechanism to supplement the weakness of the disease diagnosis framework in the ubiquitous environment and to improve the accuracy of the disease diagnosis framework. The user feedback stage is the process of compensating the system using the user's response to the predicted result upon completion of the disease diagnosis. It is the last stage of the disease diagnosis framework. In this paper, the user feedback mechanism is mainly divided into the individual user feedback mechanism and the group user feedback mechanism.

2.5.1. Individual User Feedback Mechanism

The individual user feedback mechanism was devised to reflect types of disease, such as stress, that show large deviations from person to person in the disease diagnosis framework. A user inputs the bio signals and symptoms into the disease diagnosis framework, receives the disease assessment result, and provides the feedback information of the result. The individual user feedback process is depicted in Figure 2. If the user agrees to the assessment, no feedback mechanism is operated and the information is maintained as a record for the group user feedback mechanism described below. If the user does not agree with the disease

assessment, he or she can adjust the weight factor of a bio signal or symptom that is particularly more significant to him or her.

The disease diagnosis framework generates the individualized bio signal weight vector and symptom weigh vector according to the weight factor adjusted by the user and immediately delivers a modified disease assessment according to the changed weight factor. The user checks how much impact the changed weight factor has on his/her disease assessment and selects the suitable weight vector value. Depending on the individual, such a change of weight vector value can be completed once or repeated several times. The individual user feedback mechanism is completed when the disease assessment system determines the disease based on the changed individual weight vector for the next disease assessment of the same user, as shown in Figure 2.

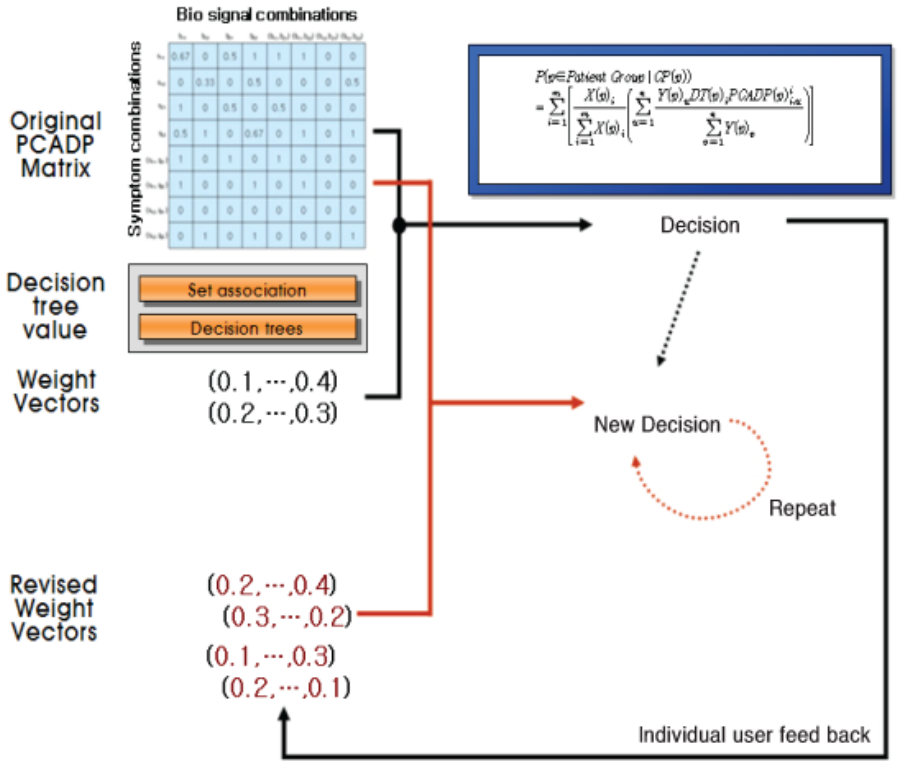


Figure 2. Individual Feedback

2.5.2. Group User Feedback Mechanism

Some data obtained from the user feedback mechanism can also be used by the new learning group to assess the disease. The group user feedback mechanism is a series of processes which advance the existing system by utilizing the newly obtained learning group data. It differs from the individual user feedback mechanism, which affects the assessment of the disease of the user but not of others, in that the feedback affects the assessment of all users. In the group user feedback mechanism, when it is confirmed that the newly obtained learning group data are more refined than the learning group data used in the existing disease assessment system, it is reflected in the PCAPDP matrix of the existing diagnosis framework to obtain a more advanced PCADP matrix. The existing disease diagnosis framework will

evolve into a better system by using the more advanced PCADP matrix. Figure 3 shows the concept of the group user feedback mechanism.

There are two issues concerning the group user feedback mechanism: the first concerns how to separate the normal group and the disease group in the data obtained through the feedback of the predicted result using the existing PCADP matrix and how to form the learning group; the other concerns how to measure the error of the existing PCADP matrix and the PCADP matrix obtained through the newly generated learning group and how to generate the disease group to be used for error measurement.

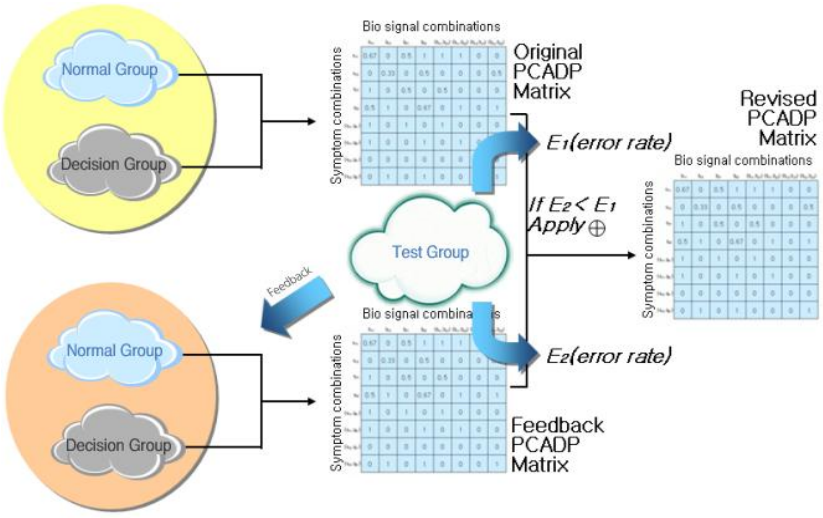


Figure 3. Group Feedback

3. Mobile U-Health Service System

3.1. Definition

Although definitions of the mobile u-health service may vary according to their points of view, we define it as a real-time service that can be obtained from a mobile terminal while we are moving. In other words, we can obtain, store, manage, and analyze mobile bio data for application to the corresponding user in order either to perform proper follow-up at the proper time or to directly cure the disease as an advanced service. In addition, the mobile u-health system is defined as an integrated scheme encompassing the bio sensors, terminal, and related software and hardware, which are needed to provide the mobile u-health service as defined above.

In this opening section of the paper, we treat the essential elements of the mobile u-health service or system in view of the service framework or platform.

3.2. Elements of the Mobile U-health Service

The mobile u-health system generally consists of the following core elements and their corresponding technologies, although there are other kinds of systems in our initiative u-health world:

- Framework for the collection of bio data

- Framework for the storage and management of bio data
- Framework for the analysis of bio data
- Framework for the mobile u-health service

Based on these core features, the structure of our mobile u-health service is represented as the diagram shown in Figure 4.

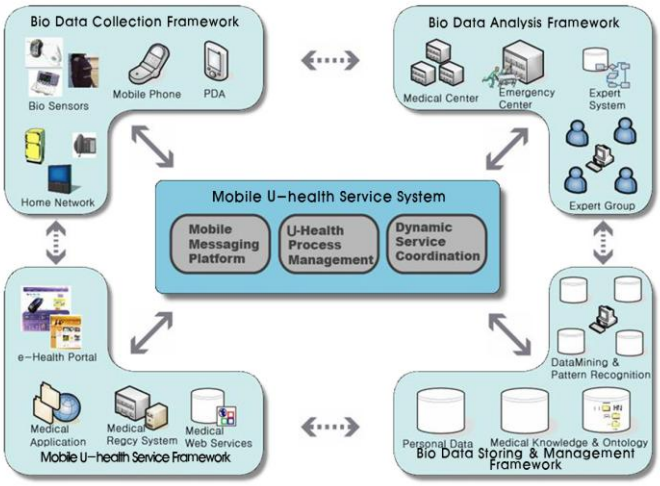


Figure 4. Structure of the Mobile U-health Service

The system periodically obtains the user's bio data and transfers them to the server using the framework for collection. The sensing output from the independent sensor installed in the terminal is transferred to the server through the terminal gateway. Although a method of direct transfer from sensor to server is possible, it would not be practical because of the expense of the sensor and its capability limit.

We can effectively store and manage the collected bio data using the framework for storage and management. In this scheme, we use temporal data management skills for the framework because the bio data are generated periodically and continuously.

The framework for analysis determines whether there are any abnormal symptoms in the user's body by the prepared analysis method applying to the new bio data obtained from the frameworks described above. In order to analyze the bio data, we apply data mining technology to detect possible abnormalities or index of health. We use a pattern-matching method, expert system concepts, and a method of supporting decision making.

Finally, the framework for the mobile u-health service is a kind of middleware for supporting the integrated service, including data collection, storage and management, and their analyses. Therefore, the elements of the u-health service described above and the various corresponding technologies are integrated into a Hub, which is named the mobile u-health service framework. In addition, the framework offers the environment for developing services and making the services to be operated on the framework.

Figure 4 shows a typical model of the mobile u-health, which consists of the core element frameworks mentioned above. The real-time bio data of a user obtained from mobile utilities are stored in the database and analyzed. Most of the time, analyses of bio data are fulfilled by the help of an expert system which collects bio data into two parts between the normal group

and the patient group, in order to enhance the accuracy of a decision. Moreover, the expert system advances continuously by studying new data as its proper learned data.

3.3. Mobile U-health Service Platform

In this section, we present the architecture of the mobile u-health service platform, including its core elements and their roles. We explain the capabilities of the platform in connection with applications, the expert system, and the related databases. Figure 5 shows the structure of the mobile u-health service platform offered from this point of view.

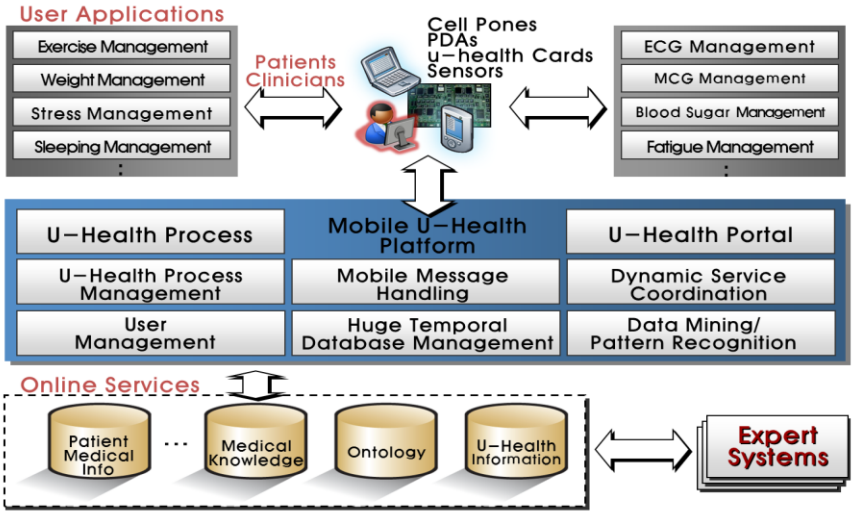


Figure 5. Mobile U-health Service Platform

The platform receives bio data as types of messages from various terminals and hands them over to the database management module for processing. In other words, the mobile message processing module connects between the moving client and the server. The bio data transferred through the mobile message processing module to the framework will be stored and managed by a large-scale temporal database management system, in which the bio data may be separated according to their users, services and, sometimes, their types of treatments.

Stored data will be used for detecting the necessary health indices by applying a data-mining or pattern-matching methodology, and then offering direct or feedback information to the expert system. Close relations between the expert system with the data mining or pattern matching module and the temporal database management module are necessary because the structure of the database varies depending on the kinds of application services.

On the other hand, the corresponding u-health application service has to be defined by a process format in order to develop the application service using the mobile platform. As shown in Figure 5, all the mobile u-health services are to be considered possible as processes for obtaining, storing, analyzing, and informing the result. In addition, the mobile u-health application services represented by the corresponding processes will be operated and controlled by the process management system, in which operational services and their monitoring stages for control will also be supported.

The user management module supports personalized service control to manage all the personal information. This module can be used in connection with the user management scheme installed in the process management system

3.4. Mobile U-health Service Scenario

Figure 6 shows a typical scenario of the mobile u-health service. In the first step of the service scenario, a user fills up the questionnaire and receives the information related to the physical symptoms and service environment, which cannot be obtained from the bio sensors, of the user. The information obtained from the questionnaire can be crucial regarding the accuracy of the disease diagnosis and provision of the proper service. After the information has been obtained from the questionnaire, the sensors attached to the mobile phone acquire the required bio signals and send them to the u-health server. In the sensing process, the graph of the bio signals, such as the pulse, can be processed and displayed on a mobile phone without having to go through the server.

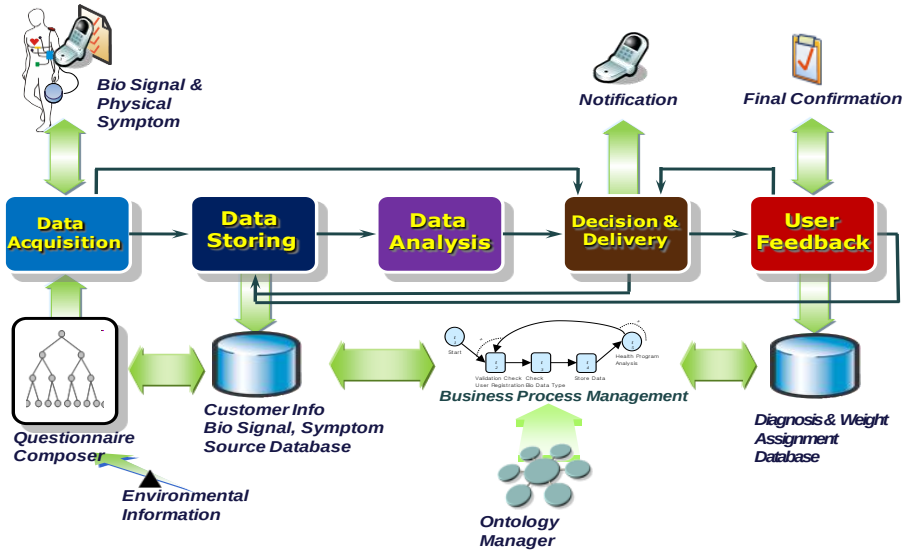


Figure 6. Mobile U-health Service Scenario

3.4.1. Bio Data Storing and Management

The framework for storing and managing bio data is a central element in the storage and management of the collected data using the sensor and the data collection framework, as shown in Figure 7. It processes u-health data, user-related data, and service specifications. In addition, we analyze disease-related issues, symptoms, and their relationships using a semantics representation model based on u-health data ontology throughout the framework. As such, we can offer a user-friendly environment for the development of various u-health services and contents in order to meet the modern requirements of users.

We can make professional service developers store their own services for various u-health applications using the ontology editor, as shown in **Fig.6**. In addition, applications developers can find the most proper service for their own u-health applications by using the service broker of the ontology manager prepared under this scheme. Throughout these processes, we can offer personalized u-health services and a reliable development environment for further or additional application of the advancing scheme.

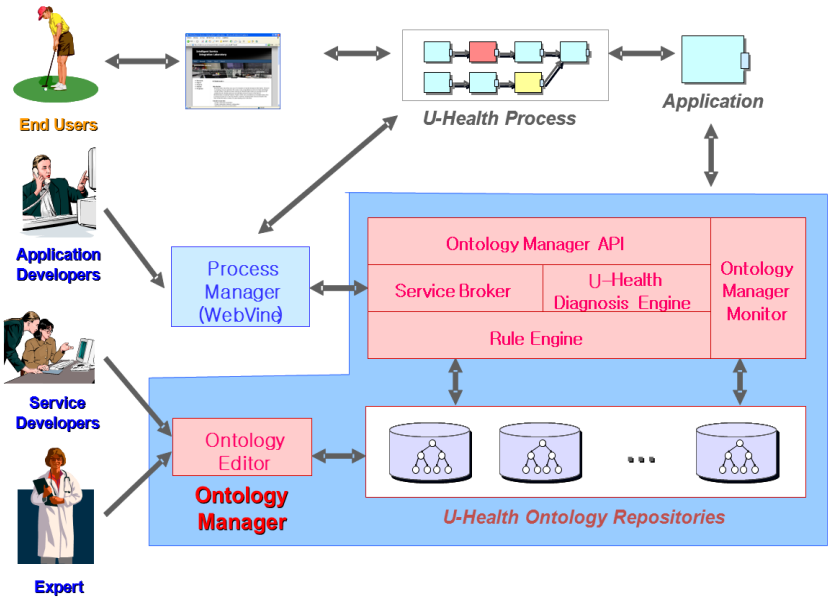


Figure 7. Bio Data Storing and Management Framework

3.4.2. Bio Data Analysis Framework

The framework for the analysis of bio data is shown in Figure 8. We can decide feces or urine about a patient by analyzing newly created data throughout the method prepared previously. As shown in Figure 5, service developers and application developers can perform their developing jobs development tasks on the link provided by the service broker of the scheme. After service developers generate service units and load them, application developers can construct corresponding service processes using the recommended service units throughout the service broker. We can realize a large process using stored process elements through the process template, while we directly realizing any sub-process by bringing service units from the service pool.

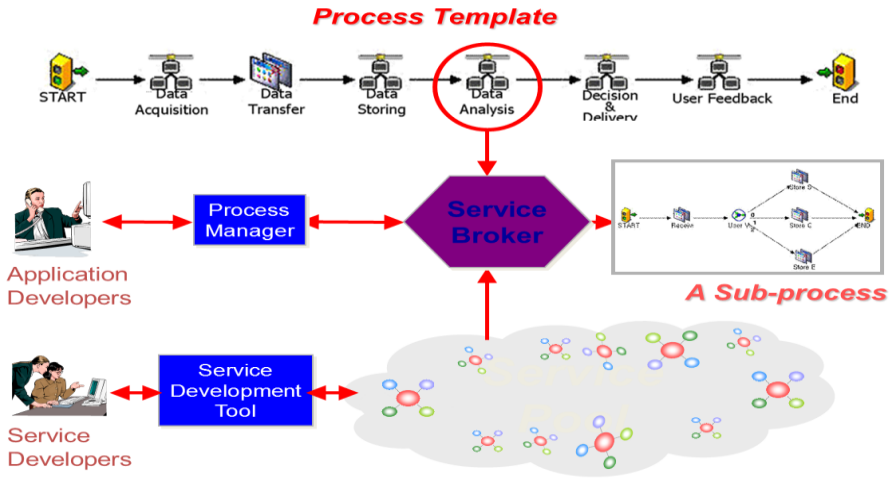


Figure 8. Bio Data Analysis Framework

3.4.3. Mobile U-health Service Framework

The mobile u-health service framework shown in Figure 9 is a kind of middleware software that supports services for data acquisition, storing and management, and analysis as an integrated service. This framework offers any kind of application service loaded on our u-health platform not only to the terminal but also to the web service. Although there are some overlaps, we can assign elements such as the client device tier, business logic tier, and data management tier to this framework, as shown in Figure 9.

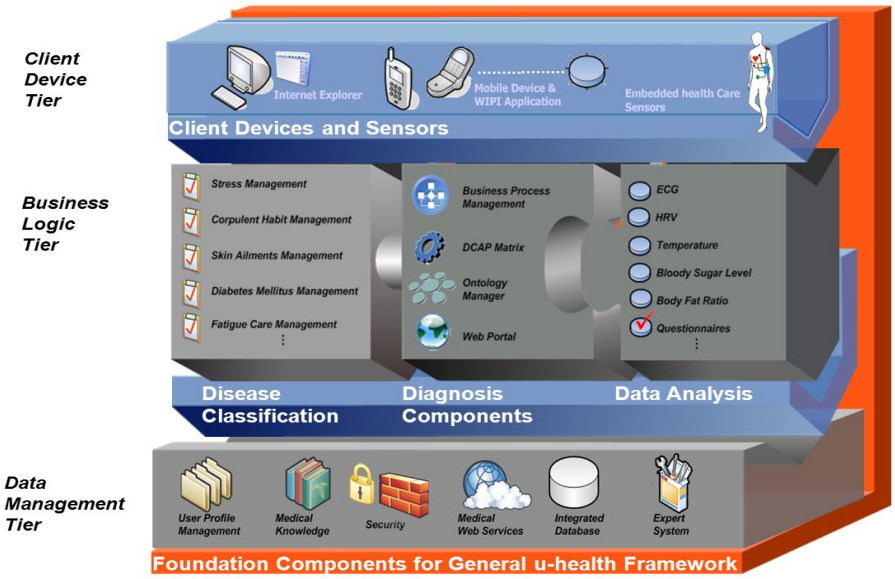


Figure 9. Mobile U-health Service Framework

4. Implementation of the Mobile U-Health Service System

4.1. Data Used for Validation

Since there were insufficient data for the validation test of stress diagnosis, only the basic validation test was performed. The stress test data used in this validation test were provided by the Stress Clinic of Prof. Jong-min Woo of Paik Hospital in Seoul. The stress data consisted of the physical records, including age, gender, height, and weight; physiological data, such as blood pressure, blood sugar level, and body temperature; RR (respiration rate) interval data measured using the vein wave meter; and the results of the SRI (stress reservation index) questionnaire. To validate the PCADP matrix, the data received from Paik Hospital were selected and analyzed for use as the learned data for the PCADP matrix.

Realizing that the stress values are significantly dependent upon age and gender, the PCADP matrix was separately built according to gender and age to build the learned data. The RR interval was analyzed via various methods to obtain the elements of the bio signals, such as SDNN (Standard Deviation of the NN intervals), LF, and HF. For the SRI, the factors of the symptoms were obtained from the score for each question by

selecting the symptoms related to the question. The 19 bio signals and symptoms listed below were used for the stress validation test.

- Bio signals: GLUCOSE, SYSTOLIC, DIASTOLIC, BODY_FAT, HF_NU, LF_NU, LF_HF, HF_MSEC_2, LF_MSEC_2, VLF_MSEC_2, TINN_SEC, HRV_INDEX, PNN50_PERCENTAGE, RMSSD_SEC, CV_PERCENTAGE, SDNN_SEC, MEAN_RR_SEC, HEART_RATE

- Symptoms: UNEASINESS, TACHYCARDIA, STIMULUS_EVASION, SOLICITUDE, SCATTERBRAINED, PALPITATIONS, OVERSENSITIVE, OUTRAGE, IRRESOLUTION, HEADACHE, HAND_TREMORS, GLOOM, EVADING_CROWD, EMPTINESS, DYSPEPSIA, DIZZINESS, CHRONIC_FATIGUE_SYNDROME, ABUSE, ABDOMINAL_PAIN

4.2. Validation

Concerning mobile U-health evaluation for the purpose of system configuration, this paper presents the possibility of a concrete and real service by implementing a program is a test bed environment destroyers. The disease diagnosis probability calculated by using a matrix of the PCADP and user feedback mechanisms, according to the available personal data to the service of the evolutionary, the diagnosis can improve the reliability.

Shown in Figure 10, first check that the authenticity (authentication?) of the terminal is sent from the data stored in the database. In the ontology manager for each bio/symptom data analysis and the data to identify possible causes, to compensate for the weight. The next phase of the PCADP matrix adjusted weights are used as input data, PCADP matrix with running-set used in the operation belong to the range of disease probabilities are calculated. Where the periodic update of the running-set is more accurate over time can be measured.

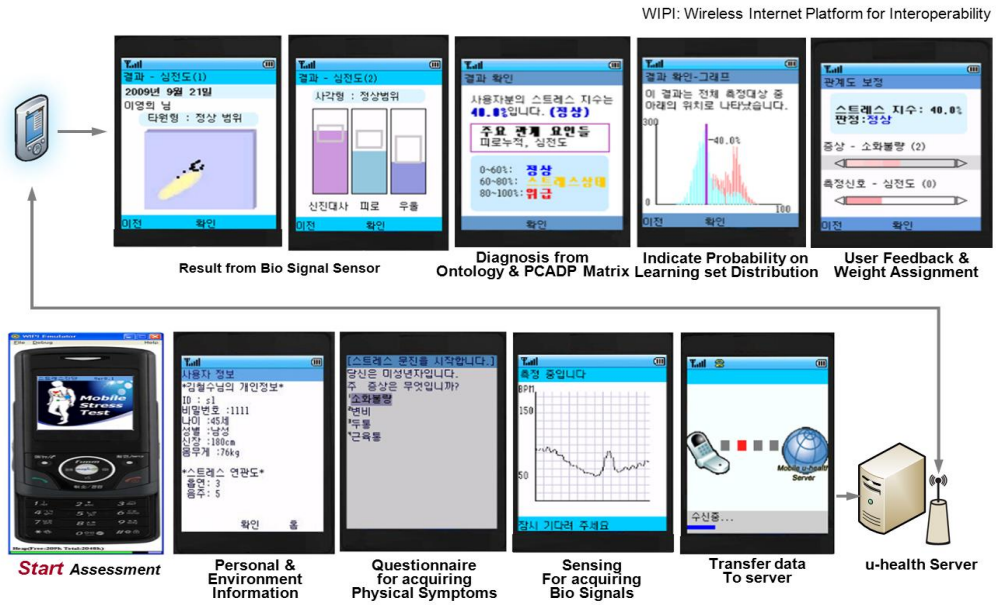


Figure 10. Implementation of the Mobile U-health Service System

4.3. Validation Test Result

The stress validation test was conducted in two ways. To develop a PCADP matrix, learned data that are clearly separated into the stress group and the normal group are required. However, the physiological data, RR interval and SRI data received from Paik Hospital were not separated into a stress group and a normal group. Thus the data had to be arbitrarily separated into two groups prior to the validation test.

The first method consists in using the SRI value to measure the actual stress level. However, the way to identify the stress with the SRI value is very strict, and there will be too few left for the stress group. Therefore, the SRI values of 18 people were divided into a certain ratio and the learned data of the stress group and the normal group were applied. In this test, the ratio was set to 15%, 30%, and 50% to compare accuracy.

The second method consists in separating the groups using the HRV features, such as SDNN or LF/HF, which are used to identify the stress group. SDNN is the feature obtained from the time domain method. It is the most easily obtained value, and people suffering from stress generally have a low value. LF/HF is the feature obtained from the frequency domain method. It reacts sensitively to sympathetic change. Generally, stress increases sympathetic activities and eventually raises the LF/HF ratio. All the tests validated the PCADP matrix by measuring the accuracy of the prediction using the r-cross validation method.

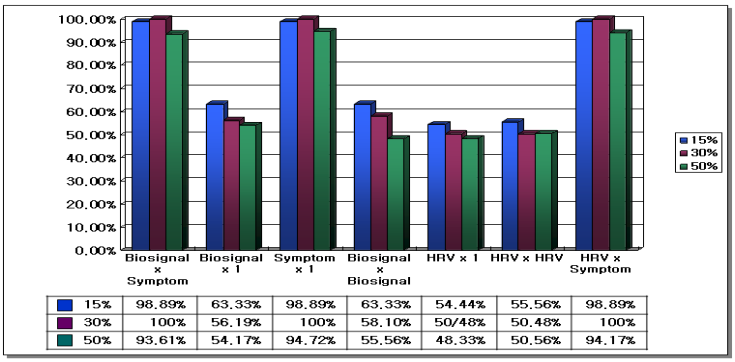
4.3.1. Identification with the SRI

The top and bottom 15%, 30%, and 50% were tested. The PCADP matrices of various structures were formed to be tested for accuracy. The following seven types of PCADP formats were analyzed.

- Biosignal x Symptom (19 x 19)
- Biosignal x 1 (19 x 1)
- Symptom x 1 (19 x 1)
- Biosignal x Biosignal (19 x 19)
- HRV x 1 (14 x 1)
- HRV x HRV (14 x 14)
- HRV x Symptom (14 x 19)

The above seven PCADP structures were validated by the top and bottom 15%, 30%, and 50% methods to identify the stress group. A total of twenty-one tests, i.e., three for each structure, were performed. Figure 11 shows the result.

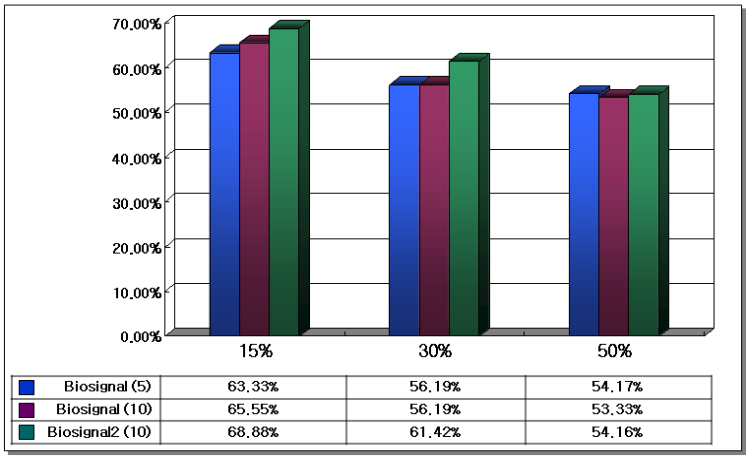
As shown in Figure 11, all the matrices showed 90% or higher accuracy, which indicates that the correlation of between the SRI and the extracted symptoms was too high. As such, the separation of the stress group and the normal group with the SRI could not be correctly validated with the learned data. Therefore, a test aimed at improving the accuracy of the Biosignal x 1 matrix (excluding the symptom) was performed. To improve its accuracy, the number of features or feature ranges could be adjusted.



15%	98.89%	63.33%	98.89%	63.33%	54.44%	55.56%	98.89%
30%	100%	56.19%	100%	58.10%	50/48%	50.48%	100%
50%	93.61%	54.17%	94.72%	55.56%	48.33%	50.56%	94.17%

Figure 11. SRI Result using PCADP

In Figure 12, Bigsinal2 means the new bio signals after reducing the number from 19 to 9. The figure in the parenthesis means the number of ranges. Accuracy was increased by increasing the number of ranges from 5 to 10 and eliminating unnecessary features.



Biosignal (5)	63.33%	56.19%	54.17%
Biosignal (10)	65.55%	56.19%	53.33%
Biosignal2 (10)	68.88%	61.42%	54.16%

Figure 12. Change in the Number Range of the Experimental Results after Biosignal

4.3.2. Identification by SDNN and LF/HF

Like the SRI, the top and bottom 30% of the SDNN and LF/HF values were separated for use as the learned data of the PCADP matrix. (In the case of SDNN, the top 30% is the normal group and the bottom 30% is the stress group. In the case of LF/HF, the top 30% is the stress group and the bottom 30% is the normal group.)

Eighteen bio signal features - after excluding one used to identify the group – out of a total of nineteen were used; nineteen symptom features were used; and five equally divided ranges were used, as in other methods.

Three tests with SDNN and three tests with LF/HF were performed for validation. Assuming that there would be a correlation among the HRV features, three tests for identifying the groups with the glucose instead of the HRV feature were added, making a total of nine tests. The results were compared with the results of the validation of the SRI. The nine tests are shown below:

- Biosignal x Symptom (18 x 19) - SDNN

- Biosignal x 1	(18 x 1)	- SDNN
- Symptom x 1	(19 x 1)	- SDNN
- Biosignal x Symptom	(18 x 19)	- LF/HF
- Biosignal x 1	(18 x 1)	- LF/HF
- Symptom x 1	(19 x 1)	- LF/HF
- Biosignal x Symptom	(18 x 19)	- GLUCOSE
- Biosignal x 1	(18 x 1)	- GLUCOSE
- Symptom x 1	(19 x 1)	- GLUCOSE

Figure 13 shows the results of the nine tests. Like the SRI, it was confirmed that both the SDNN and LF/HF have a correlation with the bio signals. Therefore, it is not suitable to identify the groups with a single method using the existing data. To obtain an accurate validation result, a virtual stress group would have to be formed and validated using each method.

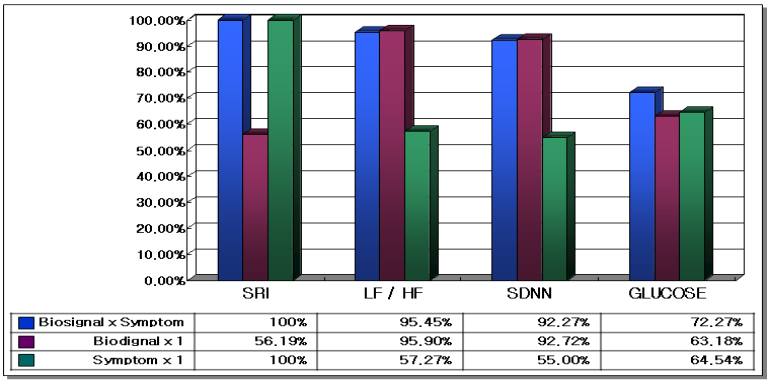


Figure 13. Test Result of the Validation of SDNN and LF/HF

4.3.3. Stress Probability Distribution Diagram

As shown in Figure 14, the PCADP matrix has the stress group and non-stress group distributions. Since these data regularly evolve, they are used as the basic data for measuring the stress probability according to the general bio signals and the questionnaire.

The prediction stage is established based on the appearance pattern of the bio signal and symptom pair constructed in the learning process. In the learning stage, separation of the correlation pattern of the bio signal data and physical symptoms of the normal group and patient group is attempted. This study assumed that the sufficiently learned data could be used at this stage to recognize the correlation.

In the learning stage, the frequency of the correlation of the bio signals and physical symptoms of the normal group and patient group are stored in the CI matrix. The CI matrix contains the pattern and its appearance frequency for the bio signals and physical symptom sets of the given normal group and patient group.

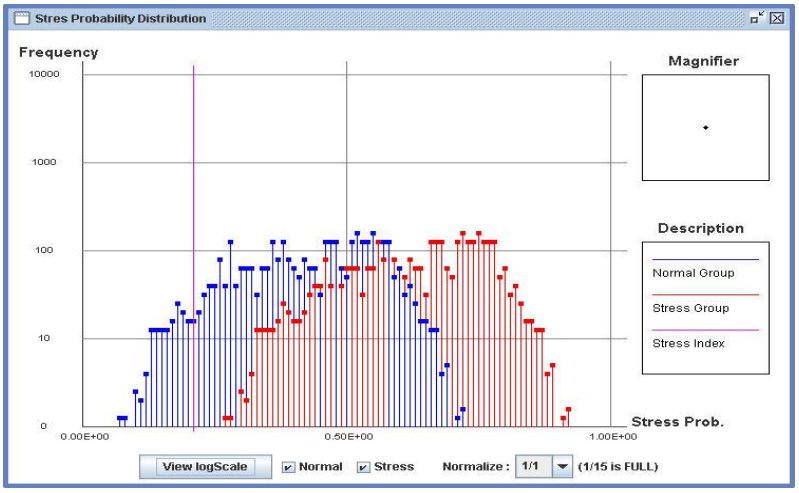


Figure 14. Stress Probability Distribution Diagram

The CI matrixes of the patient group and normal group are integrated into the PCADP matrix. Each element of the PCADP matrix represents the probability of the set pattern appearing in the patient group. In the prediction stage, the bio signals and physical symptoms of the subject of the disease diagnosis are input into the prediction system, and the prediction system calculates and presents the probability that the subject will be assessed to be in the patient group. Since the PCADP matrix is generated with the CI matrix, the contribution of a certain combination of bio signals and physical symptoms to a certain diagnosis can be easily evaluated. In Figure 15, the parts displayed in color are the elements of the set of bio signals and symptoms. The PCADP calculation equation then calculates the probability between 0 and 1.

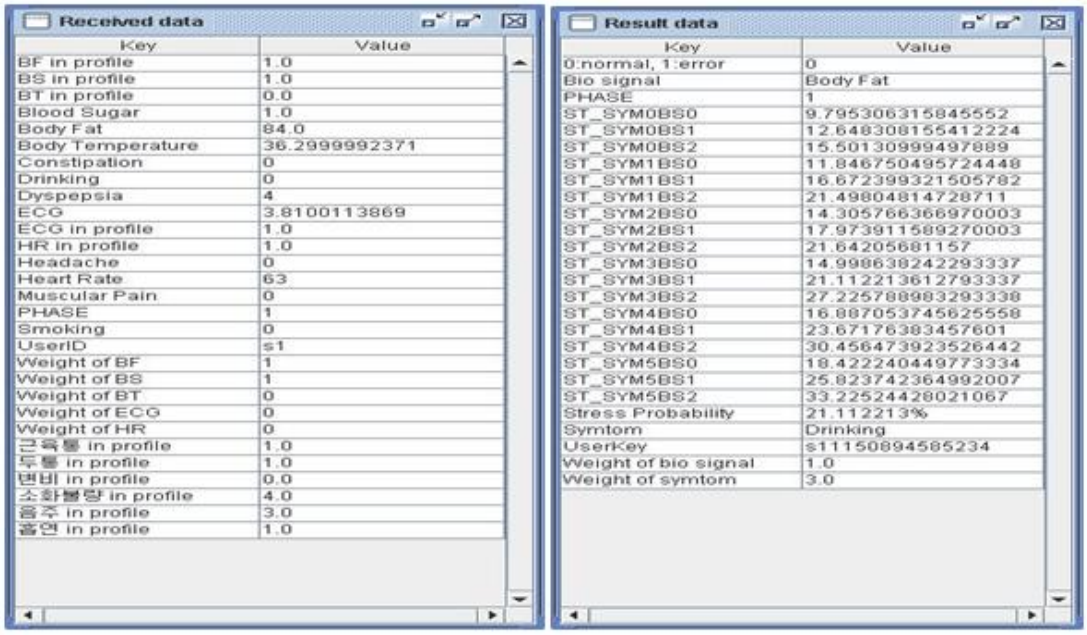


Figure 15. Health Index of a Subject Obtained from the PCADP Matrix

4.4. Comparison and Consideration

This paper proposes a PCADP algorithm, which is a personalized statistical disease prediction method suitable for the mobile u-health service environment. It established a flexible structure, real-time processing, continuous enhancement, and monitoring of the diagnosis process, which are all essential elements of a diagnosis method in the mobile u-health service environment, as the requirements. Although security is another important requirement, it was not considered in this study. However, it is an issue that must be answered if the u-health service system is to be commercialized and distributed in the future.

The existing disease prediction method through machine learning is an artificial intelligence technology designed to teach a computer to analyze the data in order to identify a new regularity and turn it into a program. Although machine learning is not the same as the statistical method, they are similar in that they analyze data to find the rule. However, the algorithm used in the machine learning method is more complex and precise than that of statistical learning.

4.4.1. Flexible Architecture

The goal of a mobile u-health service is to comprehensively check the health of the subject to help diagnose and treat him or her. Since the assessment of various diseases is needed for that, the diagnosis technique must be able to assess different diseases without a major change, and must support all the variables required for the diagnosis of each disease. The PCADP matrix proposed in this paper manages the data based on ontology to improve the flexibility of the system. The ontology structure and PCADP matrix values are dynamically changed according to the user's feedback.

4.4.2. Real-time Processing

As the mobile u-health service must enable immediate learning, diagnosis, analysis and management without the constraints of time and space, and the diagnosis method, which is the key to the service, must be optimized so that the diagnosis can be promptly executed. This paper proposes a real-time statistics-based disease prediction method. To solve the problems of the long processing time and the difficulty of finding the pattern when a large volume of data are input, it uses set association and the decision tree to reduce the volume of data, so that it can be applied to a large volume of data and improve the accuracy of prediction.

4.4.3. Continuous Enhancement

Since the majority of disease diagnosis methods depend on the nature of the learned data, the performance of the diagnosis is sensitive to any changes made to the learned data. Therefore, it is very important to effectively reflect the continuously measured data in the learned data and enhance the diagnosis method.

The mobile u-health ontology proposed in this paper is continuously enhanced with the accumulation of new data and user feedback, and the PCADP matrix is also further evolved. The changing of the PCADP matrix data for more complex and personalized analysis according to the feedback on the analysis result improves the disease prediction method.

4.4.4. Monitoring of Diagnosis Process

For comprehensive health management, the diagnosis method should not end with just assessing the disease but should also recognize a change in the user's condition and present a series of follow-up processes. Although the diagnosis method itself may not support it directly, it should be able to identify the variables that significantly affect the diagnosis and those that may not have a significant impact but which significantly deviate from normal data value, so that they can be provided as reference data for the next step.

The ontology manager monitor is a tool for using data given by the service to visually present the process of the ontology manager search, and to collect and infer the data so as to deduce a result. The ontology service was developed to assess the stress group. The symptoms ontology and ontology API were used to systematically manage the weight factors of the bio signals to assess the stress group.

5. Concluding Remarks

The mobile u-health service generally consists of four stages: the sensing stage, in which the health customer's bio signals are collected using sensors; the monitoring stage, in which the bio signals undergo preliminarily processing; the analysis stage, in which a new health index is generated from machine learning, such as pattern recognition, and data mining of the obtained data; and the feedback stage, in which the customer is notified of his/her health status. In particular, the analysis stage, whereby a new health index is calculated from the accumulated bio signals, may be considered as the key to and the (essential) platform of, the mobile u-health service

There are many problems with applying the machine learning technique, which is widely used in the conventional healthcare field, during the mobile u-health service analysis step. First, research on the mobile u-health service is just beginning, and there are very few cases where the existing techniques have been applied in the mobile u-health service environment. Second, since the machine learning technique requires a long learning period, it is not suitable for application in the mobile u-health service environment, which requires real-time disease management. Third, the various machine learning techniques that have been proposed until now do not include a way to assign the weight factors to the disease-related variables, and thus its use as a personalized disease prediction system is somewhat limited.

This paper proposes PCADP, which is an ontology-based personalized disease prediction method, to solve such problems and to interpret the bio data analysis of the mobile u-health service system as a process. Moreover, the mobile u-health service ontology framework was modeled as a semantics type in order to meaningfully express the mobile u-health data and service statement based on PCADP.

The proposed algorithm is designed to assure the reliability of the data and application service and enable the prediction of general diseases such as stress, obesity, and diabetes. When the bio signals, types and forms of the physical symptoms and weight factors of a disease are given, it constructs a matrix regardless of the disease to support the various structures and learn and diagnose quickly. It can also very easily back-track the diagnosis result. To improve the accuracy of prediction, a normalization function with consideration to the difference in size of the normal group and the patient group was

added to PCADP calculation, and the feedback mechanism was enhanced to make the personalized diagnosis more precise.

To validate the performance and efficiency of the PCADP technique proposed in this paper, the 5-cross validation method was used to measure the accuracy of the prediction. The validation of PCADP using a virtual disease group verified that the technique proposed in this paper shows much greater accuracy compared to existing methods. Moreover, the PCADP prediction method improved the flexibility and real-time attributes, which are the essential elements of any diagnosis technique in the mobile u-health environment, and showed efficiency in the continuous improvement of the monitoring and system of the diagnosis process.

Further studies are required to commercialize the mobile u-health service proposed in this paper, and there are some problems that must be resolved to make the system safer and more reliable. For the future, studies on how to manage the health data safely, and how to compress the messages exchanged between the mobile devices and the u-health server to improve the performance are needed.

References

[1] Joong-Don Kwon, *Customer Characteristics Analysis of the Curing Expert System for Dementia or Other Disabilities*, AlphaInternet Co. Ltd., **(2001)**

[2] Dong-Soo Han, In-Young Ko, and Sung-Joon Park, *A Study on the Development of the Mobile U-Health Service System*, the final report of research with the same title from ICU, Korea, **(2006)**

[3] Dong-Soo Han, In-Young Ko, and Sung-Joon Park, "Evolving Mobile U-Health Service Platform," Proceedings of Information Security Society, Vol.17, No.1, **(2007)**, pp.11-21.

[4] D. Konstantas, R. Bults, A. Van Halteren, K. Wac, V. Jones, I. Wkdya, R. Herzog, and B. Streimelweger, "Mobile Health Care: Towards a commercialization of research results," Proceedings of the 1st European Conference on eHealth-ECEH06-Fribourg, Switzerland, **(2006)** October, pp. 12-13.

[5] M. Pappas, C. Coscia, E. Dodero, G. Gianuzzi, and V. Earney, "A Mobile E-Health System Based on Workflow Automation Tools," Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems, **(2002)** June, pp.271~276.

[6] Byoung-Won Min, Yong-Sun Oh, Dong-Soo Han, and Jong-Young Ku, *Design of a Mobile U-Health Service Platform*, Proceedings of the Fall 2009 Integrated Conference, Korea Contents Association, Vol.7, No.1, **(2009)**, pp.797-801.

[7] Hyun-Serk Lee, Jeong-Ho Bak, Bo-Kyeong Sim, Hak-On Lee, Sang-Won Han, Byoung-Won Min, and Hyeun-Tae Lee, *Web-based Patient Monitoring System Using a Wireless Diaper Wetness Sensor*, Proceedings of ICCC2008, Vol.6, No.2, Korea Contents Association, **(2008)**, pp.652-660.

[8] Byoung-Won Min, Hyun-Tae Lee, and Yong-Sun Oh, *USN-Based Intelligent Urine Sensing U-Care System*, Proceedings of the Spring 2008 Integrated Conference, Korea Contents Association, Vol.5, No.2, **(2008)**, pp.598-601.

[9] Byoung-Won Min and Yong-Sun Oh, *Design of a U-Healthcare Product Using Wetness Sensor*, Proceedings of the Spring 2007 Integrated Conference, Korea Contents Association, Vol.3, No.2, **(2007)**, pp.144-147.

[10] Hyun-Gyu Park, Hye-Jung Kim, and Seung-Jae Lee, *A Transmission Management System for Signals from Living Bodies Using ZigBee*, Proceedings of the 2008 Conference, Korea Computer Society, Vol.32, No.1, **(2005)**, pp.526-528.

[11] M.A. Klopotek, *A New Bayesian Tree Learning Method with Reduced Time and Space Complexity*, Fundamenta Informaticae, Vol 49, Issue 4, **(2002)**

[12] I. Inza, M. Merino, P. Larranage, J. Quiroga, B. Sierra, M. Girala, *Feature Subset by genetic algorithms and estimation of distribution algorithms*, A case study in the survival of cirrhotic patients treated with TIPS, Artificial Intelligence in Medicine 23, **(2001)**

[13] C.F. Cooper et al., *An evaluation of machine learning methods for predicting pneumonia mortality*, Artificial Intelligence in Medicine 9, **(1997)**

[14] M.J. Fine, B.H. Hanusa, J.R. Lave, D.E. Singer, R.A. Stone, L.A. Weissfeld, C.M. Coley, T.J. Marrie, W.N. Kapoor, *Comparison of severity of illness measured in patients with community-acquired pneumonia*, J.Gen. Int. Med. **(1995)**

[15] D. H. Kim, S. Uhmm, S. W. Cho, K. B. Hahm, J. Kim, *A Predictive Model for Chronic Hepatitis Susceptibility from Single Nucleotide Polymorphism*, BIOINFO 2006, pp. 35-38 **(2006)**

Author

Byung-Won Min

He received M.S degree in computer software from Chungang University, Seoul, Korea in 2005. He received Ph.D. degree on Department of Information Communication Engineering from Mokwon University, Daejeon, Korea. He worked as a professor in the department of computer engineering, Youngdong University, Youngdong, Chungbuk, Korea, from 2005 to 2008. He joined the faculty of the Dept. of Information Communication Engineering, Mokwon University, Daejeon, Korea, in 2008. His main research interests include Ubiquitous Sensor Network and embedded system.