MULTI-MODAL WEIGHTED DENOISING CODER FOR THE MANAGEMENT OF LOST INFORMATION IN HEALTHCARE BIG DATA

SachinKumar Veerashetty

Department of Computer Science, Sharnbasva University,Klaburagi,Karnataka,India.
E-mail:sveerashetty@gmail.com

Abstract - A data-based health model includes different missing values, and the precision of a health model that requires variables that the consumer does not capture is poor. A profound learning health paradigm is designed to enhance accuracy in terms of learning weights. When a deep learning health model is applied to the user's circumstances, precision might be decreased when learning is not achieved. The paper suggested a method for predicting missing data in the field of health big data with the use of the Multi-Modal Weighted Denoising Code (MMWDC). A weighted denoising automatic encoder is used to predict incomplete information during data collecting and processing phases. The suggested approach employs autoencoders based on neural networks whose output values are estimated based on their input values. Data from the National Health and Nutrition Examination Survey (NHANES) have been utilized in this work. Based on this approach, missing data in the Individual Health Archives (IHA) can be estimated. In addition, the IHAs feature multimodality that permits the collection of data for a single item from a variety of sources. The weighted denoising autoencoder is, therefore, set up with a multi-modal configuration. A data collection with no missing value is designed using NHANES after pre-processing. A tag is set as unique information in given dataset training, and an autoencoder input is designated as a noisy input with as many arbitrary zeros as noise value. The autoencoder, therefore, learns to resemble the original label value in the form of a zero-based noise value. The reliability of the suggested technique utilizing a multi-modal weighted denoising self-coder exceeds that attained by other conventional methods when the missing information in a data set is around 26.5%. In addition, a multi-modal approach can save extra period when handling vast volumes of data in places like hospitals and universities.

Keywords: Multi-modal weighted denoising, Missing information, Big data, Healthcare,Autoencoder.

1 INTRODUCTION

Different convergence sectors are expanding based on the growth of fourth manufacturing technology, such as communication, data, and sensors. This causes significant changes in the industrial areas from intelligent farms to intelligent houses and people's everyday lives [1]. In particular, the data-based health sector accumulates enormous volumes of data, especially in many nations, corporations, and research institutes [2], through the provision of electronic records, personal health records, personal health devices, etc. To collect information, machine learning and profound learning assist initiatives such as the construction and disclosure of public-interest-related data at the national level are being carried out. Different health-related data are gathered and used from numerous devices [3].

However, the model generated by deep learning needs authentic learning. Therefore, the quickly changing user environment, health, and activity are not readily addressable [4]. Scenario variables quickly alter a person's health state; every operator has a specific device so that the arrangement is different and limits the usage of the comprehensive model[5][6]. Health demands a profound learning model to be adaptable to the many factors that are accessible, depending on the condition of the user. Therefore, instead of an extensive deep neural network, a framework suited for the user should be divided into smaller ones[7]. Composite modeling should be studied, which complicates the shortcomings of each other and optimizes the advantages of information retrieval or machine learning approaches through the integration of a general profound learning model[8].

Missing data have been dependent on the particular conditions of the subject. In addition, during integration, data duplication or omission might also occur. Soft computing offers the best approximations of the data lacking [9] to tackle this challenge. The lack of data affects information scrutiny or knowledge, and a model created from incorrect data is used less accurately[10]. Reproduced or missing values are approximated using average, median, and mode values or utilizing algorithms like regression, neural networking, Singular Value Decomposition (SVD), or Nearest Neighbor to K (K-NN)[11]-[14]. Although estimating the average, median, and fashion value is easily obtainable, it is less precise and hence less practical for its use.

Furthermore, a reversion estimate, SVD, or K-NN may reach reasonably good precision, but it requires human involvement, and substantial preparation for computational applications is required. In addition, a neural network estimate enables the model to learn characteristics by itself from the data, eliminating the need to intervene by users[15]. Therefore, this work proposes using a multi-modal weighted denoising autoencoder to predict missing information, especially missing data in IHAs, based on medical data.

The remaining document is organized as follows. Section 2 explores related works on missing data estimation for big data in healthcare. Multi-Modal Weighted Denoising Coder (MMWDC) applied to healthcare big data has been explained in section 3. Section 4 consists of the analysis and findings obtained from the proposed model. Finally, the conclusion and possible studies have been outlined in Section 5.

2 RELATED WORKS

Smart health attracts attention as the data accessible in the medical sector rises, including information on individual health, automated medical archives, general health, and genomic data [16]. In particular, a range of information is gathered without restriction from wearable devices on a portable wellbeing platform. Smart watches, intelligent bands and personal health equipment, and smartphone apps accumulate user movements, vital signs and lifelogs, locations, weather, and more. This information is directly linked to the operator's wellbeing, and much research is in progress[17].

With the industry's growth, there is an increased number of data, including temperature, precipitation, lighting, relative humidity, atmospheric pressure, ultraviolet light, pulse rate, activity, and sleep, which may be collected using mobile sensors. Health models are created utilizing data mining, machine learning, and profound learning to use this data. The NHANES [18] gathers information employing questionnaires on health, exams, and nutrition. The dietary habits, genetic predisposition, diseases, and medical archives are obtained through wellbeing inspections, and frequent checks are carried out on pulse, blood pressure, weight, height, and blood sugar. [19].

The survey gathers the regularity of food, the quantity of food, the quantity of water ingested, and how nutrients are taken via a survey. However, it has a high possible value, distant from typical IHA, as it incorporates several health-related factors. There are around 700 variables, which anyone may create and control. The correlations or norms of the variables utilized in health can be inferred when examining this with data extraction[20].

IoT and smartphones build individual health networks and disseminate enormous volumes of data produced by networks[21]. Data are also frequently gathered by consumers regarding the growth of data and communication, intelligent telephones, individual wellbeing equipment, IoT, etc.[22]. Action, place, heart rate, strain, blood pressure, weight, fatness, sleep, and other devices through intelligent clockwork, smart bands, and smartphone applications are continually gathered in different kinds of devices in everyday life[23]. When data are collected, attention is drawn to data mining and machine training, which may be utilized with limited data. Still, profound education is now being attracted by creating an environment where enormous amounts of data are produced and distributed[24]–[26].

According to the Fourth Manufacturing Revolution, information is gathered widely and the intelligent health business is constantly expanding. Due to the development of different devices, information, and communication technology, innovative health has used much information, and many profound learning models have been established [27], [28]. In this regard, it is essential to integrate the data acquired by different pathways. For a daily model, values can be gathered for one single variable at the same time.

Moreover, if observation depends on the variable, some problems are tough with one input. Different models are generated in data mining and deep learning based on the constitution, training data, and objectives. Duration, non-time, and picture type data are the dominating information for mental healthcare. Statistical analysis information with a continuous and rapid cycle was gathered, including action, heart rate, weight, and duration for sleep[29]. Data from non-time series include incoherent cycling information such as a diagnosis by the physician, frequency, family background, and surgical status. Image information reflects CT, MRI, X-ray, ultrasound, etc. [30].

3 MULTI-MODAL WEIGHTED DENOISING CODER (MMWDC) FOR ESTIMATING MISSING VALUES

The health statistical data comprises the habits, genetic predisposition, illness, and medical records of an individual. The heartbeats, blood pressure, body weight, height, and blood sugar levels are included in a health checkup. A food survey comprises the regularity of food, the volume of food, water consumption, and dietary supplements by the individual. Due to its massive amount of data stored in the IHA, NHANES is not a conventional IHA but has excellent potential [2]. Some 600 parameters, most of them prepared and maintained by individuals, comprise several health-related things. An NHANES model is therefore beneficial to an IHA.

In addition, NHANES is a multi-modal kind of information gathered through several modes. In reality, information have been gathered through health surveys, health checks, and dietary practices based on the information assortment path by each method. The lack of data is introduced under many conditions throughout the integration process. When several transactions have missing information, the results of the data analysis differ based on the procedures used for pre-processing. This calls for a suitable processing method, which minimizes the consequences of the lack of data on the results.

In the present work, a method is suggested to estimate missing information utilizing a weighted multi-modal autoencoder in the turf of large-scale healthcare. This approach predicts the lack of data by using a skilled multi-modal information autoencoder. NHANES consists of several variables, and different multi-modal combinations are achievable by the categorization technique for the variables. Better outcomes from data analysis and machine learning may be predicted through the management of missing data. The settings utilized to estimate such missing data are presented in Fig.1.

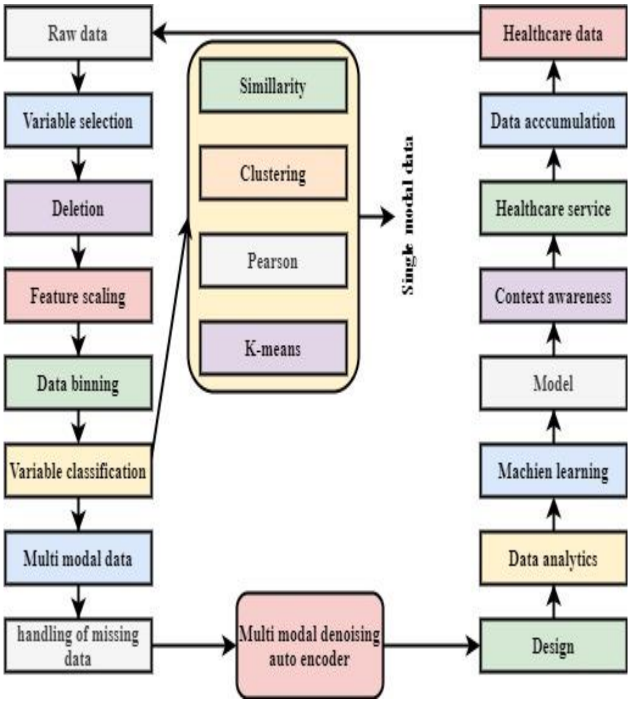

Figure 1 Framework to find the missing data in healthcare applications

3.1 Pre-processing

3.1.1 Raw Data and Selection of Variables

Information from health, health tests, and dietary studies are used in this study among the primary data published by NHANES from 2014 to 2019. In addition, the surveys are divided by only standard parameters because of annual variations in the scope of every study and exclusion of any factors that lack 20 percent or more of data in a column scan. One hundred ninety-eight elements, including pre-existing health problems, medical diagnosis, smoking, alcohol consumption, loneliness, and anxiety, have been chosen from among the primary data.

15,712 instances have been used with 189 selected parameters, except for cases having a missing information degree of 24% or above indicated by a column scan. Such replies are not deemed missing. However, the results of the data analysis could be affected by specific classifications. In many parameters, the not applicable (NA) answer has been included, but out of all 15,712 cases, the parameters with substantial amounts of NA replies (3,500 or more). Eighty-five characteristics are therefore selected and utilized as experimental evidence. NHANES includes 0 (class value) information, which requires correct processing. If there are various scales for continuous parameters, it might result in too high values or integration of weights into 0. Differences in size might result in inconsistent weight learning, which requires an additional characteristic scaling[31].

3.1.2 Feature Selection

Feature scaling is a method for normalizing the parameter ranges to be uniformly scaled. This approach assesses the impact of a specific characteristic during statistical analysis and in a neural network model via regression or clustering analysis. Therefore, while employing the feature scale, it transforms parameters to a definite type via data binning.

3.1.3 Data Binning

Binning is a method for categorizing incessant characteristics into intermissions (bins). NHANES facilitates information binning and attributes to classify variables for the various units, including age, height, and subject pressure. The parameters are classified into nine intervals, with a minimum constraint rate of 0 percent and a maximum rate of 100 percent.

Table 1 The binning and continuous parameter scaling of the NHANES data

Normalized binning	0(NA)	1(0-0.1)	2(0.1-0.2)	3(0.2-0.3)
Normalized binning	4(0.3-0.4)	5(0.4-0.5)	6(0.6-0.7)	-
Normalized binning	7(0.7-0.8)	8(0.8-0.9)	9(0.9-1.0)	-

The binning and continuous parameter scaling of the NHANES statistics have been given in Table 1. Due to the lack of data on the experimental data set to a value of 0, parameters are binned between 1 and 9.

3.2 Finding the Missing Value Using MMWDC

The weighted denoising self-encoder utilized in conventional self-encoders is a change of the learning techniques. To restore the initial noise-free data, this self-encoder augments arbitrary noise to noise-free input information. This repeats the process of altering the input value by adding noise and returning it to the original data. A random value is selected from the input before the data entry and converted to 0 by the autoencoder. The missing data is usually considered to be 0 for neural network learning. In the same way, noise is approximated to zero and reestablished to the unique information when missing information is used in denoising autoencoder training.

Therefore, when missing information are entered, a value of 0 is generated using a network trained with a non-zero projected value. Several hidden layers are necessary to represent the modality according to the information features as a neural network. Furthermore, different forms may be created using a weighted denoising autoencoder when stacking concealed layers in several lays. This is split into unattended knowledge and supervised knowledge by technique. The use of a restricted Boltzmann machine (RBM) as a deep faith network created early-weighted denoising self-encoders.

This is used to solve the huge processing, local minima and gradient issues that are disappearing. At now, usage of different propagation features and optimizers enhances learning access to the background. Weighted denoising autoencoder with supervised learning has been used in this work. Input includes data that randomly generates 25% of the missing values (0) as noise. The data on the label consists of x without missing values. The AutoEncoder is learned via supervised learning, due to one-time training with a loss of the MAE.

Fig. 2 demonstrates a weighted denoising autoencoder estimate of missing data. As seen in Fig. 2, the original document being noise-added and translated into training data. In addition, weights are learnt using the inaccuracy in the neural network output values with the real data being a label. Fig. 2 depicts a weighted denoise autoencoder composed of an input level, three hidden levels, and an output layer. The value of the transactions in the source data is around 25 percent missing. The noise ratio rises by 0.05, starting from 0.05 to 0.30 during weighted denoising autoencoder learning.

The information picked in NHANES is 14 688 without a lack of value for the trials. In the stage of noise addition, a set of noised KNHENS is thereby substituted randomly with as many input values as the noise factor. The label is currently the original noise-free information, and it is utilized for the calculation of an output error. Experimental results have been distributed to 14,688 instances at random: 75% of training data, 11% of model validation and 14% of testing dataset. There have been picked a total of 80 input nodes and 80 output nodes. With the increasing quantity of hidden layers, recurrent studies demonstrate a decreased precision and more significant loss, which generally takes place as deep knowledge is not extensive in health care.

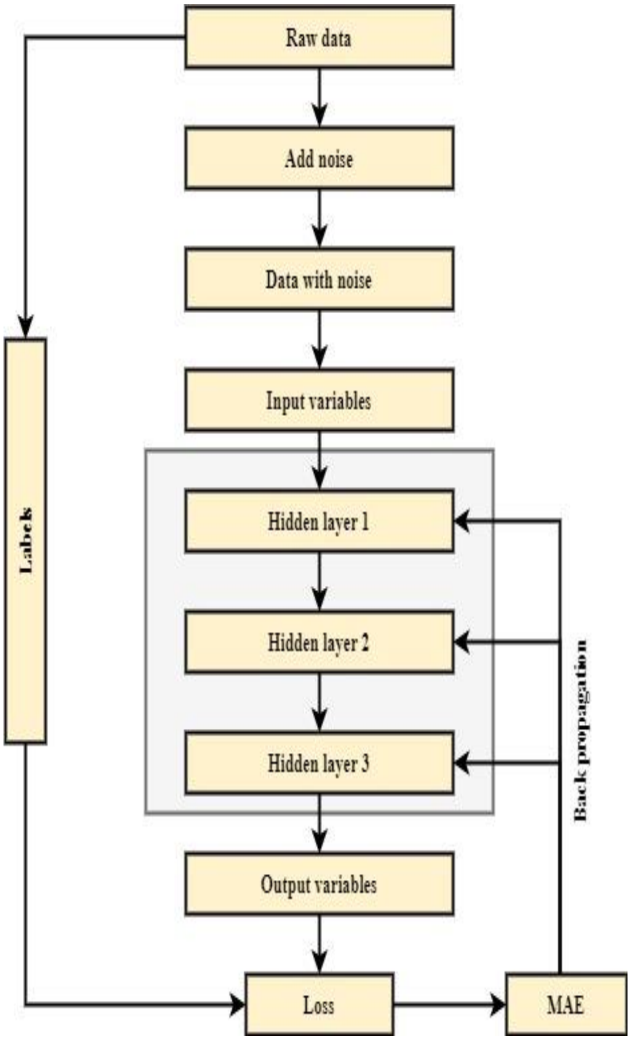

Figure 2 Assessment of missing data using the weighted denoising autoencoder

The NHANES data, therefore, reveal a more excellent organization for the neural network. The amount of hidden nodes grows to one-half of the input layer. The outcomes of multiple trials demonstrate that the precision and degradation of learning and validation data are not as high as 64 nodes.

Table 2 Autoencoder applied learning results

Model	**DAE**	**Proposed MMWDC**
Number of inpu layers	80	80
Hidden layers	64	64-32
Output layers	80	80
Variables	15,712	14,688
Distortion factor	0.15	0.15
Precision	0.9412	0.9531
Loss	0.4310	0.5121

Table 2 shows the autoencoder based functional learning results. Representations with more constraints often have more knowledge of features. The results demonstrate that when the model is designed using one hidden layer and 64 node models, the best accuracy and negligible loss can be attained. Although the effectiveness of the existing system falls with an increasing number of constraints since the number of parameters in the two models is only slightly different, the 80-64-80 model for the study has been chosen. This shows that most estimates for missing data are significant. The accuracy of the model used is 0.9531.

4 RESULTS AND DISCUSSION

The approach suggested is contrasted with current data estimate methods that were missing for validation. The testing information is managed using a multi-modal autoencoder weighted denoising (MMWDC), K-NN, SVD, and column means (c.mean) and thus arranged into the training statistics. In the c.mean procedure, a missing value is replaced by the mean of each column. The supervised learning is then used to create a model with the same machine learning method to evaluate the resultant model. The pre-processing of NHANES data is used to produce 15,712 instances for experimental evidence.

There is no missing value in the data method. The input data is erratically substituted by 0 depending on the noise factor for investigations. Each prototype is therefore supplied with the input information with a simulated missing value and actual missing value. Each of the models is generated by training data which is assessed by the test dataset. Apply the model created, which produces missing statistics randomly, and the distortion factor rises from 0.05 to 0.30 by 0.05 to the arbitrary missing information value. A 0.05 distortion factor shows that the missing data created randomly is 5%.

To compute the margin of error, the missing data is compared with the original. The actual rate is the initial trial information in the error measurement, and the projected value is the recovered data. The trial information input into every model is t, the output information is t^, and the correct response is t. The error is assessed by utilizing t as the authentic value and t^ as the projected value of the output data. Each model is tenfold evaluated according to the noise factor for performance evaluation, and the average result for the precision is computed.

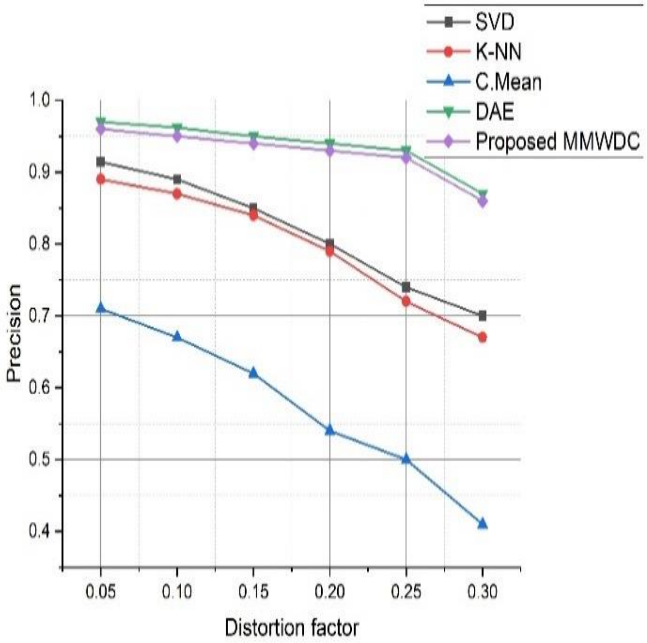

Figure 3 Comparison of various methods to estimate the missing data

The evaluation outcomes are given in Fig. 3 based on the methods used to approximate the missing information. The figure also displays the precision for each model based on the distortion feature of the missing information estimate. The precision is inferior for the c.mean method. SVD and K-NN give moderate accuracy. The precision of the proposed MMWDC is maximum, thereby showing its superior performance. Also, as the distortion factor increases, the precision drops irrespective of the method used to estimate missing information. A matching and a cluster analysis analyze the properties of the parameters.

A mixture of parameters is used to build multi-modal data using every estimating method. On this basis, you may produce three datasets: Multi-modal data based on similarities, multi-modal data based on clusters, and multi-modal data constructed on categories. In the study, we have thus constructed and assessed suitable models like an Integrated multi-modal weighted denoise autoencoder (IM WDAE), a group-based weighted denoise (GM WDAE) autoencoder, and an organized multi-modal, weighted denoise autoencoder (OM WDAE). The OM WDAE is a hierarchical architectural multi-modal data model. OM WDAE uses NHANES-classified categories to create multi-modal variables.

The number of hidden layers is set to 1/2 of input nodes to evaluate the models correctly. The Epoch values of each model are different. The maximum time is 50, and the time takes place at a concurrent time. IM WDAE has 32-epoch learning, 29-epoch learning for GM WDAE, and 28-epoch learning for OM WDAE. The analysis of similarities is a way to determine if the parameters are positive or negative. These connections are shown within a range of -1 to C1. A value nearer to -1 implies that a more malicious link moves, whereas a value nearer to C1 means that the relationship moves more in the same orientation. Figures closer to zero show a lesser reciprocal influence. A coefficient of Pearson is employed in this research, a commonly recognized and straightforward correlation coefficient. Furthermore, for user comfort and different sorts of data, several alternative correlation coefficients are accessible. Cluster analysis involves clustering a collection of vector space variables near each other.

Fig.4 shows the comparison among various modes in the WDAE for the proposed MMWDC framework. All criteria will save the number of hidden nodes, stay the same throughout multi-modal autoencoder education, reliant on the constraint setup. The multi-modal weighted self-encoder is made up of three hidden layers. To this aim, the total number of weights must be modified and each unique mode integrated. Fig. 4 displays the study findings of the autoencoder layered multimodally. The input, hidden, and the parameters in Fig. 4 define output nodes of the auto-encoder, and the OM looks to be the least that indicates that this model necessitates the least amount of resources to be used. In addition, OM displays maximum noise precision as well. This can be explained by considering real-world conditions in the groups classified in the NHANES data. If an unspecified class dataset is used to build multi-modified data, multiple results might be achieved.

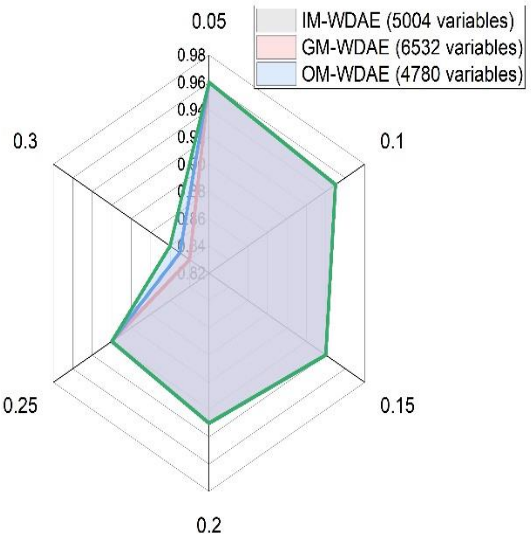

Figure 4 Comparison among various modes in the WDAE for the proposed MMWDC framework

5 CONCLUSION

The paper suggested a method for predicting missing data in big health data with the use of MMWDC. A weighted denoising automatic encoder is used to predict incomplete information during data collecting and processing phases. The suggested approach employs autoencoders based on neural networks whose output values are estimated based on their input values. Data from the NHANES have been utilized in this work. Based on this approach, missing data in the IHA can be estimated.

The weighted multi-modal autoencoder system is trained using the variables categorized with a noise factor, which is more than the accurate data estimation techniques other than the usual SVD, K-NN, and the mean column. The accuracy of the proposed system is 0.9531 when the distortion factor is 0.15. This shows that the multi-modal WDAE is more suited to a individual device than for a single-modal feature as the number of multi-modal feature constraints is almost half, with just a minor change in precision. All missing data may be made of health information depending on the user's specific scenario. In such a scenario, the missing data may be predicted using an autoencoder that has already been learned. Furthermore, depending on the set of the variables, a change in learning efficiency arises. A model setup that demonstrates the variable features enhances the learning efficiency and makes it possible to create more accurate neural network models.

References

[1] A. Gumaei, M. M. Hassan, A. Alelaiwi and H. Alsalman, "A hybrid deep learning model for human activity recognition using multi-modal body sensing data," IEEE Access, Vol. 7, pp. 99152–99160, 2019.

[2] Available online: https://www.cdc.gov/nchs/nhanes/ index.htm

[3] C. Cadena, A. Dick and I. D. Reid, "Multi-modal auto-encoders as joint estimators for robotics scene understanding," Proc. 12th Robot., Sci. Syst.,2016.

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

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

[6] H. Jung, J. Yang, J.-I. Woo, B.-M. Lee, J. Ouyang, K. Chung, and Y. Lee, "Evolutionary rule decision using similarity based associative chronic disease patients," Cluster Comput., Vol. 18, No. 1, pp. 279–291, 2015.

[7] H. Yoo and K. Chung, "Heart rate variability based stress index service model using bio sensor," Cluster Comput., Vol. 21, No. 1, pp. 1139–1149, 2018.

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

[9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, “Data Mining: Practical Machine Learning Tools and Techniques”, Morgan Kaufmann, 2016.

[10] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," J. Roy. Stat. Soc. C(Appl. Statist.), Vol. 28, No. 1, pp. 100–108, 1979.

[11] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient,"Noise Reduction in Speech Processing. Berlin, pp. 1–4, 2009.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Netw., Vol. 61, pp. 85–117, 2015.

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

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

[15] J.-C. Kim and K. Chung, "Multi-modal weighted denoising autoencoder for handling missing data in healthcare big data," IEEE Access, Vol. 8, pp. 104933–104943, 2020.

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

[17] J.-W. Baek, J.-C. Kim, J. Chun and K. Chung, "Hybrid clustering based health decision-making for improving dietary habits," Technol. Health Care, Vol. 27, No. 5, pp. 459–472, 2019.

[18] K. Chung and H. Yoo, "Edge computing health model using P2Pbased deep neural networks," Peer-to-Peer Netw. Appl., vol. 13, no. 2, pp. 694–703, Apr. 2019.

[19] K. Chung and R. C. Park, "PHR open platform based smart health service using distributed object group framework," Cluster Comput., Vol. 19, No. 1, pp. 505–517, 2016.

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

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

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

[23] M. Shang, X. Luo, Z. Liu, J. Chen, Y. Yuan, and M. Zhou, "Randomized latent factor model for high-dimensional and sparse matrices from industrial applications," IEEE/CAA J. Automatica Sinica, Vol. 6, No. 1, pp. 131–141, 2019.

[24] R. Agrawal and R. Srikant, "Mining sequential patterns," Proc. 11th Int. Conf. Data Eng, pp. 3–14,1995.

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

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

[27] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: A survey," IEEE Trans. Neural Netw., Vol. 13, No. 1, pp. 3–14, 2002.

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

[29] X. Luo, M. Zhou, S. Li, and M. Shang, "An inherently nonnegative latent factor model for high-dimensional and sparse matrices from industrial applications," IEEE Trans. Ind. Informat., vol. 14, no. 5, pp. 2011–2022, 2018.

[30] X. Luo, M. Zhou, Z. Wang, Y. Xia, and Q. Zhu, "An effective scheme for QoS estimation via alternating direction method-based matrix factorization," IEEE Trans. Services Comput., Vol. 12, No. 4, pp. 503–518, 2019.

[31] S. García, J. Luengo, and F. Herrera, “Data Preprocessing in Data Mining”, Springer, pp. 59–139, 2015.