Journal of Theoretical and Applied Information Technology
31st March 2018. Vol.96. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DESIGN AND IMPLEMENTATION OF THE EXPERT SYSTEM FOR HEALTH AND MEDICAL TREATMENT USING INTEGRATION OF BIG DATA

[1]YOO-JIN MOON, [2]EUNJOO CHOI, [3]YOUNG-HO HWANG

[1 2]Hankuk University of Foreign Studies, Seoul 02450, Korea

[3] Corresponding Author, Kunsan National University, Jeonbuk 54150, Korea

E-mail: [1]yjmoon@hufs.ac.kr, [2]lovelybeeeun@naver.com, [3]yhwang@kunsan.ac.kr

ABSTRACT

This paper designs and implements the expert system for health and medical treatment using integration of public medical big data. According to the trend that the health and medical treatment data are gradually open to the public in the world, the integrated medical information system is required for the patients. By implementation of the medical expert system, representatives of the analytical results are average treatment expenses per medical center per disease, medical institutions frequently treating patients per disease, diseases susceptibility in the age group, number of patients per disease by region, and number of the medical clinics per department and district. The expert system makes the big data helpful for the medical centers, patients, governments, the medicine importers, pharmacists and insurance agencies etc. Overall, the expert system deals with not only the fundamental analysis of the diseases per age or per gender but also the practically beneficial information for all the interested parties with analysis of the public medical treatment big data. Furthermore, the expert system can easily be extended with increasing public medical big data open.

Keywords: *Big data, Expert system, Medical expense, Insurance agency, Medical center*

1. INTRODUCTION

Big data threw one of the major challenges of information technology and statistics and has brought a great revolution from the algorithmic viewpoint and statistical framework. [1, 2, 3] Big data usually consist of massive data but they also include online data with homogeneity and heterogeneity. Recently some statistical methods have been adapted to process big data, like linear regression models, bootstrapping schemes and clustering schemes. [4, 5, 6] They are utilized with comprehensive nonparametric statistical methods working not only in single and versatile framework regression problems, but also in two-class and multi-class classification problems. [7, 8, 9, 10, 11]

Big data is quite an old term, but with advancement of computer processors and memories big data have become very popular recently in life sciences, health and medical area, security area, marketing area, agriculture area and other areas. The health and medical area has been a front-running man from large biomedical data, with the U.S. healthcare system expected to reach the zettabyte (10^{21}) size from electronic medical records to research articles in medical journals. [10, 12, 13, 14] The biomedical data have explosively produced in the hospital as well as on the mobile internet. The heterogeneous biomedical data are created from various types of data sources, which include the hospital medical data, the health data and personal behavior data published by the mobile devices, and nutrition information from social network. [15] These health big data play an important role in the data processing and analysis useful for the patients. [16, 17, 18, 19] In addition, the health big data also give chances for insurance companies, medical administrators of the hospitals and the government to take the implication into account when deciding medical management, medical facilities, medical policy and behavior. [20]

To find earlier detection of pathologies in a typical check mode of healthcare facilities, sensors connected to patients can send signals to the patients from time to time in measuring abnormal frequency physiological data. [20, 21, 22] The early detection of pathologies makes it possible for doctors to recognize the disease, start treatments earlier, potentially to let the patient have a better condition for treatment, and possibly to let the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

patient completely cured with reduced costs. [23, 24, 25] When the medical facilities generate the heterogeneous data with real-time restrictions, a big data problem may occur during processing medical algorithms for diagnosing pathologies in the real-time big data analytics system. [6, 10]

In Korea, the public medical big data are available in the websites of Public Big Data Center of Seoul City [26], Institute of Health Insurance Policy [27], National Cancer Information Center, [28], Korea Breast Cancer Cyber Center [31], and Samsung Seoul Hospital [34], etc. The public medical big data in each center were respectively processed without integration of the other public big data.

Thus, with integration of public medical big data it is necessary to implement the expert system for health and medical treatment. People would use the expert system to rationally select the medical centers, check the medical expenses and decide better treatment method, etc. The expert system creates various information through public medical big data and makes the big data helpful for the medical centers, patients, governments, the medicine importers, pharmacists and insurance agencies. [11, 29] Overall, the expert system deals with not only the fundamental analysis of the diseases per age or per gender but also the practically beneficial information with analysis of the public medical treatment big data for all the interested parties. Furthermore, the expert system can be easily extended with increasing public medical big data open.

2. PRELIMINARIES

Gradually data of the health and medical treatment area have been open to the public and developed based on information technology advancement. Especially the developed countries publicly disclose the data centered on fields of the socially high demands, which is preferentially on the way of implementing the database system. Express Scripts Holding Co., company of management for insurance and drugs, utilized big data during information processing and achieved the good result of methods for dispensing the cheaper personalized drugs for the patient. [3, 10] And McKinsey and Beyondcore conducted case analysis of the medical expense invoices and found points of decreasing medical expenses for the patients. [30]

Recent publicity of the medical big data is expected to motivate the market vitalization of the health and medical treatment area, but the public big data for the medical area are limited to the respective data processing in each center without integration of the public medical big data yet. In Korea, National Health Insurance Corporation and Daum Soft started a platform for providing people with services to forecast care of national health, which stage is still in its infancy. According to the trend that the health and medical treatment data are gradually open to the public in the world, the integrated medical information system is required for the patients. [26, 27, 28, 29, 31, 33] Thus, it is meaningful that this paper designs and implements the information system for health and medical treatment using integration of public medical big data.

3. THE SYSTEM MODEL AND ANALYSES

3.1 Entity-Relationship Diagram of the System

The research system embraces an entity-relationship model. In the entity-relationship model, an entity represents one theme or topic restricted to things that can be conveyed by a single table and a relation consists of rows and columns that have specific characteristics. In this system, the main characteristics of the relation are that cells of the relation hold a single value, the order of the rows is unimportant, the order of the columns is unimportant, each column has a unique name, a key is one or more columns of a relation used to determine a row. [1] A primary key (PK) is a candidate key chose to uniquely define a single row for the relation. By normalization of the entities and relations the functional dependencies work in the well-formed relations of the system.

Figure 1 illustrates the Entity-Relationship diagram architecture of the expert system. For the implementation of the expert system the public big data related to the medical treatment are collected from Public Big Data Center of Seoul City [26], Institute of Health Insurance Policy [27], National Cancer Information Center [28], Korea Breast Cancer Cyber Center [31], Samsung Seoul Hospital [34] and National Statistical Office [32].

The database system uses public medical big data and is composed of the following medical information: patient information, medical institution information, disease information and medical treatment information.

In the case of a medical care E-R diagram in Figure 1, each medical number is created and designated as a PK. We need to put information about the patient number, the medical institution number, the disease number, the treatment method number, fee, the treatment area, the start year of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

treatment, the recent medical year, full recovery or not, and the disease name, etc.

For the disease entity, its PK is the disease number, and the other attributes are the disease name, the relapse rate, and the five-year survival rate on the entity. In this case, we put information about cancer, cerebropathia and cardiac disorder. And the medical institution entity makes the medical institution number for PK and puts information about the medical institution name, the type, and the area.

For the method of treatment entity, the PK is the number of treatment type and the disease name. And the other attribute is the treatment method name.

For the patient entity, PK is the patient number. And the other attributes are the patient's name, the age, the sex, the address, information about death or not.

Based on the Entity-Relationship diagram of the database system in Figure 1, the expert system is processed using ORACLE and R with preprocessing by HADOOP.

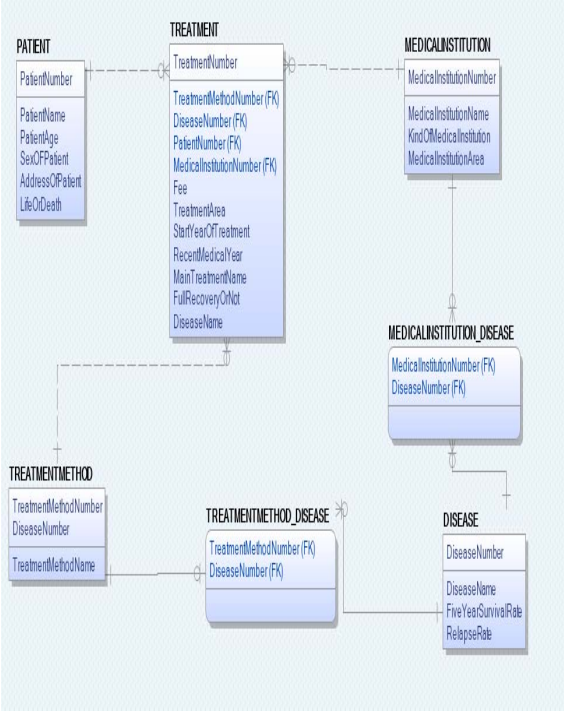


Figure 1: Entity-Relationship Diagram Architecture of the Expert System

3.2 Implementation and Analysis of the System

With implementation of the medical expert system, the various analysis methods can answer the sophisticated questions the patients want to ask, for example, comparison of treatment expenses, medical institutions frequently treating patients per disease, diseases susceptibility in the age group, and number of the medical clinics per department and district.

In addition, it can produce the actual number of patients currently suffering from the disease, allowing the government to calculate the people's insurance premiums and helping to identify those who need physical assistance of the medical sponsor. Furthermore, the number of patients per disease by region can be analyzed in real time, so if the number of patients with a specific disease in a specific area suddenly increases, the government can promptly investigate the situation and the medical institutions in the area can quickly make preparation for the situation as well.

3.2.1 Top 5 in medical institutions frequently treating patients per disease

Analysis from the health and medical system can show the order of medical institutions frequently treating patients per disease. For example, Figure 2 illustrates Top 5 in medical institutions frequently treating patients for the liver cancer, which are Severance Hospital of Yonsei University, Seoul Asan Hospital, Samsung Seoul Hospital, Seoul National University Hospital and Hanyang University Hospital in the order of frequent treatments.

In the side of liver cancer patients, they may want to know the medical institutions which have many experiences with treatment of the liver cancer. In the side of the medical institutions, they can increase or decrease the medical facilities, doctors and personnel of the liver cancer center. They may also inform people of the frequent treatments for the liver cancer and advertise the center.

```
Select MEDICALINSTITUTION.MedicalInstitutionName,
       DISEASE.DiseaseName,

From DISEASE, TREATMENT, MEDICALINSTITUTION

Where TREATMENT.MedicalInstitutionNumber
      =MEDICALINSTITUTION.MedicalInstitutionNumber
And   DISEASE.DiseaseName = 'LiverCancer'

Group By   DiseaseName,
           MedicalInstitutionName
Order By   DiseaseName,
           count( TREATMENT.PatientNumber ) Desc
Limit  5 ;
```

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MedicalInstitutionName	DiseaseName
Severance Hospital of Yonsei University	LiverCancer
Seoul Asan Hospital	LiverCancer
Samsung Seoul Hospital	LiverCancer
Seoul National Niversity Hospital	LiverCancer
Hanyang University Hospital	LiverCancer

Figure 2: Top 5 in Medical Institutions Frequently Treating Patients for the Liver Cancer

3.2.2 Average medical expenses per medical center for liver cancer

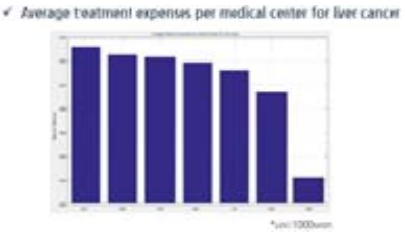
Patients may want to compare average medical expenses per medical center for his disease. Figure 3 illustrates average treatment expenses in a medical center. A medical center A36 costs Won 55,568,000 for liver cancer but the other medical center A12 charged more cost for the same disease with Won 74,295,000, which will be useful for the patient to select the medical center.

```
Select MEDICALINSTITUTION.MedicalInstitutionName,
       DiseaseName,
       round( avg(Fee)) as AverageCost

From TREATMENT inner join MEDICALINSTITUTION
On TREATMENT.MedicalInstitutionNumber
   =MEDICALINSTITUTION.MedicalInstitutionNumber

Where     DiseaseName = 'LiverCancer'

Group By  DiseaseName,
          MedicalInstitutionName;
```



Unit: Thousand Won

MedicalInstitutionName	DiseaseName	AverageCost
A52	LiverCancer	51,278
A02	LiverCancer	49,326
A36	LiverCancer	55,568
A12	LiverCancer	74,295
A45	LiverCancer	55,532
A01	LiverCancer	43,175
A13	LiverCancer	44,851
A19	LiverCancer	47,330
A20	LiverCancer	48,753

Figure 3: Average treatment expenses per medical center for liver cancer

3.2.3 Number of the medical clinics per department and district

Figure 4 shows that there are 373 Oriental medical clinics in Gangnam District and 290 in Seocho District, which is helpful for deciding a place for the doctor to open his clinic and extend a new branch to a new district.

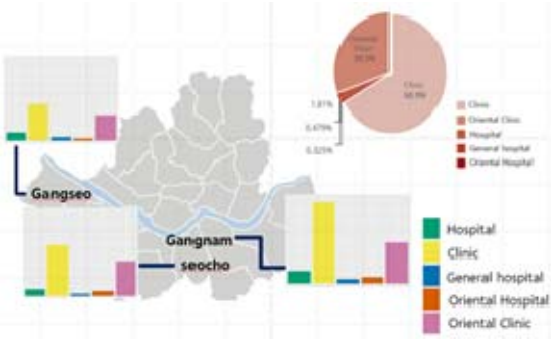
In the side of the medicine importers and pharmacists, they can recognize demands of the medicine and facilities per department and district and arrange their salespersons according to the number of the medical centers in the district.

```
Select MEDICALINSTITUTION.MedicalInstitutionArea,
       KindOfMedicalInstitution,
       count( KindOfMedicalInstitution)
         as NumberOfKindOfMedicalInstitution

From MEDICALINSTITUTION

Group By
     MEDICALINSTITUTION.KindOfMedicalInstitution,
     MEDICALINSTITUTION.MedicalInstitutionArea

Order By
 MEDICALINSTITUTION.KindOfMedicalInstitution desc,
 count(MEDICALINSTITUTION.KindOfMedicalInstitution) desc;
```



Journal of Theoretical and Applied Information Technology
31st March 2018. Vol.96. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MedicalInstitutionArea	KindOfMedicalInstitution	NumberOfInstitution
Gangnam	Oriental Clinics	373
Seocho	Oriental Clinics	290
Songpa	Oriental Clinics	231
Dongdaemun	Oriental Clinics	174
Gangdong	Oriental Clinics	167
Gangseo	Oriental Clinics	161
Nowon	Oriental Clinics	157
Gwanak	Oriental Clinics	151
Yangchun	Oriental Clinics	140

Figure 4: Number of the Medical Clinics per Department and District

3.2.4 Average medical expenses in each Seoul district for pancreatic cancer

Patients may want to compare average medical expenses in each Seoul District for his disease. Figure 5 illustrates that average treatment expense in Seodaemun District, one of Seoul districts, is Won 66,201,000 for pancreatic cancer and that in Gangnam District is Won 65,833,000, which will be useful for the patient to select the medical center. If a patient thinks the cheaper medical center with the average medical expenses for his disease, he can consider the medical center of the district close to his district.

In the side of the medical institutions, they can set the medical expenses to the acceptable average price comparing with average district prices.

```
Select TreatmentArea, DiseaseName,
       round( avg(Fee)) as AverageCost

From TREATMENT

Where      DiseaseName = 'PancreaticCancer'

Group By   TreatmentArea, DiseaseName,

Having     count(DiseaseName) >= 2
Order By   DiseaseName,
           Round( avg(Fee) ) desc;
```

	TreatmentArea	DiseaseName	AverageCost
▶	Seodaemungu	PancreaticCancer	66201
	Yeongdeungpogu	PancreaticCancer	65890
	Gangnamgu	PancreaticCancer	65833

Figure 5: Average treatment expenses in each Seoul district for pancreatic cancer

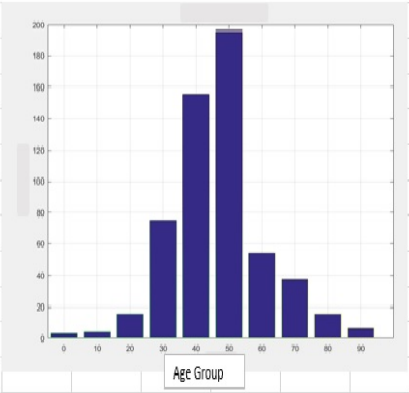
3.2.5 Number of the patients according to age group

People may exert themselves to prevent from diseases susceptible to their age group and specifically focus on examination of the diseases susceptible to the age group. Figure 6 illustrates the number of the patients according to age group. There are 15,000 patients in the twenties, 75,000 in the thirties, 155,000 in the forties and 196,000 in the fifties.

In the side of the medical institutions, they may decide the medical services and the place of the medical facilities by recognizing the number of the patients according to age group. In the side of the insurance company, they can develop the insurance product according to the age group and decide to increase the insurance fee for the frequent diseases corresponding to the age group.

```
Select
Sum(Case When PatientAge <20
              then 1 else 0 end) as UnderTwenties,
Sum(Case When PatientAge >=20 and PatientAge <30
              then 1 else 0 end) as Twenties,
Sum(Case When PatientAge >=30 and PatientAge <40
              then 1 else 0 end) as Thirties,
Sum(Case When PatientAge >=40 and PatientAge <50
              then 1 else 0 end) as Fourties,
Sum(Case When PatientAge >=50 and PatientAge <60
              then 1 else 0 end) as Fifties,
Sum(Case When PatientAge >=60
              then 1 else 0 end) as BeyondSixties

From  PATIENT ;
```



Unit: Thousand

UnderTwenties	Twnties	Thirties	Fourties	Fifties	BeyondSixties
7	15	75	155	196	112

Figure 6: Number of the Patients according to Age Group

Journal of Theoretical and Applied Information Technology
31st March 2018. Vol.96. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2.6 Number of the patients per disease

People may specifically focus on examination of the diseases according to frequency of the disease. Figure 7 provides the number of the patients per disease. In the order of frequent diseases, there are about 47,000 patients of the lung cancer, 46,000 of the gastric cancer, 32,000 of colorectal cancer, and 7,000 of cerebral infarction, etc.

In supporting medical treatment, by recognizing diseases with many patients realistic and effective supports can be organized.

```
Select  DISEASE.DiseaseName,
        count(TREATMENT.PatientNumber) as NumberOfPatient

From DISEASE, TREATMENT, PATIENT

Where TREATMENT.PatientNumber = PATIENT.PatientNumber
And   DISEASE.DiseaseNumber = TREATMENT.DiseaseNumber

Group By   DiseaseName,

Order By    count( TREATMENT.PatientNumber ) Desc
```

Unit: Thousand

DiseaseName	NumberOfPatient
LungCancer	47
GastricCancer	46
ColorectalCancer	32
LiverCancer	28
PancreaticCancer	13
Cerebral Infarction	7
Myocardial Infarction	7
Cardiac Insufficiency	6
Cerebral Thrombosis	6

Figure 7: Number of the Patients per Disease

3.2.7 Percentage of the patients per residential district

Environments including the residential area may affect the patients. Figure 8 shows percentage of the patients per district of the residence, which divides the patients of the district by the total patients. There are 6.25% patients in Nowon District, 5.89% in Gangseo District, and 5.35% in Eunpyung District, etc.

In the side of the medical institutions, they can provide residents with the number of acceptable patients and select the place where the new branch will operate successfully. In the side of the insurance companies, they can decide whether to increase their agencies and sales persons or not.

```
Select AddressOfPatient,
       count(AddressOfPatient) /
         (Select count(AddressOfPatient)
          From PATIENT)
        *100 as PercentageOfPatients

From PATIENT

Group By AddressOfPatient

Order By    count(AddressOfPatient ) Desc ;
```

Unit: %

AddressOfPatient	PercentageOfPatients
Nowon District	6.2500
Gangseo District	5.8929
Eunpyung District	5.3571
Seodaemun District	5.3532
Dongdaemun District	5.0213
Songpa District	5.0000
Gwangak District	4.8214
Seongbook District	4.6429
Gangbook District	4.4643
Gangnam District	4.2857

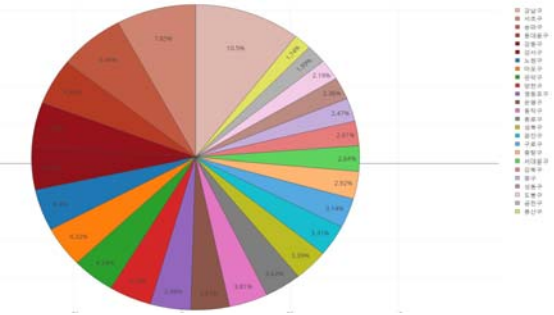


Figure 8: Percentage of the Patients per Residential District

3.2.8 The order of diseases susceptibility in the fifties

Many cancers and diseases start in the forties and fifties of the people. People may exert themselves to prevent the diseases according to their age. Figure 9 illustrates the order of diseases susceptibility in the fifties, which are colorectal cancer, lung cancer, thyroid cancer and angina in the order of frequent diseases.

.

Journal of Theoretical and Applied Information Technology
31st March 2018. Vol.96. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

```
Select  DISEASE.DiseaseName,
        count(TREATMENT.TreatmenttNumber) as NumberOfPatient

From DISEASE, TREATMENT, PATIENT

Where TREATMENT.PatientNumber = PATIENT.PatientNumber
And   DISEASE.DiseaseNumber = TREATMENT.DiseaseNumber
And   PATIENT.PatientAge >=50 And PATIENT.PatientAge  <60

Group By   DiseaseName,

Order By   count( TREATMENT.TreatmentNumber ) Desc,
           DISEASE.DiseaseName ;
```

Unit: Thousand

DiseaseName	NumberOfPatient
ColorectalCancer	110
LungCancer	55
ThyroidCancer	31
AnginaPectoris	13
HeartAttack	10
PancreaticCancer	9
Cardiomyopathy	7
valvular disease of the heart	7
RenalCancer	5
basal cell carcinoma	1

Figure 9: The Order of Diseases Susceptibility in the Fifties

3.2.9 Top 5 in high patient ratios of disease in Gangnamgu

Figure 10 provides high percentage of patients in Gangnamgu in the order of thyroid cancer, stomach cancer, colon cancer, prostate cancer, and lung cancer. It also can provide information about diseases with a high percentage of patients in each borough.

In the side of the medical institutions, if there are many patients of a certain disease in a district they can recognize demands of the medical doctors and facilities for the center and decide whether to increase them or not. Even if a medical institution does not have the center for the disease with many patients, it can decide whether to open its center or not. In the side of government, if there are many patients with a certain disease in a district, the government can start to investigate its causes.

```
Select TREATMENT.TreatmentArea,
       DISEASE.DiseaseName,
       count( TREATMENT.PatientNumber) /
         (Select count(PATIENT.AddressOfPatient) from PATIENT
          Where TREATMENT.TreatmentArea='Gangnamgu')
       as PatientRatio

From DISEASE,  TREATMENT

Where TREATMENT.TreatmentArea='Gangnamgu' AND
       DISEASE.DiseaseName=TREATMENT.DiseaseName

Group By
       DISEASE.DiseaseName

Order By count( TREATMENT.PatientNumber) desc
limit 5;
```

TreatmentArea	DiseaseName	PatientRatio
Gangnamgu	ThyroidCancer	4.8214
Gangnamgu	GastricCancer	4.1071
Gangnamgu	ColonCancer	3.3929
Gangnamgu	ProstateCancer	2.6786
Gangnamgu	LungCancer	1.7857

Figure 10: High Patient Ratio Of Disease In Gangnamgu

3.2.10 Trends of increasing patients per disease

It is helpful for the medical institution to recognize trends of increasing patients per disease. Figure 11 illustrates trends of increasing patients per disease. For the liver cancer, the new patients are steadily increasing during 2011 – 2015 from 10,000 patients to 14,000.

In the side of the medical institutions, they need to increase or decrease the medical facilities, doctors and personnel of the liver cancer and let increasing patients be treated properly in the center without sending them to the other institutions. In the side of the medicine importers and pharmacists, they can recognize demands of the increasing patients and prepare medicine and facilities by increasing importing and production.

```
Select  DiseaseName,
        StartYearOfTreatment,
        count( PatientNumber) as NumberOfNewPatients

From TREATMENT

Group By   DiseaseName,
           StartYearOfTreatment

Having     count(PatientNumber) >1
Order By   DiseaseName ;
```

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

		Unit: Thousand
DiseaseName	StartYearOfTreatment	NumberOfNewPatients
LiverCancer	2011	10
LiverCancer	2012	11
LiverCancer	2013	12
LiverCancer	2014	15
LiverCancer	2015	14
ThyroidCancer	2011	21
ThyroidCancer	2012	24
ThyroidCancer	2013	25
ThyroidCancer	2014	24
ThyroidCancer	2015	23
CerebralInfarction	2011	3
CerebralInfarction	2012	3
CerebralInfarction	2013	4
CerebralInfarction	2014	3
CerebralInfarction	2015	5
CerebralAneurysm	2011	3
CerebralAneurysm	2012	4

Figure 11: Trends of Increasing Patients per Disease

In addition, other information related to the patients, medical treatment and disease can be extracted through the expert system.

3.3 Contributions of the Research System

Until now, the public big data for the medical area are limited to the respective data processing in each center without integration of the public medical big data yet. In Korea, only with national health insurance treatments information National Health Insurance Corporation and Daum Soft started a platform for providing people with services to forecast care of national health, which stage is still in its infancy. However, recent publicity of the medical big data is expected to motivate the market vitalization of the health and medical treatment area using integration of public medical big data. Thus, the research system using integration of public medical big data will contribute to the market vitalization of the health and medical treatment area.

According to the trend that the health and medical treatment data are gradually open to the public in the world, the integrated medical information system will also contribute to the decision and selection of the patients for the better treatment. Thus, it is meaningful that this paper designs and implements the information system for health and medical treatment using integration of public medical big data.

4. DISCUSSION AND CONCLUSIONS

In the 4th Industrial Revolution, big data of health and medical treatment can be implemented into the intelligent information system and the expert system. This paper designed and implemented the expert information system for health and medical treatment utilizing integration of public medical big data.

According to the trend that the health and medical treatment data are gradually open to the public in the world, the integrated medical information system for the patients is required. People want to use the publicly available big data to rationally select the medical centers with sophisticated experiences, compare treatment expenses and decide better treatment method. With implementation of the medical expert system, representatives of the analytical results were average treatment expenses per medical center per disease, medical institutions frequently treating patients per disease, diseases susceptibility in the age group, number of patients per disease by region, and number of the medical clinics per department and district, etc. Thus, it is meaningful that this research designed and implemented the expert system for health and medical treatment using integration of public medical big data.

The expert system collated various information utilizing integration of public medical big data and made the data helpful for the medical centers, patients, governments, the medicine importers, pharmacists and insurance agencies etc. Consequently, it dealt with not only the fundamental analysis of the diseases per age or per gender but also the practically beneficial information for all the interested parties with analysis of the public medical treatment big data. Furthermore, the expert system can be easily extended with increasing public medical big data open.

The limitation of the paper is that because of the security problems and protection of personal information problems many centers avoid opening the medical data publicly. Thus, the expert system utilizes the medical big data opened by the centers, not but the full version of the medical big data. In the future, the expert system should be implemented to take care of security problems and to work without hindering protection of personal information.

5. ACKNOWLEDGMENTS

This work was supported by Hankuk University of Foreign Studies Research Fund of 2017.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES:

[1] David Kroenke and David Auer, *Database Concepts*, Pearson, 2015.

[2] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, Wei Dong, "Data Mining with Big Data", *IEEE Transactions on Knowledge & Data Engineering,* Vol. 26, No. 1, 2014, pp. 97-107.

[3] Akram Alyass, Michelle Turcotte, and David Meyre, "From Big Data Analysis to Personalized Medicine for All: Challenges and Opportunities", *BMC Medical Genomics*, Vol. 8, No. 33, 2015, pp. 1-12.

[4] Navroop Kaur and Sandeep K. Sood, "Efficient Resource Management System Based on 4Vs of Big Data Streams", *Big Data Research*, Vol. 9, Sept., 2017, pp. 98-106.

[5] Alok R. Saboo, V. Kumar, and Insu Park, "Using Big Data to Model Time-Varying Effects for Marketing Resource (Re)Allocation", *MIS Quarterly*, Vol. 40, No. 4, 2016, pp. 911-940.

[6] Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian, "Big Data Analytics in Healthcare", *BioMed Research International*, Vol. 15, No. 1, 2015, pp. 1-16.

[7] Ali Kalantari, Amirrudin Kamsin, Halim Shukri Kamaruddin, Nader Ale Ebrahim, Abdullah Gani, Ali Ebrahimi, and Shahaboddin Shamshirband, "A Bibliometric Approach to Tracking Big Data Research Trends", *Journal of Big Data*, Vol. 4, No. 30, 2017.

[8] Robin Genuer, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix, "Random Forests for Big Data", *Big Data Research*, Vol. 9, Sept., 2017, pp. 28-46.

[9] Guoliang He, Yong Duan, Rong Peng, Xiaoyuan Jing, Tieyun Qian, and Lingling Wang, "Early Classification on Multivariate Time Series", *Neurocomputing*, Vol. 149, 2015, pp. 777-787.

[10] Y. Liang and A. Kelemen, "Big Data Science and its Applications in Healthcare and Medical Research: Challenges and Opportunities", *Austin Biometrics and Biostatistics*, Vol. 3, No.1, 2016, pp. 1-9.

[11] Andreas P. Plageras, Christos Stergiou, George Kokkonis, Kostas E. Psannis, Yutaka Ishibashi, Byung-Gyu Kim, and B. Brij Gupta, "Efficient Large-scale Medical Data (eHealth Big Data) Analytics in Internet of Things", *2017 IEEE 19th Conference on Business Informatics (CBI)*, Thessaloniki, Greece, July 24-27, 2017, pp. 21-27.

[12] Helena Aidos and Ana Fred, "Discrimination of Alzheimer's Disease Using Longitudinal Information", *Data Mining and Knowledge Discovery*, Vol. 31, No. 4, 2017, pp. 1006-1030.

[13] Emilie Baro, Samuel Degoul, Regis Beuscart, and Emmanuel Chazard, "Toward a Literature-Driven Definition of Big Data in Healthcare", *BioMed Research International*, Vol. 15, No. 2, 2015, pp. 1-9.

[14] Shein-Chung Chow and Yuanyuan Kong, "On Big Data Analytics in Biomedical Research", *Journal of Biometrics & Biostatistics*, Vol. 6, No. 236, 2015, pp. 1-10.

[15] Vasileios Lampos, Elad Yom-Tov, Richard Pebody, and Ingemar J. Cox, "Assessing the Impact of a Health Intervention via User-Generated Internet Content", *Data Mining and Knowledge Discovery*, Vol. 29, No. 5, 2015, pp. 1434-1457

[16] Yelena Yesha, Vandana P. Janeja, Naphtali Rishe, and Yaacov Yesha, "Personalized Decision Support System to Enhance Evidence Based Medicine through Big Data Analytics, *2014 IEEE International Conference on Healthcare Informatics (ICHI)*, Verona, Italy, Sept. 15-17, 2014, pp. 376-376.

[17] Daniel Grunwell and Tony Sahama, "Information Accountability and Health Big Data Analytics: A Consent-Based Model", *2015 17th International Conference on E-Health Networking, Applications and Services (HealthCom.)*, Boston, USA, Oct. 14-17, 2015, pp. 195-199.

[18] Jingquan Li, "Open Medical Big Data and Open Consent and Their Impact on Privacy", *2017 IEEE International Congress on Big Data (BigData Congress)*, Honolulu, USA, June 25-30, 2017, pp. 511-514.

[19] Sharona Hoffman, "The Promise and Perils of Open Medical Data", *The Hastings Center Report*, Vol. 46, No.1, 2016, pp. 6-7.

[20] Jianwei Liu, Yong Zhang, and Chunxiao Xing, "Medical Big Data Web Service Management Platform", *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, San Diego, USA, Jan. 30-Feb.1, 2017, pp. 316-320.

[21] Zhengxing Huang, Wei Dong, Peter Bath, Lei Ji, and Huilong Duan, "On Mining Latent Treatment Patterns from Electronic Medical Records", *Data Mining and Knowledge Discovery*, Vol. 29, No. 4, 2015, pp. 914-949.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[22] Amit Sheth, "Smart Data - How You and I will Exploit Big Data for Personalized Digital Health and Many Other Activities", *2014 IEEE International Conference on Big Data*, Washington DC, USA, Oct. 27-30, 2014, pp. 2-3.

[23] Kristina Ibáñez, Maria Guijarro, Gonzalo Pajares, and Alfonso Valencia, "A Computational Approach Inspired by Simulated Annealing to Study the Stability of Protein Interaction Networks in Cancer and Neurological Disorders", *Data Mining and Knowledge Discovery*, Vol. 30, No. 1, 2016, pp. 226-242.

[24] Theodoros Mavroeidakos, Nikolaos Tsolis, and Dimitrios D. Vergados, "Centralized Management of Medical Big Data in Intensive Care Unit: A Security Analysis", *2016 3rd Smart Cloud Networks & Systems (SCNS)*, Dubai, UAE, Dec. 19-21, 2016, pp. 1-5.

[25] Javier Balladini, Claudia Rozas, Emmanuel Frati, Nestor Vicente, and Cristina Orlandi, "Big Data Analytics in Intensive Care unit: Challenges and Applicability in an Argentinian Hospital", *Journal of Computer Science & Technology*, Vol. 15, No. 2, 2015, pp. 61-69.

[26] Medical Institutions (Seoul City Public Data), http://data.seoul.go.kr/, 12/05/2017.

[27] Patient Age, Sex, National Health Insurance Service, 2011. http://www.kihasa.re.kr/common/filedown.do?seq=27637&ano=3212, 20/01/2011.

[28] Survival Rate of Main Cancer, National Cancer Information Center, 2014. http://www.cancer.go.kr/mbs/cancer/subview.jsp?id=cancer_040302000000, 11/04/2017.

[29] Patient Rate by Area, http://www.rapportian.com, 15/03/2017.

[30] National Trend, http://www.ciokorea.com/slideshow/18932?slide=3#stage_slide, 19/12/2016.

[31] Method of Cure, Korea Breast Cancer Cyber Center, 2014. http://www.kbccc.org/Controls/Main.aspx?d=02&g=HTML&c=0204, 15/03/2017.

[32] Survival Rate of Other Cancer, Korean Statistical Information Service http://kosis.kr/statisticsList/statisticsList_01List.jsp?vwcd=MT_ZTITLE&parentId=D#SubCont, 14/05/2014.

[33] Name of Disease (Cancer), National Cancer Information Center, 2014 http://blog.joins.com/media/index.asp?page=4&uid=sehee94&folder=2&pagesize=5&viewType=, 15/03/2017.

[34] Name of Disease, Samsung Seoul Medical Center, 2014. http://www.samsunghospital.com/home/healthInfo/content/contentList.do?CONT_CLS_CD=001020001013, 12/05/2014.