VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

DEVELOPMENT OF CLINICAL DECISION SUPPORT SYSTEM FOR HUMAN HEALTH MONITORING

Manickavasagam B., B. Amutha and Akanksha Chauhan
School of Computing, SRM University, Tamil Nadu, India
E-Mail: bmanickavasagam90@gmail.com

ABSTRACT

Encompasses the urgent need for health monitoring of patients. Human vital parameters such as Electrocardiography (ECG), Blood Pressure (BP), Oxygen Saturation - Sp02,Pulse Rate (bpm), and Body Temperature have been taken to predict any critical human health from Patient records of hospitals and health organization research documentation explains the parametric variation associated with the symptoms and by considering these variations the diseases can be identified to acknowledge the human body conditions and this has further been characterized into three major categories, Normal human health condition, and Maximum and Minimum level. A stochastic model has been developed to assess these variations using R Programming wherein the linear regression gives the accuracy of the values about the parameters. Using MS-SQL database is created. Further two matrices have been constructed, one having symptoms along with their associated diseases and the other having disease with the approved drugs that are available. These data values have then been fetched into the ASP.NET to create a web page for coming in handy to the patients in emergency situations or in cases where home healthcare is being done for the patient. The users can select the symptom from the webpage and on submission, the disease with which the patient is suffering will be displayed and this further directs the drugs that can be provided to the patient.

Keywords: vital body parameters, linear regression, correlation, webpage, patient monitoring system [PMS].

1. INTRODUCTION

Health monitoring is done in many ways all around the globe. To come up with something which is challenging yet easy to use is the main objective with which the research is done these days. In the modem world, human patients are continuously monitored during the observation time in the hospital [1]. It causes inconvenience to the patients staying in hospitals, treatment cost is huge as well as it leads to waste of time and human energy. To avoid these kinds of issues and to improve the monitoring facilities and further identify the health condition before the emergency arises in remote area or at patient's home can be done by using some web portals and programs. This comes in handy during an emergency or for quick reference by the attendee in remote areas or any place where there is a need.

The body condition of each person differs ranging from a healthy body to a little week and highly diseased body. There are various variations all the time and each disease condition too has a certain set of associated variations which make a disease fatal or easily treatable. In every patient's body the effect of a disease would vary. As a thorough documentation is required for the work, patient records of hospitals and health organization research documentation is procured to acknowledge the body conditions. Once the documentation from trusted websites and hospitals is got it gives a clear picture of the conditions and situations all around us.

2. BACKGROUND WORK

a) Parameter variation analysis

Different data from the hospitals or research document has been analysed. To assess a disease its body parameters are of paramount importance. The body parameters like ECG, Blood Pressure, Heart Rate, Sp02, Temperature are taken because in these only the predominant variations are seen and by looking at them we can decide as to what disease could possibly have occurred. The human body parameters play an important role in the understanding of the effects and causes of the diseases.

b) Categorization of data

In [1], Acquired data of all body parameters is analysed and characterized into three major categories, the categories being Normal human health condition, Maximum and Minimum level of the parametric changes leading to life threatening situations. The levels have been decided to categorize the diseases based on the fatalities that can occur. When the categorization is done, it enables us to assess the ranges that a disease can acquire. The disease goes from a very low parameter value to a very high parameter value. High or low parameter value decides if a person will die or will not. If the extremities have been reached on both the sides then the person will most probably die or there could be the normal body condition where nothing happens. The effects on human body have been assessed using the variations that are observed. Analysis of the body parameters in different stages has been done at micro level taking individual parameters at a time and this information has been normalized in a certain procedure to easily classify the disease further. At the micro level the number of diseases associated with the parameter variations is taken, this sets a clear example of what arc the possible outcomes of a parametric variation and this gives a lot of information regarding the prevention or cure that could be further given.

After the parameters are seen individually a combination of these parameters has been taken. The combination would include taking two parameters then three parameters and so on. The combinations are done in a random fashion to assess how many possible combinations can be made that would specify a particular. To reach to this specification is a crucial matter as it enables us to be sure about what the patient might be suffering from. There are around 18900combinations. This would help in better and accurate disease classification. This classification would help us know the severity of the disease and what stage the person is in at a particular point of time then the specific drugs that could be given to the patient would be recorded and the treatment could further be initiated.

To know what disease has occurred the symptoms associated with it are the indicators of it. Of all the available diseases, many symptoms are there but to clearly understand which disease has occurred some study has to done. We have taken some basic diseases that are not hereditary orgenetic and worked upon them.

c) Matrix construction

Two matrices have been created involving the diseases with their symptoms and diseases with their drugs. These two matrices give the visual understanding of what relations a symptom has with a disease and which all diseases are common. It also shows which all drugs are available for a disease. The created matrix gives a clear picture of how' the diseases affect the human body and how these variations are correlated with each other. The correlation is important in describing the correlation as well as unique diseases and their unique function along with their symptoms.

d) Disease storage and GUI based application

A database is required for storing all the details about the number of diseases that have been recorded and the symptoms associated with each disease. The records have been made in MSSQL. Maximum number of diseases that could be taken is done from village area [2] apart from the hereditary diseases or genetic diseases whose identification cannot be done with simple parameter variation analysis but with blood test, CT/MRI scans. Very complex diseases is not easy to analyse hence leaving them the rest have been taken. Around 5-7 databases have been created, each for a different purpose. The data of each parameter and its variation is recorded taking normal condition, death condition and some variations according to sex. The other database includes diseases with their symptoms, one having details about the parameters that are involved and how a parameter is related to a particular symptom has been mentioned, in other database we have numbered the symptoms and diseases to find the common symptoms in various diseases to finally do the linear regression. Then 2 final databases have been created one containing the symptoms and diseases, the other with diseases and available and approved drugs.

The final part is web page creation. The basic aim was to come in handy for the people in emergency situations or the patients in remote locations. Hence the web page creation is the final work. ASP.net has been used to create the webpage. Just simple selections of the symptoms have to be done by the user and the drugs for the disease that is displayed can be acquired.

3. METHODOLOGY

In the modern world, human patients are continuously monitored during the observation time in the hospital. It leads to the patients inconvenient to staying in: Hospitals, Treatment cost, Waste of time and Human energy. To avoid these kinds of issues we are trying to improve the monitoring facilities and identifying the health condition before the emergency arises in remote areas. So here we are trying to prevent the fatalities by creating a matrix involving the diseases with their symptoms. The created matrix gives a clear picture of how the diseases affect the human body and how these variations are correlated with each other. The human body parameters play an important role in the understanding of the effects and causes of the diseases. These parametric variations are measured. After measuring the variations, MS-SQL has been used to create the database to record the diseases, symptoms and associated drugs.

The basic methodology includes the usage of certain databases to record required data and tools to analyse them further to create a webpage, required things are mentioned. The main objective is to monitor patients' health condition using vital human body parameters: Temperature, Blood Pressure, Sp02 and Heart rate and to identify the fall detection before emergency. These vital human body are taken to predict disease based on variations in their ranges. The normal ranges as well as their abnormal values of these body parameters will be assessed. The data is obtained from patient records of hospitals and health organization research documentation. Acquired data of all body parameters is analysed and categorized into three major categories: Normal human health condition. Maximum and Minimum level of the parametric changes leading to life threatening situations. The effects on human body would be assessed using the variations that are observed.

Analysis of body parameters in different stages would be done at micro level taking individual parameters at a time and this information will be normalized in a certain procedure to easily classify the disease further. After the parameters are seen individually a combination of these parameters would be taken. The combination would include taking two parameters then three parameters and so on. This would help in better and accurate disease classification. This classification would help us know the severity of the disease and what stage the person is in at a particular point of time then the specific drugs that could be given to the patient would be recorded and the treatment could further be initiated. Normalization of these vital body parameter values by taking single parameters using R programming would be done. The

need of normalization arises as we need to ensure data integrity and eliminate data redundancy. Further linear regression analysis of this data was done to model a relationship between a dependent variable and an independent variable. After regression analysis, database was created which was used to develop a web page.

a) Data collection

The vital human body parameters both normal range and abnormal values along with their variation according to age were listed. This data was collected from patient record of hospitals and other health organizations. Some websites were also used to collect this data:

- www.rightdiagnosis.com
- www.healthline.com
- www.healthnet.com
- wwwMncdicinc.net.com
- www.mayoclinic.com
- www.webmed.com

b) Statistical analysis

Statistical analysis involves histogram representation and then linear and multiple regression.

i. Histogram representation:

The data of vital human body parameters along with their variations with age was represented by histograms. This representation shows density estimation and probability estimation of continuous data. For statistical analysis of data and graphical representation R studio is used. After this, combination of two parameters is normalized and then further normalization will bed one by taking three and then four parameters. Normalization is done using Normal parameter as Zero Level. The maximum value of each parameter is considered as +1 and the minimum value of the same is considered as -1. One such example of histogram representation of systolic blood pressure is shown in Figure-1 with BP on x-axis and frequency on y-axis

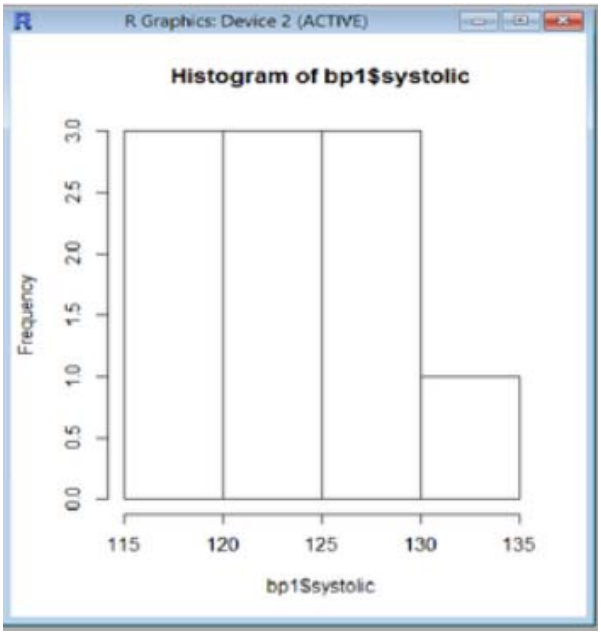

Figure-1. Histogram of BP.

ii. Range evaluation

After normalizing data, range evaluation was done to get threshold values of vital human body parameters which includes from normal range to life threatening ranges of these parameters[1,3], illustrated in Figure-2.

Range	temp(℃)	Effects
-1	24	Death
-0.9	25	death due to irregular heart beat
-0.8	26	respiratory arrest
-0.7	28	disturbed heart reathm
-0.6	31	rarely conscious/ shallow breathing
-0.5	32	hallucinations/sleepiness
-0.4	33	depressed reflexes/confusions
-0.3	34	severe shivering/blueness
-0.2	35	intense shivering / numbness
-0.1	36	mild to moderate shivering
0	36.12	normal body temperature
0.1	37	high normal
0.2	38	sweating / slight hunger
0.3	39	severe sweating / fast heart rate
0.4	40	fainting / dehydration / vomiting
0.5	41	dizziness / delinium / palpitation
0.6	42	become comatose / severe delinium / very fast heart rate
0.7	43	severe brain damage
0.8	44	almost death
0.9	44	Death
1	44	Death

Figure-2. Range wise parameter variation.

iii. Linear regression

Linear regression analysis of blood pressure was done first. In this plot, an independent variablex(age) is plotted against a dependent variable y(bp).A scatterplot was plotted between systolicblood pressure and age by using the following code: Plot (age, systolic. Normal. range, main=”scatterplof”)

Systolic normal range was the name given to column containing normal systolic blood pressure values in database. The scatterplot was plotted between age on x axis and systolic blood pressure on y axis. A straight line is obtained approximately writh initial values showing randomness. The straight line was obtained between 120-130 bp falling within age group 20-60 or above. The correlation wras also calculated for this scatter plot w'hich came positive. This shows that there is a linear association between age and systolic blood pressure. Then summary was obtained using code: summary (mod) which gives us residuals and coefficients. Residuals show random pattern obtained in the scatter plot. The intercept should be zero. It show's randomness present in the data.

In residuals, minimum and maximum values were shown. Median was also calculated along with1q and 3q. Coefficients were also calculated with intercept. Standard error was calculated as1.6174.t-value and p-value also obtained. Residual standard error was obtained as 4.4.84 on 23degrees of freedom. Multiple R-squared and adjusted R- squared values also calculated. MultipleR squared value was 0.8735 and adjusted R-squared was 0.868[4].

F-test was also done. F-Test and p value were found to be 158.8 and 8.268e-12 respectively. A regression line and a straight line were seen because it was assumed that the data is linear and the line shows the predicted y value. And the line shows which the predicted y value is. The y values are dependent and can be expressed as a linear function of the x variable. Mod and coefficient of mod is calculated subsequently.

The confidence value is generated again with the intercept. Abline and mod of it gives a straight line on the plotted points for a clear understanding and gives the line of best fit. After obtaining regression line, par and plot (mod) command was used to get four different graphs. In Figure-3, the first graph is residual v/s fitted curve. This plot showed whether linearity association criteria were met or not. If it is met then no pattern would be obtained and we get a flat line. In age v/s systolic BP plot we obtained a straight line. Next graph is normal Q-Q curve. Third one is scale location and fourth graph is about residual v/s leverage

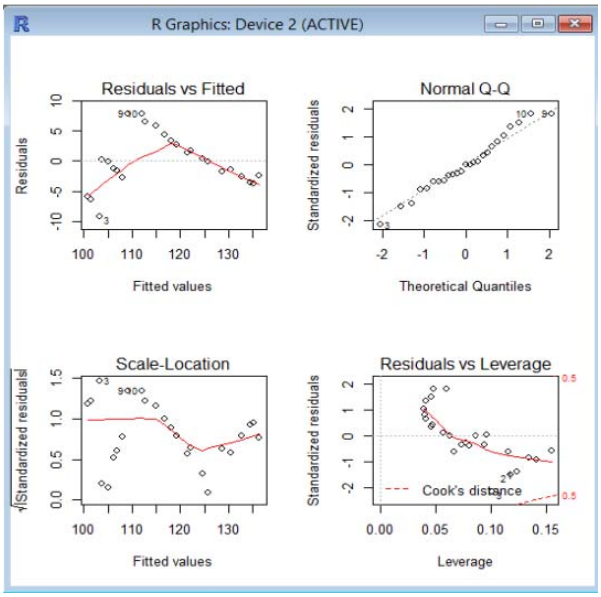

Figure-3. Statically result of scale, and residual and normal linkage values.

c) Disease identification

Around 240 diseases have been selected that are not very complex diseases. Hereditary diseases, cancerous diseases and genetically acquired diseases have been ignored as these diseases cannot be analysed just on the basis of their symptom and parameter variations. The database used to record the diseases is KEGG Human Diseases + Gene.KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances. KEGG is utilized for bioinformatics research and education, including data analysisingenomics, metagenomics, metabolomics and other omics studies, modeling and simulation in systems biology, and translational research in drug development[5].

Different categories of diseases are present in the KEGG database. From each category, we have taken the diseases and included some other diseases from Wikipedia. This collection of diseases is very through and can be used to evaluate all the different types of diseases that can occur.

d) Symptom classification

With each disease, there is some symptom associated. The information about these symptoms has been taken from the trusted websites of hospitals or research organization documents that are available. When a disease occurs, there are varied symptoms ranging from very harmful effects ornormal ones that have been a little heightened. As symptom is a departure from normal function or feeling which is noticed by a patient, reflecting the presence of an unusual state, or of a disease. A symptom is subjective, observed by the patient, and cannot be measured directly, whereas a sign is objectively observable by others. For example, paraesthesia is

asymptom (only the person experiencing it can directly observe their own tingling feeling), whereas erythema is a sign (anyone can confirm that the skin is redder than usual). Symptomsand signs are often nonspecific, but often combinations of them are at least suggestive of certain diagnoses, helping to narrow down what may be wrong. In other cases, they are specific even to the point of being pathognomonic. Here the classification of the symptoms is done and is noted

Finally, the correlation has further been seen between the different diseases symptoms, diseases and parameters. This correlation gives a clearer picture of when can a disease possibly occur (Figure-4).

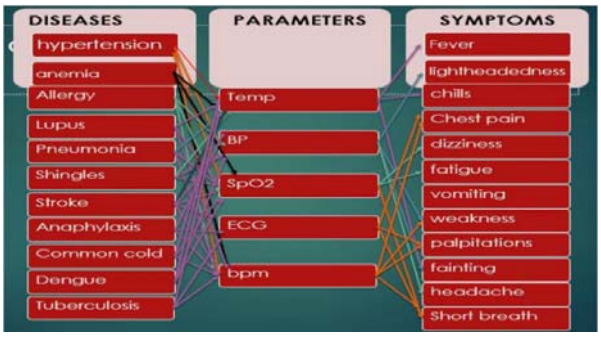

Figure-4. Correlation of the values of disease, symptom and parameters.

After this work, the diseases have been given the variable D and the symptoms have been given the variable S. Ranging from D1…Dn the diseases have been numbered as well as the symptoms have been numbered from S1…Sn, shown in Figure-5.

DISEASE	SYMPTOM
Perry syndrome(D1)	Weight loss(S1)
Appendicitis(D2)	Constipation(S1)
Eosinophilic Esophagitis(D3)	Nausea(S1)
Hypertension(D4)	Vomiting(S1)
Pneumonia(D5)	Headache(S1)
Syphilis(D6)	Fever(S1)
Anxiety Disorder(D7)	Numbness(S1)
Infectious Mononucleosis (Mono)(D8)	Loss of appetite(S1)

Figure-5. Numbering of diseases and symptoms.

Various common symptoms are present in each disease. Those common symptoms have been documented. Fever is a symptom which might be available in many other diseases example pneumonia, common cold, tonsils, etc. A disease is numbered D1…D240 and symptomS1…S2000. Then their combination would look like D1S1, D1S2, D1S3, D2S1, and D2S7 etc. their combination gives a better view which can further be used in the study.

e) Disease and drugs

The analysed diseases have been used here also. Now the addition is the drugs for theparticular diseases. For taking the drugs the database that we used was DrugBank. The Drug Bank database is a comprehensive, freely accessible, online database containing information on drugs and drug targets. The database is very informative and around 3-4 drugs for each disease have been mentioned. Though only a doctor can know which disease should be given to a patient but if some health care monitoring is being done, the user will have some idea and this information can come in handyo the user. Some drugs are also common which can be given in several health conditions but the duration and dose is decided by the practitioner. Figure 6 illustrate the disease and drugs relation.

DISEASE	DRUG
Nephronophthisis_And_Medullary_Cystic_Disease_Complex	Nephronophthisis_And_Medullary_Cystic_Disease_Complex Vandetanib
Polycystic_Kidney_Disease	Paricalcitol
Pearson's_Syndrome	Gabapentin_Enacarbil
Birt-Hogg-Dube_Syndrome	Idursulfase
Cutis_Laxa	Doxorubicin
Dermatitis_Herpetiformis	Prednisone
Natural_Killer_Cell_Deficiency	Aldesleukin

Figure-6. Drug and disease examples.

4. RESULTS

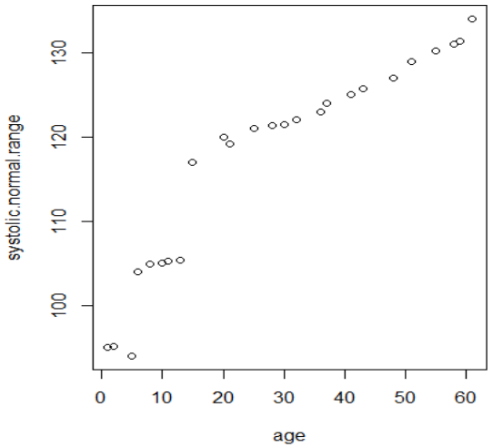

Figure-7. Correlation between age and systolic BP.

This session explains the correlation and scatter plot results. Correlation was calculated between age and systolic BP which was found to be 0.9346, Figure-7.
Residuals and coefficients are calculated, Figure-8:
Standard t-value: 0.0470
P-value: 8.268e-12
Estimated error: 0.5923
F-statistic: 158.8
Multiple R –squared: 0.8735
Adjusted R-squared: 0.868

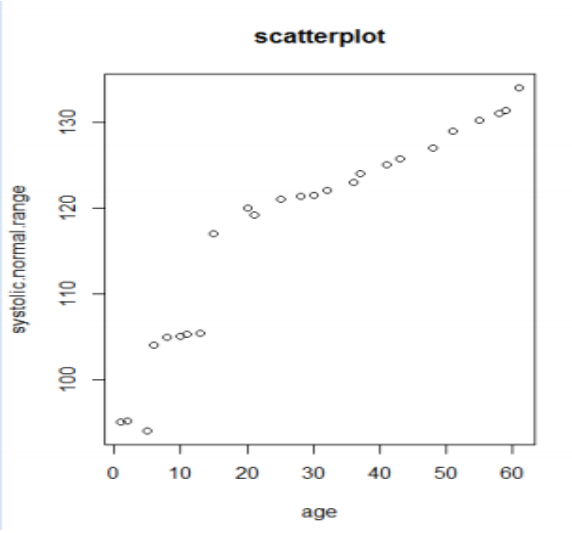

Figure-8. Summary of scatterplot.

A straight regression line is obtained which shows that the data is linear in nature, Figure-9. Multiple R squared: 0.8324, Adjusted R squared: 0.7989
F statistic: 24.83, P-value: 0.0001322

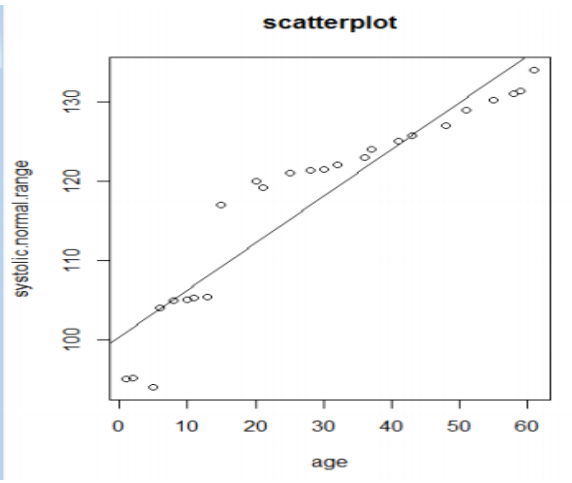

Figure-9. Regression line obtained in scatter plot.

5. CONCLUSIONS

With the aim of monitoring human health conditions, clinical support system has been developed. This support system provides patient an easy and comfortable access to predict their diseased condition with the help of symptoms with which the patient is suffering. Patients get firsthand on monitoring of their health condition. Symptoms when entered into the support system, it gives a list of predicted diseases with some prescribed drugs. This support system is apart of Expert system for clinical support in healthcare.

The first phase involves data collection of patient records of vital human body parameters fromwebsites and hospitals.

Second phase involves statistical analysis of the data with linear regression which is to prove thatour data is linear in nature.

The third phase involves development of webpage which takes symptoms as input and givespredicted diseases along with drug prescription (when doctor is not available at the moment) incase of any emergency situation.

This clinical support system may have the following future enhancements:

- A proper algorithm can be created to detect the disease with exact combination ofsymptoms.
- This program can be implemented in some devices to help health monitoring of thepatients that will be with the patient all the time.

REFERENCES

[1] Amutha, B.manickavasagam. B, *et al*. "Erection of comprehensive wellness programme for global healthcare monitoring system using AODV protocol with data clustering schema." Indian Journal of Science and Technology 8.17 (2015).

[2] St´ephane Cormier, Hac`eneFouchal and ItheriYahiaoui.CReSTIC Lab. Universit´e de Reims Champagne-Ardenne," Disease Tracking Service in Urban Areas." Published in, Wireless Communications and Mobile Computing Conference (IWCMC), 2012 8th International ,2012.

[3] Greg Kelly, "Body Temperature Variability (Part 1):A Review of the History of Body Temperature and its Variability Due to Site Selection, Biological Rhythms, Fitness, and Aging" 2006

[4] http://www.utsc.utoronto.ca/~butler/r/book.pdf

[5] http://dbpedia.org/page/KEGG?cv=1