Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 87 (2016) 227 – 232

Procedia Computer Science

Fourth International Conference on Recent Trends in Computer Science & Engineering.

Chennai, Tamil Nadu, India

Bigdata analytics on Diabetic Retinopathy Study (DRS) on real-time data set identifying survival time and length of stay

Sujatha .V[a]*, Prasanna Devi S[b], Vinu Kiran S[b],ManivannanS[c]

[a]Phd Scholar, MS University, Assistant Professor, Vellore Institute of Technology,Vellore, Tamil Nadu 632014, India
[b]Department of Computer Science & Engineering, Apollo Engineering College, Chennai 602105, India
[c]Deputy Dean, Dr.MGR Educational & Research Institute University, Chennai 600107, India

Abstract

In this paper, we had analyzed a large scale Diabetic data sets for several patients to find the length of time taken for treatment for each class of Diabetes and the risk of re-admission of diabetic patients performing Bigdata analytics, the type of diabetes and its outcome which acted as a high risk sample of patient data sets. We have collected and integrated different sources of diabetic information for several patients, from primary and secondary treatment information to administrative information, to analyze novel view of patient care processes such as type of treatments and every patient behaviors on which results multifaceted nature of chronic care that we take into our account to predict the survival factors and length of stay. Nowadays by using electronic medical equipments with high quality and high degree calibrations, we are able to gather large amounts of real-time diabetic data sets. The requires the usage of distributed platforms for making BigData analysis that results on making decisions based on available data and its trends. This type of Bigdata analysis allows geographical and environmental information of patients' enables the capability of interpreting the ethnicity of data gathered and extract new analysis to identify survival options and treatment timelines (LOS) from them.

1.Introduction

Diabetes is becoming a potential epidemic in India with more than 62 million individuals currently diagnosed with the disease[1]. It is a troubling reality, that this disease will continue to be a prevalent problem in India and throughout the world.[2]

India with 31.7 million diabetic individuals, topped the world with the highest number of people with diabetes mellitus in the year 2000, followed by China with 20.8 million individuals along with the United States having 17.7 million in second and third place respectively. According to Wild et al.[3] there will be an increase in diabetes is predicted to double globally from 171 million in 2000 to 366 million in 2030 with a maximum increase in India. Many external influences affect the development of disease throughout a country, and identification of those factors is necessary to facilitate change when facing the challenges in treating them. For a given time period, the individuals may have high or low symptoms which leads for revisiting for treatment. A class of diabetes for a particular ethnicity is could be cured on a time span.

* Corresponding author. Tel: +0-984-069-5962
*E-mail address:*sujathavishnu03@gmail.com

Peer-review under responsibility of the Organizing Committee of ICRTCSE 2016
doi:10.1016/j.procs.2016.05.153

We had implemented a model to identify high risk of readmission through Bigdata analytics that enables high opportunity with accessibility to healthcare providers based on factual data that facilitates them to develop programs to improve the quality of treatments. The proper implementation of these analytic methods over large-scale data would enable proper utilization of resources in hospitals thus reduces the readmission rate and the cost incurred due to re-hospitalization. The evolution of predictive modelling solutions are challenging for identifying patient's survival and risks of readmission in healthcare informatics and for computing the treatment timelines.

Analysis were performed and created using factual patient data sets from medical equipments and the quantified values on physical examination which acted as structured data. There is a significant increase in the digital calibrations from electronic devices and growing volume of medical and patient data sets, over the last ten years accelerated the need for efficient tools to analyse the large data for computation and analytical processing technologies. Big Data includes new data management systems, improved analytics capabilities, faster hardware which facilitates faster predictability, access through all data sets and efficient result generation.

2.Big data analysis methodology on Diabetic data

2.1.Large scale diabetic health-care data sets

Clinical data sets from hospital databases contains valuable but heterogeneous and difficult data in terms of missing values, incomplete information or inconsistent records for each patients[4]. It's high dimensionality depends upon the number of disease parameters and their number of features but also their complexity. Analyzing external factual data is more challenging than that of analysis of results of a carefully designed experiment or trial. Because one does not know on how and what type of information was collected. So, we carefully performed our analysis by knowing, it is very important to use these huge amounts of data wherever needed, to find new information that could be possibly not available or difficult to predict anywhere.

2.2.Data Assembly

This study used the patient data sets for 195 diabetic individuals as a comprehensive clinical records across various hospitals throughout Chennai and Kanchipuram, India. The Health Facts data we used was an extract of a random samples of patients, representing 5 years (2011–2016) of clinical care at 12 hospitals. This data represents stand-alone hospitals, the data contains both inpatient and outpatient data, including emergency department, for the same and other group of diabetic patients.

2.3. Identification of initial knowledge on existing data set

(1) We identified inpatient encounters where a patient is treated under a hospital admission.
(2) We identified "diabetic" encounter, where we grouped patients where diabetes was entered to the system as adiagnosis.
(3) We grouped the patients with length of stay was at least 1 day and at most 14 days.
(4) We gathered the results of laboratory tests were performed during the encounter for few parameters.
(5) We considered the type of medications were administered during the encounter.

We then identified the random samples which were of less than 23 hrs of duration. In which we observed, changes in diabetes management were less likely to have occurred. We noted that the diabetic encounters are not all encounters of diabetic patients but rather only these encounters where diabetes was coded as an existing health condition. Based on age and gender factors we are predicting the survival of patients and their length of stay for treatment.

3.Types of Diabetes

Type 1 – The pancreatic cells that generates insulin have been destroyed by the immune system of the body. There is a need of injections of insulin along with frequent blood tests and dietary restrictions for patients having Type 1 diabetes.

Type 2 – Adult-onset diabetes or the non-insulin dependent diabetes. Various organs of the body become resistant to insulin which increases the demand for insulin. Pancreas doesn't make the required amount of insulin. Most of Type 2 diabetes patients have border line diabetes where the blood glucose levels are higher than normal but not as high as a
diabetic patient.

Gestational diabetes – Occur in pregnant women due to the high sugar levels as the pancreas don't produce sufficient amount of insulin. Taking no treatment could create complications during childbirth. Diet control and taking insulin can control this form of diabetes

3.1 Symptoms and Treatment:

The common symptoms of a person suffering from diabetes are:

- Polyuria (frequent urination)
- Polyphagia (excessive hunger)
- Polydipsia (excessive thirst)
- Weight gain or strange weight loss
- Healing of wounds is not quick, blurred vision, fatigue, itchy skin, etc.

Detection of diabetes is done based on results of urine test and blood tests by checking for excess body glucose. Commonly conducted tests:

- A1C Test
- Fasting Plasma Glucose (FPG) Test
- Oral Glucose Tolerance Test (OGTT).

4.Analysis Methodology on diabetic data

We are considering the parameters mentioned below;

1. Number of times pregnant
2. Plasma glucose concentration a hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. Hour Serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height inm))
7. Diabetes pedigree function
8. Age (years)

4.1 Prediction models:

We are using multivariate quantitative statistical method which uses dependent variable along with multiple independent variables. We have made observations on the effects of patients and hospital characteristics on the diabetic in patients.

4.1.1 Power analysis:

For the given set of patient data, we are observing the study has sufficient power >=0.795. We are further querying to identify the co-morbidities of diabetes and the behavior of patient among the types of diabetes.

4.1.2 Statistical Analysis:

Patient's basic characteristics such as Age, Gender, type of diabetes with all other available characteristics were compared. Chi-square, independent t-tests and ANOVA were used to detect the actual differences between the actual outcomes and patients with the data analysis results that include, gender, length of stay, transfer of patients and death. Multivariate analysis for these predictors were used to determine the effect of our result of our prediction and its accuracy.

The frequency analysis resulted that more males were admitted in hospital than that of females. The majority of them have poor food habits and have hypertension. This can be corrected to have more survival time of these types of patients. The mean age of all patients was 63.72 (SD+- 13.33). Many of these patients were admitted through an emergency. Most of the secondary diagnosis were ranged from coronary atherosclerosis (,20%) to paroxysmal ventricular tachycardia(3.4%) which includes cardiogenic shock (1.4%) to hypotension (0.3%). We have presented the patient characteristics in Table 1 given below.

Admission Type	Emergency	66%
	Urgent	25%
	Elective	8%
	Other	0%
Length of Stay (Days)	0-12	7.7%
	13-25	13.1%
	26-38	16.1%
	39-51	11.7%
	52-64	7.9%
	65-77	5.9%
	Other	38.6%
Hospital transfer	Routine	62.8
	Transfer to short term hospital	9.8%
	Transfer to home care (Discharge)	10.2%
	Against Medical advice	0.7%
	Died in hospital	7.1%
	Died during hospitalization	7.2%

Table 1: Characteristic parameters of diabetic patients

4.1.3 Bivariate analysis on Length of stay and Patient's age

This shows that the patient's age is having impact on length of stay in hospital. Bonferroni post hoc tests were significant for patients for which we have presented the Mean and SD in the table 2 and table 3 below. The results shows that the older patients' LOS is higher than that of younger diabetic patients.

Age Group	Mean	SD
33-45 (p=0.004)	3.44	3.99
46-58	3.89	4.12
72-84 (p=0.002)	4.80	5.04

Table 2: Mean and SD for highly significant age groups

Age	Group	Mean Difference	P
20-32	33-45	-1.938	1.000
	46-58	-2.388	0.813
	59-71	-2.811	0.401
	72-84	-3.296	0.163
33.45	20-32	1.938	1.000
	46-58	-0.451	1.000
	59-71	-0.874	0.199
	72-84	-1.358*	0.004
46-58	20-32	2.388	0.813
	33-45	0.451	1.000
	46-58	0.423	0.677
	72-84	-0.907*	0.002
59-71	20-32	2.811	0.401
	33.45	0.874	0.199
	46-58	0.423	0.677

	72-84	-0.484	0.433
>=72	20-32	3.296	0.163
	33-45	1.358*	0.004
	46-58	0.907*	0.002
	59-71	0.484	0.433

Table 3: Mean and difference for LOS for all age group sample set

4.1.4 Multivariate analysis on impacts of dependent variables on length of stay (LOS):

We are observing, the patients having other diseases (such as cardiac diseases) in addition to Type 2 diabetes stayed longer in hospital or have significantly higher treatment time range than that of other patients having only diabetes variations as a result of their pre-existing disease condition as mentioned in Table 4. These patients with complications on diabetes have lesser survival time.

Variable	B	Std.Error	Beta	t	p
Gender	0.931	0.193	0.092	4.827	0.000
Age	0.039	0.008	0.113	4.794	0.000
Coronary Atherosclerosis	-1.312	0.226	-1.110	-5.797	0.000
Hypertension	-2.080	0.483	-0.080	-4.309	0.000
Adjustment. R^2=-.092, p<0.001					

Table 4: Multivariate analysis on primary factors of diabetic patients

5. Discussion

Age and Gender have significant dependence of survival time. These factor determines the difference in survival on any subset of the available covariates.Age, gender and the severity of diabetes with the type of treatment are the subsets of the covariates help to explain survival time. For example, age and gender, at time of first treatment acts as primary dependant variable to determine the relative risk of survival. Data for each patient consists of age at first admission to the hospital, sex, number of years of follow-up (years from admission to death or censoring), and patient status at the follow up time. The main goal is to compare the survival experience of these patients to the Race or No. of Treatments to determine if patients tend to have shorter lifetimes. We are seeing the results few of the patients are not exhibiting the classic symptoms of MI. They are not treated agressively and their conditions are usually complicated with they are diagnosed to MI. Since, the diagnosis happens once the disease becomes severe, the survival of is becoming lesser than men with higher amount of treatment and medications.

Based on analysis, we are observing Men have a more favorable survival experience over time than women. compared to men, we are seeing lesser number of women who have admitted. But we are observing longer length of stay. However, more number of men are admitted / diagnosed for diabetes, treated with lesser length of stay. Although MI and Type 2 Diabetes affect all ages, the disease is becoming increasingly prevalent with gender and age that shows greatest effect on elderly people. More than 18% of adults older than 65 have diabetes and the majority with Type 2 Diabetes. So, the change of medication on admission or readmission for the patients with age <30 will improve the diabetic condition and leads to more survival time than the elderly ones. We are observing there were still unexplained variances which we are seeing in this study, that could be the results of some independent variables that were not used because of the nature of database used. So, we are still analyzing this factor which we will use in our future work.

6. Conclusions

Since diabetes is being considered as one of the top priorities in medical science and health care management. The data sets from statistical models or complex pattern recognition models may be fused into a predictive models that combines data set of patients' treatment information and prognostic outcome results. This could be used in clinical decision support and diabetes surveillance to improve patient care. We are concentrating on outcomes over predictive models in screening for and the management of prevalent short term and long term complications on this disease. Several Predictive models already have been developed for management of diabetes and its complications along with the number of publications on such models has been growing over the past decade. We are observing, multiple logistic or a similar linear regression is used for prediction model development, possibly owing because of its transparent functionality. Although extensive effort has been put in to building these predictive models, still there is a remarkable scarcity of impact studies. Our approach analyzes and predicts the outcomes by considering all parameters, since each of them would have significant impact on its own.

References

[1]Seema AbhijeetKaveeshwar, Jon Cornwall, The current state of diabetes mellitus in India, Australas Med J. 2014; 7(1): 45–48. Published online 2014 January 31. doi: 10.4066/AMJ.2013.1979

[2] Joshi SR, Parikh RM ,India--diabetes capital of the world: now heading towards hypertension. J Assoc Physicians India. 2007 May; 55():323-4

[3] Wild S, Roglic G, Green A, Sicree R, King H, Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care. 2004 May; 27(5):1047-53.

[4] K. J. Cios and G. W. Moore, "Uniqueness of medical data mining," Artificial Intelligence in Medicine, vol. 26, no. 1-2, pp. 1–24, 2002.

[5] Cook CB, Naylor DB, Hentz JG, et al. Disparities in diabetes-related hospitalizations : relationship of age, sex and race/ethnicity with hospital discharges, lengths of stay and direct inpatient charges. Ethn Dis. Winter. 2006;16:126-31.