DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.5204

ISSN No. 0976-5697

Volume 8, No. 9, November-December 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

DESIGNING A CLOUD BASED FRAMEWORK FOR ENHANCING THE PERFORMANCE OF DIABETIC CLASSIFICATION USING NAÏVE BAYES CLASSIFIER

Seetha.J
Research Scholar, Department of Computer Science,
A.V.V.M. Sri Pushpam College, Poondi,
Thanjavur, India

T.Chakravarthy
Associate Professor, Department of Computer Science,
A.V.V.M. Sri Pushpam College, Poondi,
Thanjavur, India

Abstract: Diabetes Mellitus is one of the most important serious challenges in both developed and developing countries. It leads to significant medical complications including ischemic heart disease, stroke, nephropathy, retinopathy, neuropathy and peripheral vascular disease. Early identification of patients at risk of developing diabetes is a major healthcare need. Thus cloud based healthcare is the one and only solution. In this paper, we use cloud computing for the creation and organization of cloud based health care services and also use classification technique. The diagnosis of diabetes is used to classifies the individual users as "Diabetic" or "Non-Diabetic" through Naïve Bayes classifier.

Keyword: diabetes, cloud computing, Naïve Bayes classifier.

I. INTRODUCTION

Diabetes is a disease that have an impact on the way the body circulate glucose a form of sugar into energy, our bodies use glucose for nourishment. Normally, the body deliver a hormone called insulin when glucose is in the blood stream. Insulin takes the glucose (sugar) into the cells where it is either used as energy or stored. In someone with diabetes, the body either **doesn't make enough insulin** or **doesn't use the insulin properly** and too much glucose remains in the blood. Over time, high blood glucose can incite more serious problems. Diabetes has been classified into two major categories namely, type 1 and type 2 diabetes.

A. Overview of Diabetes Mellitus

In type 1 diabetes, the pancreas does not produce insulin. Glucose is unable to get into the cells, so the glucose level in the blood goes up. When the glucose level increase above normal, a person has high blood glucose, or hyperglycemia. The cause of type 1 diabetes is not known and it is not preventable with current knowledge [1].

In type 2 diabetes, the pancreas still produce insulin, but the insulin doesn't work right, or the cells can't take in the glucose. The glucose level in the blood goes up is shown in fig 1.1. A person has high blood glucose, or hyperglycemia. Type 2 diabetes is frequently associated with obesity and tends to be diagnosed in older people. It's far more common than type 1 diabetes [2].

The Major Symptoms of Diabetes

- feeling very thirsty
- urinating more frequently than usual, particularly at night
- feeling very tired
- weight loss and loss of muscle bulk
- itchiness around the genital area, or regular bouts of thrush (a yeast infection)
- blurred vision caused by the lens of your eye changing shape
- slow healing of cuts and grazes
- Fatigue (weak, tired feeling)
- Blurred vision
- Loss of consciousness (rare)
- Recurrent infections, including thrush infections

Prevention of Diabetes

- eating a healthy, balanced diet
- losing weight if you're overweight, and maintaining a healthy weight
- stopping smoking if you smoke
- drinking alcohol in moderation
- taking plenty of regular exercise

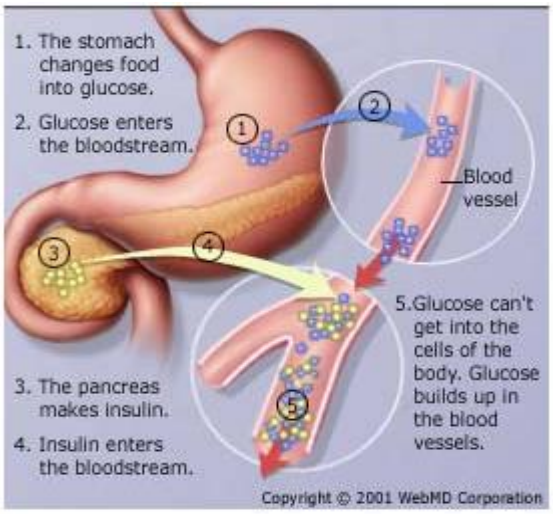

Figure 1.1 Type 2 Diabetes Affects the Body

B. Role of Cloud in Diabetes

Cloud computing is a general term for anything that involves delivering hosted services over the Internet. It can also help people stay more connected to their self-care. It is a new technology and have good performance in storing, managing, sharing and accessing information. The cloud computing based solutions in healthcare can help the physicians to stay in touch with their patients and examine their health condition effectively at a low cost. Cloud services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) [3].

Cloud Features

The additional resources as needed from the consumer requests, and similarly releases these resources when they are not needed. Different clouds offer distinct sorts of resources, e.g., processing, storage, management software, or application services [4]. Clouds are typically erected using large numbers of inexpensive machines. As a result, the cloud vendor can more easily add capacity and can more rapidly replace machines that fail, compared with having machines in multiple laboratories. Generally speaking these machines are as consistent as possible both in terms of configuration and location.

- ***Automated Backup*:** The automated backup and archival options are offer from different clouds. The cloud may move data or computation to improve responsiveness. Some clouds monitor their offerings for spiteful activity.

- ***Virtualization*:** In clouds, the Hardware resources are usually virtual; these are shared by multiple users to improve efficiency. That is, the same physical resources are supported to several lightly utilized logical resources.

- ***Parallel Computing*:** Expressing and executing easily parallelizable computations are using Map/Reduce and Hadoop frameworks, which may use hundreds or thousands of processors in a cloud.

II. RELATED WORK

The k- means algorithm approach is uses 30 diabetic data set. These data set have 10 field namely name, pregnant, plasma, skin fold, body mass index, age, serum- insulin, pres and class. The data set is classified using the k- means algorithm and attain the result may be positive or negative. This method gives better performance [5].

The innovative of hybrid model classification comprised of Bayesian classification and multilayer perceptron and classify the data as diabetic and non diabetic. The main objective of this model is to achieve high accuracy. In this method achieved 81.89% accuracy with 6 features namely pregnant, plasma glucose, triceps skin fold thickness, serum-insulin, body mass index and age. The cost and time minimized because using only six features [6].

Linear Discriminant Analysis (LDA) with Support Vector Machine and Feed Forward Neural Network to used to find data as diabetic and non diabetic. Where LDA reduces feature subset and SVM is responsible to classify the data. They have also compared SVM with feed forward Neural Network (FFNN). The combination of SVM and LDA gives better classification accuracy as 77.60% with 2 features only, these features are plasma glucose concentration and Body Mass Index [7].

III. PROPOSED ALGORITHM

In our research work, The Naive Bayes classifier is used for classification of diabetes data. This classifier is a straight forward and potent algorithm for classification task. Even if we are working on a data set with millions of records with some attributes, it is suggested to try Naive Bayes approach. This classifier gives special result when we use it for textual analysis shown in fig 3.1. Such as natural language processing. The formula for calculating the conditional probability in eqn (1).

$$\mathbf{P(H/E)} = \frac{\boldsymbol{P(E/H).\ P(H)}}{\boldsymbol{P(E)}} \text{ ---------- (1)}$$

Where,

- P(H) is the probability of hypothesis H being true. This is known as the prior probability.
- P(E) is the probability of the evidence(regardless of the hypothesis).
- P(E/H) is the probability of the evidence given that hypothesis is true.
- P(H/E) is the probability of the hypothesis given that the evidence is there.

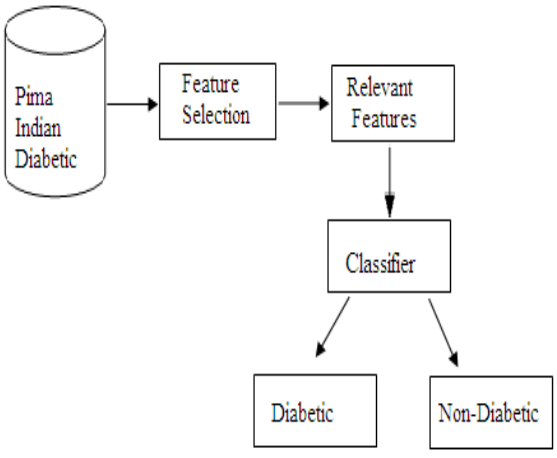

Figure 3.1 Feature Selection Model

A. Diabetes Dataset

The dataset is collected from Pima Indian diabetes data set in UCI repository which is classified under two method diabetic and non diabetic [8]. This data set consists of 8 attributes and 1 class, is shown in table 3.1.

Table 3.1 Pima Indian Diabetes Dataset

Attribute Id	Attribute Name
F1	Pregnant
F2	Plasma Glucose
F3	Diastolic Blood Pressure
F4	Triceps Skin Fold Thickness
F5	Serum-Insulin
F6	Body Mass Index
F7	Diabetes Pedigree Function
F8	Age
Class	Diabetic or Non- Diabetic

B. Performance Measures

Performance can be evaluated various measures such as classification accuracy, sensitivity and specificity[9]. These measures are evaluated using true positive (TP), true negative (TN), false positive (FP) and false negative (FN).

- ***Accuracy:***

  It refers to the closeness of a measured value to a standard or known value.

  $$\text{Accuracy} = (TP+TN) / (TP+FP+TN+FN)$$

- ***Sensitivity:***

  It refers to the ability of a test to correctly identify those with the disease (true positive rate).

  $$\text{Sensitivity} = TP/ (TP+FN)$$

- ***Specificity:***

  It is the ability of the test to correctly identify those without the disease (true negative rate).

  $$\text{Specificity} = TN/ (TN +FP)$$

Where,

False Positive (FP) Users labeled as diabetic but diagnosed as non-diabetic by the expert.

False Negative (FN) Users labeled as non-diabetic but diagnosed as diabetic by the expert.

True Positive (TP) Users labeled as diabetic and also diagnosed as diabetic by the expert.

True Negative (TN) Users labeled as non-diabetic and diagnosed as non-diabetic by the expert.

In our work, we use Naive Bayes classifier for classification of diabetes data. Totally 8 features we choose the high important features only like F8,F1,F2,F6,F5 and to eliminate the less important features like F3,F4,F7. We get better performance and the accuracy is 81.80%, the sensitivity is 71.86% and the specificity is 89.75%.

IV. EXPERIMENTAL WORK

In our implementation, we recognize the patients as diabetic or non diabetic and training data collected from pima Indian diabetic dataset. This dataset have 768 instance, our proposed work take 100 instance or subject. Out of total 100 subjects, 72 subjects are "Diabetic" and 28 subjects are "Non- Diabetic". The statistical analysis of Pima Indian diabetic dataset is shown in table 4.1.

Table 4.1 Statistical Analysis of Dataset

S. No	Attribute	Mean	Standard Deviation
1	Number of times Pregnant	3.8	3.4
2	Plasma glucose concentration	120.9	32.0
3	Diastolic Blood Pressure (mm Hg)	69.1	19.4
4	Triceps Skin fold thickness (mm)	20.5	16.0
5	Serum insulin (mu U/ml)	79.8	115.2
6	Body Mass Index (weight in kg)/(height in m)^2	32.0	7.9
7	Diabetes pedigree function	0.5	0.3
8	Age (years)	33.2	11.8

The objectives of our work is two field. First one is, it needs to correctly identify the patients as diabetic or non diabetic and finally, our work is deployed on a third party cloud computing environment such as Amazon EC2. Amazon EC2 is a IaaS cloud provider that offers lot of machine instance. So we reduce both cost and time.

A. Performance of Naïve Bayes Classifier

We utilize classification accuracy along with sensitivity and specificity measures as metrics for evaluating the performance of our classifier.

- ***Accuracy of Classification***

  The classification accuracy of a classifier is measured as the ratio of the number of subjects. It is evaluated by the formula in eqn (2),

  $$\mathbf{CA = t_c / n * 100} \text{ ---------- (2)}$$

  Where,

  t_c – represents the number of correctly classified subjects and

  n – represents the total number of subjects.

- ***Sensitivity and Specificity:***

  Sensitivity specifies the proportion of actual diabetic users, which are correctly classified.

  $$\textbf{Sensitivity} = \frac{TP}{TP+FN}$$

  Specificity is the proportion of non- diabetic users, which are correctly identified.

  $$\textbf{Specificity} = \frac{TN}{FP+TN}$$

V. RESULTS AND CONCLUSION

In this approach, we have taken Naïve Bayes classifier to find the better results in terms of accuracy, sensitivity, and specificity. We come to the conclusion, that our work has achieved the highest accuracy is 81.80% with 5 features only and achieve highest sensitivity is 71.86% and get highest specificity is 89.75%. Even this work is cost efficiency because we deploy our work in cloud. Cloud is a "pay as you scale", so reduce cost and easy to access.

REFERENCES:

[1] https :// www.nhs.uk / Conditions/ Diabetes-Type1/Pages/Introduction.aspx.

[2] http://www.who.int/mediacentre/factsheets/fs312/en/

[3] Ch Chakradhara Rao, Mogasala Leelarani and Y Ramesh Kumar, "Cloud:Computing Services And Deployment Models", International Journal of Engineering and Computer Science, Vol. 2, Issue 12, pp.3389 – 3392, Dec 2013. ISSN:2319 – 7242

[4] Sean Marston, Zhi Li , Subhajyoti Bandyopadhyay, Juheng Zhang , Anand Ghalsasi, "Cloud computing - The business perspective", Elsevier, pp. 176–189, 2010 .

[5] M.Kothainayaki and P.Thangaraj, "Clustering and classifying diabetic data sets using k- means algorithm", Journal of applied Information science, Vol. 1, Issue 1, June 2013.

[6] Amit kumar Dewangan and Pragati Agrawal "Classification of diabetes Mellitus using Machine Learning Techniques", International journal of engineering and applied science, Vol.2, Issue 5, May 2015.

[7] Parashar A, Burse K and Rawat K, "A Comparative Approach for Pima Indians Diabetes Diagnosis using LDA - Support Vector Machine and Feed Forward Neural Network", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, pp. 378-383, 2014. ISSN: 2277 128X.

[8] https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes

[9] Pankaj Deep kaur and Inderveer Chana " Cloud based intelligent system for delivering health care as a service", 2013 Elsevier, Volume 113, Issue 1, pp. 346-359, January 2014.