IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 36-42
www.iosrjournals.org

Classification of Insulin Dependent Diabetes Mellitus Blood Glucose Level Using Support Vector Machine

Anupriya K. Kamble[1], Ramesh R. Manza[2], Yogesh M. Rajput[3], Kavita A. Khobragade[4]

[1](Dept of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India)
[2](Dept of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India)
[3](Dept of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India)
[4](Department of Computer Science, Fergusson College, Pune (M.S.), India)

Abstract: *Diabetes Mellitus (DM), commonly referred as diabetes is a disorder that most of the people suffer from and which also leads to death many of the times. Type 1, Type 2 and Gestational diabetes are the types of diabetes. The database taken for present study is of patients suffering from Type 1 DM. In Type 1 diabetes a patient has to take external insulin to maintain the blood glucose level (BGL). Insulin is a hormone produced in the pancreas which maintains the BGL of a person. BGL is responsible for a person's daily activity. Within a specified target range the blood glucose level has to be maintained. Thus the blood glucose monitoring helps one to link between blood glucose, food, exercise and insulin. The readings of the blood glucose level determine the best management strategy for diabetes. The complications occurring in diabetes reduces by maintaining good blood glucose control. Four to five types of insulin are available in which our focus is on Regular insulin and NPH insulin. Regular insulin is a short acting insulin and NPH insulin is an intermediate acting insulin. The present study classifies the dataset by using the SVM classification.*

Keywords: *BGL, Diabetes, Regular insulin dose, NPH insulin dose , SVM*

I. Introduction

Diabetes Mellitus is a chronic and potentially life-threatening condition where the body loses its ability to produce insulin, or begins to produce or use insulin less efficiently, that results in blood glucose levels that are too high i.e. hyperglycaemia, commonly known as Diabetes. The types of diabetes are Type 1, Type 2 and Gestational diabetes. The beta cells in the pancreas produces a hormone namely insulin to maintain the BGL normal. Because of the deficiency of insulin the blood glucose level (BGL) of a person gets affected [1]. All the people suffering from Type 1 DM need to inject insulin from the time they are diagnosed as in Type 1 DM the body fails to produce enough of insulin and has to rely on external insulin throughout the day [2]. Increased stress and unhealthy lifestyle occurs because of the imbalance of the BGL. By which the daily life activity gets disturbed. Normal range of BGL are maintained by the insulin injected externally. It is not maintained till normal but is supposed to be at normal range.

Table I. Types Of Diabetes Mellitus and Its Causes and Effects [3] [4] [5]

Sr. No.	Type of Diabetes	Begins from	Cause	% of diab-etes mellitus cases	Mostly affected areas
1.	Type 1 DM	Pancreas failed to produce enough insulin	Unknown	10 %	North America and Europe
2.	Type 2 DM	Cells fails to respond to insulin properly (Insulin resistance)	Excessive body weight and not enough exercise	30 % 60 – 80 % 100 %	China and Japan Europe and Africa Pima Indians and Pacific Islanders
3.	Gestational diabetes	High blood sugar level of pregnant	High blood pressure without previous history of diabetes	2 – 10 %	Common

TheBGLare measured in millimoles per litre of blood (mmol/L) or milligram per deciliter of blood(mg/dL). Where, 1 mg/dL is equal to 18.018018 mmol/L. A normal pre-meal BG ranges approximately 80-120 mg/dl means 4.44-6.66 mmol/L and a normal post-meal BG ranges 80-140 mg/dl means 4.44-7.7700000000000005 mmol/L.[6]As given by 'Better Health Channel', in Type 1 DM patient fully depends upon external insulin there are five types of insulin i.e. insulin are fast or rapid-acting insulin, short-acting insulin, intermediate-acting insulin, mixed insulin and long-acting insulin which depends upon doctors prescription which and when to take for maintaining the BGL.

Present database consist of Regular insulin dose commonly known as Humulin R, NPH insulin dose and UltraLente insulin dose which is short-acitng, Intermediate acting insulin and long acting insulin respectively. The BGL readings are taken after exercise, before and after snacks, meal and dinner, also after giving insulin. From these for the current study only BGL obtained after giving Regular insulin dose and NPH insulin dose was taken into consideration and was classified by using SVM.

II. Data base

The present dataset is related to the diabetes patients who are taking their treatment on regular basis and what physiological and pathophysiological changes occurs during the treatment has been recorded. The dataset consist of 70 patients. In which the total number of treatment days differ. Diabetes patient records were obtained from two sources: an automatic electronic recording device and paper records. The automatic device had an internal clock to timestamp events, whereas the paper records only provided "logical time" slots (breakfast, lunch, dinner, bedtime). For paper records, fixed times were assigned to breakfast (08:00), lunch (12:00), dinner (18:00), and bedtime (22:00). Thus paper records have fictitious uniform recording times whereas electronic records have more realistic time stamps [6]. Attributes of Diabetes files consist of four fields per record. Each field is separated by a tab and each record is separated by a newline.

File Names and format:
1. Date in MM-DD-YYYY format
2. Time in XX:YY format
3. Code
4. Value

The Code field is deciphered as follows:
33 = Regular insulin dose
34 = NPH insulin dose
35 = UltraLente insulin dose
48 = Unspecified blood glucose measurement
57 = Unspecified blood glucose measurement
58 = Pre-breakfast blood glucose measurement
59 = Post-breakfast blood glucose measurement
60 = Pre-lunch blood glucose measurement
61 = Post-lunch blood glucose measurement
62 = Pre-supper blood glucose measurement
63 = Post-supper blood glucose measurement
64 = Pre-snack blood glucose measurement
65 = Hypoglycemic symptoms
66 = Typical meal ingestion
67 = More-than-usual meal ingestion
68 = Less-than-usual meal ingestion
69 = Typical exercise activity
70 = More-than-usual exercise activity
71 = Less-than-usual exercise activity
72 = Unspecified special event

Two types of data is present i.e. a diabetes data set and the another one is the ICU data set. For present study the Diabetes data set was taken into consideration. And only the value i.e. the blood glucose level values were useful of the attributes mentioned above. The BGL obtained after giving insulin was of the insulin formulation which had its own characteristic time of onset of effect (O), time of peak action (P) and effective duration (D). Significantly these time can be affected by many factors such as the site of injection (e.g. much more rapid absorption in the abdomen than in the thigh) or whether the insulin is a human insulin or an animal

extract. The times researcher have listed below are rough approximations and researcher is sure that an endocrinologist with different estimates will be found.
Regular Insulin: O 15-45 minutes P 1-3 hours D 4-6 hours
NPH Insulin: O 1-3 hours P 4-6 hours D: 10-14 hours
Ultralente: O: 2-5 hours. P (not much of a peak) D 24-30 hours [6].

Table II. Sample Dataset Of First 10 Values Of Patient 1

Date	Time	Code	BGL (Value)
4/21/1991	9:09	100	58
4/21/1991	9:09	9	33
4/21/1991	9:09	13	34
4/21/1991	17:08	119	62
4/21/1991	17:08	7	33
4/21/1991	22:51	123	48
4/22/1991	7:35	216	58
4/22/1991	7:35	10	33
4/22/1991	7:35	13	34
4/22/1991	13:40	2	33

From the above dataset the BGL value after giving the Regular insulin and NPH insulin was taken out separately so as to classify using the SVM algorithm.

Table III. Sample Dataset of First 5 Values of Patient 1 after Giving Regular Insulin Dose

Regular Insulin Dose		
Date	**Time**	**BGL**
4/21/1991	9:09:00 AM	9
4/21/1991	5:08:00 PM	7
4/22/1991	7:35:00 AM	10
4/22/1991	1:40:00 PM	2
4/22/1991	4:56:00 PM	7

The dataset contains 21 parameters, but these two parameters were taken as BGL reading of most of the patients these two parameters were available whereas of other parameters all patients all readings were not available.

Table IV. Sample Dataset of First 5 Values of Patient 1 after Giving Regular Insulin Dose

NPH insulin dose		
Date	**Time**	**BGL**
4/21/1991	9:09:00 AM	**13**
4/22/1991	7:35:00 AM	**13**
4/23/1991	7:25:00 AM	**13**
4/24/1991	7:52:00 AM	**14**
4/25/1991	7:29:00 AM	**14**

III. Support Vector Machine

For the statistical study, as explained above total 7 patients count of BGL reading were taken. The number of count taken per patient is given as follows:-

Table V. BGL Readings Count Per Patient Taken For Statistical Study

Sr. No.	**Patient No.**	**Count of BGL readings (X or Y)**
1.	Patient 1	139
2.	Patient 2	3
3.	Patient 3	1
4.	Patient 4	58
5.	Patient 5	72
6.	Patient 6	25
7.	Patient 7	2
Total		300

Support Vector Machine (SVM) is used when the data has exactly two classes. The data is classified in an SVM by finding the best hyperplane that separates all data points of one class from those of the other class. The hyperplane having the largest margin is the best hyperplane for SVM. The maximum width of the slab which is parallel to the hyperplane and that has no interior data points is known as the margin. In the support vector machine, the *support vectors* are the data points which are closest to the separating hyperplane. These data points are found on the boundary of the slab. The figure 3.2 explains the support vector machine in which '+' indicates data points of type 1, and '–' indicates data points of type –1. [7]

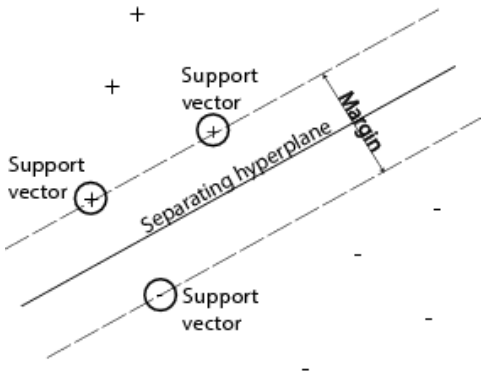

Figure: SVM Hyperplane[7]

On below sample dataset SVM classification was applied.

Table VI. Sample Dataset of BGL Readings on Which SVM Was Applied

Regular Insulin BGL Readings (X)	**NPH Insulin BGL Readings (Y)**
9	13
7	13
10	13
2	14
7	14
11	14
7	14
10	14
4	14
5	14

The data was taken as 150 each of BGL obtained after giving Regular Insulin and NPH Insulin.

IV. Nonlinear Classifier With Gaussian Kernel

Nonlinear classifier with Gaussian Kernel generates one class of points inside the unit disk in two dimensions, and another class of points in the annulus from radius 1 to radius 2. A classifier based on the data with the Gaussian radial basis function kernel is then generated. To make a strict classification the box constraint parameter is set to Inf so that no training points will be misclassified.

V. Result

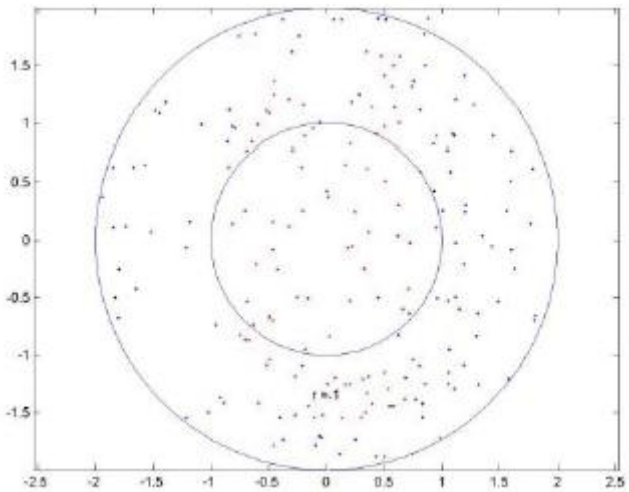

Figure II: Result of classification with SVM

As shown in Figure II the data was after applying the SVM the values of two dataset got scattered along the unit disk. The data didn't get classified properly in the circle of radius 1. Hence classification was not obtained and then the data was classified using the Kernel function where it was set to rbf and boxconstraint was set to inf. And the result obtained was as follows:-

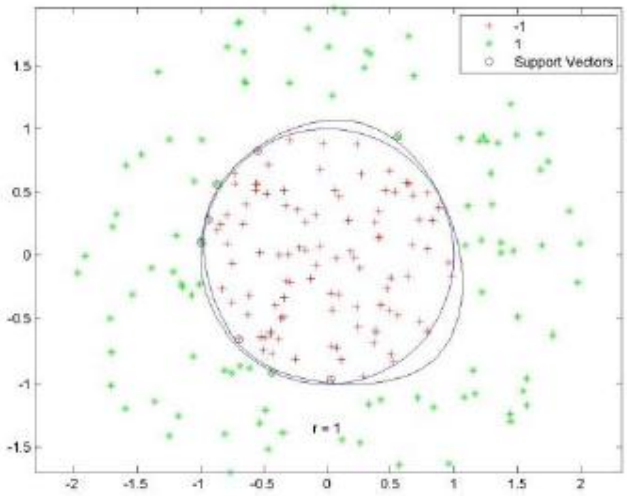

Figure III: Training with SVM using Kernel function set to rbf and boxconstraint set to inf.

When the Kernel function was set to rbf and boxconstraint to inf the data was classified using 8 support vectors. Classification was perfectly done. Then also again to verify the classification the data was further classified using SVM Kernel function set to rbf as it is only the boxconstraint here was set to default parameters. And its result obtained was as follows :-

As seen in figure 4.5 the data was well classified with SVM using Kernel function set to rbf with default parameters. The data was found to be best classified using SVM with Kernel function set to rbf and boxconstraint set to inf. But some of the data was misclassified while rounding off in the circle of radius r=1.

The line separating the two blood glucose level was found to be as given in Figure IV.

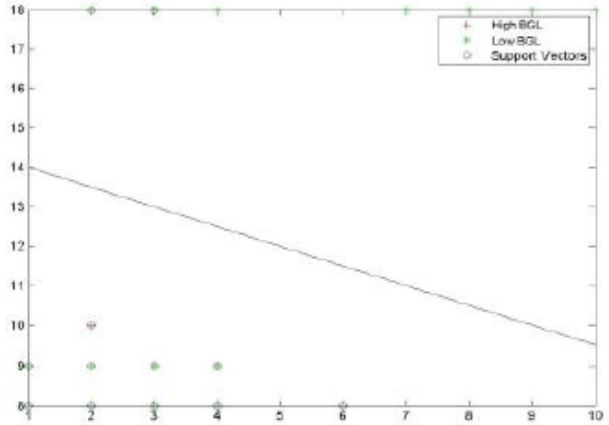

Figure IV: Line separating the blood glucose level

A conclusion section must be included and should indicate clearly the advantages, limitations, and possible applications of the paper. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extentions.

VI. Conclusion

As seen as per the results obtained it was found that SVM was best classified using SVM with Kernel function set to rbf and box constraint set to inf. The data was classified using the support vectors. A support vector was formed. From this it was concluded that in the given database the diabetes subjects suffering from

hypoglycemia was found maximum at blood glucose value 14 to 16. This technique will be useful for pathologist to find out the maximum level of subjects suffering from hypoglycemia in a wide dataset.

Acknowledgements

I am thankful to Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science for providing the database.
I am also thankful to University Grant Commission (UGC) for providing me a financial support in the form of Rajiv Gandhi National Fellowship.

References

[1]. "Better Health Channel". Copyright © 1999/2016 State of Victoria [Internet]. Cited 2016 August 27. Available from https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/diabetes-and-insulin?viewAsPdf=true.

[2]. Anupriya K Kamble, Ramesh R Manza and Yogesh M Rajput, Review on Diagnosis of Diabetes in Pima Indians, *IJCA Proceedings on National Conference on Digital Image and Signal Processing* NCDISP 2016(2):12-15, August 2016.

[3]. Shoback, Greenspan's basic & clinical endocrinology. Edited by David G. Gardner, Dolores(9th ed.) (New York:McGraw-Hill Medical, 2011).

[4]. "National Institute of Diabetes and Digestive and Kidney Diseases" [Internet]. Cited 2015 Dec. 1. Available from https://www.niddk.nih.gov/health-information/health-statistics/Pages/default.aspx#category=diabetes.

[5]. "Diabetes Fact Sheet No312". WHO[Internet]. [updated October 2013; cited 2015 Dec 1]. Available from http://www.who.int/mediacentre/factsheets/fs312/en/.

[6]. Lichman, M. 2013. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[7]. Mathworks, [Internet] Cited 2016 September 20. Available from http://in.mathworks.com/help/stats/support-vector-machines-for-binary--classification.html.