

Early Detection of Diabetic Mellitus Based On Modeling Techniques

Pydipala Laxmikanth, Bhramaramba Ravi

Abstract: Rapid growth of population in particular in elderly, signifies the issues of healthcare have become a concern. Also the lifestyle changes together with social and economical factors influences the cause of disease generation among the generated diseases, diabetic disease is mostly populated. Therefore effective measures are to be taken so that the early wakeup with regard to disease treatment helps to minimize the after effects. In this article, a novel methodology based on Gaussian mixture model built for analyzing the patients and help to identify the disease during the primitive stage. The methodology is presented based on PIMA INDIAN DATASET. The Results derived showcase efficacy of the developed method.

Index Terms: Diabetes mellitus, Gaussian mixture model, disease detection, probabilistic method, early identification

I. INTRODUCTION

Diabetes Mellitus is a kind of metabolic disease wherein the individuals suffering with this disease posses high Blood glucose. This may be due to lack of insulin production or may be due to poor response of the human cells towards insulin and thus causes the individuals with symptoms like frequent urination, thirstiness and hungriness. Most of the literature in this area have classified this disease as a metabolic disorder [1][2][3]. The conditions that trigger are mostly due to radical increase levels of insulin or decrease in insulin levels. Generally, once the disease identified, it is uncontrollable, and leads towards further complications which may lead towards mortality [4][5]. Frequently, used terminologies with respect to this disease classify it into two types, type1 and type2. In typical cases, wherein the conditions of the human body does not generate adequate hormones which allow to absorb and utilize the glucose levels, it treated as type1 disease and hormone deficiency is considered to be insulin [6] insulin resistance is generally considered as a pre-cursor for type2 diabetes and in most of the cases the factors that are assumed to have an impact includes overweight and insufficient or inactive lifestyle. Type1 diabetes leaves no clues for prompting and in most of the cases. To prevent the disease, insulin is to be advocated at regular intervals. The main impacts of improper treatment and care leads towards critical diseases like; retinopathy, neuropathy, nephropathy and even cardiac diseases [7][8][9]. In contrary type2 diabetes is considered to be less dangerous, if proper care is taken.

Revised Manuscript Received on October 20, 2019.

The cases with respect to these disease are mostly increasing not only in India but also it is observed that an overall of 20 % growth rate is observed worldwide every year in the number of patients suffering with this disease. Therefore effective steps are to be taken to identify the disease at the earlier stage and plan effective measures such that, the after impact can be minimized. Lot of activate research has been witnessed in this area and most of the works presented by various authors focus on either identifying the sensitivity of insulin and understanding the impact of exercises on plasma glucose and insulin levels [10][11]. Models are also presented for critically examining in the patients of type1 and type2 individually and suggestive methods are there by offered with regard to the physical exercises [12]. methodologies based on data mining techniques like; decision trees, Bayesian, k-nearest neighbor also highlighted in the literature to understand the impact of the disease and measures to be taken to narrow down its impact onto the other organs [13][14][15]. Models based on neural networks, optimization techniques are also quoted [16][17][18][19]. However, in majority of the works presented, least importance is given towards the early detection of the disease rather than combating the after effects of the disease. Therefore in this article, methodology is proposed which aims at early detection of the disease using Gaussian mixture models [20]. The main objective behind the choice of Gaussian mixture models includes its abilities to identify the abnormal increase of blood sugar levels having complex densities. By adjusting the means and covariance's, linear combinations can be generated with varying levels of blood levels, which helps to identify the early impact of the disease during the start stage. The remaining part of the paper is articulated in the following fashion: In section 2, the Gaussian mixture model is presented. The section 3 highlights the article by generating an alert condition. In section 4, the dataset considered is presented. The section 5 of the article, deals with the experimentation. In section 6, the results derived are analyzed. The concluding section 7 summarizes the article.

II. GAUSSIAN MIXTURE MODELS

It is a probabilistic methodology which assumes that the data is modeled using a finite mixture of Gaussian distributions. It generally considers the Bayesian densities. In one way it is assumed to be a generalization of k-means clustering coupled the covariance structure and about the cluster centers. It has a basic significant property where in case of complex data, sufficient number of Gaussian can be generated and then means and covariance's together with them linear combination can be approximated most proximately. The most specific symptoms are considered and

these symptoms are given to the model.

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \quad \text{--- (1)}$$

Where, $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Where the i^{th} component is characterized by a Gaussian distribution with mean, μ , and covariance $\boldsymbol{\Sigma}$ and mixed weight, π

This value to get included for Bayesian estimation, should be multiplied with a factor

$$P(\theta/x) = (2) \sum_{i=1}^k \pi_i N(u_i, \boldsymbol{\Sigma}_i) \quad \text{----Equation (2)}$$

III. ALERT CONDITION

Using the above section II, the patients data is collected and used for training and testing data set. During the training phase, the probability density function is calculated against each data. The same processes are implied during the testing phase and for that class where the densities matching maximum are considered to be belonging to the same class. Every patient data is considered in this model by a Gaussian Mixture Model. Each of the density functions with respect to the parameters such as low blood sugar, high blood sugar, pancreases levels are stored in the database. Whenever an individual need to assess the values attributed against these parameters are to be compared and if there is any

abnormality is to be considered to be sign for the disease i.e. if each of these values obtained in the test case are more than they values during the training set it is assumed to be a disease.

IV. DATA SET

In order to experiment, in this article, we have considered the Pima Indians Diabetes Database: The present dataset contains the medical related records which was collected from the Pima Indians and the same we can able to download from the below URL. In this data set we are going to find whether or not each and every individual patient record will have a diabetes disease within 5 years.

<https://www.kaggle.com/kumargh/pimaIndiansDiabetes.csv>

Table 1. PIMA INDIAN DATASET - Description of data set

Pregnancy in Ladies	Calculate the Count of Pregnancies
Body-Glucose	Calculate the body glucose concentration in every 2 hours interval of time from GTT.
Blood Pressure	Identifies the DBP (I,e Diastolic blood pressure) in mm Hg.
Skin Thickness	Calculate the skin fold thickness of the large muscle at the back of the upper arm (mm)
Measurement Value of Insulin	Finds the 2-Hour time interval serum insulin level in mu U/ml.
Body Weight and Index	Find out the body weight (i.e mass) and index in terms of kg/(height in m) ²
DPF	Calculate Diabetes pedigree function
Person Age	Calculate the Age of the Patient(In years)
Final Outcome	Calculate the class variable (either 0 or 1) 268 of 768 are 1, the others are 0

V. EXPERIMENTATION

In ordered to present the proposed model, each of the patients the data acquired and corresponding probabilities are identified. If the deviations with respect to the covariance and the mean values change abnormally, it is identified to be disease.

Early Detection of Diabetic Mellitus Based On Modeling Techniques

Table 2. Corresponding probabilities of individuals

Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Class	PDF
72	35	0	33.6	0.627	50	1	23.09
66	29	0	26.6	0.351	31	0	0
64	0	0	23.3	0.672	32	1	21.121
66	23	94	28.1	0.167	21	0	1.232
40	35	168	43.1	2.288	33	1	26.32
74	0	0	25.6	0.201	30	0	0.0009
50	32	88	31	0.248	26	1	32.45
0	0	0	35.3	0.134	29	0	0.0012
70	45	543	30.5	0.158	53	1	43.342
96	0	0	0	0.232	54	1	23.345
92	0	0	37.6	0.191	30	0	0.0352
74	0	0	38	0.537	34	1	32.45
80	0	0	27.1	1.441	57	0	0.1337
60	23	846	30.1	0.398	59	1	29.51
72	19	175	25.8	0.587	51	1	21.21
0	0	0	30	0.484	32	1	20.11
84	47	230	45.8	0.551	31	1	31.46
74	0	0	29.6	0.254	31	1	30.22
30	38	83	43.3	0.183	33	0	1.4458
70	30	96	34.6	0.529	32	1	19.32
88	41	235	39.3	0.704	27	0	1.6248
84	0	0	35.4	0.388	50	0	2.0168
90	0	0	39.8	0.451	41	1	19.45

80	35	0	29	0.263	29	1	20.079
94	33	146	36.6	0.254	51	1	21.343
70	26	115	31.1	0.205	41	1	24.137
76	0	0	39.4	0.257	43	1	28.168
66	15	140	23.2	0.487	22	0	2.9719
82	19	110	22.2	0.245	57	0	2.9802
92	0	0	34.1	0.337	38	0	3.7733
75	26	0	36	0.546	60	0	4.0862
76	36	245	31.6	0.851	28	1	34.093
58	11	54	24.8	0.267	22	0	4.1102
92	0	0	19.9	0.188	28	0	4.1121
78	31	0	27.6	0.512	45	0	4.1121
60	33	192	24	0.966	33	0	4.1158

VI. ANALYSIS

In this section we try to analyze and compute the similarities between the insulin and age as one relation property and another one between insulin and BMI. These correlation values try to find out the potency of these two similar variables either a positive or negative. By using the **pearson correlation coefficient** metric to test the **connotation** of the correlations, we finally make two types of graphs:

Table 3. Correlation matrix (Pearson)

Variables	Insulin	Age
Insulin	1	-0.042
Age	-0.042	1

Table 4. Confidence intervals (95%) / Lower bound

Variables	Insulin	Age
Insulin	1	-0.113
Age	-0.113	1

Table 5. Confidence intervals (95%) / Upper bound

Variables	Insulin	Age
Insulin	1	0.029
Age	0.029	1

Table 6. Coefficients of determination (Pearson)

Variables	Insulin	Age
Insulin	1	0.002
Age	0.002	1

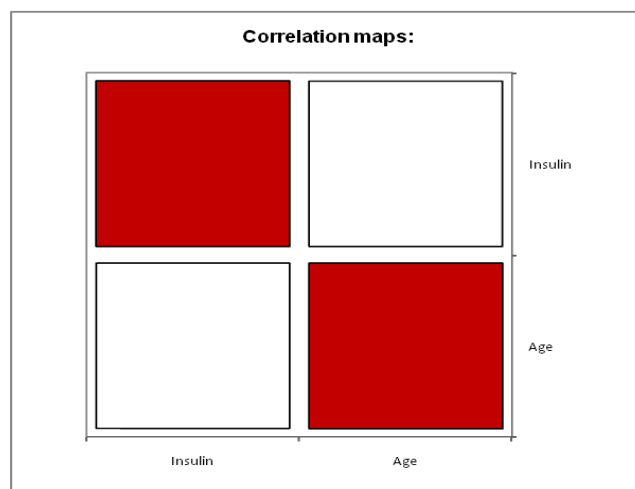


Figure 1. Correlation maps between Insulin and Age

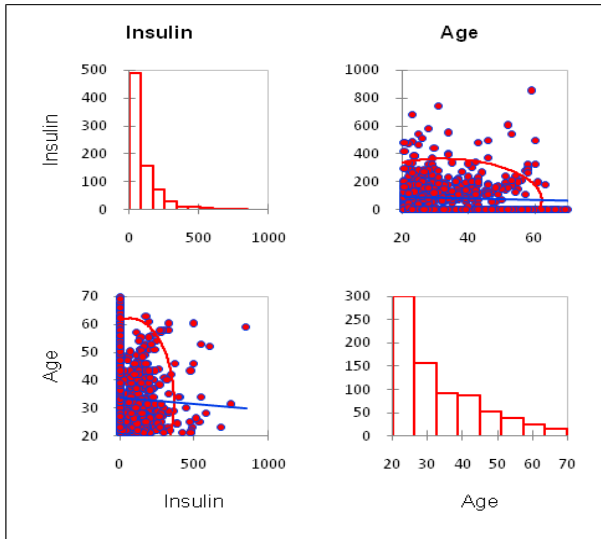


Figure 2. Represents the Comparative analysis of Age-related and insulin affected from the given dataset. Age-related shrink in tissue and activities leads to a decline in physical activities and energy expenditure, inflicting a relative increase in overweight.

Table 7. Correlation matrix (Pearson):

Variables	Insulin	BMI
Insulin	1	0.198
BMI	0.198	1

Table 8. Confidence intervals (95%) / Lower bound

Variables	Insulin	BMI
Insulin	1	0.129
BMI	0.129	1

Table 9. Confidence intervals (95%) / Upper bound

Variables	Insulin	BMI
Insulin	1	0.265
BMI	0.265	1

Table 10. Coefficients of determination (Pearson)

Variables	Insulin	BMI
Insulin	1	0.039
BMI	0.039	1

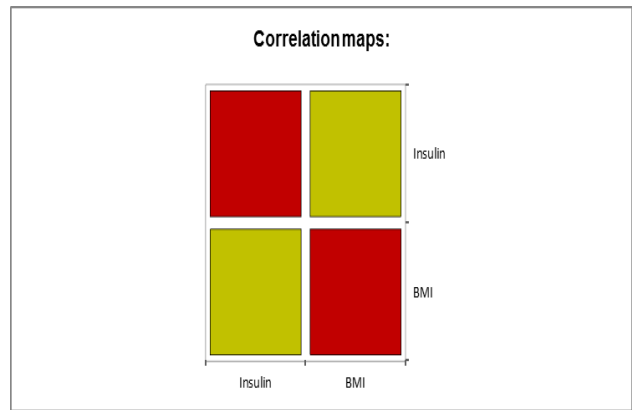


Figure 3. Correlation maps between Insulin and BMI

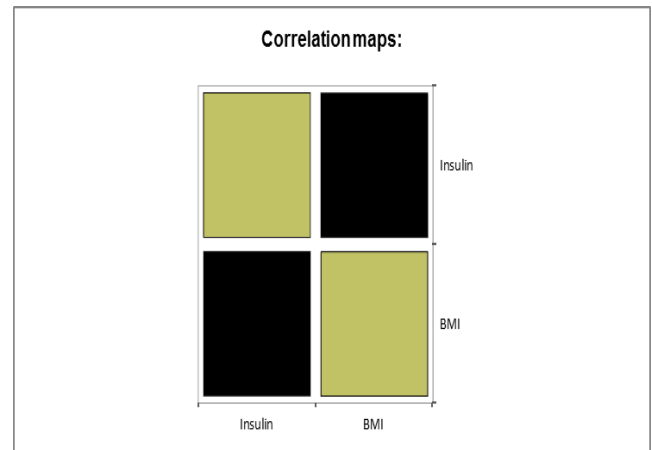


Figure 4. Correlation maps between Insulin and BMI

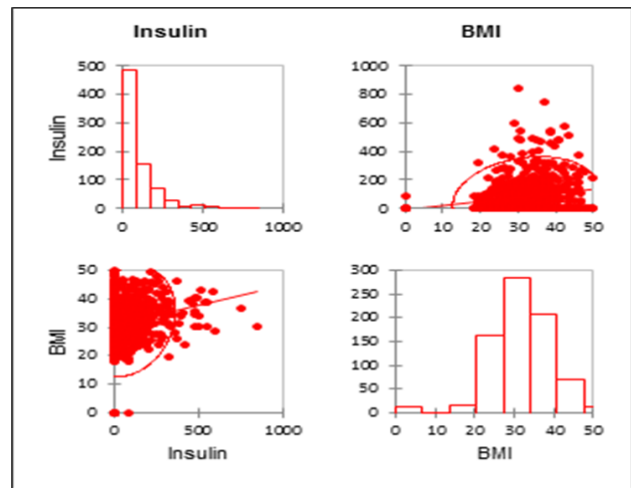


Figure 5. Therefore, in spite of of the population group, growing weight tends to get poorer insulin resistance with BMI.

VII. ILLUSTRATING THE RESULTS FOR Pearson Correlation Coefficients

The current results are the explanatory statistics for all variables (mean, standard deviation, etc). The given correlation matrix is then shown followed by the 95% lower and upper confidence



bounds for the correlation coefficients. The range of the Correlation coefficients lies between -1 and 1. Negative values imply negative correlation, and positive values denote positive correlations. Values nearer to zero indicate the nonappearance of correlation. The coefficients of determination correspond to the squared correlation coefficients. They quantify the strength of the correlation, even if it was negative or positive. We selected to exhibit only the 4 variables for which the sum of R² with supplementary variables is the highest. The color of the data points in the scatter plots reveals whether there is a positive (red) or negative correlation (blue or violet). The markings set up in the scatter plots specify the type but also the power of the relationship between two variables. All the coefficients emerge to be significant at a 0.05 significance level (values in bold). In another way, the possibility of rejecting the null hypothesis (coefficient =0) whereas this is true is less than 5%. The output of this can be recognized from a table of the p-values below (p-values < 0.0001). The correlation map is a graph resembling the diagram which uses a blue-red scale to exhibit the correlations. The blue color correlates to a correlation close to -1 and the red color dovetail to a correlation close to 1. The subsequent graph is a matrix of plots. A histogram is shown for each variable (diagonal) and a scatter plot for all consolidation of variables.

VIII. CONCLUSION

In this article a novel approach is proposed for analyzing the patient's data so as to identify the diabetic disease at the early stage based on Gaussian mixture model. This model can be into a mobile application and can be supplemented in the health bands and showcasing the blood sugar levels and thereby if any changes in these values are forcing, immediate steps can be incorporated so has to combat the disease. The proposed methodology will be very much helpful to govern the type2 diabetes mellitus in particularly.

REFERENCES

1. Zhu, J., et.al.(2015), "An improved early detection method of type-2 diabetes mellitus using multiple classifier system". in Information Sciences; PP-1-14.
2. Nguyen, T., et.al.(2015), "Classification of healthcare data using genetic fuzzy logic system and wavelets. in Expert Systems with Applications", PP-2184-2197.
3. Sanz, J.A., et.al, (2014), "Medical diagnosis of cardiovascular diseases using an interval valued fuzzy rule-based classification system". in Applied Soft Computing ; PP-103-111.
4. Jayalakshmi, T., and Santhakumaran, A.,(2010), "A novel classification method for diagnosis of diabetes mellitus using artificial neural networks". In: Data Storage and Data Engineering (DSDE), in International Conference on. IEEE; PP- 159-163.
5. Meng, X.H., et.al,(2013), "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors". The Kaohsiung journal of medical sciences;29(2):PP-93-99.
6. Lee, C.S., and Wang, M.H.(2011), "A fuzzy expert system for diabetes decision support application". Systems, Man, and Cybernetics, Part B: Cybernetics, in IEEE Transactions on ; PP-139-153.
7. Aribarg, T., et.al,(2011), "Optimizing the modified fuzzy ant-miner for efficient medical diagnosis". in Applied Intelligence :PP-357-376.
8. Ganji, M.F., and Abadeh, M.S.,(2011), "A fuzzy classification system based on ant colony optimization for diabetes disease diagnosis". in Expert Systems with Applications ;38(12):PP-14650-14659.
9. Imam, K.(2013), "Clinical features, diagnostic criteria and pathogenesis of diabetes mellitus". In Diabetes. Springer; PP- 340-355.

10. Ferenci, T., et.al,(2015), "The interrelationship of hba1c and real-time continuous glucose monitoring in children with type 1 diabetes". in Diabetes Research and Clinical Practice.
11. Jaacks, L.M.,et.al (2014), "Dietary patterns associated with hba1c and ldl cholesterol among individuals with type 1 diabetes in china". in Journal of Diabetes and its Complications;.
12. Takao, T.,et.al,(2014), "Association between hba1c variability and mortality in patients with type 2 diabetes. in Journal of Diabetes and its Complications. :PP-494-499
13. Iftene, F., et. al,(2013), "Identification of primary polydipsia in a severe and persistent mental illness outpatient population: A prospective observational study". in Psychiatry research; :PP-679-683.
14. Chhabrani, P.R., Ali, S.M., Nigam, C., Gahlot, R., Chhabrani, J.R., NK, P., "Perspective of diabetes mellitus in dentistry". in International Journal of Contemporary Dentistry 2013; 4(1).
15. Castanho, M., et.al(2013), "Fuzzy expert system for predicting pathological stage of prostate cancer". in Expert Systems with Applications; 40(2):PP-466-470.
16. Keles, A.and Yavuz, U,(2011), "Expert system based on neuro-fuzzy rules for diagnosis breast cancer". in Expert Systems with Applications ;38(5):PP-5719-5726.
17. GaldeanoF "manifestations of diabetes mellitus: clinical meaning".
18. Goyal A, Raina S "Pattern Of Cutaneous Manifestations In Diabetes Mellitus".
19. Chatterjee N "An observational study of cutaneous manifestations in diabetes mellitus Metab". 2014;
20. Nemisy NM.." Skin manifestations in Egyptian diabetic patients: a case series study". in Egypt J Dermatol Venerol. 2013; 33:56-62.
21. Timshina DK " A clinical study of dermatoses in diabetes to establish its markers" in Indian J Dermatol. 2012; 57:20-5.
22. Vijay Gupta., "A Study on Cutaneous Manifestations in Patients of Type 2 Diabetes Mellitus in a Tertiary Care Hospital.2012,5.
23. Shahzad M, and Al Moteri B., "Skin manifestations in diabetic patients attending a diabetic clinic in the Qassim region", 2011; 20:137-41.
24. Sofia Benbelkacem and Baghdad Atmani," Random Forests for Diabetes Diagnosis" in IEEE, 78-1-5386-8125-1/19, (2019).
- 25) B.V. Baiju and Dr.D. John Aravindhar(2019)," Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining" in IEEE,, PP-1-6.

AUTHORS PROFILE



P.LAXMIKANTH is pursuing Ph.D from GIT, GITAM DEEMED TO BE UNIVERSITY, and Visakhapatnam. He is received M.Tech from GITAM COLLEGE OF ENGINEERING, affiliated to Andhra University, Visakhapatnam, India in 2007 and M.Sc from SIR C.R.Reddy P.G.college Eluru, affiliated to Andhra University, India. He has published 10 papers in various reputable national/International journals and conferences. He is an active member of CSI. His research interest are Data mining, image processing.



Dr. Bhramaramba Ravi obtained her Ph.D from JNTUH in the year 2011. She has about 19 yrs of experience in engineering colleges and is currently working as a professor in the Dept. of IT, GIT, GITAM, and Visakhapatnam. She has about 32 publications in reputed Journals. Her area of interest is Data Mining and Bioinformatics.

