Dr. J. Jegathesh Amalraj *et al*, International Journal of Computer Science and Mobile Applications,
Vol.6 Issue. 7, July- 2018, pg. 78-83 **ISSN: 2321-8363**
UGC Approved Journal
Impact Factor: 5.515

Analyzing Performance of the Different Classifiers on Diabetic Dataset with Genetic Algorithm as Pre-Processor

Dr. J. Jegathesh Amalraj[1], Dr. M. Sivakumar[2]
[1]Assistant Professor, Department of Computer Science, Thiruvalluvar University Constituent College, Cuddalore, Tamilnadu, India
amal.jas@gmail.com
[2]Assistant Professor, Department of Mathematics, Thiruvalluvar University Constituent College, Cuddalore, Tamilnadu, India
sivamaths.vani_r@yahoo.com

***Abstract:** Diabetes Mellitus has emerged as a worldwide epidemic. A number of people will appreciate if diabetic diagnostic aid is provided by using a set of data with only medical information and not with any advanced medical equipment. This can make a huge positive impact in the life of lot of people. The aim of this study is to test the diabetes data in two cases as with and without pre-processing and observe the difference in the classification accuracy. Depending upon the results obtained, further decision is made. It is observed that there is increase in accuracy in classification after pre-processing since unwanted data always lead to a decrease in the performance or classification accuracy.*
***Keywords:** Classification, Pre-processing, Diabetes, Machine Learning, Genetic Algorithm, Classifiers, Accuracy.*

1. Introduction

Machine Learning is about learning structures from the data. Automated learning has fetched a greater amount of concentration in medical field due to fewer amounts [8] of time for recognition and less communication with patient, saving period for patients care. Diabetes, which is one of the chronic diseases, is caused due to the increase in blood sugar level. It can be classified into diabetes 1, diabetes 2 and gestation diabetes. Symptoms for diabetes include blurry vision, fatigue, hungry, urinary and excess thirst with weight loss or gain [9, 10]. An intelligent output can be obtained from machine learning algorithm which yields output by recognizing complex patterns. It is one of the major ways for disease classification. It was identified that the machine learning techniques can increase the early detection of disease [11].

The paper is organized as follows: section II describes the related work regarding this work. Section III explains the methodology adopted in this paper and results obtained after experimentation are discussed in Section IV. Section V is the conclusion of the paper.

2. Related Works

Discriminant analysis and the Support Vector Machine (SVM) was used for diabetes classification [1] which applied an accuracy of 82.05%. General Regression Neural Network (GRNN) [2] was developed for diabetes classification which yielded an accuracy of 80.21%. A method based on Genetic Programming [3] was proposed which comprised of three stages namely, feature selection, features generation and testing. Two

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

classifiers were used to evaluate the performance of the features. Fuzzy Neural Network (FNN) and Artificial Neural Network (ANN) [4] was developed and validated using k-fold cross validations. An accuracy of 84.24% and 86.8% was obtained by this method. An intelligent system based on Small-World Feed Forward ANN [5] (SW-FFANN) was proposed which yielded an accuracy of 91.66%. FCS-ANTMINER [6] was developed from the Ant Colony Optimization to extract a set of fuzzy rules to classify the diabetes disease which obtained an accuracy of 84.24%. Morlet Wavelet Support Vector Machine (MWSVM) and the Linear Discriminant Analysis (LDA) was used to develop an automatic diagnosis system called LDACMWSVM [7]. The accuracy was around 89.74%.

3. Methodology:

The methodology works as follows: the diabetic data set is classified using the classifier and evaluated using different metrics and the results are tabulated. And then Genetic Algorithm is applied for Pre-processing and the data set is then classified using the same classifiers and the output is also tabulated. The classifiers like Naïve Bayes, Multilayer Perceptron Network (MLPN), Radial Basis Function (RBF) Network, PART, ZeroR and J48 are selected for classifying the diabetic dataset. The evaluation metrics which are used for performance evaluation are TP Rate, FP Rate, Precision, Recall, F-Measure and accuracy.

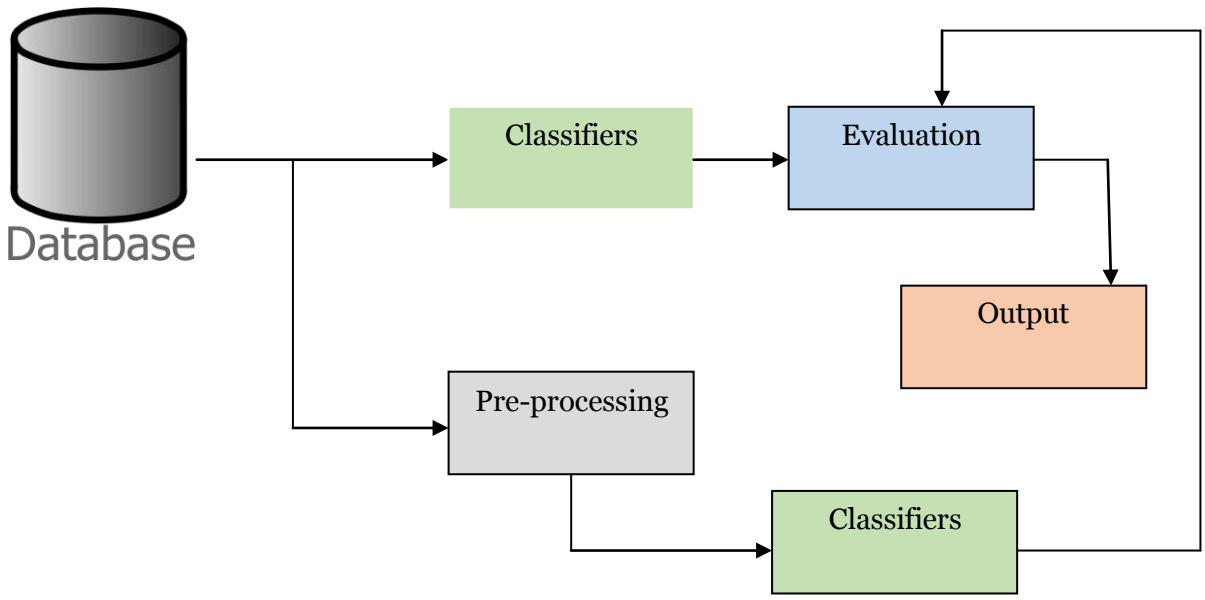

Figure 1: Proposed Methodology

4. Results:

The diabetic dataset collected from UCI [12] is loaded in MatLab [13] and experimented. The classification result obtained using the classifiers are tabulated in Table 1. It was observed from the Table 1 that the Naïve Bayesian classifier has more accuracy than the other classifiers. Figure 2 interprets the results obtained from experimentation.

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

Table 1: Classification Result obtained before pre-processing

Classifiers	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Naïve Bayes	0.76	0.31	0.76	0.76	0.76	**76.30**
MLPN	0.75	0.31	0.75	0.75	0.75	75.39
RBF	0.75	0.35	0.74	0.75	0.74	75.39
PART	0.75	0.33	0.75	0.75	0.75	75.26
ZeroR	0.65	0.65	0.42	0.65	0.51	65.10
J48	0.74	0.33	0.74	0.74	0.74	73.83

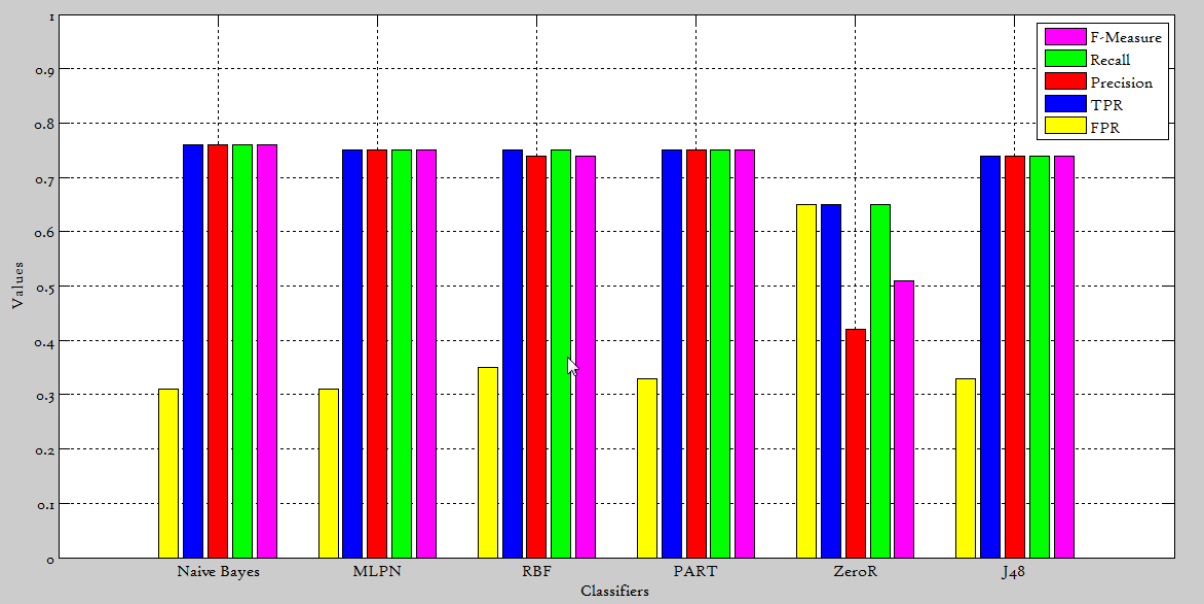

Figure 2: Classification results obtained before pre-processing

After classifying the dataset without initial level of pre-processing, the data set is again pre-processed using the Genetic Algorithm. The data set pre-processed using the Genetic Algorithm is then classified using the classifiers mentioned already. It was observed from the Table 2 that the classification accuracy and other metrics which are used to measure the performance of the classifiers have been improved after pre-processing the dataset. It was observed that the unprocessed data set may lead to certain decrease in accuracy during classification. Figure 3 depicts the result obtained after pre-processing. Figure 4 compares the improvement in accuracy before and after pre-processing. In both the case it was observed that there is difference in performance of the classifiers before and after pre-processing. After pre-processing the data set, it is observed that the classification accuracy of PART is better than the other classifiers.

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

Table 2: Classification results obtained after pre-processing

Classifiers	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
Naïve Bayes	0.78	0.29	0.78	0.78	0.78	78.21
MLPN	0.6	0.4	0.62	0.62	0.62	63.25
RBF	0.82	0.23	0.82	0.82	0.82	82.89
PART	0.86	0.14	0.86	0.85	0.85	86.23
ZeroR	0.7	0.3	0.63	0.71	0.72	70.23
J48	0.79	0.1	0.79	0.8	0.79	80.68

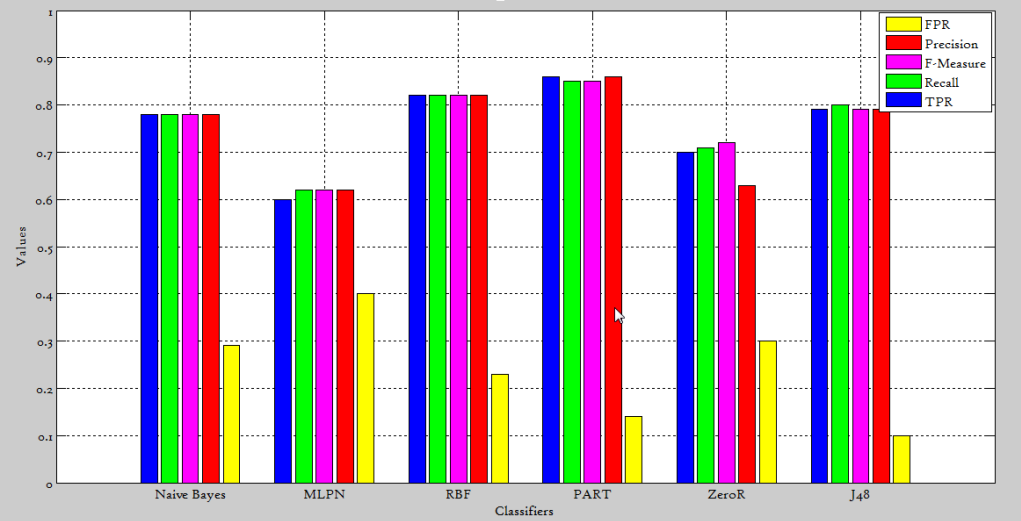

Figure 3: Classification results after pre-processing

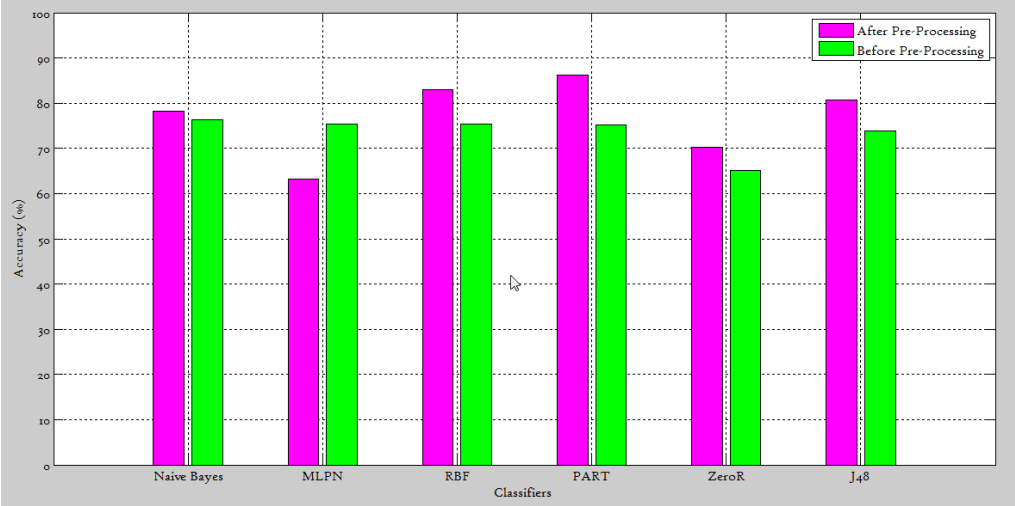

Figure 4: Comparison of Accuracy

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

5. Conclusion

From the experimentation results, it can be observed that the pre-processing can further improve the accuracy in classification. Conventional Genetic Algorithm was applied for pre-processing in this work where potential direction of this work is to advance the Genetic Algorithm in order to progress the precision and also to find a suitable classifier which will give an improved accuracy.

References

[1] K. Polat, S.Gne, A. Arslan, "A cascade learning system for classification of diabetes disease: Generialized Discriminant analysis and least support vector machine", Expert systems with applications, Vol 34, pp 482-487, 2008.

[2] K. Kayaer, T. Yold, "Medical diagnosis on Pima Indian diabetes using general regression neural networks", Proceedings of the International Conference on Artificial Neural Networks and Neural Information Processing, pp 181-184, 2013.

[3] M.W. Aslam, Z. Zhu, A.K. Nnadi, "Feature generation using genetic programming with comparative partner selection for diabetes classification", Expert Systems with Applications, Vol 40, pp 5402-5412, 2013.

[4] H. Kahramanli, N. Allahverdi, "Design of a hybrid system for the diabetes and heart disease", Expert Systems with Applications, Vol 35, pp 82-89, 2008.

[5] O. Erkaymaz, M. Ozer, "Impact of small-world network topology on the conventional artificial neural network for the diagnosis of diabetes", Choas, Solitons and Fractals, Vol 83, pp 178-185, 2016.

[6] M. F. Ganji, M.S. Abadeh, "A fuzzy classification system based on ant colony optimization for diabetes disease diagnosis", Expert Systems and Applications, Vol 38, pp 14650-14659, 2011.

[7] D. Yali, E. Donantekin, "An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier", Expert Systems with Applications, Vol 38, pp 8311-8315, 2011.

[8] Priyanka Kakria, N. K. Tripathi, and Peerapong Kitipawang, "A Real-Time Health Monitoring System for Remote Cardiac Patients Using Smartphone and Wearable Sensors". International Journal of Telemedicine and Applications, Vol 2015, pp 1-11, 2015.

[9] www.symptomchecker.webmd.com/multiple-symptoms?symptoms=fatigue%7Cfrequenturination%7Chunger%7Cincreased-thirst&symptomids=98%7C107%7C547%7C124&locations=66%7C35%7C66%7C7

[10] www.endocrineweb.com/conditions/type-2-diabetes/type-2-diabetes-symptoms

[11] Mirzaei G, Adeli A, Adeli H., "Imaging and machine learning techniques for diagnosis of Alzheimer's disease", Rev Neuroscience, Vol 27(8), pp:857-870, 2016..

[12] www.archive.ics.uci.edu/ml/datasets/diabetes

[13] www.mathworks.com/products/matlab.html

[14] Dr. Jayaraj. V, J.Jegathesh Amalraj, "Using Clustering and Indexing to Enhance Customer Relationship Management Based on Customer and Product Value Estimation - A Neural Networks Approach", International Journal of Computational Engineering & Management, 15 Issue 4, pp: 12-15, 2012.

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

Author Biography

Dr. J Jegathesh Amalraj holds the Doctoral Degree in Computer Science from Bharathidasan University, Tiruchirappalli. He has more than five years of teaching and research experience. He is currently working as Assistant Professor in the Department of Computer Science in Thiruvalluvar University Constituent College at Cuddalore, Tamilnadu, India. Presently he is doing his research on Wireless Adhoc Networks, Software Engineering and Data Mining domains.

Dr. M. Sivakumar holds the Doctoral Degree in Mathematics from Manonmaniam Sundaranar University, Tirunelveli. He has more than sixteen years of teaching and research experience. He is currently working as Assistant Professor in the Department of Mathematics in Thiruvalluvar University Constituent College at Cuddalore, Tamilnadu, India. Presently he is doing his research on Graph Labelling, Network Security and Data Mining domains.

