Prediction of Diabetes using Fused Machine Learning

Shweta.S. Nair [1], Samantha.M [1], Sabareshwar.P.S [1], Indhumathi.S [2]

[1] Student, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India.

[2] Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India.

Corresponding Author: shwetasnair0703@gmail.com

Abstract: - The healthcare assiduity is foisted with the plethora of patient data which is being supplemented each day manifold. Experimenters have been continually using this data to help the healthcare assiduity ameliorate upon the way major conditions could be handled. They're indeed working upon the way the cases could be informed timely of the symptoms that could avoid the major hazards related to them. Diabetes is one similar complaint that's growing at an intimidating rate moment. In fact, it can induce multitudinous severe damages; blurred vision, diplopia, burning extremities, order and heart failure. It occurs when sugar situations reach a certain threshold, or the mortal body can not contain enough insulin to regulate the threshold. thus, cases affected by Diabetes must be informed so that proper treatments can be taken to control Diabetes. For this reason, early vaticination and bracket of Diabetes are significant. This work makes use of Machine Learning algorithms to ameliorate the delicacy of vaticination of the Diabetes. The trials also showed that ensemble classifier models performed more than the base classifiers alone. Its result was compared with the same Dataset being applied on specific styles like arbitrary timber, Support Vector Machine, Decision Tree and Naïve Bayes bracket styles.

Key Words: - healthcare, Diabetes, Machine Learning.

I. INTRODUCTION

Multiple Openings for healthcare are created because machine literacy models have implicit for advanced prophetic analytics. There are formerly being models in machine literacy which can prognosticate the habitual illness like heart complaint, infections, and intestinal conditions. There are also many forthcoming models of machine literacy to prognosticate non-communicable conditions, which is adding further and further benefit to the field of healthcare. Experimenters are working on machine literacy models that will offer veritably early vaticination of specific complaint in a case which will produce effective styles for the forestallment of the conditions. This will also reduce the hospitalization of cases. This metamorphosis will be veritably important salutary to the healthcare associations. The most explored area is the healthcare system which uses ultramodern computing ways is in healthcare exploration.

As mentioned above the experimenters in the affiliated fields are formerly working with the healthcare association to come up with further technology ready systems. Diabetes is a complaint which reduces the body's capability to produce insulin. In other words, the body can't avenge to the hormone insulin product. This results in anomalous metabolism of carbohydrates and increased blood glucose situations. Beforehand discovery of diabetes becomes veritably important because of the reasons mentioned over. numerous people in the world are getting affected by diabetes and this number is adding day by day. This complaint can damage numerous vital organs hence the early discovery will help the medical association in treatment of it. As the number of diabetic cases is more there's an inordinate important medical information which must be maintained. With the support of adding technology, the experimenters must make a structure that store, maintain and examine this diabetic information and further see doable troubles. The blood glucose situations come too high in the body when there's diabetes. Glucose is created in the body after eating food. The hormone insulin produced in the body helps balance the glucose situations and regulate blood sugar situations, insufficiency of insulin causes Diabetes. Type 1 diabetes is a script where the body doesn't produce insulin at all to balance the sugar situations in blood. Type 2 is a diabetes type where the body produces insulin but doesn't use this

hormone fully to balance blood sugar situations. The Type 2 diabetes is most common bone. There's commodity called as pre diabetes, this is a situation where the person can have high glucose position but not that high that he she can be said to have diabetes. But the people who have pre diabetes are prone to get type 2 diabetes. This complaint can beget serious damage to numerous vital organs in the body like feathers, heart, jitters, and eyes. However, also it's known as gravid diabetes, if a woman gets this complaint during gestation. By managing our weight, mess plan and exercise we can control diabetes. One should always keep a check on its blood sugar situations.

II. MATERIALS AND METHOD

This paper is about prognosticating diabetes using a proposed fashion that consists of three phases data preprocessing, point significance, and ensembling. The dataset used is the Pima Indian Diabetes Dataset, which has multiple attributes for 768 cases. In the data preprocessing phase, missing values are removed by deleting cases that have a zero value. This process is necessary to ameliorate the quality and effectiveness of the data and reduce dimensionality. The data is also resolve into training and testing sets, and normalization is applied to bring all attributes to the same scale.

In the point significance phase, the Random Forest (RF) algorithm is used to simplify the dataset, and the Logistic Retrogression (LR) algorithm is applied to determine the most frequent attributes that contribute to prognosticating diabetes. Eventually, the proposed fashion uses ensemble literacy to combine several machine literacy models into an optimal prophetic model to ameliorate the delicacy of the vatication. In this study, the AdaBoost algorithm is used for ensemble literacy on the Pima Indian Diabetes Dataset. Ensemble literacy is a machine learning fashion that combines multiple models to ameliorate prophetic performance and reduce bias or friction. There are colorful types of ensemble literacy, similar as bagging, boosting, mounding, and more. In this study, AdaBoost is used for ensemble literacy because it's known to ameliorate bracket delicacy and reduce overfitting. Overall, the proposed fashion for diabetes vatication involves data preprocessing, point significance, and ensemble literacy using the RF and LR algorithms and the AdaBoost ensemble learning algorithm. After data preprocessing and point selection, colorful machine learning algorithms are applied to prognosticate diabetes. In this study, four different algorithms and an ensemble system are used Support Vector Machine (SVM), Decision Tree, Random Forest (RF), and Naïve Bayes. SVM is a supervised machine learning algorithm used to classify data into specific classes by creating a hyperplane that separates the classes. It can also perform retrogression. SVM can separate between data points and classify data that aren't supported by the training data. Decision Tree is an introductory type of supervised literacy algorithm used when the response variable is categorical. It creates a tree- suchlike structure that describes the bracket process grounded on the input variables. Random Forest (RF) is a well- known supervised machine learning algorithm that can perform both bracket and retrogression tasks. It uses ensemble literacy, which combines the results of multiple classifiers to ameliorate the overall delicacy. Naïve Bayes is a machine learning algorithm generally used for textbook bracket. It's simple to apply and largely effective. In summary, the proposed fashion for prognosticating diabetes involves applying different machine learning algorithms, including SVM, Decision Tree, RF, and Naïve Bayes.

III. RESULTS AND FINDING

The perpetration has successfully carried out in python programming language. The colorful ML algorithms have enforced and tested. Like DT, SVM, LR, NB and Hybrid Method. To find out the precise value is being used to get several positive results that can show a positive response, the delicacy standard can depend on four features. These features are as follows:

True Cons TP is a proper value that's meetly linked as an applicable class.

True Negatives TN is an indecorous value that's meetly linked as an unhappy class.

False Cons FP is an indecorous value imperfectly linked as an applicable class.

False Negatives FN is a proper value that's imperfectly linked as an unhappy class.

$$delicacy = ((TP\ TN))\ ((TP\ FP\ FN\ TN))$$

The perfection parameter is the parameter that specifies how numerous values determine duly

$$Precision\ TP\ (TP\ FP)$$

The recall parameter is the parameter that specifies how numerous applicable values are determined. The recall is a

value set up by the true positive value and divided by the total rate of true positive and false negative values

$$\text{Recall TP (TP FN)}$$

These are the stylish styles to find out results in dimension, but this approach isn't simple for getting the proper decision. Accordingly, the F- measure acquires a single measure for assessing results. The F- measure approach is the harmonious mean of perfection and recall

$$\text{F- Measure (2 x (perfection x recall)) (precession recall)}$$

Model	Accuracy	Precision	Recall	F-measure
SVM	93	77	91	83
DT	86	88	83	85
LR	84	79	92	85
NB	77	73	87	77
DPE(Hybrid)	95.5	85	89	87

IV. CONCLUSION

Though different models had been used for the prediction of diabetes, the accuracy of the proposed models in disease prediction has always been the main concern of researchers. Therefore, a new model is required in order to achieve higher prediction accuracy in diabetes prediction. This research proposed machine learning based diabetes decision support system by using decision level fusion. Two widely used machine learning techniques are integrated in the proposed model by using the fuzzy logic. The proposed fuzzy decision system has achieved the accuracy of 94.87, which is higher than the other existing systems. Through this diagnosis model, we can save several lives. Moreover, the death ratio of diabetes can be controlled if the disease is diagnosed and preventative measures are taken in early-stage.

REFERENCES

[1]. Abdelghani Bellaachia and Erhan Guven (2010), "Predicting Breast Cancer Survivability Using Data Mining Techniques", Washington DC 20052, vol. 6, 2010, pp. 234-239.

[2]. Azhar Rauf, Mahfooz, Shah Khusro and Huma Javed (2012), "Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity", Middle-East Journal of Scientific Research, vol. 12, 2012, pp. 959-963.

[3]. Min Chen, Yixue Hao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang (2017), "Disease Prediction by Machine Learning over Big Data from Healthcare Communities", 2017, IEEE, vol. 15, 2017, pp215-227

[4]. Akhilesh Kumar Yadav, Divya Tomar and Sonali Agarwal (2014), "Clustering of Lung Cancer Data Using Foggy K-Means", International Conference on Recent Trends in Information Technology (ICRTIT), vol. 21, 2013, pp.121-126.

[5]. Kajal C. Agrawal and Meghana Nagori (2013), "Clusters of Ayurvedic Medicines Using Improved K-means Algorithm", International Conf. on Advances in Computer Science and Electronics Engineering, vol. 23, 2013, pp. 546-552.

[6]. Chew Li Sa, Bt Abang Ibrahim, D.H., Dahliana Hossain, E. and bin Hossin, M. (2014), "Student performance analysis system (SPAS)", in Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference on, vol.15, 2014, pp.1-6.

[7]. Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", ScienceDirect, Vol. 11, issue 3, pp. 12-23, 2018.