**Related Work**

Date 10 se 2019

One of the worldwide increasing diseases is diabetes and the disease will subsequently get increasing if countries do not consider preventive measures for disease. Full Analysis and consideration of a patient are not possible nowadays by a doctor as various diseases are so widespread. And so an intelligent system is required for to consider the various factors and identify a suitable model between the different parameters. The use of intelligent systems in the diagnosis and treatment of diseases can significantly reduce medical errors leading to a decrease in both financial and human losses. The aim of this study is to compare the performance of several different classification techniques on a set of data that has been screened for diabetes.

The upcoming most popular and important approach in the field of medical research is Machine Learning. The frequency of diabetes over the years has been increasing dramatically with the aging population worldwide. Increased mortality rate and reduction in the life expectancy of elderly diabetic patients are due to growing diabetes. In clinical diagnosis problems, classification plays a vital role in further treatment of the disease. Various studies have been done on the diabetes data classification using different machine learning algorithm. In this Literature review section we are going to discuss work done on all the classification of diabetes using intelligence system, which are as follows:

**SVM**

Nahla H *et.al.* (2010) have discussed a hybrid model for medical diagnosis which integrates three different data mining and SVM. They highlighted the data mining and machine learning techniques for diagnosis, prognosis, and management of diabetes.  SVMs is employed for finding a linear hyperplane that separates the positive and negative examples with an utmost interclass distance. However, **SVM** does not provide intelligible support for the classification decisions because it is behaving like black-box models. Hence, authors presented idea of intelligible representation of the SVM’s by using data extraction so that diagnosis can be performed efficiently. The penitent dataset was taken for this work from Oman. Results on the real-life prediction of type2 diabetes dataset show a comprehensible rule set which means intelligible SVM provides a promising tool for the prediction of diabetes with prediction accuracy of 94%, sensitivity of 93%, and a specificity of 94%

Kemal Polat and Salih Gunes, *et.al* proposed A cascade learning system for classification of diabetes disease using Generalized Discriminant Analysis and Least Square Support Vector Machine.The result obtained when using LS-SVM is 78.21% classification accuracy using 10-fold cross validation, the GDA–LS-SVM obtained 82.05% classification accuracy using 10-fold cross validation. The highest accuracy obtained was 82.05% and is very promising with regards to other classification. The fallout strongly suggest that Generalized Discriminant Analysis and Least Square Support Vector Machine classifier are based on a learning method can assist in the diagnosis of Diabetes disease.

T. Santhanam and M.S Padmavathi propose a work of Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis.Datasets have been used from Pima Indians Diabetes from UCI repository and the proposed model attained an average accuracy of 98.79 % for the reduced dataset.

Seokho Kang and Pilsung Kang proposed method utilizes data selection methods to lessen the training set before constructing a collection.The model for the prescription level calculation derived by our projected method achieved about 80% classification accuracy for prescription cases.

Muhammad Waqar Aslam and  Zhechen Zhu, *et.al* proposed a model using genetic programing technique with ultimate aim to facilitate the diagnosis of diabetes automatically by figuring out if a patient has diabetes, without the need of a physician.The Pima Indian diabetes dataset from the UCI Repository of machine learning databases has been used.The result shows that GP not only improves the act or the performance but also reduces the eight input dimensions to a single dimension.

Nahla H. Barakat, and  Andrew P. Bradley, *et.al* proposed a model of Intelligible Support Vector Machines for diagnosis, prognosis, and management of diabetes many of the data mining and machine learning methods.Results on the real-life prediction of type2 diabetes dataset show a comprehensible rule set which means intelligible SVM provides a promising tool for the prediction of diabetes with prediction accuracy of 94%, sensitivity of 93%, and a specificity of 94%

Bob Zhang and B.V.K. Vijaya Kumar, *et.al* proposed detecting diabetes mellitus and non-proliferative diabetic retinopathy using tongue color, texture and geometry features.By testing through each feature individually to healthy DM, the highest average accuracy gained was 66.26% via SVM. While employing SFS with SVM, the optimal result was shown by 9 features with an average accuracy of 80.52 %. when testing for NPDR DM the best result came was 5 features with 80.33% average accuracy.

……………………………………………………………………………………………………………….

ANN

Rajeeb Dey and Vaibhav Bajpai et.al proposed the Application of artificial neural network technique for diagnosing diabetes mellitus using  ANN with backpropagation algorithm. The results validate that the network classifies with 92.5% accuracy levels which can be further increased by increasing hidden layers.

T.Jayalakshmi and Dr.A.Santhakumaran proposed Classification method for diagnosis of diabetes mellitus using artificial neural networks. Methodology used are Multi-layer perceptron(MLP) and backpropagation where our data set is incomplete KNN is applied and it shows that model can be applied to real world applications with no statistical data.

Chongjian Wang and Linlin li, et.al proposed  Evaluating risk of type2 diabetes mellitus using artificial neural network. multivariate logistic regression (MLR) with 3 layer neural network is used as methodology. ANN model is a more effective approach for identifying those at high risk of diabetes based on their lifestyle and anthropometric data.

Hasan Temurtas and Nejat Yumusak , et.al proposed The backpropagation algorithm is used for training.  The sample has 8 features, 10 fold cross-validation method for measuring accuracy which is a better option of validation of comparative study on diabetes disease diagnosis using neural network.

**IEEE**

S. Lekha proposed a model where in the experimental data has been collected from the sensory unit by Real-Time Non- Invasive Detection and Classification of Diabetes using Modified Convolution Neural Network.The CNN algorithm produces a set of feature maps by convoluting the raw data signals with a kernel-based filter. These feature extraction and classification systems are successful in diabetes recognition applications.

Konstantia Zarkogianni and Maria Athanasiou , et.al proposed a model.The models were tested using the dataset from the medical record of 560 T2DM patients and the best discrimination performance noted was up to 71.48% in terms of AUC. the results obtained indicate that a hybrid ensemble integrating both the HWNN- and SOM- which are the primary model performs well even if the small number of CVD incidents were included.

………………………………………………………………………………………………………………

**Logistic Regression**

Imran Kurt et. Al. has proposed  the classical arithmetic study between the presence and absence of CAD to examine the difference in the distribution of age variable with the basic motive was to Compare the performance of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease.

Emma Ahlqvist and Petter Storm , et.al proposed a model To compare time to medication, risk of diabetic complications and time to reaching the treatment goal, and genetic associations. A data driven analysis was done where in Cox regression technique, and logistic regression technique were used.The final result came was that the new clustering of patients with the classic diabetes classification was less than adult-onset diabetes.

The SMOTE(Synthetic Minority Oversampling Technique) was given by Manal Alghamdi and Mouaz Al-Mallah,et.al for predicting diabetes mellitus using machine learning approach This study shows that the Logistic regression classifier achieves the highest performance while the decision tree with the lowest performance.This study shows the capability of machine learning for predicting incident diabetes using cardiorespiratory fitness data

Changsheng Zhu and Christian Uwa Idemudia, et.al proposed a data mining based model that is using the Pima Indian diabetes dataset for early diagnosis and prediction of diabetes using k-means clustering and also logistic regression accuracy results.The result experimentally shows that the K-means clustering algorithm and logistic regression classifier accuracy is comparatively high than the published studies.

•Hossein Ebrahimi  & Mohammad Hassan Emamian, et.al proposed a model to determine the prevalence of diabetes and its risk factors among the middle-aged population.. According to the results of this study, increasing age increases the mean blood glucose and the prevalence of diabetes also increased accordingly.This multivariate logistic regression revealed a significant OR and the difference was statistically significant

•Mahmoud Heydari and Mehdi Teimouri, et.al aimed a model to compare the performance of several different classification techniques on a set of data that has been screened for type 2 diabetes. Based on the results obtained, it can be concluded that the efficiency of a model depends on the nature and complexity of datasets used.The most accurate method of classification in this research was an artificial neutral method with 97.18% accuracy.

•K. Saravananathan1 and T. Velmurugan analysed classification like algorithms J48, Support Vector Machines (SVM), Classification and Regression Tree CART and k-Nearest Neighbor (kNN)to best classify the input data of Diabetic dataset.The results show that the performance of the J48 technique is considerably better for the other three techniques for the classification of diabetes dataset.

•Shankaracharya and Devang Odedra,et.al reviewed artificial network and machine learning approch towards diabetes dataset with the basic aim is to sketch out the range of development and potential machine learning algorithms as diabetes diagnosis tools. and . It was suggested that within specific geographic regions models of the decision must be prepared on a dataset that intently represents the patient’s profile.

………………………………………………………………………………………………………………

**Naive Bayes**

* Rashedur M. Rahman and Farhana Afroz proposed a Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis. Methodology used includes Multilayer Perceptron, Bayes Net ,Naïve Byes, J48graft, Fuzzy Lattice Reasoning, Adaptive Neuro-Fuzzy Inference System and Performance Metrics. Naïve Bayes  classifier has the highest accuracy above all.
* Aiswarya Iyer, *et.al* proposed DIAGNOSIS OF DIABETES USING CLASSIFICATION MINING TECHNIQUES using decision tree and Naïve Bayes algorithm. The results show that 70:30 percent split for Naïve Bayes gives least error and the model prepared is quite effective.
* Dewan Md. Farid and Li Zhang *et.al* proposed DIAGNOSIS OF DIABETES USING CLASSIFICATION MINING TECHNIQUEs using Decision tree, Naïve Bayes as methodology. The NB classifier obtained average accuracy rate of 86.7%. naïve bayes tree (NBTree), rough set approach and fuzzy logic are accurate and will be implemented in future works .
* Yue Huang a and , Paul McCullagh, *et.al* gave Feature selection and classification model on type 2 diabetic patients. Methodology include naïve Bayes , decision tree learner and instance-based learner and feature selection . result: applied in real world, accurately classified models.

………………………………………………………………………………………………………………..

[Randomforests](https://docs.google.com/document/d/1eZQ56CtDJGw73WjjqU0znWFb3YaFRUH65yRYWYX3Q9I/edit?usp=sharing_eil&ts=5d688f02)

* Beatriz López and, Ferran Torrent-Fontbona *,et.al* proposed Methodology for Artificial intelligence in medicine And compared RF vs SVM and LR. The complexity, overfitting, interactions were handled by RF and it is a useful method for predictive models when compared to SVM and LR.
* Guoyan Zhaoa and , Tommi Vatanenb, *et.al* proposedIntestinal virome changes precede autoimmunity in type-1 diabetes-susceptible children Random forest analysis is used to identify disease-associated virus.(research gap)
* Attila A. Seyhan , YuryO. NuneLopez1z,*et.al* proposed the Pancreas-enriched miRNAs altered in the circulation of subjects. random forest classification is used to estimate the diagnostic odds ratio (DOR) and sensitivity analysis is implemented using the R environment.results: The prediabetic group showed a significant reduction in miR-126 and miR-146a. T2D showed elevated levels of miR-30d, miR-21, miR-34a. T1D exhibited high levels of miR-21 and miR-375.

………………………………………………………………………………………………………

**Springer**

G. Sheikhi and H. Altınçay has proposed model for The Cost Of Type 2 Diabetes Mellitus where Classification is done through Logistic Regression and Support Vector Machines. It is concluded that the burden of diabetes is not limited to a small test, it is beyond these features and needs deep investigations.

N. Yuvaraj · K. R. SriPreethaa proposed Diabetes prediction in health care systems using machine learning algorithms on hadoop cluster. The training dataset is built into a decision tree while the naïve Bayes classifier and random forests with 6% accuracy with random forests producing the highest accuracy. The Hadoop cluster performs much better when compared to the other two machine learning algorithms.

Abid Sarwar and Mehbob Ali,et.al proposed Diagnosis of diabetes type-2 using hybrid machine learning based ensemble model using 15 classifiers out of which 5 major ones are: ANN, SVM, KNN, Naïve Bayes and Ensemble. The artificial immune recognition system-2 (AIRS2) [12] had a version of AIRS2 called MAIRS2 which made use of fuzzy KNN which help improve diagnosis with 89.10% accuracy than 82.69 in AIRS2. The overall efficiency of the diagnostic tool is 97.34% which can be increased by increasing number instances

Md. Maniruzzaman & Md. Jahanur Rahman , et.al proposed an Accurate diabetes risk stratification using machine learning: role of missing value and outliers. Conventional techniques used are LDA (linear discriminant analysis), QDA(quadratic discriminant analysis), NB(naïve Bayes), GPC(Gaussian process classification), SVM(support vector machine), ANN(artificial neural network), Adaboost, LR(logistic regression) and DT(decision tree). random forests(RF) showed the most significant features and accurate diabetes prediction it works well on both non-linear and high dimensional data. RF produces very meaningful results that correlate risk factors too and yields 89% accuracy.

Pankaj Pratap Singh and Shitala Prasad et.al proposed the Classification of diabetic patient data Using The Association Rule Learning (ARL) R to implement statistics here  similar and equally accurate results are shown by both the methods ARL and NN because both are data mining techniques and are efficient.

Ambika Choudhury and Deepak Gupta proposed a model on survey of medical diagnoses of diabetes using machine learning techniques the Methodologies implemented are SVM, KNN,DT,naïve bayes,ANN and LR. result show that LR is most accurate for the classification of diabetic and nondiabetic samples.