A Comparative Study of Machine Learning Algorithm in Diabetics Risk Prediction

[1] Shweta Yadu [2] Rajshri Lanjewar [3] Vivek Kumar Sinha

[1] M. Tech Student, Dept. of CSE, Raipur Institute of Technology, Raipur, Chhattisgarh, India

[2] Assistant Professor, Dept. of CSE, Raipur Institute of Technology, Raipur, Chhattisgarh, India

[3] Head of Dept. of CSE, Raipur Institute of Technology, Raipur, Chhattisgarh, India

Corresponding Author's Email: [1] shwetayadu24@gmail.com, [2] rajeshri21@gmail.com, [3] sinha.vivekkumar7@gmail.com

Abstract— Diabetes is one of the most prevalent chronic illnesses and can affect anyone, regardless of age. Many disorders strike when the glucose or sugar level is too high. Many complications brought on by diabetes contribute to a high rate of diabetic patients being readmitted. With the aid of machine learning, the correct diagnosis can be made after the system has access to sufficient, accurate, and comprehensive information about the issue. It relies on the idea of training and testing the machine with the necessary algorithm that can produce effective results for process execution. Linear regression is one of the machine learning algorithms that we compare and analysis in this research article. Our research focuses on the accuracy, performance, and algorithm criteria that are used for medical diagnosis, and we discovered that the Bayesian Classifier, Multilayer Perceptron (MLP), K-Nearest Neighbour (kNN), Random Forest (RF), and Support Vector Machine (SVM) combination can produce the best results and most effective model for diabetic prediction. By altering the pre-trained model's layer, the outcome can also be improved.

***Keywords*-Machine Learning,Classifier,Feature Extraction, Medical Diagonisis.**

I. INTRODUCTION

One of the diseases with the fastest current growth is diabetes. According to a survey by the World Health Organization, there are 422 million people with diabetes worldwide. It also tells non communicable illnesses cause over 41 million premature deaths each year, or close to 71% of all deaths worldwide. If left unchecked, 52 million people will die each year by 2030 as a result of non-communicable diseases. Diabetes and hypertension are the most prevalent non communicable diseases, causing roughly 46.2% and 4% of all fatalities, respectively[1]. This is typically brought on by an excess of glucose, a sugar molecule generated from carbohydrates, in the blood. As food is broken down into its smallest molecules and nutrients, such as glucose, all cells take it up for the purpose of generating energy with the aid of the hormone insulin produced by the pancreas. Occasionally the body does not produce enough insulin, which prevents cells from absorbing glucose when there is no insulin present[2]. This is mostly a lifestyle disease that is challenging to diagnose in the early stages. By the time it is discovered, it is typically advanced and can only be treated with drugs, with some people also receiving insulin injections to control their blood sugar levels. When blood sugar levels are not properly controlled for an extended period of time, serious organ damage can result, including diabetic retinopathy, which results in vision loss, diabetic neuropathy, which damages the nerves, diabetic foot, as well as damage to the heart, pancreas, kidneys, and many other major organs[3].

An individual's blood sugar levels can be controlled by following a healthy diet and lifestyle. Being diagnosed with diabetes, a person must maintain a healthy lifestyle in addition to taking medication to control their blood sugar since high blood sugar levels can seriously harm their bodies. The best method to control a chronic condition like diabetes is to get frequent health screenings to look for any indications of the body's blood sugar levels becoming out of whack. Even with all of these precautions, diabetes can be difficult to diagnose in its early stages or to forecast when it will first manifest[4]. The procedure of making a medical diagnosis is quite complicated and takes a lot of time and labour. ML approaches in healthcare systems are not limited to diagnosis but can help in drug prediction, managing health-related records, medical assistant, decision-making, etc. Machine learning-based healthcare systems can assist clinicians in finding the findings extremely rapidly. In order to speed up processing and get better results, healthcare professionals and information technology specialists are currently working in the field of ML-based healthcare systems [5]. "Machine learning" (ML) is the collective name for a number of statistical techniques that allow computers to learn from experience without being explicitly programmed. Machine learning (ML) applications are fundamentally altering the healthcare sector. It is a subset of artificial intelligence (AI) technology designed to improve the effectiveness and accuracy of the work done by medical practitioners. Figure 1 illustrates how to use a machine learning algorithm to analyse a set of medical data to make disease predictions.

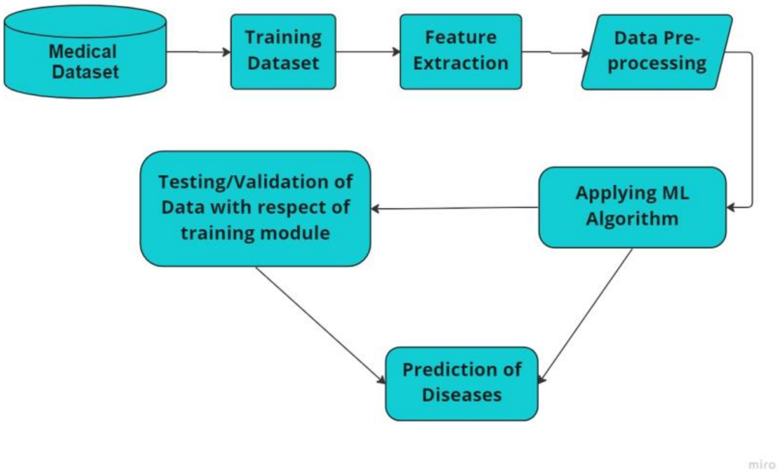

Figure 1: Approach of Machine Learning on Medical Dataset

In this study, we analyse a number of papers that use machine learning methods for diagnosing diabetics and other medical conditions. With a large data set of diabetic patients, scholars have used machine learning for expectation purposes in the medical diagnosis of various diseases. The other sections of the paper are arranged as follows: Section II contains related literature on classification and feature extraction algorithms; Section III contains some existing image classification algorithm technique; and Section IV discusses the findings.

II. RELATED WORK

Malini M et.al[7] create and implement a diabetes prediction utilising several machine learning approaches and the 768-record Pima Native Dataset to assess the efficacy of such approaches. The suggested technique for classification and ensemble learning makes use of SVM, KNN, Random Forest, Decision Tree, Logistic Regression, and Gradient Boosting classifiers. The maximum classification accuracy of 78% was achieved with Logistic Regression.
With scores of 97% and 92%, respectively, compared to 72% and 63%, or more than 25% in each category on average, the A **M Asiful Huda et.al[8]** proposed strategy beats the present results in both precision and recall. The proposed system applies classification algorithms to a variety of features from a dataset on diabetic retinopathy, including optical disc diameter, lesion-specific features (microaneurysms, exudates), or the presence of haemorrhages. After that, the traits were retrieved and employed in the final decision-making process to determine whether diabetic retinopathy was present. The prediction is then made using the Decision Tree, Logistic Regression, and Support Vector Machine. The model had an accuracy rate of 88%, which was much higher than earlier experiments.
Smitha S Prem and Umesh A.C[9] suggests a computer-aided way for classifying exudates that will aid ophthalmologists in predicting DR. The wavelet decomposition coefficient and LBP, which give texture information and frequency information, respectively, in an image, were used to establish the classifications. Several supervised classification approaches are used to evaluate the performance of classifiers. Using the DIARETDRBI dataset with 89 pictures, KNN classifier has improved performance in the proposed model with an accuracy of 94% for LBP features while ANN has improved performance with 100% accuracy for wavelet features.
A machine learning technique is utilized by **Salliah Shafi Bhat and Gufran Ahmed Ansari[10]** to diagnose diabetes and prescribe an appropriate diet for diabetic patients using a diet recommendation system (DRS). The machine learning model for the diagnosis of diabetes is created using a variety of machine learning techniques, including the probabilistic-based naive Bayes (NB), the function-based multilayer perception (MLP), and the decision tree-based Random Forests (RF). The classifier with the highest accuracy, Random Forests (RF), achieves 93%.
In the article **Usama Ahmed et.al [11]** A proposed machine learning model for predicting diabetes is presented using a combined approach. Support vector machine (SVM) and artificial neural network (ANN) models make up the conceptual framework. In order to assess if a diabetes diagnosis is accurate or not, these models analyse the dataset. The outcomes of these models serve as

the input membership function for the fuzzy model, which ultimately decides whether or not a diabetes diagnosis is made. The proposed fused ML model outperforms the previously disclosed techniques, with a prediction accuracy of 94.87%.

In this work **Sarra Samet[12]** use six supervised machine learning classification techniques are used together to detect diabetes early on, a hybrid model based on the top three findings was developed. The Pima Indians Diabetes Database is accessible through UCI's machine learning repository and is used in the studies. They are all assessed using a range of metrics. It is underlined that the hybrid model outperforms other cutting-edge techniques with an accuracy of 90.62%.

Minhaz Uddin Emon[13] attempted to predict diabetic retinopathy by using feature extraction to identify some features. The UCI machine learning repository provided the information needed for this investigation. This dataset is investigated using a variety of Machine Learning (ML) approaches in order to evaluate the performance, sensitivity, selectivity, true positive (tp), false negative (fn), and receiver operating characteristic (roc) curves. Naive Bayes, Sequential Minimal Optimisation (SMO), Logistic Regression, Stochastic Gradient Descent (SGD), Bagging Classifier, J48 Classifier, Decision Tree Classifier, and Random Forest Classifier are a few of the machine learning techniques used in this study. Logistic regression is the overall model that performs the best.

In contrast to Support Vector Machines, **S.Jyotheeswar and K.V Kanimozhi[14]** offered a study that used novel decision trees (DT) and SVM to identify diabetic retinopathy (DR). The New Decision Tree (N=10) and Support Vector Machine (N=10) algorithms are used to predict diabetic retinopathy. For detecting diabetic retinopathy, more than 50,000 digitised retinal images from the Kaggle fundus image dataset are employed. Innovative Decision Tree achieved precision of 92.8% while Support Vector Machine only achieved accuracy of 85.2%. The statistically significant difference between DT and SVM is (p=0.03). When equaled Support Vector Machine, the Innovative Decision Tree technique performs superior at detecting diabetic retinopathy.

In this work, the **M. Paliwal and P.Saraswat [15]** uses regulated machine-learning methods to the real data of potential diabetes patients, aged sixteen to ninety, as well as 520 diabetic individuals. These methods include the Naive Bayes classifier, Light-GBM, and Support-Vector Machine (SVM). In a comparison of classification and recognition accuracy, the support vector machine's performance has the highest accuracy.

III. METHODOLOGY

In this Process there are two main things which play a very important role in diabetics' prediction. They are as follows (1) Dataset Collection (2) Classification Techniques.

A. *Dataset Collection*

1. Kaggle Dataset: This dataset, which was released by the National Institute of Diabetes and Digestive and Kidney Diseases, can be used to determine a patient's risk for diabetes based on a number of diagnostic measures. The dataset includes 133 disease-related parameters and 4920 rows of symptom descriptions. Datafiles are consistent. The illness train and disease test datafiles are included in the data folders. Each file has its own set of data. We have 42 rows with regard to diseases and 133 attributes with regard to symptoms in the disease test file and 4920 rows with regard to the disease train[16].
2. PIMA Indian Diabetic Dataset: The PIMA Indian Diabetic Dataset contains records for 768 female patients, of whom 268 have diabetes and 500 have not. 768 data points and 9 qualities make up the variable into which the data is stacked[17].
   - How many times the patient has been pregnant?
   - When glucose is given to the patient, the plasma glucose concentration gauges how quickly it is absorbed from the blood.
   - The pressure exerted between two heartbeats is known as diastolic blood pressure (mm Hg).
   - Triceps Skin Thickness for Body Fat Calculation in Millimetres.
   - 2-Hours Serum Insulin: This test is used to determine when a type-2 diabetes patient should begin taking insulin to supplement oral medication (mi U/ml).
   - By dividing weight by the square of body height, the BMI is computed.
   - Diabetic Family Function: Patients are more likely to develop diabetes if their family has a history of the disease and if they have a genetic condition with those family members.
   - Patient's age, expressed in years.
   - Goal variable (zero when diabetes, one when not)
3. A dataset called Division was created to aid in the CAD of diabetic retinopathy. The Leiriade Andrade ophthalmologic department in Brazil collected 700 photographs of the eye fundus for this dataset using retinograph technology. A binary and multiclass evaluation of the diavision dataset is planned. The medical professionals have offered two diagnoses in binary for each image and four diagnoses in multiclass. The first method's goal is to identify the disease through an eye

exam, while the second method's goal is to determine the degree of DR. A smartphone (iPhone 6s, camera resolution 4608x2592 pixels, F2.2, 1.8GHz) coupled with a portable optical equipment was used to elaborate on another dataset (see Figure 1). Twenty pictures were taken with a handheld equipment[18]. For the purpose of investigations on computer-assisted diabetic retinopathy diagnoses, the Messidor is a public dataset [19] that has been created. It has 1200 photos of the eye's fundus.

4. UCI Library: In this work, a dataset for machine learning called the Diabetic Retinopathy Debrecen dataset from the University of California, Irvine (UCI) repository was used. The dataset consists of 1151 iterations with 19 features each and a binary result function that assesses if the impression demonstrates diabetic retinopathy symptoms or not. In order to assess whether or not an eye image has symptoms of diabetic retinopathy, a group of researchers from the University of Debrecen in Hungary developed a dataset with properties extracted from the test photos[13].

B. *Classification Techniques*

The classification process entails predicting the category of a given set of input data points. Classes also go by the terms targets, labels, and categories. To roughly translate input variables (X) to discrete output variables is the goal of classification predictive modelling problem (y). In machine learning, supervised and unsupervised learning are two essential subcategories[20].

1. Supervised Learning: - In supervised learning, an algorithm, an output variable (Y), and an input variable (x) are used to search for the function that maps the input to the output.

$$Y = f(X)$$

The goal is to closely approximate the mapping function so that, given a brand-new input file (x), you can accurately forecast the output variables of the data (Y). It is characterised as supervised learning because the process of an algorithm learning from a training dataset is sometimes compared to a teacher monitoring the training process[20].

2. Unsupervised learning: - Unsupervised learning is the practice of learning solely from an input file (X) with no connected output variables. Unsupervised learning aims to model the underlying structure or distribution of the data in order to learn more about it. Since there are no teachers and no correct responses, in contrast to the supervised learning previously mentioned, these are referred to as unsupervised learning activities. The intriguing structure inside the data is left up to algorithms to find and display[20].

Some of the supervised and unsupervised classification algorithms are as following: -

a. Novel Decision Tree: New Decision Tree (DT) is a widely used and one of the most efficient algorithms. It is a tree that bases its decisions on the information stored in memory. Let's train some data that might be utilized in that procedure to check the input, which is an image of the retinal fundus. Divide and conquer is the foundation upon which the special decision tree is created. Data is initially divided into subsets, then even smaller subsets, and so on until it appears impossible to divide it any more (Bibi, Mir, and Raja 2020)[14].

b. KNN: - A data item is categorised using the majority class among its K nearest neighbours in the supervised learning model K Nearest Neighbours [21]. The proximity of two points is measured by the distance between them. The number of chosen neighbours is set as K's initial value. The separation between each data point and the test data point is determined, and the data points are arranged in increasing order of their separation. The first K entries in the sorted list are taken into account. The test data point is then given a class based on the most often occurring class of these entries. KNN is different from the K-means clustering technique because K-means groups together comparable data points using an unsupervised approach.

c. Support Vector Machine: - An algorithm for supervised machine learning is the support vector machine. It is a machine learning technique used mostly for classification and regression analysis that assesses data and recognises patterns or decision boundaries within the dataset. An N-dimensional hyperplane that correctly divides the sample into numerous categories in hyperspace is what a support vector machine (SVM) classifier aims to create. The dataset's feature vector, or total number of dimensions, is what is utilised to separate different class borders into hyper-planes in a multidimensional space. SVM can handle a variety of continuous and categorical variables. The SVM's objective is to classify the two groups according to their traits. Three lines make up the model. The first is the marginal line, which is represented by the equation w.x-b=0. The closest data points for both classes are shown by the lines w.x-b=1 and w.x-b=-1. The support vectors are shown as circles on the hyperplane. In the other class, the filled circle is peculiar. In order to avoid over-fitting and obtain a classification that is almost flawless, it is disregarded. To reduce the likelihood of generalisation error, the SVM aims to maximise the perpendicular distance between the two edges of the hyper-plane. Because of their dependence on the hyper-plane, the generalisation capacity rises as the number of support vectors decreases[22].

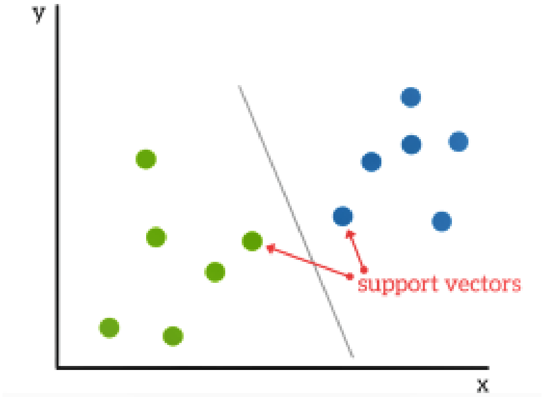

Figure 2: Support Vector Machine Graph[20]

d. Random Forest: To improve the accuracy of their predictions, ensemble models mix several different models. Random Forest is an ensemble model that generates numerous trees rather than just one [23]. The ratio of votes from each decision tree is then used to forecast the final class of the data point. As numerous decision trees that are subject to noise are combined in Random Forest, the overall impact of noise is reduced, improving the outcomes. Instead of producing only one tree, a Random Forest generates multiple trees. Each tree in the classification forest is fed with the same input in order to classify a fresh sample from a given input. Votes for each classification are referred to as categories. The classification that receives the most votes is chosen. A less mistake-prone tree becomes a good classifier, but a tree with higher similarity in the middle of the trees raises the error rate of the forest. Reduced correlation and strength result from fewer qualities, while increased correlation and strength result from more attributes[24].

e. Light Gradient Boosting Machine: Microsoft developed the LightGBM GBDT algorithm, which is open source. The parallel voting DT technique accelerates training, takes less memory, and includes improved network connection. It uses a histogram-based algorithm to maximise parallel learning. Additionally, LightGBM splits the leaf with the highest variance gain as it grows trees using the leaf-wise method. LightGBM can be distinguished from other GBDT types. LightGBM splits (small and big gradients, gi) accommodate for both weak and strong learners. In this case, the absolute values of the gradients of the training cases are ranked in descending order[25].

f. Adaboost: Adaboost is a type of iterative algorithm; its central principle is to train several weak classifiers on a training set before combining them in various ways to create a stronger classifier. The approach is implemented by altering the data distribution, which determines the weight of each sample based on the classify correction of a sample from the training set and the most recent accuracy of the overall classification. The lower classifier is then trained using the weights from the newly changed data, and eventually each training classifier is combined to create the final decision classifier[26]. An ensemble model called Adaptive Boost employs the Boosting approach. This model aids in the training of weak classifiers to create strong classifiers[25]. During training, it pays incorrectly classified samples additional attention. Then, the data is initialised with weight set to 1/N, where N stands for the overall number of occurrences. Then, as shown in, the real influence for each classifier is calculated[26].

g. Neive Bayes: The simplest classifier algorithm is Naive Bayes, which is also the fastest and most scalable when compared to more complex techniques [16]. A popular classification technique based on Bayes theory is naive Bayes. The test data will be assigned to the class with the highest posterior class probability after calculating the posterior class probability of each test data point using class conditional density estimation and class prior probability. The primary drawback of the naive Bayes approach is the estimate of the class conditional density. Data points have previously been utilised to calculate the class conditional density. In order to learn the class conditional density, we need employ uncertain data objects represented by probability distributions for uncertain classification tasks [27].

IV. RESULTS AND DISCUSSION

In order to give a complete picture of the machine learning algorithm employed and the accuracy attained by the suggested system, Table 1 giving the facts acquired from the literature review in organized form. In Table 1 a best classifier which helps to predict the diabetics can be found accurately.

Table1: Used Classification algorithms with it are obtaining accuracy.

S_No.	Reference No.	Study by	Technique Used	Obtain Accuracy
1.	[7]	M. Malini et.al	SVM, KNN, Logistic, Random Forest, Decision Tree and Gradient Boosting	78%
2	[8]	S. M. A. Huda et.al	Decision Tree, Logistic Regression and Support Vector Machine	88%,97% and 92%
3	[9]	S. S. Prem and A. C. Umesh	Support vector machine (SVM), k-nearest neighbour (KNN), decision tree, random forest (RF) and artificial neural network (ANN).	100%
4.	[10]	S. S. Bhat and G. A. Ansari	Naive Bayes, Decision Tree, and Random Forest.	93%
5.	[11]	Usama Ahmed et.al	Fused Machine Learning Decision	94.87%
6.	[12]	Sarra Samet	Ensemble Model	90.6%
7.	[13]	Minhaz Uddin Emon	Naive Bayes, SMO, SGD, Logistic, Bagging, J48, Decision Tree, and Random Forest.	75%
8.	[14]	S. Jyotheeswar and K.V Kanimozhi	Support-Vector Machine (SVM) and Novel Decision Tree	92.89%
9.	[15]	M. Paliwal and P. Saraswat	Naive Bayes classifier, Light-GBM, and Support-Vector Machine (SVM)	97.08%
10.	[18]	S. S. A. Alves et al.	Bayesian Classifier, Multilayer Perceptron (MLP), K-Nearest Neighbour (kNN), Random Forest (RF) and Support Vector Machine (SVM)	100%
11.	[23]	S. Ghane et.al	KNN, SVM, Decision Tree, Random Forest, LGBM, and Adaboost	89.85%
12.	[7]	M. Malini et.al	SVM, KNN, Logistic, Random Forest, Decision Tree and Gradient Boosting	78%
13.	[8]	S. M. A. Huda et.al	Decision Tree, Logistic Regression and Support Vector Machine	88%,97% and 92%

V. CONCLUSION AND FUTURE SCOPE

In the discipline of machine learning, disease prediction is the most difficult research problem. Early disease diagnosis is advantageous to the medical communities. The use of machine learning in the field of healthcare has been studied in this essay. The major goal of this article is to gather data on how to use the diabetic's data set with the machine learning method. Information about the machine learning algorithm used for diabetics and the degree of accuracy that was attained with regard to the disease in question has been gathered. We also provided a tabular depiction of the researcher's contribution, including the reference, author name, algorithm utilised, and accuracy level attained. With the above discussion about the classification methods of Machine Learning Techniques in Diabetic prediction, it can be concluded that by applying a combination of Bayesian Classifier, Multilayer Perceptron (MLP), K-Nearest Neighbor (kNN), Random Forest (RF) and Support Vector Machine (SVM) highest accuracy can be obtained. In future work, we can apply this technique in a gadget that end consumers might use to periodically and for no charge examine their health reports. On their own smartphones, anybody can check their report.

REFERENCES

[1] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, 'Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension', *IEEE Access*, vol. 7, pp. 144777–144789, 2019, doi: 10.1109/ACCESS.2019.2945129.

[2] M. Banchhor and P. Singh, 'Comparative study of ensemble learning algorithms on early stage diabetes risk prediction', *2021 2nd Int. Conf. Emerg. Technol. INCET 2021*, 2021, doi: 10.1109/INCET51464.2021.9456263.

[3] G. G. Warsi, S. Saini, and K. Khatri, 'Ensemble Learning on Diabetes Data Set and Early Diabetes Prediction', *2019 Int. Conf. Comput. Power Commun. Technol. GUCON 2019*, pp. 182–187, 2019.

[4] T. M. Ahmed, 'Developing a predicted model for diabetes type 2 treatment plans by using data mining', *J. Theor. Appl. Inf. Technol.*, vol. 90, no. 2, pp. 181–187, 2016.

[5] B. P. Lohani and M. Thirunavukkarasan, 'A Review: Application of Machine Learning Algorithm in Medical Diagnosis', *Proc. Int. Conf. Technol. Adv. Innov. ICTAI 2021*, pp. 378–381, 2021, doi: 10.1109/ICTAI53825.2021.9673250.

[6] '17-Significance of machine learning in healthcare_ Features, pillars and applications _ Elsevier Enhanced Reader.pdf'. .

[7] M. Malini, B. Gopalakrishnan, K. Dhivya, and S. Naveena, 'Diabetic Patient Prediction using Machine Learning Algorithm', *Proc. - 1st Int. Conf. Smart Technol. Commun. Robot. STCR 2021*, 2021, doi: 10.1109/STCR51658.2021.9588925.

[8] S. M. A. Huda, I. J. Ila, S. Sarder, M. Shamsujjoha, and M. N. Y. Ali, 'An Improved Approach for Detection of Diabetic Retinopathy Using Feature Importance and Machine Learning Algorithms', *2019 7th Int. Conf. Smart Comput. Commun. ICSCC 2019*, 2019, doi: 10.1109/ICSCC.2019.8843676.

[9] S. S. Prem and A. C. Umesh, 'Classification of Exudates for Diabetic Retinopathy Prediction using Machine Learning', *2020 IEEE 5th Int. Conf. Comput. Commun. Autom. ICCCA 2020*, pp. 357–362, 2020, doi: 10.1109/ICCCA49541.2020.9250858.

[10] S. S. Bhat and G. A. Ansari, 'Predictions of diabetes and diet recommendation system for diabetic patients using machine learning techniques', *2021 2nd Int. Conf. Emerg. Technol. INCET 2021*, pp. 1–5, 2021, doi: 10.1109/INCET51464.2021.9456365.

[11] U. Ahmed *et al.*, 'Prediction of Diabetes Empowered With Fused Machine Learning', *IEEE Access*, vol. 10, pp. 8529–8538, 2022, doi: 10.1109/ACCESS.2022.3142097.

[12] S. Samet, M. R. Laouar, and I. Bendib, 'Diabetes mellitus early stage risk prediction using machine learning algorithms', *5th Int. Conf. Netw. Adv. Syst. ICNAS 2021*, 2021, doi: 10.1109/ICNAS53565.2021.9628955.

[13] M. U. Emon, R. Zannat, T. Khatun, M. Rahman, M. S. Keya, and Ohidujjaman, 'Performance Analysis of Diabetic Retinopathy Prediction using Machine Learning Models', *Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021*, pp. 1048–1052, 2021, doi: 10.1109/ICICT50816.2021.9358612.

[14] S. Jyotheeswar and K. V. Kanimozhi, 'Prediction of Diabetic Retinopathy using Novel Decision Tree Method in Comparison with Support Vector Machine Model to Improve Accuracy', *Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2022 - Proc.*, pp. 44–47, 2022, doi: 10.1109/ICSCDS53736.2022.9760842.

[15] M. Paliwal and P. Saraswat, 'Research on Diabetes Prediction Method Based on Machine Learning', *Proc. Int. Conf. Technol. Adv. Comput. Sci. ICTACS 2022*, pp. 415–419, 2022, doi: 10.1109/ICTACS56270.2022.9988050.

[16] M. Karthik, S. G. Anuradha, K. S. Raghu Kumar, U. M. Srusti, and S. Sanjeev Kumar, 'Disease Prediction: A Case Study for Healthcare Communities Using Machine Learning', *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Sect. Int. Conf.*, pp. 1–6, 2022, doi: 10.1109/MysuruCon55714.2022.9972697.

[17] C. S. Manikandababu, S. Indhu Lekha, J. Jeniefer, and T. A. Theodora, 'Prediction of Diabetes using Machine Learning', *Int. Conf. Edge Comput. Appl. ICECAA 2022 - Proc.*, no. Icecaa, pp. 1121–1127, 2022, doi: 10.1109/ICECAA55415.2022.9936375.

[18] S. S. A. Alves *et al.*, 'A New strategy for the detection of diabetic retinopathy using a smartphone app and machine learning methods embedded on cloud computer', *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2020-July, pp. 542–545, 2020, doi: 10.1109/CBMS49503.2020.00108.

[19] T. International and C. Diabetic, 'JAMA Ophthalmology Volume 131 issue 3 2013 [doi 10.1001_jamaophthalmol.2013.1743] AbrÃ moff, Michael D.; Folk, James C.; Han, Dennis.pdf', 2013.

[20] A. Pathak and A. Dhole, 'Image classification Method in detecting Lungs Cancer using CT images A Review', *Int. J. Comput. Sci. Eng.*, vol. 9, no. 5, pp. 37–42, 2021, doi: 10.26438/ijcse/v9i5.3742.

[21] K. Taunk, S. De, S. Verma, and A. Swetapadma, 'A brief review of nearest neighbor algorithm for learning and classification', *2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019*, no. Iciccs, pp. 1255–1260, 2019, doi: 10.1109/ICCS45141.2019.9065747.

[22] S. Ghosh, A. Dasgupta, and A. Swetapadma, 'A study on support vector machine based linear and non-linear pattern

classification', *Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2019*, no. Iciss, pp. 24–28, 2019, doi: 10.1109/ISS1.2019.8908018.

[23] S. Ghane, N. Bhorade, N. Chitre, B. Poyekar, R. Mote, and P. Topale, 'Diabetes Prediction using Feature Extraction and Machine Learning Models', *Proc. 2nd Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2021*, pp. 1652–1657, 2021, doi: 10.1109/ICESC51422.2021.9532818.

[24] S. V. Patel and V. N. Jokhakar, 'A random forest based machine learning approach for mild steel defect diagnosis', *2016 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2016*, 2017, doi: 10.1109/ICCIC.2016.7919549.

[25] M. R. Machado, S. Karray, and I. T. De Sousa, 'LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry', *14th Int. Conf. Comput. Sci. Educ. ICCSE 2019*, no. Iccse, pp. 1111–1116, 2019, doi: 10.1109/ICCSE.2019.8845529.

[26] X. Shu and P. Wang, 'An Improved Adaboost Algorithm Based on Uncertain Functions', *Proc. - 2015 Int. Conf. Ind. Informatics - Comput. Technol. Intell. Technol. Ind. Inf. Integr. ICIICII 2015*, pp. 136–139, 2016, doi: 10.1109/ICIICII.2015.117.

[27] J. Ren, S. D. Lee, X. Chen, B. Kao, R. Cheng, and D. Cheung, 'Naive bayes classification of uncertain data', *Proc. - IEEE Int. Conf. Data Mining, ICDM*, no. 60703110, pp. 944–949, 2009, doi: 10.1109/ICDM.2009.90.

BIOGRAPHY

	Shweta Yadu is currently pursuing her MTech degree in Computer Science and Engineering from Raipur Institute Of Technology affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai, Chhattisgarh, India. She has completed BE in Computer Science & Engineering from ITGGU, Bilaspur Chhattisgarh, India in 2010. Her research interest fields are Data Science, Artificial intelligence, and Machine Learning.
	Rajeshri Lanjewar is currently working as an Assistant Professor in the Computer Science and Engineering Department at Raipur Institute of Technology, affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai, Chhattisgarh, India. He is having 10.9 years of experience in teaching. She has published 3 research papers in SCI, Scopus and other reputed international journals and 4 paper presented in National/International conferences.
	Vivek Kumar Sinha is currently working as an HOD in the Computer Science and Engineering Department at Raipur Institute of Technology, affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai, Chhattisgarh, India. He is having 14 years of experience in teaching. He is currently Research Scholar at Lovely Professional University Phagwara, Jalandhar, Punjab, India. He has published 21 research papers in SCI, Scopus