Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

DIABETES MONITORING USING ENSEMBLING METHODS AND DIABETIC RETINOPATHY STAGE DETECTION USING RESNET50 MODEL

Sai Madhumitha R J , Vaishnavi N K , Revathi P
Department of Computer Science and Engineering
Thiagarajar College of Engineering
Madurai, India

Abstract - **Diabetes is one of the most dangerous,chronic and threatening diseases in the world. Further,it acts as a root cause for many more disorders such as damage of blood vessels which in turn leads to heart attack,stroke,blindness, lower limb amputation,etc. One of the serious issues faced is Diabetic Retinopathy which leads to blindness. A person having diabetes for more than 20 years is more likely to be affected with diabetic retinopathy. Therefore, continuous monitoring of diabetic levels is crucial. For the sake of consultation, a patient has to visit the hospital regularly to check for his/her diabetes level which won't be a great choice in this pandemic situation. By taking advantage of Machine Learning, we can flexibility monitor diabetes levels and also detect the stage of diabetic retinopathy. This paper discusses diabetes monitoring in healthcare,with four machine learning algorithms and have ensembled them in different combinations and a deep learning model - RestNet 50 for detecting the stage of diabetic retinopathy. The input dataset is modified Pima Indian diabetes dataset and APTOS 2019 Blindness Detection which has been cleansed and preprocessed for implementation of the machine learning models. The machine learning classifiers used here are k-nearest neighbors(knn), Support Vector Machine(svm), Decision Tree(dt) and Random Forest(rf). Weighted Average ensembling of different combinations of algorithms is also performed. The best performing ensembling classifier is with accuracy, precision,sensitivity, specificity as 0.862, 0.795, 0.783 and 0.931 respectively. The ResNet50 model produces a kappa score of 0.745.**

Keywords **-** Supervised learning, Ensembling classifier, Machine learning models, Diabetes Monitoring, modified Pima Indian Diabetes dataset, k-nearest neighbors, Support Vector Machine, Decision Tree, Random Forest, Weighted Average ensembling, Confusion Matrix, Performance metrics, Diabetic Retinopathy, RestNet 50, Deep Learning.

I. INTRODUCTION

Diabetes is a disorder that is prevalent all over the world irrespective of ages. Diabetes occurs when the pancreas is no longer able to produce insulin in sufficient quantity or required amount. Insulin,a hormone secreted by pancreas lowers blood glucose level to keep it in normal range. It's actual role is to help the body cells to absorb the glucose broken down from food.

Common symptoms of diabetes are extreme hunger, frequent urination, increased thirst, unexplained weight loss, blurred vision, slow-healing sores, presence of ketones in urine, frequent infections such as gums or skin infections and also vaginal infections.

According to the American Diabetes Association, 2,83,000 youths who are under age 20 are estimated to have diabetes, which is approximately 35% of that population and also in 2014–2015,about 18,200 people with type 1 diabetes, 5,800 people with type 2 diabetes are affected.

A Government survey in India recorded 11.8% of the total population are affected by diabetes. The observation shows that the highest contribution is in between the age 70 to 79.

The major contributors in India to increase diabetes are less exercise, more screen time, usage of tobacco and alcohol, environment pollution, high blood pressure and cholesterol level, low consumption of nuts and whole grains.

Further, one of the effects of diabetes is Diabetic Retinopathy. Blood vessels in the retinas are damaged when blood sugar level rises.

Some of the symptoms of diabetic retinopathy are floaters and fluctuating vision, poor color vision. The effects are detachment of retina, glaucoma, vitreous hemorrhage. Diabetic Retinopathy contributes around 12% of blindness caused in the United States each year.

In our proposed method different machine learning classifiers such as k-nearest neighbors(knn), Support Vector Machine(svm), Decision Tree(dt), Random Forest(rf) and Weighted Average ensembling of different combinations of the four machine learning algorithms are used. ResNet 50, a deep learning model is used for Diabetic Retinopathy stage detection.

II. TYPES OF DIABETES

1. Type 1 Diabetes - It is caused by an autoimmune reaction which means the body attacks itself by mistake. It is commonly found in children and adolescents. In this case the pancreas does not or produces very little amount of insulin. Here, the patients are recommended to inject insulin daily to maintain their glucose level. The normal blood glucose level for Type 1 diabetes is 80-130 mg/dL. A value above 180 mg/dL is considered to be diabetic in male whereas in females it is above 200 mg/dL.

2. Type 2 Diabetes - 90% of overall cases are Type 2 diabetic which is commonly found in adults. In this case the insulin produced by the pancreas is not efficiently used. Here the insulin remains unused building up sugar in the bloodstream, instead of moving into the cells for energy. In most cases it is treated by pills and in severe cases insulin is injected. The normal blood glucose level for Type 2 diabetes is 70-130 mg/dL. A value above 180 mg/dL is considered to be diabetic.

3. Gestational Diabetes - Every year five million women are affected by this type of diabetes. It is widely seen in pregnant women and generally disappears after their pregnancy. It also has an impact on the fetus, and may have a risk of developing Type 2 diabetes in future. Blood glucose level less than 140 mg/dL is normal, more than 190 mg/dL is considered diabetic.

III. STAGES OF DIABETIC RETINOPATHY

1. Mild Nonproliferative Retinopathy - Here tiny bulges, leakage of fluid occur in the retina. Since it is the early stage of DR, proper treatment cures the disease.

2. Moderate Nonproliferative Retinopathy - Swelling of blood vessels occurs inside the retina. Retina is physically affected due to improper blood flow.

3. Severe Nonproliferative Retinopathy - At this stage blockage of blood vessels tends to happen. It triggers the blood vessels to develop new blood vessels. Blindness is the ultimate result of this stage.

4. Proliferative Retinopathy **-** Thin and weak blood vessels start growing in the retina. Stage 4 results in total loss of vision.

IV. LITERATURE REVIEW

Priyanka Sonar et al.[1] focuses on the PIMA Indians Diabetes (PID). The author uses Support vector machine, Artificial Neural Network, Decision Tree, Naive Bayes techniques to predict diabetes. A Comparison is done between various machine learning algorithms. A Machine learning Matrix is used to compare the models.

Muhammad Azeem Sarwar et al. [2] focuses on the PIMA Indians Diabetes(PID). The author uses K- Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), Random Forest (RF) and Logistic Regression (LR). The author compares various models and finds that SVM and KNN give the highest accuracy. The author computes a machine learning matrix where accuracy is given by : Accuracy=(TP+TN) / (P + N).

MD.Kamrul Hasan et al. [3] focuses on the PIMA Indians Diabetes(PID). The author uses k-nearest Neighbor, Random Forest, Decision Tree, Naïve Bayes, Ada Boost, XGBoost (XB) and Multilayer Perceptron. The author tries to ensemble different models for higher accuracy. As a result the author finds two boosting type classifiers (AB and XB) is the best combination for diabetes prediction.

Amani Yahyaoui et al. [4] focuses on the PIMA Indians Diabetes (PID). The author uses Decision Support Systems, Convolutional Neural Network, machine learning, Support Vector Machine, deep learning, Random Forest. The results showed that RF was more effective for classification of diabetes in all rounds of experiments.

Ms. K Sowjanya et al. [5] focuses on training data such as Age, Gender, Polydipsia, Polyphagia, Family background, Diet, Physical activity, High BP, Smoking habit, Weight loss, Height, Blurry vision and Weight (BMI), Waist circumference. J48, Naïve Bayes, SVM and Multilayer Perceptron. The results indicated that j48 was the best for the prediction

Legila Alic et al. [6] focuses on the SAHS dataset with a total of 1496 participants. It used the linear support vector machines to construct a prediction model of future development of type-2 diabetes. The outcomes of the study show that high values of glucose observed at the 2h mark during the OGTT may strongly indicate the potential risk of future development of type-2 diabetes.

Attila Mustapha, Mohemad Lachgar [7] focuses on ResNet50 and VGG-16 pretrained networks for classification of Diabetic retinopathy. They compared the models ResNet50 and VGG-16 and arrived at an accuracy of 75% and 25%.

Ratul Ghosh, Kuntal Ghosh, Sanjit Maitra [8] focuses on developing a CNN model for automating the detection and classification of stages of Diabetic Retinopathy. The model gained 85% accuracy for the classification of 5 classes(4- stages, no DR) and 95% accuracy for binary classification of DR or no DR.

V. DATASET

This paper uses the modified Pima Indian Diabetes dataset which is originally from the National Institute of Diabetes. The donor of the database is Vincent Sigillito from The Johns Hopkins University. The dataset consists of 769 patient records with 7 attributes. Table1 describes Data Statistics.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

	Age	BMI	Insulin	BloodPressure	Pregnancies	Glucose
count	614.000000	614.000000	614.000000	614.000000	614.000000	614.000000
mean	33.594463	32.208678	82.136808	69.226384	3.874593	120.895765
std	12.016168	7.914276	117.481581	19.550893	3.443637	31.809725
min	21.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	24.000000	27.500000	0.000000	62.000000	1.000000	99.000000
50%	29.000000	32.400000	37.000000	72.000000	3.000000	117.500000
75%	41.000000	36.775000	130.000000	80.000000	6.000000	141.000000
max	81.000000	67.100000	846.000000	122.000000	17.000000	198.000000

Table 1.Data Statistics

A. Attributes

The following are the attributes considered and table2 gives its description:

SN	Attribute	Description	Mean
1	Pregnancies	Number of times pregnant	3.87
2	Glucose	Plasma glucose concentration at 2 hours in an Oral Glucose Tolerance Test	120.89
3	Blood Pressure	Diastolic Blood Pressure	69.22
4	Insulin	2-hour serum Insulin	82.13
5	BMI	Body Mass Index(weight in kg/(height in inches)^2)	32.20
6	Age	Age in years	33.59
7	Outcome	Class variable(0 or 1)	-

Table 2.Attributes

This paper also uses APTOS (Asia Pacific Tele-Ophthalmology Society) 2019 Blindness Detection Dataset for Diabetic Retinopathy Detection. This dataset contains

- train.csv
- test.csv
- train_images
- test_images

The train_images and test_images folders contain scanned images of retina for training and testing respectively. The train file contains two attributes: id_code, diagnosis. The id_code is the code for the images in the train_images folder. The diagnosis attribute has values ranging from 0-4 indicating the stage of diabetic retinopathy as mentioned in table 3. The test file contains only the id_code attribute which is the code for the images in the test_images folder.

Diagnosis	Stage
0	No DR
1	Mild
2	Moderate
3	Severe
4	Proliferate DR

Table 3.Stages of DR

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

VI. DATA VISUALIZATION

Data visualization is an important part of machine learning. It helps to understand the input data in a better way. It gives an insight about the dataset and attributes used.

The figures 1,2,3 are some visualization charts for input data of Pima dataset such as Histogram, Line plot, Box plot respectively.

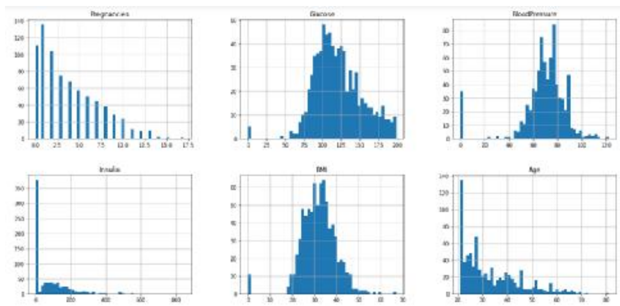

Figure 1.Histogram

Here the density plots are used to visualize how each attribute is distributed.

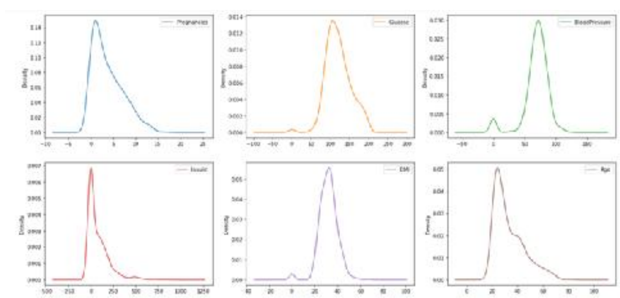

Figure 2.Line plot

```
Pregnancies        AxesSubplot(0.125,0.657941;0.227941x0.222059)
Glucose         AxesSubplot(0.398529,0.657941;0.227941x0.222059)
BloodPressure   AxesSubplot(0.672059,0.657941;0.227941x0.222059)
Insulin            AxesSubplot(0.125,0.391471;0.227941x0.222059)
BMI             AxesSubplot(0.398529,0.391471;0.227941x0.222059)
Age             AxesSubplot(0.672059,0.391471;0.227941x0.222059)
Outcome               AxesSubplot(0.125,0.125;0.227941x0.222059)
dtype: object
```

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

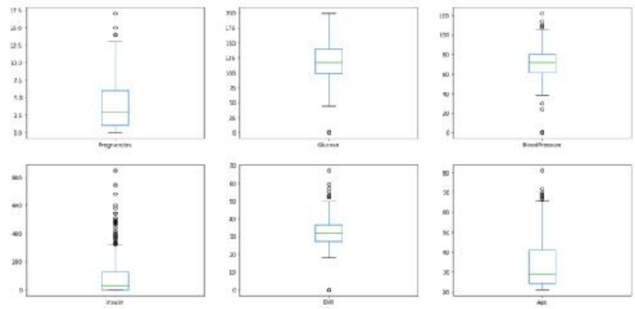

Figure 3.Box and Whisker plot

Figure 4 is the bar chart that displays the count of people affected in each stage.

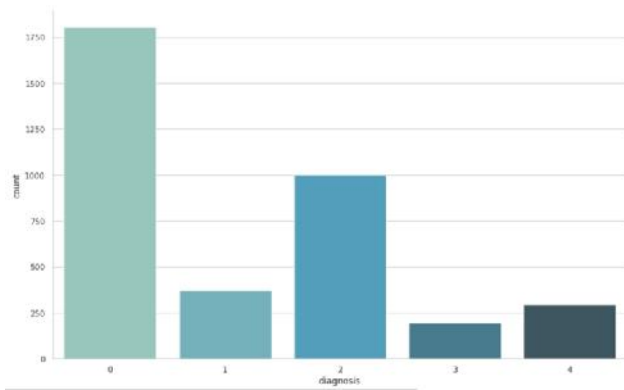

Figure 4.Barchart

VII. METHODOLOGY

Data mining is the process of finding patterns within a large dataset for predicting outcomes and solving business problems through data analysis. It helps us to arrive at more-informed decisions.Data mining uses machine learning as its technique.

Machine learning is a type of artificial intelligence that allows machines to make accurate predictions without explicitly being programmed.

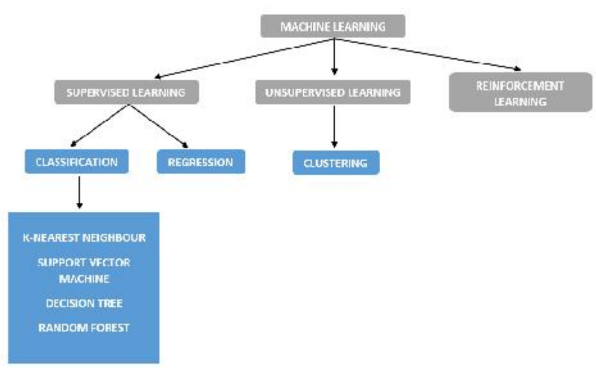

Figure 5. Subareas of ML

Supervised learning trains the machine with a labeled dataset. Classification algorithm groups objects into predetermined categories and accurately predicts target class for each new case.This paper uses classification algorithms since categorical data is used.

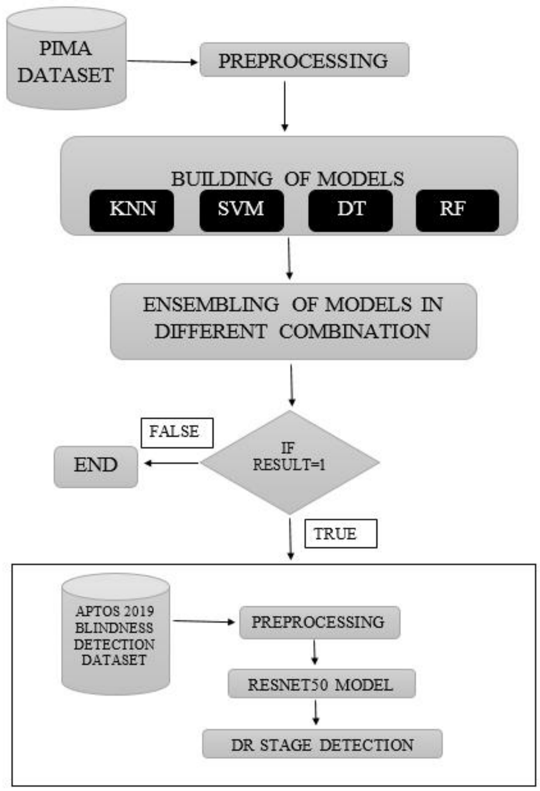

Figure 6.Workflow

This paper focuses on four classification algorithms namely k-nearest neighbors(knn), Support Vector Machine(svm), Decision Tree(dt), Random Forest(rf). Further by doing Weighted Average Ensembling these algorithms in different combinations helps to gain better results. The performance metrics considered for evaluating these algorithms are accuracy, confusion matrix, precision, sensitivity, specificity.

Figure 6 explains the workflow of the proposed method.

Neural network is a branch of machine learning and the heart of deep learning. These neural networks are similar to neurons in the brain.

Training is the important process to improve the accuracy. Initially the ResNet50 model is built and further the model is fine tuned.

VIII. PREPROCESSING

The initial step before implementing a machine learning model is preprocessing the dataset. The general steps include importing the required libraries and reading the dataset. The following are the various preprocessing steps that have been done.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

A. MISSING VALUES

Missing values refer to empty cells in a dataset, also refer to Not Available(NA). Pandas library is generally used for handling missing data. Here isnull() function is used to check for missing values and it is found that this dataset does not contain any missing values.

B. DATA IMPUTING

Data imputation fills estimated values in a particular cell when there is inconsistent data. This process substitutes the null value in attributes like Age, BMI, insulin, Glucose, Blood Pressure with the mean of the respective fields.

C. OUTLIER REJECTION

Outlier refers to irrelevant data present in the dataset. It increases the error variance. Here Robust Scaler() function is used for rejecting the outliers.

D. DATA STANDARDIZATION

Data standardization aids in making the data internally consistent and also reduces skewness of the data. The technique applied is Standard Scaler which follows standard normal distribution. The Standard Scaler() function is used to standardize data.

E. DATA PARTITIONING

The dataset used for training is modified version of PIMA dataset ie. Features such as age, BMI, Insulin, Blood Pressure, Pregnancies and Glucose alone are extracted from standard PIMA dataset along with the outcome (Presence/Absence of diabetes).

The test_train_split() function of sklearn library is used for splitting the dataset into subsets. The whole dataset is split into two parts: 80% of the dataset is used to train the model and 20% of data is set aside for testing the model.

F IMAGE RESIZING

The preprocessing in Diabetic Retinopathy is to make all images into the same size. Both train_images and test_images are resized into the same size.

IX. DIABETES MONITORING - MACHINE LEARNING MODELS

A. K-NEAREST NEIGHBORS

K-NN algorithm works by working on similarity between available data and putting the new case into the category that is most similar. This algorithm requires the least training phase when compared to other algorithms and we can seamlessly add new data.

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

PARAMETER TUNING

To increase the accuracy of the model, fixing the parameters is crucial. GridSearchCV is the sklearn library function that helps in selecting the best hyperparameters.

The best parameter values obtained are:

- n_neighbors=16
- metric=minkowski
- weights=distance
- leaf_size=2
- p=1

The KNeighborsClassifier() function is used to implement the knn model.

Table 4 is the confusion matrix for k-nn model:

	ACTUAL	
PREDICTED	97	10
	13	34

Table 4.Confusion matrix - KNN

The performance metrics are listed below
in Table 5:

PERFOMANCE METRICS	VALUES
Accuracy	0.85
Precision	0.77
Sensitivity	0.72
Specificity	0.90

Table 5.Results of KNN

B.SUPPORT VECTOR MACHINE

The SVM algorithm finds a hyperplane in n-dimensional spaces that classifies the data points distinctly. It works well in high dimensional attributes. It divides the training points into subsets therefore memory is used efficiently.

The Support Vector Classifier() function is used to implement the SVM model. The SVC classifier function holds two parameters: random_state is defined as zero and kernel type is set as its default rbf.

Random state controls the generations of random numbers for shuffling and ensures that the split will always be safe.

Table 6 is the confusion matrix for SVM model:

	ACTUAL	
PREDICTED	99	8
	21	26

Table 6.Confusion matrix - SVM

The performance metrics is listed below in table 7:

PERFOMANCE METRICS	VALUES
Accuracy	0.81
Precision	0.76
Sensitivity	0.55
Specificity	0.92

Table 7.Results of SVM

C. DECISION TREE

Decision tree identifies a common pattern from the training data and utilizes it for further conclusions. The two main nodes are decision node and leaf node. There are multiple branches for decision nodes and they are used for making decisions. Leaf nodes are branchless and they are the output of the decision nodes.

The DecisionTreeClassifier() function is used to implement the Decision tree model. Here Attribute Selection Measures (ASM) is used for finding the best attribute for the root and sub nodes. Under ASM, gini index is used as the criterion.

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

Table 8 is the confusion matrix for decision tree model:

	ACTUAL	
PREDICTED	91	16
	19	28

Table 8.Confusion matrix - DT

The performance metrics is listed below in table 9:

PERFOMANCE METRICS	VALUES
Accuracy	0.77
Precision	0.63
Sensitivity	0.59
Specificity	0.85

Table 9.Results of DT

D. RANDOM FOREST

Random forest is a flexible and simple algorithm that arrives at better results without hyperparameter tuning. In this algorithm, multiple decision trees are built and merged to gain accurate and fixed prediction.

The RandomForestClassifier() is used to implement the Random Forest model. For every sample an individual decision tree is constructed and will get the prediction results from them. Finally,voting is performed and the most voted result is selected as the final prediction.

Table 10 is the confusion matrix for random forest model:

ACTUAL

PREDICTED		
	98	9
	17	30

Table 10.Confusion matrix - RF

The performance metrics is listed below in table 11:

PERFOMANCE METRICS	VALUES
Accuracy	0.83
Precision	0.76
Sensitivity	0.63
Specificity	0.91

Table 11.Results of RF

E.ENSEMBLING

Ensembling is a technique used to combine more than one model to improve prediction. Ensembling helps us to average several models to arrive at better results. For classification algorithms, voting based ensemble techniques are used.

Weighted Average ensembling follows distributed weighting according to the priority of each algorithm. It differs from majority voting since all models have varied importance. The ensemble members are combined differently to fetch distinct accuracy. The combination that has gained the highest accuracy is considered as the best ensemble member.

X. DIABETIC RETINOPATHY - NEURAL NETWORK

Convolution Neural Network is a supervised deep learning that works on images. Combining several CNN layers results in the AlexNet model, the first convolution network which uses GPU to boost the performance. But the disadvantage here is that increasing the number of layers leads to the problem of vanishing gradient. As a result training and test error rate also increases.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

A. RESNET50 MODEL

In 2015 an alternative to AlexNet was introduced called Residual Network shortly ResNet by Microsoft Research. The Vanishing Gradient problem was solved using skip connection technique i.e. skipping a few training layers that affects the performance of the model and directly connects to the output. Keras API and tensorflow libraries are used to create ResNet50 architecture.

The model parameters such as batch_size, epochs, width, height, etc are defined. A base model for ResNet50 is built. One of the functions used to construct the network is the Dense() function. It helps in passing output of the previous layer as an input to the next layer. The model is then compiled using the optimizer() function. Repeated fine tuning has enhanced the model.

XI. RESULTS AND DISCUSSION

Comparison of the models is based the criterion such as Accuracy, Precision, Sensitivity, Specificity.The results of implementation of the four models is as follows:

Confusion matrix helps to evaluate the performance of the model. Table 12 gives the description of the elements in the confusion matrix.

	ACTUAL	
PREDICTED	Tp	Fp
	Fn	Tn

Table 12.Confusion matrix

A.ACCURACY

Accuracy simply means the number of data points correctly predicted out of all.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$$

Figure 7 helps us to compare the accuracy of the four models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

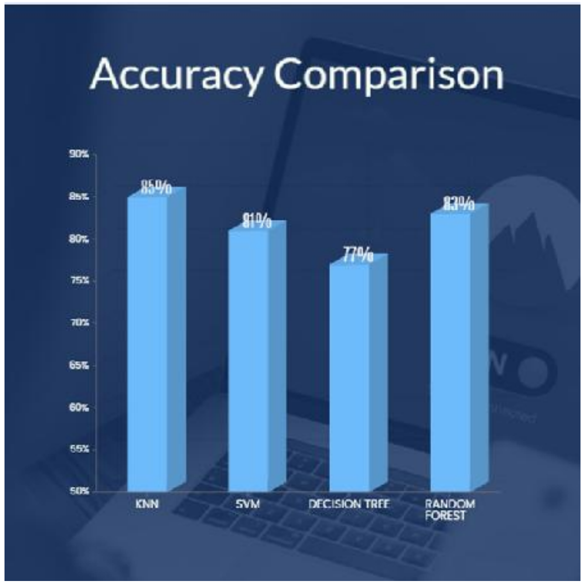

Figure 7.Accuracy Comparison

In terms of accuracy, KNN and Random Forest have gained almost similar accuracy, differing just by 2%. Among four algorithms, Decision Tree stands last with 77%.

B. PRECISION

Precision helps to measure the quality of the prediction, especially the positive prediction that the model makes.

$$Precision = \frac{TP}{TP+FP}$$

Figure 8 helps us to compare the precision of the four models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

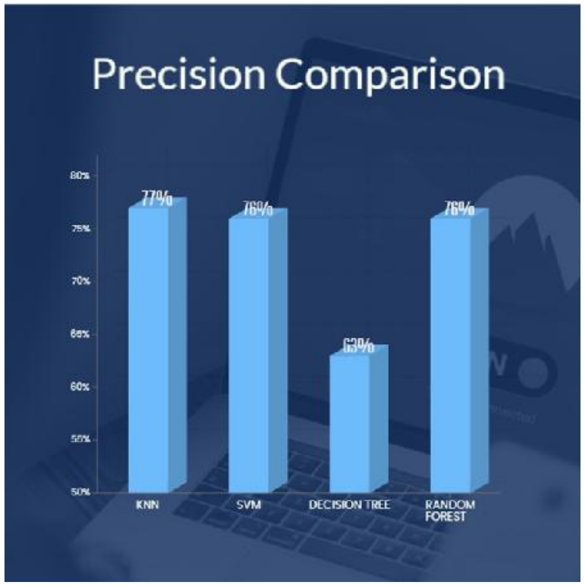

Figure 8.Precision Comparison

SVM and Random Forest have gained similar precision of 76% and for KNN the percentage is 77. Decision Tree stands last with 63%.

C. SENSITIVITY

Sensitivity helps in measuring the proportion of true positive cases out of the positive cases predicted by the model.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}$$

Figure 9 helps us to compare the sensitivity of the four models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

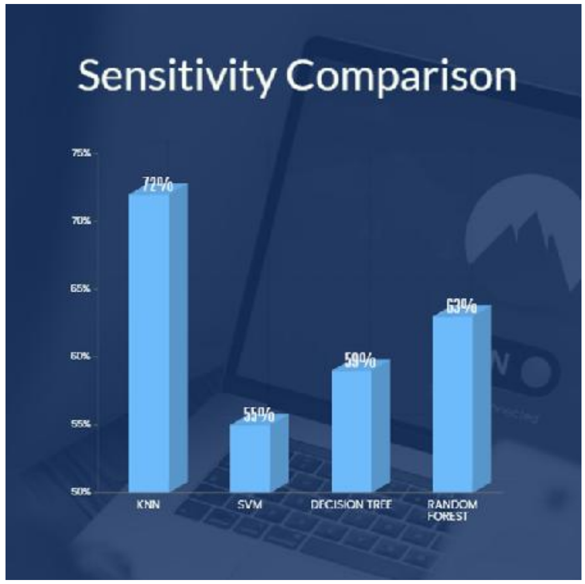

Figure 9.Sensivity Comparison

In terms of Sensitivity, KNN is better than all the other algorithms with 72%. The algorithm with least sensitivity is SVM.

D. SPECIFICITY

Specificity helps in measuring the proportion of true negative cases out of the negative cases predicted by the model.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

All the four algorithms have specificity greater than 80%. Among which Random Forest, SVM, KNN have gained 93%, 92%, 90% respectively.

Figure 10 helps us to compare the specificity of the four models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

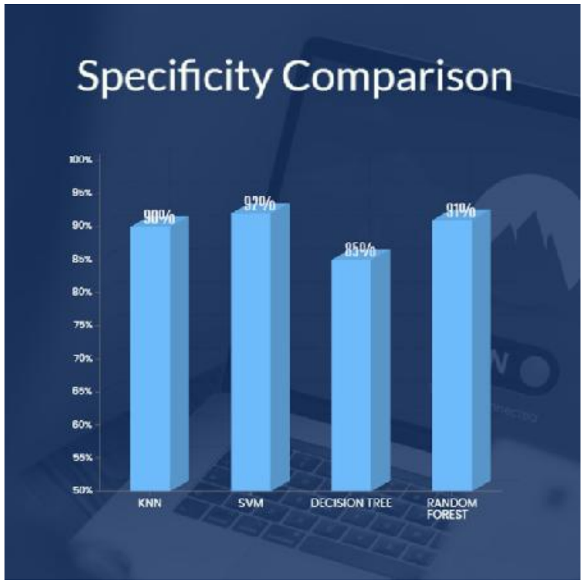

Figure 10.Specificity Comparison

Table 13 summarizes the values of Accuray, Precision, Sensitivity and Specificity of the four models.

Algorithm	Accuracy	Precision	Sensitivity	Specificity
KNN	0.85	0.77	0.72	0.9
SVM	0.81	0.76	0.55	0.92
Decision Tree	0.77	0.63	0.59	0.85
Random Forest	0.83	0.76	0.63	0.91

Table 13.Summarization of four machine learning models

Different combinations of algorithms are ensembled and Table 14 summarizes the result.

Algorithm Combination	Accuracy	Precision	Sensitivity	Specificity
KNN, DT, RF	0.862	0.795	0.783	0.931
KNN, SVM, RF	0.809	0.772	0.751	0.905
SVM, DT, RF	0.795	0.769	0.659	0.879
KNN, DT	0.824	0.786	0.674	0.885
KNN, RF	0.836	0.765	0.698	0.914

Table 14.Summarization of ensembled models

The above table shows that the combination of K-Nearest Neighbor, Decision Tree, Random Forest yields the highest values for all the performance metrics.

Figure 11 is the classification report for the ResNet50 model.

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

	precision	recall	f1-score	support
0	0.93	0.97	0.95	1805
1	0.48	0.25	0.33	370
2	0.58	0.80	0.67	999
3	0.62	0.15	0.24	193
4	0.28	0.15	0.19	295
accuracy			0.74	3662
macro avg	0.58	0.46	0.47	3662
weighted avg	0.72	0.74	0.71	3662

Figure 11.Classification Report

The model's accuracy is 74%. Figure 12 is the model accuracy graph for the ResNet50 model.

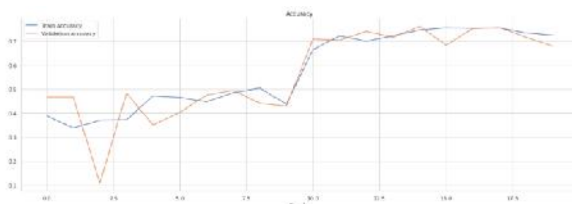

Figure 12.Model Accuracy Graph

XII. CONCLUSION

Diabetes monitoring is one of the ways that helps in reducing complications in the future. Developing machine learning models aids medical professionals to make better decisions than the regular diagnosis made.

Various machine learning models were implemented and their results are tabulated in this paper. The study shows that the ensembling of K-Nearest Neighbor, Decision Tree, Random Forest provides the maximum values for accuracy, precision, sensitivity and specificity with the values of 0.862, 0.795, 0.783 and 0.931 respectively.

The Quadratic Weighted Kappa score of the ResNet50 model is 0.745.

XIII. REFERENCE

[1] Priyanka Sonar and Prof. K. JayaMalini, "Diabetes Prediction using Different Machine Learning Approaches", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)

[2] Muhammad Azeem Sarwar, Nasir Kamal, Wajeeha Hamid and Munam Ali Shah, Prediction of Diabetes Using Machine Learning Algorithms in Healthcare, Proceedings of the 24th International Conference on Automation & Computing, Newcastle University,Newcastle upon Tyne, UK, 6-7 September 2018.

[3] MD. Kamrul Hasani, MD. Ashraful Alami, Dola Das, Eklas Hossain and Mahmudul Hasan, Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers, Volume 8, IEEEAccess.

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 1, Paper 10,January 2023

[4] Amani Yahyaoui, Akhtar Jamil, Jawad Rasheed and Mirsat Yesiltepe,
A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques, International Informatics and Software Engineering Conference(UBMYK).

[5] Ms. K Sowjanya, Dr. Ayush Singhal and Ms. Chaitali Choudhary, MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices, 2015 IEEE International Advancement.

[6] Lejla Alic, Hasan T. Abbas, Marelyn Rios, Muhammad AbdulGhani, and Khalid Qaraqe, Predicting Diabetes in Healthy Population through Machine Learning, 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).

[7]P. Suresh Kumar and V. Umatejaswi, Diagnosing Diabetes using Data Mining Techniques, International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017.

[8]Aatila Mustapha, Lachgar Mohamed, Hrimech Hamid and Kartit Ali, Diabetic Retinopathy Classification Using ResNet50 and VGG-16 Pretrained Networks, International Journal of Computer Science and Data Science, Volume 1, Issue 1, June 2017.

[9]Ratul Ghosh, Kuntal Ghosh and Sanjit Maitra, Automatic Detection and Classification of Diabetic Retinopathy stages using CNN, 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)

[10] Revathy R, Nithya B S, Reshma J J, Ragendhu and Sumithra MD, "Diabetic Retinopathy Detection using Machine Learning", International Journal of Engineering Research & Technology (IJERT), Volume 9, Issue 06, June 2020.

[11]Malik Bader Alazzam, Fawaz Alassery and Ahmed Almulihi, "Identification of Diabetic Retinopathy through Machine Learning", Hindawi Mobile Information Systems, Volume 2021, Article ID 1155116.

[12]Zhiguang Wang and Jianbo Yang, "Diabetic Retinopathy Detection via Deep Convolutional Networks for Discriminative Localization and Visual Explanation", GE Global Research,San Ramon, CA.