Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 20s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Label Prediction For Diabetic Retinopathy Using Weighted Average, Ensemble Representation, Or Mlp

P.Anitha[1], Dr. P.R. Tamilselvi[2]

[1]Research Scholar, Periyar University, Salem.

[2]Assistant Professor, Govt. Arts and Science College, Komarapalayam

Cite this paper as: P.Anitha, Dr. P.R. Tamilselvi, (2025) Label Prediction For Diabetic Retinopathy Using Weighted Average, Ensemble Representation, Or Mlp. *Journal of Neonatal Surgery*, 14 (20s), 130-136.

ABSTRACT

Diabetes type prediction is a difficult issue that is receiving more and more attention. Retinal pictures from the Indian Diabetic Retinopathy Image (IDRiD) dataset have been used to demonstrate the weighted ensemble framework for label prediction utilising different machine learning models that can be suggested to improve the prediction of diabetes. This paper proposes a weighted ensemble classifier for diabetic classification and prediction. Several Machine Learning (ML) classifiers, such as k-nearest Neighbour, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost, in addition to Multilayer Perceptron (MLP), were used to improve the prediction of diabetes. The matching Area Under ROC Curve (AUC) of the machine learning model is used to compute the weights. K-fold cross-validation, feature selection, data standardisation, filling in missing values, and outlier rejection are all important considerations for developing diabetes prediction. The grid search technique is used to improve the performance metric, AUC, during hyperparameter tweaking. The grid search technique is used to improve the performance metric, AUC, during hyperparameter tweaking. Because the ML model's AUC is independent of the class distribution, it is used as the model's weight for voting ensembling rather than accuracy. According to experimental study, the suggested system performs better than earlier state-of-the-art methods in terms of recall, precision, and f measure.

Keywords: *Diabetic, dataset, pre-processing, Feature Extraction, Ensemble Classification*

1. INTRODUCTION

Every day, the healthcare industry is currently producing vast volumes of data about its patients. The organised and unstructured formats of this amount of data make it difficult to maintain. With this type of data storage, big data can now be categorised, producing better outcomes. The effects of diabetes are depicted in Model [1]. The GRN's streamlined structure depends on precise classification in order to achieve a high outcome. They have used the multilayer perceptron (MLP) and radial basis function (RBF) for the categorisation and comparison of disorders in a specific manner in order to improve the brain structure. Back propagation training methods have been used to test them [2]. Consequently, this method increases the accuracy of its diabetes prediction.

a diabetic disorder that harms the retina as a result of high blood sugar. If treatment is not received, this may result in blindness. Tobacco use, high blood pressure, high cholesterol, poor blood sugar regulation, long-term diabetes, pregnancy, and race are risk factors.

Because of this, diabetes is regarded as one of the main causes of death worldwide. Our researchers have used the R programming technique in combination with the decision tree algorithm to develop and predict diabetics in an early manner [3]. Modern doctors can predict their patients' ailments early on with the use of this data prediction. Many novel early disease prediction algorithms were used in this study, and data mining methods help doctors make highly precise and effective predictions [4].The cause of type 2 diabetes is improper and disorganised insulin administration. They claim that the main factor contributing to the increase in retinal impairment in diabetics is type 2 diabetes. The k-clustering technique is used for this in the classification and data prediction procedures. For their classification, they employed the Indian Diabetic Retinopathy Image (IDRiD) dataset.

Excellent accuracy and less computation time should be the outcomes of the proposed classification model. They obtained classification accuracy of 86%, sensitivity of 83%, and specificity of 87%. Additionally, the area under the Receiver Operating Characteristic Curve (ROC) is 87%. [5]. They have used diabetics to predict the problem in an effort to improve results for diabetics with retinopathy. They have used the support vector regression model, a newly created method that helps forecast the value of diabetes at different periods, including before bed and at night.

Diabetes can result from blood glucose levels that are either high or too low. Thus, this method helps predict both high and low blood glucose levels in the patient. In order to compare their level to that of normal diabetics and particular glucose levels, they have calculated the accuracy of diabetics based on the four-week value. [6].a system [7] that predicts renal failure using modern artificial intelligence methods. Using AI in conjunction with machine learning algorithms and other state-of-the-art technologies, they have correctly and extremely accurately anticipated the data. Neural networks have been used for the enhanced time series. They have also used logistic regression and neural networks to improve time series and results.

It is believed to be a difficult task to examine and extract new data from a new database. This process, referred to as data mining, employs a variety of techniques, such as weighted ensemble models using KNN, Random Forest, and multilayer perceptrons, to improve the accuracy of diabetes prediction. The rest of the paper is organised as follows: related work is presented in section 2. The Weighted Ensemble classifier, a possible paradigm for diabetes classification, is described in Section 3. The experimental design and results are described in section 4. The conclusion is given in Section 5.

2. RELATED WORKS

The many models that are currently being utilised to classify diabetes are outlined and explained in this section.

A hybrid classifier for the classification of diabetes

This approach employs Cuckoo search for feature selection and normalisation for preprocessing [8]. KNN and J48 are hybrid classifiers that exploit the recovered features [9][10].The accuracy rates of these two classifier techniques are 95.5% and 96.3%, respectively. Additionally, the classification that was utilised to compare the two approaches is referred to as a hybrid model. However, this technique does not offer high performance or precision.

B. The suggested model

This section describes a Weighted Ensemble Classifier that uses machine learning techniques to efficiently classify images of diabetic retinopathy. The publicly accessible IDRiD dataset was used to train and evaluate the ML models in the following ways:

The stage of pre-processingThis suggested diabetes classification architecture incorporates a preprocessing step that comprises outlier rejection, missing value filling, standardisation, and feature selection prior to the disease being categorised and its prognosis being predicted

Ignoring irregularities

It is thought that this comment is unusual. It is computed using the feature vector, and feature values that exceed the limit are removed. Including the value that is missing

The attribute or feature mean values were used to impute the missing or null values of the specific occurrence on the feature.

Normalisation

To lessen the skewness of the data distribution, the attributes must be rescaled to a conventional normal distribution with zero mean and unit variance.Selecting attributes Using Grid Search for feature extraction produces the most distinctive features. Grid Search creates a training data set that can be as widely spaced as possible while condensing the same classes of patterns as closely as possible. By lowering the number of dimensions, grid search can identify these patterns in the data with minimal information loss. Lastly, it is composed of the differentiating characteristics' feature vectors.

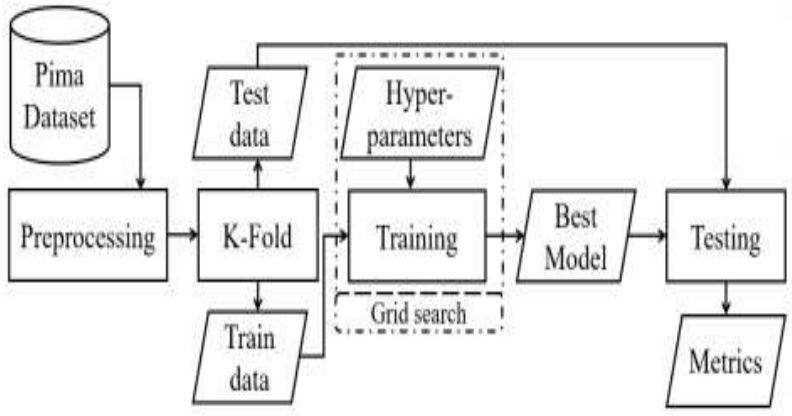

Figure 1: Process flow of the Proposed Ensemble Classifier

.

A. Ensemble Weighted Among the machine learning models that have been trained and evaluated are k-NN, DT, AB, RF, NB, and XB. One well-known method for improving performance using a collection of classifiers is to assemble the ML model. This technique, known as a weighted soft voting ensemble, combines the output of several models to increase prediction accuracy.
Machine learning algorithms like naive-bayes, decision trees, and random forest R K-NN have all been taught and evaluated using this model. Each model's inner loop will be used to modify the hyperparameter. The accuracy of forecasts can be increased by an ensemble by combining the output from several models. Table I lists the hyperprmeter changes made to each model.

Table I: Hyper Parameter of each Model

Model	**Hyper Parameter**
K-NN	Number of neighbors for queries Computing algorithm for nearest neighbors • Ball Tree (BT): Node defines a D-dimensional hypersphere or ball • KD Tree (KDT): Leaf node is a D-dimensional point • Brute: Based on the brute-force search Leaf size for BT or KDT which depends on the nature of problem Metric (Manhattan distance (L1-norm) or Euclidean distance (L2-norm))
Decision Tree	Measuring function: Gini impurity or Entropy The strategy used to choose the split at each node The minimum samples for an internal node The minimum samples for a leaf node
Naive Bayes	Portion of the largest variance of the attributes

In a decision analysis graph known as the Decision Tree DT, the outcome is displayed as a splitting rule for every unique property. The outcomes of decisions can be shown graphically and directly with this branching graph. Every property serves as a branching node, generating a rule at the branch's terminus to distinguish values from distinct classes. It looks like a tree, as its name suggests, and it concludes with a selection called the tree's leaf. The root is the most promising feature for forecasting the evolution of rules [11].

A DT is not only simple to use and intuitive, but it also produces more accurate future forecasts. New node creation is repeated until a fundamental need is not met. During the DT analysis, the leaf node is reached when the class label property is set to the highest value allowed by the rule. Decision trees can become overfit even if they are constructed with their roots at the top.

Because overfitting happens when a tree becomes too proficient with data and its leaves show low impurity, pre-pruning is the act of deleting leaves that are not significant or required for tree growth. According to pre-pruning, the base criterion must be higher than the tree's depth in order to generate a DT model. Forecast accuracy is also increased by prepruning. Since information gain aids in more accurate outcome prediction than any other criterion, it is a prerequisite for DT split.

This method determines the entropy of each feature, and the property with the lowest entropy is selected for the split. For decision trees, which are basically classification models, parameter optimisation aims to identify the set of constraints that will best optimise the model architecture.

Among the factors that must be altered in the decision tree are maximum depth, criterion, confidence, minimal gain, minimal leaf size, and minimal size for a split. In a decision tree, criteria are the first variable that can be changed. This option optimises the split value for each criterion in respect to the selected criteria and regulates the criteria used to assess the impurity of a split. Examples of split criteria include the Gini index, information gain, and gain ratio. The best splitting criteria for a decision tree's entropy and Gini index are found using the formula below. The implemented decision tree uses knowledge gain for the split criteria because of the previously described benefits.

Gini : $Gini(E) = 1-\sum_{j=1}^{c} 1P_j^2$

Entrophy : $H(E) = -\sum_{j=1}^{c} 1P_j \log p_j$

Maximum Depth is yet another fantastic decision tree function. Depending on the amount and properties of the dataset, the tree's depth changes. The more splits a tree has, the more information it will gather about the data. Therefore, a tree size of 1 to 20 has been selected, depending on the dataset. Another crucial DT parameter that affects the degree of pessimism utilised in the pruning process computation is confidence. It is estimated that the aforementioned decision tree has a 0.1 confidence level.

• The credulous Bayes

The representation of the outcomes is the primary distinction between this decision tree-based approach and others. Naïve Bayes probability, in which the rules are provided by the DT at the conclusion. Both algorithms aim at prediction [12]. Before training, ocular imaging data from diabetes patients is analysed using a Naïve Bayes model [17]. NB provides a conditional probability as well. The NB's primary advantage is its ability to handle small datasets and its superior performance over the low variance classifier, which use the Bayes theorem to ascertain whether an attribute associated with an item is feasible given the pertinent data.

$$F_i^{NB}(x) = P(x_{j=xj} | c=i)\ p(c=i)$$

Along with this, it is easy to implement and computationally low-priced. In NB all the attribute values are independent of each other, therefore, it is inexpensive in computation and separately simplifies the assumption and calculation using above mentioned equation. In Naive Bayes classifier parameter tuning and optimization is limited

- **K-NN**

K-NN is an additional machine learning method for classification that works better than Ball Tree, KD tree, and Brute hyperparameter tweaking. The method is considered instance based since it estimates the class label for the feature sample by computing the k nearest points [13]. The distance weighting estimate predicts the k nearest points in order to train feature samples as classes and forecast the class label for the sample. Let Tij be a distance measure for each class, and let X be the ideal feature vector that the nearest neighbour has to compute to represent the class. The distance is estimated using the Euclidean distance.

Distance between the samples is T_{ij} which can be represented as

$$D_i(y) = \|y - T_{ij}\|^2$$

Where T_{ij} is nearest Neighbour to Optimal feature vector.

Determine the category of the features such that category which represents most of the features. In this conclude T_{ij} belongs that particular category. In this category is collection of feature or its values to represent individual. Euclidean distance between two features in fused vector is $E_d = \sqrt{(\Delta x)2 + (\Delta y)2}$

- **Multilayer Perceptron (MLP)**

To categorize retinal images as normal or pathological, the MLPNN classifier is employed [14].

Neurones are the processing units that make up a neural network. Each neurone is connected to the others via unidirectional connections of varying weights. An input-output layer and many hidden layers make up a feedforward neural network [15]. Any MLP layer generates an N-dimensional output vector from its D-dimensional input vector. To reduce the mistake, back-propagation is used to adjust the neuron's parameters. The grid search will employ the following hyperparameters to maximise the AUC: batch size, learning rate, epoch, activation function, neurone initialisation, number of hidden layers, number of neurones in each hidden layer, loss function, and optimiser.

Feature vector has been processed with hyper parameter values on selected word vector and epoch. Further cross entropy loss function has been utilized to manage cluster seperability. Model parameter is updated to generate the cluster with minimum inter cluster distance. Updating of model parameter is given by

$$L(q_{i,j,\,Mj}) = \left(\sum_{k=0}^{n} x^k v^{n-k} \frac{nx}{1!} + \binom{n}{k} x^k v^{n-k} \frac{n(n-1)x^2}{2!}\right) - \left(\sum_{k=0}^{n} (x - \mu)\ \sum_{k=0}^{n} (x - \mu) v^k a^{n-k}\right)$$

In other words, it shows how the between-cluster distance effectively affects the within-cluster distance in the expression of cluster Membership [11].

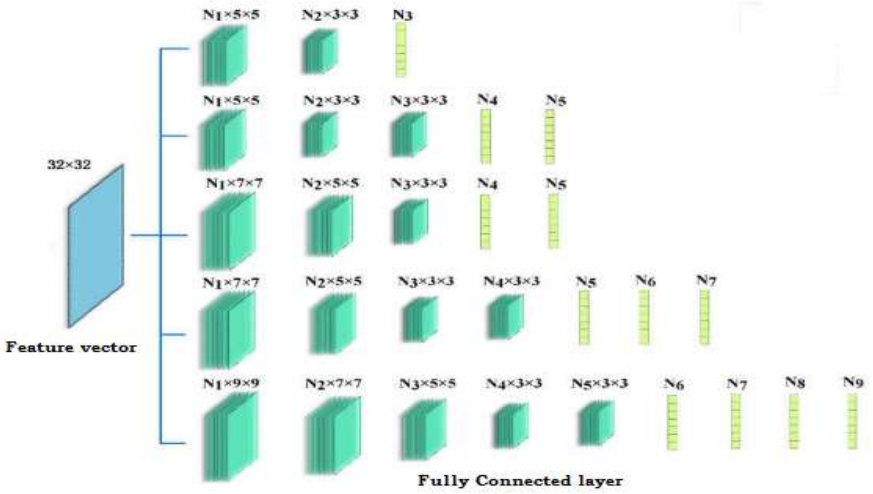

Figure 2: Processing Fully Connected Layer on Feature Vector using hyperparameter

Algorithm 1: IDRiD **Diabetics Classification using Multilayer Perceptron**

Input: Class Set C_t, Feature subset F_t

Output: Class formation

Process Initialize *Q(s,a)* arbitrarily

Repeat (for each episode):

Initialize *s*

Repeat (for each step of feature instance F_t):

Choose *a* from *s* using policy derived from *Q* for function

$$Q(s,a) = Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right]$$

***s*←*s'*;**

Until *s* is terminal

Assign the Class Label C_E to the feature instance F_t

3. EXPERIMENTAL RESULTS

The architecture has been explored using the R tool. This computing environment has different programming paradigms. It makes matrix manipulation, function charting, and data implementation easier. It also makes it easier to create interlanguage communication. A five-fold cross-fold validation performance metric was used to evaluate the model's performance. The results were produced using classification methods to maximise the accuracy of diabetes prediction [16]. With a promising accuracy of 98.07 percent, DT stands out among these classifiers as a noteworthy approach for early diabetes prediction. Classification accuracy can be expressed as the "percentage of true prediction," which is the sum of the actual positive and true negative divided by the total expected class value.

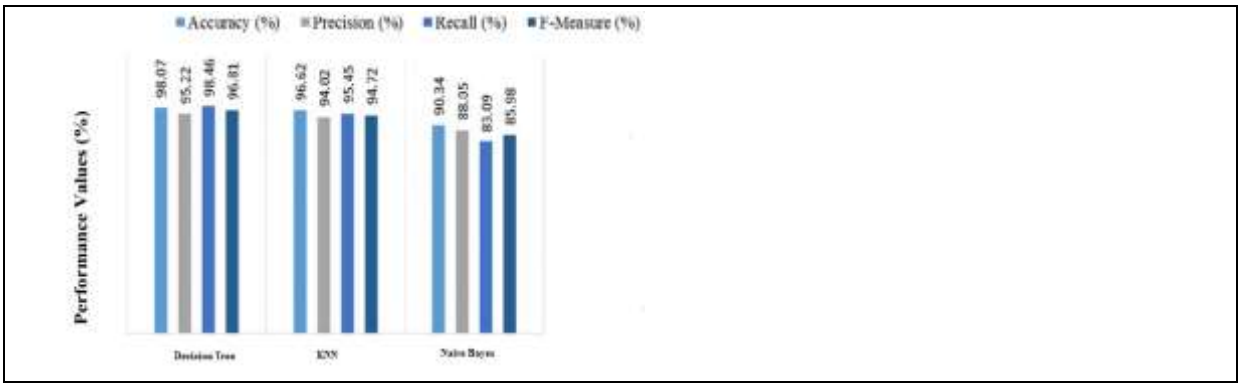

Figure 3: Results Results of the suggested model

Table II [10] shows the performance of the suggested model, including the classification and associated accuracy. Depending on the dataset, the suggested model can significantly increase classification efficiency. To determine whether the model is sufficient to handle the problem, accuracy alone is insufficient. As a result, more measures are needed to assess the classifier's performance. Class recall, class precision, and F-measure are these extra measures. The number of correctly classified qualities is known as class recall.

- Accuracy = (true Positive /true negative)*100
- Recall= TruePositives/(TruePositives + FalseNegative)
- Precision= TruePositives/(TruePositives + FalseNegative)
- F−Score = 2*((Precision * Recall)/(precision + Recall))

Table II: Performance Analysis of the weighted ensemble Classifier

Metrics	Decision Tree	KNN	Naive Bayes
Precision	95.22	94.02	88.05
Recall	98.46	95.45	83.09
F measure	96.01	94.72	85.98
Accuracy	98.07	96.62	90.34

The accuracy obtained through diverse classifiers is shown below by the confusion matrix which consists of class precision, diabetes prediction yes, diabetes prediction no, class recall.

4. CONCLUSION

This approach prevents premature death in diabetics. Early disclosure of the patient's situation is beneficial. We have developed a machine learning-based diabetes prediction method to address this issue. It starts by using the IDRiD patient dataset, which is related to diabetes. This system should be utilised for machine learning procedures after pre-processing with clustering techniques. Additionally, a grid search is used to choose features. A weighted ensemble classifier is used to classify individuals with and without diabetes. Based on this classification, we have estimated that our system should be able to predict with 97% accuracy.

REFERENCES

[1] K. A. Anant, T. Ghorpade and V. Jethani, "Diabetic retinopathy detection through image mining for type 2 diabetes," 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2017, pp. 1-5, doi:10.1109/ICCCI.2017.8117738.

[2] Cichosz, S. L., Johansen, M. D., & Hejlesen, O. (2016). Toward big data analytics: a review of predictive models in the management of diabetes and its complications. *Journal of diabetes science and technology*, *10*(1), 27-34.

[3] Rallapalli, S., & Suryakanthi, T. (2016, November). Predicting the risk of diabetes in big data electronic health Records by using a scalable random forest classification algorithm. In *2016 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 281-284). IEEE.

[4] Muni Kumar, N. (2016). Survey on map reduces based apriori algorithms in the medical field for the prediction of diabetes mellitus. *RESEARCH JOURNAL OF FISHERIES AND HYDROBIOLOGY*, *11*(4), 13-18.

[5] Shetty, S. P., & Joshi, S. (2016). A tool for diabetes prediction and monitoring using data mining technique. *International Journal of Information Technology and Computer Science (IJITCS)*, *8*(11), 26-32.

[6] Mishra, S., Chaudhury, P., Mishra, B. K., & Tripathy, H. K. (2016, March). Implementation of Feature ranking using Machine learning techniques for Diabetes disease prediction. In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies* (pp. 1-3).

[7] Joshi, R., & Alehegn, M. (2017). Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach. *International Research Journal of Engineering and Technology*, *4*(10), 426-435.

[8] Komi, M., Li, J., Zhai, Y., & Zhang, X. (2017, June). Application of data mining methods in diabetes prediction. In *2017 2nd International Conference on Image, Vision, and Computing (ICIVC)* (pp. 1006-1010). IEEE.

[9] Kumar, P. S., & Pranavi, S. (2017, December). Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. In *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)* (pp. 508-513). IEEE.

[10] Jayanthi, N., Babu, B. V., & Rao, N. S. (2017). Survey on clinical prediction models for diabetes prediction. *Journal of Big Data*, *4*(1), 26.

[11] Chen, W., Chen, S., Zhang, H., & Wu, T. (2017, November). A hybrid prediction model for type 2 diabetes using K-means and decision tree. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 386-390). IEEE.

[12] Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, *132*, 1578-1585.

[13] Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018, September). Prediction of diabetes using machine learning algorithms in healthcare. In *2018 24th International Conference on Automation and Computing (ICAC)* (pp. 1-6). IEEE.

[14] Amol Prataprao Bhatkar & G.U. Kharat, " Detection of Diabetic Retinopathy in Retinal Images Using MLP Classifier", in 2015 IEEE International Symposium on Nanoelectronic and Information Systems

[15] Eswari, T., Sampath, P., & Lavanya, S. J. P. C. S. (2015). Predictive methodology for diabetic data analysis in big data. *Procedia Computer Science*, *50*, 203-208.

[16] M. Panwar, A. Acharyya, R. A. Shafik and D. Biswas, "K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus" 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), Patna, India, 2016, pp. 132-136, doi: 10.1109/ISED.2016.7977069.

[17] Mutia Fitri Anggita & Dedi Gunawan ,"Diabetic Retinopathy Detection System Using Naïve Bayes Method" , in 2024 International Conference on Smart Computing, IoT and Machine Learning (SIML), 06-07 June 2024

[18] Yookesh, T. L., et al. "Efficiency of iterative filtering method for solving Volterra fuzzy integral equations with a delay and material investigation." Materials today: Proceedings 47 (2021): 6101-6104.