



# Tree-based homogeneous ensemble model with feature selection for diabetic retinopathy prediction

Tamunopriye Ene Dagogo-George, Hamed Adeleye Mojeed<sup>\*)</sup>, Abdullateef Oluwagbemiga Balogun, Modinat Abolore Mabayoje, Shakirat Aderonke Salihu

Department of Computer Science, Faculty of Communication and Information Sciences, University of Ilorin  
PMB 1515 Ilorin, Nigeria 240003

**How to cite:** T. E. Dagogo-George, H. A. Mojeed, A. O. Balogun, M. A. Mabayoje, and S. A. Salihu, "Tree-based homogeneous ensemble model with feature selection for diabetic retinopathy prediction," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 4, pp. 297-303, 2020. doi: [10.14710/jtsiskom.2020.13669](https://doi.org/10.14710/jtsiskom.2020.13669), [Online].

**Abstract** – *Diabetic Retinopathy (DR) is a condition that emerges from prolonged diabetes, causing severe damages to the eyes. Early diagnosis of this disease is highly imperative as late diagnosis may be fatal. Existing studies employed machine learning approaches with Support Vector Machines (SVM) having the highest performance on most analyses and Decision Trees (DT) having the lowest. However, SVM has been known to suffer from parameter and kernel selection problems, which undermine its predictive capability. Hence, this study presents homogenous ensemble classification methods with DT as the base classifier to optimize predictive performance. Boosting and Bagging ensemble methods with feature selection were employed, and experiments were carried out using Python Scikit Learn libraries on DR datasets extracted from UCI Machine Learning repository. Experimental results showed that Bagged and Boosted DT were better than SVM. Specifically, Bagged DT performed best with accuracy 65.38 %, f-score 0.664, and AUC 0.731, followed by Boosted DT with accuracy 65.42 %, f-score 0.655, and AUC 0.724 when compared to SVM (accuracy 65.16 %, f-score 0.652, and AUC 0.721). These results indicate that DT's predictive performance can be optimized by employing the homogeneous ensemble methods to outperform SVM in predicting DR.*

**Keywords** – machine learning; ensemble learning; diabetic retinopathy; decision trees

## I. INTRODUCTION

Nowadays, information technology plays an important role in providing resources and services in all life domains, most especially in the healthcare sector. The application of modern technologies such as cloud computing, big data, and sensor data is fast gaining acceptance in the healthcare domain [1]. It has been observed that the emergence and the use of machine learning techniques in healthcare can be of great benefit. Machine learning can be used to analyze and make sense of expeditious data growth in the healthcare industry [2]. Various chronic diseases, such as cardiovascular diseases, cancer, childhood pneumonia,

and diabetes, affect people worldwide. These diseases have caused several deaths. However, early diagnosis and improved access to quality and affordable healthcare are imperative to curb these diseases [2], [3].

Diabetes is one of the metabolic diseases, a condition where blood glucose levels rise over a prolonged period [4]. One of the many complications that could manifest if diabetes remains untreated is Diabetic Retinopathy (DR). DR is a diabetes complication that affects the eyes by damaging the light-sensitive tissue's blood vessels at the back of the eye known as the retina [5], [6]. At first, Diabetic Retinopathy may show no symptoms or just mild vision problems.

Over time, too much sugar in the blood can lead to the tiny blood vessels' blockage that nourishes the retina, cutting off its blood supply. As a result, the eye attempts to grow new blood vessels, but these new blood vessels do not develop properly, leak blood easily, and eventually lead to blindness. Figure 1 shows the difference between a normal and DR affected retina. Generally, DR is diagnosed by carefully investigating retinal images by an experienced ophthalmologist to determine the existence of the major symptoms; microaneurysms, hemorrhages, neovascularization, and exudates [7].

Mohammadian et al. [7] reported that in 2010, 126.6 million people were diagnosed with DR, with predictions showing a growth of up to 191 million people by 2030. They also projected that the number of people with vision-threatening DR would increase from 37.3 million to 56.3 million by 2030. Therefore, it is important to provide better diagnosis techniques in diabetes healthcare solutions to cut costs and save people's lives by applying modern technologies. Machine Learning can be adapted to automatically detect these symptoms from extracted datasets and classify the patient's condition [4], [8].

Many works have been carried out in DR diagnosis and prediction using various machine learning techniques, such as SVM, Decision Trees, Random Forest, naïve Bayes, and Gaussian Process [4], [7], [9]-[12] with SVM having the highest predictive performance [9], [10], [12]. However, the selection of adequate kernel and parameter settings are still a problem for SVM [13]-[15].

<sup>\*)</sup> Correspondence author (Hamed Adeleye Mojeed)  
Email: [mojeed.ha@unilorin.edu.ng](mailto:mojeed.ha@unilorin.edu.ng)

There is a need for an improved approach that could offer better DR predictive performance [4], [7], [11]. Also, DR predictive accuracy can still be improved by exploring other prediction techniques. Ensemble methods are believed to be more effective than individual classifiers, and as it has been widely used in other domains [16]-[18] can be explored for DR prediction.

Various machine learning approaches have been applied in the literature to predict DR from image datasets, with the majority focusing on extracting features from captured images. Gurudath et al. [19] incorporated Artificial Neural Network (ANN) and SVM to identify DR from fundus images based on blood vessel segmentation with Gaussian filtering and adaptive mask generation. The efficiency of the approaches was evaluated on 106 images from the DRIVE and DIARETB1 databases. The experimental results indicated that SVM and KNN had 98.1 % and 97.2 % accuracy, respectively.

Mahendran and Dhanasekaran [9] investigated the severity level of DR by detecting the lesion exudates to diagnose DR. It employed SVM and Probabilistic Neural Network (PNN) classification algorithms on image datasets. SVM and PNN had accuracies of 97.89 % and 94.76 %, respectively. Lachure et al. [20] used SVM and KNN for detecting retinal micro-aneurysms and exudates for automatic screening of DR using morphological operations. SVM and KNN achieved better results with specificity and sensitivity of 100 % and 90 %, respectively.

Murugeswari and Sukanesh [21] presented a study to detect abnormalities in the macula's blood vessels to prevent vision loss for diabetic patients. SVM, Cascade Neural Network (CNN), and Partial Least Square (PLS) classifiers were used for the detection of the disease in the macula using both fundus images and Optical Coherence Tomography (OCT) images. The SVM, CNN, and PLS accuracy were 98.33 %, 97.16 %, and 94.34 %.

Carrera et al. [22] presented a computer-assisted diagnosis based on the digital processing of retinal images from the Messidor database consisting of 1200 eye fundus color numerical images with eight quantitative features. An initial image processing stage was implemented to isolate blood vessels, microaneurysms, and hard exudates to extract features that can be used by a learner. Accuracies recorded from classifiers yielded 92.4 % for SVM and 92.0 % for DT.

Somasundaram and Alli [23] investigated previous research for analyzing texture discrimination capacity in fungus images to distinguish the healthy images. The feature extraction process employed did not perform well due to the high dimensionality. Therefore, to properly identify retinal features for DR diagnosis and early detection, they employed the t-distributed Stochastic Neighbor Embedding (t-SNE) method for feature extraction and Bagging ensemble classifier for DR classification. Results showed that the ensemble classifier could achieve better classification accuracy than single classification models. The work, however, focused only on the bagging ensemble method.

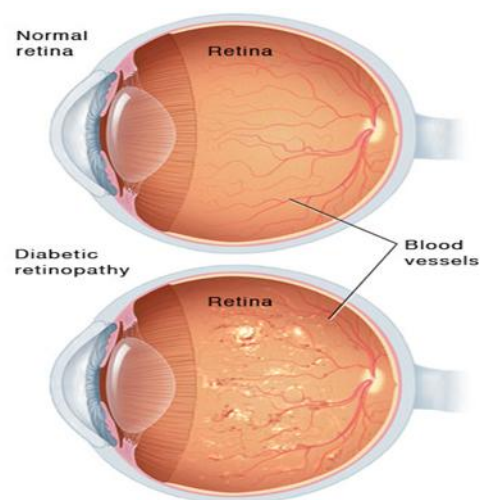


Figure 1. A normal and DR affected retina [8]

Recent studies used extracted secondary datasets available in databases to evaluate the performance of classification algorithms in predicting DR. Tsao et al. [12] presented a prediction model for DR using SVM, DT, ANN, and Logistic Regressions (LR). Experimental results demonstrated that SVM performed better than other algorithms considered with a prediction accuracy of 79.5 % and 0.839 AUC.

As various studies show different classification results, finding the most efficient DR screening classification technique is crucial. Therefore, this study proposes the use of homogenous ensemble-based machine learning techniques on DT to predict DR's presence from an image dataset. The remaining part of this paper is organized as follows: Section 2 describes the techniques used in this paper. Section 3 presents experimental results and discussions, while Section 4 concludes the paper.

## II. RESEARCH METHODS

The experimental architecture described in Figure 2 involves implementing five basic steps to achieve the study's objectives; data collection, data exploration and preprocessing, feature selection, prediction models development, and performance evaluation.

### A. Data collection

The DR dataset used in the experiment was collected from the publicly available UCI Machine Learning Repository. It contained 1151 number of instances, 19 attributes, and one numeric valued outcome variable. This dataset contains exudates, macula, and optical disc features extracted from the Messidor image set previously used to predict DR. Table 1 presents the attributes of the dataset.

### B. Data exploration and preprocessing

The dataset was analyzed using the Exploratory Data Analysis (EDA) methods in Python using Panda, Numpy, and Matplotlib packages. Panda and Numpy were used to explore data descriptive statistics and

**Table 1.** Description of DR dataset features

Variable	Feature label	Variable type	Range
X0	Binary quality assessment	Integer	0,1
X1	Result of pre-screening	Integer	0,1
X2 – X7	Results of MA detection	Real	0 – 105
X8 – X15	Results of exudates detection	Real	0 – 298
X16	The Euclidean distance between the macula center and optic disc center	Real	0 – 1
X17	The diameter of the optic disc	Real	0 – 0.2
X18	AM/FM-based detection result	Integer	0,1
X19	class label	Integer	0,1

understand the features in the data. In contrast, Matplotlib was used to visualize data by constructing plots to show correlations and differences in the data. The preprocessing stage was to ensure that all features are represented numerically. There is no data inconsistency, no occurrences of missing data, and all features are on a similar scale as features on large scales can badly influence the model.

**C. Feature extraction**

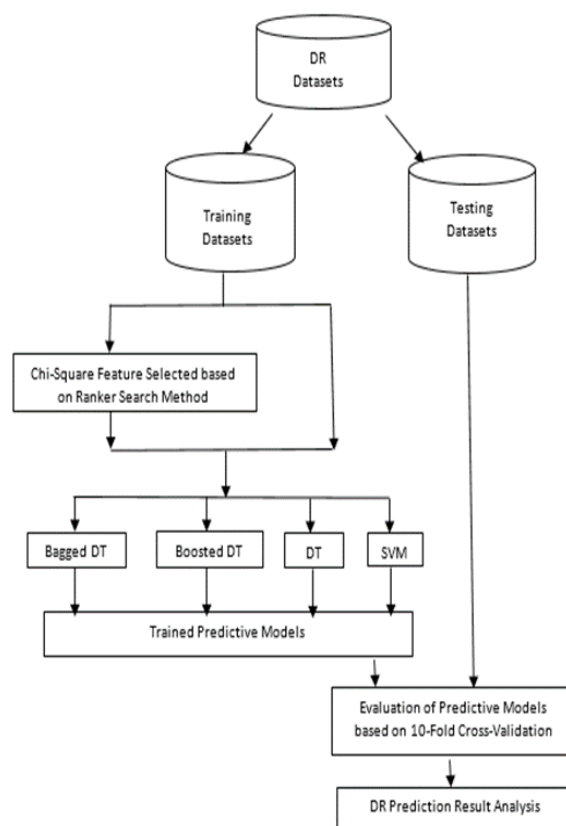
Feature selection is a dimensionality reduction technique that involves selecting relevant and irredundant features from a dataset [24]. In this study, a filter-based feature selection method, Chi-Square (CS), was incorporated to reduce the number of features used in training the model. CS is a numerical test that estimates deviation from the expected distribution of the feature set. It assumes that the feature event is independent of the class value. Metrics, such as the probability of the number of positive cases (Ppos), probability of the number of negative cases (Pneg), true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), are used to calculate the CS value.

**D. Decision Tree (DT)**

Pal et al. [10] described DT learning as a method that approximates discrete-valued target functions represented by a decision tree. Learned trees are a series of if-then statements that classify instances by sorting them down the tree from the root node to the leaf. It then assigns the classification or label for that instance. The index used to measure the degree of impurity, as defined in Eq. 1, is Entropy (H).

$$H = \sum_j -P_j \log_2 P_j \tag{1}$$

Each node in a decision tree specifies some tests on a particular attribute of the instance. Each branch of the node represents the possible values of that attribute. The DT algorithm is presented in Algorithm 1. Information Gain is used to selecting the attribute to be tested at each node. The test is then performed on the attribute that best classifies the training set and has the highest information gain. Information Gain (S, A) of an



**Figure 2.** Experimental architecture for diabetic retinopathy model

attribute A, relative to a collection of examples S, is defined in Eq. 2.

$$Gain(S, A) = H(S) - \sum_v \epsilon \text{ values}(A) \frac{|S_v|}{|S|} H(S^v) \tag{2}$$

The T model built required the use of certain parameters as different parameters offer varied performances. They include min leaf samples, number of estimators, max depth, and max features.

**E. Ensemble methods**

The ensemble method combines predictive knowledge of different base learners or different learner

---

**Algorithm 1.** Algorithm for Decision Tree

---

**Input:** Dataset and Feature set**Output:** Tree Model

Initialization: initialize an empty Root node for the tree

- 1: **if** all examples are positive, Return the single-node tree Root with label equals +
  - 2: **if** all examples are negative, Return the single-node tree Root with label equals –
  - 3: **if** the number of predicting attributes = 0,
  - 4:   **then** return the single-node tree with label = most common value of the target attribute in the examples
  - else begin**
  - 5:   A = the attribute that best classifies examples
  - 6:   Decision tree attribute for Root = A
  - 7:   **for** each positive value  $x$  of A
  - 8:     Add a new tree branch Root corresponding to the test  $A = x$
  - 9:     **Let** examples ( $x$ ) be the subset of examples that have the value  $x$  for A.
  - 10:    **if** examples ( $x$ ) is empty
  - 11:     **then** add a new leaf node below this branch with label = most common target value in the example
  - 12:     **else**, below this new branch add the subtree ID3 (examples ( $x$ ), Target attribute, Attributes – {A}).
  - end for**
  - 13:   **return** Root
  - end if**
- 

---

**Algorithm 2.** Bagging ensemble algorithm [26]

---

**Input:** A set of Classifiers ( $C_1, C_2, \dots, C_m$ )  $m < M$  $D$  a set of class labeled training tuples**Output:** A composite model

- 1: **for**  $I = k$  **do** // create  $k$  models
  - 2:   Create a bootstrap sample  $D_i$  by sampling  $D$  with replacement
  - 3:   Use  $D_i$  to derive a model  $M_i$
  - end for**
  - 4: Given a tuple  $X$ :
  - 5: **if** classification **then**
  - 6:   **let** each of the  $k$  models classify  $X$  and **return** the majority vote;
  - 7: **if** prediction **then**
  - 8:   **let** each of the  $k$  models predict a value for  $X$  and **return** the average predicted value;
- 

variations to form an improved and powerful meta-model with higher predictive performance. Piri et al. [25] stated that ensemble methods could be homogenous or heterogeneous. Homogeneous ensembles used in this work combine multiple variations of a single classifier technique. Ensembles in this category are implemented using Bagging and Boosting algorithms to develop multiple training datasets. DT uses these training datasets, and in the end, an averaging mechanism is used to make the final prediction using each classifier's outputs.

The bagging (bootstrap aggregation) ensemble method carries out a sampling procedure by merely creating various training samples from the training set by bootstrap sampling. This procedure ensures high diversity among the ensemble members and generates output that is capable of reducing bias and variability [26]. Algorithm 2 describes the ensemble bagging procedure employed in this study.

The boosting ensemble method trains many predictors in succession, and each predictor learns from its predecessor's error [27]. The base learner's training sets are chosen to adapt the learner to infer a new data pattern each time they are invoked. The Boosting method implemented in this study is Adaboost (Adaptive Boosting). The Boosting procedure is outlined in Algorithm 3.

## F. Performance metrics

Performance metrics are used in determining the performance of machine learning models. For this study, the ensemble models' performance was evaluated based on accuracy, recall, precision, f-score, and Area under curve (AUC).

Accuracy shows the fraction of predictions that the model predicts correctly right. The formula to calculate the accuracy in this study is expressed in Eq. 3. Recall measures the proportion of actual positive cases predicted as positive by the model using Eq. 4.

**Algorithm 3.** Boosting ensemble algorithm [27]

**Input:** A set of Classifiers (  $C_1, C_2, \dots, C_m$  )  $m < M$

$D$  a set of class labeled training tuples

**Output:** A composite model

```

1: for the base model,  $l = 1$  to  $m$ 
2:   ADWIN ( $D$ )
3:   for every window  $k$ 
4:     compute error rate on every window  $err_k \leftarrow \text{misclassified}_{\text{example}} / \text{total}_{\text{example}}$ 
5:     if changes detected
6:       set new sample weight  $w_i \leftarrow (1 - err_k) / err_k$ 
7:     else  $w_i \leftarrow 1$ 
8:     end if
9:     update weight
10:    learn classifier
11:  end for
12:  add next classifier  $C_{m+1}$ , drop  $C_1$  if  $m=M$ 
13: end for

```

Therefore, in this case, it refers to the model's ability to detect patients who have diabetic retinopathy correctly. Precision measures the proportion of the data points the model predicts to be relevant, which is actually relevant as given in Eq. 5.

F-score measures test accuracy. It is the harmonic mean of the precision and recall, as given in equation 6, and reaches its best value at 1. AUC shows the tradeoff between TP and FP. It provides an aggregate measure of performance across all possible classification thresholds. Parameter  $tp$  (true positive) denotes correctly identified,  $fp$  (false positive) incorrectly identified,  $tn$  (true negative) correctly rejected, and  $fn$  denotes incorrectly rejected.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (3)$$

$$Recall = \frac{tp}{tp + fn} \quad (4)$$

$$Precision = \frac{tp}{tp + fp} \quad (5)$$

$$Fscore = \frac{2tp}{2tp + fp + fn} \quad (7)$$

### III. RESULTS AND DISCUSSION

The experiments are carried out with eight (8) different learners to empirically evaluate the proposed approach's performance and compare its effectiveness against other existing approaches. The learners are DT, Bagging, Boosting, SVM, DT+FS, Bagging+FS, Boosting+FS, and SVM+FS. All experiments were performed on an HP Compaq CQ57 system running on Intel Pentium processor with 2.20 GHz speed and 4 GB of RAM. Experimental results for all classifiers are compared closely to identify similarities or differences in model predictions. Table 2 presents the preliminary results showing the  $tp$ ,  $fp$ ,  $tn$ , and  $fn$  values obtained from each model.

**Table 2.**  $tp$ ,  $fp$ ,  $tn$ , and  $fn$  values of models

Classifier	tp	fp	tn	fn
DT	374	237	367	173
Bagging	400	211	364	176
Boosting	400	211	353	187
SVM	480	131	270	270
DT + FS	374	237	369	171
Bagging + FS	400	211	367	173
Boosting + FS	409	202	365	175
SVM +FS	400	208	360	176

**Table 3.** Results of models without feature selection

Classifier	Accuracy (%)	Recall	Precision	F-score	AUC
DT	64.38	0.644	0.648	0.644	0.685
Bagging	66.38	0.664	0.666	0.664	0.731
Boosting	65.42	0.654	0.655	0.655	0.724
SVM	65.16	0.652	0.652	0.652	0.721

#### A. Results of models without feature selection

This section shows the results of prediction models without the application of the chi-square feature selection method. Table 3 presents the respective prediction models' performance (DT, Bagging, Boosting, and SVM).

It can be seen that SVM performed better than DT. This result is in accordance with other existing studies where SVM has been proven to be better than DT [4], [7], [12]. However, the homogeneous ensemble methods of DT (Bagging and Boosting) outperforms SVM. Bagged DT performed best, followed by Boosted DT when compared with SVM. The homogenous ensemble methods exploited the dependence between the base learners (DT), which in turn enhances its overall performance by weighing previously mislabeled instances of the datasets with higher weight.

## B. Results of models with feature selection (Chi-Square)

Chi-Square (CS) based on the ranker search method was further deployed with the respective prediction models. This deployment aims to improve the models' performances further as feature selection has been known to improve classifiers' performances [28], [29]. The selection of important and germane features from the dataset is obtained by log<sub>2</sub>N [24].

The predictive performances of models after employing CS as a feature selection technique is presented in Table 4. Improvements were observed when compared to the results in Table 3. It was noted that the feature selection technique had improved the predictive performance of the prediction models in all metrics. The improvement in performance metric values showed that the SVM, DT, Boosted DT, and Bagged DT works well with the feature selection technique. Specifically, the accuracy value of Boosted DT improved by 2.8 % and SVM by 1.34 %. Bagged DT (0.39 %) and SVM (0.26 %) also slightly improved accordingly. However, the homogeneous ensemble methods (Boosted DT and Bagged DT) still outperform SVM and base-classifiers DT. Bagged DT performed best, followed by Boosted DT when compared with SVM.

The improved performance recorded by all the predictive models underscores CS's effectiveness in removing redundant features and selecting more discriminating features from the feature set for enhanced classification. In other words, a much-reduced number of features were consequently used to train the predictive models, which in turn caused the observed improvement in the predictive performance.

Figure 3 shows the accuracy of each prediction model with and without the feature selection method. It can be deduced that SVM is a good predictive model as it outperforms DT. However, the optimized DT based on homogenous ensemble methods (Bagging and Boosting) is better than SVM. The ensemble technique usage is advised to be used for DR prediction as ensemble models average out biases, reduce variance, and avoid overfitting, unlike single classifier methods.

Contrary to other models' increase in performance, SVM showed an 8.18 % reduction of its AUC value. The reduction in SVM's AUC value shows that SVM had the worst separability measure compared to Bagged and Boosted DT. Nevertheless, it can be concluded from the results that feature selection positively enhanced Boosted and Bagged DT models. This improvement recorded from our result by applying CS is consistent with the results of other works, such as [21], [23], and [25] in DR prediction.

## IV. CONCLUSION

This study presented tree-based homogenous ensemble models for the prediction of DR. The bagging and boosting ensemble methods were applied to the DR dataset. The experimental results obtained confirmed that DT can increase in prediction power over SVM if ensembled. Therefore, ensemble methods increase the performance of DT and give a better predictive

Table 4. Results of models with CS feature selection

Classifier	Accuracy (%)	Recall	Precision	F-score	AUC
DT	64.55	0.646	0.650	0.646	0.687
Bagging	66.64	0.666	0.668	0.667	0.729
Boosting	67.25	0.672	0.674	0.673	0.738
SVM	66.03	0.660	0.664	0.661	0.662

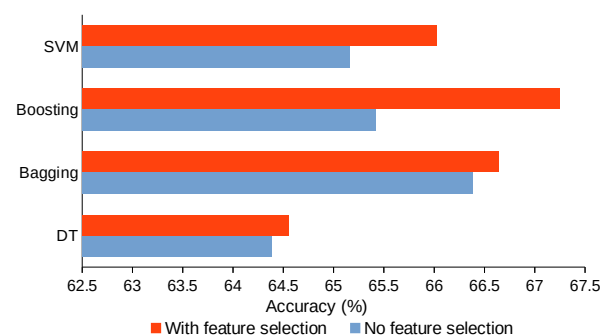


Figure 3. Accuracy of models in different stages of experiment

performance than SVM. Therefore, homogeneous ensemble methods are recommended for a more accurate diagnosis of DR. In the future, the study can be extended by considering other forms of ensemble methods (heterogeneous ensemble) and considering the data quality problems such as class imbalance and noise in DR datasets.

## REFERENCES

- [1] K. Zielinski, M. Duplaga, and D. Ingram, *Information technology solutions for healthcare*. Springer Science & Business Media, 2007. doi: [10.1007/1-84628-141-5](https://doi.org/10.1007/1-84628-141-5)
- [2] S. Dua, U. R. Acharya, and P. Dua, *Machine learning in healthcare informatics*. Springer, 2014. doi: [10.1007/978-3-642-40017-9](https://doi.org/10.1007/978-3-642-40017-9)
- [3] R. Beaglehole *et al.*, "Improving the prevention and management of chronic disease in low-income and middle-income countries: a priority for primary health care," *The Lancet*, vol. 372, no. 9642, pp. 940-949, 2008. doi: [10.1016/S0140-6736\(08\)61404-X](https://doi.org/10.1016/S0140-6736(08)61404-X)
- [4] P. S. Kumar and S. Pranavi, "Performance analysis of machine learning algorithms on diabetes dataset using big data analytic," in *International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)*, Dubai, UAE, Dec. 2017, pp. 508-513. doi: [10.1109/ICTUS.2017.8286062](https://doi.org/10.1109/ICTUS.2017.8286062)
- [5] R. Balaji, R. Duraisamy, and M. Kumar, "Complications of diabetes mellitus: A review," *Drug Invention Today*, vol. 12, no. 1, 2019.
- [6] C. Dow *et al.*, "Diet and risk of diabetic retinopathy: a systematic review," *European*

- journal of epidemiology*, vol. 33, no. 2, pp. 141-156, 2018. doi: [10.1007/s10654-017-0338-8](https://doi.org/10.1007/s10654-017-0338-8)
- [7] S. Mohammadian, A. Karsaz, and Y. M. Roshan, "A comparative analysis of classification algorithms in diabetic retinopathy screening," in *7th International Conference on Computer and Knowledge Engineering*, Mashhad, Iran, Oct. 2017, pp. 84-89. doi: [10.1109/ICCKE.2017.8167934](https://doi.org/10.1109/ICCKE.2017.8167934)
- [8] N. K. Das *et al.*, "Investigation of alterations in multifractality in optical coherence tomographic images of in vivo human retina," *Journal of Biomedical Optics*, vol. 21, no. 9, 096004, 2016. doi: [10.1117/1.JBO.21.9.096004](https://doi.org/10.1117/1.JBO.21.9.096004)
- [9] G. Mahendran and R. Dhanasekaran, "Investigation of the severity level of diabetic retinopathy using supervised classifier algorithms," *Computers & Electrical Engineering*, vol. 45, pp. 312-323, 2015. doi: [10.1016/j.compeleceng.2015.01.013](https://doi.org/10.1016/j.compeleceng.2015.01.013)
- [10] R. Pal, J. Poray, and M. Sen, "Application of machine learning algorithms on diabetic retinopathy," in *2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*, Bangalore, India, May 2017, pp. 2046-2051. doi: [10.1109/RTEICT.2017.8256959](https://doi.org/10.1109/RTEICT.2017.8256959)
- [11] P. Sonar and K. JayaMalini, "Diabetes prediction using different machine learning approaches," in *3rd International Conference on Computing Methodologies and Communication*, Erode, India, Mar. 2019, pp. 367-371. doi: [10.1109/ICCMC.2019.8819841](https://doi.org/10.1109/ICCMC.2019.8819841)
- [12] H.-Y. Tsao, P.-Y. Chan, and E. C.-Y. Su, "Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms," *BMC bioinformatics*, vol. 19, no. 9, 195, 2018. doi: [10.1186/s12859-018-2277-0](https://doi.org/10.1186/s12859-018-2277-0)
- [13] S. Cui, D. Wang, Y. Wang, P.-W. Yu, and Y. Jin, "An improved support vector machine-based diabetic readmission prediction," *Computer Methods and Programs in Biomedicine*, vol. 166, pp. 123-135, 2018. doi: [10.1016/j.cmpb.2018.10.012](https://doi.org/10.1016/j.cmpb.2018.10.012)
- [14] S. Yin and J. Yin, "Tuning kernel parameters for SVM based on expected square distance ratio," *Information Sciences*, vol. 370, pp. 92-102, 2016. doi: [10.1016/j.ins.2016.07.047](https://doi.org/10.1016/j.ins.2016.07.047)
- [15] D. Zhao, H. Liu, Y. Zheng, Y. He, D. Lu, and C. Lyu, "Whale optimized mixed kernel function of support vector machine for colorectal cancer diagnosis," *Journal of Biomedical Informatics*, vol. 92, 103124, 2019. doi: [10.1016/j.jbi.2019.103124](https://doi.org/10.1016/j.jbi.2019.103124)
- [16] A. O. Balogun, A. O. Bajeh, V. A. Orie, and W. A. Yusuf-Asaju, "Software defect prediction using ensemble learning: an ANP based evaluation method," *FUOYE Journal of Engineering and Technology*, vol. 3, no. 2, pp. 50-55, 2018. doi: [10.46792/fuoyejet.v3i2.200](https://doi.org/10.46792/fuoyejet.v3i2.200)
- [17] A. O. Balogun, A. M. Balogun, P. O. Sadiku, and L. Amusa, "An ensemble approach based on decision tree and bayesian network for intrusion detection," *Annals. Computer Science Series*, vol. 15, no. 1, pp. 82-91, 2017.
- [18] S. P. Healey *et al.*, "Mapping forest change using stacked generalization: An ensemble approach," *Remote Sensing of Environment*, vol. 204, pp. 717-728, 2018. doi: [10.1016/j.rse.2017.09.029](https://doi.org/10.1016/j.rse.2017.09.029)
- [19] N. Gurudath, M. Celenk, and H. B. Riley, "Machine learning identification of diabetic retinopathy from fundus images," in *IEEE Signal Processing in Medicine and Biology Symposium*, Philadelphia, USA, Dec. 2014, pp. 1-7. doi: [10.1109/SPMB.2014.7002949](https://doi.org/10.1109/SPMB.2014.7002949)
- [20] J. Lachure, A. Deorankar, S. Lachure, S. Gupta, and R. Jadhav, "Diabetic Retinopathy using morphological operations and machine learning," in *IEEE International Advance Computing Conference*, Bangalore, India, Jun. 2015, pp. 617-622. doi: [10.1109/IADCC.2015.7154781](https://doi.org/10.1109/IADCC.2015.7154781)
- [21] S. Murugeswari and R. Sukanesh, "Investigations of severity level measurements for diabetic macular oedema using machine learning algorithms," *Irish Journal of Medical Science (1971-)*, vol. 186, no. 4, pp. 929-938, 2017. doi: [10.1007/s11845-017-1598-8](https://doi.org/10.1007/s11845-017-1598-8)
- [22] E. V. Carrera, A. González, and R. Carrera, "Automated detection of diabetic retinopathy using SVM," in *IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing*, Cusco, Peru, Aug. 2017, pp. 1-4. doi: [10.1109/INTERCON.2017.8079692](https://doi.org/10.1109/INTERCON.2017.8079692)
- [23] S. Somasundaram and P. Alli, "A machine learning ensemble classifier for early prediction of diabetic retinopathy," *Journal of Medical Systems*, vol. 41, no. 12, 201, 2017. doi: [10.1007/s10916-017-0853-x](https://doi.org/10.1007/s10916-017-0853-x)
- [24] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, "Performance analysis of feature selection methods in software defect prediction: a search method approach," *Applied Sciences*, vol. 9, no. 13, p. 2764, 2019. doi: [10.3390/app9132764](https://doi.org/10.3390/app9132764)
- [25] S. Piri, D. Delen, T. Liu, and H. M. Zolbanin, "A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble," *Decision Support Systems*, vol. 101, pp. 12-27, 2017. doi: [10.1016/j.dss.2017.05.012](https://doi.org/10.1016/j.dss.2017.05.012)
- [26] Y. Yang, *Temporal data mining via unsupervised ensemble learning*. Elsevier, 2016. doi: [10.1016/B978-0-12-811654-8.00004-X](https://doi.org/10.1016/B978-0-12-811654-8.00004-X)
- [27] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications*. Springer, 2012. doi: [10.1007/978-1-4419-9326-7](https://doi.org/10.1007/978-1-4419-9326-7)
- [28] A. G. Akintola, A. O. Balogun, F. Lafenwa-Balogun, and H. A. Mojeed, "Comparative analysis of selected heterogeneous classifiers for software defects prediction using filter-based feature selection methods," *FUOYE Journal of Engineering and Technology*, vol. 3, no. 1, pp. 134-137, 2018. doi: [10.46792/fuoyejet.v3i1.178](https://doi.org/10.46792/fuoyejet.v3i1.178)
- [29] M. A. Mabayoje, A. O. Balogun, S. M. Bello, J. O. Atoyebi, H. A. Mojeed, and A. H. Ekundayo, "Wrapper feature selection based heterogeneous classifiers for software defect prediction," *Adeleke University Journal of Engineering and Technology*, vol. 2, no. 1, pp. 1-11, 2019.