Scientific Journal of Informatics
Vol. 11, No. 2, May 2024

p-ISSN 2407-7658 http://journal.unnes.ac.id/ journals/sji/index e-ISSN 2460-0040

Performance of Ensemble Learning in Diabetic Retinopathy Disease Classification

Anisa Nurizki[1*], Anwar Fitrianto[2], Agus Mohamad Soleh[3]
[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Sciences,
IPB University, Indonesia

Abstract.
Purpose: This study explores diabetic retinopathy (DR), a complication of diabetes leading to blindness, emphasizing early diagnostic interventions. Leveraging Macular OCT scan data, it aims to optimize prevention strategies through tree-based ensemble learning.
Methods: Data from RSKM Eye Center Padang (October-December 2022) were categorized into four scenarios based on physician certificates: Negative & non-diagnostic DR versus Positive DR, Negative versus Positive DR, Non-Diagnosis versus Positive DR, and Negative DR versus non-Diagnosis versus Positive DR. The suitability of each scenario for ensemble learning was assessed. Class imbalance was addressed with SMOTE, while potential underfitting in random forest models was investigated. Models (RF, ET, XGBoost, DRF) were compared based on accuracy, precision, recall, and speed.
Results: Tree-based ensemble learning effectively classifies DR, with RF performing exceptionally well (80% recall, 78.15% precision). ET demonstrates superior speed. Scenario III, encompassing positive and undiagnosed DR, emerges as optimal, with the highest recall and precision values. These findings underscore the practical utility of tree-based ensemble learning in DR classification, notably in Scenario III.
Novelty: This research distinguishes itself with its unique approach to validating tree-based ensemble learning for DR classification. This validation was accomplished using Macular OCT data and physician certificates, with ETDRS scores demonstrating promising classification capabilities.

Keywords: Classification, Ensemble learning, Random forest, Extra-trees, XGBoost, Double random forest, Diabetic retinopathy
Received May 2024 / **Revised** May 2024 / **Accepted** May 2024

INTRODUCTION

Diabetic retinopathy stands as a significant complication of diabetes mellitus, capable of inducing blindness if not promptly identified and treated [1], [2]. The classification of diabetic retinopathy has gained paramount importance in light of the escalating global prevalence of diabetes [3]. Early detection facilitated by precise classification holds the potential to avert severe visual impairment and blindness, thereby underscoring the pivotal role of advanced diagnostic technologies.

Ensemble models in machine learning, amalgamating multiple models to enhance predictive performance, present a promising avenue for diabetic retinopathy classification. Dietterich [4] contends that ensemble models, featuring an array of iterative classifiers, often yield more precise classification or prediction outcomes compared to individual classifiers.

Optical coherence tomography (OCT) emerges as a pivotal technology in the early detection of diabetic retinopathy. OCT furnishes quantitative insights crucial for monitoring treatment responses and the progression of diverse ocular ailments [5]–[7]. Notably, the macula can be segmented into nine subfields following the Early Treatment of Diabetic Retinopathy Study (ETDRS) protocol, pivotal for clinical investigations. This segmentation encompasses three concentric circles with respective diameters of 1 mm,

[*]Corresponding author.
Email addresses: anisanurizki@apps.ipb.ac.id (Nurizki), anwarstat@gmail.com (Fitrianto), agusms@apps.ipb.ac.id (Soleh)
DOI: [10.15294/sji.v11i2.4725](https://doi.org/10.15294/sji.v11i2.4725)

3 mm, and 6 mm for the central, inner, and outer circles [5], each subdivided into four regions: superior, temporal, inferior, and nasal [6] elaborated in Figure 1 [8].

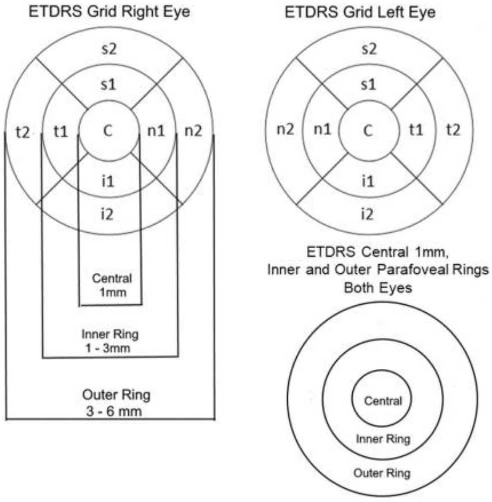


Figure 1. Nine ETDRS subfields are in the right and left eye, and inner and outer parafoveal rings are in both eyes.

Figure 1 presents ETDRS grids along with central thickness (C), inner superior (s1), inner temporal (t1), inner inferior (i1), inner nasal (n1), outer superior (s2), outer temporal (t2), outer inferior (i2), and outer nasal (n2) values. OCT measurements yield retinal thickness data for each subfield [6] with deviations from normal thickness serving as indicators of diabetic retinopathy, thus rendering OCT a valuable tool for early detection and prevention of blindness.

Recent years have witnessed a surge in research on diabetic retinopathy employing machine learning or ensemble models. Prior investigations have showcased the efficacy of various machine learning techniques in diabetic retinopathy classification. These encompass Decision Trees, Bagging, Boosting, Support Vector Machines, Neural Networks, and Deep Learning, leveraging datasets sourced from the UCI Machine Learning repository [9]–[12]or clinical data [13]. Outcomes of these studies have reported disparate levels of accuracy, underlining the potential of machine learning models in refining the accuracy of diabetic retinopathy diagnosis. Our study extends this extant research, striving to augment the accuracy and efficacy of diabetic retinopathy classification.

This study aims to classify diabetic retinopathy using a tree-based ensemble approach. We utilize Macular OCT data, emphasizing ETDRS values and information from medical certificates, to attain optimal classification results. The ensemble methods employed include Random Forest (RF), Extremely Randomized Trees (ET), Extreme Gradient Boosting (XGBoost), and Double Random Forest (DRF). We conduct a comprehensive evaluation of each model's performance to comprehend their effectiveness in identifying and classifying diabetic retinopathy, ultimately contributing to the prevention of vision loss associated with this disease.

METHODS

Prior to data analysis, the macular OCT scan data obtained from RSKM Padang Eye Center were initially entered into Microsoft Excel. Subsequently, the data were categorized into three groups based on available doctor's certificates: positive and negative for diabetic retinopathy, as well as non-diagnosis, thereby facilitating the data classification process. Following this categorization, data encoding was conducted in RStudio, converting categorical data into numeric factors. This phase is also referred to as the preprocessing stage. The subsequent step involved data exploration to assess the data's condition and identify any imbalanced classes and potential underfitting. Following this, the machine learning stage and model performance evaluation were executed. The flowchart of the data analysis stages is illustrated in Figure 2. In this study, the recall value serves as the most crucial measure of goodness-of-fit, accurately classifying positive observations [14].

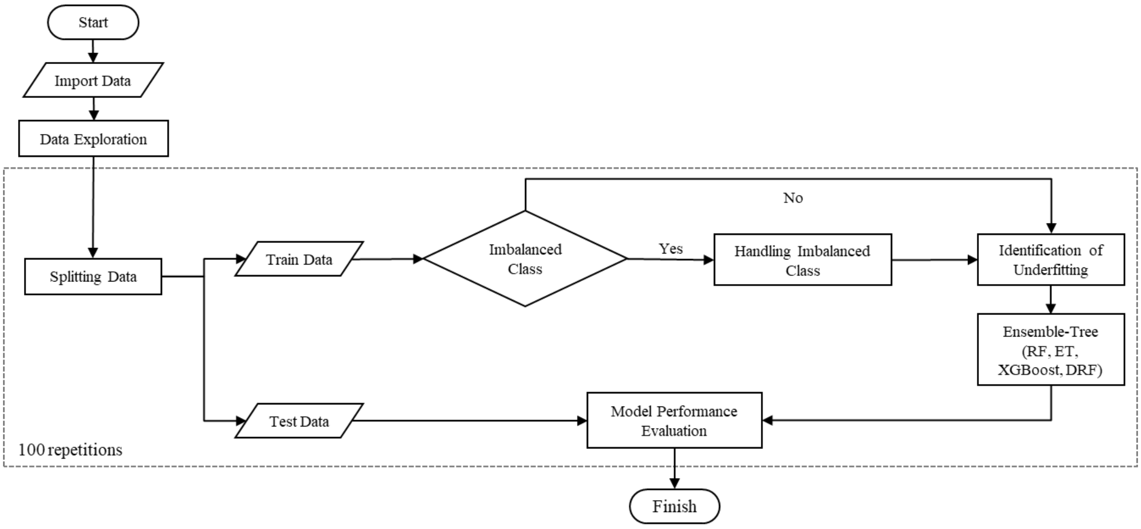
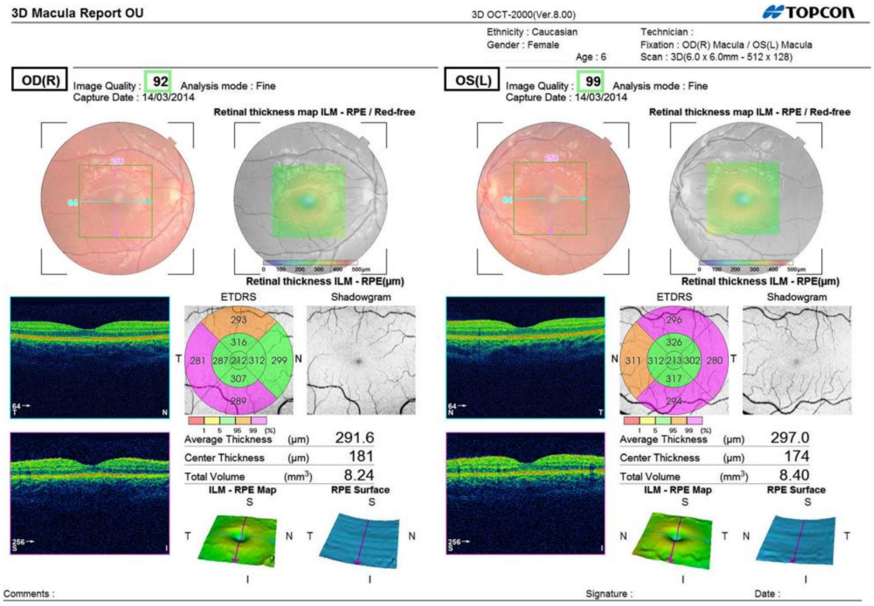

Start
Import Data
Data Exploration
Splitting Data
Train Data
Imbalanced Class
Yes
No
Handling Imbalanced Class
Identification of Underfitting
Ensemble-Tree (RF, ET, XGBoost, DRF)
Test Data
Model Performance Evaluation
100 repetitions
Finish

Figure 2. Flowchart of diabetic retinopathy data analysis stages

Data Collection

This research utilizes secondary data from the Padang Eye Center, a specialized eye hospital in Padang, West Sumatra. Data on macular OCT scan results were collected from October to December 2022. Examples of macular OCT scan results are presented in Figure 3 [15].


3D Macula Report OU

3D OCT-2000(Ver.8.00)

Ethnicity : Caucasian
Gender : Female
Age : 6

Technician :
Fixation : OD(R) Macula / OS(L) Macula
Scan : 3D(6.0 x 6.0mm - 512 x 128)

OD(R) Image Quality : 92 Analysis mode : Fine
Capture Date : 14/03/2014

OS(L) Image Quality : 99 Analysis mode : Fine
Capture Date : 14/03/2014

	OD(R)	OS(L)
Average Thickness (μm)	291.6	297.0
Center Thickness (μm)	181	174
Total Volume (mm^3)	8.24	8.40

Comments :

Signature :

Date :

Figure 3. Figure 5 example of macular OCT scan results

The dataset comprises 455 macular OCT scan results from 631 scanned eyeballs, which were entered individually into Microsoft Excel. The inputted data is subsequently analyzed using RStudio 3.4.0. There are 15 predictor variables extracted from the macular OCT scan results, as detailed in Table 1. Generally, the selected predictor variables consist of ETDRS data along with age, gender, and eye type variables. The inclusion of these variables is a novel approach in classifying diabetic retinopathy through the utilization of macular OCT scan data.

Table 1. Predictor variables

Variable	Variable Descriptions	Descriptions
x_1	Gender	1 = Male 0 = Female
x_2	Age	Numeric
x_3	Eye Type	1 = Right Eye 0 = Left Eye
x_4	Average Thickness	Numeric
x_5	Center Thickness	Numeric
x_6	Total Volume	Numeric
x_7	Foveal Thickness (C)	Numeric
x_8	Inner Superior (s1)	Numeric
x_9	Inner Nasal (n1)	Numeric
x_{10}	Inner Inferior (i1)	Numeric
x_{11}	Inner Temporal (t1)	Numeric
x_{12}	Outer Superior (s2)	Numeric
x_{13}	Outer Nasal (n2)	Numeric
x_{14}	Outer Inferior (i2)	Numeric
x_{15}	Outer Temporal (t2)	Numeric

Meanwhile, data from the doctors' notes regarding the scan results are utilized for the response variable. Among these, 140 eyeballs lack notes, categorized as non-diagnosis (healthy eyes), while 491 eyeballs possess disease notes, segmented into positive and negative for diabetic retinopathy. Four research scenarios have been formulated to facilitate effective analysis in the early prevention of retinopathy. The primary distinction among these scenarios lies in the response variable, while the predictor variables used in each scenario remain consistent. Table 2 below describes the response variables utilized in each scenario.

Table 2. Response variables

Scenario	Class used	Number of Observations	Type of Response Variables
First (I)	0 = Negative DR and non-Diagnosis 1 = Positive DR	631	Binary
Second (II)	0 = Negative DR 1 = Positive DR	491	Binary
Third (III)	0 = non-Diagnosis 1 = Positive DR	276	Binary
Fourth (IV)	0 = Negative DR 1 = non-Diagnosis 2 = Positive DR	631	Multiclass

Random Forest

Random Forest constructs a model by employing bootstrap and aggregation (bagging) techniques and selects the optimal tree separation through random feature selection. According to [16], succinctly, the algorithm for forming a Random Forest can be outlined as follows:

1) The training dataset is of size n with p explanatory variables.
2) Perform bootstrap to construct trees by drawing random samples of size n from the training dataset.
3) Randomly select m predictor variables $(m < p)$ at each tree split in the random feature selection phase. The number of m predictor variables is approximately $m \approx \sqrt{p}$ or $m \approx \frac{p}{3}$.
4) Repeat steps (2) and (3) k times to obtain k random trees.
5) Combine the prediction results from k random trees in the aggregation phase.
6) Calculate the majority vote for classification data.

Extremely Randomized Trees

Extremely Randomized Trees, commonly known as Extra-trees (ET), are among the ensemble models that employ thorough randomization. The ET model randomizes the selection of m predictor variables and split points while utilizing the entire training data to construct a tree. This randomization process aims to reduce the variance of prediction outcomes from each tree while utilizing the entire training data to mitigate bias [17], [18]. According to [18], comprehensive randomization simplifies the node-splitting procedure, resulting in reduced computation time for the ET model. The algorithm for building ET can be elucidated as follows [18]:

1) Form trees using all training data (D). Each tree is built using the same data, so $D = D_1 = D_2 = D_3 = \cdots = D_k$.
2) Select the best split using the following steps:
    a. Randomly select m predictor variables $(m < p)$ at each tree split using random feature selection. The number of m predictor variables is approximately $m \approx \sqrt{p}$ or $m \approx \frac{p}{3}$.
    b. Randomly select k split points.
    c. Repeat steps (a) and (b) until the stopping criteria are reached to obtain the prediction results from one tree.
3) Repeat steps (1) and (2) until M trees are formed.
4) Combine the prediction results from k random trees during the aggregation stage.
5) Calculate the majority vote for data classification.

Extreme Gradient Boosting

XGBoost, an acronym for Extreme Gradient Boosting, represents a refinement of the Gradient Boosting model, initially developed by Chen and Guestrin [19]. The Gradient Boosting method constructs trees sequentially or in a sequence. In boosting, there exists a term called the loss function, which evaluates the quality of the tree structure. The smaller the value of the loss function, the higher the quality of the constructed model. The loss function in Gradient Boosting is minimized using partial derivatives or gradients.

Each model is constructed in one round with a weak learner, and its prediction results are contrasted with the anticipated results. The disparity between the predicted and anticipated results is termed the error rate, which indicates the extent of the model's inaccuracies. Model parameters can be adjusted to mitigate errors in the subsequent training round. Consequently, the gradient employed in the training in the preceding iteration will influence the tree to be constructed in the subsequent iteration [20].

In XGBoost, the loss function is minimized using the second partial derivative, which also aids in comprehending gradient trends. The model training process employing XGBoost can be parallelized, leading to expedited computational processes compared to gradient-boosting models [20].

Double Random Forest

he Double Random Forest (DRF) model, developed by Han et al. [21], aims to enhance the performance of RF. The primary distinction lies in the tree construction methodology. While RF employs the bootstrap method solely at the root node, DRF applies bootstrap at every node during tree-building. This approach grants DRF a more varied set of trees, thereby augmenting the likelihood of more accurate prediction results. Succinctly, the algorithm for constructing DRF can be delineated as follows [21]:

1) Forming trees using the entire training data (D).
2) Selecting the best splitting with the following steps:
    a. At each node t, perform random sampling of size n_t^* with replacement (bootstrap), if $n_t > n \times 0.1$. If not met, skip the bootstrap.
    b. Randomly select m variables from the set of p predictor variables.
    c. Determine the best splitting criteria.
    d. Repeat steps (a) to (c) until the stopping criteria are met to obtain the prediction result of one tree.
3) Repeat steps (1) and (2) until a total of M classification trees are formed.
4) Combine the prediction results from each tree using a majority vote for classification data.

Evaluation model

A measure of model goodness is utilized to assess the performance of the model in classification tasks. Evaluating the model's efficacy involves computing the number of observations correctly predicted as positive or negative classes and the number of observations incorrectly predicted as positive or negative classes [22]. These four calculations constitute the confusion matrix in Table 3 for the binary classification scenario.

Table 3. Confusion matrix for binary classification

Actual	Prediction	
	Positive Class (1)	Negative Class (0)
Positive Class (1)	TP	FN
Negative Class (0)	FP	TN

Table 3 illustrates TP (True Positive) as the count of correctly predicted positive class members (1). TN (True Negative) represents the count of correctly predicted negative class members (0). FP (False Positive) signifies the count of positive class members (1) erroneously predicted. Finally, FN (False Negative) denotes the count of negative class members (0) inaccurately identified. Additionally, Equations (1)-(3) below outline the computations for accuracy, recall, and precision values.

1) Accuracy serves as a metric indicating the model's capability to make accurate predictions.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

2) Recall assesses the model's ability to correctly identify positive observations.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

3) Precision quantifies the proportion of predicted positive cases that are truly positive.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Binary class classification pertains to two classes in the response variable: positive and negative, whereas multi-class classification involves more than two classes. This disparity impacts the utilization of the confusion matrix as the evaluation metric for model performance. The confusion matrix for multi-class classification is depicted in Table 4 below.

Table 4. Confusion matrix for multi-class classification

		Prediction			
		Class 1	Class 2	···	Class *N*
Actual	Class 1	$C_{1,1}$	FP	···	$C_{1,3}$
	Class 2	FN	TP	···	FN
	···	···	···	···	···
	Class *N*	$C_{N,1}$	FP	···	$C_{N,N}$

The blue and yellow colors in Table 4 indicate that the present example centers on classifying class 2 against other classes. The metrics employed to assess the performance of the multi-class classification model include accuracy and recall. Equations (4)-(6) are employed to compute the accuracy, recall, and precision values for multi-class classification [23].

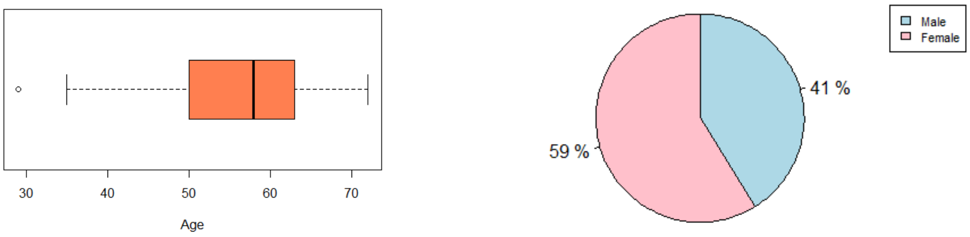
$$Accuracy = \frac{\sum_{i=1}^{N} TP(C_i)}{\sum_{i=1}^{N} \sum_{j=1}^{N} C_{i,j}} \tag{4}$$

$$Recall\ (C_i) = \frac{TP(C_i)}{TP(C_i) + FN(C_i)} \tag{5}$$

$$Precision(C_i) = \frac{TP(C_i)}{TP(C_i) + FP(C_i)} \tag{6}$$

RESULTS AND DISCUSSIONS

The distribution of data from ETDRS OCT results of retinopathy patients, as depicted in Figure 4a, reveals that the maximum age range is 72 years, with the majority of data falling between ages 50 to 63 years. The data tends to cluster around the median age of 58 years. Additionally, Figure 4a highlights one data outlier with an age of 29 years. Moreover, concerning gender distribution, as illustrated in Figure 4b, diabetic retinopathy patients are predominantly female, constituting 59% of 136 patients. In contrast, male patients afflicted with diabetic retinopathy comprise 41%. Boxplots and pie charts elucidating the distribution of diabetic retinopathy patients are presented in Figure 4.



(a) (b)

Figure 4. Patients with diabetic retinopathy (a) Boxplot by age, (b) Pie chart by gender

Identification of Imbalanced Classes in The Data

Most scenarios exhibit imbalanced classes, which can result in misclassification. Minority classes tend to be misclassified more often than majority classes [24]. The issue arises when the minority class data contains vital information, such as in ETDRS data. In scenarios I, II, and IV, the diabetic retinopathy positive class represents the minority class. Classification errors can lead to decision-making errors, particularly in the accuracy of minority class predictions [24].

Class imbalance can be identified through the calculation of the Imbalance Ratio (IR). According to [25], when $IR = 1$, the class is considered balanced, while $IR > 1$ indicates an imbalanced class. The larger the IR value, the greater the class imbalance level. For binary classes, IR is defined by Equation (7), where N_{maj} represents the number of majority classes and N_{min} represents the number of minority classes. Meanwhile, for multi-classes, IR is defined by Equation (8) [26]. Unlike Equation (7), the IR calculation utilizes the proportions of the majority class $\left(\hat{p}_{maj}\right)$ and the minority class $(\hat{p}_{min})$.

$$IR_{biner} = \frac{N_{maj}}{N_{min}} \tag{7}$$

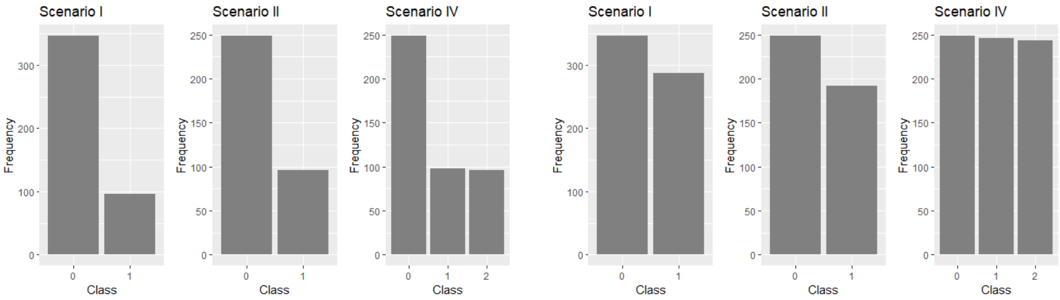
$$IR_{multi} = \frac{\hat{p}_{maj}}{\hat{p}_{min}} \tag{8}$$

The class distribution of each scenario is outlined in Table 5. In scenarios I, II, and IV, the class distributions of the response variables are imbalanced, whereas those in scenario III tend to be balanced. The imbalanced classes were addressed using the Synthetic Minority Over-sampling Technique (SMOTE) algorithm.

Table 5. Class distribution of each scenario

Skenario	Distribusi Kelas			IR
	0	1	2	
I	495	136	-	3,64
II	355	136	-	2,61
III	140	136	-	1,03
IV	355	140	136	2,61

Table 5 compares class distributions before and after implementing techniques for handling imbalanced classes. The bar chart in Figure 5 illustrates that SMOTE effectively addresses class imbalance. As per [27], [28] employing SMOTE to handle imbalanced classes can enhance model performance by generating synthetic data from minority classes



(a) (b)

Figure 5. Scenario I, II, and IV class distribution bar charts (a) before SMOTE, (b) after SMOTE

Identification of The Possibility of Underfitting

The identification of underfitting possibility is crucial for assessing the condition of the RF model during data modeling. Underfitting may occur in the RF model if the relative test accuracy value is consistently below one at each specified nodesize. The relative test accuracy value is obtained from the comparison of the accuracy value of nodesize $= 1$with the accuracy value of nodesize < 1 [21]. In this study, nodesizes of $1\,;0{,}1n;\ 0{,}09n\,;\ldots;0{,}03n$ and $0{,}02n$ are utilized, where n represents the number of training data. Evaluation results indicate that all scenarios exhibit signs of underfitting possibility. Figure 6 provides a visualization of the relative test accuracy for the data used.

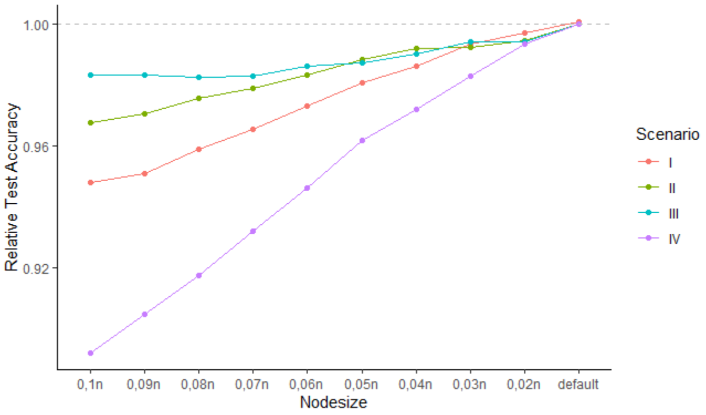


Figure 6. Relative test accuracy

Figure 6 illustrates that the relative test accuracy in each scenario is less than one. This suggests that all scenarios indicate the possibility of underfitting in the RF model. Detailed relative test accuracy values are provided in Table 6.

Table 6 Relative test accuracy

Nodesize	Scenario I	Scenario II	Scenario III	Scenario IV
0,10n	0,9479	0,9676	0,9834	0,8918
0,09n	0,9510	0,9705	0,9834	0,9044
0,08n	0,9590	0,9756	0,9827	0,9175
0,07n	0,9655	0,9788	0,9831	0,9318
0,06n	0,9732	0,9834	0,9861	0,9460
0,05n	0,9807	0,9883	0,9875	0,9619
0,04n	0,9864	0,9922	0,9903	0,9719
0,03n	0,9935	0,9925	0,9944	0,9828
0,02n	0,9973	0,9945	0,9944	0,9935

Modeling and model performance evaluation

In this research, the data is divided into two parts: 70% training data and 30% test data. The training data is utilized for machine learning, whereas the test data is employed to evaluate the model's performance.

The data is repeatedly divided for model performance evaluation until the process is iterated 100 times. The results of model performance evaluation for each ensemble tree are depicted in Figures 7, 8, and 9 below.

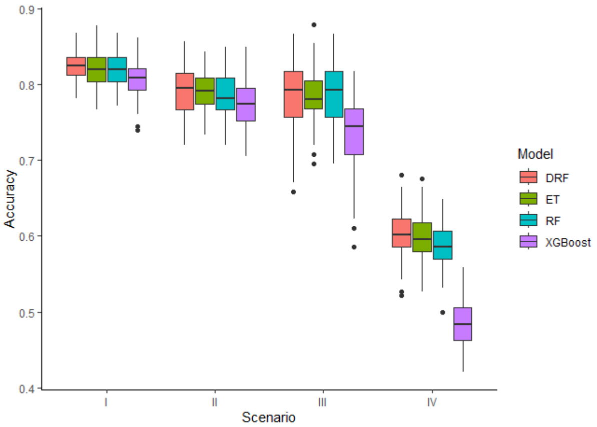
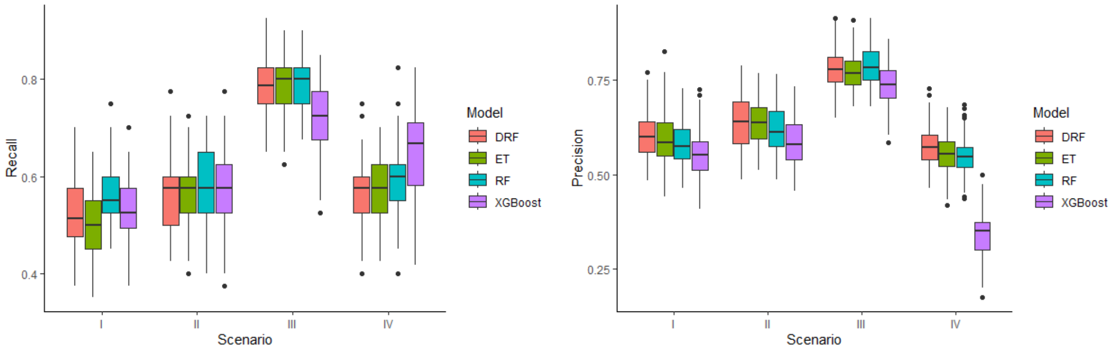


Figure 7. Accuracy of scenarios I, II, III, dan IV

Analysis of Figure 7 reveals that Scenario I exhibit higher accuracy compared to the other scenarios, while Scenario IV demonstrates the lowest accuracy. Additionally, the distribution of accuracy values in Scenario I is narrower, contrasting with Scenario III, where the distribution is more varied. Scenario I is deemed the best-performing, with accuracy values for each model of 79.27% for DRF and RF, 78.05% for ET, and 74.39% for XGBoost. Even though Scenario I received the highest rating based on accuracy, the positive class received more attention in this study, so evaluations that considered positive values, such as recall and precision, were prioritized. Consequently, the subsequent evaluation will concentrate on the recall and precision values of the model, as illustrated in Figure 8 below.



(a)

(b)

Figure 8. Evaluation measure for a positive class of scenarios I, II, III, and IV (a) Recall, (b) Precision

Previously, Scenario I was considered the best based on accuracy; however, its recall performance was less than optimal when considering recall and precision. The classification of patients positive for diabetic retinopathy in Scenario I exhibited lower recall values compared to other scenarios for each ensemble-tree model. Conversely, Scenario III demonstrated superior recall values compared to Scenarios I, II, and IV. The recall values obtained for each model in Scenario III were 80% for RF and ET, 78.75% for DRF, and 72.50% for XGBoost. Notably, due to the smaller distribution of recall values in the RF model compared to other models, the RF model emerged as the best model based on recall values. Similarly, in terms of precision values, Scenario III showed superiority over different scenarios. The precision values in Scenario III also indicated that the RF model outperformed other ensemble-tree models, achieving a precision value of 78.15%, compared to 77.64% for the DRF model, 76.74% for the ET model, and 73.68% for the XGBoost model.

In this study, the model's performance was evaluated using various measures, providing essential guidelines for determining the best scenario for classifying diabetic retinopathy diseases using Macular OCT scan data. The model performance evaluation results indicated that Scenario III tended to be more effective in classifying diabetic retinopathy diseases.

These four scenarios also presented the results of ensemble-tree model performance. However, the resulting ensemble-tree model performance did not exhibit a clear tendency toward the best model. Yet, in terms of computing time, the ET model demonstrated quick machine learning performance. The model's speed in performing machine learning varied depending on the device used; however, generally, the more sophisticated the device, the faster the computation time. In this research, a device with 8 GB RAM and an Intel Core i5 processor was utilized. The computation time of each ensemble-tree model is presented in Table 7 below.

Table 7. Computation time of ensemble-tree

Scenario	Computation time (sec)			
	RF	ET	XGB	DRF
I	0,3476	0,2293	1,2495	1,9943
II	0,2448	0,1693	1,4056	1,1055
III	0,1058	0,1022	1,1463	0,3996
IV	0,5062	0,3779	1,8484	2,6307

Table 7 shows the computation time of the ensemble-tree model during machine learning. The study revealed that the ET model efficiently performed machine learning, requiring only a short time. This finding aligns with the research results of [18], which state that the ET model can execute machine learning quickly. On average, the ET model took approximately 0.2 seconds in Scenarios I and II, 0.1 seconds in Scenario III, and 0.4 seconds in Scenario IV. Additionally, the RF model also exhibited high speed and emerged as the fastest model following the ET model.

CONCLUSION

This research encompassed several critical stages, including data collection, pre-processing, applying the SMOTE technique to handle data imbalance, and training an ensemble model to classify diabetic retinopathy using a decision tree-based ensemble approach. The models employed included RF, XGBoost, ET, and DRF. Evaluation results highlighted that Scenario III, comprising positive and undiagnosed diabetic retinopathy classes, was optimal with superior recall and precision values. Even though the RF model experiences the possibility of underfitting, the RF model can still provide the best performance. The RF model stood out with a recall value of 80% and a precision of 78.15%, while the ET model demonstrated the best performance in computing speed. These findings affirm the effectiveness of the proposed method in predicting the detection of diabetic retinopathy with high accuracy and sensitivity.

REFERENCES

[1] P. N. Dewi, F. Fadrian, and H. Vitresia, “Profil Tingkat Keparahan Retinopati Diabetik Dengan Atau Tanpa Hipertensi pada di RSUP Dr. M. Djamil Padang,” *J. Kesehat. Andalas*, vol. 8, no. 2, p. 204, 2019, doi: 10.25077/jka.v8i2.993.

[2] R. Pranata, R. Vania, and A. A. Victor, “Statin reduces the incidence of diabetic retinopathy and its need for intervention: A systematic review and meta-analysis,” *Eur. J. Ophthalmol.*, vol. 31, no. 3, pp. 1216–1224, 2021, doi: 10.1177/1120672120922444.

[3] I. D. Federation, “IDF Diabetes Atlas,” Bussels, 2021. [Online]. Available: https://diabetesatlas.org/atlas/tenth-edition/

[4] T. G. Dietterich and Oregon, “Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Models,” *Oncogene*, vol. 12, no. 2, pp. 1–15, 2000.

[5] M. S. Kim, D. Y. Kim, Y. J. Jo, and J. Y. Kim, “The Effect of Machine Aging on the Measurements of Optical Coherency Tomography,” *J. Korean Ophthalmol. Soc.*, vol. 57, no. 7, p. 1087, 2016, doi: 10.3341/jkos.2016.57.7.1087.

[6] W. Zhu, J. Y. Ku, Y. Zheng, P. C. Knox, R. Kolamunnage-Dona, and G. Czanner, “Spatial linear mixed effects modelling for oct images: Slme model,” *J. Imaging*, vol. 6, no. 6, pp. 1–16, 2020, doi: 10.3390/jimaging6060044.

[7] L. Raffa and S. S. AlSwealh, “Normative optical coherence tomography reference ranges of the optic nerve head, nerve fiber layer, and macula in healthy Saudi children,” *Saudi Med. J.*, vol. 44,

no. 12, pp. 1269–1276, 2023, doi: 10.15537/smj.2023.44.12.20230517.

[8] P. H. Scanlon, C. F. E. Norridge, P. H. J. Donachie, and Q. Mohamed, “Should we still be performing macular laser for non-centre involving diabetic macular oedema? Results from a UK centre,” *Eye*, vol. 36, no. 3, pp. 646–648, 2022, doi: 10.1038/s41433-021-01789-3.

[9] 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,” *J. Teknol. dan Sist. Komput.*, vol. 8, no. 4, pp. 297–303, 2020, doi: 10.14710/jtsiskom.2020.13669.

[10] P. Subarkah, M. M. Abdallah, S. Oktaviani, and N. Hidayah, “Komparasi Akurasi Algoritme CART dan Neural Network Untuk Diagnosis Penyakit Diabetic Retinopathy Comparison of CART Algorithm Accuracy and Neural Network for Diagnosing Diabetic Retinopathy Disease,” *Cogito Smart J.* |, vol. 7, no. 1, pp. 121–134, 2021.

[11] S. H. Abdullah, R. Magdalena, and R. Y. N. Fu’adah, “Klasifikasi Diabetic Retinopathy Berbasis Pengolahan Citra Fundus Dan Deep Learning,” *J. Electr. Syst. Control Eng.*, vol. 5, no. 2, pp. 84–90, 2022, doi: 10.31289/jesce.v5i2.5659.

[12] S. Dasari, B. Poonguzhali, M. Rayudu, K. Muthukumar, and R. Dhanalakshmi, “Binary Classification of Diabetic Retinopathy using Random Forest Classifier,” *Eur. Chem. Bull*, vol. 12, no. 1B, pp. 2276–2284, 2023, doi: 10.31838/ecb/2023.12.s1-B.226.

[13] O. I. Ogunyemi *et al.*, “Detecting diabetic retinopathy through machine learning on electronic health record data from an urban, safety net healthcare system,” *JAMIA Open*, vol. 4, no. 3, pp. 1–10, 2021, doi: 10.1093/jamiaopen/ooab066.

[14] M. Krzywicka and A. Wosiak, “Sensitivity of Standard Evaluation Metrics for Disease Classification and Progression Assessment Based on Whole-Body Imaging,” *Procedia Comput. Sci.*, vol. 225, pp. 4314–4323, 2023, doi: 10.1016/j.procs.2023.10.428.

[15] A. Muñoz-Gallego *et al.*, “Measurement of macular thickness with optical coherence tomography: impact of using a paediatric reference database and analysis of interocular symmetry,” *Graefe’s Arch. Clin. Exp. Ophthalmol.*, vol. 259, no. 2, pp. 533–545, 2021, doi: 10.1007/s00417-020-04903-5.

[16] B. Sartono and U. D. Syafitri, “Metode Pohon Gabungan: Solusi Pilihan untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal,” *Forum Stat. dan Komputasi*, vol. 15, no. 1, pp. 1–7, 2010, [Online]. Available: https://journal.ipb.ac.id/index.php/statistika/article/view/4895

[17] E. K. Ampomah, Z. Qin, and G. Nyame, “Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement,” *Inf.*, vol. 11, no. 6, 2020, doi: 10.3390/info11060332.

[18] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.

[19] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[20] S. V. Murty and R. Kiran Kumar, “Accurate liver disease prediction with extreme gradient boosting,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 2288–2295, 2019, doi: 10.35940/ijeat.F8684.088619.

[21] S. Han, H. Kim, and Y. S. Lee, “Double random forest,” *Mach. Learn.*, vol. 109, no. 8, pp. 1569–1586, 2020, doi: 10.1007/s10994-020-05889-1.

[22] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[23] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, “Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem,” *ACM Int. Conf. Proceeding Ser.*, pp. 412–419, 2021, doi: 10.1145/3453892.3461323.

[24] B. Santoso, H. Wijayanto, K. A. Notodiputro, and B. Sartono, “Synthetic over sampling methods for handling class imbalanced problems: A review,” *IOP Conf Ser Earth Env. Sci*, vol. 58, p. 012031, 2017, doi: 10.1088/1755-1315/58/1/012031.

[25] R. Zhu, Y. Guo, and J. H. Xue, “Adjusting the imbalance ratio by the dimensionality of imbalanced data,” *Pattern Recognit. Lett.*, vol. 133, pp. 217–223, 2020, doi: 10.1016/j.patrec.2020.03.004.

[26] A. N. A. Aldania, A. M. Soleh, Notodiputro, and K. Anwar, “A comparative study of CatBoost and Double Random Forest for multi-class classification,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 129–137, 2023, doi: 10.29207/resti.v7i1.4766.

[27] J. Wijaya, A. M. Soleh, and A. Rizki, “Penanganan Data Tidak Seimbang pada Pemodelan Rotation Forest Keberhasilan Studi Mahasiswa Program Magister IPB,” *Xplore J. Stat.*, vol. 2, no. 2, pp. 32–

40, 2018, doi: 10.29244/xplore.v2i2.99.

[28] D. Elreedy and A. F. Atiya, “A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance,” *Inf. Sci. (Ny).*, vol. 505, pp. 32–64, 2019, doi: 10.1016/j.ins.2019.07.070.