25_JURNAL_2023_A Bootstrap-Aggregating in Random Forest Model for Classification

By Yulia Resti

WORD COUNT	5542	TIME SUBMITTED	27-APR-2023 12:39PM
		PAPER ID	99142082

Science and Technology Indonesia
e-ISSN:2580-4391 p-ISSN:2580-4405
Vol. 8, No. 2, April 2023

Research Paper

Check for updates

A Bootstrap-Aggregating in Random Forest Model for Classification of Corn Plant Diseases and Pests

Yulia Resti[1*], Chandra Irsan[2], Jeremy Firdaus Latif[1], Irsyadi Yani[3], Novi Rustiana Dewi[1]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Sriwijaya, Indralaya, 30662, Indonesia*
[2]*Study Program of Plant Protection, Department of Plant Pest and Disease, Faculty of Agriculture, Universitas Sriwijaya, Indralaya, 30622, Indonesia*
[3]*Department of Mechanical Engineering, Faculty of Engineering, Universitas Sriwijaya, Indralaya, 30622, Indonesia*
***Corresponding author**: yulia_resti@mipa.unsri.ac.id

Abstract
Control of diseases and pests of maize plants is a significant challenge to ensure global food security, self-sufficiency, and sustainable agriculture. Classification or early detection of diseases and pests of corn plants is intended to assist the control process. Random forest is a classification model in tree-based statistical learning in making decisions. This approach is an ensemble method that generates many decision trees and makes classification decisions based on the majority of trees selecting the same class. However, tree-based methods are often unstable when small changes or disturbances exist in the learning data. Such instability can produce significant variances and affect model performance. This study classifies diseases and pests of the corn plant using a random forest method based on bootstrap-aggregating. It fits multiple models of a single random forest, then combines the predictions from all models and determines the final result using majority voting. The results showed that the bootstrap-aggregating could improve the classification of diseases and pests of maize using a random forest if the number of trees is optimal.

Keywords
Bootstrap-Aggregating, Classification, Corn Plant Disease and Pest, Decision Tree, Random Forest

Received: 4 January 2023, Accepted: 2 April 2023
https://doi.org/10.26554/sti.2023.8.2.288-297

1. INTRODUCTION

Corn is a leading commodity in the food crop sector in many countries. In addition, corn is also used as animal feed and industrial raw materials. In Indonesia, controlling diseases and pests of maize plants is a major challenge to ensure global food security, self-sufficiency, and sustainable agriculture. Classification or early detection of diseases and pests of corn plants is intended to assist the control process. In recent years, several machine-learning models have been developed for image classification (Jafarzadeh et al., 2021). Especially for classifying food crop diseases (Resti et al., 2022b; Resti et al., 2022a; Liu and Wang, 2021; Xian and Ngadiran, 2021; Syarief and Setiawan, 2020; Kasinathan et al., 2021; Mengistu, 2018). This technology is due to the low-cost use of digital images for classification (Ngugi et al., 2021). Using red, green, and blue (RGB) color features from digital images of plant diseases, especially the corn plant, best performs for most statistical machine-learning methods or models (Kusumo et al., 2018).

The random forest is a tree-based classification model for making decisions, which are conducted based on selecting the best predictor variable that produces maximum gain or minimum entropy (Gareth et al., 2013; Hastie et al., 2009). This tree-structured decision consists of a root node, internal nodes, leaf nodes, and branches. The root node is the initial, and the nodes that follow it, whether they contain branches or not, are referred to as the internal and leaf nodes, respectively (Witten and Frank, 2002). A random forest is an ensemble learning model that, during training, generates a large number of decision trees and makes classification decisions based on the majority of trees selecting the same class. Each tree is built from a unique bootstrap sample of the data (Conn et al., 2019). This model also uses recursive binary splitting to reach the final node in the tree structure it forms. Jin et al. (2020) Some cases study has implemented this method with a satisfaction performance (Singh et al., 2021; Prasojo and Haryatmi, 2021; Panigrahi et al., 2020). Although this model is intended to improve the performance of the decision tree model (Zhu et al., 2021), in some cases, the performance of this model is not satisfactory (Sahith et al., 2021), especially in digital image-based classification tasks, (Syarief and Setiawan, 2020; Kusumo et al., 2018).

However, the tree-based classification model is a statistical machine-learning method that is often unstable when there

are small changes or disturbances in the learning data. Such instability can produce large variances and affect model performance (Salman et al., 2021). Bootstrap-aggregating, or the abbreviation bagging, is a sampling-based approach using the bootstrap method and validation measurements using the aggregate method (Gareth et al., 2013). Several studies have shown that the application of bagging can improve the performance of classification models (Alelyani, 2021; Salman et al., 2021; Saifudin et al., 2020). This study aims to classify diseases and pests of corn plants using the bootstrap-aggregating-based random forest method. The data in this study formed the extracted digital image data of diseases and pests of corn plants into the RGB color space model and resized it into the same pixel structure of the matrix data.

2. EXPERIMENTAL SECTION

2.1 The Proposed Method

2.1.1 Research Data

The data in this study are digital images of corn plants' diseases and pests extracted into an RGB color space model with a 64 x 64 resolution using the OpenCV library in the Python programming language. The digital images were captured at the corn plantations of Tanjung Seteko, Tanjung Baru, and Tanjung Putus, in the Ogan Ilir Regency of South Sumatra, Indonesia, during September and October 2021. This study only uses high-quality digital image data samples (not blurry). Manual sorting obtained 4616 digital images as research samples. The data contains seven classes consisting of one class of nonpathogen, three classes of disease (leaf rust, downy mildew, and leaf blight), and three classes of pest (*Locusta*, *Spodoptera Frugiperda*, and *Heliotis Armigera*).

$$R_{64x64} = \begin{bmatrix} 55 & 49 & 117 & \dots & 45 \\ 64 & 113 & 129 & \dots & 51 \\ 80 & 92 & 113 & \dots & 73 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 67 & 82 & 110 & \dots & 58 \end{bmatrix}$$

$$G_{64x64} = \begin{bmatrix} 54 & 48 & 112 & \dots & 44 \\ 54 & 114 & 123 & \dots & 55 \\ 80 & 89 & 109 & \dots & 71 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 170 & 180 & 177 & \dots & 61 \end{bmatrix}$$

$$B_{64x64} = \begin{bmatrix} 60 & 53 & 118 & \dots & 50 \\ 65 & 114 & 127 & \dots & 56 \\ 80 & 89 & 110 & \dots & 75 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 174 & 186 & 181 & \dots & 68 \end{bmatrix}$$

The initial symptoms that occur when corn plants are attacked by leaf rust disease (LRD) are boils on both leaf surfaces,

(a) (b) (c)

Figure 1. Class of Corn Plant Disease (a) LRD (b) DWD (c) LBD

(a) (b) (c)

Figure 2. Class Composition of Corn Plant Disease and Pest (a) LP (b) SFP (c) HAP

and the leaf color becomes reddish brown. It can change to brownish-black, and the leaf texture becomes dry after developing teliospores (Mirsam et al., 2021). The chlorotic appearance on the leaves will see corn plants attacked by downy mildew disease (DWD). The plants become stunted, and the leaves' veins and color become pale. So that this disease causes a lack of production from corn plants because it can kill the growth process in corn plants, which slowly attacks the affected plants and will dry up, and eventually, the plants will die (Purwanto et al., 2017). Another disease is Leaf blight disease (LBD). This disease is caused by a fungal pathogen (Girsang et al., 2020). Symptoms caused by leaf blight are the appearance of brown spots on the surface of the leaves. Corn plants that are 20 days old have an intensity of 25% being attacked by this disease. At the beginning of the rainy season, infected plants will potentially cause high-intensity damage. These conditions will increase the sporulation of the fungus or spores in the air, which are sufficiently available so that the infection intensity of this disease is very high compared to the dry season (Sari, 2018). The three examples of this corn disease are presented in Figure 1.

Locusta pests will start attacking corn plants when the plants age 20 days, after the corn has already shoots and leaves. The pests eat the leaves of the corn plant. The intensity of locust attacks reached 70% at the age of corn 45 days after planting (Dewantara et al., 2020). Attacks by Spodoptera Frugiperda pest (SFP) can fail the formation of young shoots/leaves of plants because these pests attack the growing points of plants (Lubis et al., 2020). The other pest, Heliotis Armigera pest (HAP), eats cob hair, cob tips, and cobs of corn so that the fruit becomes damaged and unfit for sale in the mark (Megasari and Nuriyadi, 2019). Figure 2 presents examples of the digital image of three classes of corn pests.

© 2023 The Authors.

The following is an example of a structure of a 64 x 64 matrix data for each channel R, G, and B from one of the DWD digital images. The pixel value of each component R, G, and B is the average value of all entries in the matrix (Resti et al., 2022b).

2.1.2 Methods

The steps in the research consist of the following:

1. Extract data of digital images of corn plants' diseases and pests extracted into an RGB color space model with a 64 x 64 resolution using the OpenCV library in the Python programming language. Thus, the discretization of data in pixel units using Equation (1) for $X_d°$ be the d-th predictor variable which represents the color pixel values in the interval scale, (cX_d) be the number of the class interval and Range (X_d)=(max($X_d°$)-min($X_d°$)/$c(X_d°)$ (Resti et al., 2022a).

$$X_d = \text{Range}(X_d) + (X_d^\circ) \quad (1)$$

2. Splitting the data into training data and test data into a ratio of 80:20. Thus, create multiple samples of the training using bootstrap sampling (Salman et al., 2021; Kuhn and Johnson, 2013), as illustrated in Figure 3.

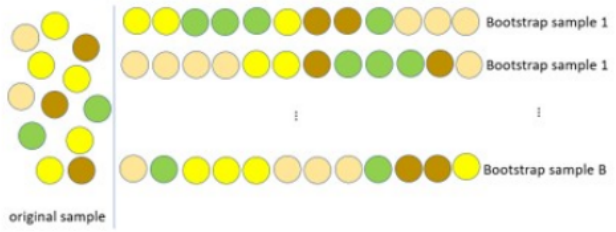

Figure 3. Illustration of Bootstrap Sampling

3. For a random forest, model the data using the following steps in each created sample (Ramasubramanian and Singh, 2016; Kuhn and Johnson, 2013).
   (a) Choose r- predictor variables randomly.
   (b) Determine the entropy of variables Y and $X_d°$, where d=1,...,r for the m-th category using Equations (2) - (3).

$$H(S(Y)) = -\sum_{j=1}^{k(Y)} p_j \log_2 p_j \quad (2)$$

For Y is the target variable that represents the types of disease and pest of corn, p_j the prior probability in the j-th type of Y, $k(Y)$ is the number of types in Y, and $S(Y)$ is the number of observations in all types Y.

$$H(X_d^m) = -\sum_{m=1}^{k(X_d)} p_m \log_2 p_m \quad (3)$$

For $S(X_d^m)$ is the number of observations in the m-th category of the X_d for all types, and p_m is the prior probability in the m-th category of X_d. The prior probability in the j-th type of Y, (p_j), is obtained as a ratio between the number of observations in the j-th type and the total observations in Y. At the same time, p_m is obtained as a ratio between the number of observations in the m-th category of X_d of the j-th type and total observations in the m-th category of X_d.

(c) Determine the gain of (Y,X_d) using Equation (4).

$$G(Y, X_d) = H(S(Y)) - \sum_{m=1}^{k(X_d)} \frac{S(X_d^m)}{S(Y)} H(X_d^m) \quad (4)$$

The predictor variable with the highest gain value is used as a node. The first node is called the root node.

(d) Repeat steps (a) – (c) until the leaf node is formed.
(e) Make classification decisions from each tree model formed using the if-then rule.
(f) Use the most votes to make the final decision on a random forest from all the trees formed from each sample, as shown in Figure 4.

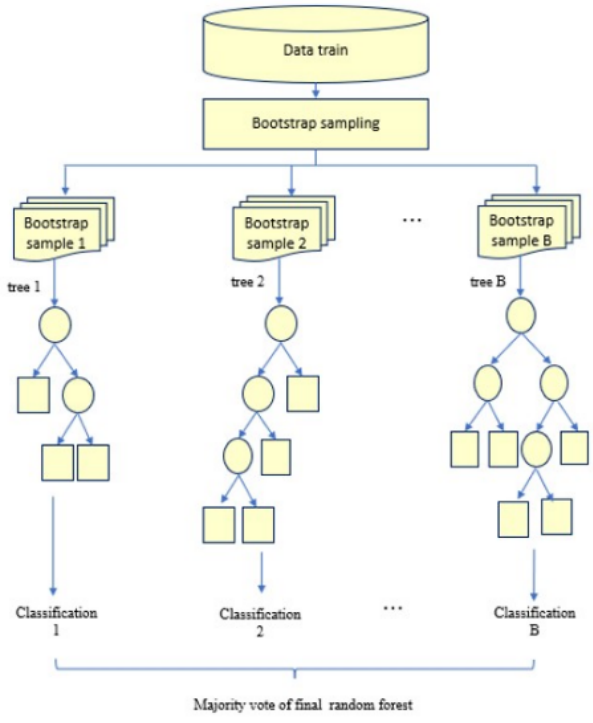

Figure 4. Illustration of Random Forest Process

(g) Create a confusion matrix table for each class (Dinesh and Dash, 2016).

© 2023 The Authors.

(h) Determining model performance for six classes based on average accuracy, precision macro, recall macro, and F1-score macro values (Sokolova and Lapalme, 2009), respectively, use Equations (5) – (8).

$$\text{Average accuracy} = \frac{\sum_{j=1}^{7} \frac{TP_j+TN_j}{TP_j+FP_j+FN_j+TN_j}}{7} \quad (5)$$

$$\text{Precision macro} = \frac{\sum_{j=1}^{7} \frac{TP_j}{TP_j+FP_j}}{7} \quad (6)$$

$$\text{Recall macro} = \frac{\sum_{j=1}^{7} \frac{TP_j}{TP_j+FN_j}}{7} \quad (7)$$

$$\text{F}_1 \text{ Score macro} = \frac{2\text{Precision macro(Recall macro)}}{\text{(Precision macro+Recall)}} \quad (8)$$

For the j-th class, let true positives (TP_j) and true negatives (TN_j) be proper classifications. False positives (FP_j) occur when an outcome is incorrectly predicted as the j-th class (or positive) when it is, in fact, not the j-th class (negative). A false negative (FN_j) occurs when a result is incorrectly predicted as not the j-th class (negative) when it is the j-th class (positive).

4. Repeat steps (2) and 3(a) – 3(f) for multiple random forests for bootstrap aggregating. Then make a final classification decision using majority voting or known as aggregating Salman et al. (2021) as described in Figure 5.
5. Next, create a confusion matrix table for each class and determine the performance model formulated in Steps 3(g) and 3(h).
6. Compare the performance of the proposed models, random forest without and with bootstrap-aggregating.

3. RESULTS AND DISCUSSION

In this study, the extraction of RGB color features from samples of corn plant pests and diseases using python was resized to 64 x 64 pixels. One sample example is presented in Figure 6.

A statistical summary of each predictor variable R, G, and B from a digital image resized with the resolution of 64 x 64 pixels is presented in Figure 7, where each of these features is denoted by X_1, X_2, and X_3.

The results of discretization into five class intervals (category) using Equation (1) are presented in Table 1. This discretization into five categories gives satisfactory performance (Resti et al., 2022b; Resti et al., 2022a).

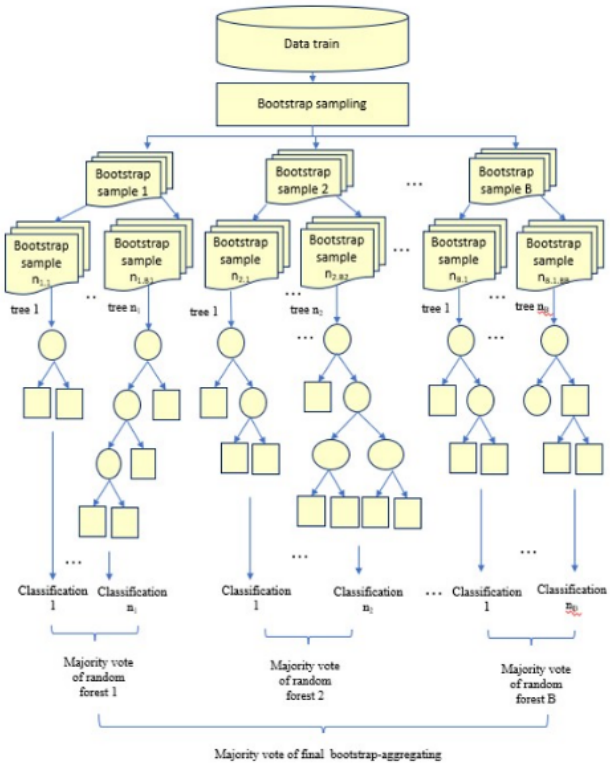

Figure 5. Illustration of Bootstrap-Aggregating of Random Forest Process

The prior probability of each class of corn disease and pest of the first tree in Figure 8 helps calculate the entropy, $H(S(Y))$, at the root node. SFP and DWD classes have the largest and smallest probability compared to other classes, respectively.

For the first bootstrap sample and the first tree of random forest, the entropy, $H(S(Y))$, and the gain, $G(Y,X_d)$, for each root dan internal node are presented in Table 2 dan Table 3, respectively. The first iteration of the random forest has the same value as the decision tree because this value is not affected by the predictor variable and has the same prior probability at the beginning of the calculation.

Variable X_1 has the highest gain, so this variable is the root node for tree formation. Next, determine the internal node for the first branch (category 1). The determination of the internal node is analogous to the root node, which is calculated based on the largest gain of the two predictor variables determined randomly. The prior probability for the first category of the two predictor variables chosen randomly other than the root node is presented in Figure 9, while the entropy of m-th category $H(X_d{}^m)$, entropy $H(S(Y))$, and gain $G(Y,X_d)$, are presented in Table 3.

In this first category, the remaining observations are only in the LBD, SFP, and HAP classes, with the largest prior prob-

© 2023 The Authors.

Table 1. Discretization of Predictor Variable

Category	Discretization Value Range		
	X_1	X_2	X_3
1	76.18 – 92.94	80.79 – 97.75	32.83 – 60.33
2	92.95 – 109.72	97.76 - 114.73	60.34 – 87.83
3	109.73 - 126.49	114.74 -131.70	87.84 – 115.34
4	126.50 – 143.27	131.71 - 148.67	115.35 – 142.84
5	143.28 – 160.05	148.68 – 165.65	142.85 – 170.36

Table 2. Entropy and Gain for Root Node of Random Forest

Variable	The Entropy of m-th Category $H(X_d{}^m)$					Entropy $H(S(Y))$	Gain $G(Y,X_d)$
	1	2	3	4	5		
X_1	0.41	1.66	2.01	2.21	0.42	2.35	0.49
X_2	0.76	1.35	2.13	2.04	2.39		0.40

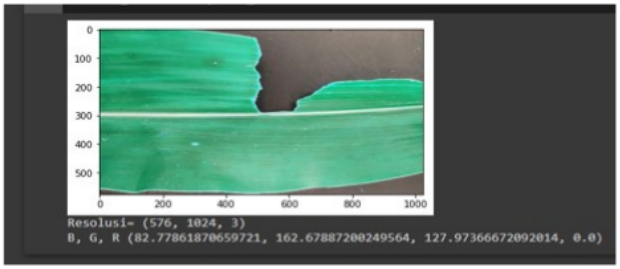

Figure 6. Sample of the Digital Image in 64 x 64 Pixels

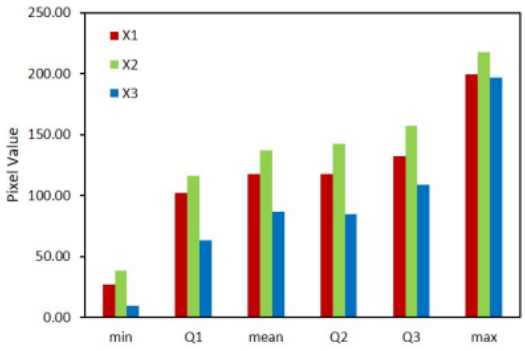

Figure 7. Summary Statistic of Predictor Variable

ability belonging to the LBD class.

Because the highest gain is owned by variable X_2, this variable is an internal node of the branching category 1 of the root node. Furthermore, the set intersection between the root node and the first internal node (Table 4) shows that the fourth and fifth categories do not have a set, so they are not included in the subsequent decision tree. Then because the remaining variable is only X_3, X_3 automatically becomes the next internal node (the second internal node) for the three rows of branches, the first, the second, and the third categories.

Table 5, which presents the intersection of the root node set, the first internal node, and the second internal node, shows that the terminal node from the first category branches to the first internal node (X_2) and the first category branching to the second internal node (X_3) with the first category branching, the SFP class. In contrast, the second and third categories are classified into the LP class. The decision of the first tree in the random forest can be seen in Figure 10.

The number of trees selected in a random forest model can affect its performance. Figure 11 shows an experiment on the number of trees in a random forest to obtain the highest model performance. Only the random forest model with two trees has lower accuracy than the model with the other number of trees which tends to be constant, starting from the number of trees is three.

The result of the classification of diseases and pests of corn plants, which constitute the majority voting of the random forest model with the highest performance, is given in a confusion matrix (Table 6). Only the DWD and HAP classes have poorly classified observations in the model with three trees.

Furthermore, the bootstrap-aggregating classification process is analogous to the random forest, but the bootstrap process takes place in two stages, as well as majority voting (Figure 5). Figure 12 presents the decision of the first random forest's first tree in bootstrap-aggregating.

Bootstrap-aggregating of random forest fits multiple models of a single random forest, then combines the predictions from all models and determines the final result using majority voting. Figure 13 shows an experiment on the number of trees and forests in bootstrap-aggregating of random forests. The highest model performance is owned by the model with six trees and six forests from experiments with 2 to 50 trees and

© 2023 The Authors.

Table 3. Entropy and Gain for Internal Node of Random Forest

Variable	The Entropy of m-th Category $H(X_d{}^m)$ 1	2	3	4	5	Entropy $H(S(Y))$	Gain $G(Y,X_d)$
X_2	0.29	0.00	0.00	0.00	0.00	0.41	0.24
X_3	0.00	0.00	0.24	0.00	0.00		0.22

Table 4. The Sets Intersection of the First Root and the First Internal Node

The Sets Intersection	Class of Corn Disease and Pest					
	LRD	DWD	LRD	LP	SFP	HAP
X_1 (1^{st} category) $\cap$ X_2 (1^{st} category)	0	0	0	18	1	0
X_1 (1^{st} category) $\cap$ X_2 (2^{nd} category)	0	0	0	11	0	0
X_1 (1^{st} category) $\cap$ X_2 (3^{rd} category)	0	1	0	0	0	0
X_1 (1^{st} category) $\cap$ X_2 (4^{th} category)	0	0	0	0	0	0
X_1 (1^{st} category) $\cap$ X_2 (5^{th} category)	0	0	0	0	0	0

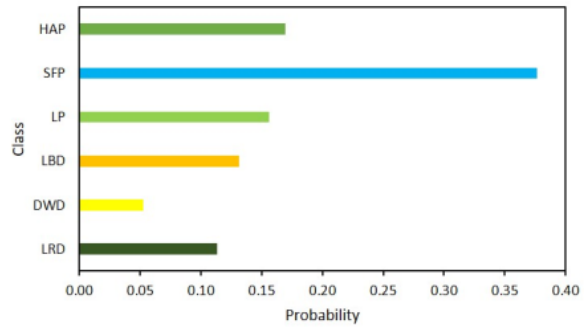

Figure 8. Prior Probability of the First Tree in a Random Forest

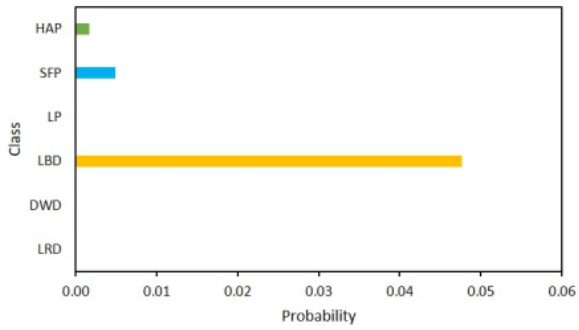

Figure 9. Prior Probability for the First Internal Node

forests, respectively.

The results of the classification of diseases and pests of maize plants using the bootstrap-aggregating of random forest with the highest performance are presented in Table 7. In the model with six forests, only one observation in the DWD class was classified in the HAP class.

The last is the performance of the proposed models that implement a random forest without and with bootstrap-aggregating in Table 8.

With the bootstrap-aggregating technique, random forest performance metrics that have increased from largest to smallest are precision macro, F1-score, recall macro and average accuracy. The improvement of the four performance metrics shows that even though the random forest is an ensemble method, its performance can still be improved using the bootstrap-aggregating technique. The number of trees in each process affects its performance; therefore, it is necessary to know the optimal number of trees for the proposed model.

The performance of the random forest model without bootstrap-aggregating tends to be constantly starting from three trees, but the number of trees smaller than three results in worse performance. At the same time, a random forest with bootstrap-aggregating tends to start from the number of trees sixteen constantly. Therefore, the number of trees smaller than sixteen produces a slightly fluctuating accuracy between 98.91%-99.78%.

Comparison of the performance of the two models proposed in this work with similar studies shows a significant increase in performance, especially the three metrics precision, recall, and F1-score (Figure 14). For example, multinomial naïve Bayes (MNB) and K-Nearest Neighbor (KNN) proposed to classify diseases and pests of maize plants Resti et al. (2022b) have demonstrated satisfactory performance as indicated by metric values of more than 85% (Aronoff, 1985), especially KNN (Mishra et al., 2016). Likewise, decision trees (DT) and fuzzy decision trees (FDT) where FDT has a significant increase compared to DT in classifying diseases and pests of

© 2023 The Authors.

Table 5. The sets Intersection of the First Root, the first Internal, and the Second Internal Node

The Sets Intersection	Class of Corn Disease and Pest					
	LRD	DWD	LBD	LP	SFP	HAP
X_1 (1^{st} category) ∩ X_2 (1^{st} category) ∩ X_3 (1^{st} category)	0	0	0	0	1	0
X_1 (1^{st} category) ∩ X_2 (1^{st} category) ∩ X_3 (2^{nd} category)	0	0	0	10	0	0
X_1 (1^{st} category) ∩ X_2 (1^{st} category) ∩ X_3 (3^{rd} category)	0	1	0	8	0	0

Table 6. Multiclass Confusion Matrix of the Best Random Forest

		Classification					
		LP	SFP	LRD	DWD	LBD	HAP
Actual	LP	20	0	0	0	0	0
	SFP	0	63	0	0	0	0
	LRD	0	0	18	0	0	0
	DWD	0	1	0	10	0	0
	LBD	0	0	0	0	21	0
	HAP	0	0	1	0	0	19

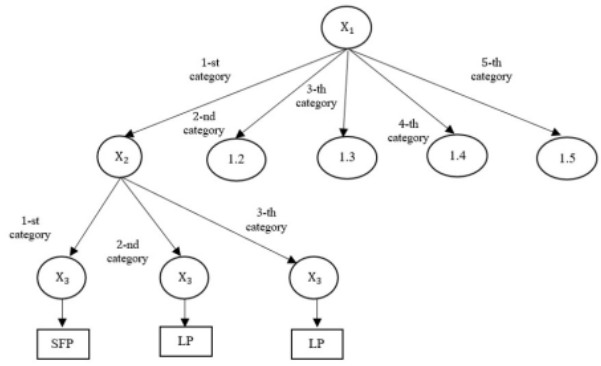

Figure 10. The Decision in the First Tree of Random Forest

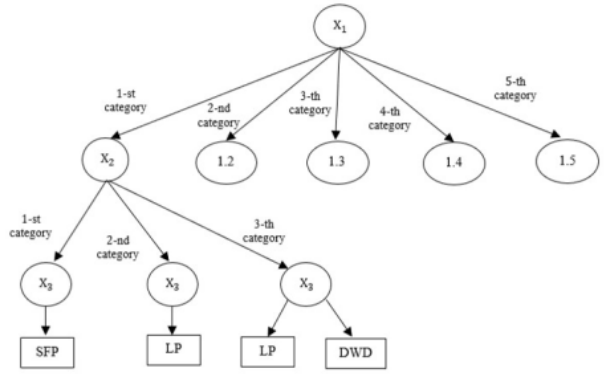

Figure 12. The Decision in the First Tree of Random Forest with Bootstrap-Aggregating

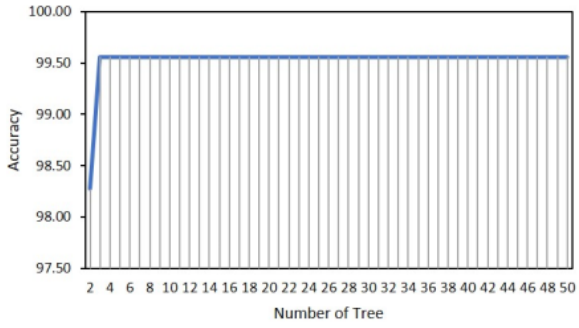

Figure 11. Random Forest Accuracy based on the Number of Trees

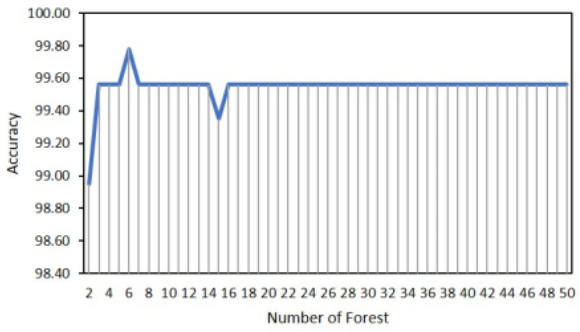

Figure 13. Random Forest with Bootstrap-Aggregating Accuracy Based on the Number of Trees and Forest

© 2023 The Authors.

Table 7. Multiclass Confusion Matrix of the Best Random Forest with Bootstrap Aggregating

		Classification					
		LRD	DWD	LBD	LP	SFP	HAP
Actual	LRD	0	0	0	20	0	0
	DWD	0	0	0	0	63	0
	LBD	18	0	0	0	0	0
	LP	0	10	0	0	0	1
	SFP	0	0	21	0	0	0
	HAP	0	0	0	0	0	20

Table 8. Model Performance of Random Forest Without and with Bagging

Method	Performance			
	Average Accuracy	Precision Macro	Recall Macro	F1-score
RF without Bagging	99.56	97.69	97.69	97.69
RF with Bagging	99.78	99.21	98.48	98.84
Enhancement	0.22	1.52	0.79	1.15

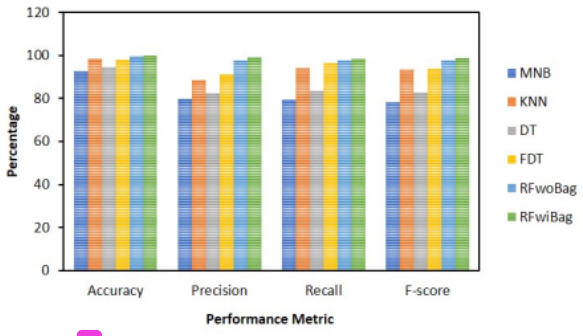

Figure 14. Comparison of the Performance of the Proposed Model with Other Studies

maize plants (Resti et al., 2022a). The increase in performance metrics obtained was 3.23% (accuracy), 8.59% (precision), 12.88% (recall), and 10.68% (F-score). However, the model performance obtained by the model proposed in this research is higher.

The main contribution of this work is that we have shown a significant improvement in the performance model of the random forest by implementing bootstrap-aggregating resampling. This robust performance increase was obtained by exploring the optimal number of trees from the random forest model without bootstrap-aggregating and with bootstrap-aggregating. The optimal number of trees can be different for each model, and this number can affect its performance.

4. CONCLUSION

Classification or early detection of diseases and pests of maize plants is intended to assist in controlling diseases and pests of maize plants. This control ensures global food security, self-sufficiency, and sustainable agriculture. Random forest is a tree-based classification model in building decisions. However, tree-based methods are often unstable when small changes or disturbances in the learning data can affect the model's performance. The bootstrap-aggregating has been implemented in the random forest to classify diseases and pests of corn plants. Even though the random forest is an ensemble method that also applies the bootstrap process to each branch, this work shows that the performance of the random forest model can still be improved using bootstrap-aggregating. Improved model performance is obtained through experiments that explore the number of trees in the related model. The number of trees that are not optimal can cause performance improvements not to be achieved. The random forest model with the optimal number of trees is compared to the random forest model, which implements bootstrap-aggregating with the optimal number of trees. The experimental results show that bootstrap-aggregating implementation in the random forest model increases when the number of trees is optimal.

5. ACKNOWLEDGMENT

This work was supported by DIPA of Universitas Sriwijaya 2022 Public Service Agency, SP DIPA-023.17.2.677515/2022, on December 13, 2021. On April 28, 2022, the Rector's Decree 0118.115/ UN9/ SB3.LP2M. PT/2022, May 17, was issued.

REFERENCES

Alelyani, S. (2021). Stable Bagging Feature Selection On Medical Data. *Journal of Big Data*, **8**(1); 1–18

Aronoff, S. (1985). The Minimum Accuracy Value as An Index

© 2023 The Authors.

Of Classification Accuracy. *Photogrammetric Engineering and Remote Sensing*, **51**(1); 99–111

Conn, D., T. Ngun, G. Li, and C. M. Ramirez (2019). Fuzzy Forests: Extending Random Forest Feature Selection for Correlated, High-dimensional Data. *Journal of Statistical Software*, **91**; 1–25

Dewantara, A. W., D. Ratna, and S. J. Santosa (2020). Kajian Macam Pupuk Hayati Terhadap Intensitas Kerusakan Hama Belalang Pada Tanaman Jagung Hitam. *Innofarm: Jurnal Inovasi Pertanian*, **22**(1); 29–35 (In Indonesia)

Dinesh, S. and T. Dash (2016). Reliable Evaluation of Neural Network for Multiclass Classification of Real-world Data. *arXiv preprint arXiv:1612.00671*, **11**(2); 30–35

Gareth, J., W. Daniela, H. Trevor, and T. Robert (2013). An Introduction to Statistical Learning: with Applications In R. *In Paper History of Documents*, **3**(5); 34–38

Girsang, W., J. Purba, and S. Daulay (2020). Uji Aplikasi Agens Hayati Tribac Mengendalikan Pathogen Hawar Daun (Helminthosporium SP.) Tanaman Jagung (Zea Mays L.). *Jurnal Ilmiah Pertanian*, **17**(1); 51–59 (In Indonesia)

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer

Jafarzadeh, H., M. Mahdianpari, E. Gill, F. Mohammadimanesh, and S. Homayouni (2021). Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and Polsar Data: A Comparative Evaluation. *Remote Sensing*, **13**(21); 4405

Jin, Z., J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **3**(4); 503–515

Kasinathan, T., D. Singaraju, and S. R. Uyyala (2021). Insect Classification and Detection In Field Crops Using Modern Machine Learning Techniques. *Information Processing in Agriculture*, **8**(3); 446–457

Kuhn, M. and K. Johnson (2013). *Applied predictive modeling*, volume 26. Springer

Kusumo, B. S., A. Heryana, O. Mahendra, and H. F. Pardede (2018). Machine Learning Based for Automatic Detection of Corn-plant Diseases Using Image Processing. *International Conference on Computer, Control, Informatics and Its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018*, **4**(5); 93–97

Liu, J. and X. Wang (2021). Plant Diseases and Pests Detection Based On Deep Learning: A Review. *Plant Methods*, **17**; 1–18

Lubis, A. A. N., R. Anwar, B. P. Soekarno, B. Istiaji, S. Dewi, and D. Herawati (2020). Serangan Ulat Grayak Jagung (Spodoptera frugiperda) pada Tanaman Jagung Di Desa Petir, Kecamatan Daramaga, Kabupatem Bogor dan Potensi Pengendaliannya Menggunakan Metarizhium Rileyi. *Jurnal Pusat Inovasi Masyarakat (PIM)*, **2**(6); 931–939 (In Indonesia)

Megasari, R. and M. Nuriyadi (2019). The Inventory of Pests and Diseases of Corn Plants (Zea mays L.) and its Control Inventarisasi Hama dan Penyakit Tanaman Jagung (Zea mays L.). *Musamus Journal of Agrotechnology Research*, **2**(1); 1–12

Mengistu, M. S. G. . M. D., A. D. (2018). An Automatic Coffee Plant Diseases Identification Using Hybrid Approaches of Image Processing and Decision Tree. *Indonesian Journal of Electrical Engineering and Computer Science*, **9**(3); 806–811

Mirsam, H., S. Suriani, N. Djaenuddin, A. T. Makkulawu, and F. Abdullah (2021). Evaluasi Ketahanan Genotipe Jagung Hibrida terhadap Penyakit Hawar Daun Maydis dan Karat Daun. *Seminar Nasional Lahan Suboptimal*, **9**(2021); 305–313 (In Indonesia)

Mishra, S., O. A. Vanli, F. W. Huffer, and S. Jung (2016). Regularized Discriminant Analysis for Multi-sensor Decision Fusion and Damage Detection With Lamb Waves. *SPIE*, **9803**; 728–741

Ngugi, L. C., M. Abelwahab, and M. Abo Zahhad (2021). Recent Advances in Image Processing Techniques for Automated Leaf Pest and Disease Recognition–a Review. *Information Processing in Agriculture*, **8**(1); 27–51

Panigrahi, K. P., H. Das, A. K. Sahoo, and S. C. Moharana (2020). Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. *Proceedings of ICCAN*, **4**(5); 659–669

Prasojo, B. and E. Haryatmi (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi dan Sistem Informasi*, **7**(2); 79–89 (In Indonesia)

Purwanto, D. S., H. Nirwanto, and S. Wiyatiningsih (2017). Model Epidemi Penyakit Tanaman: Hubungan Faktor Lingkungan Terhadap Laju Infeksi dan Pola Sebaran Penyakit Bulai (peronosclerospora maydis) pada Tanaman Jagung Di Kabupaten Jombang. *Berkala Ilmiah Agroteknologi-PLUMULA*, **5**(2); 138–152 (In Indonesia)

Ramasubramanian, K. and A. Singh (2016). Machine Learning Using R: With Time Series and Industry-Based Use Cases in R. *Apress*, **2**(321); 42–48

Resti, Y., C. Irsan, M. Amini, I. Yani, R. Passarella, and D. Zayanti (2022a). Performance Improvement of Decision Tree Model using Fuzzy Membership Function for Classification of Corn Plant Diseases and Pests. *Science and Technology Indonesia*, **7**(3); 284–290

Resti, Y., C. Irsan, M. T. Putri, I. Yani, A. Ansyori, and B. Suprihatin (2022b). Identification of Corn Plant Diseases and Pests Based on Digital Images Using Multinomial Naïve Bayes and K-Nearest Neighbor. *Science and Technology Indonesia*, **7**(1); 29–35

Sahith, R., P. V. P. Reddy, and S. Nimmala (2021). Decision Tree-based Machine Learning Algorithms to Classify Rice Plant Diseases: A Recent Study. *Advanced Aspects of Engineering Research*, **16**; 52–59

Saifudin, A., U. Nabillah, and T. Desyani (2020). Bagging Technique to Reduce Misclassification in Coronary Heart

© 2023 The Authors.

Disease Prediction Based on Random Forest. *Journal of Physics: Conference Series*, **1477**(3); 032009

Salman, R., A. Alzaatreh, H. Sulieman, and S. Faisal (2021). A Bootstrap Framework for Aggregating Within and Between Feature Selection Methods. *Entropy*, **23**(2); 200

Sari, D. M. (2018). Bakteri Antagonis Dari Sumber Air Panas Dan Uji Kemampuan Antagonisnya Terhadap Helminthosporium Turcicum (Pass.) Penyebab Hawar Daun Pada Tanaman Jagung (Zea Mays L.). *In Paper History of Documents*, **4**(5); 54–60 (In Indonesia)

Singh, A., B. Chourasia, N. Raghuwanshi, and K. Raju (2021). BPSO Based Feature Selection for Rice Plant Leaf Disease Detection with Random Forest Classifier. *International Journal of Engineering Trends and Technology*, **69**(4); 34–43

Sokolova, M. and G. Lapalme (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management*, **45**(4); 427–437

Syarief, M. and W. Setiawan (2020). Convolutional Neural Network for Maize Leaf Disease Image Classification. *Telkomnika (Telecommunication Computing Electronics and Control)*, **18**(3); 1376–1381

Witten, I. H. and E. Frank (2002). Data Mining: Practical Machine Learning Tools and Techniques With Java Implementations. *Acm Sigmod Record*, **31**(1); 76–77

Xian, T. S. and R. Ngadiran (2021). Plant Diseases Classification Using Machine Learning. *Conference Series*, **1962**(1); 012024

Zhu, J., S. Huang, Y. Shi, K. Wu, and Y. Wang (2021). A Method of Random Forest Classification based on Fuzzy Comprehensive Evaluation. *International Conference on Dependable System and Their Applications*, **4**(5); 178–183

© 2023 The Authors.

25_JURNAL_2023_A Bootstrap-Aggregating in Random Forest Model for Classification

ORIGINALITY REPORT

11%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.sciencetechindonesia.com Internet	146 words — 3%
2	www.mdpi.com Internet	139 words — 3%
3	Yulia Resti, Chandra Irsan, Adinda Neardiaty, Choirunnisa Annabila, Irsyadi Yani. "Fuzzy Discretization on the Multinomial Naïve Bayes Method for Modeling Multiclass Classification of Corn Plant Diseases and Pests", Mathematics, 2023 Crossref	63 words — 1%
4	www.researchgate.net Internet	25 words — 1%
5	Kaixiang Zhang, Xueling Wu, Ruiqing Niu, Ke Yang, Lingran Zhao. "The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China", Environmental Earth Sciences, 2017 Crossref	16 words — < 1%
6	pure.southwales.ac.uk Internet	16 words — < 1%
7	link.springer.com	

Internet

15 words — < 1%

8 "ACIT 2021 Conference Proceedings", 2021 22nd International Arab Conference on Information Technology (ACIT), 2021
Crossref

12 words — < 1%

9 Ramírez Aldana Ricardo. "Restricted or coloured graphical log-linear models", TESIUNAM, 2010
Publications

10 words — < 1%

10 Zhi Jiang, Yong Zhang, Jun Wang. "A multi-surrogate-assisted dual-layer ensemble feature selection algorithm", Applied Soft Computing, 2021
Crossref

10 words — < 1%

11 www.frontiersin.org
Internet

9 words — < 1%

12 Michaël Zamo, Liliane Bel, Olivier Mestre, Joël Stein. "Improved Gridded Wind Speed Forecasts by Statistical Postprocessing of Numerical Models with Block Regression", Weather and Forecasting, 2016
Crossref

8 words — < 1%

13 Okoth, Peter F.. "A Hierachical Method for Soil Erosion Assessment and Spatial Risk Modelling: A Case Study of Kiambu District in Kenya.", Wageningen University and Research, 2021
ProQuest

8 words — < 1%

14 T Desyani, Y Kasmayanti, A Saifudin, Yulianti. "Bagging Techniques to Reduce Misclassification of Breast Cancer Prediction Base on Gradient Boosted Trees (GBT) Algorithm", Journal of Physics: Conference Series, 2020
Crossref

8 words — < 1%

15	jame.um.ac.ir Internet	8 words — < 1%
16	"Proceedings of Data Analytics and Management", Springer Science and Business Media LLC, 2022 Crossref	6 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF