Science and Technology Indonesia
e-ISSN:2580-4391 p-ISSN:2580-4405
Vol. 7, No. 3, July 2022

Research Paper

Performance Improvement of Decision Tree Model using Fuzzy Membership Function for Classification of Corn Plant Diseases and Pests

Yulia Resti[1]*, Chandra Irsan[2], Muflika Amini[3], Irsyadi Yani[4], Rossi Passarella[5], Des Alwine Zayanti[6]

[1,3,6]Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Sriwijaya, Palembang, 30662, Indonesia
[2]Study Program of Plant Protection, Department of Plant Pest and Disease, Faculty of Agriculture, Universitas Sriwijaya, Palembang, 30662, Indonesia
[4]Department of Mechanical Engineering, Faculty of Engineering, Universitas Sriwijaya, Palembang, 30662, Indonesia
[5]Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya, Palembang, 30662, Indonesia
***Corresponding author**: yulia_resti@mipa.unsri.ac.id

Abstract
Corn is an essential agricultural commodity since it is used in animal feed, biofuel, industrial processing, and the manufacture of non-food industrial commodities such as starch, acid, and alcohol. Early detection of diseases and pests of corn aims to reduce the possibility of crop failure and maintain the quality and quantity of crop yields. A decision tree is a nonparametric classification model in statistical machine learning that predicts target variables using tree-structured decisions. The performance of this model can increase significantly if the continuous predictor variables are discretized into valid categories. However, in some cases, the result does not provide satisfactory performance. The possible cause is the ambiguity in discretizing predictor variables. The incorporation of fuzzy membership functions into the model to resolve discretization ambiguity issues. This work aims to classify diseases and pests of corn plants using the decision tree model and improve the model's performance by implementing fuzzy membership functions. The main contribution of this work is that we have shown a significant improvement in the decision tree model performance by implementing fuzzy membership functions; S-growth, triangle, and S-shrinkage curves. The proposed fuzzy model is better than the decision tree model, with an average performance increase from the largest to the smallest; kappa (12.16%), recall (11.8%), F-score (9.71%), precision (5.08%), accuracy (3.23%), specificity (1.94%), and AUC (0.49%). The combination of bias and variance generated by the proposed model is quite small, indicating that the model is able to capture data trends well.

Keywords
Classification, Corn Plant Diseases and Pests, Decision Tree, Fuzzy Membership Function

Received: 17 January 2022, Accepted: 12 May 2022
https://doi.org/10.26554/sti.2022.7.3.284-290

1. INTRODUCTION

Early detection of diseases and pests of corn aims to reduce the possibility of crop failure and maintain the quality and quantity of crop yields. Detection of food plant diseases using digital images by applying classification tasks to statistical machine learning algorithms has become popular in recent years (Resti et al., 2022; Xian and Ngadiran, 2021; Ngugi et al., 2021; Syarief and Setiawan, 2020; Rajesh et al., 2020; Kasinathan et al., 2021; Kusumo et al., 2018). This trend occurs because detection uses low-cost digital images (Ngugi et al., 2021). The choice of features from digital images is essential in classifying diseases and pests of corn because it is a distinguishing factor between classes. Image transformation using the red, green, and blue (RGB) color space model is the most informative feature compared to other features such as the scale-invariant feature transform (SIFT), the speeded-up robust features (SURF), the oriented and rotated brief (ORB), or the histogram oriented gradient (HOG) in detecting corn plant diseases. This feature also has the best accuracy in most machine learning methods applied (Kusumo et al., 2018).

A decision tree is a nonparametric classification model in statistical machine learning that predicts target variables based on tree-structured decisions. For continuous predictor variables, algorithm C4.5 of this model makes decisions by splitting the predictor variables locally (Quinlan, 1996), and the performance of this model increases significantly if the continuous predictor variables are discretized first (Dougherty et al., 1995). In many cases, the implementation of this model has performed well (Kresnawati et al., 2021; Resti et al., 2021; Hussein et al., 2020), including classifying plant diseases using digital images (Rajesh et al., 2020; Kranth et al., 2018). However, in some other cases, the implementation of this model does not provide satisfactory performance (Xian and Ngadiran, 2021; Sahith et al., 2019; García et al., 2015). The reasons that can be caused are the accuracy of method selection in preprocessing and the ambiguity in splitting features to make decisions.

The problem of ambiguity in the feature splitting, known as the discretization process, can be found in many datasets (Ferreira et al., 2015), including digital image data (Singh et al., 2021; Sutha et al., 2021; Resti et al., 2020). Since each feature in an image has a unique pixel value interval, separating digital image data processed using the RGB color space model introduces uncertainty. The fuzzy membership function can address ambiguity issues in decision-making (Dhanalakshimi et al., 2019; Bitar et al., 2016; Semra and Ersoy, 2010). In addition, no less important are the fuzzy membership functions or their combination used in the discretization process (Amini et al., 2021; Resti et al., 2020). The main contribution of this work is that we have shown a significant improvement in the decision tree model performance by implementing fuzzy membership functions; S-growth, triangle, S-shrinkage curves.

2. THE PROPOSED METHOD

2.1 Research Data

The data in this study is in the form of digital images of corn plants' diseases and pests. The digital image capture was carried out during January-March 2021 using a 12-megapixel handphone camera. The locations of the captures are in corn plantations surrounding the University of Sriwijaya, specifically in the villages of Tanjung Seteko, Tanjung Baru, and Tanjung Putus, which are located in the Ogan Ilir Regency of South Sumatra, Indonesia. This investigation resulted in 761 digital images distributed on three types of diseased and three types of pest-infested corn plants. Three types of diseased are LRD (leaf rust disease), DWD (downy mildew disease), and LBD (leaf blight disease) and three types of pest-infested are LP (*Locusta* pests), SFP (*Spodoptera frugiperda* pest), HAP (*Heliotis armigera* pest). The distribution for each class is presented in Table 1.

Figure 1 presents the examples of a digital image of each disease and pest of corn.

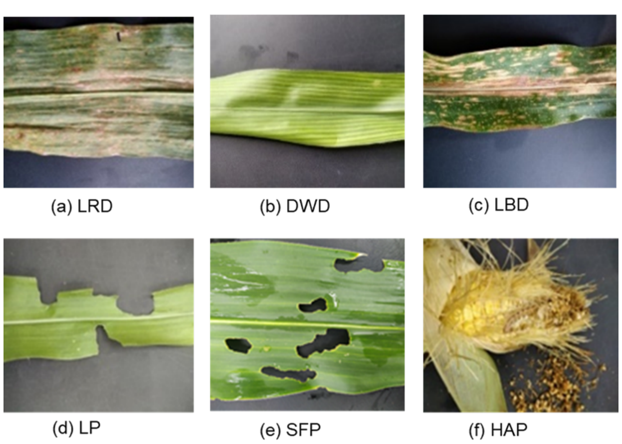

Figure 1. Class Composition of Corn Plant Disease and Pest

Table 1. Percentage of Disease and Pest

Type	Class	%
Disease	LRD	11.56
Disease	DWD	6.44
Disease	LBD	12.88
Pest	LP	14.19
Pest	SFP	39.16
Pest	HAP	15.77

2.2 Method

The method proposed in this work addresses three main points; preprocessing, implementation of fuzzy membership functions in decision tree methods, and performance measurement. In the preprocessing step, the image is cropped to highlight specific observations, especially corn plant diseases and pests. Then, the image is resized to 32×32 pixels and converted to an RGB color space model. The average value of each matrix element R, B, and B is taken as a predictor variable. Furthermore, these numerical values are discretized using (1) as (SAS Institute Inc, 1999). Suppose $X_d°$ be the d-th predictor variable which represents the color pixel values of numeric type. Variable X_d is variable X_d^o which is discretized as much as $k(X_d)$ by (Resti et al., 2022)

$$X_d = X_d^{\,0} + \text{Range}(X_d) \tag{1}$$

where

$$\text{Range}(X_d) = \frac{\max(X_d^{\,0}) - \min(X_d^{\,0})}{k(X_d)}$$

In the next step, let Y be the target variable that represents the types of disease and pest of corn, k(Y) be the number of types in Y, S(Y) be the number of observations in all types Y, $S(X_d^{\,m})$ be the number of observations in the m-th category of the X_d for all types, p_j be the prior probability in the j-th type of Y, and p_m be the prior probability in the m-th category of X_d. A digital image with a vector predictor variable X is categorized as the j-th particular type of disease or pest using the decision model based on tree-structured decisions as depicted in Figure 2.

These tree-structured decisions consist of a root, internal, leaf nodes, and branches. The root node is referred to as the initial, and the nodes that follow it that have or do not have branches are referred to as the internal and leaf nodes, respectively (Witten et al., 2011; Hastie et al., 2009). Each internal node represents X_d that is chosen based on its best information gain, while its branch represents the m-th category of X_d and the leaf node represents the k-th type of disease and pest. This procedure is repeated until all predictor variables represented by the internal node have been classified as belonging to the k -th type of disease or pest represented by the leaf node. Additionally, the generated decision tree is utilized to construct

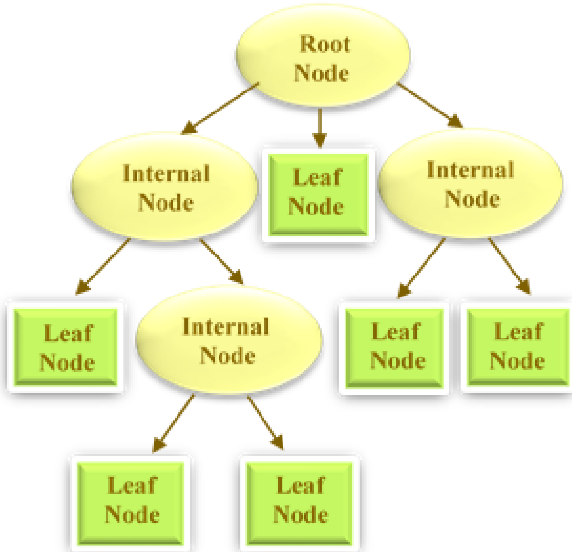

Figure 2. Representation of Decision Tree

a decision rule based on IF-THEN logic (Han et al., 2011). The information gain for the decision tree model is defined as (Mantas and Abellan, 2014),

$$G(Y, X_d) = H(S(Y)) - \sum_{m=1}^{k(X_d)} \frac{S(X_d{}^m)}{S(Y)} H(X_d{}^m) \tag{2}$$

where,

$$H(S(Y)) = -\sum_{j=1}^{k(Y)} P_j * \log_2 P_j \tag{3}$$

$$H(X_d{}^m) = -\sum_{m=1}^{k(X_d)} P_m * \log_2 P_m \tag{4}$$

Furthermore, let X is the universal set, while A denotes the fuzzy set obtained from X. The fuzzy set A in the universe X is expressed as a set of ordered pairs of x and membership function, μ_A(x),

$$A = \{(x, \mu_A(x)) \,|x \epsilon X\} \tag{5}$$

The fuzzy membership function that denoted by μ_A(x), visualizes the degree of membership of each value in a given fuzzy set, A. This function is defined as μ_A: x→[0,1], where each element of x is mapped to a value in the interval [0, 1]. Equations (6) – (8) illustrate the fuzzy membership function for the three curves depicted in Figure 3.

For the triangle curve, a is the element that has the smallest value of the domain that has the smallest membership value, b is the inflection point that is the point that has the dominant 50%, c is the element that has the greatest value of the domain that has the largest membership value, and x is the value of the predictor variable.

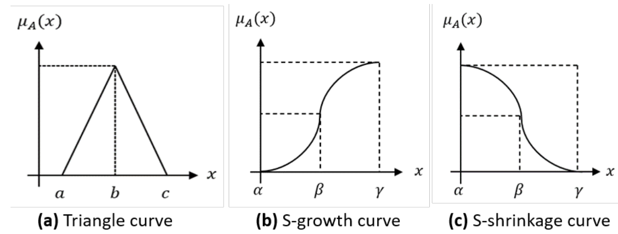

Figure 3. Fuzzy Membership Function

$$\mu_A(x) \begin{cases} 0; & x \le a \\ \frac{(x-a)}{(b-a)}; & a \le x \le b \\ \frac{(c-x)}{(c-b)}; & b \le x \le c \\ 0; & x \ge c \end{cases} \tag{6}$$

$$\mu_A(x) \begin{cases} 0; & x \le \alpha \\ 2\left(\frac{(x-\alpha)}{(\gamma-\alpha)}\right)^2; & \alpha \le x \le \beta \\ 1-2\left(\frac{(\gamma-x)}{(\gamma-\alpha)}\right)^2; & \beta \le x \le \gamma \\ 1; & x \ge \gamma \end{cases} \tag{7}$$

$$\mu_A(x) \begin{cases} 1; & x \le \alpha \\ 1-2\left(\frac{(x-\alpha)}{(\gamma-\alpha)}\right)^2; & \alpha \le x \le \beta \\ 2\left(\frac{(\gamma-x)}{(\gamma-\alpha)}\right)^2; & \beta \le x \le \gamma \\ 0; & x \ge \gamma \end{cases} \tag{8}$$

For the S-growth and S-shrinkage curves, α denotes the domain element with the lowest membership value, β denotes the inflection point, the point with the dominant 50%, and γ is the domain element with the highest membership value. The S-growth curve moves from the leftmost side, starting at the domain value with a membership degree of zero, to the far-right side, toward the domain value with a membership degree of one, and the membership function is focused on 50% of its membership value, which is frequently referred to as the inflection point. The S-shrinkage curve travels from right to left, starting with the domain value with a membership degree of one and ending with the domain value with a membership degree of zero (Dubois and Prade, 2016). The proposed model was obtained by implementing the fuzzy membership function in the discretization process.

The last step is the measurement of the performance of the classification model. A confusion matrix is a table that is frequently used to describe the performance of a classification model on a set of test data whose actual values are known. The matrix represents a straightforward cross-tabulation of observed and expected classes. Let true positives (TP) and true negatives (TN) be proper classifications. False positives (FP) occur when an outcome is incorrectly expected as yes (or positive) when it is in fact, negative (negative). A false negative occurs when a result is incorrectly projected as negative when it is actually good (FN) (Sokolova and Lapalme, 2009). For the first class of diseases and pests of corn plants, the confusion matrix is given in Table 2, and for the other classes is determined similarly.

Table 2. Confusion Matrix for The First Class of Disease and Pest Corn Plant

		Prediction Class					
	j	LRD	DWD	LBD	LP	SFP	HAP
Actual Class	LRD	TP	FN	FN	FN	FN	FN
	DWD	FP	TN	TN	TN	TN	TN
	LBD	FP	TN	TN	TN	TN	TN
	LP	FP	TN	TN	TN	TN	TN
	SFP	FP	TN	TN	TN	TN	TN
	HAP	FP	TN	TN	TN	TN	TN

$$\text{Average Accuracy} = \frac{\sum_j^J = 1\frac{TP_j+TN_j}{TP_j+FP_j+FN_j+TN}}{J} \tag{9}$$

$$\text{Precision Macro} = \frac{\sum_j^J = 1\frac{TP_j}{TP_j+FP_j}}{J} \tag{10}$$

$$\text{recall Macro} = \frac{\sum_j^J = 1\frac{TP_j}{TP_j+FN_j}}{J} \tag{11}$$

$$\text{F Score Macro} = \frac{2\text{Precision macro(Recall Macro)}}{\text{Precision Macro} + \text{Recall Macro}} \tag{12}$$

$$\text{Specificity Macro} = \frac{\sum_j^J = 1\frac{TN_j}{TN_j+FP_j}}{J} \tag{13}$$

$$\text{Kappa} = \frac{\text{Average Accuracy-r}}{1 - r} \tag{14}$$

where

$$r = \frac{a+b}{J^2}$$

$$a = \sum_{j=1}^{J} (TN_j + FP_j) \sum_{j=1}^{J} (TN_j + FN_j)$$

$$b = \sum_{j=1}^{J} (TP_j + FN_j) \sum_{j=1}^{J} (FP_j + TP_j)$$

The area under the receiver operating characteristic (ROC) curve, abbreviated as AUC, is also a good measure for measuring multiclass classification performance (James et al., 2013). The ROC curve for a given model shows the trade-off between the recall and the false positive rate (FPR), where FPR is the negation of specificity (TNR),

$$FPR = 1 - \frac{\sum_{j=1}^{J} \frac{TN_j}{TN_j+FP_j}}{J} \tag{15}$$

For all measures as given in (9)-(15), the larger values, the better the found classification model.

3. RESULTS AND DISCUSSION

The original image of the dataset, which has a resolution of 2000×3000 pixels to 6000×8000 pixels, is cropped to 700× 1000 pixels to 5000×6000 pixels and resized with the same resolution of 32×32 pixels. The examples of the digital image with a resolution of 32×32 are presented in Figure 4.

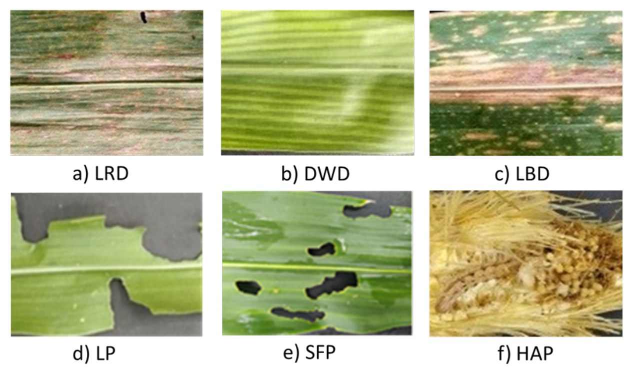

Figure 4. The Cropped and Resized Digital Image

Implementation of the fuzzy membership function in the discretization process formed by the linguistic values. For example, for three categories: dark, medium, and light, or for five categories: very dark, dark, medium, light, and very light.

Our experiment found the best discretization of predictor variables into five categories; very dark, dark, medium, light, and very light. The curve for these categories was S-growth, triangle, triangle, triangle, and S-shrinkage. The pixel value intervals for each variable R, G, and B are 79.10 – 161.64, 82.69 – 167.82, and 36.63 – 129.07.

The parameters of each fuzzy membership function in this work, as presented in Table 3, are obtained using a tuning mechanism where each model is tested by providing a range in the form of arbitrary input values for each fuzzy set (Yunus, 2018). The lower limit value in each of these intervals is the parameter α on the S-growth curve and the upper limit value is the parameter γ on the S-Shrinkage curve. The higher the pixel value range, the lighter the color category.

Table 3. Parameter of Fuzzy Membership Functions

Discretization Category	Par.	Predictor Variable R	G	B
	α	79.10	82.69	36.63
Very Dark	γ	111.25	99.27	52.49
	β	95.17	90.98	44.56
	a	82.69	84.94	42.43
Dark	c	122.33	113.74	72.82
	b	102.51	99.34	57.63
	a	85.94	91.69	44.13
Medium	c	126.35	132.43	91.34
	b	106.14	112.06	67.73
	a	86.53	92.01	45.46
Light	c	143.15	148.58	111.05
	b	114.84	120.30	78.26
	α	87.63	95.71	47.95
Very Light	γ	161.64	167.82	129.07
	β	124.64	131.77	88.51

The performance of the decision tree and the proposed model that implements fuzzy membership functions in each fold are depicted in Figure 5 and Figure 6.

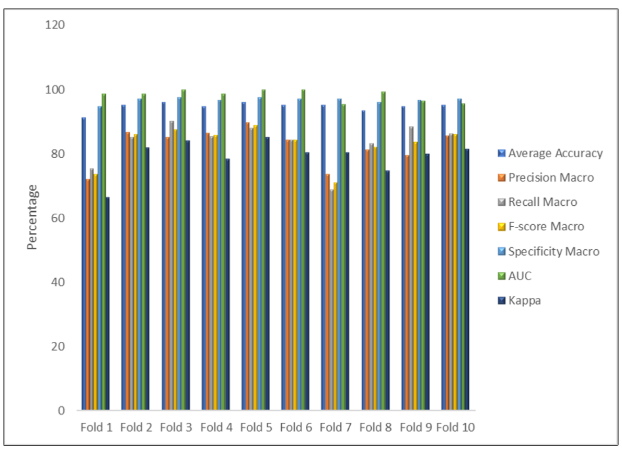

Figure 5. Performance of Decision Tree Model

Table 4 and Table 5 detail the average performance of both models. Metric measures with a higher average indicate that the metric is better than the lower. The average of all performance measures of the proposed model by implementing fuzzy membership functions into the decision tree model is higher than the original model. Three of the average performance measures of the decision tree model are greater than 94%, and the rest, four performance measures, are less than 84%. In the proposed model, almost all of the average performance measures have a value of more than 90%. Even the AUC value is greater than 99%.

Furthermore, the proposed fuzzy model is better than the decision tree model with an average performance increase from

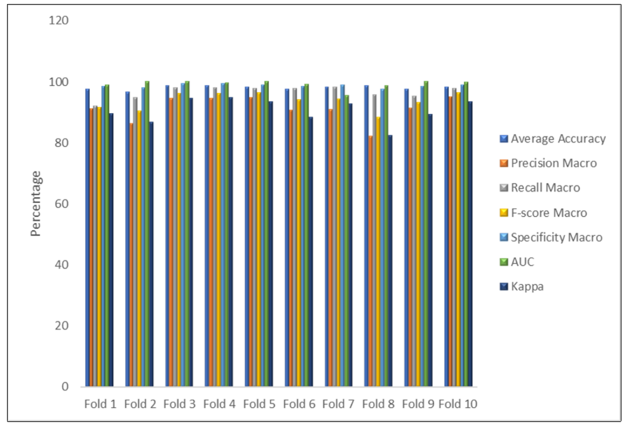

Figure 6. Performance of Decision Tree Model with The Fuzzy Membership Function

Table 4. Performance of Proposed Model

Testing	Average Accuracy	Precision Macro	Recall Macro	F-score Macro	Specificity Macro	AUC	Kappa
Mean	94.53	84.31	83.07	83.58	96.72	98.67	78.80
St.dev	1.76	6.27	5.61	5.28	1.06	2.10	6.62

the largest to the smallest; recall (12.88%), kappa (11.13%), F-score (10.68%), precision (8.59%), accuracy (3.23%), specificity (1.76%), and AUC (0.83%). In addition, referring to Ramasubramanian and Singh (2017), the performance of the decision tree and the proposed models was categorized as good (kappa 60-80%) and very good (kappa more than 80%), respectively. Referring to Mishra et al. (2016), both models' performance was categorized as excellent (AUC more than 90%). Compared to Panigrahi et al. (2020), who also proposed the decision tree model to identify corn plant disease, the result of this work is better. As shown in Panigrahi et al. (2020), the performance measures for the decision tree model are accuracy 74.35%, recall 75.00%, precision 74%, and F_1-score 75.00%. Moreover, The validation between folds in the proposed model also has a lower standard deviation in all performance metrics.

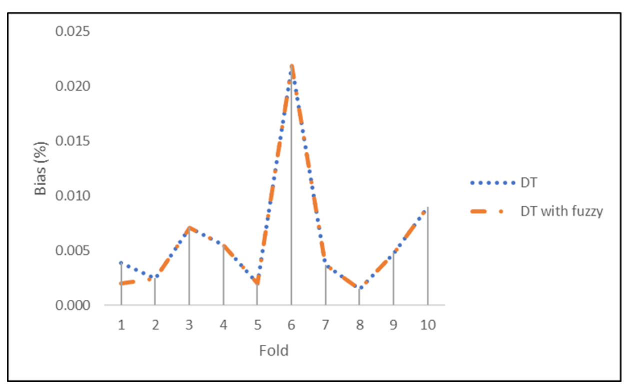

Figure 7. Bias of Proposed Model

The bias and the variance of both models for each fold

Table 5. Performance of Proposed Model

Testing	Average Accuracy	Precision Macro	Recall Macro	F-score Macro	Specificity Macro	AUC	Kappa
Mean	97.76	89.39	94.87	93.29	98.66	99.16	90.96
St.dev	0.96	6.84	5.57	2.87	0.57	1.43	4.01

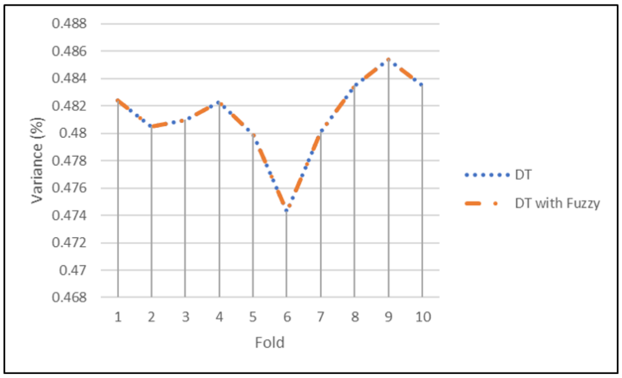

Figure 8. Variance of Proposed Model

are presented in Figure 7 and Figure 8. Bias is the difference between the average predicted results from the proposed model that is built with the test value data. The tenfold cross-validation resampling technique on both models exhibits a small bias (James et al., 2013; Rodriguez et al., 2009; Bengio and Grandvalet, 2003). These values tend to be the same in the two proposed models except for fold 1 where the bias in the FDT model is smaller. The bias difference of both proposed models is statistically significant. However, the variance of the both proposed models as shown in Figure 8 shows identical values. This value is quite large when compared to the bias (James et al., 2013; Rodriguez et al., 2009; Bengio and Grandvalet, 2003), but small enough because it is less than 0.5% so that it can be said that the proposed model, especially the one that implements the fuzzy membership function, is able to capture data trends well.

Moreover, adopting the fuzzy membership function in the suggested decision tree model only affects the prediction model's bias. The combination of bias and variance produced by the suggested model is quite small, indicating that it captures data patterns accurately. The main contribution of this work is that we have shown a significant improvement in the decision tree model performance by implementing fuzzy membership functions; S-growth, triangle, and S-shrinkage curves.

4. CONCLUSIONS

The statistical machine learning models have been implemented for identifying many objects, including food plant diseases and pests. This work classified the diseases and pests of corn plants using the decision tree model and improved the model's performance by implementing fuzzy membership functions. The function's implementation is based on the assumption that the discretization of predictor variables contains ambiguity. However, the proposed methods in this work require further refinement by adding other diseases and pests to the database and confirming the results in an uncontrolled setting. However, we recommend our proposed method to identify corn-based diseases and pests on performance measures that are mostly greater than 90%. The proposed model is better than the decision tree model, with an average performance increase from the largest to the smallest; recall (12.88%), kappa (11.13%), F-score (10.68%), precision (8.59%), accuracy (3.23%), specificity (1.76%), and AUC (0.83%). Additionally, implementing the fuzzy membership function in the decision tree model significant affects the prediction model's bias. However, the combination of bias and variance generated by the proposed model is relatively small, indicating that the model can capture data trends well. This work contribution is that the performance of the decision tree model improves by implementing fuzzy membership functions; S-growth, triangle, and S-shrinkage curves.

5. ACKNOWLEDGMENT

The publication of this paper is the result of discussions with the smart farming development team and the smart inspection system discussion group of the Universitas Sriwijaya.

REFERENCES

Amini, M., Y. Resti, and D. A. Zayanti (2021). *Identifikasi Hama dan Penyakit Pada Tanaman Jagung Menggunakan Metode Random Forest dan Fuzzy Decision Tree*. Sriwijaya University (in Indonesia)

Bengio, Y. and Y. Grandvalet (2003). No Unbiased Estimator of The Variance of k-Fold Cross-Validation. *Advances in Neural Information Processing Systems*, **16**; 1089–1105

Bitar, S., C. Campos, and C. Freitas (2016). Applying Fuzzy Logic to Estimate The Parameters of The Length-Weight Relationship. *Brazilian Journal of Biology*, **76**; 611–618

Dhanalakshimi, R., C. Geetha, and T. Sethukarasi (2019). Monitoring and Detecting Disease in Human Adults using Fuzzy Decision Tree and Random Forest Algorithm. *International Journal of Recent Technology and Engineering*, **7**(5); 93–99

Dougherty, J., R. Kohavi, and M. Sahami (1995). Supervised and Unsupervised Discretization of Continuous Features. In *Machine Learning Proceedings*. Elsevier

Dubois, D. and H. Prade (2016). Practical Methods for Constructing Possibility Distributions. *International Journal of Intelligent Systems*, **31**(3); 215–239

Ferreira, J. A., E. Soares, L. S. Machado, and R. M. Moraes (2015). Assessment of Fuzzy Gaussian Naive Bayes for Classification Tasks. In *Proceedings of The Seventh International Conferences on Pervasive Patterns and Applications (Patterns 2015). Nice, França*, **1**; 64–69

García, S., J. Luengo, and F. Herrera (2015). *Data Preprocessing in Data Mining*, Volume 72. Springer

Han, J., J. Pei, and M. Kamber (2011). *Data Mining: Concepts and Techniques*. Elsevier

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

Hussein, A. M., H. Q. Gheni, W. K. Oleiwi, and Z. Y. Hasan (2020). Prediction of Credit Card Payment Next Month through Tree Net Data Mining Techniques. *International Journal for Computing*, **19**(1); 97–105

James, G., W. Daniela, H. Trevor, and T. Robert (2013). *An Introduction to Statistical Learning: with Applications in R*. Spinger

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

Kranth, G. P. R., M. H. Lalitha, L. Basava, and A. Mathur (2018). Plant Disease Prediction using Machine Learning Algorithms. *International Journal of Computer Applications*, **18**(2); 0975 – 8887

Kresnawati, E. S., Y. Resti, B. Suprihatin, M. R. Kurniawan, and W. A. Amanda (2021). Coronary Artery Disease Prediction using Decision Trees and Multinomial Naïve Bayes with k-Fold Cross Validation. *Inomatika*, **3**(2); 172–187

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. In *International Conference on Computer, Control, Informatics and its Applications*; 93–97

Mantas, C. J. and J. Abellan (2014). Credal-C4. 5: Decision Tree based on Imprecise Probabilities to Classify Noisy Data. *Expert Systems with Applications*, **41**(10); 4625–4637

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. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Volume 9803. International Society for Optics and Photonics

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. In *Progress in Computing, Analytics and Networking*. Springer

Quinlan, J. R. (1996). Improved use of Continuous Attributes in C4. 5. *Journal of Artificial Intelligence Research*, **4**; 77–90

Rajesh, B., M. V. S. Vardhan, and L. Sujihelen (2020). Leaf Disease Detection and Classification by Decision Tree. In *International Conference on Trends in Electronics and Informatics*; 705–708

Ramasubramanian, K. and A. Singh (2017). Machine Learning using R: with Time Series and Industry-based use Cases in R, 2a. *Apress, New Delhi*

Resti, Y., F. Burlian, I. Yani, and I. M. Sari (2020). Improved the Cans Waste Classification Rate of Naïve Bayes using Fuzzy Approach. *Science and Technology Indonesia*, **5**(3); 75–78

Resti, Y., C. Irsan, M. T. Putri, I. Yani, A. Ansyori, and B. Suprihatin (2022). 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

Resti, Y., E. S. Kresnawati, N. R. Dewi, N. Eliyati (2021). Diagnosis of Diabetes Mellitus in Women of Reproductive Age using the Prediction Methods of Naive Bayes, Discriminant Analysis, and Logistic Regression. *Science and Technology Indonesia*, **6**(2); 96–104

Rodriguez, J. D., A. Perez, and J. A. Lozano (2009). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, **32**(3); 569–575

Sahith, R., P. Reddy, and S. Nimmala (2019). Decision Tree-based Machine Learning Algorithms to Classify Rice Plant Diseases. *International Journal of Innovative Technology and Exploring Engineering*, **9**(1); 5365–5368

SAS Institute Inc (1999). *SAS/GRAPH Colors*. In Sas/Graph

Semra, E. and Ö. Ersoy (2010). Comparison of ID3, Fuzzy ID3 and Probabilistic ID3 Algorithms in The Evaluation of Learning Achievements. *Journal of Computing*, **2**(12); 20–25

Singh, A. K., 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

Sutha, P., A. Nandhu Kishore, V. Jayanthi, A. Periyanan, and P. Vahima (2021). Plant Disease Detection using Fuzzy Classification. *Annals of The Romanian Society for Cell Biology*, **25**(4); 9430–9441

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

Witten, I., E. Frank, and M. Hall (2011). Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). *Complementary Literature None*

Xian, T. S. and R. Ngadiran (2021). Plant Diseases Classification using Machine Learning. In *Journal of Physics: Conference Series*, **1962**; 012024

Yunus, M. (2018). Optimasi Penentuan Nilai Parameter Himpunan Fuzzy dengan Teknik Tuning System. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, **18**(1); 21–28 (in Indonesia)

