Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 1, March 2025, pp. 59 – 66
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Implementation of Supervised Learning Method in Grapevine Leaf Classification

[1]Stevanie Aurelia Lifindra, [2]Imam Yuadi
[1]Master Human Resource Development, Post Graduate School, Airlangga University, Indonesia
[2]Department of Information and Library Science, Airlangga University, Indonesia
Email: [1]stevanie.aurelia.lifin-2022@pasca.unair.ac.id, [2]imam.yuadi@fisip.unair.ac.id

Article Info
Article history:
Received Jul 16th, 2024
Revised Sep 3rd, 2024
Accepted Nov 4th, 2024

Keyword:
Classification
Grapevine Leaf
Machine Learning
Supervised Learning Method

ABSTRACT
Grapevine leaves are a type of leaf variety that is difficult to identify because it will take time if processed manually so research will be carried out using the help of machine learning. This research aims to classify 5 varieties of grapevine leaves using orange data mining and several classification methods namely K-Nearest Neighbors (K-NN), logistic regression, random forest and support vector machine (SVM). The dataset used is 500 images and 5 classes where each class consists of 100 images, namely Ak (100), Buzgulu (100), Ala_Idris (100), Dimnit (100), and Nazli (100). The stages in the analysis process are to enter the image into orange data mining by passing several stages so that the image dataset can be processed and read on the test and score so that the confusion matrix can be obtained. The results of the research conducted using orange data mining show that classification using logistic regression gives the best results at a precision value of 0.848% and a recall value of 0.847%. This research shows that classification using orange data mining also provides good results, besides that this research can also help in the classification process so that it does not require a long time.

Corresponding Author:
Stevanie Aurelia Lifindra,
Master Human Resource Development, Post Graduate School,
Airlangga University,
Surabaya
Email: stevanie.aurelia.lifin-2022@pasca.unair.ac.id
DOI: http://dx.doi.org/10.24014/ijaidm.v8i1.31633

1. INTRODUCTION

Grapevine leaves have a high content of bioactive compounds and protein, so they can be further developed in research on grapevine leaf varieties [1]. Grapevine leaves are a type of leaf with various types of varieties and cannot be distinguished by the naked eye, besides the process of identifying the type of grapevine leaves will take a long time with a complicated classification process if done manually [2]. The application of image classification is carried out using algorithms so that it can distinguish features in digital images [3]. Classification is the process of finding a model by distinguishing data classes and is done by predicting unknown data classes [4]. Therefore, classification is used to categorize data into different classes [3].

Classification of grapevine leaf diseases is also done with several classification methods. Research by [5], [6], [7] describes the process of classifying plant diseases that focus mainly on grapevine leaves. According to [8] disease on grapevine leaves is a natural thing due to changes in climatic and environmental conditions. Overcoming this problem can be done by using a disease detection system on grapevine leaves. The Support Vector Machine (SVM) algorithm was implemented and obtained an accuracy result of 93%. Then the Convolutional Neural Network (CNN) model is widely used in the process of classifying grapevine leaf diseases. The use of this model is to facilitate farmers in the process of solving the problem of grapevine leaf disease [9]. The use of CNN models in the classification of grapevine leaf diseases using the python programming language hard librart. The accuracy result using a learning rate of 0.0001 on the CNN model is 91.37% [10]. Grapevine leaves take a long time in the identification process. Here's how to do research on

grapevine leaf classification and utilize several methods. Based on research [11] explains that grapevine leaves are an additional ingredient in traditional dishes because of their different prices and flavors for each variety. In learning and distinguishing it takes a long time. Thus machine learning is used to facilitate automatic classification. The process of classifying grapevine leaves is done using the DenseNet-30 algorithm with a dataset of 500 images and 5 different classes. The result of the accuracy value using the CNN DenseNet-30 model is 98%. Then [2] performed classification of 5 classes of grapevine leaves with the implementation of InceptionV3 architecture which is a CNN architecture. The dataset consists of 500 images and each class consists of 100 images. The results of research using the InceptionV3 model get 99.5% accuracy.

Research using Orange data mining has also been carried out for image classification by comparing several methods so as to get results with the classification with the best method. Research [3] was conducted to classify images of facial features of Indonesian citizens from various Dayak, Batak, Malay, Javanese, and Chinese tribes. The research was conducted with Orange data mining applications with Decision Tree, K-Nearest Neighbor (K-NN), AdaBoost and Support Vector Machine (SVM) algorithms. The results of the research accuracy obtained by Decision Tree 29.75%, (SVM) 37.5%, K-NN 31.55%, and AdaBoost 30.25%. Then research [12] conducted image classification of traditional musical instruments in Indonesia such as angklung, djembe, gong, gamelan, kendang, gordang, tambourine, kulintang, serunai, and sasanda. The research uses the K-NN algorithm, Random forest, and SVM. The results of the average precision value are for k-NN 92.1%, random forest 69.4%, and support vector machine 85.4%. The purpose of the study [13] was to apply image classification methods using different organized classifiers. The data set used in the study included the smallest sample of flower images and was divided into ten classes with a total of 210 images. The research was conducted using six classifiers consisting of logistic regression, neural network, AdaBoost, SVM, Random Forest and K-NN. Based on the results, it is known that the best classification accuracy is using logistic regression, neural network, AdaBoost at 99% while the lowest result is obtained using k-NN classification at 82%.

The difficulty in the classification process of grapevine leaf varieties makes researchers increasingly interested in conducting research using machine learning. The use of machine learning is defined as an automatic programming process to optimize achievement criteria using data or past examples [14]; [13]. Machine learning can be widely used because of its range of applications that can be used in the fields of telecommunications, agriculture, retail, finance, health, and bioinformation [13];[2]. In this case, deepening learning is carried out on machine learning or Orange data mining software, so that predictions can be made on image classes using several available tools. Machine learning is divided into two, namely supervised learning and unsupervised learning. The use of both machine learning can be adjusted to the needs [13]. Image classification algorithms consist of parametric and non-parametric, where both algorithms are under supervised learning [3]. Some machine learning techniques that are commonly used and have been tested [3];[10];[12];[13];[15];[16] include Logistic Regression, SVM, K-NN, Naïve Bayes, and Decision Tree.

This research was conducted to evaluate how well the use of machine learning using the K-NN algorithm, SVM, random forest, Logistic regression in the grapevine leaf classification process, besides that it can provide an easy and appropriate solution in the process of classifying grapevine leaf images. The structure of this research article consists of several parts or sections, where section 1 contains an abstract. then section 2 contains an introduction, section 3 discusses research methods, section 4 contains results and analysis, section 5 contains conclusions, section 6 contains discussion and section 7 contains a references.

2. RESEARCH METHOD

Research was conducted to classify grapevine leaves using Orange Data Mining as a tool in the feature extraction process. That way, feature extraction made using orange can be done using a deep learning model [13]. The extraction process that has been carried out can be continued at the classification stage, classification is applied to the data that has been collected so that the image can be classified based on its type. The classification process using orange is done by applying four classifiers consisting of K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Random Forest, Logistic Regression:

1. Logistic Regression

   Logistic regression classifies data by considering the outcome variable and creating a distinguishing logarithmic line. In addition, logistic regression is the most important analytical tool because it can be used to classify observations into one of two classes such as positive sentiment and negative sentiment [17]. In the process of solving problems using the logistic regression algorithm can be done using several steps [18];[19]. First determine the free and bound attributes contained in the dataset, generally the attribute has two possible values, namely (0 or 1). Second, determine the value of random attributes, namely $\beta_1, \beta_2, \ldots, \beta_p$ as a basic assumption in determining the likelihood. The third step is to calculate the predicted Y value using the following equation 1.

$$Y = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} \quad (1)$$

From equation (1) it is known that β_0 is intercept; $\beta_1, \ldots, \beta_p$ is parameters of the logistic regression algorithm; $x_1, \ldots, x_p$ is independent variable values; p is number of independent attributes; and i is number of rows in the dataset. After finding the Y value, the next step is the fourth step where the value of π(x) will be calculated using equation 2.

$$\pi(x) = \frac{\exp(Y)}{1+\exp(Y)} \quad (2)$$

From equation (2), it is known that $\pi(x)$ is the probability of an event with a good probability value of $0 \leq \pi(x) \leq 1$. The last step is to calculate the maximum log likelihood value using equation 3.

$$maks = Y_i \times \ln(\pi(x) + (1 - Y_i)) \times \ln\big(1 - \pi(x)\big) \quad (3)$$

Where Y_i is the value of the free attribute based on the predetermined class.

2. K-Nearest neighbor (K-NN)
   The k-Nearest Neighbor (k-NN) algorithm is the simplest and easiest type of classification to implement. The k-NN classification is also based on the highest number of votes. Then k specifically indicates the number of neighbors to be used. The training data used will be set on a scale so that it has an explanation of 0 and a standard deviation of 1. Calculations using the K-Nearest Neighbor (k-NN) algorithm can be done using equation 4 [19].

$$d(x, y) = \sqrt{\sum_{i=0}^{n}(x_i - y_1)^2} \quad (4)$$

3. Random Forest
   This algorithm is a development of the decision tree method. Random forest has the advantage of increasing accuracy in the event of data loss besides being used in reisting outliers and used in storing data. The process of feature selection is done to improve the performance of the classification model. This algorithm can be applied in big data with the implementation of complex parameters and is done effectively [18]. In addition, it is explained that random forest is one of the classifications consisting of trees {h(x,k), k = 1, ...} where k is known to be a vector that is randomly distributed and independent. Then each tree chooses the most popular class component with the input value x [20].

4. Support Vector Machine (SVM)
   Support Vector Machine (SVM) is widely used in solving classification and regression problems because of its accuracy against smaller computing power. In addition, SVM has a variety of use cases such as intrusion detection, text classification, face detection, and handwriting recognition [19].

The results of the four algorithms will be compared to determine which algorithm performs best in the image classification process with the smallest sample size. The research will be conducted with the following steps:

1. Dataset
   Dongwload the dataset from Kaggle with the keywords "Grapevine Leaves Image Dataset", then the dongwloaded image files are stored in the data:D computer. The original dataset contains 5 different image classes consisting of leaf types Ak, Ala_Idris, Buzgulu, Dimnit, Nazli with each class containing 100 images. Some sample images from each class can be seen in Figure 1, Figure 2, Figure 3, Figure 4, and Figure 5.

Figure 1. Ak

Figure 2. Ala Idris

Figure 3. Buzgulu

Figure 4. Dimnit

Figure 5. Nazli

2. The flow of the research process stages using Orange data mining.

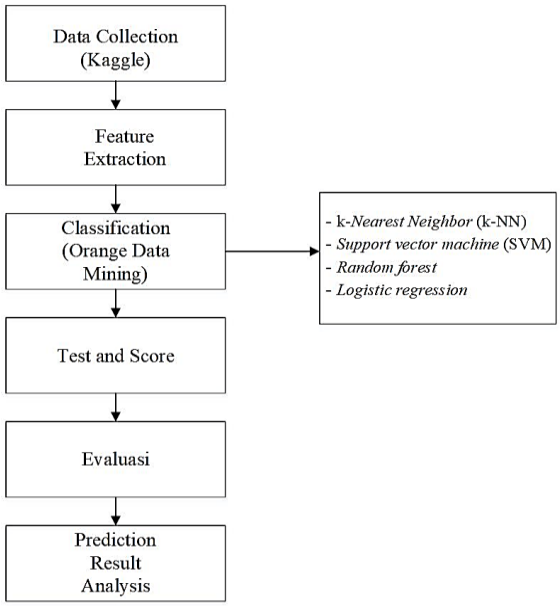

Figure 6. Modeling Using Orange

Figure 6 shows the modeling process using orange. The initial process is to collect data that will be used in the testing process using orange. The data collection process is carried out using the Kaggle WEB, where the data collected is image data of various types of grapevine leaves available on Kaggle. In the research conducted, it is limited to five types of grapevine leaves, namely Ak, Ala_Idris, Buzgulu, Dimnit, and Nazli. The next process will be feature extraction. In this process, feature extraction will be carried out, where each image that has been prepared will be calculated using the Inception v3 deep learning model. This process is incorporated in the image embedding process which is one of the tools in Orange. The function of image embedding is to generate features which will output data on the category, image name, image size, and file size of each image. Next, the classification process will be carried out using the selected algorithm, namely SVM, K-NN, Logistic regression and random forest. The classification process will get the results of each classification process based on different algorithms.

3. RESULTS AND ANALYSIS

Orange data mining is hierarchically very organized and organized, besides that orange is included in open source tools used in processing data mining such as data filtering, probability assessment, and feature assessment which will be assembled into algorithms with higher levels [20]. The research conducted using a dataset taken from kaggle consists of 500 image files with 5 different classes where each class consists of 100 images. The image data consists of images of Ak leaves (100), Ala Idris (100), Buzgulu (100), Dimnit (100), and Nazli (100).

In the research there is an implementation of four algorithms, namely K-NN where the value of the number of neighbors used k = 5. The selection of the value of k because it is easily influenced by noise data if it is too small, besides that when the value of k is too large it can cause the accuracy to decrease [3]. Furthermore, the research metric is euclidean with a weight that is uniform. In research using support vector machine (SVM), the cost value is 1 with a regression loss epsilon value equal to 0.10. Research with the support vector machine algorithm used the RBF kernel with a tolerance value of 0.0010 and an iteration limit of 100. Then the implementation of the random forest algorithm uses a number of trees of 10 with the selection of not doing split subsets of less than 5. The last algorithm used is Logistic regression with the selection of regularization type is Ridge (L2) with C equal to one (1).

The sampling technique in research with the implementation of orange machine learning applications is 5-fold cross validation and random sampling with 80% training data where there is a division of training data and testing data and will be tested using test and score evaluation. The results of random sampling calculations are carried out on grapevine leaf classes using the K-NN algorithm, SVM, random forest, Logistic regression. The results of research using 5-fold cross validation and random sampling can be seen in Tabel 1 and table 2.

Tabel 1. 5-fold cross-validation result

Algorithm	AUC	CA	F1	Precision	Recall
k-NN	0,898	0,692	0,712	0,712	0,692
SVM	0,957	0,806	0,809	0,809	0,806
Random forest	0,881	0,650	0,653	0,653	0,650
Logistic regression	0,978	0,852	0,853	0,853	0,852

Tabel 2. Random sampling results with 80% training data

Algorithm	AUC	CA	F1-score	Precision	Recall
k-NN	0,901	0,694	0,699	0,715	0,694
SVM	0,960	0,804	0,805	0,807	0,804
Random forest	0,886	0,665	0,665	0,666	0,665
Logistic regression	0,977	0,842	0,842	0,843	0,842

The results of the analysis of Tabel 1 and 2 using machine learning orange can be seen that the average result of Area Under Curve (AUC) using the k-NN algorithm is 0.899%. Then SVM with an average AUC value of 0.958%, then random forest of 0.883%, and the last is logistic regression which gets the highest AUC value of 0.977%. Furthermore, the average value of Classified Accurately (CA) shows that the K-NN algorithm gets a value of 0.693%, then followed by the SVM of 0.805% followed by random forest of 0.657% and finally logistic regression which gives the highest result of 0.847%. F1-score test results with the K-NN algorithm obtained a value of 0.705%, where with the SVM algorithm of 0.807% and with random forest of 0.659%, finally using the logistic regression algorithm obtained F1-score value with the highest value of 0.847%. In the precision results, the K-NN algorithm obtained a value of 0.713% while the SVM algorithm obtained a value of 0.808% followed by random forest of 0.659%, finally using the logistic regression algorithm obtained the highest precision value of 0.848%. The recall analysis carried out with the K-NN algorithm obtained a result of 0.693% then using the SVM algorithm obtained a precision value of 0.805% and with random forest of 0.657%, finally using logistic regression obtained the highest recall result of 0.847%.

Confusion matrix provides information in the form of actual and predicted data based on the classification results that have been carried out. The use of confusion matrix can observe the performance of each algorithm used based on the amount of data predicted correctly and incorrectly [3]. Based on the figure displayed using the confusion matrix based on the test results using 4 different combinations of predicted and actual values. There are several terms that represent the results of the classification process consisting of four terms, namely, True positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Confusion matrix in Figure 7, Figure 8, Figure 9, Figure 10 presents the results of Support vector machine (SVM), Logistic regression, k-Nearest Neighbor (k-NN) and random forest algorithms for grapevine leaf classification using orange data mining.

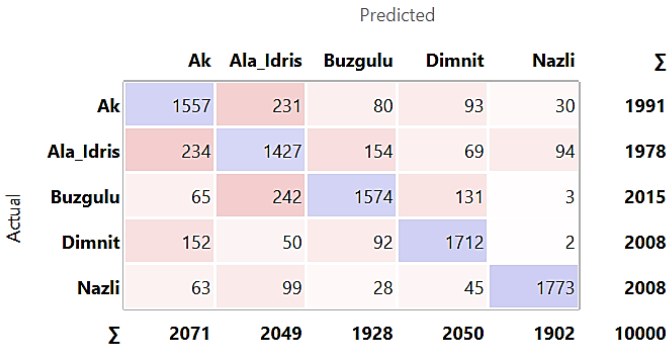

Figure 7. Support Vector Machine

It can be seen that by using the SVM algorithm, a confusion matrix is obtained with a fairly large prediction error in each class. In the Ak class there were 231 observations that were wrongly predicted as Ala-Idris. The Ala-Idris class of 234 was also incorrectly predicted as Ak. In addition, the largest prediction error value is in the Buzgulu class as much as 242 data predicted as Ala_Idris.

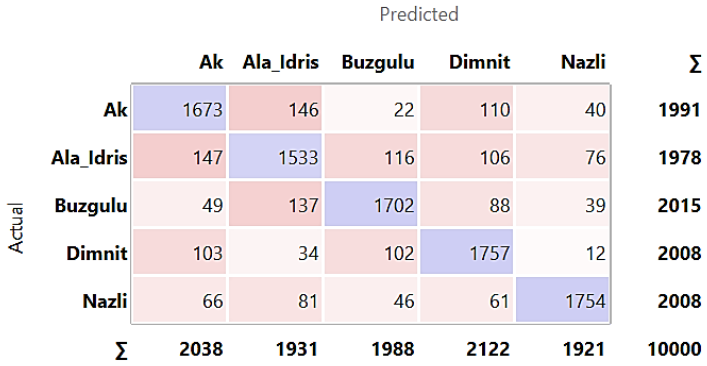

Figure 8. Logistic regression

It can be observed that by using the logistic regression algorithm, a confusion matrix is obtained with a large prediction error in each class. In the Ak class there are 146 observations that are wrongly predicted as Ala-Idris. In addition, the largest prediction error value is in the Ala-Idris class as much as 147 money data predicted as AK.

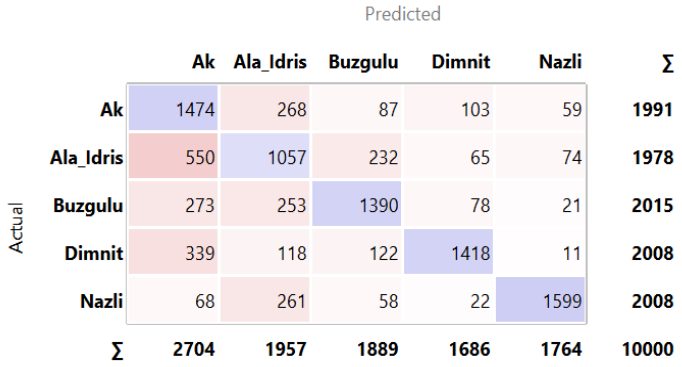

Figure 9. K-Nearest Neighbor (K-NN)

It can be seen that by using the K-NN algorithm, a confusion matrix is obtained with a large prediction error in each class. In the Ak class there are 268 observations that are wrongly predicted as Ala-Idris. In addition, the largest prediction error value is in the Ala-Idris class as much as 550 money data predicted as AK. In addition, in other classes there are also many mispredictions such as in the Buzgulu class of 273 predicted as Ak and 253 as Ala_Idris, then Dimnit is also mispredicted successively by 339 as Ak, 118 as Al_Idris, and 122 as Buzgulu.

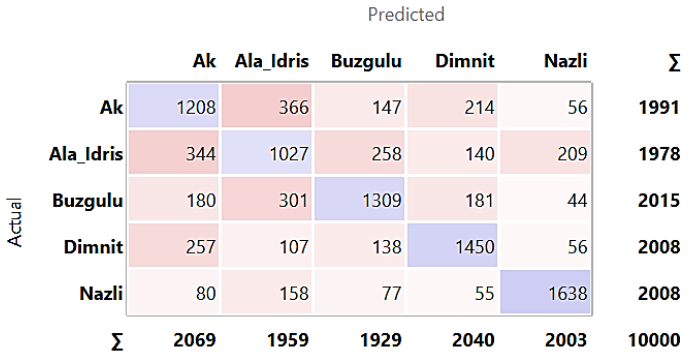

Figure 10. Random Forest

It can be seen that by using the Random Forest algorithm, a confusion matrix is obtained with a fairly large prediction error in each class. In the Ak class there are 366 observations that are wrongly predicted as Ala-Idris. In addition, the largest prediction error value is in the Ala-Idris class as much as 344 money data predicted as AK. Furthermore, there are 301 Buzgulu classes that are incorrectly predicted as Ala_Idris and Dimnit classes that incorrectly predict 257 data as Ak.

The results obtained overall using the four algorithms show that logistic regression provides good performance. Based on the use of logistic regression algorithm in image data processing is superior compared to three other algorithms namely support vector machine, k-nn, and random forest. Where the precision, recall, accuracy, and F1-score values show consistent results or remain superior compared to other algorithms. This is explained because logistic regression has several main advantages, namely ease of use, time efficiency, and easy interpretation by practitioners and researchers [23]. Then it is explained that a large sample size will give good results in the use of logistic regression [24].

The confusion matrix shows that there are many errors in the data classification process. The use of logistic regression algorithm shows that one image class, dimnit, can be predicted well. This is because the dimnit grapevine leaf image has a different shape when compared to the other four grapevine leaf shapes.

Classification of grape leaves using orange data mining gives good results in the implementation of the logistic regression algorithm. In addition, when compared to research [2];[3] using different models with the same goal of classifying grape leaves shows much better results. In the future, research using orange data mining can be developed again by applying the use of different algorithms, so that the best accuracy value can be known. In addition, a larger percentage of training data can be used to get a better accuracy value. This is because the application of the classification process with orange does not require a long time but produces results with high accuracy.

The research results can be used as a reference in performing the image classification process in a fast time. Then it cannot be ruled out that the use of orange data mining is very helpful in the process of image classification of various types of images as has been done by [12]; [13]. However, researchers hope that the development of image classification with orange data mining can still be developed to get better results.

4. CONCLUSION

This research classifies grapevine leaf types by using Orange data mining for image classification process using four different classifiers namely k-Nearest Neighbor (k-NN), Support vector machine (SVM), random forest, and Logistic regression. The leaf image dataset was divided into 5 classes and included 500 images. The average results obtained show that in terms of classified accurately (CA), Area Under Curve (AUC), and F1 the best result is logistic regression with a value of 0.977%; 0.847%; and 0.847%. In addition, the best precision and recall values are also generated from classification using logistic regression, which is 0.848% and 0.847%.

REFERENCES

[1] T. Banjanin, N. Uslu, Z. Vasic, and M. M. Ozcan, "Effect of grape varieties on bioactive properties, phenolic composition and mineral contents of different grape-vine leaves," *J. Food Process. Preserv.*, vol. 45, no. 2, pp. 0–2, 2021,

[2] N. Huda, A. Mahiruna, W. Sulistijanti, R. Chandra, and N. Santi, "Performance Analysis of InceptionV3 Convolutional Network Used for Grapevine Leaves Varieties Classification," vol. 5, no. 2, pp. 48–54, 2023.

[3] S. Hartono, H. Sujaini, and A. Perwitasari, "Komparasi Algoritma Nonparametrik untuk Klasifikasi Citra Wajah Berdasarkan Suku di," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, pp. 337–343, 2020.

[4] R. C. Gonzalez and R. E. Woods, *Digital Image Processing, 4e*. 2018.

[5] Ö. Özaltın and N. Koyuncu, "A Novel Feature Selection Approach-Based Sampling Theory on Grapevine Images Using Convolutional Neural Networks," *Arab. J. Sci. Eng.*, 2024, [Online]. Available: https://doi.org/10.1007/s13369-024-09192-2

[6] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *J. Plant Dis. Prot.*, vol. 131, no. 3, pp. 1061–1080, 2024, [Online]. Available: https://doi.org/10.1007/s41348-024-00896-z

[7] M. Koklu and I. Cinar, "Determination of Effective and Specific Physical Features of Rice Varieties by Computer Vision In Exterior Quality Inspection," *Selcuk J. Agric. Food Sci.*, vol. 35, no. 3, pp. 229–243, 2021,

[8] S. M. Jaisakthi, P. Mirunalini, and D. Thenmozhi, "Grape Leaf Disease Identification using Machine Learning Techniques," *2019 Int. Conf. Comput. Intell. Data Sci.*, pp. 1–6, 2019.

[9] R. S. Pachade, D. A. Sharma, and D. M. Patil, "Grape Leaf Disease Detection Using Deep Learning Based VGG16 Model," vol. 12, no. 5, pp. 1485–1495, 2023.

[10] M. A. Hasan, D. Riana, S. Swasono, A. Priyatna, E. Pudjiarti, and L. I. Prahartiwi, "Identification of Grape Leaf Diseases Using Convolutional Neural Network," *J. Phys. Conf. Ser.*, vol. 1641, pp. 1–7, 2020,

[11] H. A. Ahmed, H. M. Hama, S. I. Jalal, and M. H. Ahmed, "Deep Learning in Grapevine Leaves Varieties Classification Based on Dense Convolutional Network," *J. Image Graph.*, vol. 11, no. 1, pp. 98–103, 2023,

[12] H. Sujaini, "Sistem Analisis Citra Alat Musik Tradisional dengan Metode k-Nearest Neighbor , Random Forest , dan Support Vector Machine," *J. Sist. Inf. Bisnis*, vol. 02, pp. 185–191, 2019.

[13] V. Sajwan and R. Ranjan, "Classifying Flowers Images by using Different Classifiers in Orange," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 1057–1061, 2019,

[14] A. Fauzi, R. Supriyadi, and N. Maulidah, "Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest," vol. 2, no. 1, 2020.

[15] R. Ardiansyah, L. A. S. I. A, and A. S. Rachman, "Klasifikasi Penyakit Daun Anggur Menggunakan YoloV4 dan

OpenCV Pada Aplikasi WEB," 2019.
[16] A. G. Sooai, P. A. Nani, and N. M. R. Mamulak, "Klasifikasi Citra Daun Anggur Menggunakan SVM Kernel Linear," *Journa Inf. Technol. Comput. Sci.*, vol. 8, no. 1, pp. 19–26, 2024,
[17] A. B. Musa, "Logistic Regression Classification for Uncertain Data," *J. Math. Stat. Sci.*, vol. 2, no. 2, pp. 1–6, 2022.
[18] A. Indrasetianingsih, F. Fitriani, and P. Junita, "Klasifikasi Indeks Pembangunan Gender Di Indonesia Tahun 2020 Menggunakan Supervised Machine Learning Algorithms," vol. 4, no. September, pp. 129–136, 2021,
[19] N. D. Azzahra, Ambarwati, A. Desiani, S. I. Maiyanti, and I. Ramayanti, "Perbandingan Algoritma K-Nearest Neighbor Dan Logistic Regression Dalam Klaasifikasi Penyakit Kanker Serviks," *J. Energy*, vol. 14, no. 1, pp. 1–8, 2024, [Online]. Available: https://ejournal.upm.ac.id/index.php/energy/ Jurnal
[20] A. N. Yesa, H. A. Siregar, M. Z. Raditya, and I. Permana, "Comparison of Classification Algorithm Performance for Diabetes Prediction Using Orange Data Mining," *Indones. J. Data Sci.*, vol. 4, no. 3, pp. 176–182, 2023.
[21] A. Navlani, A. Fandango, and I. Idris, *Python Data Analysis*, Third Edit. Birmingham-Mumbai: Packt Publishing All, 2021.
[22] J. Demsar, T. Curk, C. Gorup, and M. Mozina, "Orange : Data Mining Toolbox in Python," *J. Mach. Learn. Res.*, vol. 14, pp. 2349–2353, 2013.
[23] M. I. Gunawan, D. Sugiarto, and I. Mardianto, "Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Seacrh pada Algoritma Logistic Regression," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, pp. 280–284, 2020.
[24] S. A. T. Al Azhima, D. Darmawan, N. F. A. Hakim, I. Kustiawan, M. Al Qibtiya, and N. S. Syafei, "Hybrid Machine Learning Model Untuk Memprediksi Penyakit Jantung Dengan Metode Logistic Regression Dan Random Forest," *J. Teknol. Terpadu*, vol. 8, no. 1, pp. 40–46, 2022, [Online]. Available: https://journal.nurulfikri.ac.id/index.php/jtt

BIBLIOGRAPHY OF AUTHORS

Stevanie Aurelia Lifindra (stevanie.aurelia.lifin-2022@pasca.unair.ac.id) was born in Bajawa. I am a student of Universitas Airlangga, Graduate School, Department of Human Resource Development with a concentration in Data Analytics. I have an interest in data science, so my research focus is related to image data processing using machine learning with orange data mining.

Imam Yuadi (imam.yuadi@fisip.unair.ac.id) is a lecturer at the Department of Information and Library Science, FISIP Unair since 1999. Imam continued his doctoral studies in Information Management from National Chiao Tung University, Taiwan in 2017. He specializes in digital libraries, digital forensics, and image processing. Imam's interest in Digital Forensics began at the beginning of his doctoral studies which discussed the differences in printouts from various printers that can be seen through digital analysis. Armed with this initial knowledge, Imam began to explore Digital Forensics both in terms of theory and practice.