

Economic Potential of Using Machine Learning and Haar Wavelet Transform to Reduce Losses in the Industrial Sector and Enhance Economic Sustainability

Yessi Jusman^{1*}, Alfinto Maulana¹, Adam Umarov³

¹Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah

²Faculty of Electrical Engineering and Technology, Universiti Malaysia Perlis, 02600 Perlis,

³Kadyrov Chechen State University, Grozny, Russia

Abstract. Oil palm plants are essential as they produce palm fruit that can be processed into edible oil—an essential human need. However, these plants are often infected with diseases, negatively impacting crop productivity and the quality of the oil produced. These diseases are caused by mushrooms, bacteria, viruses, and pests that can spread rapidly and damage the leaves. Therefore, early detection of oil palm leaf disease plays a crucial role in reducing the negative impact on crops and significant economic losses. This study aims to design a system to classify the types of leaf diseases of oil palm plants using texture feature extraction (Haar Wavelet Algorithm) and machine learning-based classification algorithms (Cubic SVM, Medium Gaussian SVM, Quadratic SVM, Cosine KNN, Fine KNN, and Weighted KNN). Cubic SVM yielded the highest training result with an averages accuracy of 81.54% and an average time of 48.135 seconds. However, Medium Gaussian SVM outperformed other models during testing, producing an accuracy of 87%, precision of 81%, recall of 81 %, specificity of 90%, and F-score of 81%.

1 Introduction

Palm oil is one of the agricultural products playing a critical role in the global vegetable oil industry. Ready-made foods and biodiesel fuels are only two examples of the many products utilizing palm oil. In addition, it is also applied as a raw material in the food, pharmaceutical, and cosmetic industries. In this regard, Indonesia is the world's top producer of palm oil, with a production of 46.98 million tonnes in 2023 [1]. As has been highlighted, the country's palm oil production reached 46 million tonnes for the marketing year of 2022/2023 [2]. The title is set in bold 16-point Arial, justified. The first letter of the title should be capitalised with the rest in lower case. You should leave 22 mm of space above the title and 6 mm after the title.

* Corresponding author: yjusman@umy.ac.id

Moreover, the area dedicated to palm oil in Indonesia extended by 13.5 million hectares, with a rise of 3 million hectares in the marketing year of 2000/2001 [2]. Accordingly, the palm oil industry has had a significant economic impact on the country. This sector provides direct employment for nearly 4 million people, often in remote rural areas with a scarcity of alternative employment [3]. This industry has also contributed to reducing rural poverty and improving the country’s infrastructure. Hence, the palm oil industry has been a vital source of income for millions of people in Indonesia, with the average income of smallholders being seven times that of farmers involved in the subsistence production of food crops.

Unfortunately, oil palm plants are often affected by diseases, adversely affecting crop productivity and the quality of the oil produced. These diseases are caused by mushrooms, bacteria, viruses, and pests that can spread quickly and damage the leaves. Therefore, early detection of oil palm leaf disease is essential in lessening the negative impact on crops and significant economic losses. However, manual disease detection methods commonly deployed by medical personnel or agricultural specialists are often time-consuming, costly, and require a fairly high level of expertise [4].

With the rapid development of the times, advances in technology and science are also increasing, and a deep understanding of the various aspects of science is the key to the survival of science. These advances in technology can help solve complex problems quickly and with accurate results. Especially in agriculture, today’s detection system technology for plant diseases has been widely applied in diseases on tea leaves [6], orange plant leaves [7], aromatic leaves [8], coconut tree leaves [9], as well as banana, nut, rose, and lemon leaves [4,5].

In recent years, machine learning and deep learning techniques have been applied to diagnosing oil palm leaf diseases. Machine learning has depicted the ability to achieve excellent accuracy, ranging from 86% to 99.56% in diagnosing certain leaf diseases [10-13]. Furthermore, previous studies on the use of deep learning for classifying palm leaf diseases have disclosed similar results, yielding accuracy between 87% and 99.67% [6], [14-17].

This study was conducted as an innovative approach to developing methodologies using image processing and machine learning in response to the demand for oil palms as raw materials for human needs and the existing research in this field. This research utilized image processing techniques during the pre-processing stage, Haar wavelet transform level 2 algorithms for feature extraction, and K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms for classification.

2 Methodology

This section describes the system design stages for the classification of leaf diseases in oil palm plants using MATLAB. Figure 1 illustrates the research stages.

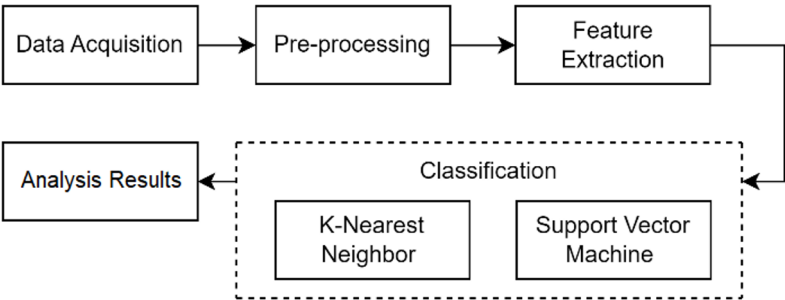


Fig. 1. Research stages.

2.1 Data Acquisition

This study utilized data in the form of images of oil palm plant leaves obtained from the Kaggle website with a total of 2,631: 470 images of the brown spot class, 958 images of the white scale class, and 1,203 images of the healthy class [18]. Visually, the oil palm leaves of the brown spot category have nearby brown stacks, while those in the white scale category have white stacks, and the healthy leaves are green without stacks. These images had a resolution of 350 x 192 pixels and were in JPG format. Figure 2 exhibits the differences in images across classes.

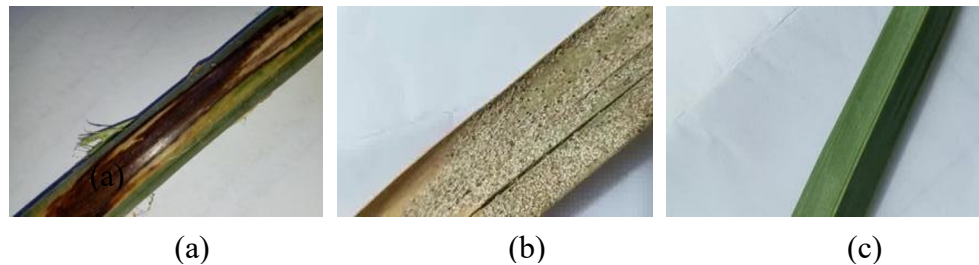


Fig. 2. Oil palm plant leaf images: (a) Brown Spots, (b) White Scale, (c) Healthy.

2.2 Pre-processing

Pre-processing was performed to improve image quality. This phase was divided into four parts: image rotation, augmentation, grayscale, and image enhancement. Image rotation rotated images to the same size from 192 x 350 pixels to 350 x 192 pixels. Augmentation was conducted to create a new variation on the images of the brown spot class using a horizontal flip, resulting in 940 images in this category. Grayscale was adopted to convert RGB images to gray ones. Image enhancement was performed to boost image quality and escalate the clarity of features on images using the *adapthisteq* function in MATLAB.

2.3 Feature Extraction

Feature extraction takes a feature from an image to obtain a feature value. This study utilized Haar wavelet transform level 2—a feature extraction method based on texture. Haar wavelet transform is the simplest and relatively easiest method discovered by Alfred Haar in 1909 [19]. Haar wavelet utilizes two filters: the Low Pass Filter (LPF) and the High Pass filter (HPS) [20]. The formulas for the LPF and the HPS coefficients of the Haar wavelet are as equations 1 and 2:

Low Pass Filter : $h_0 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$ (1)

High Pass Filter : $h_1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)$ (2)

Description:

h0 – Tapis Low Pass

h1 – Tapis High Pass

2.4 Classification

The classification in this study adopted two algorithms: SVM and KNN. Three models with the highest accuracy for each method were selected for SVM (Cubic SVM, Medium Gaussian SVM, Quadratic SVM) and KNN (Cosine KNN, Fine KNN, and Weighted KNN). SVM is a learning system using a hypothetical space in a vast feature space. The classification is carried out by finding a model that differentiates classes of data with the aim of estimating the unknown class of an object using a separator field (hyperplane) [21]. To solve nonlinear problems, SVM was modified by inserting kernel functions, allowing data to be transferred to larger dimensions (feature space). The equations 3-6 are some of the most commonly applied kernel functions [22]:

$$\text{Kernel Linear} \quad K(x, y) = x, y \quad (3)$$

$$\text{Kernel Polynomial} \quad K(x, y) = (x \cdot y + c)^d \quad (4)$$

$$\text{Kernel Gaussian RBF} \quad K(x, y) = \exp\left(\frac{-||x-y||^2}{2 \cdot \sigma^2}\right) \quad (5)$$

$$\text{Kernel Sigmoid} \quad K(x, y) = \tanh(\sigma(x, y) + c) \quad (6)$$

KNN is a machine learning algorithm for classification and regression. It works by searching for the nearest neighbor value of a new data point and then taking the label (class) of most of the closest neighbors as a prediction of the label of the new data point [23]. It employs the Euclidean distance measurement with the equations 7 and 8:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (7)$$

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (8)$$

Description:

- dissimilarity measure
- data variable
- variable at the center point

3 Results

Image data undergoing pre-processing served as input to perform feature extraction. Haar wavelet transform level 2 utilized 535 features from three image classes to run feature extraction. The feature extraction results, as summarized in Table 1, were saved in an Excel file. Table 1 illustrates the results of feature extraction using the Haar Wavelet method on oil palm leaf images, showing a clear distinction in the feature values across different classes. The values for Feature 1 are significantly different among Class 1, Class 2, and Class 3, demonstrating the ability of Haar Wavelet transformation to capture unique characteristics in the leaf images that correspond to each class. This pattern of distinct feature values continues across Feature 2 and the subsequent features, further emphasizing the effectiveness of this method in distinguishing between the classes. The significant differences in feature values across the classes provide a robust basis for accurate classification, as the extracted features serve as reliable indicators for identifying and categorizing the different types of oil palm leaves. This distinct separation between the classes in the feature space highlights the potential of Haar Wavelet features to improve the performance of subsequent classification algorithms.

Table 1. Results of Haar Wavelet Level 2 Feature Extraction.

Haar Wavelet Level 2 Feature	Average ± Standard deviation		
	Class 1 (Brown Spots)	Class 2 (White Scale)	Class 3 (Healthy)
1	95.7 ± 3.7	96.8± 3.5	96.5 ± 3.9
2	0.4 ± 0.3	0.4 ± 0.5	0.5 ± 0.9
3	0.7 ± 0.4	0.5 ± 0.5	0.7 ±1
4	1.4 ± 1.5	0.9 ± 1.2	0.9 ± 1.3
:	:	:	:
532	19.7 ± 17.7	14.4 ± 13.1	17.9 ± 17.8
533	19.1 ± 17.1	13.9 ± 12.8	17.3 ± 16.6
534	19.1 ± 17.6	13.5 ± 12.8	17.4 ± 17.5
535	8E-15 ± 2.6E-15	9E-15 ± 4E-15	1E-14 ± 4E-15

Table 2. Training Accuracy Results.

Run	Accuracy (%)					
	Cosine KNN	Fine KNN	Weight ed KNN	Cubic SVM	Medium Gaussian SVM	Quadratic SVM
Run 1	74.9	77.6	72.4	81.4	79.8	79.1
Run 2	74.9	77.1	72.7	81.0	79.8	79.0
Run 3	75.3	77.3	72.3	82.2	79.8	78.9
Run 4	75.0	77.4	72.5	81.5	79.5	78.7
Run 5	74.6	77.7	72.9	81.1	80.0	79.3
Run 6	75.1	77.8	73.1	81.4	80.1	79.1
Run 7	75.6	77.9	72.7	81.5	79.9	79.1
Run 8	75.5	77.9	72.7	82.2	79.6	79.5
Run 9	75.4	77.4	72.5	81.6	80.3	79.1
Run 10	75.9	77.2	72.1	81.5	79.8	79.1
Average± St Dev	75.2±0.4	77.5±0.3	72.6±0.3	81.5±0.4	79.9±0.2	79.1±0.2

Feature extraction results were employed as the classification input. The classification was performed using KNN and SVM. Three models with the highest accuracy for each method were identified, KNN (Cosine KNN, Fine KNN, and Weighted KNN) and SVM (Cubic SVM, Medium Gaussian SVM, Quadratic SVM). The classification underwent two stages: training and testing. Table 2 exhibits the training accuracy results.

Table 2 displays the classification performance results of various KNN and SVM models. The average accuracy achieved by each KNN model (Cosine KNN, Fine KNN, and Weighted KNN) is 75.2%, 77.5%, and 72.6%, respectively, while for the SVM models (Cubic SVM, Medium Gaussian SVM, and Quadratic SVM), the average accuracy is 81.5%, 79.9%, and 79.1%. Therefore, the highest average accuracy among the KNN models was achieved by Fine KNN at 77.5%, while the highest among the SVM models was achieved by Cubic SVM at 81.5%. The highest accuracy for each KNN model was as follows: Cosine KNN achieved the highest accuracy of 75.9% on the tenth run, Fine KNN achieved the highest accuracy on the seventh and eighth runs at 77.9%, and Weighted KNN obtained the highest accuracy of 73.1% on the sixth run. Meanwhile, for the SVM models, Cubic SVM achieved the highest

accuracy of 82.2% on the eighth run, Medium Gaussian SVM achieved 80.3% on the ninth run, and Quadratic SVM achieved 79.5% on the eighth run.

Table 3. Training Time Results.

Run	Time (s)					
	Cosine KNN	Fine KNN	Weighted KNN	Cubic SVM	Medium Gaussian SVM	Quadratic SVM
Run 1	14.2	15.6	17.2	37.1	43.1	25.8
Run 2	12.9	15.6	14.0	23.6	23.2	27.9
Run 3	11.8	14.4	10.0	52.4	24.4	24.6
Run 4	14.7	14.9	9.7	52.5	23.3	25.5
Run 5	14.3	16.0	11.3	52.6	22.8	23.3
Run 6	13.8	15.1	10.8	52.5	24.0	26.5
Run 7	17.2	16.8	12.1	51.8	22.7	25.1
Run 8	14.6	16.6	12.3	53.0	26.	28.0
Run 9	15.5	15.0	10.4	53.6	28.8	24.3
Run 10	14.5	16.5	11.4	52.4	28.3	25.1
Average±St Dev	14.4±1.4	15.6±0.8	11.9±2.2	48.1±9.9	26.7±6.1	25.6±1.5

Table 3 displays the training times for each KNN and SVM model to compare which model has the fastest computational time. The average training times for the three KNN models (Cosine KNN, Fine KNN, and Weighted KNN) are 14.4s, 15.6s, and 11.9s, respectively, while for the three SVM models (Cubic SVM, Medium Gaussian SVM, and Quadratic SVM), the average training times are 48.1s, 26.7s, and 25.6s. Based on these average training times, the Weighted KNN model is identified as the fastest among the KNN methods with a training time of 11.9s, while among the SVM methods, the Quadratic SVM model is the fastest, with a training time of 25.6s. Both Cosine KNN and Fine KNN recorded their fastest training times on the third run, at 11.8 seconds and 14.4 seconds, respectively. Meanwhile, Weighted KNN achieved its quickest training time on the fourth run, with 9.7 seconds. For the SVM models, Medium Gaussian SVM had the fastest training time of 22.7 seconds on the seventh run, followed by Quadratic SVM with 23.3 seconds on the fifth run, and Cubic SVM with 23.6 seconds on the second run.

The blue-colored tables in Figure 3 depict the number of images classified correctly by the models. In contrast, the red ones demonstrate the number of images classified incorrectly by the models. The matrix confusion results revealed that Cubic SVM obtained the highest score with 846 images in Class 1 in True Class, 862 images in Class 2 in True Class, and 1,083 images in Class 3 in True Class. On the third run, 707 images were correctly classified as Class 1, 675 images as Class 2, and 912 images as Class 3. The greatest training result for each model was utilized as a brain in testing. Table 4 portrays the best matrix performance test result on each model.



Fig. 3. The best of the training confusion matrix: (a) Cosine KNN, (b) Fine KNN, (c) Weighted KNN, (d) Cubic SVM, (e) Medium Gaussian SVM, and (f) Quadratic SVM.

After evaluating the performance of various KNN models using the Haar wavelet transform level 2 matrix, it was found that the Fine KNN model outperformed the others, achieving the highest classification metrics: an accuracy of 83%, a precision of 76%, a recall of 74%, a specificity of 88%, and an F-score of 74%. These results indicate that Fine KNN is particularly effective in correctly identifying true positives while maintaining a strong balance between sensitivity and specificity. On the other hand, the Weighted KNN model demonstrated the lowest performance among the KNN variants, with an accuracy of 78%, a

precision of 72%, a recall of 70%, a specificity of 85%, and an F-score of 68%. This suggests that while Weighted KNN still performs reasonably well, it struggles more with balancing false positives and false negatives compared to Fine KNN.

Table 4. The Best Results of Testing.

Model	Performance Matrix (%)				
	Accuracy	Precision	Recall	Specificity	F-score
Cosine KNN	79%	73%	72%	84%	72%
Fine KNN	83%	76%	74%	88%	74%
Weighted KNN	78%	72%	70%	85%	68%
Cubic SVM	85%	78%	78%	89%	78%
Medium Gaussian SVM	87%	81%	81%	90%	81%
Quadratic SVM	84%	77%	77%	88%	77%

When comparing KNN to SVM models, the Medium Gaussian SVM emerged as the top performer, delivering an impressive accuracy of 87%, with a recall of 81%, a specificity of 90%, and an F-score of 81%. The superior performance of the Medium Gaussian SVM highlights its robustness in handling the complexities of the dataset, particularly in distinguishing between different classes with a high degree of precision and reliability. This comparison underscores the effectiveness of SVM, especially the Medium Gaussian variant, in leveraging the features extracted through Haar wavelet transformation for accurate classification tasks.

4 Conclusion

The feature extraction using the Haar Wavelet method effectively differentiated the classes, as the feature values between classes were significantly distinct, making it a strong foundation for the classification stage. The highest average training accuracy was achieved by the Fine KNN model at 77.5% and by the Cubic SVM model at 81.5%. Additionally, the Weighted KNN model demonstrated superior training efficiency, achieving the fastest training time of 9.7 seconds, indicating that KNN outperformed SVM in terms of computational speed. Furthermore, the Medium Gaussian SVM model delivered the best testing performance with an accuracy of 87%, precision of 81%, recall of 81%, specificity of 90%, and an F-score of 81% for the Haar wavelet transform feature extraction, underscoring its effectiveness in classification tasks

The Acknowledgements, this research is supported by Universitas Muhammadiyah Yogyakarta through the Center of Artificial Intelligence and Robotic Studies, Research and Innovation Institute.

References

1. Statista, “Production volume of palm oil in Indonesia from 2014 to 2023.” [Online]. Available: <https://www.statista.com/statistics/706786/production-of-palm-oil-in-indonesia/>

2. “Foreign Agricultural Service Global Market Analysis International Production Assessment Division Indonesia Palm Oil: Historical Revisions Using Satellite-Derived Methodology,” 2023.

3. M. Russell, "Palm oil: Economic and environmental impacts." [Online]. Available: <http://www.europarl.europa.eu/thinktank>
4. V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, Mar. 2017, doi: 10.1016/j.inpa.2016.10.005.
5. I. Ahmad, Y. Rahmanto, R. I. Borman, F. Rossi, Y. Jusman, and A. D. Alexander, "Identification of Pineapple Disease Based on Image Using Neural Network Self-Organizing Map (SOM) Model," in *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022, pp. 12–17. doi: 10.1109/ICE3IS56585.2022.10010110.
6. U. Hairah, A. Septiarini, N. Puspitasari, E. Romiyanto, N. A. Hadiwijaya, and D. Nurcahyono, "Analysis of EfficientNet Architecture Performance for Classifying the Tea Leaves Diseases," in *Proceedings of the 9th International Conference on Computer and Communication Engineering, ICCCE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 223–228. doi: 10.1109/ICCCE58854.2023.10246097.
7. Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Computers and Electronics in Agriculture*, vol. 153. Elsevier B.V., pp. 12–32, Oct. 01, 2018. doi: 10.1016/j.compag.2018.07.032.
8. G. Dhingra, V. Kumar, and H. D. Joshi, "A novel computer vision based neutrosophic approach for leaf disease identification and classification," *Measurement (Lond)*, vol. 135, pp. 782–794, Mar. 2019, doi: 10.1016/j.measurement.2018.12.027.
9. S. Bharathi and P. Harini, "Early Detection of Diseases in Coconut Tree Leaves," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 1265–1268. doi: 10.1109/ICACCS48705.2020.9074357.
10. E. P. Atika, A. Sunyoto, and E. T. Luthfi, "Genetic Algorithm and K-Nearest Neighbors for Oil Palm Leaf Disease Classification," in *ICOIACT 2022 - 5th International Conference on Information and Communications Technology: A New Way to Make AI Useful for Everyone in the New Normal Era*, Proceeding, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 447–451. doi: 10.1109/ICOIACT55506.2022.9971854.
11. A. Septiarini, H. Hamdani, T. Hardianti, E. Winarno, S. Suyanto, and E. Irwansyah, "Pixel quantification and color feature extraction on leaf images for oil palm disease identification," in *7th International Conference on Electrical, Electronics and Information Engineering: Technological Breakthrough for Greater New Life, ICEEIE 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICEEIE52663.2021.9616645.
12. A. Nor, I. Masazhar, M. Kamal, S. Alam, and S. Malaysia, "Measurement and Applications (ICSIMA) 28," 2017.
13. A. F. Aji et al., "Detection of Palm Oil Leaf Disease with Image Processing and Neural Network Classification on Mobile Device," *International Journal of Computer Theory and Engineering*, pp. 528–532, 2013, doi: 10.7763/ijcte. 2013.v5.743.
14. H. Hamdani, A. Septiarini, A. Sunyoto, S. Suyanto, and F. Utaminigrum, "Detection of oil palm leaf disease based on color histogram and supervised classifier," *Optik (Stuttg)*, vol. 245, Nov. 2021, doi: 10.1016/j.ijleo.2021.167753.

15. J. H. Ong, P. Ong, and W. K. Lee, "Image-based Oil Palm Leaf Disease Detection using Convolutional Neural Network," *Journal of Information and Communication Technology*, vol. 21, no. 3, pp. 383–410, Jul. 2022, doi: 10.32890/jict2022.21.4.
16. A. Septiarini, H. Hamdani, E. Junirianto, M. S. S. Thayf, G. Triyono, and Henderi, "Oil Palm Leaf Disease Detection on Natural Background Using Convolutional Neural Networks," in *Proceeding - IEEE International Conference on Communication, Networks and Satellite, COMNETSAT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 388–392. doi: 10.1109/COMNETSAT56033.2022.9994555.
17. S. Firdaus, M. Nasution, and F. Fahmi, "Initial Design For Utilizing Machine Learning In Identifying Diseases In Palm Oil Plant," in *Proceeding - ELTICOM 2023: 7th International Conference on Electrical, Telecommunication and Computer Engineering: Sustainable and Resilient Communities with Smart Technologies*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 95–99. doi: 10.1109/ELTICOM61905.2023.10443120.
18. "Date Palm data," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/hadjerhamaidi/date-palm-data>
19. A. S. Lewis and G. Knowles, "Image Compression Using the 2-D Wavelet Transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 244–250, 1992, doi: 10.1109/83.136601.
20. M. Bhagat and D. Kumar, "Performance evaluation of PCA based reduced features of leaf images extracted by DWT using random Forest and XGBoost classifier," *Multimed Tools Appl*, vol. 82, no. 17, pp. 26225–26254, Jul. 2023, doi: 10.1007/s11042-023-14370-9.
21. Y. Jusman, J. H. Lubis, A. N. N. Chamim, and S. N. A. M. Kanafiah, "Feature Extraction Performance to Differentiate Spinal Curvature Types using Gray Level Co-occurrence Matrix Algorithm," in *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 337–341. doi: 10.1109/ICOIACT50329.2020.9332067.
22. I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, "The effect of gamma value on support vector machine performance with different kernels," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, pp. 5497–5506, 2020, doi: 10.11591/IJECE.V10I5.PP5497-5506.
23. Y. Jusman, J. H. Lubis, S. N. A. M. Kanafiah, and M. I. Yusof, "Comparison of spine curvature images classification using support vector machine and K-nearest neighbors," *AIP Conf Proc*, vol. 2499, no. 1, p. 080001, Nov. 2022, doi: 10.1063/5.0105008