Jurnal Teknologi Riset Terapan (JATRA)
ISSN 2986-7169 , Vol 2, No 1, 2024, 27-36 https://doi.org/10.35912/jatra.v2i1.4598

Technology-Based Classification of Clerodendrum paniculatum Using CNN and Confusion Matrix

Pandu Wijaya[1], Alvin Reihansyah Makarim[2], Meizano Ardhi Muhammad[3], Cela Febriyani[4], Vezan Hidayatullah[5], Resty Annisa[6]
University,Bandar Lampuung, Lampung, Indonesia[1,2,3,4,5,6]
pandu.wijaya21@students.unila.ac.id[1], alvin.reihansyah21@students.unila.ac.id[2], meizano@eng.unila.ac.id[3], cela.febriyani21@students.unila.ac.id[4], vezan.hidayatullah21@students.unila.ac.id[5], resty.annisa@eng.unila.ac.id[6]

Riwayat Artikel
Received on 07 Desember 2023
1st Revision 12 December 2023
2nd Revision 23 December 2023
3rd Revision 26 December 2023
Accepted on 04 January 2024

Abstract
Purpose: This study aims to develop a classification system for the Clerodendrum paniculatum plant (Bunga Pagoda), focusing on its key parts—stems, flowers, leaves, and trees—using the Convolutional Neural Network (CNN) algorithm. The objective is to support conservation efforts and facilitate digital data grouping through technology-based classification.
Methodology: The research involved collecting a dataset of images representing different parts of the Clerodendrum paniculatum plant. These images were then used to train a CNN model. The training process included 200 epochs to optimize performance. The model's accuracy and performance were evaluated using a confusion matrix to measure classification success across the plant's various parts.
Results: The CNN model achieved its highest accuracy of 97.78% when trained for 200 epochs. The results indicated a significant improvement in evaluation metrics compared to models trained with fewer epochs. The mo del successfully classified the plant parts with high precision, demonstrating its robustness and reliability for rare plant classification.
Conclusions: This study confirms that the CNN algorithm is effective in classifying the parts of the Clerodendrum paniculatum plant. Increasing the number of training epochs substantially enhances the model's performance, making it a practical tool for digital plant conservation initiatives.
Limitations: The study is limited by its reliance on a specific dataset, which may not encompass all possible variations of the Clerodendrum paniculatum plant under different environmental conditions.
Contributions: This research contributes to digital plant conservation by developing a CNN-based classification system for rare plants. It highlights the importance of deep learning in biodiversity preservation and provides a foundation for future AI-driven botanical studies.

Keywords: *Clerodendrum Paniculatum, Convolutional Neural Network (CNN), Plant Clasification, Tropical Forest.*

How to Cite: Wijaya, P., Makarim, A. R., Muhammad, M. A., Febriyani, C., Hidayatullah, V., Annisa, R. (2024). Technology-Based Classification of Clerodendrum paniculatum Using CNN and Confusion Matrix. *Jurnal Teknologi Riset Terapan*, 2(1), 27-36.

1. Pendahuluan

Clerodendrum paniculatum plants are one of the rare plant species that grow in tropical forest areas, and have great potential in the world of botany (Nona et al., 2024; Weny.J.A Musa, 2017) Each part of this plant, from the stem, leaves, flowers, to trees, has different morphological characteristics. Plant classification based on these parts is important, not only for scientific purposes, but also to support efforts to preserve and group digital data. However, manual classification efforts often face challenges due to the complexity of morphology and visual variations in each part of the plant(Hartono et al., 2020).

In the current era of technological development, Artificial Intelligence (AI) and Machine Learning (ML)-based approaches(Helm et al., 2020) are increasingly being applied to overcome the challenges of object classification(Khan & Al-Habsi, 2020; Penumuru et al., 2020), including in the realm of botany. One method that is widely used for image-based classification is the Convolutional Neural Network (CNN)(Pornpanomchai & Pornpanomchai, 2021; Pratiwi et al., 2021) The CNN algorithm is able to recognize visual patterns in images and produce accurate classification models(Tama & Santi, 2023).

Digital image-based plant classification research has been widely developed, such as the study by Khorram, Vahidi, and Ahmadian (2020), developed plant classification using CNN on leaf datasets with high accuracy, but the focus was only on leaves without considering other plant parts such as stems or flowers. Meanwhile, Zhu, Abdalla, Tang, and Cen (2022) applied deep learning for image-based plant disease classification, but the research was oriented towards disease detection, not the classification of plant parts as a whole (link). In contrast to these studies, this study focuses on the classification of multi-parts (stems, leaves, flowers, and trees) of the rare plant Clerodendrum paniculatum as a whole with a CNN model, which strengthens its novelty in supporting comprehensive digital conservation of rare plants (Melvi, Nurhayati, Batubara, Septama, & Ulvan, 2023).

This study focuses on collecting technology-based datasets for *Clerodendrum paniculatum* plants and developing classification models using the CNN algorithm. This model is trained with images that include the stems, leaves, flowers, and trees of the *Clerodendrum paniculatum* (Abade et al., 2020; Krichen, 2023; Mohammad Mustafa Taye, 2023) plant. The use of confusion matrix as an evaluation tool is expected to provide a clear view of the classification performance, including precision, recall, f1-score, and overall accuracy metrics. The performance of the model tested in, this study also tested the performance of the model with epoch number settings, 200 epochs, to obtain optimal accuracy. With this evaluation, it is hoped that this classification system can make a significant contribution to efforts to conserve and automatically group rare plant data, especially for *Clerodendrum paniculatum* plants (Mardiono, Nanra, & Rican, 2023).

2. Literature Review

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms that have proven highly effective in image classification tasks. CNNs mimic the human visual system by using layers of interconnected neurons to detect spatial hierarchies of features from input images (Riesna, Pujianto, Efendi, Nugroho, & Saputra, 2023). The architecture typically consists of convolutional layers, pooling layers, and fully connected layers that work together to extract and classify complex patterns in image data. CNNs have become the foundation of many applications in plant recognition and environmental monitoring due to their robustness in handling large and complex datasets (Melvi, Ulvan, Sidiq, & Batubara, 2023).

Image-based plant classification plays a vital role in biodiversity conservation and agricultural management. Accurate identification of plant parts such as leaves, stems, flowers, and roots are essential for taxonomic studies, disease detection, and preservation of rare species. Traditional manual classification methods are time-consuming and error-prone, particularly when dealing with large-scale datasets. The integration of machine learning, especially CNNs, offers a scalable and precise alternative for plant classification tasks (Nanni, Maguolo, & Pancino, 2020).

The training process in CNN-based classification relies on large datasets to optimize feature extraction and improve model accuracy. The number of training epochs, data augmentation techniques, and hyperparameter tuning significantly impact model performance (Yando, Panusunan, & Fauzan, 2023). Longer training periods generally allow the model to learn more complex patterns, but they must be balanced with considerations of overfitting and computational cost. Therefore, carefully selecting the number of epochs is crucial to achieving optimal results in image classification tasks (Sudeepa, Aithal, Rajinikanth, & Satapathy, 2020).

Digital conservation efforts increasingly rely on artificial intelligence to document and preserve biodiversity. The application of CNNs to classify rare plant species supports not only ecological preservation but also the creation of structured digital archives for research and education. As environmental threats and biodiversity loss continue to escalate, leveraging advanced computational tools such as CNNs offers a promising pathway for sustainable biodiversity monitoring and management (Zhou, Majeed, Naranjo, & Gambacorta, 2021).

3. Research Methods

The method used includes technology-based dataset collection from each part of the plant, followed by model training using the CNN algorithm. Evaluation is carried out with a confusion matrix to measure classification performance, including precision, recall, and f1-score

3.1 Data Collection

The data used in this study are digital images of *Clerodendrum paniculatum* plants. The data collection process is carried out manually by taking pictures of plants in the Gita Persada Butterfly Park using a cellphone camera.The usage of cellphone camera are meant to replicate activities of biodiversity researcher. After collection, the images are stored in digital format with a standard resolution suitable for further processing purposes.

3.2 Data Preparation

Dataset of *Clerodendrum paniculatum* contains 450 images,which divided at random into training data (314 Images) validation data (45 Images), and test data (91 Images).

3.3 Model Training

The CNN model designed with several convolutional, pooling, and fully connected layers. The details of the model architecture used:

1. Convolutional Layers: Using convolutional layers to extract features from images. The kernel size used is 3x3, with ReLU activation function to maintain non-linearity.
2. Pooling Layers: Using MaxPooling2D to reduce the dimensionality of features and maintain important information. Fully Connected Layers: After several layers of convolution and pooling, the features are flattened and passed to a dense fully connected layer for the classification process.
3. Dropout: To prevent overfitting, dropout is applied in several fully connected layers with a rate of 0.5. Output Layer: Using the Softmax activation function on the output layer to generate the predicted probability of each class.

3.4 Model Evaluation

This study uses a Convolutional Neural Network (CNN) model to classify *Clerodendrum paniculatum* plant parts into four classes: stems, flowers, leaves, and trees. With a dataset consisting of manually taken images, the model was trained for 200 epochs to achieve optimal performance

4. Result and Discussion

4.1 Data Collection

The data collection process was carried out by taking images of *CleroDendrum paniculatum* plants in Gita Persada Butterfly Park. The dataset obtained consisted of 450 images grouped into four classes based on plant parts, namely stems, flowers, trees, and leaves. As shown in table

Table 1. Dataset of *Clerodendrum paniculatum*

Class of Plant part	Amount of Images
Stem	100

Flowers	200
Tree	50
Leaf	100
Total	450

The sample images of *Clerodendrum paniculatum* plant display on figure 1 stems,figure 2 flowers,figure 3 leaf,figure 4 tree.

(figure 1)

(figure 2)

(figure 3)

(figure 4)

4.2 Data Preparation

The data preparation process aims to prepare data with a total of 450 images, divided into 314 images or around 70% as training data, 45 images or around 10% as validation data and 91 images or around 20% as testing dataset. Based on the data above, it consists of 4 classes or 4 types of parts of the *Clerodendrum paniculatum* plant. Of all 4 types of plant parts, each stem part has 100 images, the flower part has 200 images, the tree part has 50 images and the leaf part has 100 images.

Table 2. Data preparation of *Clerodendrum paniculatum*

Class of Plant part	Training	Validation	Test
Stem	68	6	26

Flowers	143	21	36
Tree	33	6	11
Leaf	70	12	18
Total	314	45	91

4.3 Training

4.3.1 Computational Specification

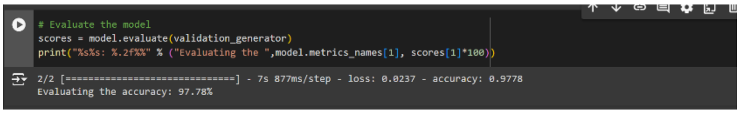
The model was trained using Google Colab with access to GPU to accelerate the training process. The use of a batch size of 32 and the Adam optimizer with an initial learning rate of 0.001 were used to accelerate convergence.

4.3.2 Training Configuration

The model was trained with 200 epoch configurations.Based on the experimental results, the configuration with 200 epochs gave the best results with the highest accuracy of 97.78% on the validation dataset(Darmatasia & A. Muhammad Syafar, 2023).

4.4 Evaluation

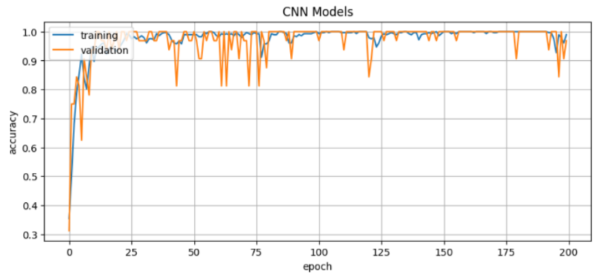
Based on the results of the model evaluation on the test dataset consisting of 91 images, the following evaluation metrics were obtained:



4.4.1 Accuracy and Loss Graph

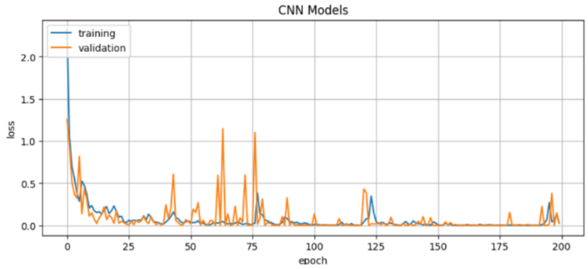
Based on the two graphs displayed, it can be seen that the training results of the Convolutional Neural Network (CNN) model in terms of loss and accuracy for 200 epochs, both on training and validation data(figure)

Based on the accuracy graph with the x-axis (number of epochs) and the y-axis (accuracy level with a scale of 1:100) it shows that the graph between training and validation is running significantly increased because the difference between the training and validation graphs runs in line with the difference in fluctuation distance which is directly proportional to what can be seen when entering the graph starts to run significantly without any difference in fluctuation with an average touching 90%(Bijalwan et al., 2022).



Clerodendrum paniculatum Model Accuracy for 200 epoch

The loss graph (figure) with the x-axis (number of epochs) and the y-axis (accuracy level with a scale of 1:100) shows that the loss level only reaches its highest point at the beginning of the epoch, namely in the epochs from -1 to 25, after passing the 25th epoch the loss level decreases significantly and the fluctuation level between training and validation is not far apart, this can happen because based on the previous accuracy graph it has a high accuracy value so that it reduces the loss level on the loss graph(Bijalwan et al., 2022).



4.4.2 Precision,recall,and f1-score

Based on the image, the evaluation results of the classification model for the *Clerodendrum panicaulatum* plant show several performance metrics, namely precision, recall, f1-score, and support for each class, as well as an overall accuracy of 93%. Only 77% of the total stem class objects were successfully identified correctly(Yacouby & Axman, 2020).

Table 3. Performance metrics of Clerodendru paniculatum

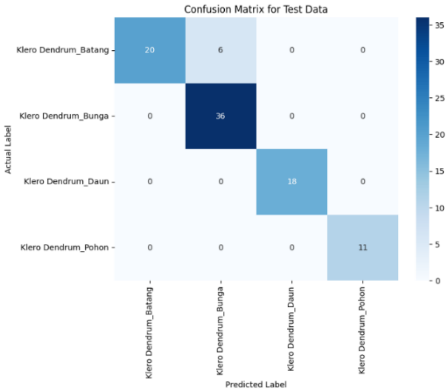
	Precision	recall	f1-score	support
Klero Dendrum_Batang	1.00	0.77	0.87	26
Klero Dendrum_Bunga	0.86	1.00	0.92	36
Klero Dendrum_Daun	1.00	1.00	1.00	18
Klero Dendrum_Pohon	1.00	1.00	1.00	11

Accuracy			0.93	91
Macro avg	0.96	0.94	0.95	91
Weighted avg	0.94	0.93	0.93	91

1. Klero Dendrum_Batang Class:
   a. Precision: 1.00 (the model is able to correctly identify all predictions for the stem class).
   b. Recall: 0.77 (only 77% of the total objects of the stem class were correctly identified).
   c. F1-Score: 0.87 (a combination of precision and recall indicating good overall performance but lower recall).
   d. Support: 26 (number of samples for the stem class).
2. Klero Dendrum_Bunga Class:
   a. Precision: 0.86 (almost 86% of all predictions for the flower class were correct).
   b. Recall: 1.00 (the model is able to correctly identify all objects of the flower class).
   c. F1-Score: 0.92 (a combination of precision and recall, indicating very good performance).Support: 36 (number of samples for the flower class).
3. Klero Dendrum_Daun Class:
   a. Precision: 1.00 (all predictions for the leaf class are correct).
   b. Recall: 1.00 (all leaf class objects are correctly identified).
   c. F1-Score: 1.00 (perfect performance).
   d. Support: 18 (number of samples for the leaf class).
4. Klero Dendrum_Pohon Class:
   a. Precision: 1.00 (all predictions for the tree class are correct).
   b. Recall: 1.00 (all tree class objects are correctly identified).
   c. F1-Score: 1.00 (perfect performance).
   d. Support: 11 (number of samples for the tree class).

Overall Performance Metrics:
1. Accuracy: 0.93 (accuracy is 93%, indicating that out of all predictions made, 93% of them are correct).
2. Macro avg: Precision (0.96), Recall (0.94), F1-Score (0.95) are calculated as simple averages over all classes.
3. Weighted avg: Precision (0.94), Recall (0.93), F1-Score (0.93) are calculated as weighted averages based on the number of samples in each class.

4.4.3 Confusion Matrix
Based on the displayed Confusion Matrix(Bijalwan et al., 2022), the performance of the CNN model in classifying test data for parts of the *Clerodendrum paniculatum* plant can be seen.



1. klero Dendrum_Batang Class:
   a. Correct prediction: 20 (diagonal, meaning 20 images of the stem class were correctly identified as stems).
   b. Wrong prediction: 6 images from the stem class were incorrectly predicted as the flower class. No images were incorrectly classified as leaves or trees.
   c. This is consistent with the recall of 0.77 shown previously (20/26 stem samples were correctly identified).
2. Klero Dendrum_Bunga Class:
   a. Correct prediction: 36 flower images were correctly identified as flowers (100% recall, because all images from the flower class were correctly classified).
   b. There were no incorrect predictions for the flower class.
3. Klero Dendrum_Daun Class:
   a. Correct prediction: 18 leaf images were correctly identified as leaves (100% recall).
   b. No leaf images were misclassified.
4. Klero Dendrum_Pohon Class:
   a. Correct prediction: 11 tree images were correctly identified as trees (100% recall).
   b. No tree images were misclassified

The results showed that the model trained for 200 epochs provided the highest accuracy in classifying each plant part, with significant improvements in evaluation metrics compared to lower epoch counts. This model is able to recognize *Clerodendrum paniculatum* plant parts with sufficient accuracy to be applied in rare plant classification.

5. Conclusion
Based on the research results, several conclusions that can be drawn are, the creation of a model made using the Convolution Neural Network algorithm type is able to provide a good level of accuracy in the training process/training dataset with the highest level of accuracy obtained at 200 epochs, namely reaching an accuracy level of 97.87%.

Limitations and Future Study

The study is limited by its reliance on a specific dataset, which may not encompass all possible variations of the Clerodendrum paniculatum plant under different environmental conditions. Further testing with larger and more diverse datasets is needed to generalize the model's applicability.

Acknowledgements

The authors gratefully acknowledge the contribution of Gita Persada Butterfly Park and grant supported by LPPM Universitas Lampung.

References

Abade, A. S., Ferreira, P. A., & Vidal, F. de B. (2020). *Plant Diseases recognition on images using Convolutional Neural Networks: A Systematic Review* (No. arXiv:2009.04365). arXiv. http://arxiv.org/abs/2009.04365

Bijalwan, V., Semwal, V. B., & Gupta, V. (2022). Wearable sensor-based pattern mining for human activity recognition: Deep learning approach. *Industrial Robot: The International Journal of Robotics Research and Application*, *49*(1), 21–33. Https://doi.org/10.1108/IR-09-2020-0187

Darmatasia & A. Muhammad Syafar. (2023). IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI TANAMAN RIMPANG SECARA VIRTUAL. *Jurnal Instek*, *8*(1).

Hartono, A., Adlini, M. N., Ritonga, Y. E., Tambunan, M. I. H., Nasution, M. S., & Jumiah, J. (2020). IDENTIFIKASI TUMBUHAN TINGKAT TINGGI (PHANEROGAMAE) DI KAMPUS II UINSU. *Jurnal Biolokus*, *3*(2), 305. Https://doi.org/10.30821/biolokus.v3i2.755

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, *13*(1), 69–76. Https://doi.org/10.1007/s12178-020-09600-8

Khan, A. I., & Al-Habsi, S. (2020). Machine Learning in Computer Vision. *Procedia Computer Science*, *167*, 1444–1451. Https://doi.org/10.1016/j.procs.2020.03.355

Khorram, Asghar, Vahidi, Bahman, & Ahmadian, Bahram. (2020). Computational analysis of adhesion between a cancer cell and a white blood cell in a bifurcated microvessel. *Computer Methods and Programs in Biomedicine, 186*, 105195.

Krichen, M. (2023). Convolutional Neural Networks: A Survey. *Computers*, *12*(8), 151. Https://doi.org/10.3390/computers12080151

Mardiono, D. A., Nanra, S., & Rican, D. (2023). Rancang Bangun Pengaman Pintu Menggunakan RFID Dengan Mikrokontroler Atmega 328. doi:https://doi.org/10.35912/jatra.v1i1.1872

Melvi, M., Nurhayati, N., Batubara, M. A. M., Septama, H. D., & Ulvan, A. (2023). Unjuk Kerja Teknologi Akses Jamak TD-CDMA dan TD-SCDMA pada Infrastruktur Jaringan High Altitude Platform Stations. *Jurnal Teknologi Riset Terapan, 1*(1), 51-59. doi:10.35912/jatra.v1i1.1790

Melvi, M., Ulvan, A., Sidiq, M. R., & Batubara, M. A. M. (2023). Rancang Bangun Sistem Monitoring Ketinggian Muka Air Laut Menggunakan Arduino Pro Mini dan NodeMCU ESP8266. doi:https://doi.org/10.35912/jatra.v1i1.1794

Mohammad Mustafa Taye. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *MDPI*, *11*(3), 52. Https://doi.org/10.3390/computation11030052

Nanni, Loris, Maguolo, Gianluca, & Pancino, Fabio. (2020). Insect pest image detection and recognition based on bio-inspired methods. *Ecological Informatics, 57*, 101089. doi:https://doi.org/10.1016/j.ecoinf.2020.101089

Nona, R., Ramadhani, F., & Andrea, R. (2024). Development "Sincatensa", a Plant Sensing System Application in Land Dry Areas Using Qr-Code. *TEPIAN*, *5*(1), 30–34. Https://doi.org/10.51967/tepian.v5i1.3023

Penumuru, D. P., Muthuswamy, S., & Karumbu, P. (2020). Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. *Journal of Intelligent Manufacturing*, *31*(5), 1229–1241. Https://doi.org/10.1007/s10845-019-01508-6

Pornpanomchai, C., & Pornpanomchai, V. (2021). *Plant Leaf Image Recognition Based on Convolutional Neural Network*. *37*(2), 79–92.

Pratiwi, H. A., Cahyanti, M., & Lamsani, M. (2021). IMPLEMENTASI DEEP LEARNING FLOWER SCANNER MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK. *Sebatik*, *25*(1), 124–130. Https://doi.org/10.46984/sebatik.v25i1.1297

Riesna, D. M. R., Pujianto, D. E., Efendi, A. J. I., Nugroho, B. A., & Saputra, D. I. S. (2023). Identifikasi Platform dan Faktor Sukses dalam Manajemen Proyek Teknologi Informasi. *Jurnal Teknologi Riset Terapan, 1*(1), 1-9. doi:10.35912/jatra.v1i1.1458

Sudeepa, KB, Aithal, Ganesh, Rajinikanth, V, & Satapathy, Suresh Chandra. (2020). Genetic algorithm based key sequence generation for cipher system. *Pattern Recognition Letters, 133*, 341-348. doi:https://doi.org/10.1016/j.patrec.2020.03.015

Tama, A. M., & Santi, R. C. N. (2023). Klasifikasi Jenis Tanaman Hias Menggunakan Metode Convolutional Neural Network (CNN). *INTECOMS: Journal of Information Technology and Computer Science*, *6*(2), 764–770. Https://doi.org/10.31539/intecoms.v6i2.7002

Weny.J.A Musa. (2017). *ISOLASI SENYAWA ANTIFEEDANT DARI TUMBUHAN CLERODENDRUM PANICULATUM*. ZAHIR publishing.

Yacouby, R., & Axman, D. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, 79–91. Https://doi.org/10.18653/v1/2020.eval4nlp-1.9

Yando, J. R., Panusunan, P., & Fauzan, F. (2023). Penggunaan Filler Tanah (Silt) sebagai Perencanaan Campuran Aspal Beton AC-WC. doi:https://doi.org/10.35912/jatra.v1i1.1873

Zhou, Zheng, Majeed, Yaqoob, Naranjo, Geraldine Diverres, & Gambacorta, Elena MT. (2021). Assessment for crop water stress with infrared thermal imagery in precision agriculture: A review and future prospects for deep learning applications. *Computers and Electronics in Agriculture, 182*, 106019. doi:https://doi.org/10.1016/j.compag.2021.106019

Zhu, Yueming, Abdalla, Alwaseela, Tang, Zheng, & Cen, Haiyan. (2022). Improving rice nitrogen stress diagnosis by denoising strips in hyperspectral images via deep learning. *Biosystems Engineering, 219*, 165-176. doi:https://doi.org/10.1016/j.biosystemseng.2022.05.001