Journal of Research in Artificial Intelligence for Systems and Applications (RAISA)
2025, Vol. 01, No. 02, E-ISSN: 3090-3157

Original Research

Disease Classification in Cauliflower Plants Using Vgg19 Architecture and Support Vector Machine (SVM) + Lime

Utut Ardiansah[1], Tio Dharmawan[2]

[1]Departement of Computer Science, University of Jember

Abstract

Cauliflower (Brassica oleracea L. botrytis) is a cool-season vegetable rich in fiber, vitamin B, and phytonutrients that provide significant health benefits including cardiovascular protection and cancer risk reduction. Manual monitoring of plant diseases is extremely difficult and time-consuming, making early disease detection crucial for efficient cauliflower cultivation in the agricultural sector. This study aims to classify diseases in cauliflower plants using VGG19 architecture combined with Support Vector Machine (SVM) and Local Interpretable Model-Agnostic Explanations (LIME). The dataset consists of 7,360 digital images covering three disease types (downy mildew, black rot, bacterial spot rot) and healthy plants. Results show that the VGG19+SVM model achieved 99.86% accuracy, outperforming standalone VGG19 (99.46%). LIME analysis successfully visualized critical areas underlying the model's predictions. These findings demonstrate the effectiveness of combining deep learning and machine learning for plant disease detection, while enhancing model interpretability through visual explanations.

Keywords

Cauliflower, VGG19, SVM, LIME, disease classification

1. Introduction

Agriculture plays a vital role in global food security, with vegetable production being essential for meeting the nutritional needs of the growing world population. Cauliflower (*Brassica oleracea* L. botrytis), a member of the Brassicaceae family, is one of the most important vegetable crops worldwide due to its high nutritional value, containing essential fiber, vitamin B, and phytonutrients that reduce cardiovascular disease risk and support vital organ functions [1]. However, cauliflower cultivation faces significant challenges from various plant diseases that can severely impact crop yield and quality.

Traditional manual disease detection methods in agriculture are labor-intensive, time-consuming, and often require specialized expertise, making them impractical for large-scale farming operations [2]. The conventional approach of visual inspection by agricultural experts is not only costly but also prone to human error and subjective interpretation. Furthermore, early-stage disease detection is crucial for effective treatment, yet manual identification of subtle disease symptoms often occurs too late to prevent significant crop losses.

Recent advances in computer vision and deep learning have opened new possibilities for automated plant disease detection systems. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in image classification tasks, making them particularly suitable for agricultural applications [3]. Several researchers have explored the application of pre-trained CNN architectures for plant disease classification with promising results.

Rajab et al. [4] successfully implemented VGG-16 and VGG-19 architectures for grapevine leaf disease

*Corresponding author: Utut Ardiansah

Email addresses:

u.ardiansah@gmail.com (Utut Ardiansah), tio.pssi@unej.ac.id (Tio Dharmawan)

Received: 19 Juli 2025; **Accepted:** 25 Juli 2025; **Published:** 30 Juli 2025

classification, achieving accuracy rates of 99.6% and 100% respectively across five different grape species. Their work demonstrated the effectiveness of deep learning approaches in agricultural disease detection. Similarly, Nguyen et al. [5] applied VGG-19 with transfer learning and image segmentation techniques for tomato leaf disease classification, processing 16,010 HSV-segmented tomato leaf images and achieving 99.72% accuracy. These studies highlight the potential of VGG architectures for plant disease recognition tasks.

In medical imaging applications, Rajinikanth et al. [6] developed a customized VGG-19 network combined with handcrafted features for brain tumor detection using 2D MRI slices. Their comparative evaluation of various deep learning architectures, including AlexNet, VGG16, VGG19, ResNet50, and ResNet101, revealed that VGG-19 combined with SVM-RBF classifier achieved superior classification accuracy exceeding 97% across different imaging modalities. This research demonstrates the effectiveness of combining CNN feature extraction with traditional machine learning classifiers.

The interpretability of deep learning models remains a critical concern in agricultural applications, where understanding the decision-making process is essential for building trust among farmers and agricultural experts. Eray Önler [7] addressed this challenge by implementing LIME (Local Interpretable Model-Agnostic Explanations) to explain CNN-based predictions for cassava leaf disease classification. Their work demonstrated how LIME could visualize which parts of leaf images influenced the model's classification decisions, providing valuable insights into the model's behavior across five disease classes including Cassava Bacterial Blight, Cassava Brown Streak Disease, Cassava Green Mite, Cassava Mosaic Disease, and healthy leaves.

Despite these advances, cauliflower disease detection remains underexplored compared to other crops. Cauliflower is particularly susceptible to several devastating diseases including Downy Mildew caused by *Peronospora parasitica*, Black Rot caused by *Xanthomonas campestris* pv. *campestris*, and Bacterial Spot Rot caused by *Pseudomonas syringae* pv. *maculicola* [8]. Each of these diseases presents distinct visual symptoms that can be challenging to differentiate, especially in early stages, making automated classification systems highly valuable for farmers.

Current research gaps include the limited application of hybrid approaches that combine the feature extraction capabilities of deep CNN architectures with the robust classification performance of traditional machine learning algorithms like Support Vector Machines (SVM). While VGG-19 has proven effective as a feature extractor, its integration with SVM classifiers for cauliflower disease detection has not been thoroughly investigated. Additionally, the interpretability aspect of such hybrid models using explanation techniques like LIME requires further exploration to ensure practical applicability in agricultural settings.

The integration of explainable AI techniques with deep learning models is particularly important in agricultural applications where end-users need to understand and trust the system's decisions. LIME provides a model-agnostic approach to generate local explanations for individual predictions, making it suitable for understanding how different regions of plant images contribute to disease classification decisions. This interpretability is crucial for agricultural practitioners who need to validate automated diagnoses and make informed treatment decisions.

This research addresses the critical need for an automated, accurate, and interpretable cauliflower disease classification system by investigating the comparative performance of VGG-19 alone versus VGG-19 combined with SVM classifier, both enhanced with LIME explanations. The study focuses on classifying four categories: three disease classes (Downy Mildew, Black Rot, and Bacterial Spot Rot) and one healthy class, using a comprehensive dataset of 7,360 digital images. The innovative contribution of this work lies in the systematic comparison of these approaches and the integration of explainable AI techniques to provide interpretable results that can build trust and facilitate adoption in real-world agricultural applications.

2. Research Method

This study employs an applied research approach aimed at developing and evaluating deep learning models for automated cauliflower disease classification. The research follows a comparative experimental design to assess the performance of two distinct approaches: VGG-19 standalone model and VGG-19 combined with Support Vector Machine (SVM) classifier, both integrated with Local Interpretable Model-Agnostic Explanations (LIME) for model interpretability [1].

The experimental framework is designed to systematically compare the classification performance of these approaches across three disease classes and one healthy class. The research design incorporates rigorous evaluation metrics including accuracy, precision, recall, and F1-score to ensure comprehensive performance assessment [2].

2.1 Data Acquisition and Dataset Description

Cauliflower plants are susceptible to several diseases including downy mildew caused by Peronospora parasitica which manifests as white, yellow, or brownish spots on upper leaf surfaces with fine grayish downy fungal growth on the lower surface [3]. Black rot caused by Xanthomonas campestris pv. campestris which initially appears as dull yellow irregular spots along leaf margins that develop into distinctive V-shaped lesions with the wide portion at the leaf edge [4]. Bacterial spot rot caused by Pseudomonas syringae pv. maculicola which creates lesions on the cauliflower head that expand and crack upon air exposure, releasing a slimy exudate that darkens to brown or black [5]. In contrast, healthy cauliflower plants produce compact, fresh white heads representing immature flower clusters at the terminal end, typically reaching 0.5 meters in height with large rounded cabbage-like leaves, and the white head remains firm, dense, and free from discoloration or lesions [6].

(a) (b) (c) (d)

Figure 1 (a) downy mildew; (b) black rot; (c) bacterial spot rot; (d) healthy cauliflower.

The dataset utilized in this research was obtained from Mendeley Data repository, specifically the "VegNet: An extensive dataset of cauliflower images to recognize the diseases using machine learning and deep learning models" collection [7]. This publicly available dataset contains pre-augmented digital images of cauliflower plants, providing a robust foundation for machine learning model development.

The dataset comprises 7,360 digital images distributed across four classes:

Table 1. Number of Datasets

Class	Images
Downy Mildew	1,770
Black Rot	1,800
Bacterial Spot Rot	1,730
Healthy Cauliflower	2,060

All images in the dataset have been preprocessed and augmented using standard computer vision techniques to enhance dataset diversity and model generalization capability [8]. The dataset is accessible through the following URL: https://data.mendeley.com/datasets/t5sssfgn2v/3.

2.2 Data Preprocessing and Partitioning

The dataset preprocessing pipeline follows established practices in deep learning for image classification tasks [9]. The complete dataset was partitioned using an 80:10:10 ratio, allocating 80% for training, 10% for validation, and 10% for testing purposes. This partitioning strategy ensures adequate data for model training while maintaining independent validation and test sets for unbiased performance evaluation [10]. Before entering into training, the model is adjusted by several parameters, namely:

Table 2. Table Parameter

Parameter	Configuration
Rescale	1./255
Batch size	32
Image Ratio	224x224
Epoch	10
Learning rate	0.0001

2.3 VGG-19 + SVM Hybrid Architecture

The VGG-19 architecture serves as the foundation for both experimental approaches in this study [11]. The hybrid approach combines VGG-19 as a feature extractor with SVM as the final classifier [12]. The implementation involves removing the fully connected layers from VGG-19 and extracting features from the last convolutional layer. These extracted features are then fed into an SVM classifier configured with One-vs-One (OvO) strategy for multiclass classification [13].

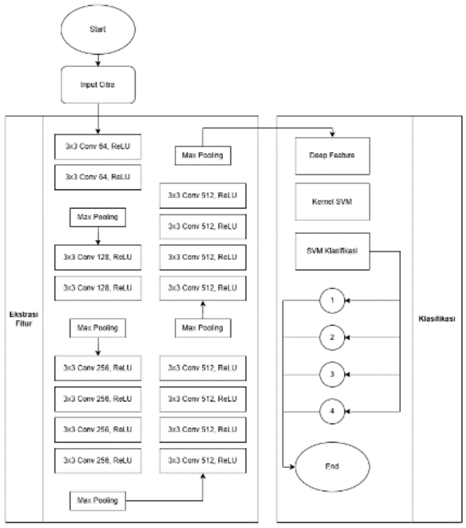

Figure 2 Model Architecture

2.4 Confusion Matrix

The evaluation framework employs comprehensive metrics to assess model performance across multiple dimensions [14]. Performance evaluation is conducted on both validation and test sets to ensure robust assessment of model generalization capabilities. After all models have been successfully executed. The results will be evaluated with a confusion matrix.

Table 3. Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual	True	TP	FN
Class	False	FP	TN

There are four types of assessments: first, true positive, which means that both the actual and predicted classes are correct. Second, true negative, which means that the predictive model is the real value but the actual is negative. Third, false negative, meaning that the predicted model is incorrect but the actual value is positive. Lastly, false positive, meaning that the predicted model is incorrect but the actual value is negative. The most common indicator include precision, recall, f1-score, and accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{F1-Score} = \frac{2XPrecisionXRecall}{Precision+Recall} \tag{4}$$

2.5 LIME

Local Interpretable Model-Agnostic Explanations (LIME) is integrated into both approaches to provide visual explanations of model predictions [15]. LIME generates local explanations by perturbing the input image and observing the impact on model predictions, highlighting image regions that contribute most significantly to classification decisions.

3. Results and Analysis

The results obtained with several stages have been an experiment. The standalone VGG-19 model has been run on patience 3 and VGG-19 + SVM hybrid architecture run on patience 5. In this experiment the class weight is used only on the train data before the model is trained to balance the dataset.

3.1 VGG19

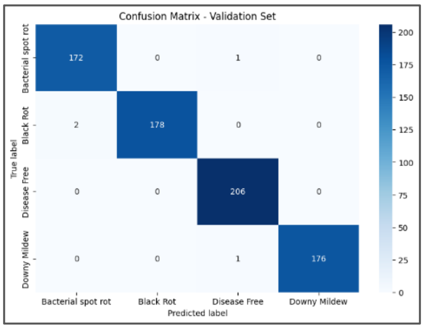

Figure 3 Confusion Matrix VGG19 (Data Validation)

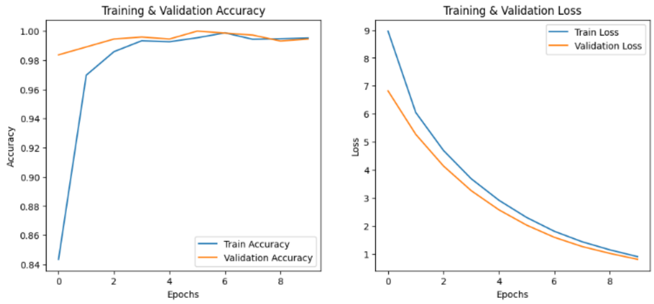

Figure 4 VGG19 Accuracy and Loss Graph (Data Validation)

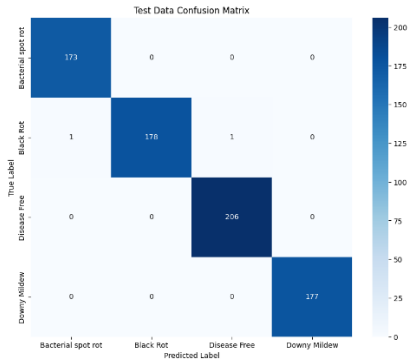

Figure 5 Confusion Matrix VGG19 (Data Test)

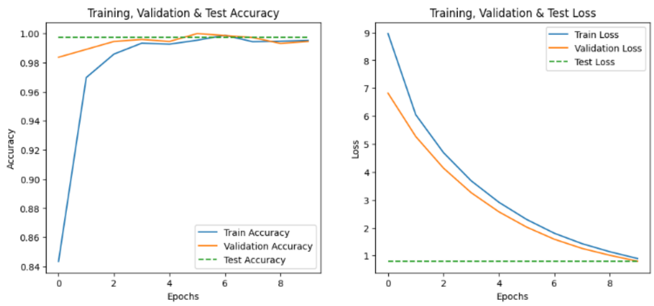

Figure 6 VGG19 Accuracy and Loss Graph (Data Test)

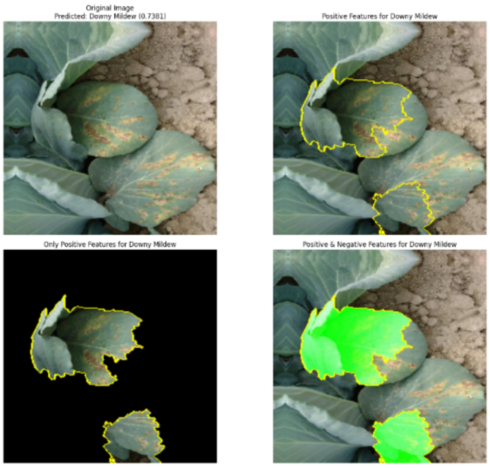

Figure 7 LIME Result for Downy Mildew Class

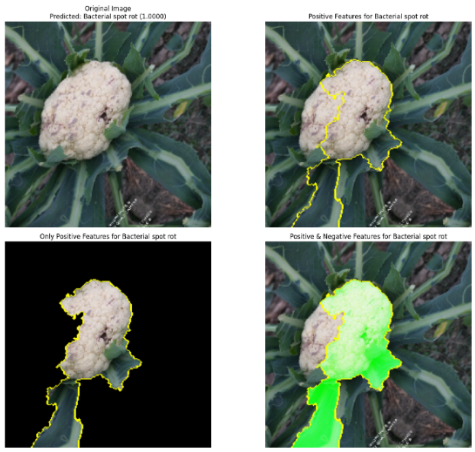

Figure 8 LIME Result for Bacterial Spot Rot Class

3.2 VGG19 + SVM

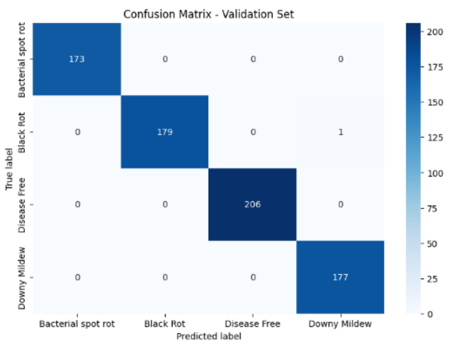

Figure 9 Confusion Matrix VGG19 + SVM (Data Validation)

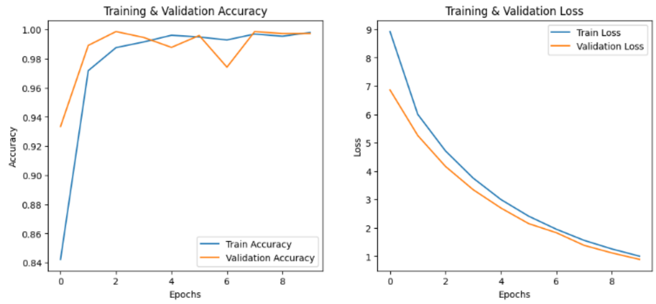

Figure 10 VGG19 + SVM Accuracy and Loss Graph (Data Validation)

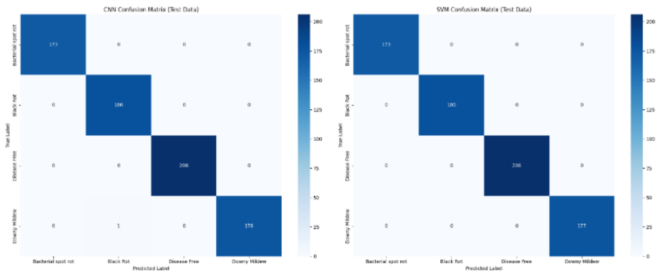

Figure 11 Confusion Matrix VGG19 + SVM (Data Test)

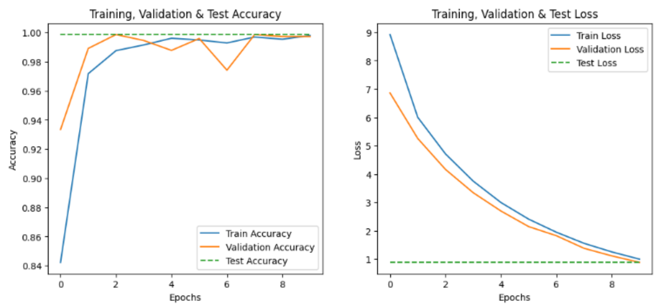

Figure 12 VGG19 + SVM Accuracy and Loss Graph (Data Test)

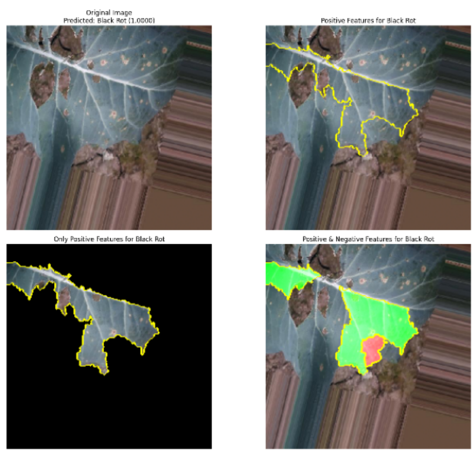

Figure 13 LIME Result for Black Rot Class

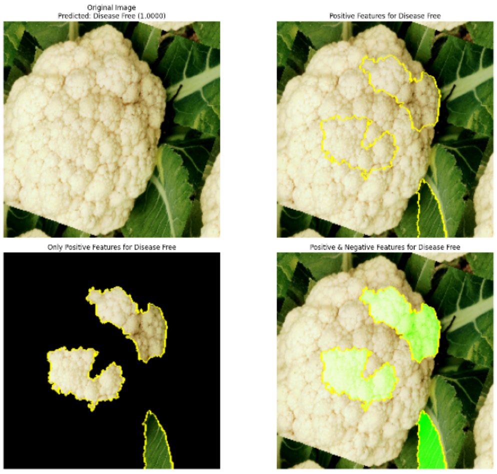

Figure 14 LIME Result for Healthy Cauliflower Class

3.3 Evaluation Matrix

Table 4 VGG19 (Data Validation)

Class	Evaluation Matrix (%)			
	Accuracy	Precision	Recall	F1-Score
Bacterial Spot Rot	99.6	99	99	99
Black Rot	100	100	99	99
Healthy Cauliflower	99.7	99	100	100
Downy Mildew	100	100	99	100
Average	99.8	99.5	99.25	99.5

Table 5. VGG19 (Data Test)

Class	Evaluation Matrix (%)			
	Accuracy	Precision	Recall	F1-Score
Bacterial Spot Rot	99.6	99	100	100
Black Rot	100	100	99	99
Healthy Cauliflower	99.7	100	100	100
Downy Mildew	100	100	100	100
Average	99.8	99.75	99.75	99.75

Table 6. VGG19 + SVM (Data Validation)

Class	Evaluation Matrix (%)			
	Accuracy	Precision	Recall	F1-Score
Bacterial Spot Rot	99.6	100	100	100
Black Rot	100	100	99	100
Healthy Cauliflower	99.7	100	100	100
Downy Mildew	100	99	100	100
Average	99.8	99.75	99.75	100

Table 7. VGG19 + SVM (Data Test)

Class	Evaluation Matrix (%)			
	Accuracy	Precision	Recall	F1-Score
Bacterial Spot Rot	99.6	100	100	100
Black Rot	100	99	100	100
Healthy Cauliflower	99.7	100	100	100
Downy Mildew	100	100	99	100
Average	99.8	99.75	99.75	100

4. Conclusion

This research successfully designed and implemented a hybrid CNN-SVM model with VGG19 architecture as a feature extractor integrated with LIME for detection and classification of diseases in cauliflower plants, where the implementation utilized VGG19 for automatic feature extraction with traditional fully connected layers replaced by SVM algorithm as classifier and integrated with LIME to provide model interpretability, demonstrating that the combination of VGG19 and SVM for cauliflower disease detection achieved superior performance compared to conventional VGG19 models with accuracy, precision, recall, and F1-Score values of 99.8%, 99.75%, 99.75%, and 100% respectively, while interpretability analysis using LIME showed satisfactory results in identifying critical features that contribute to classification decisions.

Acknowledgements

The authors would like to express their sincere gratitude to their parents for their unwavering support and encouragement throughout this research. We extend our appreciation to our supervisors and faculty members for their invaluable guidance, constructive feedback, and continuous support during this study. Special thanks are also given to our colleagues and friends who provided assistance and valuable discussions that contributed to the successful completion of this work.

References

All references must be relevant and up-to-date. References should be formatted in IEEE style. Please ensure consistency in the reference format – see examples below (10 pt):

[1]. Li, Y., Xue, J., Wang, K., Zhang, M., & Li, Z. (2022). Surface Defect Detection of Fresh-Cut Cauliflowers Based on Convolutional Neural Network with Transfer Learning. Foods, 11(18). https://doi.org/10.3390/foods11182915

[2]. O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. http://arxiv.org/abs/1511.08458

[3]. Hsu, C.-W., & Lin, C.-J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. In IEEE TRANSACTIONS ON NEURAL NETWORKS (Vol. 13, Issue 2).

[4]. Rajab, M. A., Abdullatif, F. A., & Sutikno, T. (2024a). Classification of grapevine leaves images using VGG-16 and VGG-19 deep learning nets. Telkomnika (Telecommunication Computing Electronics and Control), 22(2), 445–453. https://doi.org/10.12928/TELKOMNIKA.v22i2.25840

[5]. Nguyen, T. H., Nguyen, T. N., & Ngo, B. V. (2022). A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease. AgriEngineering, 4(4), 871–887. https://doi.org/10.3390/agriengineering4040056

[6]. Rajinikanth, V., Raj, A. N. J., Thanaraj, K. P., & Naik, G. R. (2020). A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection. Applied Sciences (Switzerland), 10(10). https://doi.org/10.3390/app10103429

[7]. Eray ÖNLER. (2022). Explaining image classification model (CNN-based) predictions with lime. World Journal of Advanced Engineering Technology and Sciences, 7(2), 275–280. https://doi.org/10.30574/wjaets.2022.7.2.0176

Biography

Utut Ardiansah is a student from University of Jember studying at the faculty of computer science.

Tio Dharmawan is lecturer who teaches the topic of computer vision thesis and helps to prepare this thesis as supervisor.