MIoSA : Machine Intelligence for Societal Advancement
Volume 1, (3) 2026, p. 110-118
E-ISSN : XXXX-XXXX

Comparative performance analysis of convolutional neural network and support vector machine for automatic object recognition based on computer vision

Ridho Bagus Satrio [a,1,*], Ilyas Nurokhim [a,2], Fandi Dwi Putra [a,3], Sayudha Hanif Saputra [a,4], Raditya Duta Pratama[a,5]

[a] Department of Informatic Engineering, STMIK Amikom Surakarta, Sukoharjo, Indonesia
[1*] ridho.130514@mhs.amikomsolo.ac.id, [2]ilyas.130512@mhs.amikomsolo.ac.id, [3]fandy.130522@mhs.amikomsolo.ac.id, [4]sayudha.130527@mhs.amikomsolo.ac.id, [5]raditya.130549@mhs.amikomsolo.ac.id
* corresponding author

Article Info

Article History:

Received July 28, 2026
Revised August 5, 2026
Accepted August 12, 2026

Keywords :
CIFAR-10, Computer Vision, Convolutional Neural Network, Object Recognition, Support Vector Machine.

Abstract

Object recognition is one of the fundamental applications of computer vision and plays an important role in various intelligent systems. This study aims to implement and compare the performance of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for automatic object recognition using the CIFAR-10 and ImageNet (subset) datasets. The novelty of this study lies in providing a comprehensive comparative evaluation of CNN and SVM across datasets with different levels of image complexity while simultaneously analyzing classification performance and computational efficiency using multiple evaluation metrics. The CNN model was implemented using the PyTorch framework with the Adam optimizer, whereas the SVM model employed the Radial Basis Function (RBF) kernel. Performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and computational time. Experimental results show that CNN consistently outperformed SVM on both datasets. On CIFAR-10, CNN achieved an accuracy of 91.8%, compared with 83.6% for SVM. On the ImageNet subset, CNN obtained an accuracy of 79.6%, whereas SVM achieved 65.4%. CNN also demonstrated higher precision, recall, and F1-score, indicating superior feature extraction and generalization capabilities for complex image classification tasks. In contrast, SVM required lower computational resources and shorter training time, making it suitable for resource-constrained environments. These findings demonstrate that CNN is the preferred approach for high-accuracy object recognition, while SVM remains a practical alternative when computational efficiency is prioritized. The results provide practical guidelines for selecting appropriate classification methods based on dataset complexity and available computational resources.

1. Introduction

The development of Artificial Intelligence (AI) and computer vision has encouraged the application of automated object recognition systems in various sectors, such as security, healthcare, manufacturing industry, autonomous vehicles, and e-commerce[1]. Object recognition systems allow computers to automatically identify and calcify objects in images and videos, improving efficiency, accuracy, and speed in the decision-making process. This capability makes the object recognition system one of the essential components in the development of intelligent systems that require real-time visual analysis to support various applications[2].

Various classification algorithms have been developed to improve the performance of object recognition systems, including Convolutional Neural Network (CNN) and Support Vector Machine (SVM)[3]. CNN is a deep learning algorithm that is able to automatically learn the representation of features

MIoSA : Machine Intelligence for Societal Advancement
Volume 1, (3) 2026, p. 110-118
E-ISSN : XXXX-XXXX

through a series of convolutional layers, thus demonstrating high performance in images with complex characteristics[4]. In contrast, SVM is an effective machine learning algorithm in classifying high-dimensional data with a relatively limited number of samples and has lower computational complexity than deep learning methods. Therefore, both algorithms are widely used as the main approaches in various object recognition studies[5].

Although studies have applied or compared CNN and SVM in image classification tasks, most studies still focus on using one type of dataset or evaluating only models based on classification accuracy[6]. In fact, other evaluation metrics, such as precision, recall, F1-score, confusion matrix, and computational time, also play an important role in describing the quality and efficiency of a classification model. In addition, there is still limited research evaluating both algorithms on datasets with different levels of complexity, so empirical evidence regarding the consistency of CNN and SVM performance in various data characteristics is still inadequate[7].

The difference in the characteristics of CNN and SVM shows that the selection of algorithms is not only determined by the degree of accuracy, but is also influenced by the complexity of the dataset, computational efficiency, and implementation needs in real applications[8]. Therefore, the evaluation of the two algorithms using datasets with different levels of complexity is necessary to obtain a more comprehensive understanding of the advantages, limitations and conditions of application of each method.

Based on the results of the study, this study conducted a comparative analysis of CNN and SVM algorithms using two public datasets that have different characteristics, namely CIFAR-10 and ImageNet (subset)[9]. The use of the two datasets aims to evaluate the consistency of the performance of the two algorithms at different levels of data complexity while identifying the influence of different data characteristics as well as identifying the influence of dataset characteristics on classification results and computational efficiency[10].

This research provides three main contributions. First, perform a comparative analysis between CNN and SVM on two data sets with different levels of complexity[11]. Second, evaluating the performance of the two algorithms using various performance metrics, including accuracy, precision, recall, F1 score, matrix confusion and computational time, resulting in a more comprehensive evaluation compared to research that used only one evaluation metric[12]. Third, to provide recommendations on the selection of the most appropriate algorithm based on the characteristics of the dataset and the need for computational efficiency in the implementation of a computer vision-based object recognition system[13].

Thus, this study aims to analyze and compare the performance of CNN and SVM algorithms in computer vision-based object recognition systems using datasets with different levels of complexity[14]. The results of the study are expected to provide a more comprehensive understanding of the advantages and limitations of each algorithm and become a reference for researchers and practitioners in choosing classification methods that are in accordance with the characteristics of the dataset and computing needs in various computer vision applications[15].

2. Research Methodology

This study uses an experimental quantitative method to analyze and compare the performance of the Convolutional Neural Network (CNN) algorithm with the Support Vector Machine (SVM) in a computer vision-based object recognition system. The quantitative method was chosen because of its focus on the numerical data analysis of the test results of the model, while the experimental design helped assess the performance of both algorithms through a series of structured steps. The research stages include dataset collection, data pre-processing, CNN and SVM modeling, training, testing, performance evaluation with various classification metrics, and comparative analysis of the results produced. The overall flow of the research is shown in Figure 1.

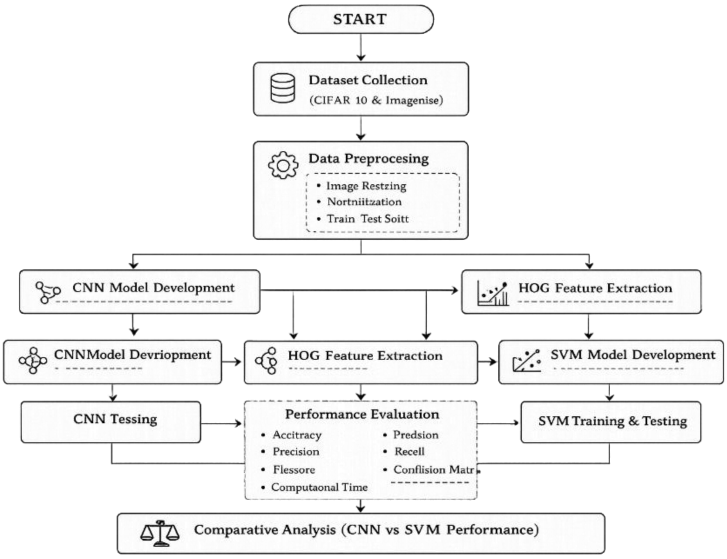

Fig. 1. Research Methodology

2.1 Dataset

This study uses two public datasets that are widely used as benchmarks in computer vision research, namely CIFAR-10 and the ImageNet subset. The use of two datasets with different levels of complexity aimed to evaluate the consistency of CNN and SVM performance on varied image characteristics. The CIFAR-10 dataset contains 60,000 32 × 32-pixel color images divided into 10 classes, with 50,000 images used for training and 10,000 images for testing. This study uses a subset of ImageNet that includes 10 categories of objects selected to reflect variations in visual complexity. The subset was chosen to reduce the computational load without sacrificing the representativeness of the dataset. All images are processed with uniform input resolution as required by the CNN model as well as the extraction feature on the SVM.

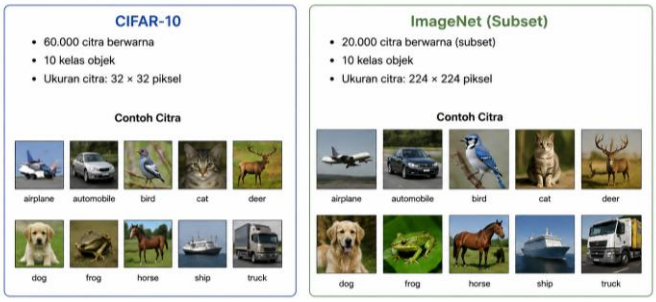

Fig. 2. Dataset

Table 1. Use of Datasets

Dataset	Image	Classes	Resolution
CIFAR-10	60.000	10	32 x 32
ImageNet Subset	20.000	10	224 x 224

2.2 Preprocessing Data

Before the model is trained, all images undergo a pre-data processing stage to ensure that the format and quality match the needs of each algorithm. In the CIFAR-10 dataset, the image resolution remained at 32 × 32 pixels, while the image on ImageNet (subset) was resized to 224 × 224 pixels to match the size of the CNN model's input.

MIoSA : Machine Intelligence for Societal Advancement
Volume 1, (3) 2026, p. 110-118
E-ISSN : XXXX-XXXX

Then, all pixel values are converted to a range of 0–1 by dividing each pixel value by 255. This normalization is intended to improve learning stability, accelerate convergence during training, and reduce the impact of scale differences between features.

In order for the CNN model to have better generalization capabilities, data augmentation was carried out on the training set using Random Horizontal Flip and Random Rotation. This approach adds to the diversity of training examples, thereby reducing the potential for overfitting and strengthening the model's resilience to changes in object orientation in the image.

Since the SVM algorithm cannot handle image data in the form of a two-dimensional matrix directly, each image is first converted into a one-dimensional vector (flatten) before being trained and classified. The stages of data cleansing applied to both datasets are summarized in Table 2.

Table 2. Preprocessing Data

Stages	CIFAR-10	ImageNet (Subset)	Description
Dataset	60,000 images, 10 classes	20,000 images, 10 classes	Research dataset
Resize	32 × 32 pixel	224 × 224 pixels	Standardize image sizes
Normalization	Pixel value 0–1	Pixel value 0–1	Speed up the training process
Data Augmentation	Random Horizontal Flip, Random Rotation	Random Horizontal Flip, Random Rotation	Adding training data variety
Data Representation	Tensor (CNN), Flatten for SVM	Tensor (CNN), Flatten for SVM	Customize the model input format
Data Sharing	50.000 train, 10.000 test	12.000 train, 4.000 validation, 4.000 test	Model training and evaluation

2.3 CNN Model Development

The CNN model was created using the PyTorch framework. The model structure includes three convolutional layers each of which is followed by the activation of the Rectified Linear Unit (ReLU) as well as the Max Pooling layer to extract the features sequentially. After that, the obtained features are streamed to the Fully Connected layer before the classification process with Softmax activation.

The model training used the Adam optimizer with a learning rate of 0.001, batch size 32, and 30 epochs. The loss function chosen is CrossEntropyLoss because it is suitable for multiclass classification. This combination of parameters is chosen to keep the training process stable and the computation time efficient.

Table 3. CNN Training Parameters

Parameter	Value
Framework	PyTorch
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epoch	30
Loss Function	CrossEntropyLoss

2.4 SVM Model Development

The SVM model was created using the Scikit-learn library which uses the Radial Base Function (RBF) kernel. Before the training stage begins, the entire image is converted into a one-dimensional vector (flatten) so that it can be processed by the SVM algorithm.

The RBF kernel was chosen because it can model the nonlinear relationships between data, making it suitable for image classification problems. The C=10 parameter is set to balance the model's generalization capabilities with classification accuracy, while Gamma is set to Scale so that the value adjusts automatically to the data distribution.

MIoSA : Machine Intelligence for Societal Advancement
Volume 1, (3) 2026, p. 110-118
E-ISSN : XXXX-XXXX

Fig. 3. CNN and SVM Implementation Comparison

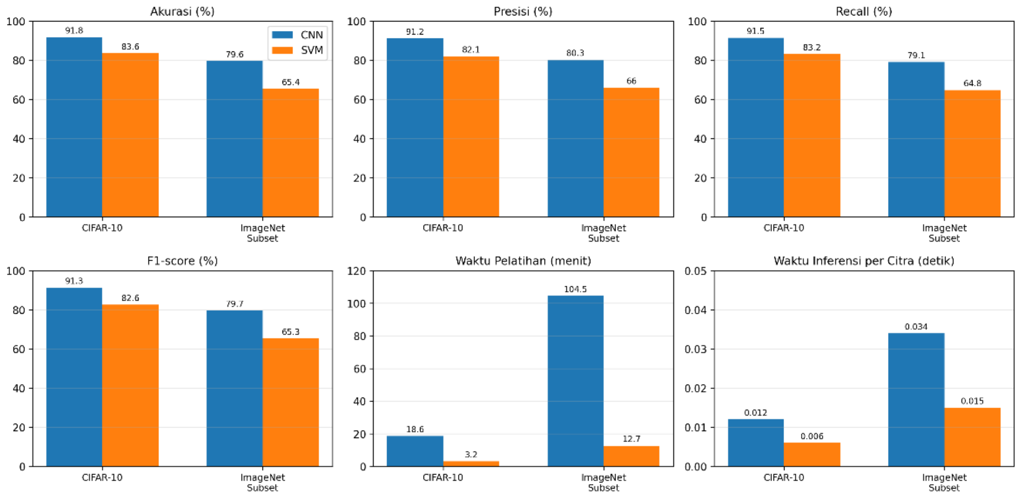

Table 4. CNN and SVM Comparison

Dataset	Accuracy (%)		Precision (%)		Recall (%)		F-1 Score (%)		Training Time (Minutes)		Inference Time (Seconds)	
	CNN	**SVM**	**CNN**	**SVM**	**CNN**	**SVM**	**CNN**	**SVM**	**CNN**	**SVM**	**CNN**	**SVM**
CIFAR-10	91.8	83.6	91.2	82.1	91.5	83.2	91.3	82.6	18.6	3.2	0.012	0.006
ImageNet (Subset)	79.6	65.4	80.3	66.0	79.1	64.8	79.7	65.3	104.5	12.7	0.034	0.015

2.5 Experimental Setup

All experiments were conducted in Python. The CNN model is implemented through PyTorch, while the SVM model is built using Scikit-learn. The pre-processing and data management stage utilizes Torchvision, and the results of the experiment are visualized using Matplotlib. All experiments were run with a consistent configuration of parameters on each dataset to ensure a fair comparison.

2.6 Experimental Setup

The performance evaluation of the two models was carried out using a number of classification metrics, including Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Computational Time. Accuracy measures the percentage of accurate predictions relative to the total test data. Precision describes how accurate the model is in classifying a class, while Recall assesses the model's ability to find all the data that does belong to that class. The F1-score serves as an indicator of the balance between Precision and Recall.

The Confusion Matrix is also used to analyze the pattern of misclassification of each class, thus giving an idea of the distribution of the predictions produced by the model. On the other hand, Computational Time is used to assess the efficiency of the training and testing process on CNN and SVM. The combination of these two metrics provides a more thorough assessment of the quality and efficiency of both algorithms. The evaluation metric is calculated by Equations (1)–(4) below:

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

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

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

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

With TP (True Positive) is the amount of data that is classified as a positive class, TN (True Negative) is the amount of data that is correctly classified as a negative class, FP (False Positive) is the amount of negative data that is incorrectly classified as positive, while FN (False Negative) is the amount of positive data that is incorrectly classified as negative.

3. Results and Discussion

3.1 Experimental Results

The experiment was conducted using two public datasets, namely CIFAR-10 and ImageNet (subset), to evaluate and compare the performance of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) algorithms in image classification tasks. The evaluation was carried out using several metrics, namely accuracy, precision, recall, and F1-score, so that the quality of the classification of each algorithm could be analyzed more comprehensively.

Table 5. Classification Performance Comparison

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN	CIFAR-10	91,8	91,2	91,5	91,3
SVM	CIFAR-10	83,6	82,1	83,2	82,6
CNN	ImageNet (Subset)	79,6	80,3	79,1	79,7
SVM	ImageNet (Subset)	65,4	66,0	64,8	65,3

Based on Table 5, CNN produces a superior classification to SVM on all evaluation metrics in both datasets. In CIFAR-10, CNN achieved an accuracy of 91.8% while SVM was only 83.6%, a difference of about 8.2%. In addition to higher accuracy, CNN also obtained better precision, recall, and F1-score, indicating the model's ability to identify objects more consistently in each class. The performance differences become more pronounced on the ImageNet subset. CNN managed to achieve an accuracy of 79.6% while SVM was only 65.4%. A more significant decline in SVM suggests that feature representations using flattening do not adequately capture complex visual characteristics, especially when the variety of objects, textures, lighting, and backgrounds increases. On the other hand, CNN can still maintain a fairly good performance because it performs feature extraction automatically through the convolution layer. In conclusion, these findings confirm that CNN has better generalization capabilities than SVM in the face of increased dataset complexity.

3.2 Confusion Matrix Analysis

In addition to relying on quantitative measures, classification quality assessment is also carried out through a confusion matrix to see the exact prediction distribution and misclassification of each object class. Based on the results of the confusion matrix, CNN produces a higher number of True Positive predictions than SVM in almost all object classes, both in the CIFAR-10 dataset and in the ImageNet subset. This indicates that CNN is able to learn more discriminating representations of features so that it can distinguish objects with similar visual characteristics. In contrast, SVM shows a higher rate of misclassification, especially in classes that have similar shapes, colors, or texture patterns. This is due to the use of a flattening feature representation, which transforms the image into a one-dimensional vector without maintaining a spatial relationship between pixels. As a result, the important visual information needed to distinguish objects cannot be utilized optimally during the classification process. These findings confirm that the automatic feature extraction mechanism in CNN provides a significant advantage over the simple pixel representation approach used by SVM.

Volume 1, (3) 2026, p. 110-118
E-ISSN : XXXX-XXXX

3.3 Computational Time Analysis

In addition to evaluating the quality of classification, this study also compares the computational efficiency of the two algorithms based on training time and inference time. The test results are presented in Table 6.

Table 6. Computational Time Comparison

Model	Dataset	Training Time (minutes)	Inference Time (seconds/images)
CNN	CIFAR-10	18,6	0,012
SVM	CIFAR-10	3,2	0,006
CNN	ImageNet (Subset)	104,5	0,034
SVM	ImageNet (Subset)	12,7	0,015

Based on Table 6, SVM requires a much shorter training time than CNN in both datasets. On CIFAR-10, SVM takes only about 3.2 minutes, while CNN takes 18.6 minutes. The same pattern is also seen on ImageNet (subset), where CNN takes up to 104.5 minutes of training, while SVM takes only about 12.7 minutes.

The difference in computational time is due to the CNN learning mechanism which involves the process of parameter optimization through forward propagation and backpropagation at each network layer. In contrast, SVM only optimizes the separation function (hyperplane) so that the computational complexity is relatively lower.

Although it takes longer training time, the classification performance improvements obtained by CNN show that the additional computational costs provide a comparable benefit to the quality of the classification results.

3.4 Discussion

Research shows that Convolutional Neural Network (CNN) consistently provides higher classification accuracy than Support Vector Machine (SVM) in both datasets tested. This advantage is mainly due to CNN's ability to automatically extract features through convolutional layers, so it can learn complex visual representations without the need for manual feature engineering. In contrast, SVMs rely on flattened representations of pixels, which are less able to maintain spatial relationships between pixels, so their performance decreases as data complexity increases.

The widening performance gap in the ImageNet subset indicates that the complexity of the data has a greater impact on traditional machine learning algorithms than on deep learning techniques. The diversity of objects, textures, lighting, and backgrounds in ImageNet makes classification tasks more difficult, giving CNN's ability to study features in layers gives them a more significant advantage over SVM.

However, computational time analysis shows a trade-off between classification quality and computational efficiency. CNN produces higher accuracy, but it requires a longer training time because the learning process involves many network parameters. In contrast, SVMs have shorter training and inference times, making them suitable for use in environments with limited computing resources or when datasets are relatively small.

The study found results consistent with previous studies that stated that CNN is superior in classification than conventional machine learning algorithms on complex image data sets because it can extract features automatically. However, these findings also indicate that SVM still has advantages in terms of computing efficiency, making it still relevant for applications that require a quick training process and have limited hardware resources. In general, the study emphasizes that the selection of algorithms should be tailored to the characteristics of the data and the needs of implementation. CNN is more recommended for object recognition applications that demand high accuracy on datasets of great complexity, while SVM can be an effective alternative when computational efficiency is a top priority.

4 Conclusion

This study compares the performance between Convolutional Neural Network (CNN) and Support Vector Machine (SVM) in a computer vision-based object recognition system using two data sets that have different levels of complexity, namely CIFAR-10 and the ImageNet subset. The results of the experiment indicated that CNN consistently produced superior classification performance over SVM on all evaluation metrics, including accuracy, precision, recall, and F1-score. This advantage indicates that CNN's ability to automatically extract features creates more effective visual representations, resulting in better generalization capabilities, especially on high-complexity data. In contrast, SVM shows advantages in computing efficiency with shorter training times, making it a suitable choice for environments with limited computing resources. The main contribution of this study is the presentation of a comparative evaluation between CNN and SVM using two datasets with different levels of complexity and various evaluation metrics, thus providing a more comprehensive understanding of the trade-off between classification accuracy and computational efficiency. These findings are expected to be a reference for researchers and practitioners in determining the classification method that best suits the characteristics of the dataset and the need for implementation in various computer vision applications. Further research is suggested to test more advanced CNN architectures, such as ResNet, EfficientNet, or Vision Transformer (ViT), as well as perform hyperparameter optimization and apply more diverse data augmentation techniques. In addition, testing on larger, more varied datasets is also needed to assess performance consistency and improve the model's generalization capabilities.

References

[1] L. Bamatraf, R. Akbar, Z. Rahman, B. Afifah, and N. Nurkholis, “Developing a Supply Chain Management Application for Laying Hens with Integrated Egg Quality Detection Based on Computer Vision (Case Study of a Laying Hen MSME),” *Indones. J. Innov. Stud.*, vol. 27, no. 1, Dec. 2025, doi: https://doi.org/10.21070/ijins.v27i1.1873.

[2] D. Putrie Lestari *et al.*, *Results of Applied Research on Computer Vision, Artificial Intelligence, and Embedded Systems*. 2017.

[3] Oriza Sativa, "Comparison Of The Classification Performance Of Support Vector Machine (SVM)," 2023.

[4] N. K. Negoro, E. Utami, and A. Yaqin, "Classification Of Detection Of Mask Use Using The Convolutional Neural Network Method," *JIPI (Journal of Ilm. Researcher. and Informed Learning.*, vol. 8, no. 2, pp. 664–674, May 2023, doi: https://doi.org/10.29100/jipi.v8i2.3748.

[5] K. R. Widyana, R. Regasari, and M. Putri, "Implementation of MobileNetV2 and Support Vector Machine (SVM) for Tangerine Fruit Automatic Sorting System Based on Raspberry Pi 4B Based on Quality Level," 2026. [Online]. Available: http://j-ptiik.ub.ac.id

[6] A. Peryanto, A. Yudhana, and R. Umar, "Image Classification Using Convolutional Neural Network and K Fold Cross Validation," 2020. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[7] T. Liu, A. Abd-Elrahman, J. Morton, and V. L. Wilhelm, “Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system,” *GIScience Remote Sens.*, vol. 55, no. 2, pp. 243–264, 2018, doi: https://doi.org/10.1080/15481603.2018.1426091.

[8] A. S. Ritonga and I. Muhandhis, "Analysis and Implementation of Viola-Jones and CNN Methods on Real-Time Sleepiness Detection Systems," *J. Ilm. IT CIDA Disem. Technology. Inf.*, vol. 10, no. 2, 2024.

[9] I. Fatahna, P. Desi Kusuma Sari, N. Kamilah, R. Wulanningrum, and W. Cahyo Utomo, "Implementation of Computer Vision on the Quality of Milk Bananas Using the YOLOv8n Method Based on WebApps," 2025. [Online]. Available: https://jurnal.polsri.ac.id/index.php/jupiter/article/view/8633/3151

[10] I. Fatahna, D. P. Pamungkas, and D. W. Widodo, "Hybrid Optimization Yolov9c-Vgg16 For Local Orange Classification In Automatic Sorting System In The Agricultural Industry," *Rabit J. Teknol.*

and Sist. Inf. Univrab, vol. 10, no. 2, pp. 217–227, Jul. 2025, doi: https://doi.org/10.36341/rabit.v10i2.6024.

[11] B. Intan, A. Taqwa Martadinata, and A. Qodir, "Ensemble Fusion-Stacking Model for Deep Feature-Based Salak Variety Classification," *Bull. Inf. Technol.*, vol. 7, no. 1, pp. 68–74, 2026, doi: https://doi.org/10.47065/bit.v5i2.2625.

[12] C. Sothe *et al.*, “Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data,” *GIScience Remote Sens.*, vol. 57, no. 3, pp. 369–394, 2020, doi: https://doi.org/10.1080/15481603.2020.1712102.

[13] A. Raisha Agma, "Application of Computer Vision for Object Detection and Facial Recognition," 2025. [Online]. Available: http://pustakajurnal.web.id/index.php/jkti

[14] G. A. W. Satia, E. Firmansyah, and A. Umami, "Design of a disease identification system in oil palm leaves (Elaeis guineensis Jacq.) with the Deep Learning Convolutional Neural Networks algorithm," *J. Ilm. Pertan.*, vol. 19, no. 1, pp. 1–10, Mar. 2022, doi: https://doi.org/10.31849/jip.v19i1.9556.

[15] A. Sentas *et al.*, “Performance evaluation of support vector machine and convolutional neural network algorithms in real-time vehicle type classification,” *Lect. Notes Data Eng. Commun. Technol.*, vol. 17, pp. 934–943, 2018, doi: https://doi.org/10.1007/978-3-319-75928-9_86.