Joinforcs: Journal of informatics and information system
Volume 1, (4) 2026, p. 150-158
E-ISSN : XXXX-XXXX

Performance evaluation of CNN and SVM in computer vision-based animal image classification

Dimas Andreas Mustofa[a,1,*], Muhammad Iqbal Tsabitul Azmi[a,2], Aldo Paristyo[a,3], Muhammad Rizky Perdana[a,4], Muhamad Hanif Kurnia Hadi[a,5]
[a] Informatics study program, STMIK AMIKOM Surakarta, Surakarta, Indonesia
[1*] dimas.130518@mhs.amikomsolo.ac.id , [2] muhammad.130521@mhs.amikomsolo.ac.id , [3] aldo.130543@mhs.amikomsolo.ac.id , [4] muh.130537@mhs.amikomsolo.ac.id, [5] muhammad.130538@mhs.amikomsolo.ac.id
* corresponding author

Article Info

Article History:

Received: July 29, 2026
Revised: August 5, 2026
Accepted: August 13, 2026

Keyword:
background removal, comparative analysis, convolutional neural network, image classification, support machine vector

Abstract

Manual animal image classification has limitations in terms of efficiency and consistency, particularly when dealing with large-scale image datasets. This study aims to compare the performance of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel for animal image classification in a computer vision system. The experiment was conducted using the Animals-10 dataset, which consists of 26,179 images, with an 80:20 ratio for training and testing data. Two experimental scenarios were evaluated, namely the original dataset and the dataset processed using background removal, to examine the impact of image preprocessing on the performance of both algorithms. The results show that CNN achieved the highest classification accuracy of 70.68%, while SVM with the RBF kernel achieved 49.50% on the background-removed dataset. Background removal improved the accuracy of SVM by 25.21%, whereas CNN showed only a 0.21% improvement. These findings indicate that background removal is more beneficial for improving SVM performance than CNN, suggesting that image preprocessing techniques should be selected according to the characteristics of the classification algorithm.

1. Introduction

Classification is the process of grouping objects based on specific traits or characteristics, thereby facilitating identification and analysis. With the advancement of computer vision, image classification is now applied in various fields, such as agriculture, healthcare, industry, and animal species recognition. By utilizing image processing techniques and artificial intelligence, systems can automatically identify objects based on the visual information contained in digital images [1]. In animal identification, image classification plays a vital role because many species bear a visual resemblance to one another, making manual methods inefficient and prone to errors, especially when processing large amounts of data [2], [3].

To build an image classification system, various strategies have been employed to achieve optimal performance. Two commonly chosen techniques include Convolutional Neural Networks (CNNs) as a deep learning method and Support Vector Machines (SVMs) as a machine learning algorithm. CNNs can automatically extract image features through a series of hierarchical convolutions, enabling them to identify complex visual patterns. On the other hand, SVMs are known for their strong classification capabilities on high-dimensional data and require less computational resources [4], [5], [6]. In designing image classification systems, various approaches have been applied to achieve optimal performance. Two commonly chosen methods include Convolutional Neural Networks (CNNs) as a deep learning technique and Support Vector Machines (SVMs) as a machine learning algorithm. CNNs automatically extract image features through a series of layered convolutions, enabling them to recognize complex visual patterns.

Joinforcs: Journal of informatics and information system
Volume 1, (4) 2026, p. 150-158
E-ISSN : XXXX-XXXX

Conversely, SVMs are known for their strong classification capabilities on high-dimensional data with lower computational demands [7], [8], [9].

In addition to the choice of algorithm, the quality of the input data also affects the performance of the classification system. The image preprocessing stage aims to improve image quality before it is processed by the model using techniques such as resizing, normalization, grayscale conversion, and background removal. This process can eliminate irrelevant information and help the model focus on extracting features of the main object [10]. In deep learning methods, CNNs can automatically derive feature representations through convolutional layers. Meanwhile, SVMs rely on feature representations generated from the preprocessing stage, so the quality of the input images significantly affects classification performance [11], [12], [13].

Based on previous research, the majority of comparisons between CNNs and SVMs still utilize objects with relatively simple visual characteristics, such as leaves, clothing, plant diseases, or text data [5], [7], [8], [9], [13]. Studies evaluating both algorithms using image datasets of animals with diverse natural backgrounds remain limited. Furthermore, the impact of background removal techniques on the performance of CNNs and SVMs on the same dataset has not been thoroughly investigated. This situation presents an opportunity to evaluate the effectiveness of the preprocessing stage in both classification approaches.

To address gaps in the literature, this study presents a comparison between CNN and SVM with a Radial Basis Function (RBF) kernel on two types of datasets: original images and images with the background removed. This method allows for a more objective analysis of the impact of preprocessing on the performance of each algorithm in classifying animal images [14], [15].

This study aims to evaluate and compare the performance of the CNN and SVM algorithms in classifying animal images by considering accuracy, computational efficiency, and the impact of background removal techniques on the performance of both methods. It is hoped that the results of this study can serve as a reference in selecting an image classification approach that is appropriate for the data characteristics in the development of computer vision systems.

2. Research Methodology

This study uses an experimental method to evaluate and compare the performance of the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) algorithms in the task of classifying animal images using computer vision. The experimental process consists of several steps, including data collection, image preprocessing, dataset splitting, model training, testing, and performance evaluation.

a. Data Acquisition: The dataset used is Animals-10, downloaded from Kaggle in .jpeg format. The dataset contains 26,179 images divided into ten animal categories: dogs, chickens, sheep, elephants, cats, horses, butterflies, spiders, cows, and squirrels. This dataset was chosen because it offers a wide variety of objects, textures, colors, and backgrounds, thereby providing a more complex image classification scenario compared to datasets with uniform backgrounds.

b. Image Preprocessing : The preprocessing stage aims to improve the quality of the input data to meet the requirements of each algorithm and to reduce irrelevant information prior to the model training process [10]. This study uses two testing scenarios—the original dataset and the dataset resulting from background removal—to evaluate the effect of background removal on classification performance.
   1. In the CNN pipeline, images are retained in RGB format, then resized to 128 × 128 pixels and normalized to the range 0–1 by dividing each pixel value by 255. This image size was chosen to achieve a balance between feature representation quality and computational efficiency during the training process.
   2. In the SVM model, images are first converted to grayscale to reduce data complexity. Next, the images are resized to 64 × 64 pixels, followed by a flattening process so that each image is represented as a vector with 4,096 features. All pixel values are then normalized to the range of 0–1 before being used as model input data.

c. Data Splitting : The dataset was split into a training set and a test set in an 80:20 ratio. Approximately 20,947 images were allocated for training, while 5,232 images were allocated for testing. The split was performed randomly to create a representative sample distribution, ensuring that training and evaluation could proceed objectively.

d. Model Training : Training was conducted separately for each algorithm using the prepared training data.
   1. The CNN model was built using a sequential architecture consisting of three convolutional layers with 32, 64, and 128 filters, respectively, using a 3 × 3 kernel size and the Rectified Linear Unit (ReLU) activation function. Each convolutional layer is followed by 2 × 2 MaxPooling2D to reduce the feature dimensions while retaining the most representative information. After the Flatten process, the features are processed by a Dense Layer consisting of 128 neurons with the ReLU activation function, followed by the application of a 0.5 Dropout rate to reduce the risk of overfitting. The output layer uses 10 neurons with a Softmax activation function, which generates probabilities for each image class [14]. Training uses the Adam optimizer, the Categorical Cross-Entropy loss function, a batch size of 32, and runs for 20 epochs.
   2. The Support Vector Machine (SVM) model is implemented using the Scikit-Learn library with a Radial Basis Function (RBF) kernel. The RBF kernel was chosen because it can model nonlinear relationships between features, making it suitable for classifying images with complex visual characteristics. The model was trained using preprocessed data to obtain a hyperplane capable of optimally separating each image class.

e. Experimental Environment: All experiments were run in Google Colaboratory using the Python 3 programming language. The CNN was implemented using the TensorFlow/Keras libraries, while the SVM was implemented using Scikit-Learn. The training process utilized NVIDIA T4 GPU acceleration with 12.7 GB of system memory to improve computational efficiency.

f. Model Evaluation: The evaluation was conducted using test data to compare the performance of the two algorithms based on accuracy, precision, recall, F1-score, training time, and testing time. The evaluation metrics were calculated based on a confusion matrix consisting of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

$$Accuracy \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision \frac{TP}{TP + FP}$$

g. Comparative Analysis: The final stage of the research involves a comparative analysis of the CNN and SVM test results across both dataset scenarios. This analysis considers accuracy, precision, recall, F1 score, training time efficiency, and testing time to assess the impact of background removal techniques on each algorithm. The findings of this analysis will serve as a reference in selecting the method with the optimal performance for classifying animal images using computer vision. The research flowchart is shown in Figure 1.

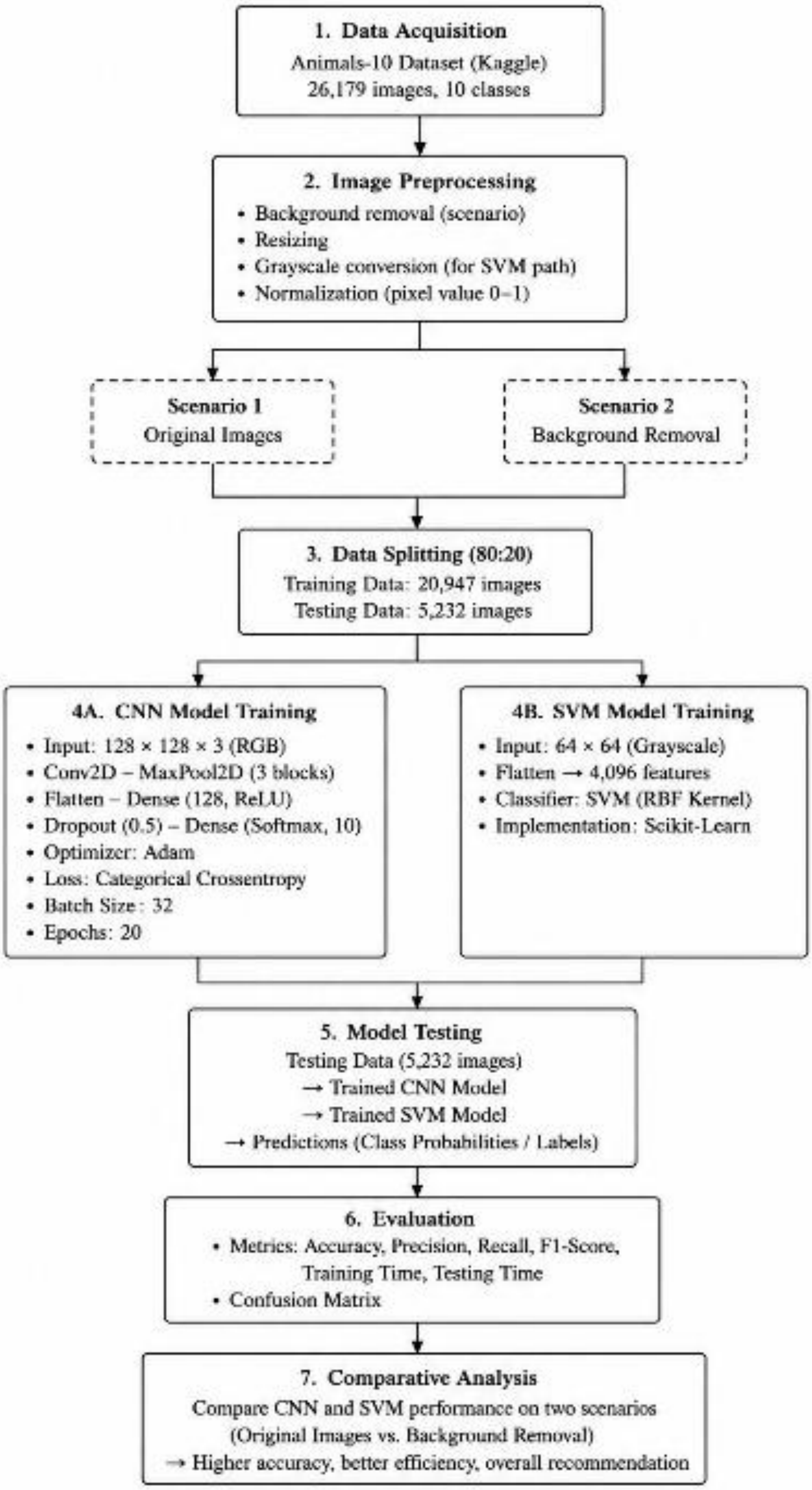

Figure 1. Research Flowchart

Joinforcs: Journal of informatics and information system
Volume 1, (4) 2026, p. 150-158
E-ISSN : XXXX-XXXX

3. Results and Discussion

3.1 Classification Performance

The performance of the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) algorithms was evaluated using two test scenarios: the original dataset and the dataset after background removal. The evaluation focused on accuracy, training time, and testing time to determine the impact of preprocessing techniques on the performance of each algorithm.

Table 1. Performance Comparison of CNN and SVM

Algorithm	Dataset Condition	Accuracy	Training Time	Test Time
CNN	Original dataset	70.47	12 min 08 s	8 s
CNN	No background	70.68	13 min 14 s	14 s
SVM(RBF)	Original dataset	24.29	3250.96 s	319.90 s
SVM(RBF)	No background	49.50	2016.18 s	518.54 s

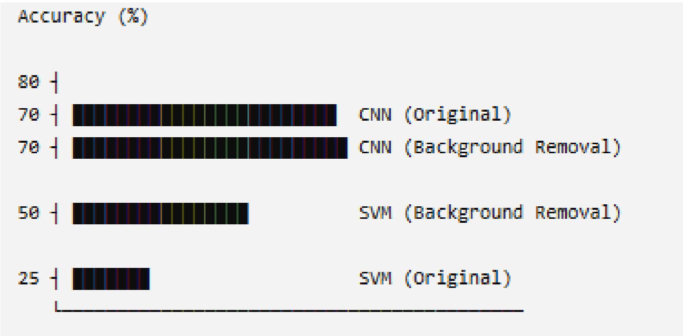

Figure 2. Accuracy Comparison Between CNN and SVM

Based on Table 1 and Figure 2, it can be seen that the CNN algorithm produces superior classification results compared to SVM in both testing scenarios. The CNN's accuracy was 70.47% on the original dataset and increased to 70.68% after the background was removed. This 0.21% increase indicates that the background removal process has a relatively small impact on the CNN's performance. In contrast, the SVM with an RBF kernel showed a much greater improvement in performance following the preprocessing stage. The model's accuracy increased from 24.29% on the original dataset to 49.50% on the background-free dataset—a 25.21% increase. These findings confirm that the quality of the input images has a more significant impact on SVM performance than on CNN performance.

From a computational time perspective, the CNN training process was completed in approximately 12–13 minutes, while the SVM took between 33 and 54 minutes. This difference is due to the use of GPU acceleration in the CNN implementation during training, whereas the SVM, which must optimize high-dimensional data, requires a longer computation time. During the testing phase, the CNN also demonstrated significantly shorter inference times compared to the SVM. Overall, the test results revealed that the CNN provided more consistent classification performance on both datasets, while the SVM experienced a significant increase in accuracy after applying the background removal technique. These findings indicate that the preprocessing stage affects the two algorithms differently, in line with their respective feature extraction mechanisms.

3.2 Experimental Result

This study evaluates the impact of background removal methods on image classification accuracy using convolutional neural networks (CNNs) and support vector machines (SVMs) with Radial Basis Function (RBF) kernels. The experimental findings reveal that the effects of preprocessing vary across algorithms, indicating that the characteristics of each model significantly influence how visual data in images is utilized. Figure 3 presents several examples of images before and after the background removal process used in this study.

Joinforcs: Journal of informatics and information system
Volume 1, (4) 2026, p. 150-158
E-ISSN : XXXX-XXXX

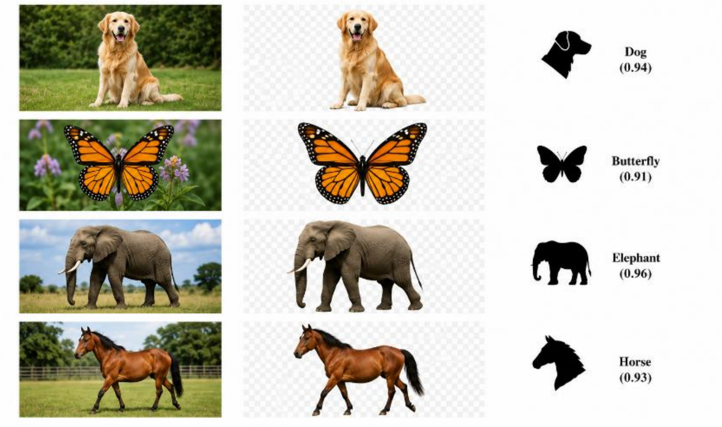

Figure 3. Example of Background Removal

In the CNN algorithm, the use of background removal only increased accuracy from 70.47% to 70.68%, a 0.21% increase. This small improvement indicates that the CNN can still maintain its classification performance even when images contain a background. This is due to the hierarchical feature extraction mechanism in the convolutional layers, which allows the model to learn object characteristics such as shape, texture, and visual patterns while reducing the influence of less relevant information. Thus, the presence of a background does not significantly impact the CNN's ability to recognize animal objects.

Unlike CNN, the SVM algorithm showed a much more significant improvement in performance after background removal. Accuracy rose from 24.29% to 49.50%, representing an increase of 25.21%. These findings indicate that SVM is more sensitive to the quality of input data because its classification process relies on feature representations generated during the preprocessing stage. With the background removed, the amount of information unrelated to the main object is reduced, making the features used to form the hyperplane more representative. This allows the model to achieve better class separation compared to when the images still contain varied backgrounds.

The difference in response between the two algorithms indicates that the success of preprocessing techniques is highly dependent on the learning mechanism used. CNNs can learn feature representations automatically, so changes in the background have only a limited impact on classification results. In contrast, SVMs do not perform automatic feature extraction, so the quality of the input feature representations becomes the primary factor determining model performance.

This study found that removing the background is an effective preprocessing step for improving the performance of SVM algorithms in animal image classification, while its impact on CNNs is barely noticeable. Therefore, the choice of preprocessing method should be tailored to the characteristics of the algorithm used so that the classification process can achieve optimal performance.

3.3 Computational Efficiency

In addition to assessing classification accuracy, this study also compares the computational efficiency of the CNN and SVM algorithms based on training time and testing time. The purpose of this evaluation is to provide an overview of the computational requirements of each algorithm in the animal image classification process. Figure 4 shows a comparison of training and testing times for CNN and SVM under both experimental scenarios.

Joinforcs: Journal of informatics and information system
Volume 1, (4) 2026, p. 150-158
E-ISSN : XXXX-XXXX

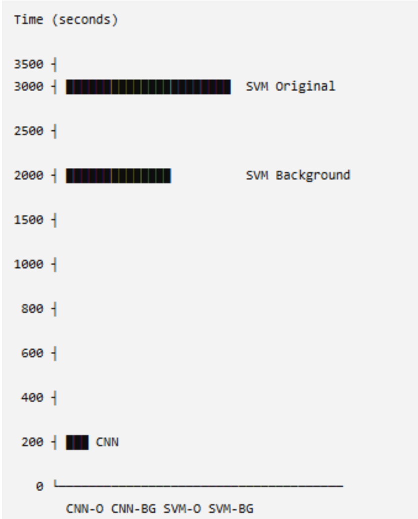

Figure 4. Comparison of Training and Testing Time

According to the test results, training the CNN model took approximately 12–13 minutes, while the SVM model took between 33 and 54 minutes. This difference indicates that CNNs can complete the training process faster than SVMs under the same experimental configuration. Although the CNN architecture is more complex, training leverages GPU acceleration with TensorFlow, allowing computations to be performed in parallel and more efficiently.

In the SVM algorithm, the optimization stage must construct a hyperplane using all the features resulting from preprocessing. As the amount of data and the number of feature dimensions increase, the optimization process becomes more complex, leading to longer training times. This was evident in tests using the original dataset, which required a longer training time compared to the dataset that had undergone background removal.

From the perspective of testing time, the CNN demonstrated superior performance with inference times ranging from 8 to 14 seconds, while the SVM required approximately 320 to 519 seconds. This difference indicates that CNNs are more efficient at making predictions on the test data within the computational environment used in this study. The evaluation shows that CNNs not only achieve higher classification accuracy but also offer better computational efficiency compared to SVMs on the Animals-10 dataset. Therefore, CNNs are more suitable for application in image classification systems that require efficient training and prediction processes, while SVMs can still be a viable option when simpler features or relatively limited data are available.

3.4 Discussion

This study shows that convolutional neural networks (CNNs) produce superior classification results compared to support vector machines (SVMs) under all testing conditions. CNNs maintain a nearly constant accuracy rate both on the original dataset and after background removal. In contrast, SVMs exhibit a significant increase in accuracy after the background is removed. The results indicate that the effectiveness of the preprocessing step depends on the characteristics of the algorithm used.

This difference in performance stems from how each algorithm learns. CNNs automatically extract features through convolutional layers, allowing them to learn visual representations in a hierarchical manner—from basic features such as edges and textures to more complex patterns. This advantage makes CNNs more flexible in handling variations in background and lighting. On the other hand, SVMs are highly dependent on the quality of the preprocessed features, so a complex background can reduce the model's ability to form an optimal hyperplane. Therefore, removing the background results in a more significant performance improvement for SVMs compared to CNNs.

The findings of this study are consistent with a number of previous studies showing that CNNs excel at image classification because they can automatically extract features, while SVM performance is heavily influenced by the quality of the input features and the preprocessing stage. However, this study makes a further contribution by testing both algorithms on the Animals-10 dataset under two different image conditions: with and without background removal. This approach provides a clearer picture of the impact of preprocessing techniques on each algorithm in the animal image classification process.

Based on the implementation, the study's findings indicate that CNNs not only achieve higher accuracy but also demonstrate superior computational efficiency in the experimental environment used. This suggests that CNNs are better suited for image classification systems that require high accuracy and fast inference. Conversely, the increase in accuracy achieved by SVMs after background removal indicates that this algorithm still has potential for use under certain conditions, particularly when supported by more discriminative feature representations or when data is limited.

This study has several limitations. First, the experiment used only one dataset, namely Animals-10, so it cannot yet assess the model's ability to generalize to other datasets. Second, the study tested only one CNN architecture configuration and one type of SVM kernel, so it does not yet demonstrate the algorithm's performance under different configurations. Third, the evaluation remains focused on classification metrics and computational efficiency without including statistical analysis to compare the performance differences between the two models.

For future research, the evaluation can be improved by using a more diverse dataset, comparing various state-of-the-art CNN architectures such as ResNet, EfficientNet, or MobileNet, and investigating various SVM kernels and more representative feature extraction methods. Furthermore, the use of statistical tests on the experimental results can provide more convincing evidence regarding the significance of performance differences between algorithms.

The summary provided in Table 2 confirms that CNNs consistently achieve superior classification performance and computational efficiency compared to SVMs on the Animals-10 dataset. Meanwhile, the substantial improvement observed in SVMs after background removal highlights the importance of image preprocessing for traditional machine learning approaches. These findings provide practical insights for selecting the appropriate classification method based on the characteristics of the dataset and application requirements.

Table 2. Summary of CNN and SVM Performance

Aspect	CNN	SVM (RBF)
Classification Accuracy	Higher	Lower
Effect of Background Removal	Small improvement (0.21%)	Significant improvement (25.21%)
Training Time	Faster	Slower
Testing Time	Faster	Slower
Feature Learning	Automatic	Depends on preprocessing
Computational Complexity	Higher model complexity, efficient with GPU	Simpler model, higher computation on high-dimensional data

4. Conclusion

This study compared the performance of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel for animal image classification using the Animals-10 dataset under two experimental scenarios, namely the original dataset and the dataset processed with background removal. The experimental results demonstrate that CNN consistently outperformed SVM in terms of classification accuracy and computational efficiency. CNN achieved the highest accuracy of 70.68%, while SVM achieved 49.50% after the application of the background removal preprocessing technique.

The experimental findings also indicate that the impact of image preprocessing differs depending on the characteristics of the classification algorithm. Background removal provided only a marginal improvement for CNN (0.21%) because the network is capable of automatically extracting discriminative visual features through hierarchical convolutional layers. In contrast, SVM showed a substantial improvement (25.21%) after background removal, indicating that traditional machine learning methods are more dependent on the quality of input feature representations than deep learning approaches.

Overall, this study demonstrates that CNN is a more suitable approach for animal image classification involving complex visual characteristics, while SVM can still provide competitive performance when supported by appropriate preprocessing techniques. Future research may consider evaluating additional CNN architectures, investigating alternative SVM kernels, applying feature extraction methods prior to classification, and validating the proposed approach using larger and more diverse image datasets to improve model generalization.

REFERENCES

[1] R. Febrian, N. Putra, P. Manurung, M. Syifauz, Z. Arrifki, and D. Oktafiani, "Seminar Nasional Amikom Surakarta (Semnasa) 2025 Klasifikasi Citra Pakaian Menggunakan Convolution Neural Network (CNN) Berbasis MobileNetV2".

[2] N. Kristiawan and M. Sugeng, "Klasifikasi Jenis Hewan pada Kebun Binatang dengan Menggunakan Metode Deep Neural Network," *Jurnal Infokam*, vol. XVIII, no. 1, pp. 46–51, Mar. 2022, doi: https://doi.org/10.53845/infokam.v18i1.317.

[3] M. Murdifin and S. Uyun, "Klasifikasi Hewan Anjing, Kucing, dan Harimau Menggunakan Metode Convolutional Neural Network (CNN)," 2025. doi: https://doi.org/10.14421/jiska.2025.10.3.331-340

[4] R. Hidayat, Y. S. Sy, T. Sujana, M. Husnah, H. T. Saputra, and F. Okmayura, "Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 2, pp. 161–168, Sep. 2024, doi: https://doi.org/10.37148/bios.v5i2.152.

[5] S. Mulyana and I. Muthmainnah, "Analisis Perbandingan Tingkat Akurasi Algoritma Cnn Dan Svm Dalam Klasifikasi Pada Daun Gedi, Daun Pepaya Dan Daun Ubi," 2024. doi: https://doi.org/10.36040/jati.v8i4.10144

[6] A. Faizah, S. Imron, A. Rahel, P. Rewur, J. N. Makasunggal, And H. Saputro, "Jurnal Informatik Edisi ke-21, Nomor 1," 2025. doi: https://doi.org/10.52958/iftk.v21i1.11091

[7] M. Saputra, V. J. Maulana, D. I. Budiman, and F. Sugiarto, "Deteksi awal penyakit tanaman berbasis citra daun: Studi perbandingan algoritme SVM dan CNN," 2025.

[8] P. A. Prayesy, "Studi Perbandingan Metode Support Vector Machine, Random Forest, Dan Convolutional Neural Network Untuk Klasifikasi Penyakit KULIT," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 1, pp. 70–76, Jan. 2025, doi: https://doi.org/10.69916/jkbti.v4i1.214.

[9] R. Mardianto, Stefanie Quinevera, and S. Rochimah, "Perbandingan Metode Random Forest, Convolutional Neural Network, dan Support Vector Machine Untuk Klasifikasi Jenis Mangga," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 63–71, May 2024, doi: https://doi.org/10.52158/jacost.v5i1.742.

[10] A. Simarmata, A. Putra, and A. Husein, "Penerapan Metode Computer Vision Dalam Klasifikasi Buah Jeruk Menggunakan Teknik Image Pre-Processing," *Data Sciences Indonesia (DSI)*, vol. 3, no. 2, pp. 108–114, Jun. 2024, doi: https://doi.org/10.47709/dsi.v3i2.4010.

[11] N. K. Negoro, E. Utami, and A. Yaqin, "Klasifikasi Deteksi Penggunaan Masker Menggunakan Metode Convolutional Neural Network," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 664–674, May 2023, doi: https://doi.org/10.29100/jipi.v8i2.3748.

[12] B. W. Kurniadi, H. Prasetyo, G. L. Ahmad, B. Aditya Wibisono, and D. Sandya Prasvita, *Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah*. 2021.

[13] N. Simanjuntak and A. H. Muhammad, "Analisis Perbandingan Algoritma SVM dan CNN dalam Mendeteksi Website Judi Online Berdasarkan Konten Teks," *Bulletin of Computer Science Research*, vol. 5, no. 4, pp. 361–371, Jun. 2025, doi: https://doi.org/10.47065/bulletincsr.v5i4.586.

[14] D. Aditya Putra, F. Tri Anggraeny, and C. Aji Putra, "Seminar Nasional Informatika Bela Negara (SANTIKA) Klasifikasi Citra Ras Anjing Menggunakan Metode Convolutional Neural Network (CNN)".

[15] B. Baso, N. Huda, U. Timor, K. Timor Tengah Utara, and P. Korespondensi, "Hybrid Algoritma Convolutional Neural Network Dengan Support Vector Machine Untuk Klasifikasi Jenis Tenun Timor Hybrid Algorithm Of A Convolutional Neural Network With A Support Vector Machine For The Classification Of The Type Of Timor Weaving Features," Vol. 12, No. 6, Pp. 1233–1242, 2025.