OPEN ACCESS

Volume 10 (4), October 2026, 1431-1443

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v10i4.6905

Pikachu Image Classification Using Neural Network and Support Vector Machine with Painters, VGG-16, and Inception-V3 Feature Representations

Muthia Andriana Putri [1*], Imam Yuadi [2]

[1,2] Universitas Airlangga, Surabaya City, East Java Province, Indonesia.*

a r t i c l e i n f o

Article history:
Received 6 February 2026
Received in revised form
3 March 2026
Accepted 25 April 2026
Available online October
2026.

Keywords:
Image Classification; Transfer Learning; Convolutional Neural Networks; Support Vector Machine; Feature Representation.

Kata Kunci:
Klasifikasi Citra; Transfer Learning; CNN; SVM; Fitur Visual.

a b s t r a c t

This study aims to evaluate the effectiveness of multiple feature representations in classifying Pikachu images into three distinct visual categories: anime, action figures, and hand-drawn illustrations. The primary challenge lies in limited data availability and the high visual variability across styles, resulting in significant inter-class similarity and intra-class diversity. To address this issue, the study employs a transfer learning approach utilizing pre-trained Convolutional Neural Networks (CNNs), namely VGG-16 and Inception-V3, alongside painterly feature descriptors. The dataset comprises 351 images collected from open-access sources with balanced class distribution. Extracted features are subsequently classified using Support Vector Machines (SVM) and shallow Neural Networks. The findings demonstrate that integrating deep semantic features with artistic representations significantly improves classification accuracy compared to single-feature approaches. These results highlight the critical role of hybrid feature engineering and classifier selection in achieving robust image classification performance under data-constrained conditions.

a b s t r a k

Penelitian ini bertujuan untuk mengevaluasi efektivitas berbagai representasi fitur dalam klasifikasi citra Pikachu berdasarkan tiga kategori visual, yaitu anime, action figure, dan gambar tangan. Permasalahan utama terletak pada keterbatasan data dan kompleksitas variasi gaya visual yang menyebabkan tingginya kemiripan antar kelas serta keragaman dalam kelas. Penelitian ini menggunakan pendekatan transfer learning melalui model Convolutional Neural Network (CNN) pralatih, yaitu VGG-16 dan Inception-V3, serta fitur artistik (painters features). Dataset terdiri dari 351 citra yang dikumpulkan dari sumber terbuka dengan distribusi seimbang. Proses klasifikasi dilakukan dengan mengintegrasikan fitur yang diekstraksi ke dalam algoritma Support Vector Machine (SVM) dan jaringan saraf tiruan. Hasil penelitian menunjukkan bahwa kombinasi fitur mendalam dan fitur artistik mampu meningkatkan akurasi klasifikasi secara signifikan dibandingkan pendekatan tunggal. Temuan ini menegaskan pentingnya pemilihan representasi fitur dan algoritma klasifikasi dalam kondisi data terbatas untuk meningkatkan kinerja sistem klasifikasi citra.

Corresponding Author. Email: muthia.andriana.putri-2025@pasca.unair.ac.id [1].

1. Introduction

Image classification is one of the most dynamic fields of study in computer vision. Its versatility facilitates many applications such as object detection, multimedia processing and analysis, and comprehension of digital text. An unprecedented increase in the visual information available on the internet in recent years calls for powerful and refined classification models capable of addressing the complexity of visual information. Handcrafted visual feature extraction techniques fail to capture the complexity of visual patterns and styles. As a result, most researchers rely on deep learning techniques for feature extraction, particularly Convolutional Neural Networks (CNNs), which automatically create hierarchical features from raw image data. Consequently, CNNs are able to produce state-of-the-art results on a variety of large-scale datasets such as ImageNet (M. Chen *et al.*, 2020; Jastrzebska, 2022; Liu *et al.*, 2020). Despite the advances described, it is still sometimes impossible to train deep CNN models from scratch. In many real-world situations, the availability of extensive annotated datasets, as well as the requisite significant computational power, is lacking. In response to this issue, the academic community has favored a novel technique in the last decade—transfer learning.

This approach, in essence, allows us to exploit learning from large datasets to solve other, related problems. Therefore, in situations where large datasets cannot be used, models such as VGG-16 and Inception-V3 are utilized as feature extractors. The process of transfer learning, particularly the use of frozen or previously trained convolutional layers, enables us to obtain meaningful representations of the data, which is particularly useful given the limited amount of training data available (Raghu *et al.*, 2021; Zhang *et al.*, 2020). Along with feature extraction, the selection of the right classification algorithm is critical to the overall system performance. While end-to-end deep neural networks are very common, numerous works show that even classical machine learning techniques can effectively classify the deep features obtained from CNNs. One of these techniques, the Support Vector Machine (SVM), is particularly noted for its strength in high-dimensional feature spaces and generalizes well in cases where the training data is sparse. In parallel, shallow neural networks are less expensive to implement than deep architectures and offer more computationally manageable nonlinear decision boundaries (H. Chen *et al.*, 2023; Law *et al.*, 2020). Based on preliminary experiments, this study selects SVM and Neural Network classifiers because they consistently outperform Random Forest and Gradient Boosting models in terms of classification accuracy. Considering this methodological framework, this study focuses on the classification of Pikachu images divided into three distinct categories: anime, action figures, and hand-drawn images. Anime, action figures, and hand-drawn images all represent the same character; however, each collection possesses a distinct type of visual style. Anime images typically consist of smooth surfaces and stylized outlines, action figures display three-dimensional representations and realistic lighting, and hand-drawn images tend to present peculiar shapes and artistic distortions. These visual contrasts contribute to inter-class similarity and intra-class diversity, thus offering complexity to the classification of the data, making it an optimal challenge for testing the robustness of features (Iqbal *et al.*, 2021; Visalli *et al.*, 2021).

For this purpose, the dataset for this study contains 351 images of Pikachu, with 117 images for each category, to maintain an even distribution of the classes. The images were collected from freely accessible sites such as Google Images and Pinterest to represent diverse artistic styles, as well as different backgrounds and image qualities. Such an approach to addressing the dataset creation problem accurately represents the challenges associated with image classification tasks, where the input images come from numerous heterogeneous and uncontrolled sources. This study utilizes three distinct image embedding techniques—Painters features, VGG-16, and Inception-V3—to proficiently capture both stylistic and semantic information. Painters features are used to show artistic and stylistic traits, which are especially important for anime and hand-drawn pictures. Conversely, VGG-16 and Inception-V3 deliver profound semantic representations via transfer learning, with VGG-16 providing consistent deep features and Inception-V3 encapsulating multi-scale spatial information (Nnamoko *et al.*, 2022; Shen *et al.*, 2023).

This study systematically assesses the impact of various feature representations on classification performance by integrating these embeddings with SVM and Neural Network classifiers. The primary aim of this paper is to evaluate the efficacy of Painters, VGG-16, and Inception-V3 feature representations in conjunction with Support Vector Machine and Neural Network classifiers for the classification of Pikachu images. This study offers empirical insights into the amalgamation of artistic and deep feature representations with classical classifiers through extensive experimental evaluation. In the end, the results help us better understand how to classify stylized images when there isn't much data available. Lastly, this paper is structured as follows: In Section 2, we discuss the research methodology, which includes how we prepared the dataset, how we extracted features, and how we built classification models. Section 3 presents the results of the experiments and discusses them. The main findings of the paper are summarized in Section 4, which is the last section. There are acknowledgments in Section 5, an author's note in Section 6, and a list of references in Section 7.

2. Research Methodology

The research utilizes a specifically created dataset of Pikachu images aimed at demonstrating different visual representations of the same character. The dataset consists of 351 images, equally divided into three categories: Pikachu anime, Pikachu action figure, and Pikachu hand-drawn. Each category contains 117 images (Figure 1). The even class distribution ensures that the classification accuracy is not skewed toward one class, allowing the models to be assessed equally.

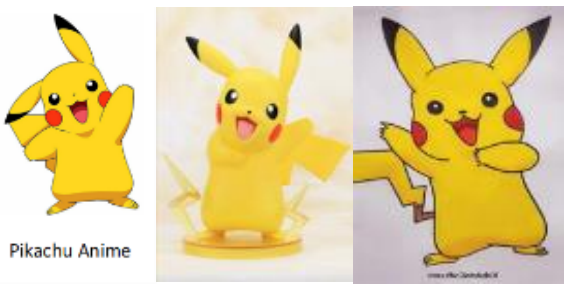

Figure 1. Three Images of Pikachu from Various Sources

Anime images are two-dimensional and resemble cartoons, characterized by simple textures and stylized borders. Action figures, on the other hand, are plastic, three-dimensional representations of Pikachu often depicted in real-life photographs where light, depth, and shadows are visible. The hand-drawn category consists of images created by hand, featuring irregular formations, artistic anomalies, and varying techniques. These contrasts result in significant differences across categories while maintaining a common identity for Pikachu anime, Pikachu figure, and Pikachu hand-drawn images, rendering the dataset ideal for testing feature robustness (Chaturvedi *et al.*, 2020; Zhao *et al.*, 2021). The images were sourced from publicly accessible platforms, primarily Google Images and Pinterest, which are two of the most common places to find visual data for exploratory image classification studies. This method of data collection ensures that the backgrounds, poses, resolutions, and artistic styles vary, closely resembling how images differ in the real world (Khan *et al.*, 2020). Prior to the experiments, each image was manually reviewed to ensure correct labeling and good image quality.

Data Processing

Before extracting features, the input images underwent a data processing stage to standardize them. To ensure compatibility with deep learning models, all images were resized to match the input sizes required by the models, making them suitable for VGG-16 and Inception-V3 architectures (Pinaya *et al.*, 2020; Tsipras *et al.*, 2020). To stabilize the learning process and reduce lighting discrepancies, the pixel values were normalized. Images were also organized into folders based on their category—anime, action figure, or hand-drawn art. This structured organization simplifies the feature extraction pipeline and ensures consistent labeling across all experiments. We refrained from using aggressive data augmentation to preserve the original artistic qualities of each image, which are crucial for evaluating the effectiveness of stylistic feature representations. This preprocessing method aims to retain important visual information while eliminating irrelevant variations. By maintaining uniform preprocessing across all embedding methods, any differences in classification performance can be attributed to the feature representations and classifiers rather than inconsistencies in the input data.

Image Embedding and Feature Extraction

This study employs three image embedding methods—Painters features, VGG-16, and Inception-V3—to capture both the style and meaning of the images. Painter-based features illustrate artistic and stylistic traits such as color distribution, stroke patterns, and texture, which are particularly useful for distinguishing between anime and hand-drawn images. Artistic feature representations have proven effective in style-aware image analysis tasks. VGG-16, a convolutional neural network with several stacked convolutional layers and small receptive fields, is utilized to extract both artistic and deep semantic features. In this research, VGG-16 is used as a static feature extractor by removing the final classification layers and utilizing the output from the last convolutional block as a feature vector. This approach allows for obtaining high-level visual representations without requiring fine-tuning of the model (M. Chen *et al.*, 2020). Additionally, Inception-V3 is employed to extract deep features at various scales. Its design incorporates parallel convolutional operations with different kernel sizes, enabling it to efficiently capture spatial and contextual information at multiple scales. Inception-V3 is applied in a transfer learning context with frozen weights, similar to VGG-16. Previous research has indicated that Inception-based embeddings often outperform simpler architectures in managing intricate visual variations (Deng *et al.*, 2009; Kumar *et al.*, 2020).

Classification Models

To identify the most effective classification method, we initially tested four models: Support Vector Machine (SVM), Random Forest, Neural Network, and Gradient Boosting. We evaluated these models using features derived from Painters, VGG-16, and Inception-V3 embeddings. The evaluation results demonstrated that SVM and Neural Network consistently exhibited better accuracy and more stable performance compared to Random Forest and Gradient Boosting. The Support Vector Machine classifier aims to identify an optimal hyperplane that maximizes the margin between classes in a high-dimensional feature space. Formally, given training data $\{(x_i, y_i)\}_{i=1}^{N}$, SVM solves the following optimization problem (Cortes & Vapnik, 1995):

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

where $w$is the weight vector, $b$is the bias term, $\xi_i$are slack variables, and Ccontrols the trade-off between margin maximization and classification error. SVM is particularly effective for high-dimensional deep feature vectors. The Neural Network classifier employed in this study is a feedforward model that maps input features to output class probabilities through non-linear transformations. Given an input feature vector x, the network computes:

$$y = f(Wx + b)$$

where $W$and $b$represent weights and biases, and $f(\cdot)$is a non-linear activation function such as ReLU. Neural networks are well suited for modeling complex, non-linear relationships between features and class labels (Goodfellow *et al.*, 2016).

Experimental Workflow

The experimental workflow follows a structured path that begins with preparing and preprocessing the dataset, followed by feature extraction using Painters, VGG-16, and Inception-V3 (Figure 2). The extracted feature vectors serve as inputs for the selected classifiers, specifically SVM and Neural Network. This modular workflow facilitates easy comparison of different feature-classifier combinations in a systematic manner. By maintaining uniformity in preprocessing and evaluation settings, the effects of each embedding method and classifier on classification performance can be directly assessed. This pipeline-based approach is commonly utilized in studies that compare image classification (Tselentis & Papadimitriou, 2023).

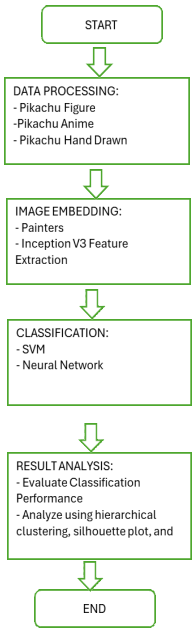

Figure 2. Flowchart of Research Method

Result Analysis

To evaluate classification performance, standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis are employed. These metrics provide supplementary insights into model efficacy, especially in multi-class classification contexts (Powers, 2011). In addition to classification metrics, hierarchical clustering and silhouette analysis are utilized to examine the intrinsic structure of the feature representations. Hierarchical clustering illustrates how similar samples are to one another, while silhouette scores measure the separability of clusters. These analyses offer a more comprehensive understanding of the effectiveness of various embeddings in distinguishing the three Pikachu image categories, extending beyond mere classification accuracy (Raswa *et al.*, 2025; Singh *et al.*, 2021).

3. Results and Discussion

Results

Classification Performance Overview

In this section, we present an overview of the classification performance achieved through the various models employed in our study. The primary objective was to assess how effectively each model could classify Pikachu images across the three distinct categories: anime, action figures, and hand-drawn illustrations. To quantify the performance, we utilized several standard metrics, including accuracy, precision, recall, and F1-score. These metrics provide a multi-faceted view of the models' capabilities, allowing us to identify not only the overall accuracy but also how well each model performs in recognizing specific classes. Furthermore, we conducted a confusion matrix analysis to visualize the classification results, which highlights the instances of true positives, false positives, true negatives, and false negatives. This analysis enables us to pinpoint areas where the models excel and where they struggle, offering insights into the strengths and weaknesses of each classification approach. Overall, the results indicate that the integration of artistic features with deep learning models significantly enhances classification accuracy. By comparing the performance of Support Vector Machine (SVM) and Neural Network classifiers, we aim to demonstrate the effectiveness of combining various feature representations in achieving robust classification outcomes. Through this comprehensive analysis, we hope to elucidate the effectiveness of different embedding techniques and their impact on the classification of stylized images, ultimately contributing valuable insights to the field of image classification.

Table 1. SVM and Neural Network Result

Model	Feature Extraction	AUC	CA	F1	Prec	Recall	MCC
SVM	Painters	0.996	0.960	0.960	0.962	0.960	0.941
	Inception V3	0.993	0.957	0.957	0.958	0.957	0.936
	VGG-16	0.975	0.892	0.892	0.894	0.892	0.839
Neural Network	Painters	0.995	0.980	0.980	0.980	0.980	0.970
	Inception V3	0.995	0.949	0.948	0.949	0.949	0.923
	VGG-16	0.977	0.912	0.911	0.913	0.912	0.869

We used six standard metrics to see how well the classification models worked: Area Under the Curve (AUC), Classification Accuracy (CA), F1-score, Precision, Recall, and Matthews Correlation Coefficient (MCC). These metrics give a full picture of how well a model works, especially when there are more than one class. Table III shows the results of using three different feature extraction methods—Painters, Inception-V3, and VGG-16 (Table 1)—with Support Vector Machine (SVM) and Neural Network (NN) classifiers. In general, the results show that both SVM and Neural Network models do a good job of classifying data when they are used with the right feature representations. However, there are significant differences in performance between feature extraction methods, which shows how important it is to choose the right embeddings for stylized image classification tasks.

Performance of Support Vector Machine

When using the SVM classifier, the best results are with Painters features. The accuracy (CA) and F1-score are both 0.960, and the AUC is 0.996, which is very high. The MCC value of 0.941 further confirms that the predicted and true class labels are very similar, which means that the classification performance is strong and balanced across all classes. These findings indicate that Painters features successfully encapsulate stylistic attributes that are distinctly delineated by the SVM decision boundary. When you combine SVM with Inception-V3 features, you get great results, with an accuracy and F1-score of 0.957 and an AUC of 0.993. The results are still very competitive, even though they are a little lower than the Painters-based configuration. This shows that Inception-V3 gives rich semantic representations that are good for SVM classification. The slight drop in MCC (0.936) means that there is only a small amount of overlap between classes when only deep semantic features are used. On the other hand, when VGG-16 features are used with SVM, the performance drops a lot, with accuracy and F1-score dropping to 0.892 and MCC dropping to 0.839. This performance gap shows that VGG-16 features may not work as well for stylized Pikachu images as Painters and Inception-V3 embeddings. Because VGG-16 has a uniform and deeper structure, it may not be able to show the small stylistic differences that are common in anime and hand-drawn images.

Performance of Neural Network

The Neural Network classifier works best when it uses Painters features. It has an accuracy and F1-score of 0.980, an AUC of 0.995, and an MCC of 0.970. These values are the best results from all the configurations that were tested. This shows that the Neural Network is very good at modeling non-linear relationships in stylistic feature representations. The consistently high precision and recall values also point to balanced performance across all classes. When combined with Inception-V3 features, the Neural Network achieves an accuracy of 0.949 and an MCC of 0.923. Although this performance is slightly lower than that obtained with Painters features, it remains strong and demonstrates the capability of Neural Networks to learn from deep semantic embeddings. The reduction in performance compared to Painters may indicate that stylistic cues play a more dominant role than semantic features in distinguishing between the three Pikachu image categories. Similar to the SVM results, VGG-16 features yield the weakest performance for the Neural Network classifier, with an accuracy of 0.912 and an MCC of 0.869. While still acceptable, these results further support the observation that VGG-16 embeddings are less suited for capturing the stylistic diversity present in the dataset compared to Painters and Inception-V3 features.

Hierarchical Clustering Analysis

Hierarchical clustering analysis was used to learn more about the intrinsic structure and separability of the feature representations that Painters, Inception-V3, and VGG-16 extracted. Hierarchical clustering offers an unsupervised view of how image samples are naturally organized according to feature similarity, in contrast to supervised classification. The dendrograms that come out of this let you see how well the pikachu anime, pikachu figure, and pikachu hand-drawn images are separated from each other and how close they are to each other.

Clustering Pattern Using Painters Features

The dendrogram made from Painters features shows a clear and well-defined clustering structure. Most samples that belong to the same visual category tend to group together in small groups, especially for the hand-drawn and anime classes. At higher linkage distances, these two groups are mostly separate, which

shows that the Painters representation captures a strong stylistic difference. This behavior indicates that Painter's features are proficient in encoding artistic elements such as stroke patterns, texture irregularities, and color distributions, which are prevalent in stylized images. At lower linkage levels, anime and figure samples show some overlap. This is because they share visual traits like color dominance and consistent character shape. But these overlaps are still small, and the overall separation of the clusters is still strong. This clustering pattern is in line with the high classification performance that was achieved when Painters features were used with both SVM and Neural Network classifiers. The Neural Network configuration, in particular, had the highest accuracy and MCC.

Clustering Pattern Using Inception-V3 Features

The dendrogram that uses Inception-V3 features shows a moderate level of cluster separation. In general, samples from the same class tend to group together, but the boundaries between clusters are not as clear as they are with Painters features. Anime and hand-drawn images, in particular, show a lot of mixing in different parts of the dendrogram. This suggests that deep semantic features alone may not be able to fully capture stylistic differences. Still, figure images are grouped together fairly well, making them more coherent than the other two classes. The three-dimensional structure and realistic lighting in action figure images are what make this possible. The deep convolutional filters in Inception-V3 do a great job of capturing these images. The observed clustering behavior elucidates the reason Inception-V3, in conjunction with SVM and Neural Network, attains robust yet marginally inferior classification performance compared to Painter-based embeddings.

Clustering Pattern Using VGG-16 Features

Using VGG-16 features to make a dendrogram shows the weakest cluster separation of the three embedding methods. It is common for samples from different categories to be combined at short linkage distances, which shows that the features are very similar across classes. There is a lot of overlap between anime, figure, and hand-drawn images, which suggests that VGG-16 embeddings have trouble telling the difference between different styles in this dataset. The architectural features of VGG-16, which uses uniform convolutional blocks, may be to blame for this behavior. This is because it may focus on general object-level features instead of more specific stylistic cues. Consequently, the hierarchical clustering structure seems more disjointed and less comprehensible. These results are consistent with the reduced classification accuracy, F1-score, and MCC noted for VGG-16 in both SVM and Neural Network classifiers.

Multidimensional Scaaling (MDS) Analysis

We used Multidimensional Scaling (MDS) to show how similar the Pikachu image representations were that we got from Painters, Inception-V3, and VGG-16 features. MDS makes it easy to see how the three categories—pikachu anime, pikachu figure, and pikachu hand-drawn—are separated from each other and how close they are to each other by projecting high-dimensional feature vectors into a two-dimensional space. The spatial arrangement of points in each visualization indicates the efficacy of the feature representations in maintaining significant distances among image samples.

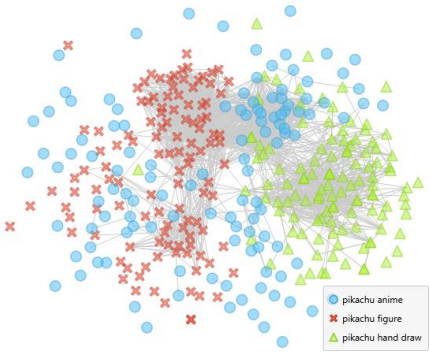

Figure 3. MDS Result Using Painters Features

The MDS visualization based on Painters features shows clear and well-structured separation among the three Pikachu categories. The pikachu figure class forms a tight, dense group, which shows that the members of that class are very similar to each other. At the same time, the pikachu anime and pikachu hand-drawn samples are in different parts of the embedding space, with only a small amount of overlap (Figure 3). Anime and hand-drawn samples are somewhat similar, probably because they both have flat color areas and stylized outlines, but the overall cluster boundaries are still very clear. This

distribution indicates that Painters features proficiently encapsulate stylistic and artistic characteristics essential for differentiating various visual representations of the same object. The clear separation of clusters seen in this MDS plot is in line with the high classification accuracy and MCC that Painters-based models got in the supervised tests.

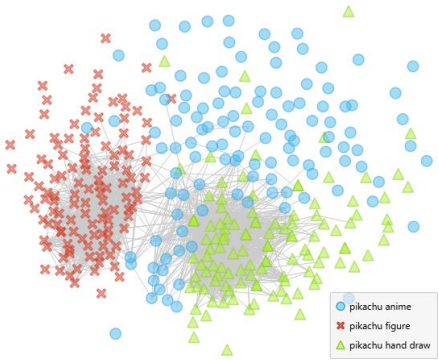

Figure 4. MDS Result Using Inception-V3 Features

On the other hand, the MDS visualization based on Inception-V3 features shows only moderate class separation. The samples of the Pikachu figure are still fairly well grouped, making a recognizable cluster that is only partially separated from the other categories. This indicates that Inception-V3 proficiently captures three-dimensional structure, shading, and semantic object characteristics typically found in action figure imagery (Figure 4). There is, however, a clear overlap between anime and hand-drawn samples of Pikachu. In some areas of the MDS space, these two categories seem to be mixed up. This suggests that deep semantic features alone may not be sensitive enough to small stylistic differences. Inception-V3 gives you rich semantic embeddings, but its representations tend to focus more on object identity than artistic style. This is why they are less distinct than Painters features. This observation aligns with the marginally reduced classification performance exhibited by Inception-V3-based models.

MDS Analysis Using VGG-16 Features

The MDS plot shows that the three classes are the least separated. There are a lot of samples from the pikachu anime, the pikachu figure, and the pikachu hand-drawn that are spread out and overlap a lot in the embedding space. There are no clear boundaries between clusters, and many samples from different groups look like they are close to each other. This distribution indicates that VGG-16 features have difficulty maintaining discriminative information pertaining to artistic and stylistic variations (Figure 5). The model seems to encode general object-level features instead of more specific visual cues, which makes it hard to tell where the samples are in the MDS space. The absence of distinct separation directly elucidates the diminished accuracy, F1-score, and MCC achieved utilizing VGG-16 features for classification with both SVM and Neural Network classifiers.

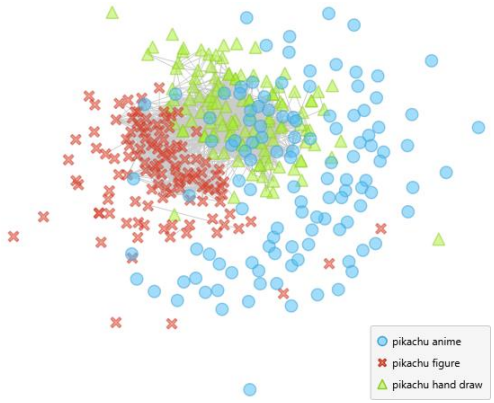

Figure 5. MDS Result Using VGG-16 Features

Silhouette Analysis

We used silhouette analysis to get a numerical measure of how well the clusters were separated using different feature representations, such as Painters, Inception-V3, and VGG-16. The silhouette coefficient tells you how well each sample fits into its assigned cluster compared to other clusters. The values can be anywhere from −1 to 1. Higher positive values mean that the clusters are more cohesive and separate, while values close to zero or negative mean that there is overlap or misassignment. In this research, silhouette plots offer an unsupervised assessment of the discriminative efficacy of each feature embedding.

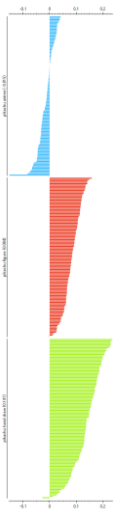

Figure 6. Silhouette Using Painters

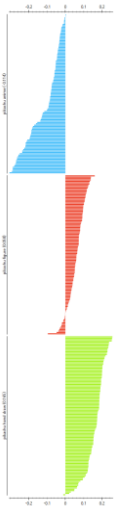

Figure 7. Silhouette Using Inception-V3 Features

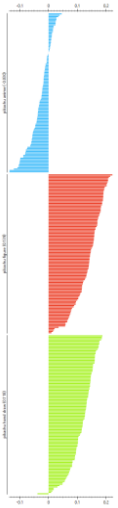

Figure 8. Silhouette Using VGG-16 Features

Silhouette Analysis Using Painters Features

The silhouette plot derived from Painter's features exhibits the most robust clustering structure among the three representations. Most samples from all three classes—pikachu anime, pikachu figure, and pikachu hand draw—show positive silhouette values, which means that the clusters are assigned correctly (Figure 6). The average silhouette scores back this up even more, with the pikachu figure getting the highest average score (0.139) and the pikachu hand draw getting the second highest (0.118). These fairly high numbers show that the groups within the same class are close together and are clearly separate from other groups. The average silhouette score for the pikachu anime class is lower (0.030), which means that it is only partially similar to hand-drawn samples. However, most anime samples still have good silhouette values, which means that stylistic differences are mostly kept, even though there is some confusion between these two visually similar groups. The Painters-based silhouette distribution shows that artistic and stylistic features create strong cluster cohesion and separation, especially for figure and hand-drawn images. This finding is completely in line with the higher classification accuracy and MCC that Painters features achieved in the supervised experiments. This shows that they are effective for classifying stylized images.

Silhouette Analysis Using Inception-V3 Features

The silhouette plot derived from Inception-V3 features demonstrates a moderate degree of clustering quality in comparison to Painter's features (Figure 7). Most samples still have positive silhouette values, which means that the clusters are generally useful. However, the separation between classes is not as clear. The average silhouette score for Pikachu hand-drawn art is 0.141, which means that this group has a lot of cohesion within its own class. The pikachu figure, on the other hand, has an average silhouette value of 0.084, which shows that the clusters are moderately compact. The pikachu anime class, on the other hand, has a lower average silhouette score of 0.050, which means that it is more similar to other categories, especially hand-drawn images. This distribution shows that Inception-V3 is good at capturing semantic and structural features, especially for hand-drawn and figure images. However, it is not as good at picking up on small stylistic differences

between anime and hand-drawn representations. These silhouette patterns align with the supervised classification outcomes, indicating that Inception-V3 attains satisfactory yet suboptimal performance relative to Painters-based embeddings.

Silhouette Analysis Using VGG-16 Features

The silhouette plot made with VGG-16 features shows the weakest clustering structure of the three feature representations (Figure 8). Some samples have positive silhouette values, but a lot of the points are close to zero. This means that the cluster assignment is unclear and there is a lot of overlap between classes. The average silhouette scores back up this finding even more. The pikachu figure got an average score of only 0.058, which shows that the classes are not very cohesive. The average silhouette score for the Pikachu anime class is 0.114, which is a little higher than the average for the Pikachu hand draw class, which is 0.165. Still, the overall silhouette distribution is less stable and more spread out than it is for Painters and Inception-V3. The inconsistent silhouette values across classes indicate that VGG-16 has difficulty consistently preserving discriminative stylistic information. This finding directly elucidates the diminished classification accuracy, F1-score, and MCC noted for VGG-16 in both SVM and Neural Network classifiers.

Discussion

This study examines the efficacy of various feature representations Painters, Inception-V3, and VGG-16 when integrated with Support Vector Machine (SVM) and Neural Network (NN) classifiers for the classification of Pikachu images. By synthesizing findings from supervised classification metrics and multiple unsupervised analyses, including hierarchical clustering, multidimensional scaling (MDS), and silhouette evaluation, this discussion uses supervised classification metrics and several unsupervised analyses, such as hierarchical clustering, multidimensional scaling (MDS), and silhouette evaluation, to give a full picture of how feature characteristics affect classification performance. The supervised classification results show that Painters-based features consistently do better than deep CNN embeddings, especially when used with the Neural Network classifier. The Painters–NN setup gets the best accuracy, F1-score, and MCC, which shows that it can handle different visual representations of the same object very well. Inception-V3 performs competitively but slightly worse, while VGG-16 consistently gets the worst results with both SVM and NN classifiers. These results indicate that stylistic information is more significant than purely semantic features in this dataset, which is characterized by artistic and visual style variations rather than solely object identity. The hierarchical clustering analysis offers unsupervised validation for these findings. Dendrograms based on Painters features exhibit dense intra-class clusters and distinct inter-class separation, especially between hand-drawn and figure images. Inception-V3 makes clusters that are somewhat coherent, with a clear overlap between anime and hand-drawn images. VGG-16, on the other hand, shows clusters that are broken up and overlap. This pattern is very similar to the supervised results, which means that better intrinsic cluster separability leads to better classification results.

The MDS visualizations give us more information by showing how feature embeddings keep sample distances in a smaller two-dimensional space. The features of painters create clusters that are easy to understand and separate, while the embeddings of Inception-V3 show some overlap between categories that look similar. VGG-16 embeddings, on the other hand, have the most spread-out and overlapping distributions, which shows that they don't work well for stylized image classification. These spatial patterns elucidate the observed discrepancies in performance within the classification experiments. The silhouette analysis quantitatively reinforces these findings by measuring cluster cohesion and separation across all feature representations. Painters features exhibit the most consistent and well-balanced cluster separation, with average silhouette scores of 0.139 for pikachu figure, 0.118 for pikachu hand draw, and 0.030 for pikachu anime, indicating strong intra-class cohesion and clear inter-class separation, particularly for figure and hand-drawn images. In comparison, Inception-V3 features show moderate clustering quality, achieving average silhouette scores of 0.141 for pikachu hand draw, 0.084 for pikachu figure, and 0.050 for pikachu anime, suggesting that semantic features capture structural information effectively but are less sensitive to stylistic differences between anime and hand-drawn images. Conversely, VGG-16

features demonstrate the weakest and most uneven clustering behavior, with average silhouette scores of 0.165 for pikachu hand draw, 0.114 for pikachu anime, and only 0.058 for pikachu figure, indicating substantial overlap and unstable cluster boundaries. These silhouette patterns closely mirror the supervised classification results and provide strong unsupervised validation of the observed performance trends. The consistency across classification metrics, dendrogram structures, MDS plots, and silhouette scores enhances the validity of the conclusions. The results show that feature representations that capture artistic and stylistic traits work better for this particular classification task than general deep semantic embeddings. Deep CNN models like Inception-V3 are still useful for getting object-level semantics, but they work better when used with datasets that have a lot of artistic variation when they are used with complementary stylistic representations. This study shows how important it is to match feature extraction methods with the dataset's visual features. Combining Painters features with Neural Network or SVM classifiers is the best way to classify Pikachu images when there isn't much data. These results provide useful information for designing hybrid image classification systems and could help with future research on stylized image recognition tasks that are not part of this study.

4. Conclusion

This study examined the efficacy of various feature representations and classification models for Pikachu image classification, categorizing them into three visually distinct groups: anime, action figures, and hand-drawn images. The study offers a thorough evaluation of the impact of feature characteristics on classification performance in stylized image datasets by integrating Painters-based stylistic features with deep transfer learning models (VGG-16 and Inception-V3) and assessing them through Support Vector Machine (SVM) and Neural Network (NN) classifiers. The experimental outcomes consistently indicate that Painters features surpass deep CNN embeddings, especially when integrated with the Neural Network classifier. This setup has the best classification accuracy, F1-score, and Matthews Correlation Coefficient, which means it works well and evenly across all image categories. Inception-V3 performs better than VGG-16 and is better at capturing semantic and structural information, especially for action figure images. VGG-16, on the other hand, consistently performs worse, which suggests that it is not very sensitive to the artistic and stylistic differences in the dataset. These results are strongly corroborated by various unsupervised analytical methods, including hierarchical clustering, multidimensional scaling (MDS), and silhouette evaluation, all of which demonstrate consistent patterns of cluster separability that reflect the supervised classification outcomes. The convergence of supervised and unsupervised evidence strengthens the reliability of the conclusions and underscores the significance of stylistic feature representations in managing visually diverse image collections.

The results have several important practical implications. First, for applications that use artistic, cartoon, or hand-drawn images, like digital content moderation, fan-art categorization, intellectual property management, and visual search in creative platforms, stylistic feature representations are much better than purely semantic deep features. Second, the strong performance of classical classifiers combined with fixed feature extractors shows that it is possible to get high classification accuracy without needing a lot of computing power or a lot of training data. This makes the proposed method good for situations where resources are limited. Third, the hybrid framework shown in this study can be easily used for other stylized object recognition tasks besides Pikachu images. This means that it can be used in real-world systems where visual style changes a lot. In general, this study shows how important it is to match feature extraction methods to the visual properties of the data. For datasets characterized by artistic variation rather than solely semantic distinctions, hybrid methodologies that amalgamate stylistic attributes with traditional classifiers provide a pragmatic, efficient, and efficacious solution, especially in scenarios with constrained data availability. Despite its contributions, this study has several limitations. The dataset size is relatively small and focuses on a single character, which may limit generalizability to other objects or domains. In addition, the use of fixed feature extractors without fine-tuning may restrict the full potential of deep models. Future research could

explore larger and more diverse datasets, investigate feature fusion strategies, and incorporate fine-tuning or attention-based mechanisms to further enhance classification performance. Extending the proposed framework to broader stylized image domains may also provide valuable insights into its applicability in real-world scenarios.

5. Acknowledgment

The authors would like to express their sincere gratitude to all parties who contributed to the completion of this research. Special appreciation is extended to colleagues and peers who provided valuable feedback and constructive suggestions during the research process. Their insights significantly helped improve the quality and clarity of this work. The authors also acknowledge the use of publicly available image sources, including Google Images and Pinterest, which were utilized for dataset collection in this study. Also the use of Chat GPT to elaborate and enhance the wording. These resources played an important role in enabling the exploration of stylized image classification. Finally, the authors would like to thank the academic community and open-source contributors whose tools and frameworks supported the experimental analysis conducted in this research.

6. References

Chaturvedi, S. S., Tembhurne, J. V., & Diwan, T. (2020). A Multi-Class Skin Cancer Classification Using Deep Convolutional Neural Networks. Multimedia Tools and Applications, 79(39–40), 28477–28498. https://doi.org/10.1007/s11042-020-09388-2.

Chen, H., Wang, Y., Guo, J., & Tao, D. (2023). VanillaNet: the Power of Minimalism in Deep Learning. Advances in Neural Information Processing Systems, 36, 7050–7064.

Chen, M., Chen, W., Chen, W., Cai, L., & Chai, G. (2020). Skin Cancer Classification with Deep Convolutional Neural Networks. Journal of Medical Imaging and Health Informatics, 10(7), 1707–1713. https://doi.org/10.1166/jmihi.2020.3078.

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Imran, S., Naqvi, R. A., Sajid, M., Malik, T. S., Ullah, S., Moqurrab, S. A., & Yon, D. K. (2023). Artistic Style Recognition: Combining Deep and Shallow Neural Networks for Painting Classification. Mathematics, 11(22), 4564. https://doi.org/10.3390/math11224564.

Iqbal, M. A., Wang, Z., Ali, Z. A., & Riaz, S. (2021). Automatic Fish Species Classification Using Deep Convolutional Neural Networks. Wireless Personal Communications, 116(2), 1043–1053. https://doi.org/10.1007/s11277-019-06634-1.

Jastrzebska, A. (2022). Time Series Classification Through Visual Pattern Recognition. Journal of King Saud University - Computer and Information Sciences, 34(2), 134–142. https://doi.org/10.1016/j.jksuci.2019.12.012.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A Survey of The Recent Architectures of Deep Convolutional Neural Networks. Artificial Intelligence Review, 53(8), 5455–5516. https://doi.org/10.1007/s10462-020-09825-6.

Kumar, N., Kaur, N., & Gupta, D. (2020). Major Convolutional Neural Networks in Image Classification: A Survey (pp. 243–258). https://doi.org/10.1007/978-981-15-3020-3_23.

Law, S., Seresinhe, C. I., Shen, Y., & Gutierrez-Roig, M. (2020). Street-Frontage-Net: Urban Image Classification Using Deep Convolutional Neural Networks. International Journal of Geographical Information Science, 34(4), 681–

707. https://doi.org/10.1080/13658816.2018.1555832.

Liu, S., Niles-Weed, J., Razavian, N., & Fernandez-Granda, C. (2020). Early-Learning Regularization Prevents Memorization of Noisy Labels. Advances in Neural Information Processing Systems, 33, 20331–20342.

Nnamoko, N., Barrowclough, J., & Procter, J. (2022). Solid Waste Image Classification Using Deep Convolutional Neural Network. Infrastructures, 7(4), 47. https://doi.org/10.3390/infrastructures7040047.

Pinaya, W. H. L., Vieira, S., Garcia-Dias, R., & Mechelli, A. (2020). Convolutional Neural Networks. In Machine Learning (pp. 173–191). Elsevier. https://doi.org/10.1016/B978-0-12-815739-8.00010-9.

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2021). Do Vision Transformers See Like Convolutional Neural Networks? Advances in Neural Information Processing Systems, 34, 12116–12128.

Raswa, F. H., Lu, C.-S., & Wang, J.-C. (2025). HistoFS: Non-IID Histopathologic Whole Slide Image Classification via Federated Style Transfer with RoI-Preserving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 30251–30260).

Shen, X., Wang, Y., Lin, M., Huang, Y., Tang, H., Sun, X., & Wang, Y. (2023). DeepMAD: Mathematical Architecture Design for Deep Convolutional Neural Network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6163–6173.

Singh, R., Bharti, V., Purohit, V., Kumar, A., Singh, A. K., & Singh, S. K. (2021). MetaMed: Few-Shot Medical Image Classification Using Gradient-Based Meta-Learning. Pattern Recognition, 120, 108111. https://doi.org/10.1016/j.patcog.2021.108111

Tselentis, D. I., & Papadimitriou, E. (2023). Driver Profile and Driving Pattern Recognition for Road Safety Assessment: Main Challenges and Future Directions. IEEE Open Journal of Intelligent Transportation Systems, 4, 83–100. https://doi.org/10.1109/OJITS.2023.3237177

Tsipras, D., Santurkar, S., Engstrom, L., Ilyas, A., & Madry, A. (2020). From ImageNet to Image Classification: Contextualizing Progress on Benchmarks. Proceedings of Machine Learning Research, 9625–9635.

Visalli, F., Bonacci, T., & Borghese, N. A. (2021). Insects Image Classification Through Deep Convolutional Neural Networks. In Progresses in Artificial Intelligence and Neural Systems (pp. 217–228). https://doi.org/10.1007/978-981-15-5093-5_21.

Zhang, Y., Gao, J., & Zhou, H. (2020). Breeds Classification with Deep Convolutional Neural Network. Proceedings of the 2020 12th International Conference on Machine Learning and Computing, 145–151. https://doi.org/10.1145/3383972.3383975.

Zhao, J., Fang, Y., & Li, G. (2021). Recurrence along Depth: Deep Convolutional Neural Networks with Recurrent Layer Aggregation. Advances in Neural Information Processing Systems, 34, 10627–10640.