Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 9 No. 3 (2025) 435 - 443 e-ISSN: 2580-0760

Classification of Red Foxes: Logistic Regression and SVM with VGG-16, VGG-19, and Inception V3

Brian Sabayu[1*], Imam Yuadi[2]

[1]Master's Program Human Resource Development-Data Analytics, Graduate School, Universitas Airlangga, Surabaya, Indonesia

[2]Department of Information and Library Science, Faculty of Social and Political Sciences, Universitas Airlangga, Surabaya, Indonesia

[1]brian.sabayu-2024@pasca.unair.ac.id, [2]imam.yuadi@fisip.unair.ac.id

Abstract

Deep learning models demonstrate a high degree of accuracy in image classification. The task of distinguishing between various sources of red fox images—such as authentic photographs, game-captured images, hand-drawn illustrations, and AI-generated images—raises important considerations regarding realism, texture, and style. This study conducts an evaluation of three deep learning architectures: Inception V3, VGG-16, and VGG-19, utilizing images of red foxes. The research employs Silhouette Graphs, Multidimensional Scaling (MDS), and t-Distributed Stochastic Neighbor Embedding (t-SNE) to assess clustering and classification efficiency. Support Vector Machines (SVM) and Logistic Regression are utilized to compute the Area Under the Curve (AUC), Classification Accuracy (CA), and Mean Squared Error (MSE). The MDS plots and t-SNE data clearly demonstrate the capability of the three deep learning models to distinguish between the image categories. For game-captured images, VGG-16 and VGG-19 demonstrate quite outstanding performance with silhouette scores of 0.398 and 0.315, respectively. This study explores the enhancement of classification accuracy in logistic regression and support vector machines (SVM) through the refinement of decision boundaries for overlapping categories. Utilizing Inception V3, an artificial intelligence-generated image silhouette score of 0.244 was achieved, demonstrating proficiency in image classification. The research highlights the challenges posed by diverse datasets and the efficacy of deep learning models in the classification of red fox images. The findings suggest that integrating deep learning with machine learning classifiers, such as logistic regression and SVM, may improve classification accuracy.

Keywords: red fox images; image classification; deep learning models

How to Cite: B. Sabayu and I. Yuadi, "Hybrid Classification of Red Foxes: Logistic Regression and SVM with VGG-16, VGG-19, and Inception V3", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 3, pp. 435 - 443, May 2025.
*Permalink/*DOI: https://doi.org/10.29207/resti.v9i3.6356

Received: February 4, 2025
Accepted: April 12, 2025 2025
Available Online: May 24, 2025

1. Introduction

In the realm of complex image classification tasks, such as the categorization of photographs depicting wild animals, hybrid classification techniques—which integrate deep learning models with machine learning algorithms—are gaining prominence. The red fox, in particular, presents a unique challenge due to its diverse habitats and morphological similarities to other species. This classification process typically overwhelms traditional classification techniques; thus, an appropriate hybrid solution is needed [1]. This study uses deep learning features like VGG-16, VGG-19, and Inception V3 to investigate a hybrid classification framework by combining Logistic Regression and Support Vector Machines (SVM) models. For photo identification applications, convolutional neural networks (CNNs) like VGG-16 and Inception V3 might show remarkable feature extraction accuracy [2]. The proposed model combines their expertise in feature extraction with the classification accuracy of SVM and Logistic Regression to increase accuracy and generalizability. The shortcomings of individual classifiers are mitigated and workable solutions to wildlife classification issues are provided by combining various approaches.

The primary uses of image analysis of red foxes are ecological monitoring, behavioral research, and species identification. However, it is challenging to accurately identify red fox pictures due to their wide geographic range and complex morphological characteristics. Prior to this, red foxes were distinguished from other canids using labor-intensive and error-prone manual feature extraction and photographic analysis methods. The red fox recognition is now automatic and extremely accurate thanks to Convolutional Neural Networks (CNNs), a type of deep learning. Photos taken in different lighting conditions have been used to extract distinguishing features like body form and fur patterns using pre-trained models like VGG-16 and Inception V3 [3]. CNN features combined with machine learning classifiers like SVM have significantly improved classification performance in datasets with unequal species representation [4]. These advancements have improved our ability to monitor red fox populations, supporting biological research and conservation efforts.

Image recognition and species identification have been much improved via classification and deep learning. Prior studies found convolutional neural networks like AlexNet, which changed feature extraction for difficult datasets [5]. Inception V3 improved classification performance by means of factorized convolutions and additional classifiers, therefore highlighting its relevance in wildlife research where precise species identification is vital [6]. Most research has focused on CNN-based categorization, thereby disregarding the general significance of hybrid techniques integrating conventional machine learning models. Although SVMs have shown strong performance in species classification—especially when coupled with handcrafted or deep learning components—their success in hybrid systems for red fox images classification is yet unknown. Logistic regression is a basic component of hybrid models, even if its simplicity is due to its strong interpretability and efficiency with linearly separable datasets. Recent studies suggest that in challenging classification tasks involving imbalanced datasets, the integration of CNNs with machine learning classifiers, including SVM and logistic regression, increases resilience and generalization [7]. This work intends to present a more exact and efficient solution employing a new hybrid approach integrating CNN-based feature extraction with SVM and logistic regression for red fox image classification.

This study offers a hybrid classification strategy combining a Support Vector Machine (SVM) with Logistic Regression by using pre-trained deep learning architectures VGG-16, VGG-19, and Inception V3. CNN-based feature extraction combined with the interpretability and efficiency of SVM and Logistic Regression improves computational efficiency and classification accuracy, unlike most methods, which basically depend on fixed deep learning or conventional machine learning algorithms. By use of the hierarchical feature extraction powers of VGG-16, VGG-19, and Inception V3, the proposed method detects complicated morphological changes in red foxes across many image categories: real-world, game-captured, hand-drawn, and artificially created images. Furthermore, using hybrid machine learning methodologies with reduced classification thresholds enhances generalizability from imbalanced datasets. The findings unequivocally demonstrate the substantial advancements that hybrid machine learning and deep learning have led to in the development of categorization systems.

Therefore, the integration of Convolutional Neural Networks (CNN) with Conventional Classifiers is made possible by the strong and scalable solution that is provided by assisting in the identification of animal images in AI-generated image analysis and ecological monitoring. This method addresses issues like visual similarity to other species and environmental unpredictability; hence, it improves classification accuracy and robustness. The next parts of the research deal with the use of the hybrid model and the approaches for data preparation and collection. We will first demonstrate the assessment measures for model performance, followed by a careful analysis of the outcomes using conventional methods. The results are discussed in the next section of the research, coupled with their consequences for wildlife monitoring and suggested approaches for automated species categorization in the next projects.

Machine learning and deep learning technologies have tremendously expanded research on automated species categorization by providing more scalable and consistent approaches for wildlife monitoring [8]. Early attempts in this sector mostly relied on traditional machine learning methods, such as Support Vector Machines (SVM) and Logistic Regression, which conducted classification with manually created features obtained from images with customized tools. Although their effectiveness with smaller and more clearly specified datasets was good, these algorithms suffered greatly when applied to vast and diverse wildlife datasets distinguished by significant variety in animal anatomy, lighting conditions, and environmental factors [9]. CNNs changed the domain by allowing automatic and powerful feature extraction directly from unprocessed visual input, hence enhancing classification accuracy and efficiency. Pretrained models using deep architecture with small convolutional filters—VGG-16 and VGG-19—are often utilized for feature extraction. These models capture intricate visual features yet can efficiently generalize throughout a spectrum of activities and contexts. These systems exhibit amazing species detection and classification by way of camera-captured images in wildlife monitoring [10]. Although they are effective, managing imbalanced datasets and complex ecological circumstances remains difficult; thus, more research on hybrid algorithms combining CNNs with conventional machine learning classifiers is essential to increase classification performance.

Furthermore, hybrid models have shown to increase resilience in real-world settings where normal fluctuations in lighting, posture, and environmental variables exist [11]. Since their bodily forms and some of their visual characteristics cross those of other species, few research has concentrated on the difficulty in classifying red foxes. This work uses the benefits of hybrid models combining contemporary CNNs with conventional classifiers to enable individuals to recognize red foxes and offer a consistent approach to animal monitoring [12]. Common and flexible carnivores are found all throughout North America, Europe, and Asia. The red fox lives in most of the cities to meadows and woodlands, the red fox finds perfect habitat all around. Often opportunistic feeders, consuming small animals, birds, fruits, and even human food waste, this animal is the red fox for its pointed ears, bushy tail, and beautiful reddish-orange fur. Their success in rural and urban settings—where they often dwell next to people—has come from their capacity to adapt to different situations [13]. Red foxes are a vital participant in preserving ecological balance as they regulate the numbers of tiny animals, therefore influencing ecosystems.

The fox is one of the few species that survive in the wide-open landscape of Hyrule in The Legend of Zelda: Breath of the Wild. Designed and released by Nintendo, the game is part of the much-loved Zelda series and premiered 2017 for the Nintendo Switch and Wii U. Players in this open-world action-adventure game rouse Link from a long hibernation to stop the horrible Calamity Ganon. Among the several creatures' players discover in the diverse surroundings improving the immersive and dynamic environment of the game are foxes. Usually wary and retreating as the player approaches, foxes in Zelda: Breath of the Wild might be seen prowling forests. Under direction by Hidemaro Fujibayashi and created by Eiji Aonuma, the game was praised for its creative mechanics, meticulous attention to detail, and adaptability it gives players in exploring and interacting with the environment [14].

Artificial intelligence red fox images show an interesting junction of artificial intelligence and wildlife representation whereby machine learning algorithms—especially Generative Adversarial Networks (GANs)—are used to generate realistic or creative representations of this well-known animal. Extensive databases of real-world fox images allow training on these artificial intelligence systems to learn to create complete visual representations, so capturing the special features of the red fox, including its reddish-orange fur, bushy tail, and sharp facial expressions [15]. Environmental simulation, gaming, and instructional tools are among the applications for such AI-generated photos; they help to produce lifelike representations without asking for actual animal photography. Therefore, modeling numerous scenarios or settings involving red foxes helps to better wildlife research and conservation activities by means of thorough study of animal behavior and ecosystems [16]. As artificial intelligence develops, the precision and inventiveness in creating creatures like the red fox may grow, therefore offering even more chances for their use in many other spheres.

Red fox Tod interacts with hound dog Copper in Disney's 1981 animated film The Fox and the Hound. The story looks at personal qualities that set people apart, social expectations, and the challenges of growth. Although Tod's inner complexity is clear as he tries to keep his love for Copper despite their basic differences, his conduct is defined by traditional fox traits like cunningness and arrogance. The red fox in the movie stands for both urban environments and wild surroundings as well as young innocence. Viewers have especially related to the critically praised picture The Fox and the Hound for decades, which investigates ideas of friendship, loyalty, and the conflict between nature and nurture [17].

2. Methods

Four sets of one hundred pictures per made up of the diversified collection used in the study. There were 400 images in total. The first group consists of one hundred actual red foxes in their natural environment, which properly represents the species. Figure 1 presents the example of the captured true nature of the red fox with its pointed ears, bushy tail, and reddish-orange fur. The second collection consists of one hundred images taken from a video game featuring foxes among other animals. These pictures show digitally created foxes put in a fantastical setting.

Figure 1. Four Images of Red Fox from Various Sources

Seen in the third set of one hundred images, Tod and Vixey, the red fox characters, uses a hand-drawn animation style in the Disney movie "The Fox and the Hound". These pictures emphasize the range of emotions the animal experiences and provide a realistic and imaginative picture of it, therefore underlining its sensitivity. One hundred photos of red foxes produced by artificial intelligence using generative adversarial networks (GANs), an ML method make up the collection. Using fictional materials, these computer-generated images aim to almost perfectly resemble real foxes. This study is to examine efficient processing and classification of many image formats to increase understanding of the problems and accuracy of fox photo categorization from numerous sources.

Figure 2 shows the process of the 400 images that have been collected. The image processing was carried out using the Orange Data Mining application by performing data extraction on the image collection and then conducting image embedding using Inception V3, VGG-16, and VGG-19. Next, it is classified using Support Vector Machine and Logistic Regression until the image collection can be analyzed using t-SNE, MDS, and Silhouette plot.

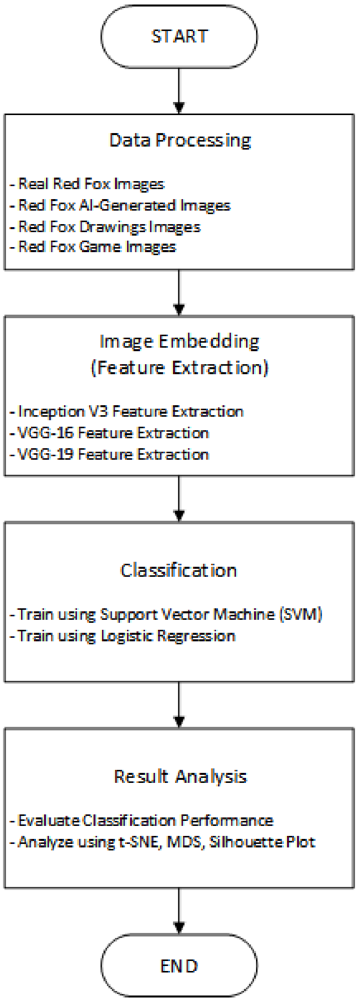

Figure 2. Flowchart of Research Method

2.1 Data Processing

This phase produces a diverse array of images for feature extraction, utilizing four distinct types of images: game-captured images, artificial intelligence-generated images, drawing images, and real red fox images. Artificial intelligence images are generated using generative models such as GANs [15]. Game images are captured by in-game conditions that are created by computer design; real red-fox images are those that are genuine and not produced by human-made tools.

The act of drawing images or using digital paints is an expression of artistic expression. The classification model's ability to generalize effectively across multiple image sources is ensured by the acceptance of a diverse range of input types by these categories [18]. The model can more accurately classify and identify images from a variety of backgrounds because of this spectrum.

2.2 Image Embedding (Feature Extraction)

This process transforms images into numerical values appropriate for classification. Three deep learning models—VGG-16, VGG-19, and Google Inception V3—are utilized to accomplish this goal. VGG-16 utilizes a sixteen-layer Convolutional Neural Network (CNN) to extract fundamental hierarchical features from images, including edges, textures, and complex patterns [19]. Despite comprising 19 layers, VGG-19 is comparable and can generate more accurately defined images.

Google Inception V3, a sophisticated CNN architecture, achieves computational efficiency and efficacy on intricate datasets by utilizing diverse filter widths in each convolutional layer to capture a spectrum of features across multiple layers [20]. Each technique generates embeddings from images by extracting features that serve as numerical representations, maintaining essential image information while minimizing dimensionality, thus improving assessment and comprehension by machine learning models.

2.3. Classification

The objective is to group acquire attributes utilizing machine learning models. This study employs two prevalent classification techniques: logistic regression and support vector machine (SVM). Support Vector Machine (SVM) is a supervised machine learning technique that is particularly effective in classifying high-dimensional data, such as image embeddings, by identifying the optimal hyperplane to separate data points into distinct classes [21].

Primarily used in binary classification, logistic regression assesses the probability of a sample belonging to a particular class and can be extended to multiclass scenarios through one-versus-all techniques [22]. Utilizing the embeddings generated by the VGG-16, VGG-19, and Inception V3 models, these classifiers will categorize images into designated classifications such as game, artificial intelligence, hand-drawn, and real-world images. The simplicity and efficacy of Support Vector Machines and Logistic Regression render them excellent alternatives for this classification problem.

2.4 Result Analysis

This study evaluates classification models across diverse image domains using metrics including F1 score, recall, accuracy, and precision. Utilized dimensionality reduction techniques, namely t-SNE and MDS, to visually depict high-dimensional embeddings. This indicated that unique clustering patterns could be identified, allowing each classifier to differentiate between image categories. The silhouette plot, which quantitatively evaluated clustering quality via intra-cluster cohesion and inter-cluster separation, further corroborated the models [23]. This methodology evaluated SVM and logistic regression across four categories: real red fox images, hand-drawn, AI-generated images, and game images.

The study reveals inconsistencies in model effectiveness, with certain models outperforming others for image categories due to the distinctive characteristics of the underlying CNN architectures. This study assesses the advantages and disadvantages of each classifier through quantitative metrics and visual representations, thereby aiding in the selection of the most appropriate model for various input types. The thorough investigation uncovers significant new insights into classifier behavior, suggesting that image categorization presents challenges for future research efforts [24].

3. Results and Discussions

Regularly outperforming three models with different architectures—Inception V3, VGG-16, and VGG-19—is logistic regression. This holds outside of design.

Table 1. Logistic Regression (LR) and Support Vector Machine (SVM) Results

Model	Feature Extraction	AUC	CA	F1	Prec	Rcl	MCC
LR	Inception V3	1.00	0.99	0.99	0.99	0.99	0.98
	VGG-16	1.00	1.00	1.00	1.00	1.00	0.99
	VGG-19	1.00	0.99	0.99	0.99	0.99	0.98
SVM	Inception V3	1.00	0.98	0.98	0.98	0.98	0.98
	VGG-16	1.00	0.97	0.97	0.97	0.97	0.96
	VGG-19	1.00	0.96	0.96	0.96	0.96	0.95

Table 1 shows the logistic regression approach used in Inception V3, which yielded some quite amazing results. These have an MCC of 0.983, an AUC of 1.000, CA of 0.989, F1, accuracy and recall of 0.987. Logistic regression performed rather well in the VGG-16 architecture with a 1.000 for area under the curve (AUC), 0.995 for confidence interval (CA), and 0.995 for F1, recall, accuracy, and MCC, logistic regression. Getting these results requires attaching great degrees of accuracy. Logistic regression kept on outperforming VGG-19, as in past example, with an AUC of 1.000, a CA of 0.985, and F1, precision, recall, and MCC all at 0.985. SVM did remarkably and peaked in Inception V3 with an AUC of 0. 999 and CA of 0.983. This was the best result thus far generated. Though it constantly displayed rather low values, it kept producing outstanding results. Logistic regression was the best among all the models and setups. Though usually robust, the Support Vector Machine (SVM) routinely performed worse than logistic regression. This was noted under several benchmarks, including accuracy and area under the curve (AUC).

3.1 Results

The t-SNE visualization of the four image categories (Figure 4, 5 and 6): AI-generated images (blue circles), drawings from "The Fox and the Hound" (red crosses), game images from "The Legend of Zelda: Breath of the Wild" (green triangles), and real-world fox images (orange crosses). Four different image types—drawings, game images, artificial intelligence-generated images, and real-world images of foxes—are analyzed inception V3, VGG-16, and VGG-19 to show their clustering performance.

T-SNE visualizations are applied to underline these performances. Figure 3 shows the Inception V3 clusters have some overlap between the game-captured and AI-generated images, which would point to feature mixing. Moreover, the created images of the game and their own self-generated images only somewhat overlap. Moreover, the created images and real-life pictures of the game have a certain resemblance.

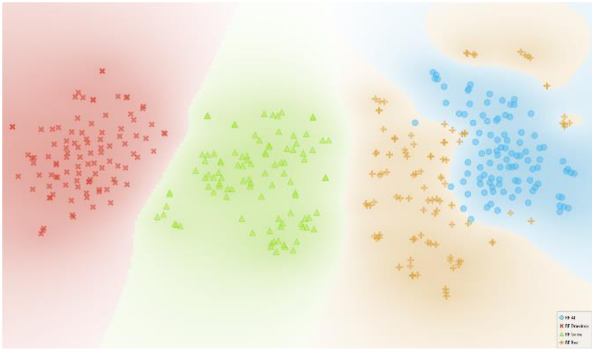

Figure 3. t-SNE result using Google Inception V3

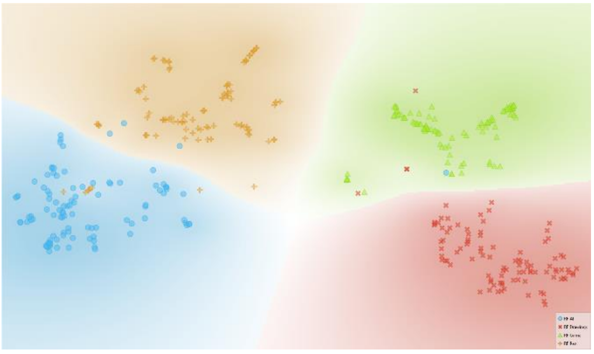

Figure 4. t-SNE result using VGG-16

Though there are a few obvious outliers, the VGG-16 method (Figure 4) enhances separation, especially between game images and drawings. This is rather evident when one compares the two image formats. This is particularly pertinent while differentiating the two varieties of images.

Clustering the several categories helps to make this possible. The arrangement makes this feasible. Having this strategy in place will help us to reach our target.

Figure 5 shows VGG-19 is the most advanced model since it shows the most compact and unique clusters with the least overlap. This shows its extraordinary feature extraction and category differentiation capacity. Though it beats the Inception V3 model generally, in a few areas the most unique VGG-19 model is rather less effective. Next comes the VGG-19 model then the VGG-16 model. Though all three models can fairly separate the categories, the VGG-19 model offers the best separation. This is still the case even if all three models can clearly separate. This is a broad overview.

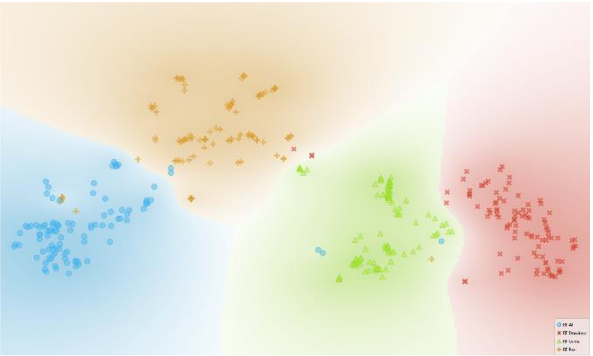

Figure 5. t-SNE result using VGG-19

The scatter plot analysis exposes several approaches to illustrating red foxes (Figure 6-8). orange plus signs (Real-world images), green triangles (game), red crosses (drawings), and blue circles (AI-Generated images). By images of red foxes, models Inception V3, VGG-16, and VGG-19 can generate the MDS (Multidimensional Scaling). The outputs greatly exhibit various clustering trends.

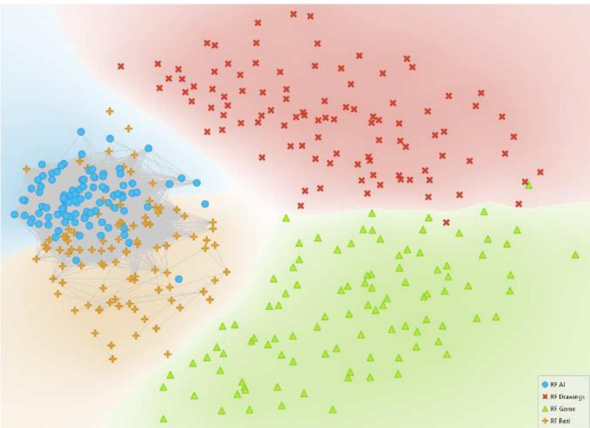

Figure 6. MDS result using Google Inception V3

Figure 6 illustrates the reliability and efficacy of the Inception V3 classification systems, which adeptly distinguish between various image categories, such as illustrations, game-captured images, and real-world photographs, through well-defined clusters. Figure 7 demonstrates that VGG-16 effectively categorizes original data points into four distinct groups: artificial intelligence-generated images, real-world images, gaming images, and drawings, utilizing exceptional feature extraction and classification capabilities..

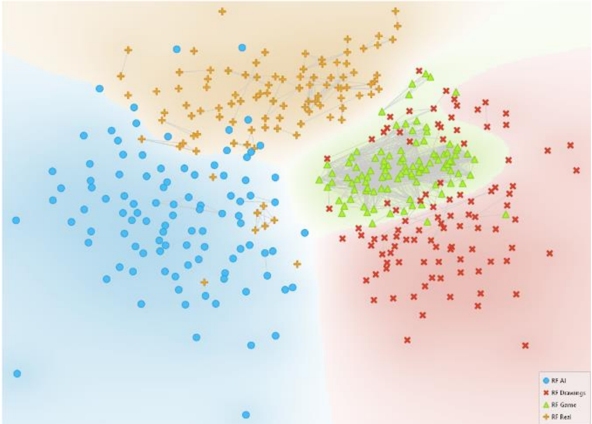

Figure 7. MDS result using VGG-16

The results define several clusters for the same categories and show the stability of VGG-19 over several image datasets, shown in Figure 8. Three closely spaced models show the amazing arrangement of several red fox image variants. While all models run effectively, little variations in cluster alignment and cohesiveness could affect the generalizability and feature extraction details. Artificial intelligence research, game development, wildlife monitoring, image classification among other uses show promise for deep learning models including Inception V3, VGG-16, and VGG-19.

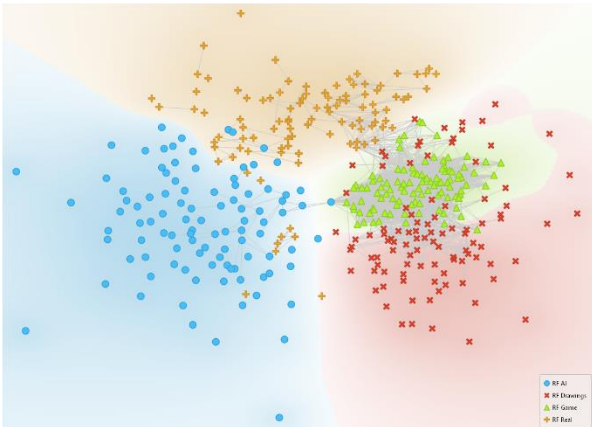

Figure 8. MDS result using VGG-19

The MDS (Multidimensional Scaling) results for Inception V3, VGG-16, and VGG-19 show their ability to classify red fox images over many datasets when tightly matched with their respective t-SNE values. Particularly true when considering AI-generated images and in-game graphics, there are some commonalities shown by the clear clusters revealed by Inception V3. VGG-16 improves cluster refinement and more varied groupings, so facilitating better extraction and classification of features. Getting perfect separation by well-defined small clusters shows how remarkably effectively VGG-19 classifies. All three models show remarkable ability in differentiating between four types of images: those produced artificially, drawings, images from games, and images from the real world. This is true despite little variations in cluster separation and compactness as well as feature extraction capacity. As the MDS results show, digital art analysis, artificial intelligence development, and wildlife monitoring are just a few of the several disciplines that can benefit from the accuracy in image classification of the models.

Various images of red foxes reveal varying degrees of clustering efficiency in the silhouette plots generated by the Inception V3, VGG-16, and VGG-19 models. With a silhouette score of 0.244, Figure 9 shows the Inception V3 beats other AI-generated image clustering systems displaying notable intra-cluster similarity and clear difference from other clusters. A score of 0.013 denotes flaws in the images of the game. Figure 10 shows VGG-16 accuracy in spotting game images with its amazing silhouette score (0.398). Drawings clusters based on a 0.002 score produces less than perfect results.

VGG-19 shows clustering difficulty with a score of roughly 0.007 (Figure 11); yet it performs effectively in classifying game images with a score of 0.315. From both natural and synthetic images, the two VGG models generate unsatisfactory results. Every model shows superiority in some domains: Since all models except for Inception V3 show poor performance in this regard, improved accuracy in clustering drawings is obviously needed. VGG-16 and VGG-19 perform better on game images in some other respects as well.

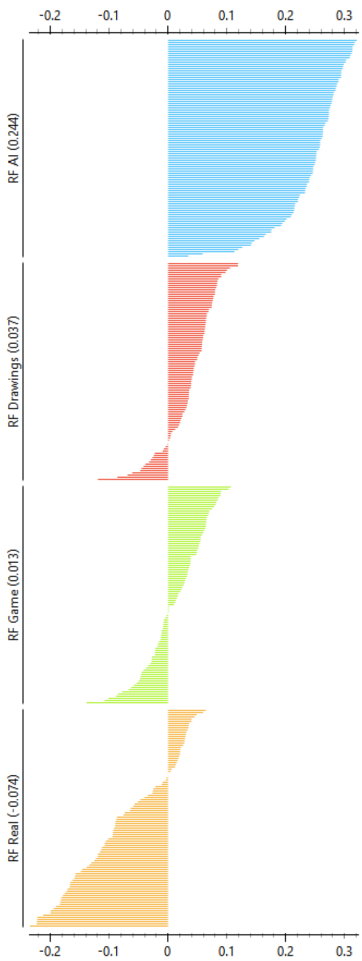

Figure 9. Silhouette Plot result using Google Inception V3

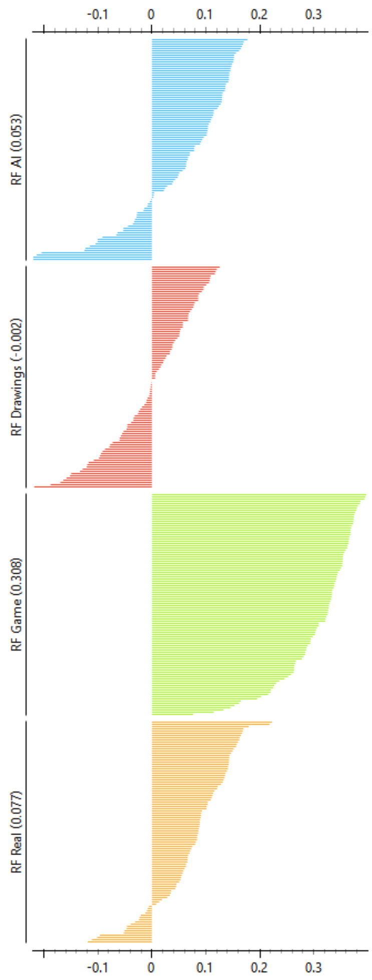

Figure 10. Silhouette Plot result using VGG-16

This comparison highlights the need to choose a suitable model depending on the specific demand of the image classification employment. By means of the Silhouette Plot, the clustering efficacy of Inception V3, VGG-16, and VGG-19 is revealed in an analysis of the data for several red fox image categories. The silhouette score of 0.244 for Inception V3 shows great efficiency in AI-generated images; on game images, its score of 0.013 indicates rather challenging results. VGG-16 achieves the highest clustering accuracy for game images (0.398 silhouette score), while struggling separating drawings (yielding a score of 0.002). With game images (0.315), the results show VGG-19 performs rather well; with clustered drawings (0.007), it performs poorly. About the grouping of real-generated and synthetic images, both VGG models exhibit very poor performance. Though VGG-16 and VGG-19 shine with gaming images, the statistics show that Inception V3 is the best option for AI-generated image tasks. The need for improved feature extraction in this field is shown by the impossibility of any model to precisely cluster drawings. This stresses the need to customize the model to meet the requirements of the image classification job.

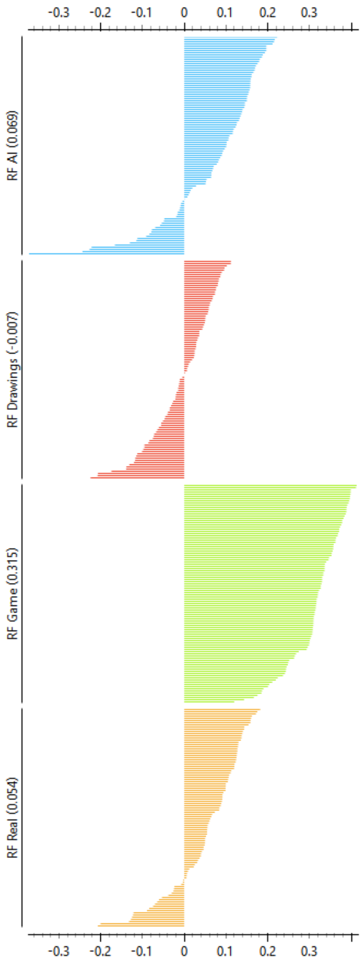

Figure 11. Silhouette Plot result using VGG-19

3.2 Discussions

This study categorizes red fox images utilizing Deep Learning models Inception V3, VGG-16, and VGG-19. Although category affects performance, the findings indicate that these models effectively differentiate between real-world, game, drawing, and artificial intelligence-generated images. Convolutional neural networks (CNNs) extracted extensive features from various image types and demonstrated efficacy in image classification tasks. Various filter sizes utilized in each layer result in Inception V3 exhibiting exceptional image accuracy in artificial intelligence. This corroborates previous research demonstrating the efficacy of multi-scale CNN architecture in handling challenging datasets [25].

VGG-16 and VGG-19 demonstrated significant accuracy in identifying game images due to their intricate architecture. Since none of the three models successfully clustered the drawing images, abstract or stylized images appear to impede feature extraction [7]. Deep learning models can identify this, thereby corroborating prior research that demonstrates sophisticated non-photorealistic content. Particularly for VGG-16 and VGG-19, which exhibited subpar performance on these datasets, the silhouette plot analysis revealed distinct clustering deficiencies.

Although all models exhibit remarkable clustering abilities, the t-SNE and MDS visualizations reveal that Inception V3 is particularly proficient at differentiating images generated by artificial intelligence, highlighting its superiority in handling synthetic media. Although these models are adequate, the ambiguous clusters identified in VGG-16 and VGG-19 indicate a possible lack of nuance for specific image categories in their feature extraction processes. The paper underscores the importance of tailored datasets and domain-specific training to improve classification accuracy in specialized fields, including artificial intelligence research [26].

Research indicates that image classification is essential for assessing animal populations, thereby rendering deep learning models pertinent for practical applications such as wildlife conservation. Artificial intelligence-generated images, particularly those produced by GANs, offer innovative opportunities in virtual environments and content creation [27]. Studies indicate that improved feature extraction, particularly for abstract image categories, yields more resilient models for diverse classification tasks.

This study elucidates the advantages and disadvantages of various deep learning models in the identification of red fox images. This facilitates the exploration of innovative concepts, such as hybrid models and attention mechanisms, to address contemporary challenges in image classification. Future research should concentrate on augmenting generalization and classification accuracy through feature extraction techniques, while simultaneously improving model efficiency across a broader spectrum of image categories [28].

4. Conclusions

This study presents the efficacy of deep learning models—inception V3, VGG-16, and VGG-19—in classifying red fox images into four separate categories: actual red fox images, game-captured, hand-drawn, and AI-generated. Despite variations in model performance indicated by Silhouette Scores, Inception V3 demonstrated significant efficacy in AI-generated images (0.244), whereas VGG-16 and VGG-19 achieved superior outcomes for game-captured images (0.398 and 0.315, respectively). T-SNE and MDS visualizations facilitated the validation of multiple cluster setups. Lowering decision thresholds will enhance the efficacy of Logistic Regression and Support Vector Machines (SVM) in hybrid classification systems, hence improving classification accuracy. Despite these advancements, the diversity of datasets posed challenges for categorization, since models demonstrated varying capacities based on the image type. To improve generalization, further investigation into adaptive learning methodologies, including transformer-based models and self-supervised learning, is crucial. Improving model resilience requires the incorporation of a broader range of information, including motion blur, background fluctuations, and changes in illumination. Future research should investigate integrated approaches that combine different deep learning architectures to improve stability and accuracy across diverse image categories. This research highlights the significance of deep learning in digital art analysis, AI-generated content classification, and wildlife monitoring, while also laying the groundwork for future advancements in automated image recognition.

Acknowledgements

The authors declare no conflict of interest.

References

[1] Z. Miao, Z. Liu, K. M. Gaynor, M. S. Palmer, S. X. Yu, and W. M. Getz, "Iterative human and automated identification of wildlife images," *Nat Mach Intell*, vol. 3, no. 10, pp. 885–895, Oct. 2021, doi: 10.1038/s42256-021-00393-0.

[2] Y. S. Vang, Z. Chen, and X. Xie, "Deep Learning Framework for Multi-class Breast Cancer Histology Image Classification," Feb. 2018, doi: https://doi.org/10.1007/978-3-319-93000-8_104.

[3] Y. Rahmati, "Artificial Intelligence for Sustainable Urban Biodiversity: A Framework for Monitoring and Conservation," Dec. 2024, [Online]. Available: http://arxiv.org/abs/2501.14766

[4] J. Wan and Y. Ma, "Multi-scale Spectral-Spatial Remote Sensing Classification of Coral Reef Habitats Using CNN-SVM," *J Coast Res*, vol. 102, no. sp1, Dec. 2020, doi: 10.2112/SI102-002.1.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[6] S. Christin, É. Hervet, and N. Lecomte, "Applications for deep learning in ecology," *Methods Ecol Evol*, vol. 10, no. 10, pp. 1632–1644, Oct. 2019, doi: 10.1111/2041-210X.13256.

[7] H. Chen and C. Haoyu, "Face Recognition Algorithm Based on VGG Network Model and SVM," *J Phys Conf Ser*, vol. 1229, no. 1, p. 012015, May 2019, doi: 10.1088/1742-6596/1229/1/012015.

[8] S. Binta Islam, D. Valles, T. J. Hibbitts, W. A. Ryberg, D. K. Walkup, and M. R. J. Forstner, "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images," *Animals*, vol. 13, no. 9, p. 1526, May 2023, doi: 10.3390/ani13091526.

[9] C. Duhart, G. Dublon, B. Mayton, and J. Paradiso, "Deep Learning Locally Trained Wildlife Sensing in Real Acoustic Wetland Environment," 2019, pp. 3–14. doi: 10.1007/978-981-13-5758-9_1.

[10] S. Sharma, S. Neupane, B. Gautam, and K. Sato, "Automated Multi-Species Classification Using Wildlife Datasets Based On Deep Learning Algorithms," *Materials, Methods & Technologies*, vol. 17, no. 1, pp. 103–117, 2023, doi: 10.62991/MMT1996359772.

[11] Y. Ohtsubo, T. Matsukawa, and E. Suzuki, "Semi-Supervised Few-Shot Classification with Deep Invertible Hybrid Models," May 2021, [Online]. Available: http://arxiv.org/abs/2105.10644

[12] J. Wang and Y. Jiang, "A Hybrid convolution neural network for the classification of tree species using hyperspectral imagery," *PLoS One*, vol. 19, no. 5, p. e0304469, May 2024, doi: 10.1371/journal.pone.0304469.

[13] Rashmi Ashtagi, "Fusion of AI Techniques: A Hybrid Approach for Precise Plant Leaf Disease Classification," *Journal of Electrical Systems*, vol. 20, no. 1s, pp. 850–861, Mar. 2024, doi: 10.52783/jes.836.

[14] K. Watabe, "Immersion and Manipulation of Time in The Legend of Zelda: Breath of the Wild (2017) and The Legend of Zelda: Tears of Kingdom (2023)," in *Panoramic and Immersive Media Studies Yearbook*, De Gruyter, 2024, pp. 55–76. doi: 10.1515/9783111335575-004.

[15] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2019, pp. 4396–4405. doi: 10.1109/CVPR.2019.00453.

[16] Onyebuchi Nneamaka Chisom, Preye Winston Biu, Aniekan Akpan Umoh, Bartholomew Obehioye Obaedo, Abimbola Oluwatoyin Adegbite, and Ayodeji Abatan, "Reviewing the role of AI in environmental monitoring and conservation: A data-driven revolution for our planet," *World Journal of Advanced Research and Reviews*, vol. 21, no. 1, pp. 161–171, Jan. 2024, doi: 10.30574/wjarr.2024.21.1.2720.

[17] T. Price, "A Cauldron of Chaos and Cultivation: Rediscovering Disney Animation of the 1980s," Thesis, Chapman University, Orange, CA, 2019. doi: 10.36837/chapman.000112.

[18] T. Yu, Y. Liu, H. Liu, J. Chen, and X. Wang, "ParaLkResNet: an efficient multi-scale image classification network," *Vis Comput*, vol. 40, no. 7, pp. 5057–5066, Jul. 2024, doi: 10.1007/s00371-024-03508-x.

[19] T. Talaei Khoei, H. Ould Slimane, and N. Kaabouch, "Deep learning: systematic review, models, challenges, and research directions," *Neural Comput Appl*, vol. 35, no. 31, pp. 23103–23124, Nov. 2023, doi: 10.1007/s00521-023-08957-4.

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946

[21] T. He, T. Wang, R. Abbey, and J. Griffin, "High-Performance Support Vector Machines and Its Applications," May 2019, [Online]. Available: http://arxiv.org/abs/1905.00331

[22] B. Ghasemkhani, K. F. Balbal, and D. Birant, "A New Predictive Method for Classification Tasks in Machine Learning: Multi-Class Multi-Label Logistic Model Tree (MMLMT)," *Mathematics*, vol. 12, no. 18, p. 2825, Sep. 2024, doi: 10.3390/math12182825.

[23] N. Ahmad and A. B. Nassif, "Dimensionality Reduction: Challenges and Solutions," *ITM Web of Conferences*, vol. 43, p. 01017, Mar. 2022, doi: 10.1051/itmconf/20224301017.

[24] X. C. Li, L. L. Liu, and L. K. Huang, "Comparison Of Several Remote Sensing Image Classification Methods Based On Envi," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3/W10, pp. 605–611, Feb. 2020, doi: 10.5194/isprs-archives-XLII-3-W10-605-2020.

[25] K. Kansal and S. Sharma, "Predictive Deep Learning: An Analysis of Inception V3, VGG16, and VGG19 Models for Breast Cancer Detection," 2024, pp. 347–357. doi: 10.1007/978-3-031-56703-2_28.

[26] L. Li, "Application of deep learning in image recognition," *J Phys Conf Ser*, vol. 1693, no. 1, p. 012128, Dec. 2020, doi: 10.1088/1742-6596/1693/1/012128.

[27] L. Bothmann *et al.*, "Automated wildlife image classification: An active learning tool for ecological applications," *Ecol Inform*, vol. 77, p. 102231, Nov. 2023, doi: 10.1016/j.ecoinf.2023.102231.

[28] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artif Intell Rev*, vol. 57, no. 1, p. 11, Jan. 2024, doi: 10.1007/s10462-023-10631-z.