Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

Classification of Roronoa Zoro Anime, Cosplay, and Action Figure Images Using VGG16 and Inception V3 with Logistic Regression and Support Vector Machine to Improve Popular Culture Object Recognition

Denaldy Oktavian Noor Rizki*[1], Imam Yuadi[2]

[1]Master's Program Human Resource Development-Data Analytics, Graduate School, Airlangga University, Indonesia
[2]Department of Information and Library Science, Faculty of Social and Political Sciences, Airlangga University, Indonesia

Email: [1]denaldy.oktavian.noor-2025@pasca.unair.ac.id

Received : Dec 3, 2025; Revised : Feb 7, 2026; Accepted : Feb 13, 2026; Published : Jun 15, 2026

Abstract

The diversity of visual representations of anime characters across anime scenes, cosplay photographs, and action figure images poses challenges for automated image classification due to variations in pose, lighting, background, and visual style. This study aims to develop a robust image classification system for the character Roronoa Zoro using deep learning–based feature extraction combined with classical classification algorithms. The method employs VGG16 and Inception V3 as feature extractors, followed by classification using Logistic Regression and Support Vector Machine. The dataset comprises three classes (anime, cosplay, and action figure), processed through image resizing, normalization, and data augmentation. Performance was evaluated using accuracy, F1-score, Area Under Curve (AUC), Matthews Correlation Coefficient (MCC), confusion matrix, silhouette plot, and multidimensional scaling. The experimental results show that Inception V3 combined with Logistic Regression achieved the best performance, with an AUC of 0.993, accuracy of 95.7%, F1-score of 0.957, and MCC of 0.935, outperforming VGG16 with Logistic Regression, which achieved 91.7% accuracy and an AUC of 0.986. Visualization-based evaluation indicates that Inception V3 produces more separable feature representations, particularly in distinguishing cosplay images from anime and action figure categories. This research demonstrates the effectiveness of multi-model feature extraction and classification for improving recognition performance in character-based image classification tasks and contributes empirically to the application of hybrid deep feature–machine learning approaches in computer vision.

Keywords : *Anime Character Classification, Image Classification, Inception V3, Logistic Regression, Support Vector Machine.*

1. INTRODUCTION

Machine learning has become a core technology in modern information systems, particularly in image classification, pattern recognition, and intelligent decision support. Among various machine learning paradigms, deep learning based on Convolutional Neural Networks (CNN) has demonstrated outstanding performance in extracting discriminative visual features from complex image data. Architectures such as VGG16 and Inception V3 have been widely adopted due to their ability to model hierarchical visual representations and achieve state-of-the-art accuracy in many benchmark image classification tasks. These models have been successfully applied in diverse application domains, including medical imaging, biometric recognition, and object detection [1], [2], [3].

In recent years, the rapid growth of digital content and popular culture has created new challenges in visual information processing, particularly in the automatic recognition of fictional and animated characters. Anime characters are distributed through multiple visual domains, including two-

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

dimensional animation, three-dimensional action figures, and real-world cosplay photography. The visual variations in artistic style, lighting, pose, costume, and physical structure significantly increase the complexity of automated character recognition. As a result, conventional single-domain image classification approaches often struggle to generalize across heterogeneous visual representations [4], [5].

Several recent studies have explored deep learning-based image classification using pretrained CNN architectures. Ahmed et al. applied transfer learning using CNN models for histopathology image classification and reported significant improvements in classification accuracy [6]. Hosni et al. demonstrated that VGG16 and Inception-based models provide robust feature representations for medical image prediction tasks [3]. Bakasa and Viriri employed VGG16 as a feature extractor combined with machine learning classifiers for cancer prediction [2]. Doungpaisan and Khunarsa reported that Inception V3 outperforms other pretrained CNN models in weather image classification [5]. Habibi and Yuadi compared VGG-based and Inception-based feature extraction combined with Logistic Regression and Support Vector Machine for image analysis and showed that deep feature embeddings significantly improve classification performance [7]. These studies confirm the effectiveness of deep CNN architectures in complex visual classification problems.

In addition, hybrid learning approaches that combine deep feature extraction with classical machine learning classifiers have received increasing attention. Awad et al. demonstrated that Logistic Regression achieves competitive performance when applied to deep feature embeddings for medical image classification [8]. Ganesh and Babu showed that Support Vector Machine outperforms Logistic Regression in complex classification problems involving high-dimensional feature spaces [9]. Allo et al. reported that the comparative performance of Logistic Regression and Support Vector Machine depends strongly on the quality of feature representation [10]. These findings indicate that integrating CNN-based feature extraction with traditional classifiers can provide a flexible and computationally efficient classification framework.

Despite the rapid progress of image classification research, most existing studies are still concentrated on medical imaging, biometric recognition, weather image analysis, and natural object classification, while research on anime character classification remains very limited, especially in heterogeneous visual domains. Previous works generally rely on a single visual domain and rarely examine cross-domain generalization across different representation forms. In addition, systematic comparisons between different pretrained CNN architectures and multiple machine learning classifiers for anime character recognition are still scarce. This study addresses these gaps by proposing a multi-model image classification framework for anime character recognition using deep learning–based feature extraction and machine learning classifiers.

The novelty of this research lies in three main contributions: (1) the construction of a heterogeneous dataset consisting of anime illustrations, cosplay photographs, and three-dimensional action figure images of the same character, Roronoa Zoro; (2) a comprehensive comparative evaluation of two pretrained CNN architectures, VGG16 and Inception V3, as feature extractors; and (3) the integration of deep feature representations with multiple machine learning classifiers, namely Logistic Regression and Support Vector Machine, to systematically investigate their effectiveness for cross-domain anime character classification.

Therefore, the objective of this study is to develop and evaluate an image classification system for Roronoa Zoro using VGG16 and Inception V3 for feature extraction combined with Logistic Regression and Support Vector Machine for classification. The study aims to analyze classification performance using multiple evaluation metrics and visual analytics techniques, including confusion matrix, silhouette analysis, and multidimensional scaling, in order to provide a robust framework for digital character recognition in popular culture applications.

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

2. METHOD

This study applies a machine learning–based image classification framework to recognize the anime character Roronoa Zoro from three visual domains: anime images, cosplay photographs, and action figure images. The overall research workflow is organized into four sequential stages: dataset collection, preprocessing and data augmentation, feature extraction and classification, and performance evaluation. The complete research pipeline is illustrated in Figure 1.

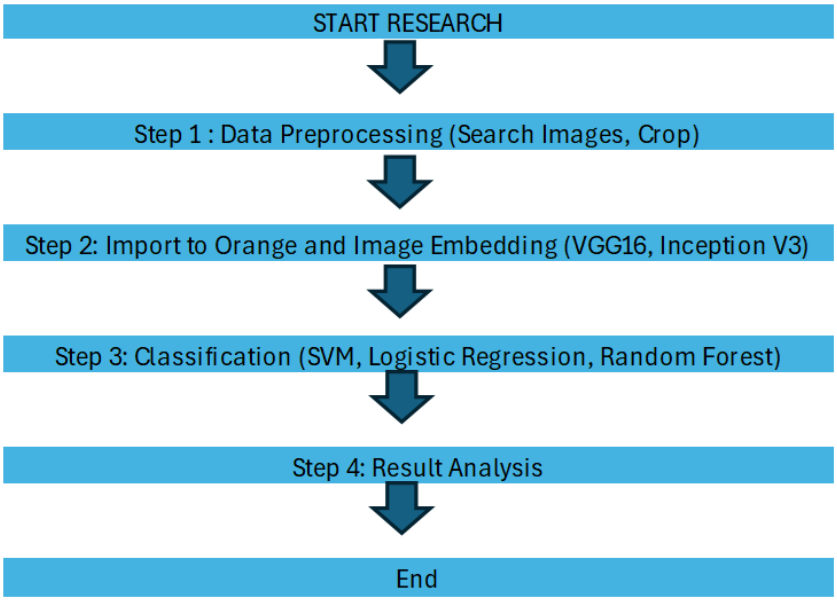

Figure 1. Research Workdlow

2.1. Dataset Collection

The dataset used in this study consists of images of the anime character Roronoa Zoro collected from three online sources representing different visual domains. The first dataset contains anime screenshots obtained from an official anime image repository (https://bit.ly/495gqkV). The second dataset consists of cosplay photographs collected from a cosplay image platform (https://bit.ly/4hRoacF), representing real-world interpretations of the character. The third dataset contains images of Roronoa Zoro action figures obtained from an online merchandise gallery (https://bit.ly/49cBUfS), representing three-dimensional object representations.

All collected images were manually filtered to ensure data quality based on the following criteria: (1) clear visibility of the character, (2) frontal or semi-frontal pose, (3) sufficient image resolution, and (4) minimal occlusion. Images that did not meet these criteria were excluded. The final dataset was constructed to ensure balanced class distribution across the three categories.

Each image was resized to 224 × 224 pixels to ensure compatibility with the VGG16 and Inception V3 architectures, which require fixed-size input images [11], [12].

2.2. Preprocessing and Data Augmentation

In the preprocessing stage, all images used in this study were first processed to ensure data quality and compatibility with the employed deep learning architectures. The dataset consists of photographs of the anime character Roronoa Zoro obtained from three different visual domains, namely anime images, cosplay photographs, and action figure images. Each image was cropped to remove irrelevant background regions and to focus the observation area on the main object, namely the character Roronoa Zoro. This step was performed to reduce visual noise that could negatively affect the feature extraction process.

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

After cropping, all images were normalized and resized to a resolution of 224 × 224 pixels. This standardization is required to ensure consistent input dimensions and compatibility with the VGG16 and Inception V3 architectures, enabling stable and efficient model training and precise feature extraction. Size compatibility is crucial to ensure that the CNN models process images correctly and generate reliable feature representations [13]. Let the original image be denoted as I0I_0I0 and the preprocessed image as III, then the normalization process is defined as:

$$I = \frac{I_0}{255} \quad (1)$$

To increase data diversity and reduce overfitting, data augmentation was applied using random rotation ($\theta \in [-\alpha, \alpha]$), horizontal flipping, random cropping, and zooming. Augmented images can be expressed as:

$$I' = T(I) \quad (2)$$

where $T(\cdot)$ denotes a geometric transformation operator applied to the input image.

Before classification, the pre-processing phase incorporated data augmentation techniques to enhance dataset diversity, including random rotations, horizontal flipping, random cropping, and zoom transformations. These operations increase visual variation in the dataset and improve the model's ability to generalize visual patterns under different viewing conditions. The application of data augmentation has been shown to effectively enhance dataset diversity and strengthen model robustness against complex visual variations, thereby improving overall classification performance [14].

Subsequent to preprocessing and feature extraction using VGG16 and Inception V3, the resulting image embeddings were input into a machine learning classification system employing Logistic Regression and Support Vector Machine (SVM) methods. The integration of deep feature extraction with classical machine learning classifiers enables effective learning from high-dimensional feature representations and improves the precision of anime character recognition across heterogeneous visual domains [8].

2.3. Processing

2.3.1 Feature Extraction (Image Embedding)

Image embedding plays a central role in computer vision and deep learning by transforming raw image data into compact and discriminative feature representations. In this study, image embedding is performed using two state-of-the-art Convolutional Neural Network (CNN) architectures, namely VGG16 and Inception V3. These models are widely recognized for their strong capability in extracting high-level semantic features from complex visual patterns and have been extensively adopted in image classification and feature extraction tasks.

Let an input image be denoted as:

$$I \in \mathbb{R}^{224\times224\times3} \quad (3)$$

The feature extraction function using a pretrained CNN $\Phi(\cdot)$ is formulated as:

$$\mathbf{f} = \Phi(I), \mathbf{f} \in \mathbb{R}^d \quad (4)$$

where $\mathbf{f}$ denotes the extracted feature vector and $\boldsymbol{d}$ is the embedding dimension. Two CNN architectures are used: VGG16 and Inception V3, both initialized with ImageNet pretrained weights (transfer learning). VGG16 is characterized by its deep and uniform architecture, consisting of sequential convolutional layers with small receptive fields that enable the model to learn hierarchical

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

visual representations effectively. In contrast, Inception V3 employs a more sophisticated multi-branch architecture that performs parallel convolutions with different kernel sizes, allowing the network to capture multi-scale features within the same layer. This architectural design enables Inception V3 to model both fine-grained and global visual characteristics more efficiently.

Both models are implemented using a transfer learning strategy with pretrained weights from the ImageNet dataset, which provides a strong initialization for feature extraction on limited target data. Transfer learning has been shown to significantly improve classification performance and reduce the risk of overfitting, particularly in scenarios with relatively small or heterogeneous datasets [6].

Previous studies have demonstrated the effectiveness of VGG16 and Inception V3 in various complex image classification tasks. Hosni et al. reported that these models achieve strong predictive performance in postoperative visual acuity estimation from retinal images [3], while Chen et al. showed their effectiveness in detecting malignant prostate tumors from MRI data [15]. In the domain of emotion recognition, embeddings generated by VGG and Inception architectures were shown to provide robust discriminative features [7]. Furthermore, Inception V3 achieved a classification accuracy of 96.1 percent in weather image classification tasks, outperforming several other pretrained CNN models [5].

The adaptability of these architectures across diverse visual domains highlights their suitability for heterogeneous image datasets. VGG16 excels in extracting consistent structural patterns through its deep sequential design, while Inception V3 is particularly effective in capturing complex visual variations through its multi-scale feature learning mechanism. As a result, both models provide complementary strengths for image embedding and serve as reliable feature extractors for subsequent classification using machine learning algorithms. Therefore, the combination of VGG16 and Inception V3 as feature extraction backbones provides a robust and scalable embedding framework for the classification of anime, cosplay, and action figure images of Roronoa Zoro, supporting accurate and generalizable character recognition across heterogeneous visual domains [4], [16], [17].

2.3.2 Classification Models

In the development of image classification systems, Logistic Regression and Support Vector Machine (SVM) are two widely used machine learning techniques due to their effectiveness and computational efficiency. Logistic Regression is a probabilistic statistical model that estimates the relationship between input features and class membership by modeling the posterior probability of each class. This method is commonly applied because of its simplicity, high interpretability, and stable performance across various application domains, including both medical and non-medical classification tasks [18], [19].

$$P(y = 1|f) = \frac{1}{1+e^{-(W^T f+b)}} \qquad (5)$$

Where,

- $\boldsymbol{P(y=1|f)}$ = posterior probability of the positive class
- $\boldsymbol{y}$ = class label
- f = feature vector
- $\boldsymbol{w}$ = weight vector
- $\boldsymbol{b}$ = bias term
- $\boldsymbol{e}$ = Euler's number (≈ 2.71828)

The optimization objective minimizes the binary cross-entropy loss:

$$\mathcal{L}_{LR} = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\widehat{y_i}) + (1 - y_i)\log(1 - \widehat{y_i}] \qquad (6)$$

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

Where,

- $\mathcal{L}_{LR}$ = Logistic regression loss
- N = number of samples
- y_i = true label of sample i
- $\hat{y}_i$ = predicted probability for sample i

Support Vector Machine, on the other hand, is a margin-based classifier designed to construct an optimal decision boundary by maximizing the separation margin between classes. SVM is well known for its capability to handle high-dimensional and complex feature spaces. Through the application of kernel functions, SVM can effectively model non-linear decision boundaries, making it more adaptable for complex classification problems compared to Logistic Regression, which typically follows a linear decision structure [9], [20].

$$\frac{1}{2}\|W\|^2 = \frac{1}{2}\left(W_1{}^2 + W_2{}^2\right) \quad (7)$$

Where W demonstrate as weiht vector. Several recent studies have demonstrated the complementary strengths of these two classification approaches. Jhee et al. evaluated Logistic Regression and SVM for predicting preeclampsia and showed that both models exhibit competitive performance depending on feature complexity and dataset characteristics [19]. Handayani reported that Logistic Regression achieved higher accuracy than SVM in heart disease classification, indicating its suitability for relatively less complex feature distributions [18]. Conversely, Ganesh and Babu found that SVM outperformed Logistic Regression in COVID-19 patient prediction tasks involving high-dimensional medical features, highlighting its advantage in more complex classification scenarios [9].

In general, an image classification pipeline consists of three main stages: data acquisition, feature extraction, and model training. The quality of feature representation plays a crucial role in determining classification performance. In this study, deep feature embeddings extracted from VGG16 and Inception V3 are used as input for both Logistic Regression and SVM classifiers. Cross-validation is applied to ensure model stability and generalization. Previous comparative studies indicate that while Logistic Regression offers better model interpretability, SVM often achieves higher accuracy on complex datasets through the use of appropriate kernel functions [20]. Therefore, evaluating both classifiers provides a comprehensive assessment of classification performance under different feature representations [21].

2.4 Evaluation

Model performance is evaluated using Accuracy, Precision, Recall, F1-score, Area Under Curve (AUC), and Matthews Correlation Coefficient (MCC). These metrics are selected to provide a comprehensive assessment of classification performance under potential class overlap and class imbalance conditions commonly found in heterogeneous image datasets. Accuracy measures overall correctness, Precision reflects prediction exactness, Recall captures sensitivity toward relevant instances, F1-score balances Precision and Recall, AUC evaluates the model's discriminative capability across decision thresholds, and MCC provides a robust correlation measure between predicted and actual classes, even under imbalanced class distributions.

The mathematical formulations of the evaluation metrics are defined as follows:

$$Akurasi = \frac{TP + TN}{TP+TN+FP+FN} \times 100\% \quad (8)$$

$$Presisi = \frac{Tp}{TP+FP} \times 100\% \quad (9)$$

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (10)$$

$$F1\ Score\ = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (11)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (12)$$

Where,

- TP = True Positive (number of correctly predicted positive samples)
- TN = True Negative (number of correctly predicted negative samples)
- FP = False Positive (number of negative samples incorrectly predicted as positive)
- FN = False Negative (number of positive samples incorrectly predicted as negative)

In addition to the scalar metrics, a confusion matrix is employed to analyze misclassification patterns across classes. The confusion matrix provides a detailed breakdown of correct and incorrect predictions, enabling the computation of Accuracy, Precision, Recall, F1-score, and MCC, as well as revealing specific error tendencies between visually similar categories (e.g., cosplay versus anime representations). This analysis is crucial for assessing the reliability of the classification system in cross-domain character recognition scenarios. The confusion matrix used in this study is illustrated in Table 1.

Table 1. Confusion Matrix for Binary Classification Model

Actual Class \ Predicted Class	Positive Class	Negative Class
Positive Class	TP (True Positive)	FP (False Positive)
Negative Class	FN (False Negative)	TN (True Negative)

To provide qualitative insights into the separability of feature representations, silhouette coefficient and multidimensional scaling (MDS) are applied to the embedded feature space. The silhouette coefficient is defined as:

$$\boldsymbol{s_i} = \frac{\boldsymbol{b_i - a_i}}{\boldsymbol{\max\{a_i, b_i\}}} \quad (14)$$

Where,

- $\boldsymbol{s_i}$= silhouette value of sample $\boldsymbol{i}$
- $\boldsymbol{a_i}$ = average intra-class distance of sample $\boldsymbol{i}$
- $\boldsymbol{b_i}$ = minimum average inter-class distance of sample i

These evaluation strategies jointly provide quantitative performance measures and qualitative visual evidence of feature discriminability across heterogeneous visual domains, enabling a comprehensive assessment of the proposed multi-model image classification framework.

2.5 Analysis

This study evaluates the performance of deep learning–based image classification models using the VGG16 and Inception V3 architectures as feature extractors. Both models were selected due to their proven effectiveness in high-dimensional visual feature learning and their strong performance across a wide range of image classification benchmarks. The evaluation focuses on the ability of each model to distinguish between three visual categories of the character Roronoa Zoro, namely Action Figure, Anime, and Cosplay.

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

Figure 2. Picture of Rorono Zoro in Action Figure, Anime and Cosplay

To assess classification performance, a confusion matrix was employed to analyze the distribution of predicted labels across the three categories. The dataset was carefully curated from verified online sources and underwent a validation process to ensure balanced class representation and consistent image quality. This design ensures that the evaluation reflects the true discriminative capability of the models rather than dataset bias.

The confusion matrix analysis shows that the Anime category achieved the highest classification accuracy, with 94 images correctly classified and only 4 instances misclassified as Action Figure. This result indicates that the model is highly effective in learning the distinctive visual patterns of two-dimensional anime illustrations, which are characterized by unique color compositions, line structures, and shading styles.

The Cosplay category also demonstrates strong classification performance, with 90 images correctly identified. However, a small number of misclassifications occurred, with 6 images incorrectly predicted as Action Figure and 4 images misclassified as Anime. These errors can be attributed to visual similarities between cosplay photographs and action figure images, particularly in costume texture, pose, and lighting conditions.

Overall, the results confirm that both VGG16 and Inception V3 exhibit strong discriminative capabilities for multi-domain image classification. The observed misclassifications suggest that images with similar visual attributes across domains remain challenging for automated recognition systems. These findings are consistent with previous studies demonstrating that deep learning models are effective in capturing complex visual features for object classification tasks **[7]**. The analysis highlights the importance of dataset diversity and feature robustness in improving classification performance across heterogeneous visual domains.

3 RESULT

This section presents the experimental results of the proposed image classification framework for recognizing the character Roronoa Zoro across three visual domains: Anime, Cosplay, and Action Figure. The results are reported following the methodological stages described in Section 2, starting from model performance evaluation, followed by confusion matrix analysis, feature distribution assessment, and low-dimensional visualization.

3.1 Model Performance Evaluation

Table 1 summarizes the performance of four machine learning classifiers Logistic Regression, Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN) combined with two deep feature extraction models, namely VGG16 and Inception V3. The evaluation metrics include Area Under Curve (AUC), Classification Accuracy (CA), Precision, Recall, F1-score, and Matthews Correlation Coefficient (MCC), which together provide a comprehensive assessment of classification reliability and robustness.

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

Table 2. Research Model Performance

Model	Feature Extraction	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	VGG-16	0.986	0.917	0.917	0.917	0.917	0.875
SvM		0.975	0.890	0.891	0.893	0.890	0.836
Random Forest		0.950	0.847	0.848	0.851	0.847	0.771
kNN		0.952	0.843	0.845	0.861	0.843	0.772
Logistic Regression	Incception V3	0.993	0.957	0.957	0.957	0.957	0.935
SvM		0.992	0.947	0.947	0.948	0.947	0.920
Random Forest		0.907	0.790	0.791	0.794	0.790	0.686
kNN		0.972	0.897	0.898	0.907	0.897	0.849

Using VGG16 as the feature extractor (Table 2), Logistic Regression achieved the highest AUC value of 0.986 and the highest classification accuracy of 0.917, outperforming SVM (0.890), Random Forest (0.847), and KNN (0.843). In addition, Logistic Regression recorded superior F1-score, Precision, Recall, and MCC values of 0.917, 0.917, 0.917, and 0.875, respectively. These results indicate that Logistic Regression is highly effective in exploiting VGG16 feature embeddings for multi-class anime character recognition.

When Inception V3 was employed as the feature extractor, overall classification performance improved for all models. Logistic Regression achieved the best performance with an AUC of 0.993 and a classification accuracy of 0.957. SVM and Random Forest also showed strong performance with identical CA values of 0.947, while KNN obtained the lowest accuracy of 0.897. The improvement in AUC from 0.986 (VGG16) to 0.993 (Inception V3) and the increase in accuracy from 91.7 percent to 95.7 percent demonstrate that Inception V3 provides more discriminative and robust feature representations for heterogeneous image domains.

The Precision and Recall metrics further confirm the reliability of Logistic Regression. Using Inception V3, Logistic Regression achieved Precision and Recall values of 0.957, indicating excellent capability in minimizing both false positive and false negative predictions. Similarly, the F1-score of 0.957 demonstrates strong balance between Precision and Recall. The MCC value of 0.935 further confirms high-quality classification performance under class-balanced conditions.

3.2 Confusion Matrix Analysis

To further analyze class-wise prediction behavior, confusion matrices were generated for both feature extraction models.

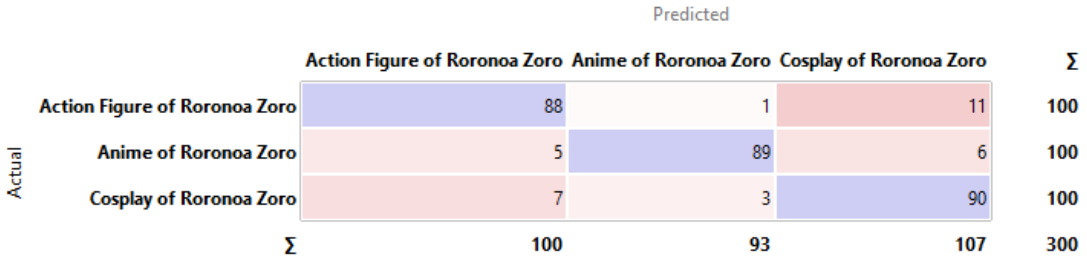

Figure 3. Confussion Matrix result (VGG-16)

Figure 3 presents the confusion matrix obtained using VGG16. The diagonal values indicate correct predictions, with 88 images correctly classified as Action Figure, 89 as Anime, and 90 as Cosplay. The model shows strong overall performance; however, minor misclassifications are observed. One Action Figure image was incorrectly predicted as Cosplay, and five Anime images were misclassified as Action Figure. These errors suggest that visual similarity in costume structure and pose may introduce ambiguity in the classification process.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

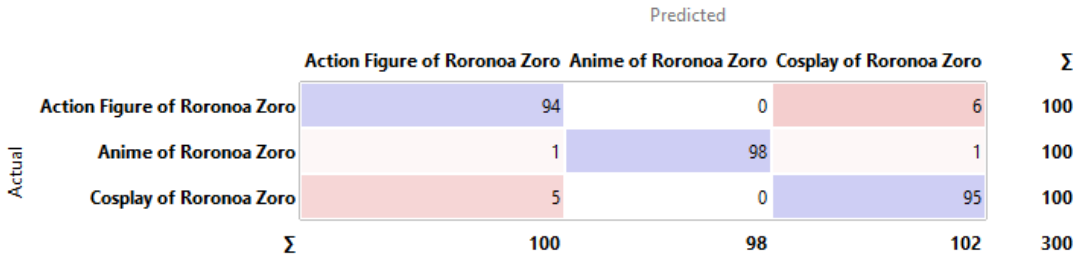

Figure 4. Confussion Matrix result (Inception V3)

Figure 4 shows the confusion matrix generated using Inception V3. The model achieves higher classification accuracy, with 94 correct predictions for Action Figure, 98 for Anime, and 95 for Cosplay. Only a small number of misclassifications occurred, indicating that Inception V3 produces more discriminative feature embeddings. Compared to VGG16, the total number of misclassified samples is reduced, confirming that Inception V3 improves class separability and generalization.

These confusion matrix results directly support the quantitative evaluation in Table 2 and confirm that Inception V3 enhances the reliability of anime character recognition across heterogeneous image categories.

3.3 Feature Distribution Analysis Using Silhouette Plot

To evaluate the separability of feature embeddings, silhouette analysis was performed.

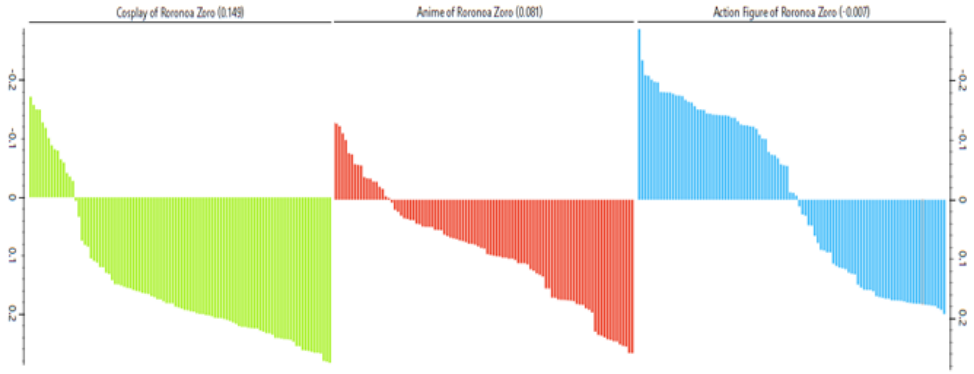

Figure 5. Silhouette Plot result (VGG-16)

Figure 5 shows the silhouette plot generated from VGG16 feature embeddings. The Action Figure and Cosplay categories exhibit predominantly positive silhouette values, indicating strong cluster cohesion. However, the Anime category demonstrates more dispersed values, suggesting partial overlap with other classes. This explains the misclassification patterns observed in the VGG16 confusion matrix.

Figure 6. Silhouette Plot result (Inception V-3)

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

Figure 6 presents the silhouette plot obtained from Inception V3 embeddings. The Action Figure and Cosplay categories show consistently high positive silhouette values, indicating well-separated clusters. The Anime category exhibits greater variation, reflecting stylistic similarities between anime and action figure images. Nevertheless, the overall silhouette distribution for Inception V3 is more compact and stable than that of VGG16, indicating superior feature separability.

These results demonstrate that Inception V3 generates more discriminative embeddings, leading to improved classification stability and robustness.

3.4 Multidimensional Scaling Visualization

To visualize feature similarity in a low-dimensional space, Multidimensional Scaling (MDS) was applied.

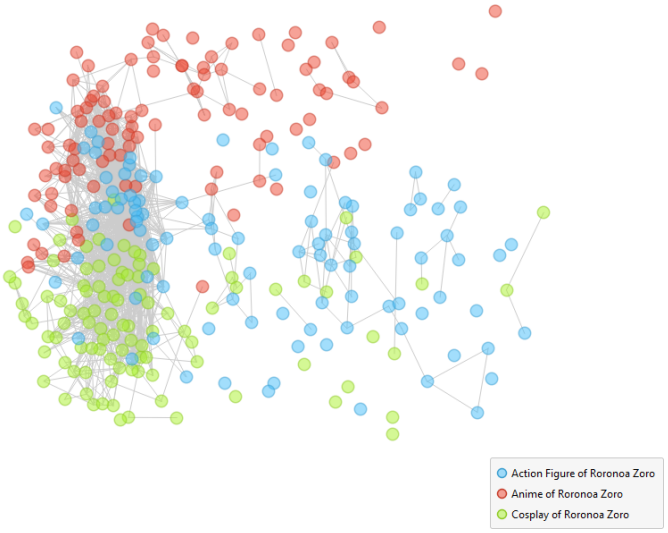

Figure 7. MDS result (VGG-16)

Figure 7 illustrates the MDS projection of VGG16 embeddings. The Cosplay cluster is clearly separated, while the Anime and Action Figure clusters exhibit noticeable overlap. This indicates that VGG16 has limitations in distinguishing two-dimensional anime illustrations from three-dimensional action figures due to similarities in color composition, pose geometry, and outline structures.

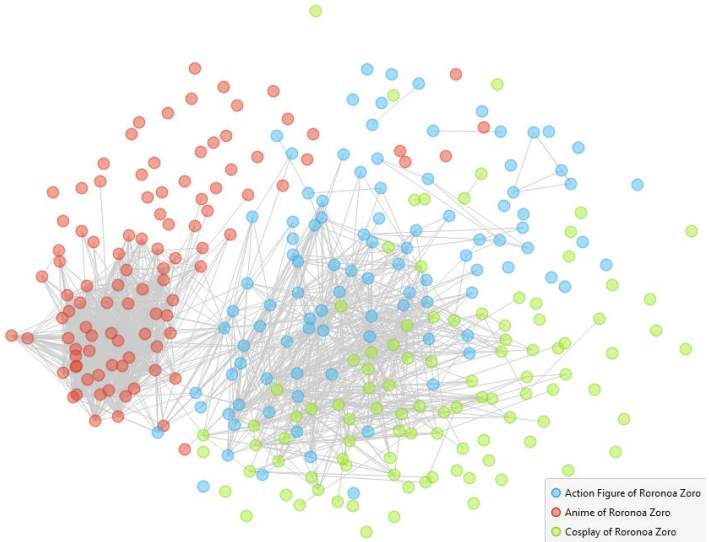

Figure 8. MDS result (Inception V-3)

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

Figure 8 presents the MDS visualization for Inception V3 embeddings. The Cosplay category forms a distinct cluster, and the Anime and Action Figure categories exhibit clearer separation compared to VGG16. This result demonstrates the effectiveness of Inception V3 in capturing multi-scale visual features and preserving semantic structure across different visual domains.

The improved cluster separation in Figure 8 directly explains the higher classification accuracy and AUC values achieved by Inception V3 in Table 1.

3.5 Quantitative Interpretation of Results

The experimental results demonstrate that the proposed framework successfully classifies multi-domain character images with high accuracy. The combination of Inception V3 and Logistic Regression achieves the best overall performance, with an accuracy of 95.7 percent, an AUC of 0.993, an F1-score of 0.957, and an MCC of 0.935. These values indicate strong discriminative capability, balanced prediction performance, and robust generalization.

The confusion matrix, silhouette analysis, and MDS visualization consistently indicate that Cosplay images are the easiest category to distinguish due to strong human facial and texture cues, while Anime and Action Figure images present greater classification challenges because of their stylistic similarities. These findings highlight the importance of multi-scale feature extraction and dataset diversity in improving cross-domain generalization.

The results validate the effectiveness of the proposed deep learning and machine learning integration for multi-domain anime character classification. Inception V3 provides superior feature embeddings, while Logistic Regression offers strong classification robustness. The quantitative analysis confirms that the proposed system successfully achieves the research objective of accurate and reliable digital character recognition across heterogeneous visual domains.

4 DISCUSSIONS

The experimental results demonstrate that the proposed image classification framework based on deep learning feature extraction and machine learning classification is effective in recognizing multi-domain representations of the character Roronoa Zoro. The confusion matrix analysis confirms that both VGG16 and Inception V3 are capable of learning discriminative visual patterns from heterogeneous image sources, including anime illustrations, cosplay photographs, and action figure images. However, Inception V3 consistently outperforms VGG16 in terms of classification accuracy, cluster separability, and robustness across visually similar categories. This finding aligns with the architectural design of Inception V3, which is optimized to capture multi-scale visual features through parallel convolutional paths [22].

The superior performance of Inception V3 observed in this study is consistent with several previous works across different image classification domains. Xu et al. (2025) reported that Inception V3 achieved higher accuracy than VGG-based models in weather image classification due to its stronger ability to represent spatial patterns at different resolutions. Similar performance gains were also reported by Alaca in medical image analysis, where Inception-based architectures showed better generalization on complex visual structures compared to conventional CNN backbones [23]. The present findings extend these results into the domain of anime character recognition, which combines artistic two-dimensional illustrations and real-world three-dimensional object representations. This extension indicates that multi-scale feature extraction is not only effective for natural and medical images but also robust for culturally constructed visual content.

The silhouette analysis and multidimensional scaling (MDS) visualization further confirm that Inception V3 generates more compact and well-separated feature clusters than VGG16. This result is in line with Al-Omari et al. who demonstrated that deep CNN feature embeddings significantly improve

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

cluster separability in histopathology image classification [24]. Moreover, the effectiveness of combining deep feature extraction with classical machine learning classifiers observed in this study corroborates the findings of Bilal et al. who showed that CNN features combined with Logistic Regression and Support Vector Machine can achieve competitive performance with lower computational cost than end-to-end deep learning models [23]. In the present study, Inception V3 combined with Logistic Regression achieved an accuracy of 95.7 percent and an AUC of 0.993, indicating that high discriminative power can be obtained through hybrid architectures.

The observed misclassifications between the Anime and Action Figure categories reveal persistent challenges in cross-domain visual recognition. This phenomenon is consistent with Lhermitte et al. who reported that visually similar textures and object shapes can lead to classification ambiguity even when deep learning models are applied [25]. In the context of anime character recognition, stylistic similarities between two-dimensional illustrations and three-dimensional figurines, such as pose geometry, costume design, and color patterns, generate overlapping visual cues that remain difficult to disentangle. This suggests that future research may benefit from incorporating domain adaptation or contrastive learning to reduce feature overlap across visual domains.

From the perspective of informatics and computer science, this study provides empirical evidence that hybrid deep learning–machine learning architectures can serve as an efficient alternative to fully end-to-end deep neural networks for real-world visual recognition systems. The demonstrated performance of Inception V3 combined with Logistic Regression shows that high-accuracy image classification can be achieved with lower computational complexity, making this approach suitable for deployment in resource-constrained environments such as mobile applications, edge devices, and multimedia content management systems. This contributes to the development of scalable computer vision systems within applied informatics, particularly in domains where computational resources and energy efficiency are critical design constraints.

Furthermore, this research broadens the application domain of computer vision in cultural and creative informatics. While most existing studies focus on medical imaging, surveillance, and industrial inspection, the present work highlights the feasibility of intelligent visual recognition in digital entertainment, virtual communities, and multimedia information retrieval systems. By addressing the classification of anime characters across heterogeneous visual domains, this study contributes to the growing body of knowledge in multimedia informatics and cultural computing, and demonstrates how computer vision techniques can be adapted to support digital content analysis in popular culture contexts.

The results not only confirm the effectiveness of deep learning–based feature extraction for multi-domain anime character recognition but also strengthen the theoretical and practical understanding of hybrid classification frameworks in computer vision. The findings reinforce prior evidence on the advantages of multi-scale CNN architectures [26] and extend their applicability to cultural visual data. This study thus contributes to the advancement of image representation learning, efficient visual classification pipelines, and applied multimedia informatics, while opening new research directions in intelligent digital content analysis and cross-domain visual recognition.

Therefore, the objective of this study is to develop and evaluate an image classification system for Roronoa Zoro using VGG16 and Inception V3 as feature extractors combined with Logistic Regression and Support Vector Machine as classifiers, and to empirically demonstrate that multi-scale deep feature extraction (Inception V3) provides superior discriminative representations for multi-domain visual data. The study aims to analyze classification performance using quantitative metrics (accuracy, precision, recall, F1-score, AUC, and MCC) and visual analytics techniques (confusion matrix, silhouette analysis, and multidimensional scaling), in order to propose an efficient and scalable

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

framework for digital character recognition within multimedia informatics and cultural computing applications.

5 CONCLUSION

This study presents a structured and empirically validated image classification framework for recognizing the anime character Roronoa Zoro across three heterogeneous visual domains, namely Anime, Cosplay, and Action Figure. The results demonstrate that deep learning–based feature extraction using VGG16 and Inception V3, when combined with classical machine learning classifiers, is effective in capturing discriminative visual representations from complex and stylistically diverse image data. The confusion matrix, MDS visualization, and silhouette analysis consistently show that Inception V3 provides superior feature separability and robustness in handling cross-domain visual variations involving two-dimensional illustrations, three-dimensional objects, and real-world photography.

The main scientific contribution of this research lies in providing empirical evidence that hybrid deep learning–machine learning architectures can achieve high recognition performance while maintaining computational efficiency. By demonstrating that Inception V3 combined with Logistic Regression attains an accuracy of 95.7 percent and an AUC of 0.993, this study contributes to the methodological development of efficient computer vision pipelines, particularly for scenarios where full end-to-end deep models are impractical due to resource constraints. This contribution strengthens the understanding of feature representation learning and classifier integration within applied computer vision research.

From the perspective of informatics, this study advances the design of scalable and deployable visual recognition systems by showing that high-performance image classification can be achieved through lightweight classifiers operating on robust deep features. This has practical implications for the development of intelligent multimedia systems, digital content management platforms, and interactive applications that require reliable visual understanding with limited computational resources. Furthermore, by situating the problem within digital entertainment and cultural content analysis, this research expands the application domain of computer vision beyond conventional industrial and medical contexts, reinforcing the role of informatics in supporting creative and cultural industries through intelligent content understanding.

For future research, several specific directions are recommended. First, future studies should construct a multi-character benchmark dataset that includes multiple anime characters with controlled inter-class similarity levels to evaluate generalization performance under higher visual ambiguity. Second, domain adaptation or contrastive representation learning should be explored to explicitly reduce feature overlap between visually similar domains, such as Anime and Action Figure, which were observed to produce misclassifications. Third, multimodal learning approaches that integrate visual features with textual metadata (e.g., character profiles, attributes, and narrative context) can be investigated to improve semantic discrimination across domains. Fourth, comparative experiments involving Vision Transformers and self-supervised pretraining should be conducted to assess their effectiveness in learning domain-invariant representations for character recognition tasks. Finally, future work should implement and evaluate the proposed framework in real-time environments, such as mobile or web-based platforms, to measure latency, scalability, and user-level performance in practical deployment scenarios.

ACKNOWLEDGEMENT

Thank you to Mr. Imam Yuadi as my supervisor and also to my colleagues who have helped in writing this article.

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

REFERENCES

[1] R. Nakasi, E. Mwebaze, A. Zawedde, J. Tusubira, B. Akera, and G. Maiga, "A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models," *SN Applied Sciences*, vol. 2, no. 7, 2020, doi: 10.1007/s42452-020-3000-0.

[2] W. Bakasa and S. Viriri, "VGG16 feature extractor with extreme gradient boost classifier for pancreas cancer prediction," *Journal of Imaging*, vol. 9, no. 7, p. 138, 2023, doi: 10.3390/jimaging9070138.

[3] S. Hosni *et al.*, "Prediction of postoperative visual acuity in rhegmatogenous retinal detachment using OCT images," *IEEE Access*, vol. 11, pp. 135435–135448, 2023, doi: 10.1109/ACCESS.2023.3338362.

[4] A. Ciptaningrum, R. Apriyanto, D. Prakoso, R. Yudha, and M. Echsony, "Data-driven anomaly control detection for railroad lines using Sobel filter and VGG-16 model, Res-Net50, InceptionV3," *Journal Geuthee of Engineering and Energy*, vol. 2, no. 1, pp. 31–46, 2023, doi: 10.52626/joge.v2i1.17.

[5] P. Doungpaisan and P. Khunarsa, "A comparative study of pre-trained models for image feature extraction in weather image classification using Orange Data Mining," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 37, no. 1, p. 241, 2025, doi: 10.11591/ijeecs.v37.i1.pp241-249.

[6] S. Ahmed *et al.*, "Transfer learning approach for classification of histopathology whole slide images," *Sensors*, vol. 21, no. 16, p. 5361, 2021, doi: 10.3390/s21165361.

[7] W. Habibi and I. Yuadi, "Evaluating logistic regression and SVM for image analysis using VGG-16, VGG-19, and Inception V3 features," *Jurnal Ilmiah Teknologi Dan Rekayasa*, vol. 30, no. 2, pp. 136–155, 2025, doi: 10.35760/tr.2025.v30i2.14056.

[8] F. Awad, M. Hamad, and L. Alzubaidi, "Robust classification and detection of big medical data using advanced parallel k-means clustering, YOLOv4, and logistic regression," *Life*, vol. 13, no. 3, p. 691, 2023, doi: 10.3390/life13030691.

[9] K. Ganesh and C. Babu, "Predictive analysis for identifying survival of COVID-19 patients using support vector machine over logistic regression," *APC*, 2022, doi: 10.3233/APC220032.

[10] C. Allo, L. Putra, N. Paranoan, and V. Gunawan, "Comparing logistic regression and support vector machine in breast cancer problem," *Jambura Journal of Probability and Statistics*, vol. 4, no. 1, pp. 1–8, 2023, doi: 10.34312/jjps.v4i1.19246.

[11] E. Corcoran, K. Hosseini, L. Siles, S. Kurup, and S. Ahnert, "Automated dynamic phenotyping of whole oilseed rape (Brassica napus) plants from images collected under controlled conditions," *Frontiers in Plant Science*, vol. 16, 2025, doi: 10.3389/fpls.2025.1443882.

[12] F. Mohanty and C. Dora, "An optimized KELM approach for the diagnosis of COVID-19 from 2D-SSA reconstructed CXR images," *Optik*, vol. 244, p. 167572, 2021, doi: 10.1016/j.ijleo.2021.167572.

[13] Y. Tang *et al.*, "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *NPJ Digital Medicine*, vol. 3, no. 1, 2020, doi: 10.1038/s41746-020-0273-z.

[14] S. Chatterjee and Y. Byun, "Voting ensemble approach for enhancing Alzheimer's disease classification," *Sensors*, vol. 22, no. 19, p. 7661, 2022, doi: 10.3390/s22197661.

[15] Q. Chen, S. Hu, P. Long, L. Fang, Y. Shi, and Y. Li, "A transfer learning approach for malignant prostate lesion detection on multiparametric MRI," *Technology in Cancer Research & Treatment*, vol. 18, 2019, doi: 10.1177/1533033819858363.

[16] H. Hoefling *et al.*, "Histonet: A deep learning-based model of normal histology," *Toxicologic Pathology*, vol. 49, no. 4, pp. 784–797, 2021, doi: 10.1177/0192623321993425.

[17] Y. Wang *et al.*, "Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning," *IEEE Access*, vol. 8, pp. 27779–27792, 2020, doi: 10.1109/ACCESS.2020.2964276.

[18] F. Handayani, "Komparasi support vector machine, logistic regression dan artificial neural network dalam prediksi penyakit jantung," *Jurnal Edukasi Dan Penelitian Informatika*, vol. 7, no. 3, p. 329, 2021, doi: 10.26418/jp.v7i3.48053.

Vol. 7, No. 3, June 2026, Page. 2561-2576
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5516

[19] J. Jhee *et al.*, "Prediction model development of late-onset preeclampsia using machine learning-based methods," *PLoS One*, vol. 14, no. 8, p. e0221202, 2019, doi: 10.1371/journal.pone.0221202.

[20] J. Yee, S. Oh, S. Kim, J. Kim, J. Chung, and H. Gwak, "Machine learning approaches for predicting bisphosphonate-related osteonecrosis in women with osteoporosis using VEGFA gene polymorphisms," *Journal of Personalized Medicine*, vol. 11, no. 6, p. 541, 2021, doi: 10.3390/jpm11060541.

[21] A. Gullapalli and V. Mittal, "Early detection of Parkinson's disease through speech features and machine learning: A review," in *Machine Learning for Healthcare Applications*, 2021, pp. 203–212. doi: 10.1007/978-981-16-4177-0_22.

[22] I. Khalil, A. Mehmood, H. Kim, and J. Kim, "OCTNet: A Modified Multi-Scale Attention Feature Fusion Network with InceptionV3 for Retinal OCT Image Classification," *Mathematics*, vol. 12, no. 19, p. 3003, Sep. 2024, doi: 10.3390/math12193003.

[23] Y. Alaca, "Machine learning via DARTS-Optimized MobileViT models for pancreatic Cancer diagnosis with graph-based deep learning," *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, p. 81, Feb. 2025, doi: 10.1186/s12911-025-02923-x.

[24] O. Al-Omari, O. Alkhatib, and T. Al-Omari, "CNN-Based Automated Detection of Metastatic Cancer in Histopathology Images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24478–24485, Aug. 2025, doi: 10.48084/etasr.10888.

[25] E. Lhermitte, M. Hilal, R. Furlong, V. O'Brien, and A. Humeau-Heurtier, "Deep Learning and Entropy-Based Texture Features for Color Image Classification," *Entropy*, vol. 24, no. 11, p. 1577, Oct. 2022, doi: 10.3390/e24111577.

[26] S. Ghadai, X. Y. Lee, A. Balu, S. Sarkar, and A. Krishnamurthy, "Multi-resolution 3D CNN for learning multi-scale spatial features in CAD models," *Comput. Aided Geom. Des.*, vol. 91, p. 102038, Nov. 2021, doi: 10.1016/j.cagd.2021.102038.