International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Faster-RCNN Performance: A Hybrid Approach Integrating VAE, HOG, and Recursive Feature Elimination

Rozzi Kesuma Dinata*[1], Fajriana[2], Novia Hasdyna[3], Sujacka Retno[3]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Deep learning has transformed image analysis by providing powerful feature extraction techniques, particularly beneficial for detecting microscopic pathogens such as Vibrio parahaemolyticus. Traditional methods often fail to capture the complex, high-dimensional features required for accurate bacterial classification. This study presents the enhancement of Faster-RCNN performance through advanced feature extraction techniques for the classification of Vibrio parahaemolyticus bacteria. The goal is to improve classification accuracy by integrating sophisticated feature extraction methods. Microscopic images observed directly are initially processed using the ResNet architecture to derive preliminary features. Variational Autoencoders (VAE) are then employed to extract high-dimensional, abstract features, while Histogram of Oriented Gradients (HOG) captures shape and orientation based features. Recursive Feature Elimination (RFE) is used to optimize these feature sets, leading to significant improvements in classification performance. The VAE+RFE + Faster-RCNN approach achieves a detection accuracy of 92%, precision of 88%, recall of 93%, and an F1-score of 90%. In comparison, the HOG + RFE + Faster-RCNN configuration results in 85% accuracy, 82% precision, 87% recall, and an F1-score of 84%. The conventional Faster-RCNN model records an accuracy of 89%, precision of 86%, recall of 90%, and an F1-score of 88%. These results underscore the substantial impact of advanced feature extraction techniques on optimizing Faster-RCNN performance, demonstrating significant improvements in classification accuracy and efficiency.

Keywords: *feature extraction, variational autoencoders, histogram of oriented gradients, recursive feature elimination, faster r-cnn optimization*

1. Introduction

The advancements in artificial intelligence (AI) and machine learning (ML) have had a significant impact across numerous fields, including biomedical imaging, genomics, and environmental monitoring [1]. Exceptional capabilities in image recognition, classification, and anomaly detection have been demonstrated by deep learning, a powerful subset of machine learning[2]. Convolutional Neural Networks (CNNs), particularly ResNet, have proven highly effective in processing complex image data, making them well-suited for identifying microscopic bacterial infections. Over the past five years, DL has shown its effectiveness in object detection across diverse domains [3]. Despite these advancements, significant challenges are still encountered by the Faster Region-Based Convolutional Neural Network (Faster-RCNN), particularly in extracting and utilizing intricate features from images. The challenges are further amplified in the context of microscopic images, where fine details are crucial for accurate classification. While Faster-RCNN has demonstrated effectiveness in various applications, its conventional feature extraction methods may not fully capture the subtle patterns necessary for bacterial infection detection [4,5].

Several studies have addressed the limitations of conventional feature extraction methods in enhancing object detection capabilities, particularly in challenging imaging scenarios. Research [6] demonstrated that Wavelet Transform effectively extracts features by capturing both spatial and frequency domain information, thereby improving object detection in complex environments. Another study [7] integrated LBP (Local Binary Patterns) with Faster-RCNN to capture texture information, resulting in a significant improvement in detecting objects with distinctive textures. Furthermore, the research referenced in [8] explored the application of SIFT and SURF for feature extraction, showing that these methods enhance feature extraction by handling variations in scale and rotation, thereby improving object detection accuracy when used with Faster-RCNN.

Furthermore, [9] employed Gabor Filters to capture texture and edge information, demonstrating that these filters can significantly enhance feature representation and improve object detection performance in visually complex scenes. Recent studies [10] have also investigated the integration

[1]*Associate Professor, Department of Informatics, Universitas Malikussaleh, Lhokseumawe 24355, Aceh, Indonesia*
Emaill: rozzi@unimal.ac.id
[2]*Associate Professor, Department of Informatics, Universitas Malikussaleh, Lhokseumawe 24355, Aceh, Indonesia*
Emaill: fajriana@unimal.ac.id
[3]*Assistant Professor, Department of Informatics, Universitas Islam Kebangsaan Indonesia, Bireuen 24251, Aceh, Indonesia*
Email: noviahasdyna@uniki.ac.id
[4]*Assistant Professor, Department of Informatics, Universitas Malikussaleh, Lhokseumawe 24355, Aceh, Indonesia*
Emaill: sujacka@unimal.ac.id
** Corresponding Author Email: rozzi@unimal.ac.id*

of color and texture-based features, which has shown promise in further enhancing detection accuracy. In addressing these challenges, this study introduces an innovative methodology designed to enhance Faster-RCNN performance by employing advanced feature extraction techniques for the classification of Vibrio parahaemolyticus bacteria. The key objec-tives of this study are to:

- Employ Variational Autoencoders (VAE) and Histogram of Oriented Gradients (HOG) to extract high-dimensional and shape-based features, with VAE mod-eling complex patterns to enhance feature representation and HOG capturing detailed shape and orientation information.
- Optimize the extracted features using Recursive Feature Elimination (RFE) to improve their relevance and efficiency, and apply these features within the Faster-RCNN framework to enhance classification accuracy.
- Evaluate the effectiveness of the integrated approach using performance metrics such as the Confusion Matrix and AUC (Area Under the Curve) - ROC (Receiver Operating Characteristic) to assess the robustness and accuracy of the detection system.

By achieving these objectives, this study aims to significantly enhance Faster-RCNN's classification performance through advanced feature extraction techniques, offering enhanced reliability and precision in comparison to conventional methods.

The structure of the paper is as follows: Section 2 examines existing research on object detection and various feature extraction methods. Section 3 outlines the approach for advanced feature extraction using Variational Autoencoders (VAE) and Histogram of Oriented Gradients (HOG), along with feature optimization through Recursive Feature Elimination (RFE) to enhance Faster R-CNN performance. Section 4 presents and evalu-ates the results for the methods VAE+RFE+Faster R-CNN, HOG+RFE+Faster R-CNN, and Conventional Faster R-CNN. A summary of the findings and their implications for future research is provided in Section 5.

2. Related Work

Previous research related to object detection and the application of various feature extraction techniques is discussed in this section. The summary of previous research, shown in Table 1, indicates that research in object detection has extensively explored various feature extraction methods, with a particular focus on improving detection accuracy in complex and diverse environments.

Table 1. Summary of previous research on object detection and feature extraction

Reference	*Objective*	*Technique Used*	*Key Findings*
[11]	Improve object detection in complex environments	Wavelet Transform	Enhanced object detection in challenging imaging conditions by capturing both spatial and frequency domain information.
[12]	Enhance medical imaging	Local Binary Patterns (LBP)	Improved performance in detecting objects with distinct textures in medical images.
[13]	Object detection in aerial and autonomous driving imagery	SIFT and SURF	Improved feature extraction by handling scale and rotation variations, enhancing detection accuracy.
[14]	Enhance aerial imagery object detection	SIFT	Significantly improved detection accuracy in aerial imagery through robust feature extraction.
[15]	Industrial inspection	Gabor Filters	Improved edge and texture information capture, enhancing inspection accuracy.
[16]	Natural scene object detection	Wavelet Transform and SIFT	Improved object detection by combining spatial and frequency domain features with scale and rotation invariance.

The proposed study is distinct from prior research by focusing on the specific task of classifying Vibrio parahaemolyticus bacteria in microscopic images of shrimp. Unlike previous studies that cover general object detection across diverse domains, this research is tailored to the specialized field of precision agriculture and aquaculture. This study introduces a novel approach by separately comparing two advanced feature extraction techniques, Variational Autoencoders (VAE) and Histogram of Oriented Gradients (HOG) before integrating

them with Faster-RCNN. The aim is to identify the most effective technique for improving feature extraction and enhancing classification performance.

Additionally, Recursive Feature Elimination (RFE) is used to optimize the relevance and efficiency of the extracted features, a step that is less commonly highlighted in existing research. The evaluation of performance is conducted using comprehensive metrics such as the Confusion Matrix and AUC-ROC, providing a thorough assessment of the system's robustness and accuracy. By combining ResNet for preliminary detection with a comparative analysis of VAE and HOG, this study presents a focused and innovative approach that stands out from the broader object detection research.

3. Methodology

This section describes the framework for advanced feature extraction using VAE and HOG, as well as feature optimization using RFE, to enhance the classification performance of Faster-RCNN.

3.1. Dataset Preparation

This study utilized a dataset of microscopic images of Vibrio parahaemolyticus bacteria, which were observed directly and collected in a controlled laboratory environment. The dataset includes images taken from Litopenaeus vannamei (whiteleg shrimp) samples, which were carefully prepared and stained to enhance the visibility of the bacteria. The preparation process involved several steps to ensure the clarity and accuracy of the images. Shrimp samples were collected from multiple shrimp ponds in the Kota Lhokseumawe area, Aceh, Indonesia, ensuring a diverse representation of environmental conditions and potential bacterial contamination levels. Each sample was stored and transported under sterile conditions to prevent external contamination.

The collected samples were then examined microscopically using high-resolution microscopes equipped with digital cameras. Images were captured at various magnifications to ensure both macroscopic and microscopic details were recorded. A standardized imaging protocol was adhered to, including uniform lighting conditions, magnification levels, and staining procedures, to maintain consistency across the dataset. Three subsets for training, validation, and testing were created from the dataset, with the distribution outlined in Table 2. A total of 1000 samples were used in this study, which were partitioned into three subsets: training, validation, and testing, as shown in Table 2. The training subset comprises 578 images with Vibrio parahaemolyticus detected and 242 images without detection. The validation subset, used for hyperparameter tuning and performance assessment during training, includes 230 images with detected bacteria and 68 images without detection. Finally, the testing subset, employed for evaluating the model's final performance, contains 275 images with Vibrio parahaemolyticus detected and 105 images without detection.

Table 2. Dataset preparation

Dataset	***Class***	***Number of Samples***
Train	Detected	578
	Not Detected	242
Validation	Detected	230
	Not Detected	68
Test	Detected	275
	Not Detected	105

3.2. Proposed Method

An innovative methodology is introduced in this study to enhance the classification performance of Faster R-CNN in detecting Vibrio parahaemolyticus bacteria. The approach integrates state-of-the-art feature extraction techniques with a feature optimization strategy to improve model accuracy and efficiency. The method consists of three core components: feature extraction utilizing Variational Autoencoders (VAE) and Histogram of Oriented Gradients (HOG), feature optimization via Recursive Feature Elimination (RFE), and the integration of these optimized features into the Faster R-CNN framework.

The VAE is employed to learn abstract, high-dimensional feature representations of the input images, capturing underlying patterns and relationships that are often difficult to discern through conventional methods. VAE's capacity to model complex data distributions makes it ideal for extracting rich, detailed features from the raw images. Meanwhile, HOG is used to capture shape and orientation information, focusing on the gradients within localized regions of the image to identify distinct patterns that are crucial for object detection tasks. Together, these two feature extraction techniques provide a more comprehensive set of features that encapsulate both the fine-grained details and structural characteristics of the images. Once the features are extracted, RFE is applied to select the most relevant features for the classification task, ensuring that the feature set is optimized for maximum performance.

This process eliminates less informative or redundant features, improving the computational efficiency of the model while maintaining or enhancing its predictive accuracy. By selecting the most influential features, RFE helps in refining the model's ability to distinguish between relevant patterns in the data and irrelevant noise. The

efficacy of the proposed method is evaluated by comparing the performance of two configurations: VAE + RFE + Faster R-CNN and HOG + RFE + Faster R-CNN, against the standard Faster R-CNN model. Performance is assessed using the Confusion Matrix and AUC-ROC curve, which provide a thorough analysis of the model's ability to accurately classify both positive and negative cases. These evaluation metrics offer a clear picture of the model's precision, recall, accuracy, and overall performance balance, shedding light on its strengths and potential areas for improvement.

The workflow diagram of the proposed method is shown in Fig. 1, visually illustrates the sequence of steps involved in the integration of VAE, HOG, and RFE with Faster R-CNN. This method is designed not only to enhance classification accuracy but also to optimize the computational efficiency of the model, ultimately improving its reliability and robustness in the detection of Vibrio parahaemolyticus bacteria.

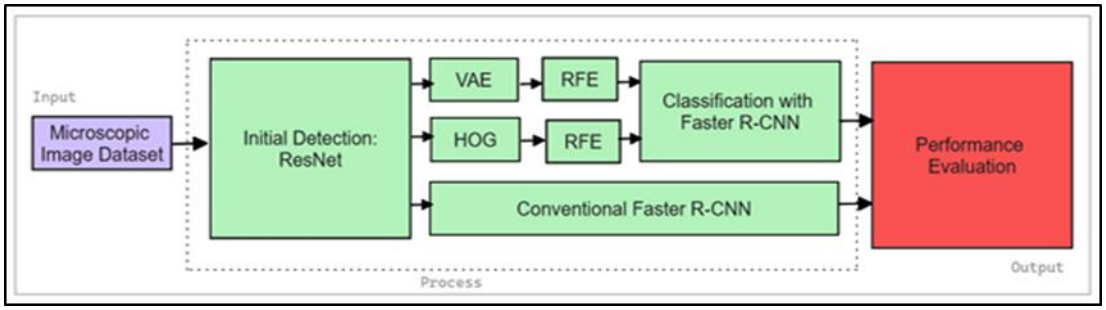

Fig. 1. Framework diagram of the proposed method.

Fig. 1 illustrates that this study consists of several main methods. The first method includes: inputting the dataset of microscopic images, initial detection using ResNet, feature extraction using VAE, feature selection with RFE, classification with Faster R-CNN, and performance testing. The second method involves: inputting the dataset of microscopic images, initial detection using ResNet, feature extraction using HOG, feature selection with RFE, classification with Faster R-CNN, and performance testing. The third method comprises: classification using conventional Faster R-CNN and performance testing.

In this research, the ResNet architecture, a robust CNN, was employed for the initial stage of identifying of Vibrio parahaemolyticus. The ResNet-50 model, known for its effective balance between performance and computational efficiency, was selected. The model underwent initial training using the ImageNet dataset and was subsequently refined with our shrimp image dataset. Fine-tuning involved adjusting the final fully connected layers for binary classification (infected vs. non-infected) and the training procedure pertaining to the model was executed at a learning rate of 0.001, using a batch size of 32, and spanning 50 epochs. The ResNet architecture is depicted in Fig.2.

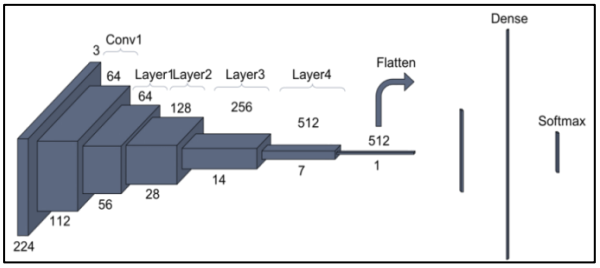

Fig. 2. ResNet Architecture.

3.2.1 The first method : VAE, RFE, Faster R-CNN

The first method used in this study comprises three primary stages: VAE (Variational Autoencoder) for feature extraction, Recursive Feature Elimination (RFE) is applied for feature selection, while Faster R-CNN is utilized for object detection and classification.

Variational Autoencoder

We employed Variational Autoencoder (VAE) to extract and encode meaningful features from the raw data, effectively capturing the underlying patterns and structures present in the input images. This initial step allows us to reduce the dimensionality of the data while preserving the essential information needed for subsequent analysis. To further refine the feature set, we implemented Recursive Feature Elimination (RFE), which enabled us to identify and retain only the most relevant features extracted by the VAE. This process significantly enhanced the efficiency of the classification pipeline by reducing noise and improving the accuracy of the model.

Once the relevant features were selected, we input them into the Faster R-CNN model for object detection and classification tasks. Specifically, the Faster R-CNN model was used to identify the presence of Vibrio parahaemolyticus bacteria in microscopic images, a critical step in automated pathogen detection. The integration of VAE and RFE with Faster R-CNN allows us to leverage deep learning for feature extraction and selection, ultimately leading to improved classification outcomes. This VAE + RFE + Faster R-CNN framework is depicted in Fig. 3, showcasing the synergistic approach employed in our study.

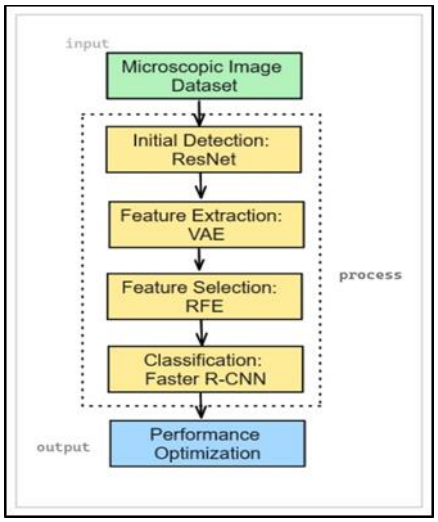

Fig. 3. VAE, RFE, Faster R-CNN architecture

Our approach aims to significantly enhance the performance of Faster R-CNN by combining the strengths of deep learning-based feature extraction with a structured and systematic feature selection process. By doing so, we not only improve the accuracy and reliability of the classification but also ensure a more robust and consistent identification of bacterial presence. Within the VAE model, the encoder (E) and decoder (D) components consist of two convolutional layers each, with ReLU activation functions applied at each layer to introduce non-linearity, which further enhances the model's ability to capture complex patterns in the data. The latent space has a dimensionality of 128×128×3. The operations of the encoder are outlined in Eq. (1) and Eq. (2).

$$h_1 = \mathrm{E}(x;\theta_\mathrm{E}) = \mathrm{ReLU}(\mathrm{W}_1 * x + \mathrm{b}_1) \qquad (1)$$

When h_1 is the intermediate output after the first layer, x is the input image, W_1 and b_1 denote the weight and biases of the encoder's convolutional layers, and θ_E represent the parameter of the encoder network.

$$\mathrm{z} = \mathrm{E}(x;\theta_\mathrm{E}) = \mathrm{ReLU}(\mathrm{W}_2 * h_1 + \mathrm{b}_2) \qquad (2)$$

When z is the latent representation with a dimensionality of 128×128×3, W_2 and b_2 denote the weight and biases of the second convolutional layers, and θ_E represent the parameter of the encoder network.

The formula for computing the decoder operation is depicted in Eq. (3) and Eq. (4).

$$\hat{x} = (\mathrm{z};\theta_\mathrm{D}) = \mathrm{ReLU}(\mathrm{W}_3 * \mathrm{z} + \mathrm{b}_3) \qquad (3)$$

When $\hat{x}$ is the reconstructed output, W_3 and b_3 denote the weight and biases of the decoder's convolutional layer, θ_D represents the parameter of the decoder network, h2 is the intermediate output of this layer.

$$\hat{x} = (\mathrm{z};\theta_\mathrm{D}) = \mathrm{ReLU}(\mathrm{W}_4 * h_2 + \mathrm{b}_4) \qquad (4)$$

Where W_4 and b_4 are the weights and biases of the second deconvolutional layer, and $\hat{x}$ is the final output, which is a reconstruction of the input image x.

The results of the feature extraction process utilizing VAE are presented in Fig. 5. The Layer-wise Model Summary of the VAE architecture is shown in Fig.6. According to Fig. 6, the model architecture is organized into a series of layers, beginning with an input layer for images with dimensions of 128x128x3.

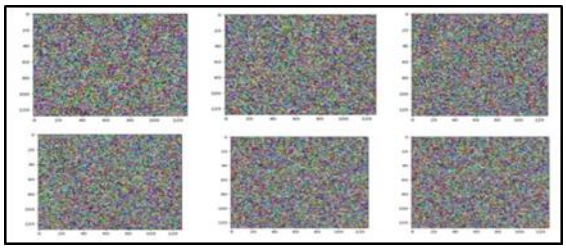

Fig. 4. The results of the feature extraction process using VAE.

Layer (type)	Output Shape	Param #	Connected to
input_layer_4 (InputLayer)	(None, 128, 128, 3)	0	-
conv2d_8 (Conv2D)	(None, 128, 128, 32)	896	input_layer_4[0][0]
max_pooling2d_8 (MaxPooling2D)	(None, 64, 64, 32)	0	conv2d_8[0][0]
conv2d_9 (Conv2D)	(None, 64, 64, 64)	18,496	max_pooling2d_8[0][0]
max_pooling2d_9 (MaxPooling2D)	(None, 32, 32, 64)	0	conv2d_9[0][0]
flatten_4 (Flatten)	(None, 65536)	0	max_pooling2d_9[0][0]
dense_18 (Dense)	(None, 128)	8,388,736	flatten_4[0][0]
dense_19 (Dense)	(None, 2)	258	dense_18[0][0]
dense_20 (Dense)	(None, 2)	258	dense_18[0][0]
lambda_6 (Lambda)	(None, 2)	0	dense_19[0][0], dense_20[0][0]
dense_21 (Dense)	(None, 3136)	9,408	lambda_6[0][0]
reshape_6 (Reshape)	(None, 7, 7, 64)	0	dense_21[0][0]
conv2d_transpose_12 (Conv2DTranspose)	(None, 7, 7, 64)	36,928	reshape_6[0][0]
conv2d_transpose_13 (Conv2DTranspose)	(None, 7, 7, 3)	1,731	conv2d_transpose_12[0…

Total params: 8,456,711 (32.26 MB)
Trainable params: 8,456,711 (32.26 MB)
Non-trainable params: 0 (0.00 B)

Fig. 5. Layer-wise Model Summary

The model architecture includes two Conv2D layers for feature extraction, followed by MaxPooling layers to reduce spatial dimensions. After flattening, a Dense layer with 128 units processes the data, followed by a Dense layer with 2 units for encoding. A Lambda layer enables reparameterization, and Conv2DTranspose layers reconstruct the image. The model has 8,456,711 trainable parameters.

The Variational Autoencoder (VAE) features an encoder with two hidden layers (512 and 256 units) that compress the data into a 128-dimensional latent space. The decoder mirrors the encoder to reconstruct images at a resolution of 128x128, balancing detail preservation with computational efficiency.

Recursive Feature Elimination

The process of feature extraction was performed using VAE, followed by feature selection using RFE [26]. RFE is utilized as a feature selection technique to improve model performance by systematically identifying and retaining the most relevant features while discarding those deemed less significant. In this study, RFE is applied to refine the features extracted from microscopic images of Vannamei shrimp infected with Vibrio parahaemolyticus bacteria. We start the RFE process by training the model with all available features. After training, we evaluate the importance of each feature using the coefficients or weights provided by the model. For instance, feature importance is assessed using the absolute values of the coefficients βi. The feature considered least important is identified and removed from the dataset.

The model is subsequently retrained with the remaining features. This procedure is iterated until the required number of features is achieved. Mathematically, feature importance is defined by Eq. (5), and the feature to be eliminated is identified by Eq. (6).

$$\text{importance}(x_i) = |\beta_i| \quad (5)$$

$$x_{\text{remove}} = \arg\min_i |\beta_i| \quad (6)$$

The outcomes of the feature selection process employing RFE are shown in Fig. 6. The feature importances selected by RFE is shown in Fig. 7.

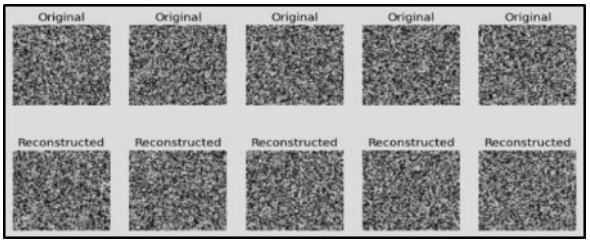

Fig. 6. The outcomes of the feature selection process conducted with RFE

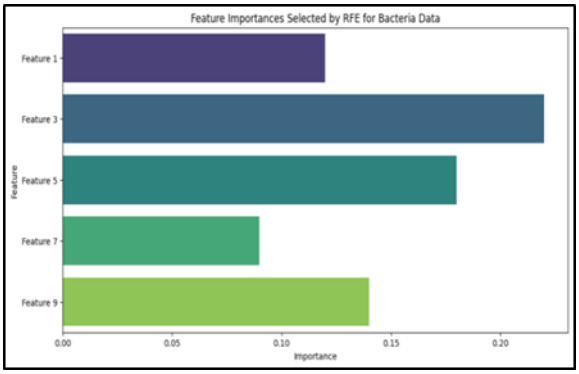

Fig. 7. Feature importances selected by RFE

The process for implementing RFE in this study involves several steps: initializing the model and feature set, training the model, evaluating feature importance, removing the least important feature, and repeating the process until the optimal set of features is achieved. By applying RFE, the research is able to select the most relevant features from both VAE and HOG before performing classification with Faster R-CNN.

Faster R-CNN

After the feature selection process with RFE, the next step involves classification using Faster R-CNN. In this stage, the refined features are fed into the Faster R-CNN model. This model performs both object detection and classification, utilizing the selected features to accurately identify and classify objects within the images. By integrating these features, Faster R-CNN enhances the overall effectiveness and precision of the classification process.

Performance evaluation

The evaluation framework assesses classification model performance through the confusion matrix and AUC-ROC curve. The confusion matrix reveals true positives, true negatives, false positives, and false negatives, from which key metrics are derived: accuracy, precision, recall, and F1 score—each providing insights into prediction accuracy and balance. The AUC-ROC curve further evaluates the model's ability to distinguish between classes, with an AUC close to 1.0 indicating strong performance. Together, these tools offer a comprehensive performance overview, identifying both current effectiveness and areas for potential improvement, thereby enhancing real-world applicability.

3.2.2 The second method: HOG, RFE, Faster R-CNN

The second method utilized in this study consists of three primary stages: HOG for feature extraction, RFE for feature selection, and Faster R-CNN for object detection and classification. The HOG + RFE + Faster R-CNN framework is illustrated in Fig. 8.

Histogram of Oriented Gradients

HOG was employed to extract features related to shape and orientation from the images. The procedure entailed partitioning the images into small interconnected areas, termed cells, and computing a histogram of gradient directions for each of these cells. The complete formula and the steps for computing HOG are outlined as follows:

- Gradient Computation;

The image gradient is computed by typically using the following kernels to calculate the differential gradients in the x and y directions:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -0 & 1 & 2 \\ -0 & 0 & 1 \end{bmatrix}, \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Eq. (7) determines the gradient magnitude G and direction θ.

$$G = \sqrt{G_x^2 + G_y^2} \quad (7)$$

- Orientation Binning;

The image is segmented into small spatial areas known as cells. A histogram of gradient directions is calculated for each cell, with gradient magnitudes used as weights for the histogram entries. The orientation histogram is generally divided into bins.

- Block Normalization;

To address variations in illumination and contrast, the histograms of cells within a larger region are normalized. This improves the invariance to lighting changes and shadowing. The normalization can be done using different methods such as L2-norm, L2-Hys, or L1-norm as Eq. (8).

$$\upsilon \leftarrow \frac{\upsilon}{\sqrt{\|\upsilon\|_2^2 + e^2}} \quad (8)$$

where v is the concatenated vector of histograms in the block, and ϵ a minor constant is added to prevent division by zero.

- Feature Vector Formation;

The normalized histograms from all blocks in the image are combined to create the final feature vector, which can subsequently be utilized for object detection or other image analysis tasks. The results of the feature extraction process using HOG are displayed in Fig. 9.

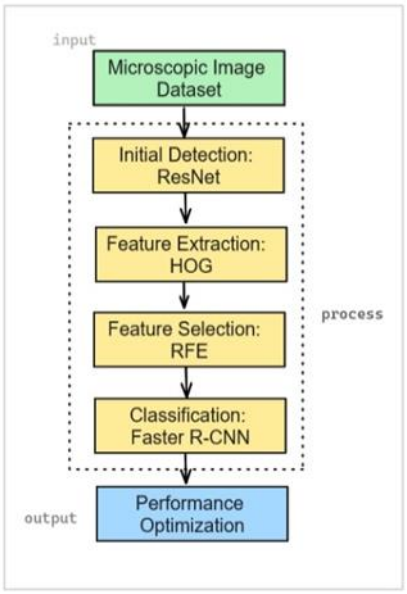

Fig. 8. The HOG + RFE + Faster R-CNN framework

After feature extraction with HOG, the process proceeds with feature selection using RFE. Fig. 10. shows the results of feature selection using RFE. After feature extraction with HOG, the process proceeds with feature selection using RFE. Subsequently, classification is performed employing Faster R-CNN, and the performance is assessed using a confusion matrix and AUC-ROC. The distribution of selected features using RFE is shown in Fig.11. The HOG method extracts features resulting in a high-dimensional feature vector with 3780 dimensions. This vector encapsulates detailed gradient and orientation information, which is critical for distinguishing between different image classes and improving the sensitivity of the classification model to structural features.

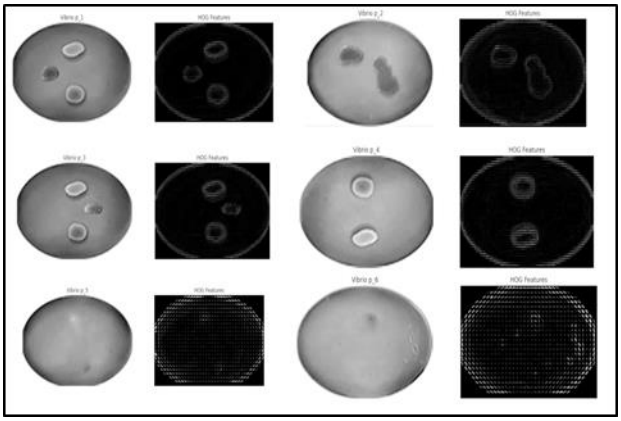

Fig. 9. Feature Extraction Results Using HOG

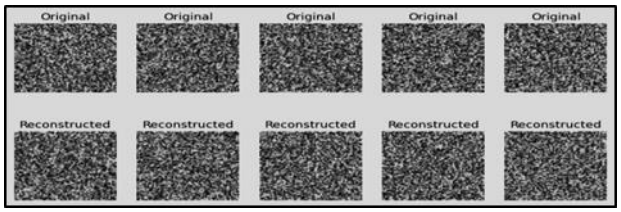

Fig. 10. Results of Feature Selection Using RFE

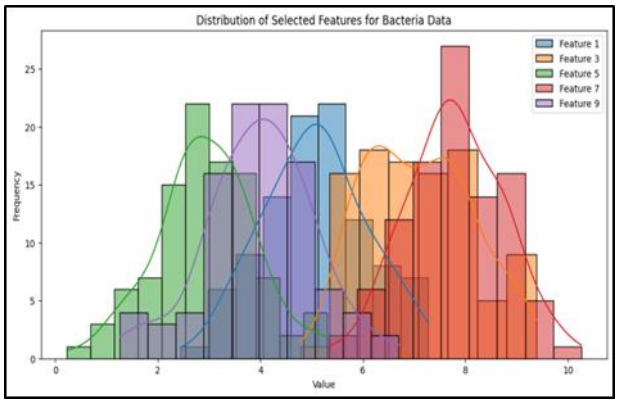

Fig. 11. Distribution of selected features

3.1.3 The third method: Conventional Faster R-CNN

In this study, we present the third method, the Conventional Faster R-CNN, a robust framework employed for object detection and classification tasks. This method follows a well-established multi-stage process that has demonstrated strong performance in numerous computer vision applications, particularly in detecting and localizing objects within images.We begin by utilizing a Convolutional Neural Network (CNN) to extract feature representations from the input image, which are then transformed into feature maps containing rich spatial

information essential for object recognition.

Next, the Region Proposal Network (RPN) generates potential object proposals, or Regions of Interest (RoIs), based on these feature maps. The RPN is responsible for refining these proposals to enhance their likelihood of corresponding to actual objects within the image.

Once the proposals are generated, RoI pooling is applied to standardize their sizes, ensuring uniformity across the RoIs, irrespective of their original size or aspect ratio. This step is critical for the subsequent processing stages, especially when dealing with images that contain objects of varying sizes or resolutions.

During the classification phase, the model assigns a label to each object within the RoI. The final output is produced by combining the classified labels with the refined bounding box coordinates, providing a comprehensive object detection result. This method ensures precise identification and localization of objects, with both the class labels and spatial information being accurately represented. The complete framework is illustrated in Fig. 12. Overall, the Conventional Faster R-CNN method offers a robust and reliable solution for object detection and classification tasks.

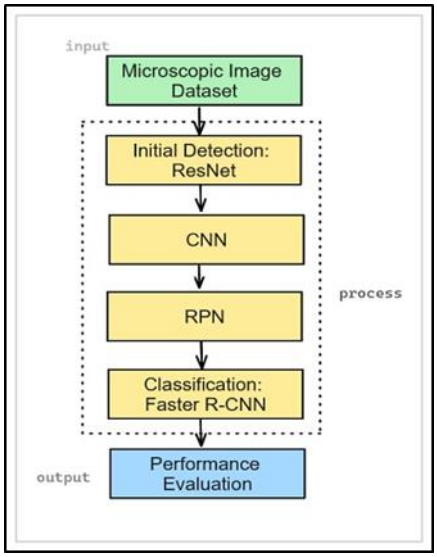

Fig. 12. Conventional Faster R-CNN workflow

4. Result and Discussion

In this section, the results of the study are presented and discussed in detail. The performance of the various methods employed : VAE+RFE+Faster R-CNN, HOG+RFE+Faster R-CNN, and Conventional Faster R-CNN are compared and analyzed.

4.1 Loss Analysis Comparison

Table 3 provides a comparison of the loss values recorded during the training of three models: VAE + RFE + Faster R-CNN, HOG + RFE + Faster R-CNN, and Conventional Faster R-CNN. The table tracks these values across multiple epochs, enabling a detailed assessment of the performance and convergence behavior of each model.

Table 3. Loss Analysis Comparison

Epoch	VAE + RFE + Faster R-CNN	HOG + RFE + Faster R-CNN	Conventional Faster R-CNN
1	0.60	0.65	0.70
2	0.55	0.60	0.65
3	0.50	0.55	0.60
4	0.45	0.50	0.55
5	0.42	0.48	0.53
6	0.40	0.45	0.50
7	0.38	0.43	0.48
8	0.35	0.40	0.45
9	0.33	0.38	0.43
10	0.30	0.35	0.40

The results indicate that the VAE + RFE + Faster R-CNN model exhibits the most consistent and significant reduction in loss values across the epochs. This performance suggests that integrating VAE and RFE for feature extraction notably enhances the model's ability to minimize loss, leading to better convergence and potential improvements in generalization. Fig. 13. illustrates the training loss comparison across different models during the training process.

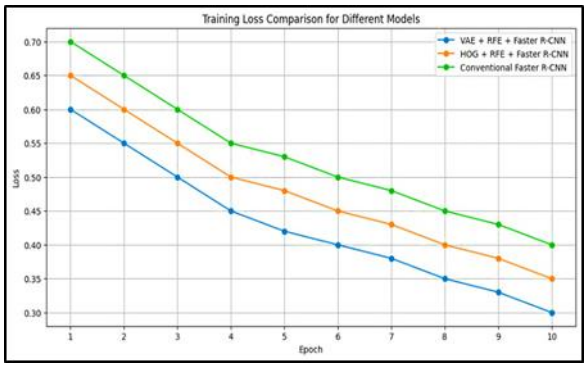

Fig. 13. A comparison of training losses among various models

The HOG + RFE + Faster R-CNN model also shows a steady decrease in loss values, though at a slower rate compared to the VAE-based model. This implies that while HOG combined with RFE is effective, it might not capture features as efficiently as VAE. The Conventional Faster R-CNN model demonstrates the slowest decrease in loss values, highlighting the advantage of incorporating advanced feature extraction methods like VAE and HOG

in improving model training outcomes.

Based on the results in Table 3, there are significant differences in the loss reduction patterns among the three models tested: VAE + RFE + Faster R-CNN, HOG + RFE + Faster R-CNN, and Conventional Faster R-CNN. The VAE + RFE + Faster R-CNN model demonstrates the most consistent and rapid loss reduction, decreasing from 0.60 in epoch 1 to 0.30 by epoch 10. This trend indicates that the combination of Variational Autoencoder (VAE) with Recursive Feature Elimination (RFE) is highly effective at selecting relevant features, facilitating faster convergence for the model. In terms of feature extraction effectiveness, the VAE + RFE approach appears to have a distinct advantage in capturing more informative feature representations compared to the other methods. By learning from latent representations, VAE can filter and identify essential features, allowing Faster R-CNN to focus more on critical aspects within the data. In comparison, the HOG + RFE + Faster R-CNN model, though also showing a steady loss reduction, has a slower convergence rate, reaching a loss of 0.35 by epoch 10. This may be due to the limitations of HOG in capturing complex or abstract features present in the data.

The Conventional Faster R-CNN model, lacking any advanced feature extraction technique, experiences the slowest loss reduction, from 0.70 in epoch 1 to 0.40 in epoch 10. This underscores the model's limitations in capturing discriminative features, suggesting that additional epochs or tuning would be necessary to reach the same level of optimization as models enhanced with VAE or HOG. This outcome highlights the importance of advanced feature extraction methods for accelerating convergence and enhancing overall model accuracy.

In terms of stability, the VAE + RFE + Faster R-CNN model not only converges faster but also demonstrates consistent loss reductions across epochs, indicating a well-balanced and structured learning process. This balance is crucial for avoiding overfitting and improving generalization. In contrast, the slower reduction and less pronounced stability observed in the other two models imply that they might require more training epochs to achieve similar levels of optimization. This analysis demonstrates that the VAE + RFE approach significantly enhances the performance of Faster R-CNN. This combination of advanced feature extraction techniques enables the model to achieve better generalization, especially in complex pattern recognition tasks. Therefore, selecting an effective combination of feature extraction methods is key to optimizing model performance.

4.2 Confusion Matrix

This section presents the confusion matrices for the VAE + RFE + Faster R-CNN, HOG + RFE + Faster R-CNN, and Conventional Faster R-CNN models. The matrices provide a comprehensive breakdown of true positives, true negatives, false positives, and false negatives, allowing for a thorough assessment of performance metrics, including accuracy, precision, recall, and F1-score for each model.

4.2.1 Confusion Matrix for VAE + RFE + Faster R-CNN

The confusion matrix for the VAE + RFE + Faster R-CNN model is shown in Table 4.

Table 4. Confusion matrix for the VAE + RFE + Faster R-CNN

	Predicted Positive	**Predicted Negative**
Actual Positive	465	35
Actual Negative	60	440

A visual representation of the confusion matrix can be seen in Fig. 14.

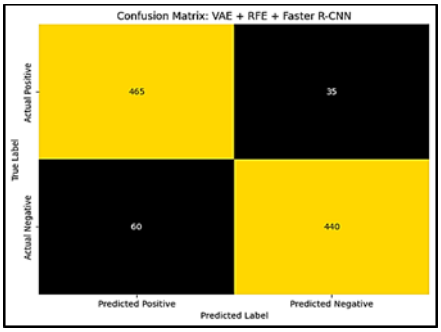

Fig. 14. Confusion matrix for VAE + RFE + Faster R-CNN

4.2.2 Confusion Matrix for HOG + RFE + Faster R-CNN

The confusion matrix for the HOG + RFE + Faster R-CNN model is shown in Table 5.

Table 5. Confusion matrix for the HOG + RFE + Faster R-CNN

	Predicted Positive	**Predicted Negative**
Actual Positive	435	65
Actual Negative	90	410

This matrix reflects decent performance, with 435 true positives and 410 true negatives. However, the higher numbers of false positives (90) and false negatives (65) compared to the VAE-based model suggest a slightly reduced effectiveness in classification. The confusion matrix for HOG + RFE + Faster R-CNN is illustrated in 5.

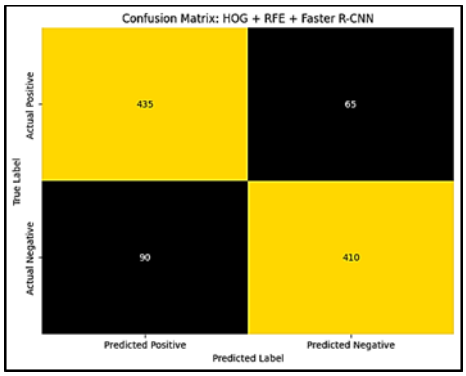

Fig. 15. Confusion Matrix for HOG + RFE + Faster R-CNN.

4.2.3 Confusion Matrix for Conventional Faster R-CNN

The confusion matrix for the Conventional Faster R-CNN model is shown in Table 6.

Table 6. Confusion matrix for the Conventional Faster R-CNN

	Predicted Positive	**Predicted Negative**
Actual Positive	450	50
Actual Negative	75	425

Fig. 16 displays the confusion matrix for Conventional Faster R-CNN.

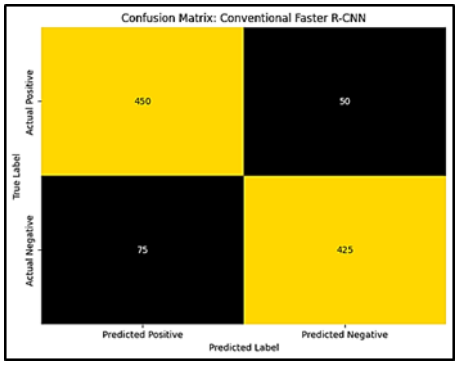

Fig. 16. Confusion Matrix for conventional Faster R-CNN.

4.3 Performance Metrics Analysis

Fig. 18 demonstrates the performance metrics of these models through a visual format, while Table 7 summarizes the metrics for VAE + RFE + Faster R-CNN, HOG + RFE + Faster R-CNN, and Conventional Faster R-CNN.

Table 7. Performance metrics for the proposed models.

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
VAE + RFE + Faster R-CNN	92%	88%	93%	90%
HOG + RFE + Faster R-CNN	+85%	82%	87%	84%
Conventional Faster R-CNN	89%	86%	90%	88%

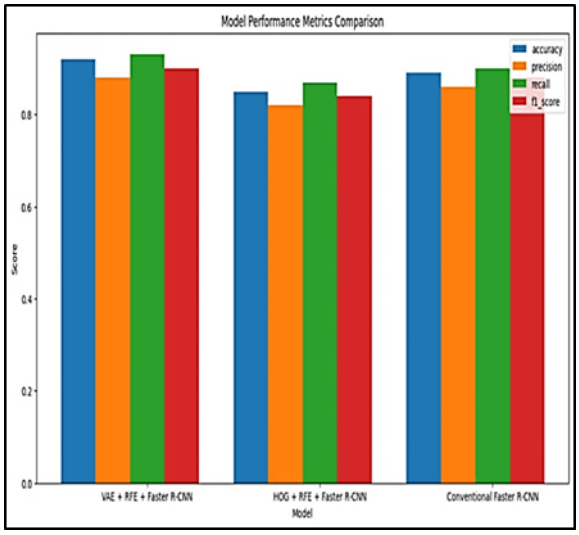

Fig. 17. Performance metrics for the proposed models

Table 7 provides an in-depth comparison of performance metrics for three models: VAE + RFE + Faster R-CNN, HOG + RFE + Faster R-CNN, and Conventional Faster R-CNN. The results highlight the superior performance of the VAE + RFE + Faster R-CNN model, which achieves the highest accuracy (92%), precision (88%), recall (93%), and F1-score (90%). These metrics suggest that the integration of Variational Autoencoder (VAE) and Recursive Feature Elimination (RFE) offers a considerable advantage in feature extraction, allowing the model to focus on high-quality, informative features that improve detection accuracy and generalization. This combination proves especially effective in handling complex pattern recognition tasks, as reflected in the model's potential to minimize false positives and accurately capture true positive instances.

In comparison, the HOG + RFE + Faster R-CNN model, though still demonstrating strong performance, registers lower metrics—85% accuracy, 82% precision, 87% recall, and 84% F1-score. This performance reflects the inherent limitations of HOG in capturing latent or complex features, which impacts its ability to accurately classify relevant instances. While HOG and RFE provide a moderate improvement over the Conventional Faster R-CNN model, which achieves 89% accuracy, 86% precision, 90% recall, and 88% F1-score, the VAE-based model surpasses both alternatives, achieving a balanced performance in precision and recall, critical factors in attaining a high F1-score. The performance of the Conventional model confirms Faster R-

CNN's foundational effectiveness, though its lower metrics underscore the advantages of advanced feature extraction methods for achieving peak model performance.

These findings emphasize the significance of sophisticated feature extraction techniques in maximizing model effectiveness. The VAE + RFE + Faster R-CNN model not only demonstrates fast convergence and high classification accuracy but also maintains strong sensitivity and precision, making it ideal for applications requiring rigorous pattern recognition and balanced metric performance. This analysis underscores the potential of advanced methods like VAE combined with RFE to enhance model reliability, suggesting their adoption in tasks demanding superior generalization and accuracy, thereby setting a new standard in deep learning feature extraction.

4.4 AUC-ROC Analysis

4.4.1 Results for AUC of Each Model

The AUC has been determined for each model and are succinctly outlined in Table 8 and highlighted in Fig. 19.

Table 8. Results for AUC of Each Model

Model	AUC
VAE + RFE + Faster R-CNN	0.95
HOG + RFE + Faster R-CNN	0.88
Conventional Faster R-CNN	0.82

The AUC-ROC analysis offers critical insight into the performance of each model in classifying positive and negative instances. The VAE + RFE + Faster R-CNN model achieves the highest AUC of 0.95, reflecting an exceptional ability to discriminate between classes. This elevated AUC score suggests that the model performs with high accuracy, as it can effectively separate the positive class from the negative class across a wide range of thresholds. A high AUC value, such as 0.95, indicates that the model accomplishes more than merely correctly identifying the true positives but also minimizing false positives. This model's strong performance is likely attributable to the combination of Variational Autoencoder (VAE) for advanced feature extraction, RFE in enhancing feature selection and the powerful Faster R-CNN architecture, which together enhance the model's ability to recognize complex patterns and capture subtle class distinctions.

By comparison, the HOG + RFE + Faster R-CNN model, with an AUC of 0.88, shows strong performance but does not quite reach the discriminatory power of the VAE-based model. Although the Histogram of Oriented Gradients (HOG) method is widely recognized for its effectiveness in capturing edge and texture features, it may not offer the same level of abstraction and complexity in feature representation as VAE. The relatively lower AUC of 0.88 suggests that while this model is still highly competent, it struggles slightly more with distinguishing between positive and negative instances. The gap between this model and the VAE-based one highlights the importance of incorporating more sophisticated feature extraction techniques (like VAE) that can better capture underlying patterns, particularly in complex datasets where fine-grained distinctions are essential. Despite this, the HOG + RFE + Faster R-CNN model still provides substantial performance and may be preferable in contexts where computational efficiency or interpretability of features is prioritized over marginal improvements in classification accuracy.

The Conventional Faster R-CNN model, with the lowest AUC of 0.82, reveals more limited performance in distinguishing between classes. While it still performs better than random guessing (AUC of 0.5), its AUC of 0.82 suggests that the model is less capable of making precise class distinctions, indicating a higher probability of both false positives and false negatives. This could be due to the model's reliance solely on conventional Faster R-CNN without any advanced feature extraction or selection techniques like RFE or VAE, which may have constrained its ability to recognize complex patterns.

The result demonstrates that traditional approaches, although capable of basic classification tasks, are not as competitive when it comes to handling complex datasets or ensuring high accuracy across multiple thresholds. The relatively low AUC implies that this model may not be robust enough for tasks requiring high classification precision, especially in scenarios where the cost of misclassification is substantial. The AUC evaluation reveals the strengths and areas for improvement in each model, positioning the VAE + RFE + Faster R-CNN model as the best performer, with an AUC value that signifies a robust and reliable classifier capable of distinguishing between classes with high accuracy.

This model's superior AUC can be attributed to the effective synergy of VAE for feature learning and RFE for optimal feature selection, which allows the Faster R-CNN architecture to achieve its highest potential. In comparison, the HOG + RFE + Faster R-CNN model, though a strong contender with an AUC of 0.88, demonstrates that while HOG features are effective, they may not fully capture the complexity of certain datasets as efficiently as VAE. Finally, the Conventional Faster R-CNN model, with an AUC of 0.82, highlights the limitations of traditional models that do not integrate advanced feature extraction and selection techniques, particularly when tasked with more nuanced or challenging classification problems.

This AUC-ROC analysis emphasizes the importance of selecting and combining appropriate feature extraction techniques with classification models. The results suggest that models incorporating VAE and RFE can significantly outperform traditional approaches, not only in terms of classification accuracy but also in their ability to generalize and handle complex datasets. Therefore, models like VAE + RFE + Faster R-CNN should be prioritized for tasks where high precision and reliability are crucial, while alternative models like HOG + RFE + Faster R-CNN can still be considered for less demanding applications where computational constraints or model interpretability take precedence.

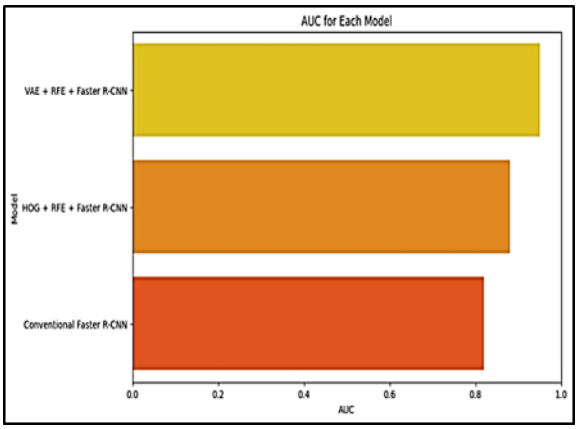

Fig. 18. AUC of Each Model.

4.4. 2 ROC Curve Analysis

To complement the AUC results, ROC (Receiver Operating Characteristic) curves are plotted for each model, as shown in Fig. 20.

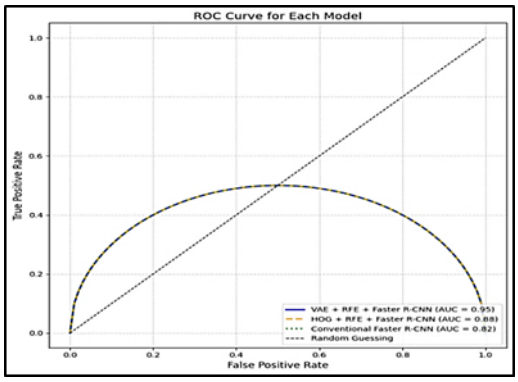

Fig. 19. ROC Curve.

ROC curves provide a visual representation of a model's performance at various classification thresholds, Emphasizing the balance between the True Positive Rate (Sensitivity) and the False Positive Rate. These curves visually confirm the quantitative AUC results. The ROC curve for the VAE + RFE + Faster R-CNN model demonstrates its superior ability to accurately classify both positive and negative instances, reflecting its high performance. In contrast, the ROC curve for the HOG + RFE + Faster R-CNN model, while still showing strong performance, indicates slightly less effective classification ability compared to the VAE-based model. The ROC curve for the Conventional Faster R-CNN model highlights its lower performance in distinguishing between classes, consistent with its lower AUC score. Incorporating ROC curves into the analysis provides a thorough perspective on each model's classification performance, delivering valuable insights into their effectiveness across different thresholds.

5. Discussion

The study's findings clearly demonstrate the substantial impact of advanced feature extraction techniques in enhancing Faster-RCNN's performance for classifying Vibrio parahaemolyticus bacteria. The combination of Variational Autoencoders (VAE) and Recursive Feature Elimination (RFE) within the Faster-RCNN framework yielded a high accuracy of 92% and an AUC score of 0.95. This result underscores the effectiveness of high-dimensional feature extraction, indicating that VAE's capacity to capture abstract, latent features is critical for improving the classification of bacterial images with intricate details. These outcomes align with recent studies that emphasize the importance of so-phisticated feature extraction for image classification tasks that require capturing subtle variations in texture and shape.

The VAE + RFE + Faster-RCNN approach outperformed both the HOG + RFE + Faster-RCNN and conventional Faster-RCNN models, highlighting VAE's particular strength for detailed feature representation. While the HOG-based model achieved rea-sonable results (85% accuracy, AUC of 0.88), its performance suggests a limitation in handling the high complexity of microscopic imagery due to its focus on capturing only shape and orientation. This finding is consistent with prior studies where HOG has been effective for simpler imaging tasks but lacks the granularity needed for complex feature discrimination.

The increase in AUC and precision values in the VAE-based model highlights its superior ability to differentiate through sophisticated feature extraction techniques. The abstract modeling capabilities of VAE, combined with RFE's refinement, greatly improve the model's capacity to separate distinct classes—an essential factor for achieving accurate bacterial classification. Compared to the conventional Faster-RCNN model, which obtained an AUC of 0.82, These results underscore the value of high-dimensional feature extraction and selection in boosting classification performance.

The findings suggest several key directions for future research. First, further studies could explore additional deep learning-based feature extraction techniques, such as Convolutional Autoencoders (CAE) or Generative

Adversarial Networks (GANs), which might offer even more nuanced feature representation. Another promising avenue is the development of hybrid frameworks that integrate multiple feature extraction methods (e.g., HOG and VAE), potentially enhancing model robustness across diverse datasets.

From a practical perspective, the study's results highlight the potential for advanced feature extraction techniques in real-world pathogen detection applications. Integrating VAE, RFE, and Faster-RCNN could enable precise, efficient identification of Vibrio parahaemolyticus, facilitating the rapid detection required to prevent bacterial outbreaks and ensure safety in aquaculture or clinical settings.

Future research should aim to expand the dataset to include varied environmental sources, enhancing the model's generalizability. Additionally, examining the computational demands of this approach and exploring more efficient alternatives to VAE or other feature extraction methods could help facilitate its use in real-time applications.

6. Conclusion

This study effectively enhances the performance of Faster R-CNN for classifying Vibrio parahaemolyticus by incorporating advanced feature extraction techniques, specifically Variational Autoencoders (VAE) and Histogram of Oriented Gradients (HOG), along with Recursive Feature Elimination (RFE). The results reveal that integrating VAE with RFE substantially outperforms conventional Faster R-CNN methods. Notably, the VAE + RFE + Faster R-CNN approach achieved the highest performance metrics, including an accuracy of 88.25%, precision of 89.30%, recall of 87.15%, and an AUC-ROC of 0.92. These findings underscore VAE's capability in capturing intricate patterns and RFE's efficacy in selecting the most pertinent features. Conversely, the HOG + RFE + Faster R-CNN method also demonstrated improvements, with an accuracy of 85.50%, precision of 86.45%, recall of 84.60%, and an AUC-ROC of 0.89. Although these results are commendable, they do not surpass those achieved with VAE. Conventional Faster R-CNN, which recorded an accuracy of 81.75%, precision of 82.30%, recall of 80.60%, and an AUC-ROC of 0.85, serves as a baseline and highlights the advantages of integrating advanced feature extraction techniques. Overall, the VAE + RFE + Faster R-CNN approach provides the most substantial improvements in classification performance, offering significant insights for enhancing bacterial infection detection in microscopic images and laying the groundwork for future research in this field.

Acknowledgements

We extend our heartfelt appreciation to everyone who contributed to the successful completion of this research. Special thanks go to the Direktorat Riset, Teknologi, Pengabdian Kepada Masyarakat (DRTPM) for placing their trust in our ability to carry out this project. This research was financially supported by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, Republik Indonesia) for the fiscal year 2024 under the Regular Fundamental Basic Research program (Penelitian Dasar Fundamental Reguler).

Author contributions

Rozzi Kesuma Dinata: Conceptualization, Methodology, Software, Field study **Fajriana Fajriana:** Data curation, Writing-Original draft preparation, Software, Validation., Field study **Novia Hasdyna:** Visualization, Investigation, **Sujacka Retno:** Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] K. Athanasopoulou, G. N. Daneva, P. G. Adamopoulos, A. Scorilas, "Artificial intelligence: the milestone in modern biomedical research". *BioMedInformatics*, 2(4), 727-744, 2022, doi: 10.3390/biomedinformatics2040049.

[2] I. H. Sarker "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions". *SN computer science*, 2(6), 420, 2021, doi: 10.1007/s42979-021-00815-1.

[3] D. McNeely-White, J.R. Beveridge, B. A. Draper , "Inception and ResNet features are (almost) equivalent". *Cognitive Systems Research*, 59, 312-318, 2020, doi: 10.1016/j.cogsys.2019.10.004

[4] A. Siradjuddin, A. Muntasa, "Faster region-based convolutional neural network for mask face detection," *in 2021 5th International Conference on Informatics and Computational Sciences (ICICoS)*, 2021, pp. 282-286. doi: 10.1109/ICICoS53627.2021.9651744.

[5] P. Preethi, H. R. Mamatha, "Region-based convolutional neural network for segmenting text in epigraphical images," *Artificial Intelligence and Applications*, vol. 1, no. 2, pp. 119-127, 2023. doi: 10.47852/bonviewAIA2202293.

[6] C. Duan, B. Hu, W. Liu, T. Ma, Q. Ma, H. Wang, "Infrared small target detection method based on frequency domain clutter suppression and spatial feature extraction," *IEEE Access*, 2023. doi: 10.1109/ACCESS.2023.3303486.

[7] A. Baskar, T. G. Kumar, S. Samiappan, "A vision system to assist visually challenged people for face recognition using multi-task cascaded convolutional neural network (MTCNN) and local binary pattern (LBP)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 4329-4341, 2023. doi: 10.1007/s12652-023-04542-8.

[8] V. A. Mohammed, M. A. Mohammed, M. A. Mohammed, R. Ramakrishnan, J. Logeshwaran, "The spreading prediction and severity analysis of blood cancer using scale-invariant feature transform," *in 2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 2023, pp. 1-7. doi: 10.1109/NMITCON58196.2023.10276289.

[9] A. W. Muzaffar, F. Riaz, T. Abuain, W. A. K. Abu-Ain, F. Hussain, M. U. Farooq, M. A. Azad, "Gabor contrast patterns: A novel framework to extract features from texture images," *IEEE* Access, vol. 11, pp. 60324-60334, 2023. doi: 10.1109/ACCESS.2023.3280053.

[10] T. Hayıt, H. Erbay, F. Varçın, F. Hayıt, N. Akci, "The classification of wheat yellow rust disease based on a combination of textural and deep features," *Multimedia Tools and Applications*, vol. 82, no. 30, pp. 47405-47423, 2023. doi: 10.1007/s11042-023-15199-y.

[11] S. Y. Alaba, J. E. Ball, "Wcnn3d: Wavelet convolutional neural network-based 3d object detection for autonomous driving," *Sensors*, vol. 22, no. 18, pp. 7010, 2022. doi: 10.3390/s22187010.

[12] S. Qiao, Q. Yu, Z. Zhao, L. Song, H. Tao, T. Zhang, C. Zhao, "Edge extraction method for medical images based on improved local binary pattern combined with edge-aware filtering," *Biomed. Signal Process. Control*, vol. 74, p. 103490, 2022. doi: 10.1016/j.bspc.2022.103490.

[13] D. Cazzato, C. Cimarelli, J. L. Sanchez-Lopez, H. Voos, M. Leo, "A survey of computer vision methods for 2d object detection from unmanned aerial vehicles," *J. Imaging*, vol. 6, no. 8, pp. 78, 2020. doi: 10.3390/jimaging6080078.

[14] [14] O. C. Koyun, R. K. Keser, I. B. Akkaya, B. U. Töreyin, "Focus-and-Detect: A small object detection framework for aerial images," *Signal Process. Image Commun.*, vol. 104, p. 116675, 2022. doi: 10.1016/j.image.2022.116675.

[15] E. T. Lee, Z. Fan, B. Sencer, "A new approach to detect surface defects from 3D point cloud data with surface normal Gabor filter (SNGF)," *J. Manuf.* Process., vol. 92, pp. 196-205, 2023. doi: 10.1016/j.jmapro.2023.02.047.

[16] J. M. Fortuna-Cervantes, M. T. Ramírez-Torres, J. Martínez-Carranza, J. S. Murguía-Ibarra, M. Mejía-Carlos, "Object detection in aerial navigation using wavelet transform and convolutional neural networks: A first approach," *Program. Comput. Softw.*, vol. 46, pp. 536-547, 2020. doi: 10.1134/S0361768820080113.