ISSN 2355-0082

Evaluating the Impact of Particle Swarm Optimization Based Feature Selection on Support Vector Machine Performance in Coral Reef Health Classification

Jessica Carmelita Bastiaans[1], James Hartojo[2], Ricardus Anggi Pramunendar[3], Pulung Nurtantio Andono[4]
[1,2,3,4] Informatics Engineering, Dian Nuswantoro University, Semarang, Indonesia
[1]jessicacarmelita2004@gmail.com, [2]jameshartojo@gmail.com, [3]ricardus.anggi@dsn.dinus.ac.id, [4]pulung@dsn.dinus.ac.id

Accepted 1 September 2024
Approved 19 November 2024

***Abstract*— This research explores improving coral reef image classification accuracy by combining Histogram of Oriented Gradients (HOG) feature extraction, image classification with Support Vector Machine (SVM), and feature selection with Particle Swarm Optimization (PSO). Given the ecological importance of coral reefs and the threats they face, accurate classification of coral reef health is essential for conservation efforts. This study used healthy, whitish, and dead coral reef datasets divided into training, validation, and test data. The proposed approach successfully improved the classification accuracy significantly, reaching 85.44% with the SVM model optimized by PSO, compared to 79.11% in the original SVM model. PSO not only improves accuracy but also reduces running time, demonstrating its effectiveness and computational efficiency. The results of this study highlight the potential of PSO in optimizing machine learning models, especially in complex image classification tasks. While the results obtained are promising, the study acknowledges several limitations, including the need for further validation with larger and more diverse datasets to ensure model robustness and generalizability. This research contributes to the field of marine ecology by providing a more accurate and efficient coral reef classification method, which can be applied to other image classifications.**

***Index Terms*— Coral Reef Classification; Histogram of Oriented Gradients (HOG); Machine Learning; Particle Swarm Optimization (PSO); Support Vector Machine (SVM)**

I. INTRODUCTION

Indonesia as an archipelago that has the second longest coastline reaching more than 95,000 km2 where more than 60% of its territory is the ocean, and its geographical location between the Indian and Pacific Oceans produces a very rich and diverse marine biodiversity. More than 39,500 km2 or as much as 16% of the world's coral reefs are found in Indonesia [1]. Coral reefs are home to marine biodiversity up to more than 6000 species of fish that are very large and unique [2]. Coral reefs play an important role in maintaining biodiversity, preventing coastal erosion, and promoting business trade.

However, coral reefs are experiencing population decline due to overexploitation, ecosystem damage, and climate change causing abrasion [3], [4]. Nearly 25% of Indonesia's 270 million people live and do activities in coastal areas within 30 km of coral reefs, based on this percentage, it affects up to 95% of coral reefs in Indonesia are currently in threatened status where more than 35% are in high or very high threat levels [1]. Meanwhile, about 43% of observations of 324 coral reefs in Indonesia are damaged and even endangered while only 6.48% are still in very good condition and only 5.48% of coral reefs have high status from the results of a survey at 985 stations conducted by the Oceanographic Research Center of LIPI in 2008 [1].

Based on research conducted by several scientists, almost 50% of coral reefs will be destroyed by 2030 [5] Therefore, the destruction of coral reefs affects the health of marine life and also decreases the livelihood of people who depend on it. One of the diseases experienced by coral reefs is bleaching. Coral reefs bleach due to ocean acidification and global warming, which is a serious threat to the earth's ecosystem [6].

With the advancement of technology in image processing, researchers began to conduct research on coral reefs using a variety of machine learning and deep learning algorithms. Such as, the use of CNN (Convolutional Neural Network) Algorithm [7] in the classification of small datasets of coral texture images, data augmentation techniques and transfer learning approaches, using variations of ResNet and ImageNet to improve coral reef image processing results [8], [9]. YOLOv4 (You Only Look Once) was also used in determining coral reef disease for the computer vision algorithm training process. An incremental methodology was developed in three training stages to evaluate accuracy, by modifying different parameters [2], [10], [11].

But on the other hand, by proposing learning models for species classification in underwater images, as well as many factors such as lighting, color, shape, structure, etc. it was found that the performance of CNN models in image classification decreased significantly [12], [13]. Image classification analysis comparison has also been conducted out with a small sample ImageNet Dataset, obtained SVM accuracy results of 0.85 and CNN 0.82 with runtime for SVM 1.05 minutes and CNN 2.05 minutes [14]. Deep Learning models often require more computational resources for training, and these models are less interpretable due to their black box nature [15], [16]. So, based on the research that has been done, SVM is superior in handling small datasets with the support of a fast runtime process compared to CNN [14], [15], [17], [18].

Previous researchers used three algorithms to model coral reef bleaching areas with 3 models, the results showed that SVM was the most effective classification model with 88.85% accuracy, followed by decision tree and Naïve Bayes with 80.25% and 71.34% accuracy [19]. Other researchers also compared SVM with other machine learning, namely KNN, where SVM is the right choice for the classification of larger datasets [20]. Coral reef researchers in Kapota Atoll (Wakatobi National Park, Indonesia), Harapan Island and Kelapa Island of Kepulauan Seribu Indonesia, Palmyra Atoll also used the SVM (Support Vector Machine) algorithm method to classify and recognize characteristic images of coral reefs [21], [22], [23], [24]. As for the development of other research that aims to classify bleached and healthy coral reefs, this research also uses the SVM (Support Vector Machine) Classifier method and receives input from grouping features based on the similarity of coral reef characteristics [6].

The author also concluded that the Histogram Of Oriented Gradients (HOG) feature provides better results than SIFT and SC in its use with the SVM algorithm model to perform coral reef classification [25]. The SVM algorithm has the advantage of using raw image data as feature vectors, especially for natural random textures where characterization is difficult to obtain. This is relevant to the classification of coral reef images that are rich in patterns, colors, shapes, and textures [26], [27]. Based on previous research, it is an appropriate target to use various HOG feature extraction methods and PSO optimization to determine the best features based on the particle set so as to achieve the best accuracy of the SVM model [28]. The application of this strategy is considered suitable as it aims to obtain a higher level of precision and accuracy. The contribution of this research is validated by comparing the classification results of healthy, whitish and dead coral reef images for: (1). test the use of PSO parameters on the effect of SVM classification performance; (2). investigate the relationship of HOG extraction with PSO optimization; (3). Identify solutions and feature combinations needed to achieve accurate performance of healthy, whitish and dead coral reef image classification.

This document is organized in the following order: Section 1, regarding the background of this proposed research. Section 2, relates the research method and discusses the experimental strategy used by the authors. Section 3 details the evaluation and experimental findings. Section 4 contains the conclusion of the research.

II. METHOD

Contents Based on previous research, it shows that the selection of machine learning algorithms and image preprocessing affects the accuracy and classification of coral reef images. The characteristics of the detected coral reefs will determine the label of each processing dataset collected. The dataset also has an influence for machine learning to learn the uniqueness of the coral reef itself. This research uses a unique approach to recognize the characteristics of healthy, whitish and dead coral reefs.

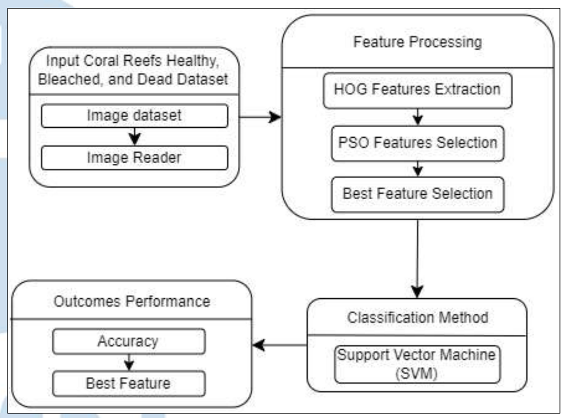

Fig. 1. Research proposed diagram

As shown in Fig. 1, the classification stage begins with the process of collecting coral reef datasets labeled healthy, bleach and dead which will be used as input data for the training process. This classification technique is based on several categories extracted through Histogram Oriented Gradients (HOG), including histograms of color, texture, shape, and standard deviation. HOG itself has superior characteristics in the form of edge structure representation, shape, and adjustable level of variation [29]. The extraction results performed by HOG will then be optimized through the Particle Swarm Optimization (PSO) algorithm feature selection. PSO will take the best value from the optimization results for the classification of coral reef images by PSO-based Support Vector Machine. PSO is a method of finding the best combination of features that will be classified by the SVM model. In the special case of coral reef texture classification, Support Vector Machine allows excellent class separation even when the feature vector size is large and the number of training samples is limited [26], [27].

A. Data Collection

In this study we conducted two initial stages, namely data collection and recognition of coral reef characteristics. Bleached Health Dead Corals Dataset used in this study in the form of images of healthy, whitish, and dead coral reefs (https://www.kaggle.com/datasets/sonainjamil/bhd-corals).

Fig. 2. Bleached coral reef images

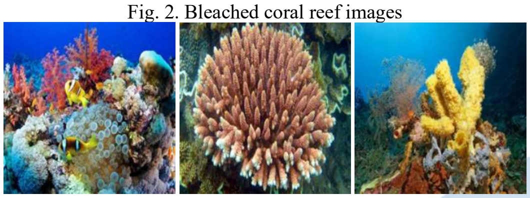

Fig. 3. Healthy coral reef images

Fig. 4. Dead coral reef images

The dataset is divided into three parts, namely training, testing, and validation. Training data includes 576 images of whitish coral reefs, 569 healthy coral reefs, and 120 dead coral reefs, while validation data includes 72 images of whitish coral reefs, 71 healthy coral reefs, and 15 dead coral reefs, then for testing data includes 72 images of whitish coral reefs, 72 healthy coral reefs, and 15 dead coral reefs. The set of image data contained in the training file will be processed so as to obtain information on the unique characteristics of healthy, whitish and dead coral reefs. HOG helps the process of extracting coral reef image recognition and removing noise around the detected image.

B. Histogram of Oriented Gradients (HOG) for Features Extraction

Histogram of Oriented Gradients (HOG) is an extraction feature that helps in coral reef image recognition [30]. In the image processing process, grouping pixel gradient values based on the directional orientation of each part of the local structure and shape characteristics of the image [31]. HOG will convert the input image into a feature vector representation that reflects the gradient orientation [31] in various parts of the coral reef image. To perform the image feature extraction stage, it is necessary to read the dataset from the folder and set parameters for dynamic input.

The feature extraction function in this project will receive an image parameter containing a coral reef image that has been converted to grayscale format using the OpenCV library. This image input is dynamic, where the image dimensions, namely height and width, are extracted automatically. Based on the varying image dimensions, the cell size is calculated by dividing the height and width of the respective image by 9. This parameter will determine how large the size of each cell is where the gradient histogram will be calculated. Then, a fixed value of (2,2) is determined, which means that each block consists of 2x2 cells. The blocks are used as histogram normalization in some cells to increase the robustness against lighting changes. In addition, the gradient orientation in each cell is divided into 9 bins, which is useful for determining the number of gradient orientation intervals calculated in each cell.

After setting the parameters for image processing with HOG, for each pixel in the coral reef image, the gradient is calculated in the x and y directions. This gradient reflects the change in pixel intensity, which can be interpreted as a shape feature.

$$G_x = \frac{\partial I}{\partial x}, \quad G_y = \frac{\partial I}{\partial y} \tag{1}$$

Where I is the intensity of the image, and Gx and Gy represent the change in intensity in the horizontal and vertical directions, respectively (1).

After calculating the image gradient with the formula above, the next step is to determine the magnitude and orientation of the gradient at each pixel.

$$G = \sqrt{G_x^2 + G_y^2} \tag{2}$$

$$\theta = atan2(G_{y,}\ G_x) \tag{3}$$

The magnitude G represents the strength of the intensity change, while the orientation θ determines the direction of the change (2), (3). With the pre-set parameters, HOG will calculate each cell (small region in the image) which will then represent the gradient orientation distribution within a cell.

HOG leverages the use of block normalization to make features more resilient to lighting changes with *L2-Hys* normalization. This normalization is applied to ensure that the magnitude of the gradient vector does not affect feature detection, and helps to improve the quality of the extracted features by reducing sensitivity to lighting and contrast differences, so that the features remain well distributed throughout the image.

$$v' = \frac{v}{\sqrt{||v||_2^2 + \epsilon^2}} \tag{4}$$

Where v is the unnormalized feature vector of the block, and ϵ is a small value to prevent division by zero. v' is the normalized feature vector (4). After all the blocks in the image are normalized, the feature vectors of all the blocks are combined into one large feature vector that represents the image as a whole.

C. Particle Swarm Optimization (PSO) for Features Selection

The feature selection method in the SVM classification model uses Particle Swarm Optimization (PSO) which is an optimization algorithm adopted from the social behavior of animal group movements such as birds flying in flocks and groups of coral reefs [28], [32]. Each particle in PSO will move in the solution space to determine the combination as a potential solution from many available features, so that PSO can choose the best features to be used in SVM model training in order to maximize model accuracy [32], [33].

The process begins with initializing the parameters of the number of epochs and population size. Where the particles will be generated randomly in the solution space, then each particle determines one solution in the form of a binary vector that will be carried out in the feature selection process. Based on the given solution, the selected feature columns are taken from the training data and validation data. PSO will generate a new solution randomly when no features are selected. The selected features are then standardized to ensure that each feature has the same scale [34].

Next, the SVM model is trained based on a subset of features from the retrieved binary vectors to measure the accuracy of the model on the test data. The validation accuracy value obtained is then returned as the "fitness" value of the solution, where the fitness function is a calculation that determines the suitability or objective value of a solution. This research utilizes the classification method to evaluate fitness performance. Then, for the best solution results that have been achieved will be stored by personal best (*pbest*) with the aim of updating the particle position. At each iteration, the particle will compare the fitness value of the current position with the personal best (*pbest*) value. Then an update is made to the global best (*gbest*) value which refers to the best solution among all personal bests (*pbest*) of each particle in a particular iteration population. In each iteration, the global best (*gbest*) is calculated based on the fitness comparison of each personal best (*pbest*). The best solution (a subset of features) found during the optimization process is used to train the final model on training data and tested on test data.

Based on the process previously described, the formula for calculating the position and velocity of particles is generated as below.

$$x_i(t+1) = x_i(t) v_i(t+1) \quad (5)$$

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot rand_1 \cdot (pbest_i - x_i(t)) + c_2 \cdot rand_2 \cdot (gbest - x_i(t)) \quad (6)$$

In PSO, the representation of the solution using a position vector $x_i(t)$ indicates that each particle can have a potential solution in the search space, while the velocity vector $v_i(t)$ determines how fast and in which direction the solution will move in the next iteration. The combination of the two creates a search mechanism that helps the algorithm reach the optimal solution. Then, the personal best (pbset) affects the movement of individual particles, while the global best (gbest) is used to affect the movement of the entire population (5). The following is a description of the use of formulas that affect particle position and velocity updates (5), (6).

- Particle velocity *i* at iteration *t+1* is initialized with $v_i(t)$
- The position of particle *i* at iteration *t* is initialized with $x_i(t)$
- *w* is the inertia factor
- $c_1$dan c_2 are acceleration factors
- Random values between *0* and *1* are assigned to $rand_1$ and $rand_2$

$$fitness = accuracy(SVM_{model}, X_{selected}, Y_{val} \quad (7)$$

It then evaluates potential solutions on each particle representation. Each particle selects which features to use based on its position. $X_{selected}$ is the feature selected based on the particle position, this solution is then tested on the SVM model and evaluated based on accuracy on the validation set (7). By utilizing the feature solution information obtained by personal best (*pbest*) and global best (*gbest*), it can adaptively adjust the movement of particles to approach the most optimal solution.

D. Support Vector Machine (SVM) for Classification Model

Support vector machines (SVM) are binary classifiers that estimate the optimal separating hyperplane that maximizes the margin between two classes [26], [27], [31]. This coral reef image research classification uses the Support Vector Machine method because of the advantages in object- and pixel-based classification methods that have high accuracy [20], [35]. The preprocessing step is followed by training the Support Vector Machine (SVM) classifier [23]. The data splitting used in this study involves 3 categories using the ratio (80% Training: 10% Testing: 10% Validation).

Support Vector Machine receives input from HOG extractions in the form of color, structure, shape, texture, size, and gradient features from coral reef images. The author implemented the SVM model using the *C* parameter to control large margins and misclassification. A larger value of *C* influences the model to classify all samples correctly.

Many mapping functions are available, including linear, polynomial, and radial basis (RBF) kernel functions. Polynomial and RBF kernel functions are commonly used depending on the training dataset [19], it should be noted that the RBF kernel can be considered as a restricted version of the generalized Gaussian version, where the Gaussian matrix is restricted to a unit matrix multiplied by a scale factor [36]. In this study, the RBF kernel is applied to the Support Vector

Machine (8) [35] to handle the non-linearity problem (not linearly separable) by mapping the data to a higher dimension [37]. With gamma set to value scale so that it can automatically calculate the gamma value based on the number of features and data variations (9). there is a formula applied to coral reef classification where :

$$K(X_i, X_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (8)$$

$$\gamma = \frac{1}{2\sigma^2} \quad (9)$$

- x_i and x_j are two input data vectors
- $\|x_i - x_j\|^2$ is the squared Euclidean distance between the two input data vectors
- γ is a gamma parameter that determines how far the influence of a training sample goes

Then, the SVM model is trained using training data with appropriate labels, and tries to find the most optimal hyperplane that separates the two classes with the largest margin. By maximizing the margin, the SVM model tries to have good generalization ability on data that has never been seen before. Here is the SVM formula to minimize the loss function while still ensuring maximum margin (10),

$$\min_{w,b}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i\right) \quad (10)$$

- $\frac{1}{2}\|w\|^2$ is the regularization part that tries to minimize the norm of the weight vector w, thus ensuring the margin remains large.
- C is a regularization parameter that controls the change between large margins and misclassification.
- ξ_i i is a slack variable that allows the feature vector data to be inside the margin or on the wrong side of the hyperplane, usually occurring when the data cannot be perfectly separated.

After the SVM model is trained, an evaluation is conducted on the training data to see how well the model learns to predict labels and recognize coral reef image patterns. The SVM model will be integrated into the feature selection process with PSO to select the best combination of features that produce the highest accuracy. With the support of PSO will find the optimal feature subset, so that the performance of the SVM model is maximized. The following is the decision formula used as a label prediction from new data (11),

$$y(x) = sign(w \cdot x + b) \quad (11)$$

- *y(x)* is the class prediction for data *x*
- $sign\ (z)$ to return *+1* if $z > 0$ and *-1* if $z < 0$

E. Performance Evaluation

Based on the classification performed by the Support Vector Machine (SVM) algorithm, the accuracy results are measured with a division of 80% training data, 10% test data, and 10% validation data. The trained model is used to predict the test data. Then, comparing the adjustment between the prediction results with the actual labels, namely healthy, whitish and dead reefs (12).

$$Accuracy = \frac{ypred}{ytrue} \times 100 \quad (12)$$

The process of classifying healthy, whitish and dead coral reefs using the Support Vector Machine (SVM) model produces an accuracy value of 85.44%, PSO helps the model in processing and finding the best features from the many combinations of particles available. Where, there will be a combination of the best features selected from all columns. The author decides to set a *feature solution ≥ 0.5* which will then be selected randomly, because the solution has a big influence on the best feature combination to improve model accuracy and speed up the feature selection process and model training (13), (14).

$$Features = Number\ of\ Columns \quad (13)$$

$$solutions \geq 0.5 \quad (14)$$

III. Result and Discussion

A. Result of the Histogram Oriented of Gradients (HOG)

Histogram Oriented of Gradients (HOG) extraction helps the classification process of coral reef images from a total of 2304 extraction features performed as in table 1. There are 158 images processed in the use of the HOG feature extraction method.

Table 1 Feature extraction results with HOG

No	X0	X1	X2	X3	X4	X5	X6	...	X2303
1	0.137	0.004	0.014	0.008	0.029	0.014	0.059	...	0.265
2	0.078	0.082	0.053	0.147	0.322	0.132	0.077	...	0.079
3	0.089	0.080	0.051	0.182	0.263	0.085	0.052	...	0.000
4	0.222	0.098	0.206	0.269	0.275	0.275	0.231	...	0.122
5	0.230	0.027	0.015	0.007	0.011	0.004	0.019	...	0.254
6	0.175	0.060	0.083	0.121	0.247	0.162	0.110	...	0.243
7	0.167	0.062	0.099	0.165	0.247	0.247	0.247	...	0.186
...	...	...	...	...	...	...	...	...	...
158	0.263	0.309	0.309	0.309	0.277	0.088	0.083	...	0.182

Then these features will help the PSO method in finding and selecting the best combination of features to find the best fitness value, where each iteration will compare the fitness value of the current position with the personal best (pbest) value obtained.

B. Feature Selection Results with Particle Swarm Optimization (PSO)

Experimental iterations were conducted 5 times with a population size parameter range of 5 to 25 iterations. Based on the PSO process that has been carried out, the best combination of solutions is obtained by taking the solution value set by the author, namely the *feature solution* ≥ *0.5*. The selected solution represents the best feature columns, which are then used to form the optimal feature combination that will be used in the image classification stage..

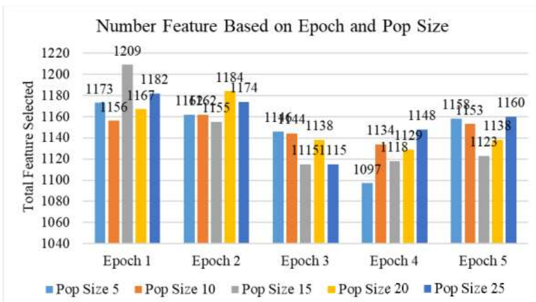

Fig. 5. Comparison total feature selected of each epoch and pop size

Fig. 5 visualizes the effect of epoch and population size on the number of features selected after selection. Looking at the graph above, larger population sizes, such as population size 25 often produce the largest number of features, while smaller populations, such as population sizes 10 and 15, tend to have fewer features. The number of selected features also tends to decrease as epochs increase, although there is movement between populations. Larger population sizes, especially in the final epochs, correlate with the best accuracy, indicating the important role of population size in the optimization process. The two-factor Analysis of Variance (ANOVA) results without replication show that the epoch factor has a significant effect on the number of features selected after selection, with an F-value of 7.8589554 and a P-value of 0.0010557, which is smaller than 0.05. This means that the number of selected features varies significantly between different epochs. In contrast, the population size factor did not show a significant effect on the number of selected features, with an F-value of 0.3041953 and a P-value of 0.8708985, which is much larger than 0.05. Thus, population size does not significantly affect the number of features selected in this selection process.

In Fig. 6, the scatter plot results show the relationship between the number of features selected and accuracy. The X-axis, the total number of features selected after the PSO selection process ranges from 1080 to 1220, while the Y-axis shows the accuracy of the model which reaches between 80% to 86%. It can be seen that the highest accuracy, about 85%, is achieved when the number of features is in the range of 1140 to 1160, although some other points also show high accuracy at different numbers of features. This shows that there is no linear correlation between the number of features and accuracy, so increasing or decreasing the number of features does not always have a consistent impact on accuracy. The use of PSO in feature selection proved to be effective in finding the optimal number of features that can achieve the best accuracy without having to use all the features, thus improving the efficiency of the model and reducing computational complexity.

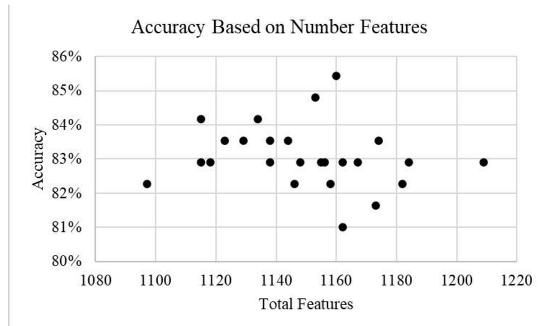

Fig. 6. Comparison accuracy based on total feature selected

The author also conducted an ANOVA analysis to show that the variation between columns, i.e. total features selected and validation accuracy, has a significant effect with an F-value of 50665.6 and a P-value of 2E-41, well below the 0.05 threshold. This indicates that the number of selected features strongly influences accuracy validation. In contrast, the variation between rows represents the iterations with an F-value of 0.99975 and a P-value of 0.50025, which means there is no significant difference between iterations. This result indicates that the main factor affecting accuracy is the number of features selected, while the difference between iterations has no significant impact. Based on the explanation that has been described, the best initial accuracy evaluation results are obtained at epoch 5 population size 25 with a total of 1160 features as in table 2.

TABLE 2 FEATURE SELECTION RESULTS WITH PSO

No	X0	X1	X5	X8	X9	X12	X13	...	X2303
1	0,137	0,004	0,014	0,273	0,169	0,039	0,062	...	0.265
2	0,078	0,082	0,132	0,244	0,039	0,201	0,322	...	0.079
3	0,089	0,080	0,085	0,129	0,085	0,228	0,263	...	0.000
4	0,222	0,098	0,275	0,117	0,247	0,275	0,275	...	0.122
5	0,230	0,027	0,004	0,285	0,110	0,020	0,019	...	0.254
6	0,175	0,060	0,162	0,112	0,093	0,142	0,247	...	0.243
7	0,167	0,062	0,247	0,075	0,176	0,133	0,210	...	0.186
...	...	...	...	...	...	...	...	...	...
158	0,263	0,309	0,088	0,117	0,177	0,309	0,204	...	0.182

From the initial evaluation results, the best accuracy in the classification of healthy, whitish, and dead coral reefs produced solutions with values {0.817, 0.917, 0.145, 0.077, 0.343, 0.564, 0.324, 0.230, 0.810, 0.670, ... , 0.585, 0.310, 0.813, 0.856, 0.675}. From a total of 2303 features, the selected features that have been selected are 1160 combinations of features, namely {0, 1, 5, 8, 9, 12, 13, 14, 16, 17, 20, 21, 22, 24, 25, 31, 33, 34, 35, 36, ... , 2278, 2279, 2283, 2284, 2287, 2288, 2289, 2290, 2291, 2295, 2298, 2299, 2301, 2302, 2303}, these results are randomly selected by PSO.

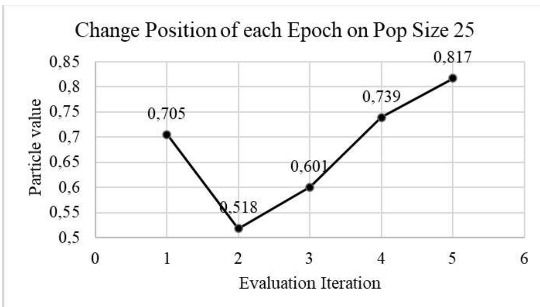

Fig. 7. Position movement of each epoch on pop size 25

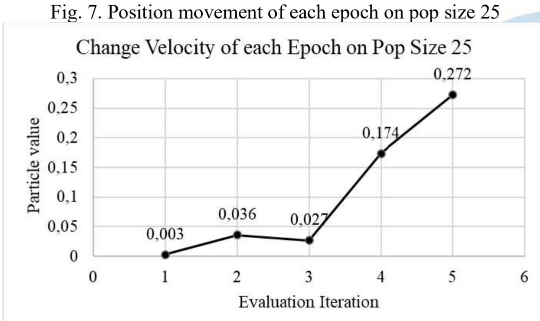

Fig. 8. Movement velocity of each epoch on pop size 25

The following Fig. 7 is a graph of changes in the position of particles in PSO at each epoch with a population size of 25 showing a significant shift at the beginning of the process. In the first epoch, it starts with a relatively high particle position of 0.705, but experiences a fairly significant decrease in epoch 2 with a position of 0.518. This decrease indicates that the particle is exploring the search space to find a more optimal solution. After that, there was a gradual increase in epochs 3 and 4 with positions of 0.601 and 0.739 respectively, until finally reaching the best position at epoch 5 with a value of 0.817. This increase in position indicates that the particles are getting closer to the optimal solution as time goes by.

Meanwhile, the particle velocity in Fig. 8 change graph shows that the particles start with a very low velocity at the first epoch, which is 0.003. This speed increased slightly at epoch 2 to 0.036, but again dropped at epoch 3 with a value of 0.027, indicating that the particles were exploiting deeper around the temporary solution. A significant change occurred at epoch 4, where the velocity jumped to 0.174 and then peaked at epoch 5 with a value of 0.272. Where the acceleration of particle movement that occurs is useful for optimizing its position. This increase shows that the particle moves faster and more intensively to find a better solution at the final stage of the optimization process.

The relationship between position change and velocity shows that higher velocities in the last epochs encourage particles to find more optimal positions. When the particle speed reaches its peak at epoch 5, the particle position also reaches the highest value, which coincides with the highest accuracy of 85.44%. This shows that more intensive particle movement speed helps to find better solutions in coral reef classification. Thus, the combination of increased particle speed and position significantly contributed to the optimal result at the last epoch.

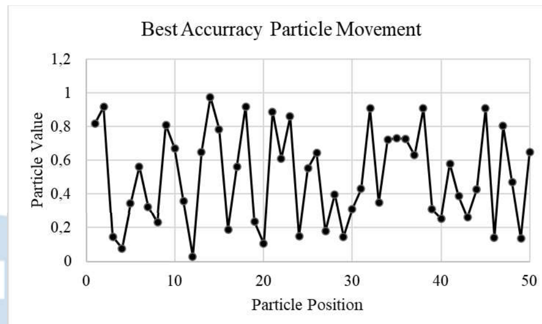

Fig. 9. Particle position movement based on best accuracy

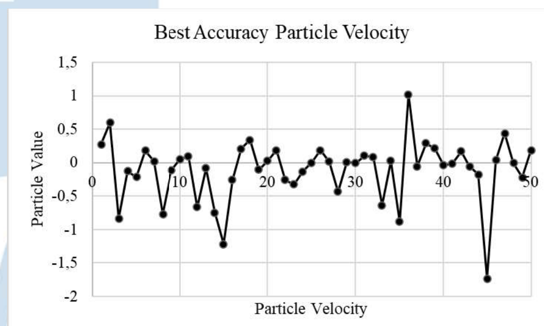

Fig. 10. Particle velocity movement based on best accuracy

Then, the author also looks at how the movement of the velocity and position displacement when it is in the best accuracy, namely epoch 5 population size 25 with 23 particles. In Fig. 9, the graph of particle position changes experiences a significant shift, but remains in a more stable range than the velocity graph, with position values ranging from 0.2 to 1.2. At some points, such as the 5th and 46th positions, the position peaks around the value of 1.2, but also drops to lower values, such as at the 10th and 35th positions. This shifting pattern indicates that particle 23 remains in the process of optimizing the solution space, with more controlled changes in position than in velocity.

And in Fig. 10 the velocity change graph shows a fairly dynamic velocity shift. Where, the particle velocity is around the zero value with some significant peaks, such as at the 30th velocity which reaches a value above 1.5, and a drastic decrease at the 44th velocity which reaches a value close to -2. This pattern explains that the particles experience rapid changes in acceleration, both in terms of increases and decreases. Such rapid changes in velocity usually reflect intensive exploration of the solution space, where particles move quickly to various points in an attempt to find an optimal solution.

Rapidly moving velocities indicate that the particle is conducting an intensive search, while more stable positions indicate that the particle is focusing on an area that the algorithm considers optimal. The peaks in the velocity graph go hand in hand with larger position changes, indicating that as the particle moves quickly, the position will continue to move to a new, more significant position. At epoch 5 of population size 25, the combination of dynamic speed and stable position helped the particle find the optimal solution in the coral reef classification process.

C. Accuracy Evaluation

PSO is a feature selection method that researchers use to find the best combination of features so as to find the most optimal solution. The use of PSO feature selection improves the performance of the SVM algorithm model by producing the best accuracy of 85.44% at epoch 5 population size 25 as in Fig. 11, while the lowest accuracy is 81.01% at epoch 2 population size 10 and for the best testing accuracy obtained is 80% as in Fig. 12. Researchers used Analysis of Variance (ANOVA) to show statistical analysis conducted on validation accuracy data and testing accuracy data from the classification of healthy, whitish, and dead coral reefs. Two main factors that affect the results of the analysis in this ANOVA are the number of epochs (experiments) and population size in the SVM model used.

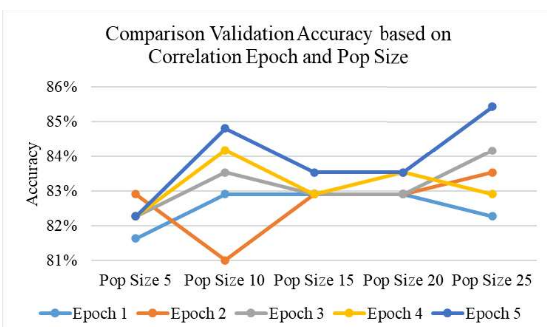

Fig. 11. Comparion validation accuracy based on correlation epoch and pop size

In the ANOVA results of the validation accuracy data, it was found that the F-value for the epoch factor was 2.41026, followed by a P-value of 0.09213. It is also explained that the F-crit value is 3.00692, where this value is relevant at a certain significance level of 0.05. From the results of the analysis it can be seen that, the F-value is smaller than the F crit and the P-value is greater than the general significance level (0.05). It can be concluded that the variation in the number of epochs has no significant effect on the validation accuracy. Then for the population size F-value of 2.08974, with a P-value of 0.12977 and an F crit value of 3.00692. Similar to the epoch factor, where the F-value is smaller than the F crit and the P-value is greater than the general significance level (0.05), indicating that variations in population size have no significant effect on validation accuracy. Based on the results of the ANOVA analysis of the validation accuracy data, it shows that changes in these two factors do not have a significant impact on the validation accuracy results of the coral reef classification model. In other words, there was no significant difference in validation accuracy based on variations in the number of epochs or population size used in the classification model.

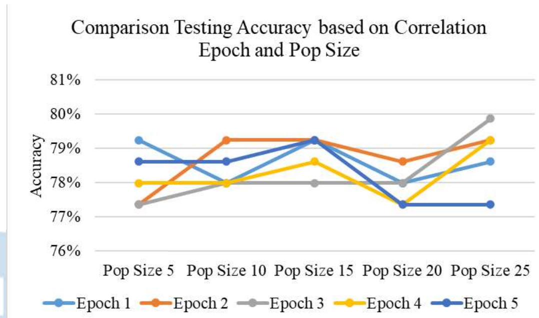

Fig. 12. Comparion testing accuracy based on correlation epoch and pop size

Furthermore, in the ANOVA results of testing accuracy data, the F value for the epoch factor is 0.54874, with a P-value of 0.70264 and a relevant F crit value of 3.00692. Then in the population size section, the F value is 1.84838, with a P-value of 0.16896, and an F crit value of 3.00692. Similarly, the results of the ANOVA analysis of the validation accuracy data show that changes in both factors do not have a large enough impact to be considered significant on the test accuracy results of the coral reef classification model used. In other words, there was no significant difference in testing accuracy based on variations in the number of epochs or population size used. This result is consistent with the results of the validation accuracy data, which also showed that both components did not significantly impact the performance of the model.

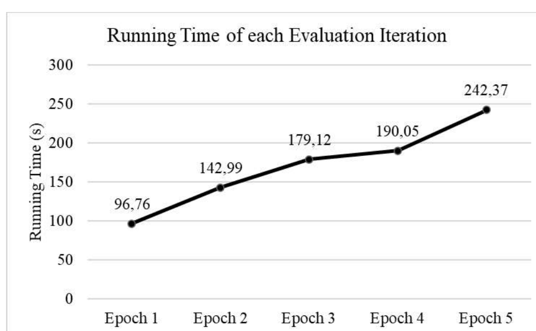

Fig. 13. Running time of each evaluation iteration

In Fig. 13 the graph above shows an increase in the running time of the SVM model with PSO optimization for coral reef classification as the epoch increases, from 96.76 seconds at Epoch 1 to 242.37 seconds at Epoch 5. This increase is due to the increasing complexity of the model and deeper exploration of the solution space by PSO to find the optimal solution. This shows that the optimization process gets more complex as time goes by.

D. Comparison with Ordinary Models

By using the same dataset and data division of 80:10:10, researchers conducted a comparison between the original SVM model and the SVM model optimized using PSO (Particle Swarm Optimization) to see how much influence PSO has on accuracy and running time speed in the classification of healthy, whitish, and dead coral reefs. In the original SVM model, the accuracy result is 79.11% with a running time speed of 916.37 seconds. However, after optimization with PSO, the accuracy of the model increased by 6.33%, reaching 85.44%. In fact, the lowest accuracy of the SVM model results that have been optimized with PSO is still higher than the accuracy of the original SVM model, which is 81.01%. The difference in running time also provides a significant difference of 674 seconds.

This experiment shows that the PSO optimization algorithm has a significant effect on the performance of the SVM model, both in terms of accuracy improvement and running time efficiency. This result confirms that PSO not only improves the model's ability to recognize and classify patterns in coral reef datasets more accurately, but also speeds up the computational process, which is very important in the context of real-time or large-scale applications..

IV. CONCLUSION

This study successfully demonstrated that the combination of feature extraction method with Histogram of Oriented Gradients (HOG), image classification with Support Vector Machine (SVM), and feature selection with Particle Swarm Optimization (PSO) significantly improved the accuracy in the classification of healthy, whitish and dead coral reef images. Experimental results show that the use of PSO successfully increases the accuracy of the SVM model to 85.44%, which is a substantial improvement compared to the original SVM model without optimization, which only achieves an accuracy of 79.11%. In addition, PSO also reduced the running time required for the classification process making it computationally efficient. PSO also showed excellent performance in performing feature selection, by effectively reducing the number of features used in training the model, which originally had a total of 2303 features, at best accuracy only used a total of 1160 features. This not only speeds up the training process but also increases the accuracy of the model in predicting coral reef classification. Although the results achieved are quite satisfactory, this study has some limitations, especially in terms of generalizing the model to new data. Further research using larger and varied datasets is recommended to ensure the robustness and generalization of the resulting model.

ACKNOWLEDGMENT

All praise to God Almighty for all His mercy and grace so that this research can be completed properly. We would also like to thank Dian Nuswantoro University for the support that has been given during this research process. Without continuous help and support, this research could not be completed perfectly. We hope that the results of this research can provide benefits and have a positive impact on the wider community, as well as being a meaningful contribution in the field of science and technology.

REFERENCES

[1] A. Triwibowo, "STRATEGI PENGELOLAAN EKOSISTEM TERUMBU KARANG DI WILAYAH PESISIR," *Jurnal Kelautan dan Perikanan Terapan (JKPT)*, vol. 1, p. 61, Jan. 2023, doi: 10.15578/jkpt.v1i0.12048.

[2] S. Villon *et al.*, "A Deep learning method for accurate and fast identification of coral reef fishes in underwater images," *Ecol Inform*, vol. 48, pp. 238–244, Nov. 2018, doi: 10.1016/j.ecoinf.2018.09.007.

[3] E. Bollati, C. D'Angelo, D. I. Kline, B. G. Mitchell, and J. Wiedenmann, "Development of a multi-excitation fluorescence (MEF) imaging method to improve the information content of benthic coral reef surveys," *Coral Reefs*, vol. 40, no. 6, pp. 1831–1847, Dec. 2021, doi: 10.1007/s00338-021-02101-1.

[4] S. A. Harahap, N. A. Shabrina, N. P. Purba, and M. L. Syamsuddin, "The patterns of changes in coral reef coverage in the Seribu Islands National Park, Jakarta, Indonesia," 1994. [Online]. Available: www.worldnewsnaturalsciences.com

[5] J. Borbon, J. Javier, J. Llamado, E. Dadios, and R. K. Billones, "Coral Health Identification using Image Classification and Convolutional Neural Networks," in *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/HNICEM54116.2021.9731905.

[6] S. Jamil, M. Rahman, and A. Haider, "Bag of features (Bof) based deep learning framework for bleached corals detection," *Big Data and Cognitive Computing*, vol. 5, no. 4, Dec. 2021, doi: 10.3390/bdcc5040053.

[7] C. Jackett *et al.*, "A benthic substrate classification method for seabed images using deep learning: Application to management of deep-sea coral reefs," *Journal of Applied Ecology*, vol. 60, no. 7, pp. 1254–1273, Jul. 2023, doi: 10.1111/1365-2664.14408.

[8] A. Gómez-Ríos, S. Tabik, J. Luengo, A. S. M. Shihavuddin, and F. Herrera, "Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks," *Knowl Based Syst*, vol. 184, Nov. 2019, doi: 10.1016/j.knosys.2019.104891.

[9] A. Gómez-Ríos, S. Tabik, J. Luengo, A. S. M. Shihavuddin, B. Krawczyk, and F. Herrera, "Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation," *Expert Syst Appl*, vol. 118, pp. 315–328, Mar. 2019, doi: 10.1016/j.eswa.2018.10.010.

[10] G. A. Bautista-Hernández, D. A. Jimenez-Nixon, and A. M. Reyes-Duke, "Coral Reef Disease and Bleaching Indentification through Computational Vision Algorithm," in *Proceedings of the 2022 39th IEEE Central America and*

Panama Student Convention Conference, CONESCAPAN 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/CONESCAPAN56456.2022.9959584.

[11] S. Mittal, S. Srivastava, and J. P. Jayanth, "A Survey of Deep Learning Techniques for Underwater Image Classification," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 10, pp. 6968–6982, Oct. 2023, doi: 10.1109/TNNLS.2022.3143887.

[12] Y. Pei, Y. Huang, Q. Zou, H. Zang, X. Zhang, and S. Wang, "Effects of Image Degradations to CNN-based Image Classification," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.05552

[13] S. Sood, H. Singh, M. Malarvel, and R. Ahuja, "Significance and Limitations of Deep Neural Networks for Image Classification and Object Detection," in *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1453–1460. doi: 10.1109/ICOSEC51865.2021.9591759.

[14] M. Madanan, S. S. Gunasekaran, and M. A. Mahmoud, "A Comparative Analysis of Machine Learning and Deep Learning Algorithms for Image Classification," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 2436–2439. doi: 10.1109/IC3I59117.2023.10398030.

[15] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit Lett*, vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.

[16] M. Burman, R. Singh, R. Das, S. Gauraha, and S. Kumar, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *International Journal of Scientific Research in Engineering and Management*, 2023, doi: 10.55041/IJSREM23473.

[17] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. Gnrsn Prudhvith, and N. Kumar, "Image Classification using SVM and CNN," in *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*, Institute of Electrical and Electronics Engineers Inc., Mar. 2020. doi: 10.1109/ICCSEA49143.2020.9132851.

[18] H. Hasan, H. Z. M. Shafri, and M. Habshi, "A Comparison between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models for Hyperspectral Image Classification," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1755-1315/357/1/012035.

[19] N. Boonnam, T. Udomchaipitak, S. Puttinaovarat, T. Chaichana, V. Boonjing, and J. Muangprathub, "Coral Reef Bleaching under Climate Change: Prediction Modeling and Machine Learning," *Sustainability (Switzerland)*, vol. 14, no. 10, May 2022, doi: 10.3390/su14106161.

[20] N. Ani Brown Mary and D. Dharma, "Coral reef image classification employing Improved LDP for feature extraction," Nov. 01, 2017, *Academic Press Inc.* doi: 10.1016/j.jvcir.2017.09.008.

[21] B. Hamuna, S. Pujiyati, J. L. Gaol, and T. Hestirianoto, "Spatial distribution of benthic habitats in Kapota Atoll (Wakatobi National Park, Indonesia) using remote sensing imagery," *Biodiversitas*, vol. 24, no. 7, pp. 3700–3707, 2023, doi: 10.13057/biodiv/d240706.

[22] Hartoni, V. P. Siregar, S. Wouthuyzen, and S. B. Agus, "Object based classification of benthic habitat using Sentinel 2 imagery by applying with support vector machine and random forest algorithms in shallow waters of Kepulauan Seribu, Indonesia," *Biodiversitas*, vol. 23, no. 1, pp. 514–520, Jan. 2022, doi: 10.13057/biodiv/d230155.

[23] J. J. Gapper, H. El-Askary, E. Linstead, and T. Piechota, "Coral reef change detection in remote Pacific Islands using support vector machine classifiers," *Remote Sens (Basel)*, vol. 11, no. 13, Jul. 2019, doi: 10.3390/rs11131525.

[24] N. W. Prabowo, V. P. Siregar, and S. B. Agus, "KLASIFIKASI HABITAT BENTIK BERBASIS OBJEK DENGAN ALGORITMA SUPPORT VECTOR MACHINES DAN DECISION TREE MENGGUNAKAN CITRA MULTISPEKTRAL SPOT-7 DI PULAU HARAPAN DAN PULAU KELAPA," *Jurnal Ilmu dan Teknologi Kelautan Tropis*, vol. 10, no. 1, pp. 123–134, Apr. 2018, doi: 10.29244/jitkt.v10i1.21670.

[25] S. Villon, M. Chaumont, G. Subsol, S. Villéger, T. Claverie, and D. Mouillot, "Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and HOG+SVM methods," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2016, pp. 160–171. doi: 10.1007/978-3-319-48680-2_15.

[26] A. Mehta, E. Ribeiro, J. Gilner, and R. Van Woesik, "CORAL REEF TEXTURE CLASSIFICATION USING SUPPORT VECTOR MACHINES," 2007.

[27] E. Ribeiro, A. Mehta, J. Gilner, and R. Van Woesik, "CORAL REEF TEXTURE CLASSIFICATION USING SUPPORT VECTOR MACHINES," 2007. [Online]. Available: https://www.researchgate.net/publication/221416032

[28] Z. Lu, "Enhanced Accuracy Enabled by Particle Swarm Optimization in Classification Application," in *Proceedings - 2020 International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 146–149. doi: 10.1109/ICAICE51518.2020.00034.

[29] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection." [Online]. Available: http://lear.inrialpes.fr

[30] N. Ani Brown Mary and D. Dharma, "A novel framework for real-time diseased coral reef image classification," *Multimed Tools Appl*, vol. 78, no. 9, pp. 11387–11425, May 2019, doi: 10.1007/s11042-018-6673-2.

[31] F. T. Anggraeny, B. Rahmat, and S. P. Pratama, "Deteksi Ikan Dengan Menggunakan Algoritma Histogram of Oriented Gradients," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 15, no. 2, p. 114, Sep. 2020, doi: 10.30872/jim.v15i2.4648.

[32] C.-J. Tu, L.-Y. Chuang, and C.-H. Yang, "Feature Selection Using PSO-SVM," Feb. 2007.

[33] G. Jin, P. Jin-Ye, and L. Zhan, *Application of Improved PSO-SVM Approach in Image Classification*. IEEE, 2010.

[34] P. N. Andono, G. F. Shidik, D. P. Prabowo, D. H. Yanuarsari, Y. Sari, and R. A. Pramunendar, "Feature Selection on Gammatone Cepstral Coefficients for Bird Voice Classification Using Particle Swarm Optimization," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 254–264, 2023, doi: 10.22266/ijies2023.0228.23.

[35] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, and D. Kriegman, "Automated annotation of coral reef survey images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1170–1177. doi: 10.1109/CVPR.2012.6247798.

[36] K. Thurnhofer-Hemsi, E. López-Rubio, M. A. Molina-Cabello, and K. Najarian, "Radial basis function kernel optimization for Support Vector Machine classifiers," Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.08233

[37] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology (Singapore)*, vol. 13, no. 5, Oct. 2021, doi: 10.1007/s41870-017-0080-1.