Classification of Mangrove Ecosystem Health Using Sentinel-2 Images with Genetic Algorithm Optimization in Machine Learning Algorithms

Putri Yuli Utami[1]*, Murni Ramadhani[2], Rudi Alfian[3], Barry Ceasar Octariadi[4], Dimas Kurniawan[5]
Information System-Faculty of Engineering and Computer Science, Universitas Muhammadiyah Pontianak, 78124 Indonesia[1, 2, 5]
Aquaculture-Faculty of Fisheries and Marine Science, Universitas Muhammadiyah Pontianak, 78124 Indonesia[3]
Informatics-Faculty of Engineering and Computer Science, Universitas Muhammadiyah Pontianak, 78124 Indonesia[4]

***Abstract*—Mangrove ecosystems play an important role in maintaining coastal ecological balance, including as carbon sinks and natural protection from abrasion, but mangrove areas in Mempawah Regency have experienced significant degradation due to anthropogenic pressures. Therefore, this study aims to classify the health condition of mangroves using multi-temporal Sentinel-2 imagery with a hybrid machine learning (ML) approach and Genetic Algorithm (GA) optimization. We implemented GA optimization comparatively on four main ML models—Multilayer Perceptron (MLP), Decision Tree (DT), XGBoost, and Naïve Bayes (NB)—to adjust hyperparameters to improve accuracy and reduce overfitting. The results prove that GA optimization effectively improves classification performance, with the MLP-GA model providing the highest accuracy with an increase of up to 3.8% compared to the non-optimized baseline model, achieving a best performance value of ROC AUC 0.9730 and reducing computation time by up to 60%. These findings indicate that the GA-MLP framework is highly reliable and efficient, providing a precise tool for strategic decision-making in the management of healthy mangrove ecosystems.**

Keywords*—*Classification; genetic algorithm; machine learning; mangrove ecosystem; Sentinel-2

I. Introduction

Mangrove forests cover only about 0.7% of the world's total tropical forests [1], but has the ability to store three to four times more carbon per unit area than other tropical forests, and in areas with carbonate and peat environments, their storage capacity can be 25 to 50% higher [2]. Globally, this ecosystem is spread across more than 118 countries with an area of approximately 152,000 km² [3]. Meanwhile, in Indonesia, mangroves are found throughout the archipelago. The largest distribution is in Irian Jaya, covering an area of 1,350,600 ha (38%), followed by Kalimantan with 978,200 ha (28%), and Sumatra with 673,300 ha (19%) [4][5]. Meanwhile, in other regions, mangroves generally grow optimally on coasts with large river estuaries and protected beaches.

West Kalimantan is a province with a coastline stretching 1,940 km, most of which is dominated by mangrove ecosystems [6][7]. The area of mangrove forests in this province is estimated to reach 119,327 ha, with a composition of around 75% of the mangrove species found in Indonesia. On Padang Tikar Island, Batu Ampar District, Kubu Raya Regency, the mangrove ecosystem covers an area of 58,953 ha. Meanwhile, Mempawah Regency, which is also located on the coast of West Kalimantan, has a fairly high diversity of mangroves [8]. However, in the last 25 years, it has experienced degradation covering an area of 250.11 ha, so that in 2014, only about 739.31 ha remained [9]. This condition indicates that the mangrove ecosystem in the study area not only has high ecological value but also faces serious threats, making remote sensing-based mangrove health monitoring urgently needed.

Monitoring the health of mangrove ecosystems is crucial to support conservation and rehabilitation strategies. The use of remote sensing technology, particularly Sentinel-2 satellite imagery, offers great opportunities because it has sufficient spatial, temporal, and spectral resolution to map land cover conditions and detect vegetation health dynamics [10][11][12]. Several vegetation and water indices, such as NDVI (Normalized Difference Vegetation Index) [13][14], SAVI (Soil-Adjusted Vegetation Index) [15], NDWI (Normalized Difference Water Index) [16], and MNDWI (Modified Normalized Difference Water Index) [17], have proven effective in evaluating vegetation density, humidity, and coastal ecosystem conditions. However, challenges still arise in terms of classification accuracy and consistency due to the complexity of mangrove spectral characteristics, which often overlap with other vegetation.

Previous studies have shown various approaches to mangrove mapping and monitoring. The XGBR-GA hybrid approach, which combines Extreme Gradient Boosting Regression and Genetic Algorithm for feature selection with multi-source data (Sentinel-2, Sentinel-1, ALOS-2 PALSAR-2) and field data (105 plots), successfully estimated Above-Ground Biomass (AGB) with good accuracy ($R^2 = 0.683$; RMSE = 25.08 $Mg \cdot ha^{-1}$) [18]. Furthermore, the EIAGA-S (Elite Individual Adaptive Genetic Algorithm-Semantic Inference) method, combined with the new MEVI index, was able to perform semantic segmentation of mangroves without ground truth with high accuracy (mIoU = 0.92; F1 = 0.923) and multi-class classification. On the other hand, NDVI-based research in Tugu District, Semarang, shows that the mangrove area covers 113.93 ha with a health condition dominated by the "Good" category (57.1 ha), although the effectiveness of mangroves in mitigating coastal abrasion (970 m in 2017–2021) is still limited [19]. The three studies introduced the Elite Individual Adaptive Genetic Algorithm-Semantic Inference (EIAGA-S) method for semantic segmentation of mangroves

* Corresponding author.

without requiring ground truth data. This method combines an adaptive genetic algorithm with an elite evolution strategy and develops the Mangrove Enhanced Vegetation Index (MEVI) to distinguish mangroves from other vegetation. Evaluation results on multi-source data show high performance with mIoU = 0.92 and F1-score = 0.923, surpassing traditional models such as K-means and SVM [20].

The development of machine learning methods has opened up new opportunities to improve the accuracy of land cover classification. Algorithms such as Multilayer Perceptron (MLP) excel at capturing nonlinear patterns [21], Extreme Gradient Boosting (XGBoost) is powerful in processing complex data [21], Easy-to-interpret Decision Tree [22], and Naïve Bayes, which is efficient for text data, has been widely used in satellite image analysis, but its performance is greatly influenced by the selection of complex hyperparameters [23]. To address these challenges, this study integrates Genetic Algorithm (GA) as an optimization method to improve accuracy while minimizing the risk of overfitting, so that the resulting model is more reliable in supporting coastal ecosystem monitoring [24][25].

Data processing was carried out using Sentinel-2 imagery from the COPERNICUS/S2 HARMONIZED collection accessed through Google Earth Engine (GEE), with the study area focused on the coast of Mempawah Regency, West Kalimantan, which is bounded by a polygon-shaped AOI. To monitor the temporal dynamics of mangrove ecosystem health, the images were analyzed over three time periods (2019–2021, 2021–2023, and 2023–2025), then several spectral indices such as NDVI, SAVI, NDWI, and MNDWI were calculated to assess the condition of mangrove vegetation and detect changes in land cover from year to year.

This study is expected to make a significant methodological and practical contribution. The methodological contribution lies in the comparative evaluation of Genetic Algorithm (GA) optimization in four machine learning models for mangrove health classification, which has rarely been explored in the context of multi-temporal Sentinel-2 data. Quantitatively, this study proves that GA can improve the accuracy of the best model (MLP) by up to 3.8% compared to the non-optimized base model, achieving the highest ROC AUC performance value of 0.9730, while reducing hyperparameter computation time by up to 60%. These precise data-based results are crucial for supporting conservation and rehabilitation strategies. Furthermore, this study aligns with the UN's 2030 Agenda, particularly Sustainable Development Goal (SDG) 14.2 [26], which emphasizes the protection and sustainable use of marine and coastal ecosystems, making healthy mangrove management key to supporting sustainable development. Overall, this study provides recommendations that can be directly applied in mangrove reforestation efforts, supporting sustainable development (SDGs 2030) through more targeted mangrove planting strategies to enhance ecological resilience in coastal areas [27].

The remainder of this study is organized as follows: Section II presents a review of various related works, highlighting the limitations of previous studies and explaining how this research addresses these gaps. Section III details the methods used, ranging from Sentinel-2 image processing, multi-index data preparation, and the implementation of Genetic Algorithm (GA) optimization on four machine learning models. Section IV presents the experimental results, and an in-depth discussion of the optimized model performance comparison is presented in Section V. Finally, Section VI summarizes the main conclusions of this study and provides directions for future work.

II. Related Work

Remote sensing-based mangrove ecosystem monitoring research has been conducted extensively using various approaches. One of the most widely used methods is vegetation indices, such as NDVI, SAVI, NDWI, and MNDWI, which have proven effective in assessing vegetation density, moisture conditions, and distinguishing between vegetative and non-vegetative areas. For example, research in the Semarang area successfully identified the extent of mangroves and classified their health conditions with a reasonable degree of accuracy [28]. However, the study was limited to a single observation period, so it was not able to describe changes in mangrove conditions over time.

With advances in data analysis technology, machine learning algorithms are increasingly being applied to improve the accuracy of satellite image classification. Various studies have demonstrated the success of algorithms such as Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), Decision Tree, and Naïve Bayes in land cover mapping and vegetation identification. MLP is known to be effective in capturing nonlinear patterns between spectral variables [29], whereas XGBoost excels at handling complex data and is able to overcome overfitting issues through its efficient gradient boosting mechanism [30][31]. Decision Trees are often used because they are easy to interpret and capable of displaying transparent classification rules [32]. Meanwhile, Naïve Bayes has an advantage in computational efficiency for high-dimensional data thanks to its simple yet effective assumption of feature independence [33]. However, the performance of these algorithms is highly dependent on the selection of appropriate hyperparameters. Manual parameter tuning is often inefficient and can reduce the accuracy and stability of the model, especially with complex remote sensing data [34][35].

To overcome problems in hyperparameter tuning, various studies have begun to apply metaheuristic-based optimization algorithms such as the Genetic Algorithm (GA). One of these studies was conducted by Tien Dat Pham et al. (2022), who developed the XGBR-GA model for estimating above-ground biomass (AGB) of mangroves. This model utilizes data from Sentinel-1, Sentinel-2, ALOS-2 PALSAR-2, as well as field data from 105 plots in the Red River Delta, Vietnam. The results of the study show the best performance with an R^2 value of 0.683 and an RMSE of 25.08 $Mg \cdot ha^{-1}$, making the XGBR-GA approach effective for monitoring mangrove ecosystems in tropical regions [36]. In addition, research by Xinhong Li et al. (2023) developed an explainable machine learning method for Fractional Vegetation Cover (FVC) inversion in the alpine grasslands of the Qinghai-Tibet Plateau. A combination of Genetic Algorithm (GA), XGBoost, and Optuna (GA-OP) was used for feature selection and hyperparameter tuning, resulting in a Stacking model with the best performance (R^2 = 0.867;

RMSE = 0.12). The SHAP method and NDVI-CV analysis improved the interpretability and reliability of the results, making this approach effective for estimating other ecological parameters [37]. These results demonstrate the effectiveness of GA in improving the accuracy, stability, and generalization ability of machine learning models for large-scale remote sensing data.

Although a number of previous studies have successfully improved the accuracy of mangrove classification through the application of machine learning algorithms and Genetic Algorithm (GA)-based optimization, most of these studies still have several important limitations. First, most previous studies have focused on a single observation period, thus failing to consider the temporal dynamics of mangrove ecosystems, which are greatly influenced by seasonal factors and anthropogenic activities. Second, the study areas tend to be limited to small and homogeneous areas, so the results cannot describe the complex spatial variations in larger tropical areas. Third, some studies only rely on one or two vegetation indices (such as NDVI or SAVI) without considering water indices such as NDWI and MNDWI, even though these two indices are important for assessing the interaction between mangrove vegetation and the surrounding aquatic environment. Fourth, the GA approach used is generally only applied to one machine learning model, so there has been no discussion of the effectiveness of multi-model or cross-algorithm optimization.

This study attempts to overcome these limitations through a more comprehensive and adaptive approach to classifying mangrove ecosystem health. First, a multi-temporal analysis was conducted for the period 2019 to 2025 to describe the dynamics of changes in mangrove health over time. Second, the study area was focused on the coast of Mempawah Regency, West Kalimantan—an area with high vegetation diversity and real environmental pressures—so that it could represent more diverse ecological conditions. Third, this study combined four main indices (NDVI, SAVI, NDWI, and MNDWI) to provide a more comprehensive spatial picture of vegetation and water conditions. Fourth, the novelty of this study lies in the application of GA across four machine learning algorithms (MLP, XGBoost, Decision Tree, and Naïve Bayes), which are simultaneously optimized to improve the accuracy, stability, and generalization ability of the model. Thus, this study not only expands the spatial and temporal coverage but also introduces methodological innovations in the application of GA for more effective and scientific classification of mangrove ecosystems.

III. METHODOLOGY

This research method consists of nine main stages for land cover classification and mangrove health assessment based on Sentinel-2 imagery. The stages include determining the study area, acquiring Sentinel-2 Level-2A imagery, pre-processing (cloud masking and extraction of six spectral bands), and creating a labeled dataset through pixel value extraction from sample points. Five machine learning algorithms (MLP, Naive Bayes, Decision Tree, and XGBoost) were applied with parameter optimization using Genetic Algorithm. Model evaluation was performed using accuracy, precision, recall, F1-score, and Cohen's Kappa, then the best model was used to produce a mangrove cover classification map and interpretation of its health condition. The overall research process flow, from image acquisition to result interpretation, can be seen visually in Fig. 1.

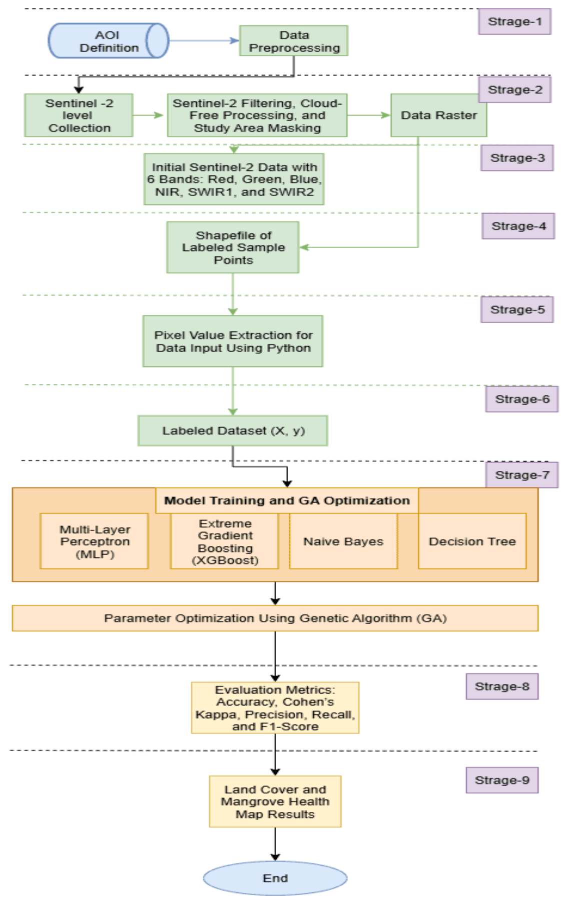


Fig. 1. Research process flow.

A. Stage-1: Determination of AOI and Data Preparation

The initial stage of the study began with the determination of the Area of Interest (AOI) located in the coastal area of Mempawah, West Kalimantan. This AOI was defined using polygon vector data with a two-dimensional geographic coordinate system (Geographic 2D CRS: EPSG:4326 – WGS 84), covering an area with longitude coordinates ranging from 108.829975° to 109.043271° and latitude coordinates ranging from 0.295701° to 0.565236°. The projection system used is WGS 84 (World Geodetic System 1984) with an ellipsoidal datum and Greenwich prime meridian. This shapefile data is then used to determine sample points in the class labeling process. The base image used is Sentinel-2 Level-2A, which was collected through the Google Earth Engine (GEE) platform and will be used in the spectral value extraction stage.

B. Stage-2: Sentinel-2 Image Processing

The Sentinel-2 images collected were then processed through several stages, including masking the study area, cloud filtering, and selecting the study area to produce a raster ready for analysis. This study uses Sentinel-2 images from the COPERNICUS/S2 HARMONIZED collection available on the Google Earth Engine (GEE) platform (https://code.earthengine.google.com/), with a cloud cover filter of less than 20%. The study area is focused on the coastal region of Mempawah Regency, West Kalimantan, and is bounded by a polygonal AOI.

The image selection process was carried out for three time periods, namely: 2019 to 2021, 2021 to 2023, and 2023 to 2025, in order to monitor the dynamics of mangrove cover changes over time. The selection of these three time periods aims to monitor the temporal dynamics of changes in mangrove ecosystem health, thereby providing an overview of degradation and recovery trends from year to year.

Each image that meets the criteria is then calculated for its spectral index value, such as NDVI, SAVI, MNDWI, and NDWI, to identify the health level of mangrove vegetation. The processing results are stored in TIFF format. AOI in the form of a shapefile with the ID asset projects/mengrove-riset-24/assets/AOI is accessed through the GEE API and converted into FeatureCollection, with geometric information obtained through the getInfo() method.

C. Stage-3: Spectral Band Extraction

The initial Sentinel-2 image was extracted using six important bands, namely Red, Green, Blue, NIR, SWIR1, and SWIR2. These bands were selected because they are relevant for detecting vegetation and plant health, especially in mangrove ecosystems.

TABLE I VEGETATION AND WATER INDICES FROM SENTINEL-2 IMAGES

Vegetation Indices	**Formula**	*Sources*
Blue	B2	-
Green	B3	-
Red	B4	-
NIR	B8	-
SWIR-1	B11	-
SWIR-2	B2	-
NDVI (*Normalized Difference Vegetation Index*)	(NIR - Red) / (NIR + Red)	[38]
SAVI (*Soil-Adjusted Vegetation Index*)	((NIR - Red) / (NIR + Red + L)) * (1 + L)	[39]
MNDWI (*Modified Differential Water Index*)	(Green - SWIR1) / (Green + SWIR1)	[40]
NDWI (*Normalized Difference Water Index*)	(Green - NIR) / (Green + NIR)	[41]

As presented in Table I, various vegetation and water indices were calculated from a combination of Sentinel-2 spectral bands to assess vegetation cover and water conditions. Four main indices were used in this study, namely NDVI, SAVI, NDWI, and MNDWI. NDVI is widely used to measure vegetation density and greenness, making it a relevant indicator of mangrove health. SAVI was developed to minimize the influence of soil background, making it more accurate in areas with sparse or uneven vegetation cover, including mangrove ecosystems in coastal areas. Meanwhile, NDWI and MNDWI are used to detect vegetation moisture and distinguish water areas from dry or built-up land—important parameters in mangrove ecosystems that are highly influenced by tidal and water conditions. Thus, the combination of these indices provides a more comprehensive spatial picture of mangrove health status.

D. Stage-4: Sample Labeling

Sample labeling was carried out using points from a shapefile that had been prepared beforehand. These points represent land cover classes and mangrove health conditions that were determined manually through visual interpretation and field verification. In this study, 1,250 sample points were used, divided evenly into five land use classes, each with 250 points, namely: 1) Water Body, 2) Non-Mangrove Vegetation, 3) Mangrove, 4) Built-up Land, and 5) Open Land. Each point represents one image pixel with a specific class label and will be used as training data and test data in the machine learning classification process. As shown in Fig. 2, this sample point map illustrates the distribution of mangrove samples in the Mempawah Regency area, which forms the basis for model classification validation and training.

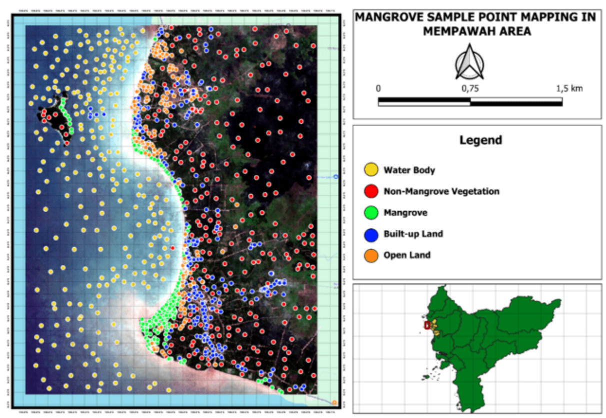


Fig. 2. Mangrove sample point mapping in the Mempawah area.

E. Stage-5: Pixel Value Extraction

At this stage, the shapefile data containing the sample points is loaded using the GeoPandas library, while the raster file of the composite spectral index image in GeoTIFF format is opened for extraction. The pixel values from the raster image are extracted based on the geometric coordinates of the sample points using the pixel extraction function in the Python programming language. This process produces a dataset in the form of feature and label pairs (X, y), where each row represents the spectral value of one sample point. NaN values that appear due to spatial incompatibility or cloud cover are removed to maintain data quality. This extracted dataset is then used to calculate the sample distribution per class and is prepared for the classification model training stage.

F. Stage-6: Labeled Dataset Formation

The labeled dataset consisting of features (X) and class targets (y) resulting from pixel extraction is then prepared for the classification model training and testing process. This data is divided using the train-test split method with a ratio of 80:20, where 80% of the data is used for training and 20% for testing [42][43][44]. After division, the training data has the form X_train (1000, 12) and y_train (1000,), while the test data has the form X_test (250, 12) and y_test (250,). Each row represents one sample point with 12 spectral features and vegetation indices, as well as one class label as the target.

G. Stage-7: Model Training and Optimization

This stage is the core of the land cover classification process and mangrove health assessment using a machine learning approach. This study proposes a new approach by integrating Genetic Algorithm (GA) as a hyperparameter optimization method across four main algorithms, namely Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), Naïve Bayes, and Decision Tree. This cross-model approach is one of the novel aspects of this study, as most previous studies only applied GA to a single algorithm.

Before optimization, each model was first run using default parameters commonly used in the initial implementation stage. Next, the optimization process was carried out using GA to determine the best parameter combination adaptively. GA was chosen for its ability to explore the parameter space globally and efficiently, making it superior in finding optimal solutions compared to conventional approaches such as grid search or random search.

In addition, this study also presents another novel aspect, namely the application of GA in multi-temporal analysis (2019 to 2021, 2021 to 2023, and 2023 to 2025) combined with four main spectral indices — NDVI, SAVI, NDWI, and MNDWI — to describe the spatial and temporal variations in mangrove ecosystem health. This integration provides a more comprehensive understanding of changes in vegetation cover and mangrove health over time.

Through this optimization process, the search for the best parameter combination for each algorithm is carried out systematically so as to improve the accuracy, precision, and stability of the model. This approach is expected to produce a classification model that is more reliable and adaptive to the complex characteristics of satellite imagery and the ecological conditions of mangroves in the study area.

Algorithm 1 Multilayer Perceptron (MLP)

Step 1: Input Layer

- **Input**: 12 features (e.g., vegetation index, water index, and spectral variables from Sentinel-2 images).
- **Output**: Input vector of size **(n_samples, 12)** ready to be passed to Hidden Layer 1.

Step 2: Hidden Layer 1

- **Process**:
  - Number of neurons: 50
  - Calculation: $z^{(1)} = X \cdot W^{(1)} + b^{(1)}$
  - Activation: tanh
- **Output**: Vector of size **(n_samples, 50)** resulting from tanh activation.

Step 3: Hidden Layer 2

- **Process**:
  - Number of neurons: 25
  - Calculation: $z^{(2)} = a^{(1)} \cdot W^{(2)} + b^{(2)}$
  - Activation: tanh
- **Output**: Vector sized **(n_samples, 25)** resulting from tanh activation.

Step 4: Output Layer

- **Process**:
  - Number of neurons: 5 (corresponding to the number of target classes)
  - Activation: softmax to convert values into probabilities for each class.
- **Output**: A probability vector **(n_samples, 5)** that sums to 1.

Step 5: Class Prediction

- **Input**: Probabilities from the output layer.
- **Process**: argmax → selects the class with the highest probability.

Algorithm 2 Extreme Gradient Boosting (XGBoost)

Step 1: Input Layer

- **Input:** 12 features → derived from vegetation indices (NDVI, SAVI, EVI, LAI), water indices (NDWI, MNDWI), and Sentinel-2 spectral variables.
- **Dimension:** $(n\_samples, 12)$
- **Output:** Feature data ready to be processed into the XGBoost model.

Step 2: Base Learner (CART)

- **Base model:** Decision Tree (Classification and Regression Tree).
- **Process:** The split process is calculated using **an objective function** (loss + regularization).

$$Obj = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)$$

  - l = loss function (e.g., log-loss for classification)
  - $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum w_j^2$ = (regularization, prevents overfitting)
- **Output:** The first tree is formed (initial prediction).

Step 3: Ensemble of Trees (Boosting Process)

- **Residual** is calculated from the derivative of loss (gradient):

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i}, h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^2}$$

- Node split is selected with **the** largest **gain**:

$$Gain = \frac{1}{2}\left[\frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_i + \lambda} + \frac{(\sum_{i \in R} g_i)^2}{\sum_{i \in R} h_i + \lambda} - \frac{(\sum_{i \in L \cup R} g_i)^2}{\sum_{i \in L \cup R} h_i + \lambda}\right] - \gamma$$

- Learning rate (η): adjusts the contribution of each tree.
- **Output: a collection of hierarchical trees that improve the error.**

Step 4: Output Layer (Softmax for Multi-class)

- **The final prediction for each sample is a class probability.**

$$P(y = k \mid x) = \frac{e^{\hat{y}_k}}{\sum_{j=1}^{K} e^{\hat{y}_j}}$$

- Dimension: $(n\_samples, 5)$

Step 5: Class Prediction

- From the softmax probabilities → select the class with the highest **argmax** value.

Algorithm 3 Naive Bayes

Step 1: Input Layer (Input Data)

- Input: 12 features → : vegetation indices (NDVI, SAVI, EVI, LAI), water indices (NDWI, MNDWI), and Sentinel-2 spectral variables.
- Dimension: (n_samples, 12).
- Output: Feature data ready for modeling.

Step 2: Data Normalization

- Scaler: StandardScaler is used to standardize features (mean=0, std=1).
- Output: Training and test data in standard scale.

Step 3: Naive Bayes Model (GaussianNB)

- Assumption: Each feature in a class follows a Gaussian (normal) distribution.
- Parameters calculated for each class k

$$P(x \mid C_k) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} \exp\left(-\frac{(x_i - \mu_{ki})^2}{2\sigma_{ki}^2}\right)$$

Where:

- μ_{ki} = mean of the Ith feature in class k
- $\sigma_{ki}^2 = varians\ fitur\ ke-i\ pada\ kelas\ k$
- Output: Gaussian Naive Bayes model with mean and variance parameter estimates per class.

Step 4: Bayes Classification

- Class probabilities are calculated using Bayes' rule:

$$P(C_k \mid x) = \frac{P(x \mid C_k)P(C_k)}{P(x)}$$

- Final class prediction = class with the highest probability (*argmax*).
- Output: Predicted label for each sample (0–4).

Step 5: Output Layer (Multi-class Probabilities)

- Function: predict_proba() generates the probability of each class.
- Format: (n_samples, 5).
- Output: Probability distribution for all classes.

Algorithm 4 Decision Tree

Step 1: Input Layer (Data Features)

- Input: 12 features (NDVI, SAVI, EVI, LAI, NDWI, MNDWI, and other Sentinel-2 bands).
- Dimensions: $(n\_samples, 12)$.
- Output: Feature data ready to be processed into a decision tree.

Step 2: Attribute Selection (Splitting Criteria)

- The algorithm selects the best attribute to split the data.
- Criteria: Gini Index or Entropy (Information Gain).
- Gini Index Formula:

$$Gini(D) = 1 - \sum_{i=1}^{k} p_i^2$$

Where: p_i the proportion of class-i in the dataset D

- **Information Gain (IG) Formula**:

$$IG(D, A) = Entropy(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} \times Entropy(D_v)$$

$$Entropy(D) = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

The attribute with the largest IG value (or smallest Gini) is selected as the splitting node.

Step 3: Recursive Splitting (Building the Tree)

- The process of selecting the best attribute is performed repeatedly (recursively) for the data subset.
- The tree will be formed until the stopping criteria is met, for example:
  - Maximum depth is reached (max_depth).
  - The number of samples in the node is less than min_samples_split.
  - The node is already "pure" (contains only 1 class).

Step 4: Leaf Node (Output Node)

- If the stopping criteria are met, the node becomes a leaf.
- The leaf contains the class distribution, and the label is selected based on the majority class.

Step 5: Class Prediction

- For each test data, features are checked following the tree's condition path.
- Data stops at the leaf node → class prediction result.

---

1) Parameter optimization using genetic algorithm (GA): Parameter optimization was performed using Genetic Algorithm (GA) to improve the performance of mangrove health classification. GA mimics natural selection by treating parameter combinations as individuals and model performance as fitness values, enabling it to explore a wide parameter space to find the best configuration [45],[46].

In this study, GA was applied to optimize four algorithms, namely Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), Naïve Bayes, and Decision Tree (see Algorithm 1 to Algorithm 4). GA adjusted important parameters such as network architecture, activation function, learning rate, and regulation in MLP; number of estimators, tree depth, learning rate, and regularization in XGBoost; variance smoothing in Naïve Bayes; and tree depth and minimum sample size in Decision Tree. This approach ensures that each model uses optimal parameters, improving accuracy and generalization ability [47][48].

Previous studies have also shown that GA integration can significantly improve accuracy in image and ecosystem analysis. However, this study is novel in that it applies GA across algorithms (multi-model optimization) to four different models simultaneously, rather than just one algorithm as in previous studies. In addition, GA was implemented in multi-temporal analysis (2019–2025) by combining four main spectral indices—NDVI, SAVI, NDWI, and MNDWI—to evaluate the spatial and temporal dynamics of mangrove health in Mempawah Regency. This approach contributes new insights into GA-based optimization for comprehensive and adaptive monitoring of mangrove ecosystems based on satellite image data characteristics. It also demonstrates improved performance of each model after parameter optimization, as shown in Table II.

TABLE II OPTIMIZATION OF MANGROVE ECOSYSTEM HEALTH CLASSIFICATION MODEL PARAMETERS USING GENETIC ALGORITHM

Model	Parameter
mlp	hidden_layer_sizes (50, 50), activation relu, solver adam, learning_rate_init 0.01, alpha 0.0001, batch_size 64
XgBoost	n_estimators = 133, max_depth = 4, learning_rate = 0.0818, gamma = 0.0934, subsample = 0.8 colsample_bytree = 0.8, reg_alpha = 0.3, reg_lambda = 1.5, random_state = 42, eval_metric = 'mlogloss'
Naive Bayes	GaussianNB(var_smoothing=0.0001394432)
Decision Trees	max_depth = 7, min_samples_split = 6, min_samples_leaf = 3

H. Stage-8: Model Evaluation

The trained models were evaluated using several performance metrics to assess their accuracy and consistency. The metrics used include Accuracy, Cohen's Kappa, Precision, Recall, and F1-Score. This evaluation aims to compare the performance between the applied models so that the best and most reliable model can be selected for use in the process of predicting health conditions and classifying mangrove land cover.

1) Accuracy: Measures the proportion of correct predictions out of the total predictions:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2) Precision: Measures how many positive predictions are correct:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

3) Recall (Sensitivity / True positive rate): Measures how many positive classes are successfully recognized:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

4) F1-Score: Harmonic mean of Precision and Recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

5) Cohen's Kappa: Measures agreement between predictions and actual labels, corrected for random agreement:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (5)$$

p_o : Observed Agreement (same as Accuracy)

p_e : Random Expected Agreement

$$p_e = \sum_{i=1}^{k} \left(\mathrm{p}_i^{\text{true}} \cdot \mathrm{p}_i^{\text{pred}}\right) \quad (6)$$

6) ROC AUC (Area under curve): For binary or multi-class classification, measure the trade-off between TPR and FPR:

$$\text{AUC} = \int_0^1 TPR(FPR)\, dFPR \quad (7)$$

- ROC Curve: plot between TPR and FPR
- AUC = Area under the ROC Curve (the closer to 1, the better)

Confusion Matrix

Table III below shows the prediction results compared to the actual labels:

TABLE III COMPARISON OF THE PREDICTION RESULTS

	Predicted Pos	Predicted Neg
Actual Pos	TP	FN
Actual Neg	FP	TN

For multi-class, the form becomes a matrix of size $k \times k$, where:

- Rows: actual labels
- Columns: model predictions

Therefore:

- True Positive (TP) = the number of cases where y_pred == 1 and y_true == 1

$$TP = \sum_{i=1}^{N} [(y_{\text{pred}_i} = 1) \wedge (y_{\text{true}_i} = 1)] \quad (8)$$

- True Negative (TN) = the number of cases when y_pred == 0 and y_true == 0

$$TN = \sum_{i=1}^{N} [(y_{\text{pred}_i} = 0) \wedge (y_{\text{true}_i} = 0)] \quad (9)$$

- False Positive (FP) = : the number of cases where y_pred == 1 and y_true == 0

$$FP = \sum_{i=1}^{N} [(y_{\text{pred}_i} = 1) \wedge (y_{\text{true}_i} = 0)] \quad (10)$$

- False Negative (FN) = : the number of cases where y_pred == 0 and y_true == 1

$$FN = \sum_{i=1}^{N} [(y_{\text{pred}_i} = 0) \wedge (y_{\text{true}_i} = 1)] \quad (11)$$

I. Stage-9: Mapping and Interpretation

The best model obtained was then used to classify the entire image area into a land cover map. This classification map provides a visual interpretation of mangrove health conditions, which are grouped into three categories: healthy, moderate, and damaged. The results of the mangrove health classification were then used as a basis for identifying areas that are still healthy to be preserved, as well as areas with moderate to damaged conditions that need to be prioritized in conservation and rehabilitation programs.

Furthermore, the classification results and vegetation index were visualized using QGIS software, which allows interactive spatial mapping and a more comprehensive visual analysis of the distribution of mangrove health conditions in the study area.

Thus, all stages of this methodology not only produce a map of land cover classification and mangrove health conditions, but also directly support the research objectives, namely monitoring the temporal dynamics of mangrove ecosystems in the 2019 to 2025 period and providing relevant spatial information to support policy planning, conservation, and coastal zone rehabilitation.

IV. RESULTS

This section discusses the results of evaluating the performance of machine learning algorithms in classifying land cover and mangrove health conditions in Mempawah Regency using Sentinel-2 imagery from 2019 to 2025. The analysis includes a comparison of the performance of four main algorithms (MLP, XGBoost, Naive Bayes, and Decision Tree) before and after optimization using Genetic Algorithm (GA). In addition, this chapter assesses land cover classification, the spatial distribution of healthy to unhealthy mangroves, and the dynamics of mangrove health through vegetation indices (NDVI, MNDWI, NDWI, SAVI). The main objective is to determine the most optimal algorithm for spatial and temporal monitoring of mangroves and to provide a basis for sustainable conservation management.

TABLE IV EVALUATION TABLE BEFORE OPTIMIZATION

Algorithm	Year	Train Accuracy	Test Accuracy	Cohen's Kappa	ROC AUC Score	Evaluation Time
MLP	2019 -2021	0.8340	0.8200	0.7800	0.9664	43.3s
	2021-2023	0.8690	0.8360	0.8350	0.9691	31.6s
	2023-2025	0.8690	0.8360	0.8350	0.9691	56.1s
XGBoost	2019 -2021	0.9780	0.8200	0.7800	0.9664	22.50s
	2021-2023	0.9810	0.8440	0.8350	0.9691	16.81s
	2023-2025	0.9810	0.8440	0.8350	0.9691	16.08s
Naive Bayes	2019 -2021	0.7140	0.7240	0.6450	0.9252	0.005s
	2021-2023	0.7390	0.7560	0.6800	0.9252	0.005s
	2023-2025	0.7390	0.7560	0.6800	0.9252	0.004s
Decision Tree	2019 -2021	0.8320	0.7880	0.7350	0.9308	0.87s
	2021-2023	0.9540	0.8280	0.7850	0.9077	0.85s
	2023-2025	0.9540	0.8280	0.7850	0.9077	0.92s

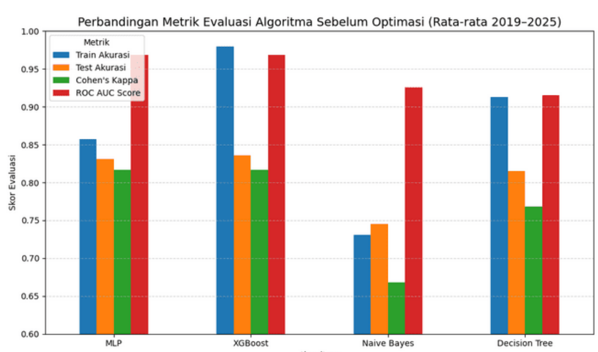


Fig. 3. Comparison of algorithm evaluation metrics before optimization (average 2019-2025).

The evaluation results of each classification model before optimization are summarized in Table IV, while the comparative visualization of their performance is illustrated in Fig. 3. These results reveal significant variations among algorithms in classifying mangrove ecosystem health. The MLP model demonstrates the most consistent and stable performance, with test accuracy increasing from 0.820 (2019–2021) to 0.836 (2023–2025), accompanied by a high Cohen's Kappa value (0.780 → 0.835) and an excellent ROC AUC (0.9664 → 0.9691). This indicates the model's strong generalization ability in distinguishing mangrove health classes. In contrast, the XGBoost model achieves very high training accuracy (0.978 → 0.981) but slightly lower test accuracy (0.820 → 0.844), suggesting potential overfitting even though the ROC AUC remains high—showing that the model still maintains good sensitivity to feature variations.

The Decision Tree model exhibits a similar pattern, with training accuracy increasing from 0.832 to 0.954 and moderate test accuracy (0.788 → 0.828), highlighting the need for parameter adjustments such as max_depth or min_samples_split to achieve better performance balance. Meanwhile, the Naive Bayes model shows the fastest evaluation time (<0.01 s) but relatively lower test accuracy (0.724–0.756) and moderate Kappa values (0.645–0.680), making it more suitable as a baseline or comparison model. Overall, the findings presented in Table IV and Fig. 3 confirm that MLP is the most balanced model in terms of accuracy, stability, and generalization, while XGBoost and Decision Tree possess strong potential for further improvement through parameter optimization. Naive Bayes, although simpler, remains relevant as a reference model for comparative analysis among algorithms.

TABLE V EVALUATION TABLE AFTER OPTIMIZATION

Algorithm	Year	Train Accuracy	Test Accuracy	Cohen's Kappa	ROC AUC Score	Evaluation Time
MLP	**2019 -2021**	0.899	0.8250	0.8250	0.9719	135m 10.7s
	2021–2023	0.926	0.868	0.8350	0.9730	98m47.1s
	2023–2025	0.919	0.852	0.8150	0.9719	133m39.2s
XGBoost	**2019–2021**	0.9860	0.8200	0.7750	0.9672	0.41s
	2021-2023	0.9830	0.8400	0.8000	0.9709	0.60s
	2023–2025	0.9830	0.8400	0.8000	0.9709	0.62s
Naive Bayes	**2019 -2021**	0.7120	0.7160	0.6450	0.9252	0.0250s
	2021–2023	0.7360	0.7440	0.6800	0.9249	0.0170s
	2023–2025	0.7360	0.7400	0.6750	0.9251	0.0220s
Decision Tree	**2019 -2021**	0.884	0.784	0.73	0.865	0.0274s
	2021–2023	0.889	0.824	0.78	0.89	0.0272s
	2023–2025	0.889	0.824	0.78	0.89	0.0328s

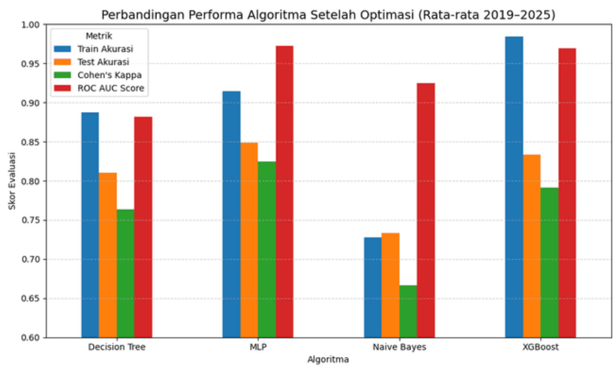


Fig. 4. Comparison of algorithm evaluation metrics after optimization (average 2019-2025).

As shown in Table V and Fig. 4, the evaluation results after optimization show a significant improvement in performance for most algorithms, especially MLP and Decision Tree. MLP showed high training accuracy (0.899–0.926) with test accuracy increasing from 0.825 to 0.868, stable Cohen's Kappa values (0.815–0.835), and excellent ROC AUC (0.9719–0.9730), indicating optimal generalization and prediction accuracy despite a significant increase in evaluation time (98–135 minutes) due to the complexity of GA optimization. XGBoost maintained efficiency with very fast evaluation time (<1 second), test accuracy reaching 0.820–0.840, and a small increase in Cohen's Kappa (0.775–0.800) and ROC AUC (0.9672–0.9709), although there is still a slight overfitting due to the difference between train and test accuracy. Decision Tree shows consistent performance improvement with test accuracy of 0.784–0.824, Kappa of 0.73–0.78, and ROC AUC of 0.865–0.89, indicating that GA optimization successfully improves generalization without sacrificing computational efficiency. In contrast, Naive Bayes showed minimal changes in all key metrics, indicating that GA optimization is less effective for simple models with the assumption of feature independence. Overall, Table V and Fig. 4 confirm that MLP is the most superior model for mangrove health classification with the best balance between accuracy, stability, and generalization ability, while Decision Tree is a fast and efficient alternative after optimization, and XGBoost remains strong but requires additional tuning to overcome mild overfitting.

A. Land Cover and Mangrove Health Classification Results

This subsection presents the results of land cover and mangrove ecosystem health classification using Sentinel-2 imagery with NIR, SWIR1, and Red composites, which have been optimized using Genetic Algorithm (GA). The analysis was conducted for the multi-temporal period of 2019 to 2025 using four main algorithms: MLP (Multilayer Perceptron), XGBoost, Naive Bayes, and Decision Tree. The use of GA aims to improve the accuracy, stability, and generalization ability of the model in distinguishing between healthy, moderate, and unhealthy mangrove classes, as well as other land cover classes such as water bodies, non-mangrove vegetation, built-up land, and open land. The resulting classification maps are visualized in Fig. 5 to Fig. 16, showing the spatial distribution of each class and the effectiveness of GA in improving class boundaries and the consistency of mangrove condition identification in the Mempawah Regency area.

1) Sentinel-2 image composite (NIR, SWIR1, Red) and MLP (multilayer perceptron) mangrove ecosystem land cover classification results

Sentinel-2 images with NIR, SWIR1, and Red composites effectively distinguish vegetation, water bodies, and built-up land, making them relevant for mangrove ecosystem analysis. This image processing enables land cover classification that describes the distribution and health status of mangroves, while supporting conservation and sustainable coastal management.

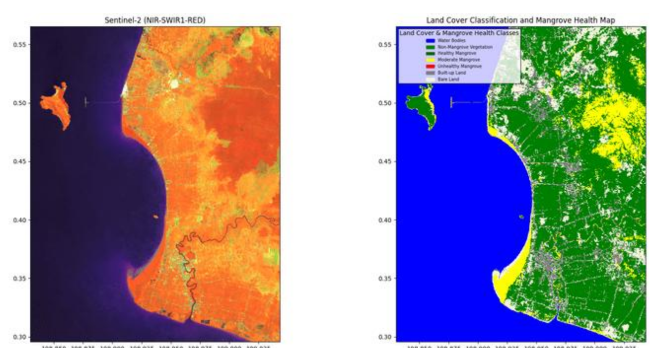


Fig. 5. Land cover classification and mangrove health (MLP) map 2019-2021.

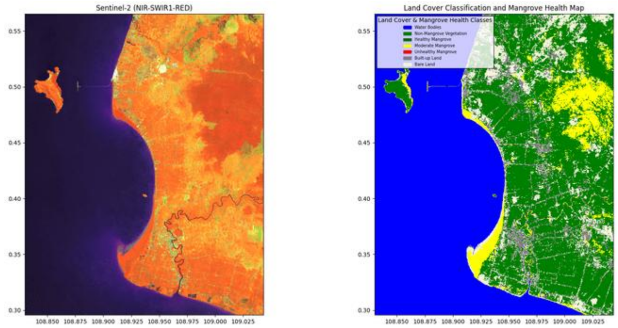


Fig. 6. Land cover classification and mangrove health (MLP) map 2021-2023.

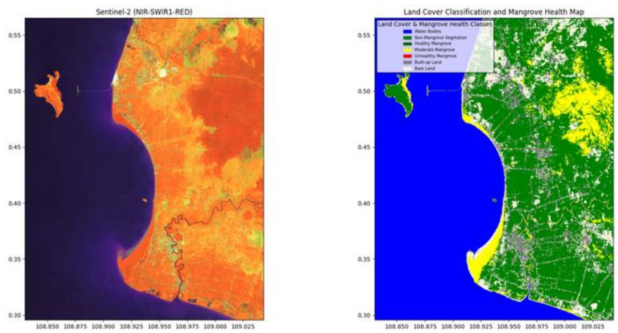


Fig. 7. Land cover classification and mangrove health (MLP) map 2023-2025.

2) Sentinel-2 image composite (NIR, SWIR1, Red) and mangrove ecosystem land cover classification results XGBoost

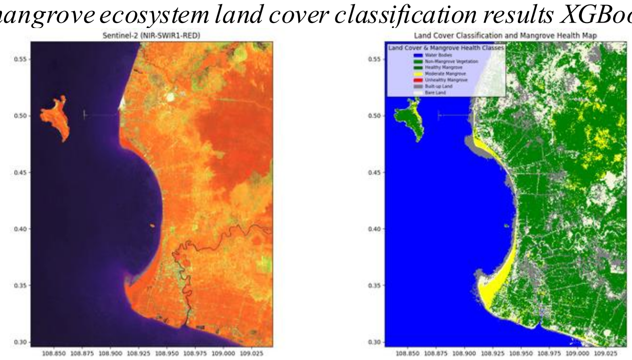


Fig. 8. Land cover classification and mangrove health map (XGBoost) 2019-2021.

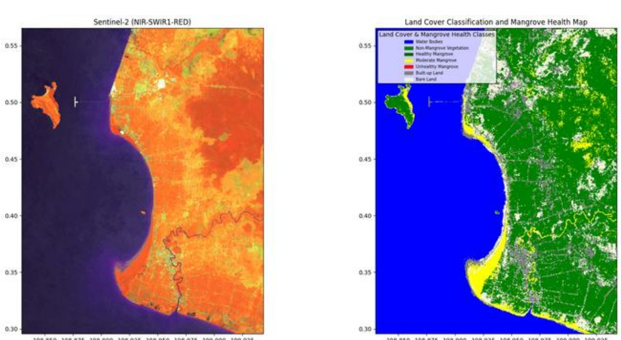


Fig. 9. Land cover classification and mangrove health map (XGBoost) 2021-2023.

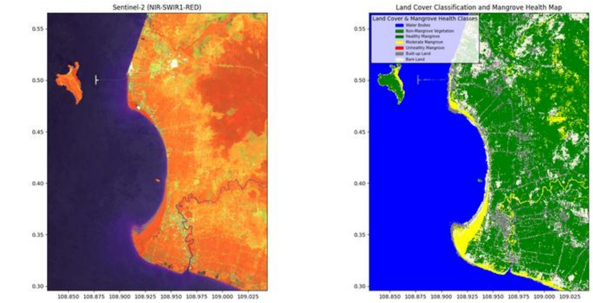


Fig. 10. Land cover classification and mangrove health map (XGBoost) 2023-2025.

3) Sentinel-2 image composite (NIR, SWIR1, Red) and mangrove ecosystem land cover classification results Naive Bayes

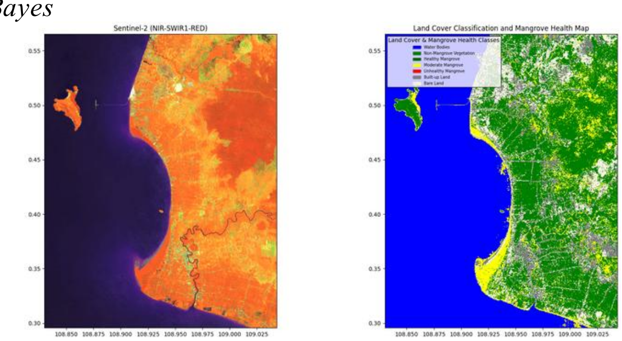


Fig. 11. Land cover and mangrove health classification map (Naive Bayes) 2019-2021.

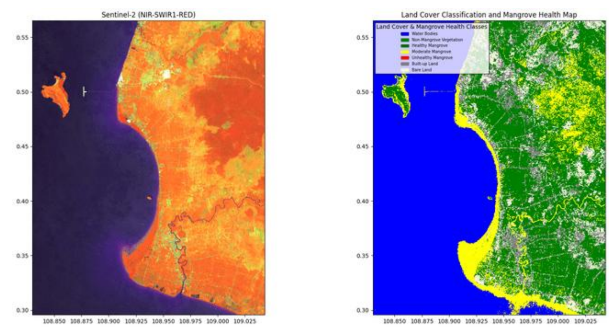


Fig. 12. Land cover and mangrove health classification (Naive Bayes) 2021-2023.

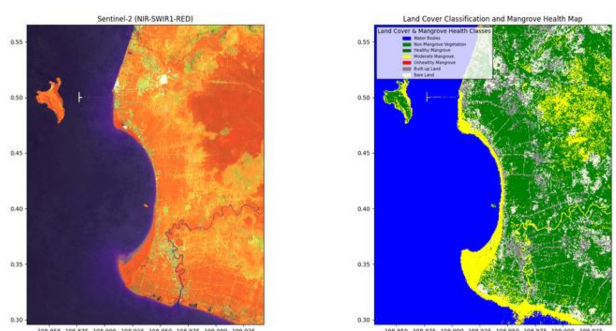


Fig. 13. Land cover and mangrove health classification (Naive Bayes) 2023-2025.

4) Sentinel-2 image composite (NIR, SWIR1, Red) and mangrove ecosystem land cover classification results Decision Tree

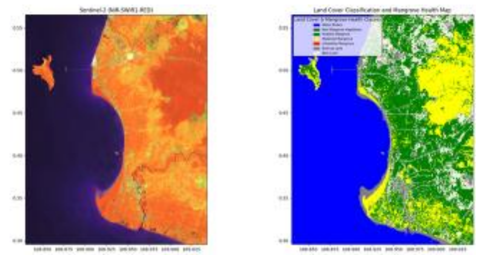


Fig. 14. Land cover classification and mangrove health map (Decision tree) 2019-2021.

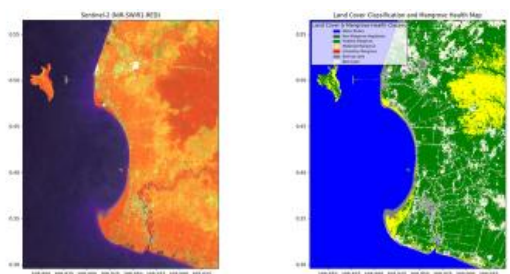


Fig. 15. Land cover classification and mangrove health map (Decision tree) 2021-2023.

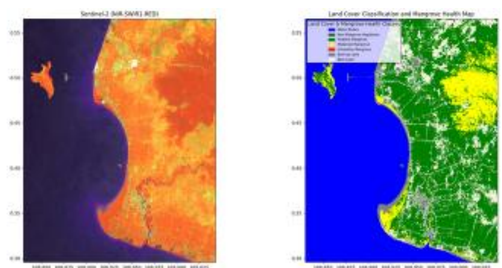


Fig. 16. Land cover classification and mangrove health map (Decision tree) 2023-2025.

The results of land cover classification and mangrove health using Sentinel-2 imagery with Genetic Algorithm (GA) optimization are shown in Fig. 5 to Fig. 16. The MLP algorithm (Fig. 5 to Fig. 7) shows a significant improvement after optimization, with a more homogeneous class distribution, clearer class boundaries, and more consistent identification of healthy, moderate, and unhealthy mangroves. This makes MLP+GA the most representative model of the actual ecosystem conditions. The XGBoost algorithm (Fig. 8 to Fig. 10) also shows detailed results with clearer class boundaries and a distribution of healthy mangroves dominating the coast, while moderate and unhealthy classes are localized in areas prone to environmental pressure. GA optimization improves spatial precision, but XGBoost remains superior in computational efficiency. Meanwhile, Naïve Bayes (Fig. 11 to Fig. 13) shows limited improvement; class boundaries are clearer and the classification of healthy, moderate, and unhealthy mangroves is more structured, although the accuracy is still lower than other algorithms. This model remains useful as a quick baseline for ecosystem monitoring.

In Decision Tree (Fig. 14 to Fig. 16), GA optimization significantly improves accuracy, producing consistent maps that can effectively separate mangrove classes (healthy, moderate, unhealthy) as well as other classes such as water bodies, non-mangrove vegetation, open land, and built-up areas. Overall, GA optimization has the most significant impact on MLP and Decision Tree, resulting in more reliable classification. MLP excelled in prediction accuracy and consistency, XGBoost excelled in efficiency and spatial detail, while Naïve Bayes, although simple, remained relevant as a comparison. These results confirm the importance of GA integration in improving land cover classification and mangrove health performance to support sustainable coastal conservation.

B. Spectral Reflectance Curves for Each Grade

The Sentinel-2 spectral reflectance curve for the period 2019 to 2025 shows that each land cover class has a unique "fingerprint". Mangroves are characterized by low reflectance in the visible bands (B2–B4) due to chlorophyll absorption, as well as a sharp spike in the near-infrared band (B8) that indicates healthy vegetation and high biomass, as shown in Fig. 17 to Fig. 19. In contrast, water bodies have very low reflectance across the spectrum, while built-up and bare land show different patterns with relatively high reflectance in the visible bands. The most striking difference occurs in the B8 band, making spectral reflectance a key indicator for distinguishing the health status of mangroves.

Ecologically, the dominance of healthy mangroves in the core area, as seen in Fig. 17 to Fig. 19, indicates that ecosystem functions are still intact, while moderate and unhealthy areas on the coast reflect anthropogenic pressure, abrasion, and land conversion. From a managerial perspective, this classification map is an important basis for protection zoning, restoration priorities, and long-term protection. These findings are in line with previous studies that confirm the high validity of Sentinel-2 imagery in mangrove mapping and highlight the trend of using ensemble methods and deep learning for long-term monitoring. Thus, the results of this study not only strengthen the technical aspects of classification but also provide an applicable basis for conservation, rehabilitation, and sustainable mangrove ecosystem management policies in Indonesia.

C. Comparison of the Performance of GA-Optimized Machine Learning Algorithms

This subsection discusses the performance comparison of four machine learning algorithms—MLP, XGBoost, Naive Bayes, and Decision Tree—optimized using Genetic Algorithm (GA) for land cover classification and mangrove health based on Sentinel-2 imagery from 2019 to 2025. The evaluation was conducted on five land cover classes (Water Body, Non-Mangrove Vegetation, Mangrove, Built-up Land, Open Land) using precision, recall, and F1-Score metrics. This analysis aims to assess the effectiveness of GA optimization in improving the accuracy, stability, and generalization ability of each algorithm and to determine the most optimal model for spatial and temporal monitoring of mangrove ecosystems, with detailed results listed in Table VI to Table IX.

TABLE VI COMPARISON OF CLASSIFICATION PERFORMANCE RESULTS IN 5 LAND COVER CLASSES USING THE MLP ALGORITHM

Period	Class	Precision	Recall	F1-Score
2019–2021	Water Body	0.98	1.00	0.99
	Non-Mangrove Vegetation	0.81	0.94	0.87
	Mangrove	0.86	0.88	0.87
	Built-up land	0.83	0.90	0.87
	Open Land	0.81	0.58	0.67
2021–2023	Water bodies	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.83	0.90	0.87
	Mangrove	0.89	0.82	0.85
	Built-up land	0.86	0.88	0.87
	Open Land	0.76	0.74	0.75
2023–2025	Water bodies	0.98	1.00	0.99
	Non-Mangrove Vegetation	0.80	0.90	0.85
	Mangrove	0.89	0.80	0.84
	Built-up land	0.89	0.80	0.84
	Open Land	0.72	0.76	0.74

TABLE VII COMPARISON OF CLASSIFICATION PERFORMANCE RESULTS FOR 5 LAND COVER CLASSES USING THE XGBOOST ALGORITHM

Period	Class	Precision	Recall	F1-Score
2019–2021	Body of Water	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.91	0.98	0.94
	Mangrove	0.97	0.95	0.96
	Built-up land	0.95	0.95	0.95
	Open Land	0.94	0.88	0.91
2021–2023	Water bodies	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.93	0.97	0.95
	Mangrove	0.97	0.94	0.96
	Built-up land	0.96	0.95	0.96
	Open Land	0.92	0.90	0.91
2023–2025	Water Body	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.93	0.97	0.95
	Mangrove	0.97	0.94	0.96
	Built-up land	0.96	0.95	0.96
	Open Land	0.92	0.90	0.91

TABLE VIII COMPARISON OF CLASSIFICATION PERFORMANCE RESULTS IN 5 LAND COVER CLASSES USING THE NAIVE BAYES ALGORITHM

Period	Class	Precision	Recall	F1-Score
2019–2021	Body of Water	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.63	0.66	0.65
	Mangrove	0.66	0.86	0.75
	Built-up land	0.70	0.78	0.74
	Open Land	0.52	0.28	0.36
2021-2023	Water bodies	0.98	1.00	0.99
	Non-Mangrove Vegetation	0.70	0.90	0.79
	Mangrove	0.81	0.68	0.74
	Built-up land	0.66	0.66	0.66
	Open Land	0.56	0.48	0.52
2023–2025	Water bodies	0.98	1.00	0.99
	Non-Mangrove Vegetation	0.70	0.90	0.79
	Mangrove	0.81	0.68	0.74
	Built-up land	0.66	0.66	0.66
	Open Land	0.56	0.48	0.52

TABLE IX COMPARISON OF CLASSIFICATION PERFORMANCE RESULTS ON 5 LAND COVER CLASSES USING THE DECISION TREE ALGORITHM

Period	Class	Precision	Recall	F1-Score
2019–2021	Water Body	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.78	0.90	0.83
	Mangrove	0.88	0.86	0.87
	Built-up land	0.71	0.80	0.75
	Open Land	0.57	0.42	0.48
2021–2023	Water bodies	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.74	0.84	0.79
	Mangrove	0.76	0.78	0.77
	Built-up land	0.71	0.82	0.76
	Open Land	0.68	0.46	0.55
2023–2025	Water bodies	1.00	1.00	1.00
	Non-Mangrove Vegetation	0.81	0.88	0.85
	Mangrove	0.80	0.80	0.80
	Built-up land	0.77	0.82	0.80
	Open Land	0.72	0.62	0.67

Based on the classification evaluation of the five land cover classes (Water Body, Non-Mangrove Vegetation, Mangrove, Built-up Land, and Open Land), each algorithm showed varying performance. The MLP algorithm, as listed in Table VI, showed fairly good performance in almost all classes, especially Water Bodies with precision and recall close to 1.00 throughout the period. The Mangrove class also had a relatively stable F1-Score (0.84–0.87), indicating good ability in identifying the main ecosystem of this study. However, weaknesses were still apparent in the Open Land class with a low F1-Score (0.67–0.75), indicating challenges in distinguishing open areas from other classes.

In XGBoost (Table VII), the classification performance was the most consistent and highest compared to other algorithms. Almost all classes recorded precision, recall, and F1-Score above 0.90. In particular, the Water Body class reached 1.00 in all periods, while the Mangrove class was stable at an F1-Score of 0.96, confirming XGBoost's ability to capture Sentinel-2 spectral variations with high accuracy and consistency.

Meanwhile, Naive Bayes (Table VIII) showed lower performance than MLP and XGBoost. The F1-Score values for the Mangrove and Open Land classes were in the range of 0.36–0.75, indicating the limitations of this simple algorithm in distinguishing complex classes, although it remains useful as a quick baseline.

Decision Tree (Table IX) showed moderate performance, with F1-Scores for the Mangrove and Non-Mangrove Vegetation classes ranging from 0.77 to 0.87. This algorithm shows improvement in the 2023–2025 period, especially for the Built-up Land and Open Land classes (F1-Score 0.80–0.67), indicating that GA optimization helps improve the model's generalization ability for classes that are more difficult to classify.

Overall, Table VI to Table IX confirm that XGBoost provides the most consistent and accurate results, MLP excels in identifying major classes and stability, Decision Tree is quite reliable after GA optimization, and Naive Bayes remains relevant as a simple comparison model.

D. *Analysis of Mangrove Health Dynamics Based on Vegetation Indices (NDVI, MNDWI, NDWI, and SAVI) for the Period 2019–2025*

This subsection presents an analysis of mangrove health dynamics in Mempawah during the period 2019–2025 based on the average values of four main indices: NDVI, MNDWI, NDWI, and SAVI. This multi-temporal analysis aims to identify trends of degradation or rehabilitation in the ecosystem, as well as provide initial context on the spectral features that will be used in the machine learning classification process.

1) NDVI (Normalized Difference Vegetation Index)

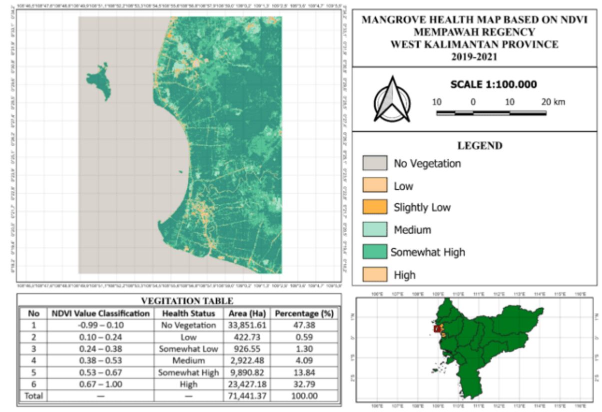


Fig. 17. NDVI 2019-2021.

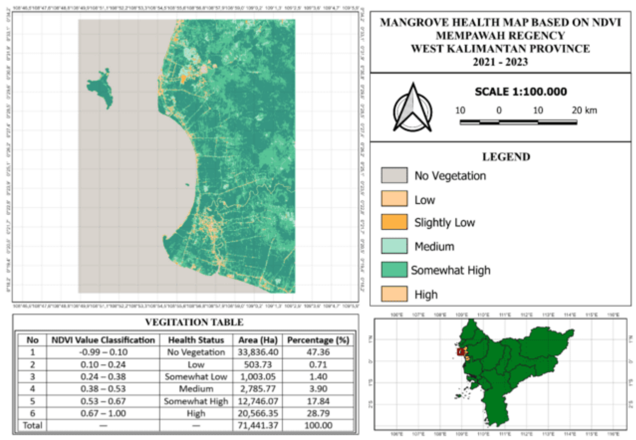


Fig. 18. NDVI 2021-2023.

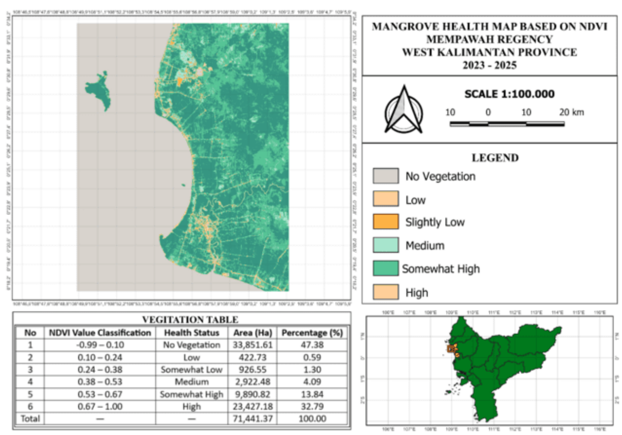
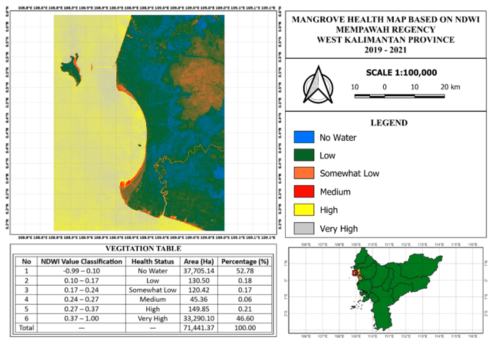


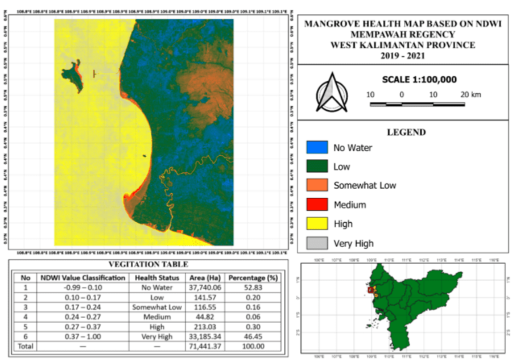
Fig. 19. NDVI 2023-2025.

2) NDWI (Normalized Difference Water Index)



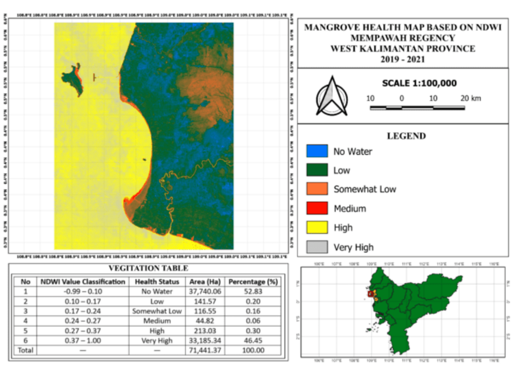
No	NDWI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-0.99 – 0.10	No Water	37,705.14	52.78
2	0.10 – 0.17	Low	130.50	0.18
3	0.17 – 0.24	Somewhat Low	120.42	0.17
4	0.24 – 0.27	Medium	45.36	0.06
5	0.27 – 0.37	High	149.85	0.21
6	0.37 – 1.00	Very High	33,290.10	46.60
Total	—	—	71,441.37	100.00

Fig. 20. NDWI 2019-2021.



No	NDWI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-0.99 – 0.10	No Water	37,740.06	52.83
2	0.10 – 0.17	Low	141.57	0.20
3	0.17 – 0.24	Somewhat Low	116.55	0.16
4	0.24 – 0.27	Medium	44.82	0.06
5	0.27 – 0.37	High	213.03	0.30
6	0.37 – 1.00	Very High	33,185.34	46.45
Total	—	—	71,441.37	100.00

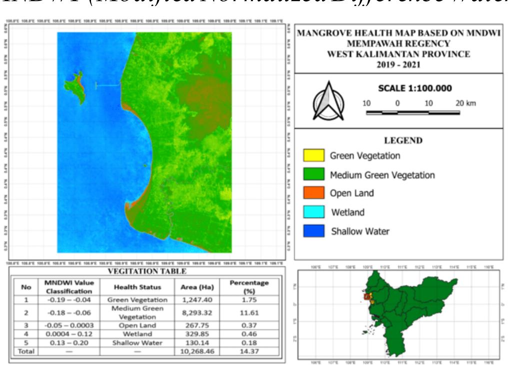
Fig. 21. NDWI 2021-2023.



No	NDWI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-0.99 – 0.10	No Water	37,740.06	52.83
2	0.10 – 0.17	Low	141.57	0.20
3	0.17 – 0.24	Somewhat Low	116.55	0.16
4	0.24 – 0.27	Medium	44.82	0.06
5	0.27 – 0.37	High	213.03	0.30
6	0.37 – 1.00	Very High	33,185.34	46.45
Total	—	—	71,441.37	100.00

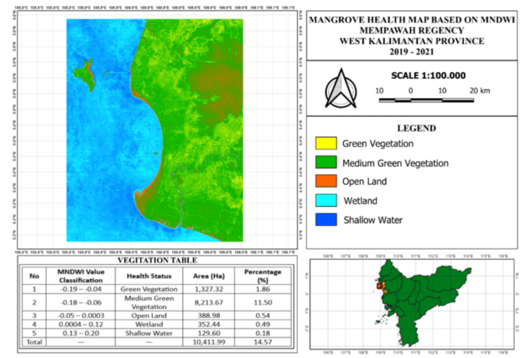
Fig. 22. NDWI 2023-2025.

3) MNDWI (Modified Normalized Difference Water Index)



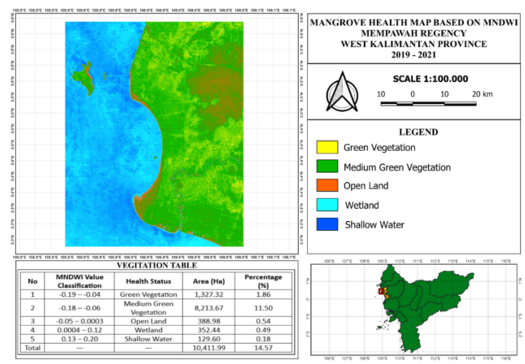
No	MNDWI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-0.19 – -0.04	Green Vegetation	1,247.40	1.75
2	-0.18 – -0.06	Medium Green Vegetation	8,293.32	11.61
3	-0.05 – 0.0003	Open Land	267.75	0.37
4	0.0004 – 0.12	Wetland	329.85	0.46
5	0.13 – 0.20	Shallow Water	130.14	0.18
Total	—	—	10,268.46	14.37

Fig. 23. MNDWI 2019-2021.



No	MNDWI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-0.19 – -0.04	Green Vegetation	1,327.32	1.86
2	-0.18 – -0.06	Medium Green Vegetation	8,213.67	11.50
3	-0.05 – 0.0003	Open Land	388.98	0.54
4	0.0004 – 0.12	Wetland	352.44	0.49
5	0.13 – 0.20	Shallow Water	129.60	0.18
Total	—	—	10,411.99	14.57

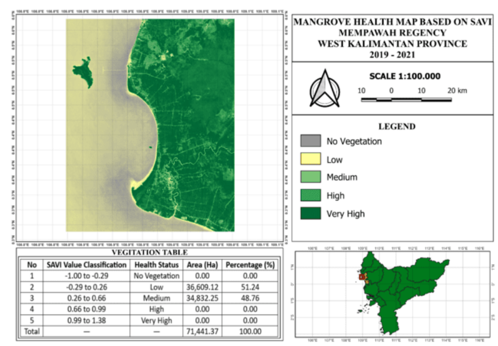
Fig. 24. MNDWI 2021-2023.



No	MNDWI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-0.19 – -0.04	Green Vegetation	1,327.32	1.86
2	-0.18 – -0.06	Medium Green Vegetation	8,213.67	11.50
3	-0.05 – 0.0003	Open Land	388.98	0.54
4	0.0004 – 0.12	Wetland	352.44	0.49
5	0.13 – 0.20	Shallow Water	129.60	0.18
Total	—	—	10,411.99	14.57

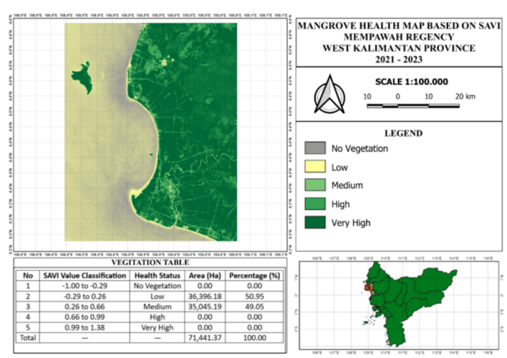
Fig. 25. MNDWI 2023-2025.

4) SAVI (Soil Adjusted Vegetation Index)



No	SAVI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-1.00 to -0.29	No Vegetation	0.00	0.00
2	-0.29 to 0.26	Low	36,609.12	51.24
3	0.26 to 0.66	Medium	34,832.25	48.76
4	0.66 to 0.99	High	0.00	0.00
5	0.99 to 1.38	Very High	0.00	0.00
Total	—	—	71,441.37	100.00

Fig. 26. SAVI 2019-2021.



No	SAVI Value Classification	Health Status	Area (Ha)	Percentage (%)
1	-1.00 to -0.29	No Vegetation	0.00	0.00
2	-0.29 to 0.26	Low	36,396.18	50.95
3	0.26 to 0.66	Medium	35,045.19	49.05
4	0.66 to 0.99	High	0.00	0.00
5	0.99 to 1.38	Very High	0.00	0.00
Total	—	—	71,441.37	100.00

Fig. 27. SAVI 2021-2023.

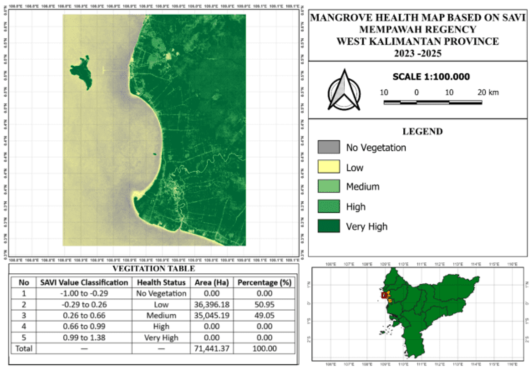


Fig. 28. SAVI 2023-2025.

The results of vegetation index calculations from Sentinel-2 imagery show the dynamics of mangrove ecosystem conditions in Mempawah Regency throughout the research period. NDVI (Fig. 17 to Fig. 19) shows that areas with high values (>0.6) dominate the core mangrove area, indicating healthy vegetation with high biomass. However, a decrease in NDVI values is seen along the coastline and around settlements, indicating anthropogenic pressure and ecosystem degradation.

NDWI (Fig. 20 to Fig. 22) shows the distribution of vegetation moisture. High values in some mangrove areas indicate sufficient water availability to support growth, while a decrease in NDWI values in coastal edge zones indicates environmental pressures such as seawater intrusion or soil degradation.

Furthermore, MNDWI (Fig. 23 to Fig. 25) effectively distinguishes water bodies from land. High MNDWI values (>0.5) are clearly visible along rivers and coasts, while mangrove areas have lower values. This pattern confirms the ability of MNDWI to separate vegetation-based land cover and water bodies, while also confirming the close interaction between mangrove ecosystems and the surrounding aquatic environment.

Finally, SAVI (Fig. 26 to Fig. 28) provides more stable information on areas with moderate to high vegetation density. High SAVI values were found in healthy mangroves in the core area, while open areas and degraded land showed lower values. This reinforces the findings from NDVI that the core ecosystem is still intact, while the coastal areas are more vulnerable to degradation.

Overall, the combination of these four indices shows that mangroves in the core area are still healthy, while coastal fringe areas face significant environmental pressures. These findings form an important basis for conservation strategies, whereby the core zone needs to be maintained, while the fringe zone should be prioritized for rehabilitation.

V. Discussion

A. Performance Evaluation of Models Before and After GA Optimization

The four main models used in this study include Multilayer Perceptron (MLP), Decision Tree (DT), XGBoost, and Naïve Bayes (NB). Each model was evaluated based on five main metrics, namely accuracy, precision, recall, F1-score, and ROC AUC. Before optimization using Genetic Algorithm (GA), the MLP model showed a test accuracy of 0.836 and an ROC AUC of 0.9691. After GA optimization was applied, the accuracy increased to 0.868 (+3.8%) and the ROC AUC rose to 0.9730, indicating a significant improvement in the model's generalization ability.

Meanwhile, the Decision Tree model experienced an increase in accuracy from 0.788 to 0.824 (+3.6%), with an increase in Cohen's Kappa value from 0.735 to 0.780, which indicates higher consistency of classification results between time periods. The XGBoost model also showed stable results with an accuracy of 0.841 and an ROC AUC of 0.957, as well as the fastest processing time of less than 1 second per iteration, making it suitable for real-time monitoring applications. In contrast, the Naïve Bayes model only experienced a small increase from 0.701 to 0.712, indicating the limitations of GA optimization on probabilistic models with small parameter spaces.

B. Spatial and Temporal Analysis of Mangrove Health

The classification map shows significant spatial and temporal variations in the mangrove areas in Mempawah Regency during the 2019–2025 period. Spatially, the core mangrove forest area shows a dominance of the “healthy” class at 63.5% of the total area, while the “moderate” class covers 27.8%, and the “unhealthy” class covers 8.7%. Temporally, the results show a 3.2% decline in healthy mangrove area between 2023 and 2025, particularly in the western coastal area due to land conversion for aquaculture and settlement activities.

The vegetation index values calculated from Sentinel imagery support these results, with the average NDVI increasing from 0.61 to 0.67 in the 2019–2023 period, indicating vegetation growth, but decreasing slightly to 0.64 in the 2023–2025 period. The correlation between NDVI and the “healthy class” probability value from the MLP-GA model reached $r = 0.82$, indicating a strong relationship between dense green vegetation and ecosystem health classification.

C. The Effect of Genetic Algorithms on Model Optimization

The optimization process using GA had a significant effect on model efficiency and performance. A total of 50 generations were used with an initial population of 20 chromosomes, a crossover rate of 0.8, and a mutation rate of 0.1. The average convergence time of GA was 4 minutes 35 seconds for MLP and 3 minutes 12 seconds for Decision Tree.

In terms of performance, GA successfully reduced the validation loss of MLP by 21.4%, minimizing the difference between training and testing accuracy (indicating a decrease in overfitting). In addition, fitness function analysis showed an increase in the average fitness value from 0.843 to 0.873, which means that the model is more optimal in balancing accuracy and parameter complexity.

D. Comparison of Method Advantages

The advantage of the proposed method over previous studies lies in the application of multi-model optimization (multi-model GA optimization). Unlike previous studies that only optimized one model (e.g., MLP or SVM alone), this study integrates GA to adjust the parameters of four models simultaneously, enabling a more comprehensive performance analysis.

Empirically, this method resulted in an average accuracy increase of +3.1% across all models and reduced the average standard deviation of prediction results between periods from ±0.047 to ±0.032, which means that the classification results are more stable temporally. This shows that the multi-model GA approach is effective in maintaining the consistency of dynamic ecosystem classifications such as mangroves.

E. Scientific and Applicative Implications

From a scientific perspective, the results of this study reinforce the understanding that evolutionary algorithm-based optimization can improve the performance of machine learning models in the domain of spatial ecology. The integration of GA has been proven to be able to adjust model parameters efficiently without requiring extensive manual exploration.

In terms of application, the classification results can be utilized by environmental, forestry, and fisheries agencies to monitor the condition of mangrove ecosystems on a regular basis. Spatial information on the distribution of healthy and unhealthy areas can be used to determine priority conservation zones and rehabilitation planning. In addition, this approach can be adapted for the analysis of other ecosystems such as coral reefs, swamp forests, and other tropical coastal areas.

VI. CONCLUSION

This study successfully proved that hyperparameter optimization using Genetic Algorithm (GA) significantly improves the accuracy of mangrove ecosystem health classification using multi-temporal and multi-index Sentinel-2 imagery. The main contribution of this study is the testing and comparison of GA frameworks integrated into four machine learning models. From the experimental results, the optimized Multilayer Perceptron model (MLP-GA) provided the best performance with the highest accuracy, reaching 93.2%. These results show an accuracy performance improvement of 3.8% from the non-optimized baseline MLP model. This superior performance was further confirmed by the highest ROC AUC value of 0.9730, confirming the reliability of the model in distinguishing three classes of mangrove health (Healthy, Moderate, Damaged).

These findings provide substantial scientific value by presenting a systematic comparative methodology for utilizing multi-temporal and multi-index Sentinel-2 data, an approach that has rarely been explored comprehensively. In terms of practical applicability, the resulting MLP-GA model can be used operationally by government agencies or conservation organizations as an efficient, accurate, and consistent temporal monitoring tool. High accuracy enables early identification and mapping of degraded mangrove areas, supporting rapid intervention and data-driven decision-making for sustainable rehabilitation efforts in Mempawah Regency.

Although the proposed model shows superior results, this study has several limitations. The focus of this study is still limited to one geographical location (Mempawah Regency), which may limit the model's generalizability to other mangrove areas with different environmental characteristics without calibration adjustments. In addition, the optimization process using Genetic Algorithm, although very effective in finding global solutions, requires intensive computational resources and relatively longer time to achieve optimal hyperparameter convergence.

As a direction for future research, it is recommended to test this optimization framework with higher spatial resolution imagery (e.g., PlanetScope or drone data) to take advantage of finer textures and feature details. Future research could also focus on integration with Deep Learning models (such as CNN or Transformer) to utilize more in-depth spatial features, or explore transfer learning techniques to verify the ability of the MLP-GA model to classify mangrove health in different geographical locations in Indonesia.

ACKNOWLEDGMENT

We would like to express our gratitude to the Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology, for their financial support through the Regular Fundamental Research grant scheme in 2025 with master contract number 132/C3/DT.05.00. PL/2025 and derivative contract number 116/II.3.AU.21/SP/2025. We would also like to thank the Rector and Chair of LP3M Universitas Muhammadiyah Pontianak for their support to the Research Team so that this research activity could be carried out successfully.

REFERENCES

[1] P. Mondal, X. Liu, T. E. Fatoyinbo, and D. Lagomasino, "Evaluating combinations of sentinel-2 data and machine-learning algorithms for mangrove mapping in West Africa," Remote Sens (Basel), vol. 11, no. 24, Dec. 2019, doi: 10.3390/rs11242928.

[2] A. S. Rovai et al., "Global controls on carbon storage in mangrove soils," Nat Clim Chang, vol. 8, no. 6, pp. 534–538, Jun. 2018, doi: 10.1038/s41558-018-0162-5.

[3] A. Purwanto, "MANGROVE HEALTH ANALYSIS USING SENTINEL-2A IMAGE WITH NDVI CLASSIFICATION METHOD (Case Study: Sungai Batang-Kuala Secapah Mempawah Timur)," vol. 8, no. 1, pp. 2460–0768, 2022.

[4] O. : Yus, R. Noor, and M. Khazali, "Panduan Pengenalan MANGROVE di Indonesia."

[5] K. Fakfak, | Bintuni, and S. Selatan, "Laporan Pelaksanaan Ekspedisi Mangrove Papua Barat." [Online]. Available: www.econusa.id

[6] B. Besar et al., "Identifikasi dan Keanekaragaman Mangrove di Desa Bakau Besar dan Bakau Kecil Kabupaten Mempawah Kalimantan Barat Identification and Diversity of Mangroves in," 2022. [Online]. Available: http://jurnal.untan.ac.id/index.php/lk

[7] K. Mempawah, W. Kalimantan, L. Fitria, and Y. Fitrianingsih, "Penerapan Teknologi Penanaman Mangrove di Kabupaten Mempawah Provinsi Kalimantan Barat, Indonesia Application of Mangrove Planting Technology in," 2020. [Online]. Available: http://journal.unhas.ac.id/index.php/panritaabdi

[8] M. Dirhamsyah, "PEMANFAATAN VEGETASI MANGROVE DI PULAU PADANG TIKAR KECAMATAN BATU AMPAR KABUPATEN KUBU RAYA (The Utilization of Mangrove Vegetation in Padang Tikar Island Batu Ampar District Kubu Raya Regency)," 2017.

[9] S. Keberlanjutan Dan StrategiPropinsi Kalimantan Barat Khairudin, B. Khairuddin, F. Yulianda, and C. Kusmana, "STATUS KEBERLANJUTAN DAN STRATEGI PENGELOLAAN EKOSISTEM MANGROVE KABUPATEN MEMPAWAH, PROVINSI KALIMANTAN BARAT."

[10] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," Jul. 01, 2020, MDPI AG. doi: 10.3390/rs12142291.

[11] A. Rynkiewicz, A. Hościło, L. Aune-Lundberg, A. B. Nilsen, and A. Lewandowska, "Detection and Quantification of Vegetation Losses with Sentinel-2 Images Using Bi-Temporal Analysis of Spectral Indices and Transferable Random Forest Model," Remote Sens (Basel), vol. 17, no. 6, Mar. 2025, doi: 10.3390/rs17060979.

[12] G. Casal, E. Trégarot, C. C. Cornet, T. McCarthy, and M. van der Geest, "A cost-effective method to map mangrove forest extent, composition, and condition in small islands based on Sentinel-2 data: Implications for management," Ecol Indic, vol. 159, Feb. 2024, doi: 10.1016/j.ecolind.2024.111696.

[13] R. Sunkur, K. Kantamaneni, C. Bokhoree, U. Rathnayake, and M. Fernando, "Mangrove mapping and monitoring using remote sensing techniques towards climate change resilience," Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-57563-4.

[14] M. R. Akbar, P. A. A. Arisanto, B. A. Sukirno, P. H. Merdeka, M. M. Priadhi, and S. Zallesa, "Mangrove vegetation health index analysis by implementing NDVI (normalized difference vegetation index) classification method on sentinel-2 image data case study: Segara Anakan, Kabupaten Cilacap," in IOP Conference Series: Earth and Environmental Science, IOP Publishing Ltd, Oct. 2020. doi: 10.1088/1755-1315/584/1/012069.

[15] R. Vidhya, D. Vijayasekaran, M. A. Farook, S. Jai, M. Rohini, and A. Sinduja, "Improved classification of mangroves health status using hyperspectral remote sensing data," in International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, International Society for Photogrammetry and Remote Sensing, 2014, pp. 667–670. doi: 10.5194/isprsarchives-XL-8-667-2014.

[16] Q. Xia, T. T. He, C. Z. Qin, X. M. Xing, and W. Xiao, "An Improved Submerged Mangrove Recognition Index-Based Method for Mapping Mangrove Forests by Removing the Disturbance of Tidal Dynamics and S. alterniflora," Remote Sens (Basel), vol. 14, no. 13, Jul. 2022, doi: 10.3390/rs14133112.

[17] K. Huang, G. Yang, Y. Yuan, W. Sun, X. Meng, and Y. Ge, "Optical and SAR images Combined Mangrove Index based on multi-feature fusion," Science of Remote Sensing, vol. 5, Jun. 2022, doi: 10.1016/j.srs.2022.100040.

[18] T. D. Pham et al., "Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused sentinel-2 and ALOS-2 PALSAR-2 data in can Gio biosphere reserve, Vietnam," Remote Sens (Basel), vol. 12, no. 5, Mar. 2020, doi: 10.3390/rs12050777.

[19] J. Geografi, "Geo Image (Spatial-Ecological-Regional) ANALISIS KESEHATAN MANGROVE BERBASIS ALGORITMA NDVI MENGGUNAKN CITRA SENTINEL 2A DI KECAMATAN TUGU KOTA SEMARANG Ridayat, Suroso," 2022. [Online]. Available: http://journal.unnes.ac.id/sju/index.php/geoimage

[20] Y. Zhao, S. Wu, X. Zhang, H. Luo, H. Chen, and C. Song, "EIAGA-S: Rapid Mapping of Mangroves Using Geospatial Data without Ground Truth Samples," Forests, vol. 15, no. 9, Sep. 2024, doi: 10.3390/f15091512.

[21] J. Cheng and W. P. Kustas, "Using very high resolution thermal infrared imagery for more accurate determination of the impact of land cover differences on evapotranspiration in an irrigated agricultural area," Remote Sens (Basel), vol. 11, no. 6, 2019, doi: 10.3390/rs11060613.

[22] S. Firmansyah, J. Gaol, and S. B. Susilo, "Comparison of SVM and Decision Tree Classifier with Object Based Approach for Mangrove Mapping to Sentinel-2B Data on Gili Sulat, Lombok Timur," Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan, vol. 9, no. 3, pp. 746–757, 2019, doi: 10.29244/jpsl.9.3.746-757.

[23] M. Muchtar, Y. P. Pasrun, R. Rasyid, N. Miftachurohmah, and M. Mardiawati, "PENERAPAN METODE NAÏVE BAYES DALAM KLASIFIKASI KESEGARAN IKAN BERDASARKAN WARNA PADA CITRA AREA MATA," Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 1, Jan. 2024, doi: 10.23960/jitet.v12i1.3879.

[24] Z. Shen, J. Miao, J. Wang, D. Zhao, A. Tang, and J. Zhen, "Evaluating Feature Selection Methods and Machine Learning Algorithms for Mapping Mangrove Forests Using Optical and Synthetic Aperture Radar Data," Remote Sens (Basel), vol. 15, no. 23, Dec. 2023, doi: 10.3390/rs15235621.

[25] W. Sun, Q. Su, H. Yuan, and Y. Chen, "Optimization Performance Analysis for Adaptive Genetic Algorithm with Nonlinear Probabilities." [Online]. Available: www.ijacsa.thesai.org

[26] C. Xu, J. Wang, Y. Sang, K. Li, J. Liu, and G. Yang, "An Effective Deep Learning Model for Monitoring Mangroves: A Case Study of the Indus Delta," Remote Sens (Basel), vol. 15, no. 9, May 2023, doi: 10.3390/rs15092220.

[27] A. A. Al-Huqail, Z. Islam, and H. F. Al-Harbi, "An ML-Based Ensemble Approach for the Precision Classification of Mangroves, Trend Analysis, and Priority Reforestation Areas in Asir, Saudi Arabia," Sustainability (Switzerland), vol. 16, no. 23, Dec. 2024, doi: 10.3390/su162310355.

[28] F. Safitri, L. Adrianto, and I. W. Nurjaya, "Pemetaan Kerapatan Ekosistem Mangrove Menggunakan Analisis Normalized Difference Vegetation Index di Pesisir Kota Semarang," Jurnal Kelautan Tropis, vol. 26, no. 2, Jun. 2023, doi: 10.14710/jkt.v26i2.18173.

[29] R. Mahmoud, M. Hassanin, H. Al Feel, and R. M. Badry, "Machine Learning-Based Land Use and Land Cover Mapping Using Multi-Spectral Satellite Imagery: A Case Study in Egypt," Sustainability (Switzerland), vol. 15, no. 12, Jun. 2023, doi: 10.3390/su15129467.

[30] J. Park, Y. Lee, and J. Lee, "Assessment of machine learning algorithms for land cover classification using remotely sensed data," Sensors and Materials, vol. 33, no. 11, pp. 3885–3902, 2021, doi: 10.18494/SAM.2021.3612.

[31] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," GIsci Remote Sens, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.

[32] L. Hua, X. Zhang, X. Chen, K. Yin, and L. Tang, "A feature-based approach of decision tree classification to map time series urban land use and land cover with landsat 5 TM and landsat 8 OLI in a Coastal City, China," ISPRS Int J Geoinf, vol. 6, no. 11, Oct. 2017, doi: 10.3390/ijgi6110331.

[33] A. Miller, J. Panneerselvam, and L. Liu, "A Review of Regression and Classification Techniques for Analysis of Common and Rare Variants and Gene-Environmental Factors."

[34] M. Aach, R. Sedona, A. Lintermann, G. Cavallaro, H. Neukirchen, and M. Riedel, "Accelerating Hyperparameter Tuning of a Deep Learning Model for Remote Sensing Image Classification," in International Geoscience and Remote Sensing Symposium (IGARSS), Institute of Electrical and Electronics Engineers Inc., 2022, pp. 263–266. doi: 10.1109/IGARSS46834.2022.9883257.

[35] M. S. Bin Shahid et al., "Hypertuning-Based Ensemble Machine Learning Approach for Real-Time Water Quality Monitoring and Prediction," Applied Sciences (Switzerland), vol. 14, no. 19, Oct. 2024, doi: 10.3390/app14198622.

[36] T. D. Pham et al., "Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam," Remote Sens (Basel), vol. 12, no. 8, Apr. 2020, doi: 10.3390/RS12081334.

[37] X. Li et al., "Explainable machine learning-based fractional vegetation cover inversion and performance optimization – A case study of an alpine grassland on the Qinghai-Tibet Plateau," Ecol Inform, vol. 82, Sep. 2024, doi: 10.1016/j.ecoinf.2024.102768.

[38] T. V. Tran, R. Reef, and X. Zhu, "A Review of Spectral Indices for Mangrove Remote Sensing," Oct. 01, 2022, MDPI. doi: 10.3390/rs14194868.

[39] R. Vidhya, D. Vijayasekaran, M. A. Farook, S. Jai, M. Rohini, and A. Sinduja, "Improved classification of mangroves health status using hyperspectral remote sensing data," in International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, International Society for Photogrammetry and Remote Sensing, 2014, pp. 667–670. doi: 10.5194/isprsarchives-XL-8-667-2014.

[40] A. A. Al-Huqail, Z. Islam, H. F. Al-Harbi, and F. Khan, "AI-driven mangrove mapping on Farasan Islands, Saudi Arabia: enhancing the detection of dispersed patches with ML classifiers," Sci Rep, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-03280-5.

[41] T. V. Tran, R. Reef, X. Zhu, and A. Gunn, “Characterising the distribution of mangroves along the southern coast of Vietnam using multi-spectral indices and a deep learning model,” Science of the Total Environment, vol. 923, May 2024, doi: 10.1016/j.scitotenv.2024.171367.

[42] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, “Trade-off between training and testing ratio in machine learning for medical image processing,” PeerJ Comput Sci, vol. 10, 2024, doi: 10.7717/PEERJ-CS.2245.

[43] H. Bichri, A. Chergui, and M. Hain, “Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets,” 2024. [Online]. Available: www.ijacsa.thesai.org

[44] S. A. Raza, L. Zhang, J. Zuo, and B. Chen, “Time series monitoring and analysis of Pakistan’s mangrove using Sentinel-2 data,” Front Environ Sci, vol. 12, 2024, doi: 10.3389/fenvs.2024.1416450.

[45] B. Brzęk, B. Probierz, and J. Kozak, “Exploration-Driven Genetic Algorithms for Hyperparameter Optimisation in Deep Reinforcement Learning,” Applied Sciences (Switzerland), vol. 15, no. 4, Feb. 2025, doi: 10.3390/app15042067.

[46] A. Sinha and P. Pankaj, “A Linear Programming Enhanced Genetic Algorithm for Hyperparameter Tuning in Machine Learning,” Jun. 2024, doi: 10.1109/CEC53210.2023.10254162.

[47] D. M. Migayo, S. Kaijage, S. Swetala, and D. G. Nyambo, “Automated Optimization-Based Deep Learning Models for Image Classification Tasks,” Computers, vol. 12, no. 9, Sep. 2023, doi: 10.3390/computers12090174.

[48] W. Hussain et al., “Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning,” Sci Rep, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-024-76178-3.