e-JURNAL PENYELIDIKAN DAN INOVASI
ISLAMIC INNOVATION CENTER FOR TECHNOLOGY, UIS
https://ejpi.uis.edu.my
e-ISSN 2289-7909

VOL. 12 NO. 5, SPECIAL EDITION, (DECEMBER, 2025): page 272-283
DOI: 10.53840/ejpi.v12i5.323

Received: October 27, 2025 | Accepted: November 26, 2025 | Published: December 31, 2025

Analysis of Mangrove Health Predictions Using the Random Forest Regression Method

Analisis Prediksi Kesehatan Mangrove Menggunakan Metode Random Forest Regression

Nurheni Dewi Pranita[1], Erwin Hermawan[2], Sahid Agustian Hudjimartsu [3]
Nurdin Sulistiyono[4]
[1,2,3] Informatics Engineering, Faculty of Technology & Science, Universitas Ibn Khaldun Bogor
Jl. Sholeh Iskandar, RT.01/RW.10, Kedungbadak, Kec. Tanah Sereal, Bogor City, West Java 16162 Indonesia.
[4] Department of Forest Conservation, Faculty of Forestry, Universitas Sumatera Utara, Medan Indonesia.
Email: [1]nurhenidewipranita@gmail.com, [2]rwien.gfm@gmail.com, [3]shudjimartsu@uika-bogor.ac.id, [4]nurdinsulistyono@usu.ac.id

ABSTRACT

Mangrove forests are important ecosystems in coastal areas that play a role in maintaining environmental balance and protecting the coast from erosion and abrasion. However, these ecosystems have experienced a decline in quality and quantity due to human activities, such as land conversion for industry and plantations. This observe aims to are expecting the health circumstance of mangroves inside the Belawan Sicanang region, Medan metropolis, North Sumatra, the usage of the Random woodland Regression (RFR) approach. The records utilized include multispectral imagery incorporating the NDVI, NDRE, and ARVI flora indices, as well as Mangrove health Index (MHI) values. The analysis process was carried out through several stages, namely processing drone image data, oversampling using the SMOTE method, and applying RFR models to predict mangrove health. The prediction results showed that 68.9% of the area was classified as healthy, 22.3% moderate, and 8.8% unhealthy, with good model accuracy indicated by an R-squared (R^2) value of 0.757 and a Root Mean Squared Error (RMSE) of 8.515. This take a look at demonstrates that the Random woodland Regression method is effective in predicting and mapping mangrove fitness conditions with an excessive degree of accuracy. The recommendation of this study is to increase the sample during field data collection so that the research results can be more accurate and good.

Keywords: Mangrove Health Index; Mangrove Health; Random Forest Regression

ABSTRAK

Hutan bakau merupakan ekosistem penting di kawasan pantai yang berperanan dalam mengekalkan keseimbangan alam sekitar dan melindungi pantai daripada hakisan dan lelasan. Walaubagaimanapun, ekosistem ini telah mengalami kemerosotan dalam kualiti dan kuantiti akibat aktiviti manusia, seperti

penukaran tanah untuk industri dan perladangan. Pemerhatian ini bertujuan untuk menjangkakan keadaan kesihatan bakau di wilayah Belawan Sicanang, metropolis Medan, Sumatera Utara, penggunaan pendekatan Regresi Hutan Rawak (RFR). Rekod yang digunakan termasuk imejan berbilang spektrum yang menggabungkan indeks flora NDVI, NDRE dan ARVI, serta nilai Indeks kesihatan Bakau (MHI). Proses analisis dijalankan melalui beberapa peringkat iaitu memproses data imej dron, pensampelan berlebihan menggunakan kaedah SMOTE, dan mengaplikasi model RFR untuk meramal kesihatan bakau. Keputusan ramalan menunjukkan bahawa 68.9% kawasan dikelaskan sebagai sihat, 22.3% sederhana, dan 8.8% tidak sihat, dengan ketepatan model yang baik ditunjukkan oleh nilai R-squared (R^2) 0.757 dan Root Mean Squared Error (RMSE) sebanyak 8.515. Semakan ini menunjukkan bahawa kaedah Regresi hutan rawak berkesan dalam meramal dan memetakan keadaan kecergasan bakau dengan tahap ketepatan yang berlebihan. Cadangan kajian ini adalah menambah sampel semasa pengumpulan data lapangan supaya hasil kajian lebih tepat dan baik.

Kata kunci: Indeks Kesihatan Bakau; Kesihatan Bakau; Regresi Hutan Rawak

INTRODUCTION

Mangrove forests are transitional ecosystems between land and sea that are widely found in coastal areas. Its existence is influenced by various land and sea factors, making it one of the ecosystems that are very important for the balance of the environment (Suyadi *et al*., 2021). The health status of these ecosystems is often assessed through canopy density and the number of individuals per hectare, which are grouped into three categories: dense, moderate, and sparse. The solid and medium category is considered to be in good condition, while the category is rarely included in damaged condition (Suyuti, 2019).

Indonesia has one of the largest mangrove distributions in the world, supported by its long coastline and geographical position in the tropics (Doni Nurdiansah & I Wayan Eka Dharmawan, 2021). One of the areas with great mangrove potential is Belawan Sicanang Village, Medan City, North Sumatra, with an area of ±895.24 ha. However, the area has experienced a decline in quality due to land conversion for industry, settlements, ponds, and oil palm plantations. This condition encourages the need for deeper identification of mangrove health (Aspan et al., 2023).

Remote sensing technology provides a solution to efficiently monitor the health status of mangroves (Minh et al., 2022). This method allows the analysis of vegetation through indices such as NDVI, NDRE, and ARVI, which are integrated with Mangrove Health Index (MHI) to describe vegetation conditions (Singgalen, 2023). In this context, the algorithm Random Forest Regression (RFR) is a promising method due to its ability to handle multidimensional data and provide high accuracy (Tridawati et al., 2023).

Previous studies has proven that Random forest Regression is able to generating correct predictions across a huge range of cases. As an example, established that spectral modifications inside the crimson-part band can improve the accuracy of mangrove forest LAI predictions. (Zhu et al., 2017). Used RFR to estimate plant biomass with vegetation index and texture metrics from UAV imagery (Liu et al., 2019). In addition, successfully leveraged RFR to predict daily cases and death rates due to COVID-19 with high accuracy (Özen, 2024).

Based on the problems that have been described, the purpose of the study is to predict the distribution of mangrove health and evaluate its accuracy using the Random Forest Regression method based on the value of the Mangrove Health Index (MHI). This research is expected to provide important information for the management of mangrove ecosystems in the Belawan Sicanang area, Medan City, North Sumatra.

PROBLEM STATEMENT

Based on the introduction, the formulation of the problems in this study is:

1. How to get a prediction of the distribution of mangrove health through Random Forest Regression?
2. How to find out the accuracy test for Mangrove Health using the Random Forest Regression Method?

LITERATURE REVIEW

Table 1: Literature review

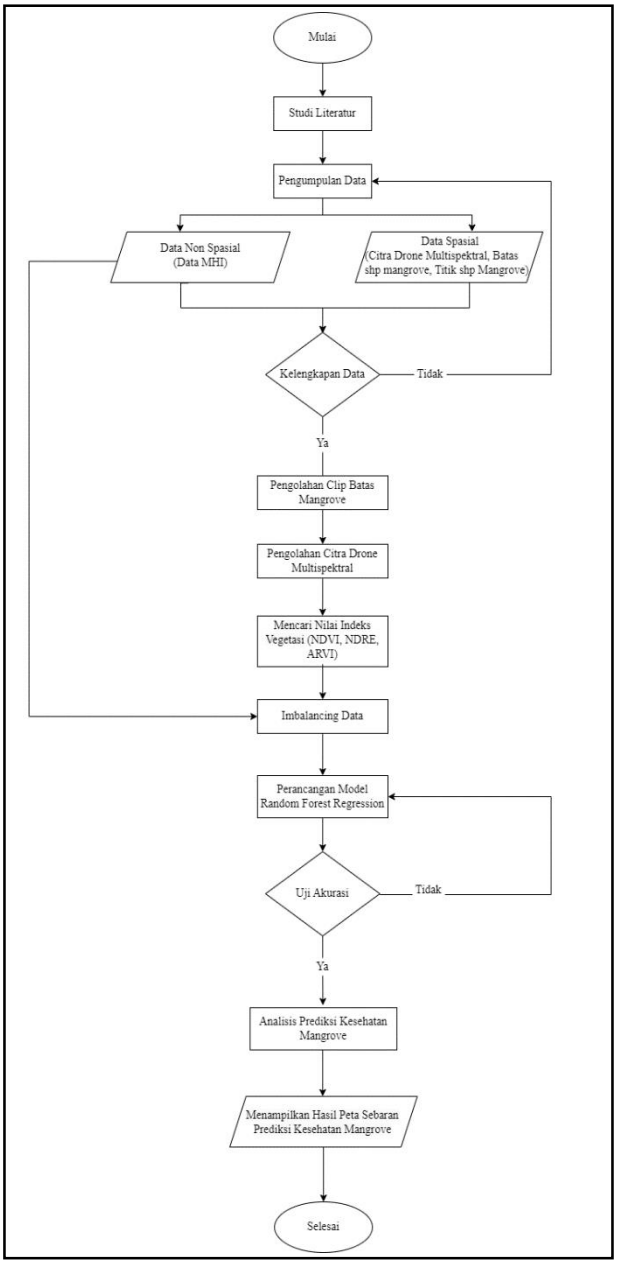
No	Author Name, Journal Title, and Year	Intention	Conclusion
1	Nurdiansah & Dharmawan, 2021 "Community Structure and Mangrove Health Conditions on Middleburg Island - Miossu, West Papua".	Combination application Remote sensing analysis with analysis mangrove community structure and MHI able to give a better picture About the health of the mangrove community on Middleburg-Miossu Island. Just less of the 5% of mangroves on Middleburg Island Miossu that have health conditions poor community.	Discussing Mangrove Health with the Mangrove Health Index (MHI).
2	H. Aspan, 2023 "Socialization of the Development of Mangrove Forest Tourism in Belawan Sicanang by Implementing Information Systems in an Effort to Improve the Community's Economy".	Belawan Sicanang is one of the villages in the Medan City area that has a fairly large mangrove ecosystem, which is ± 895.24 ha.	Explaining the area of Mangrove Forest in Belawan Sicanang.
3	Xie et al., 2021 "Comparison Of Random Forest and Multiple Linear Regression Models For Estimation Of Soil Extracellular Enzyme Activities In Agricultural Reclaimed Coastal Saline Land".	The results showed that the RF model is a more robust modeling approach than the MLR model estimating amylase and urease activity because of its superiority in dealing with nonlinear and hierarchical relationships between extracellular enzyme activity soils and environmental variables, which provide lower error indices (MAE and RMSE) and high R2 values.	Explains that the RF model is a more robust modeling to estimate amylase and urease activity.
4	Liu et al., 2019 "Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images"	Describing the Estimating winter oilseed radish biomass using vegetation index and texture metrics derived from UAV multispectral images, the accuracy of the AGB predicted by the RF regression model using Variables and texture metrics (RMSE = 274.18 kg/ha for the validation dataset) was slightly higher than the results of the PLSR model (RMSE = 284.09 kg/ha for the validation dataset.	Discuss the performance and texture metrics derived from multispectral images of UAVs in estimating AGB using the Random Forest method.

No	Author Name, Journal Title, and Year	Intention	Conclusion
5	Ozen, 2024 "Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey"	When random forest regression was used, testing data associated with daily cases was predicted with an accuracy of 92.30% and with an R2 of 0.9893. In addition, daily mortality was predicted with an accuracy of 91.39% and with an R2 of 0.9834. Therefore, it can be concluded that random forest regression is an efficient and potent method to predict the value of daily cases and deaths due to Covid-19 in the future for Turkish data.	Discussing the prediction of covid 19 death cases in Turkey using the Random Forest Regression method.

METHODOLOGY

This research method has several steps as shown in Figure 1.

Figure 1: Frame of mind



Data Preparation

This stage is the procedure of gathering statistics related to the research. The data collection section is completed to facilitate the analysis of mangrove forest health. records sorts are divided into two components: spatial and non-spatial. Spatial statistics includes drone imagery, mangrove point spacing data captured using a DJI Phantom four Multispectral Drone, and mangrove spacing limitations from the Ministry of surroundings and Forestry. In the meantime, non-spatial records is received from mangrove pattern information containing Mangrove fitness Index (MHI) values.

Data Processing

This stage aims to determine how to process the data used in this study, by processing drone imagery into vegetation indices (NDVI, NDRE, ARVI).

Oversampling Data

In this stage, statistics balancing is done the use of the synthetic Minority Oversampling method (SMOTE). SMOTE works via figuring out the k nearest acquaintances for every pattern inside the minority magnificence. artificial statistics are then generated via duplicating the minority elegance samples and their acquaintances according to the required oversampling percent, and then the new samples are decided on randomly.

Random Forest Regression Model

Random woodland Regression is a device learning set of rules able to robust prediction and estimation. It makes use of the concept of randomly combining decision bushes. This method is supported with the aid of a gridsearch cross-validation set of rules, which combines grid seek and move-validation to pick out premiere parameters. Then, via acquiring the parameter value from the quantity of bushes with the assist of the gridsearchCV algorithm, it's far then endured by means of studying the MHI fee prediction version the use of the Random woodland Regression technique.

RFR Model Accuracy

This stage is carried out to measure the performance of the *Random Forest Regression method.* In the results of the accuracy of the RFR model, there are several metrics in it, including the mtry parameter, which is the number of independent variables used to build a tree in each iteration, RMSE to measure the disparity between the values predicted by the example and the observed values, and R-Squared (R2) which represents the proportion of variants of independent variables described by independent variables in the model

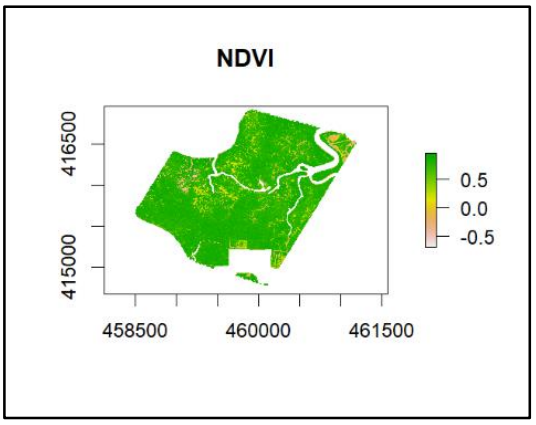
FINDING AND DISCUSSIONS

Data Processing

The multispectral imagery obtained using the DJI Phantom 4 Multispectral drone is processed to produce three main vegetation indices, namely:

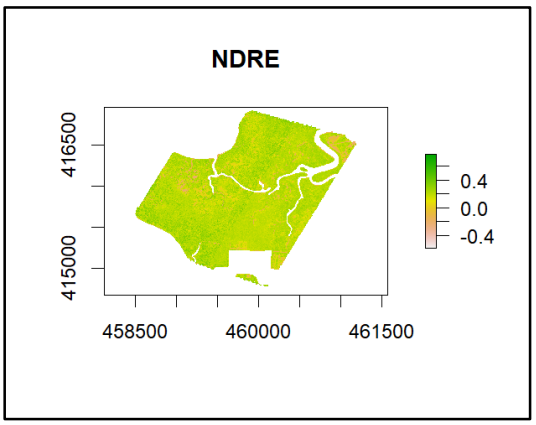
- NDVI: Indicates the greenery level of vegetation based on the difference in red (Red) and near-infrared (NIR) light reflections. The higher the NDVI value, the healthier the vegetation with a high rate of photosynthesis. The results of the NDVI calculation are shown in Figure 2.

Figure 2: NDVI Index Results



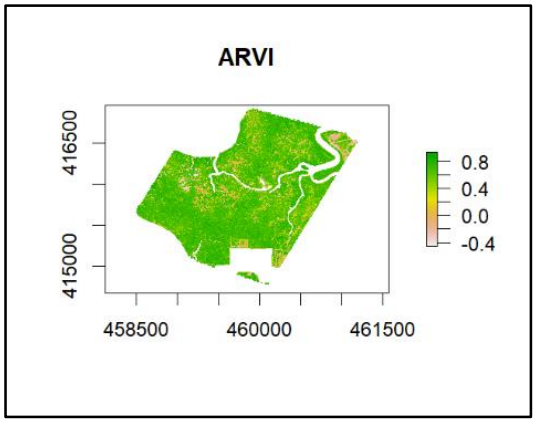
- NDRE: Indicates the level of chlorophyll content using *Red Edge* and NIR wavelengths. This index is more effective than NDVI for mature or high-chlorophyll vegetation. The results of the NDRE calculation are shown in Figure 3.

Figure 3: NDRE Index Results



- ARVI: Which is used to reduce atmospheric influences, such as fog or pollution, on light reflection. ARVI utilizes blue (*Blue*), red, and NIR bands for atmospheric correction. The results of the ARVI calculation are shown in Figure 4.

Figure 4: ARVI Index Results



Vegetation Index Extraction

To find out the value of each index, extraction was carried out using R-Studio software. The index results will have different values for each index and are shown in Table 2.

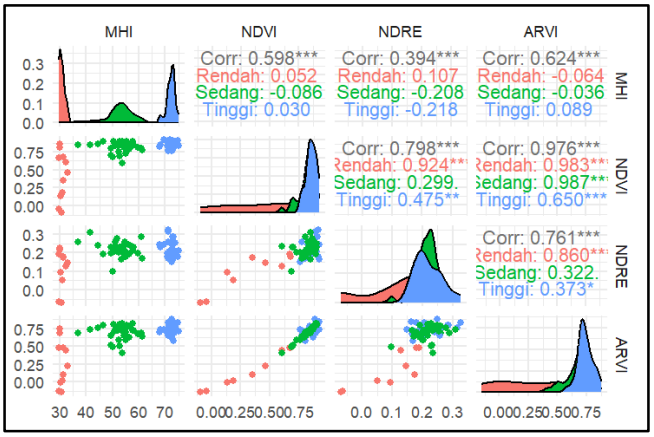
Table 2: Index Value Results

MHI	NDVI	NDRE	ARVI
30,156	0,811	0,216	0,687
30,572	-0,096	-0,069	-0,148
29,861	0,860	0,224	0,749
29,747	0,671	0,190	0,474
31,136	0,185	0,051	0,011
29,736	-0,045	-0,064	-0,128
31,267	0,679	0,181	0,474
…	…	…	…
51,772	0,864	0,278	0,757
49,429	0,797	0,199	0,649
49,485	0,702	0,232	0,516
53,516	0,846	0,165	0,724
54,853	0,821	0,256	0,712

Oversampling Data

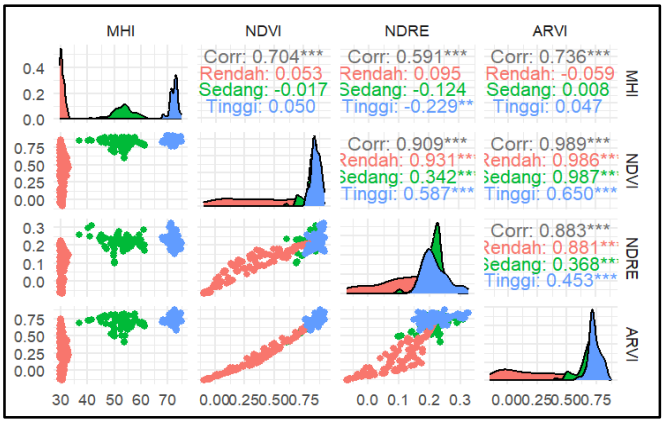
Mangrove health data based on the Mangrove Health Index (MHI) has an imbalance in distribution in the unhealthy, moderate, and healthy categories. In overcoming this, the Synthetic Minority Oversampling Technique (SMOTE) method was carried out to create synthetic data in the minority category. SMOTE works by looking for the closest neighbours in the minority category and randomly generating a new sample based on those values. There is a difference between the initial data and the data after the process of using the SMOTE method which is shown in the form of graphic images in Figure 5 and Figure 6.

Figure 5: Initial Data Graph



Look at the initial data graph a little more before using the smote technique.

Figure 6: SMOTE Data Graph



After smote, you can see the difference that the graph has a different point because there is an addition to that point. This graph helps to understand the relationship between various vegetation indices and MHI. And how the relationship differs based on the level of vegetation.

Application of Random Forest Regression

Random Forest Regression functions as a means of predicting the value of the Mangrove Health Index (MHI). However, before using the Random Forest Regression method to make predictions, hyperparameter optimization is carried out first. The best parameters such as the number of trees (*ntree*) and independent variables per tree (*mtry*) are determined using *GridSearchCV*. The *GridSearchCV* process is shown in Figure 7.

Figure 7: The *gridsearchCV* process

```
# Load the data
data <- read_excel("C:/Users/Nurheni Dewi/SKRIPSI/Belawan Mangrove/hasil_smote_RST4.xlsx")
# Inspect the data
sample <- data[, c(1, 2, 3, 4)]
View(data)
MHI <- data[1]

# Define training control
set.seed(123)
metric <- "RMSE"
control <- trainControl(method = "cv", number = 5, search = "grid")
tunegrid <- expand.grid(.mtry = c(1:10))
```

Figure 8: Best Model Search Process

```
# Loop through different ntree values
ntree_values <- c(50, 100, 150, 200, 250 , 300, 350, 400,450, 500, 550,
                  600, 650, 700, 750, 800, 850, 900, 950, 1000)
best_model <- NULL
best_rmse <- Inf
best_ntree <- NULL

for (ntree in ntree_values) {
  set.seed(123)
  model <- train(MHI ~ ., data = sample, method = "rf", metric = metric,
                 trControl = control, tuneGrid = tunegrid, ntree = ntree)

  if (min(model$results$RMSE) < best_rmse) {
    best_model <- model
    best_rmse <- min(model$results$RMSE)
    best_ntree <- ntree
  }
}

print(best_model)
print(paste("Best ntree:", best_ntree))
plot(best_model)
```

Figure 8: Is the selection of the number of trees in a random forest with the model with the lowest RMSE value selected as the best model, and the optimal number of trees is determined through a grid search. The results of the gridsearchCV process show that the best model has ***mtry* = 1** and ***ntree* = 100.** The results are shown in Figure 9.

Figure 9: Random Forest Regression Model Results

```
Random Forest

396 samples
  3 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 317, 316, 316, 317, 318
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
   1    8.515296  0.7574350  5.951945
   2    8.622321  0.7510242  5.845427
   3    8.543793  0.7557086  5.709133
   4    8.640314  0.7498250  5.745489
   5    8.686037  0.7471058  5.795516
   6    8.662688  0.7483567  5.776575
   7    8.629347  0.7504073  5.747980
   8    8.655189  0.7488929  5.774467
   9    8.568054  0.7535694  5.713213
  10    8.699517  0.7459000  5.817348

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 1.
> print(paste("Best ntree:", best_ntree))
[1] "Best ntree: 100"
```

Mangrove Health Predictions

After the Random Forest Regression model was applied, the predicted value of mangrove health was calculated based on three vegetation indices: NDVI, NDRE, and ARVI. The results of these predictions are grouped into three categories of mangrove health:

- Unhealthy: MHI < 33.33%
- Medium: 33.33% 33.33% ≤ MHI < 66.67%
- Healthy: MHI ≥ 66.67%

The results of the calculation of mangrove health predictions are shown in Table 3.

Table 3: Mangrove Health Predictions

No	Prediction Results
1	47,86
2	58,93
3	61,29
4	60,45
5	59,27
6	47,81
7	58,14
8	50,68
9	45,22
10	58,87
..	…
387	65,89
388	68,63
389	70,28
390	65,82
391	70,78
392	64,10
393	71,78
394	71,35
395	68,06
396	70,77

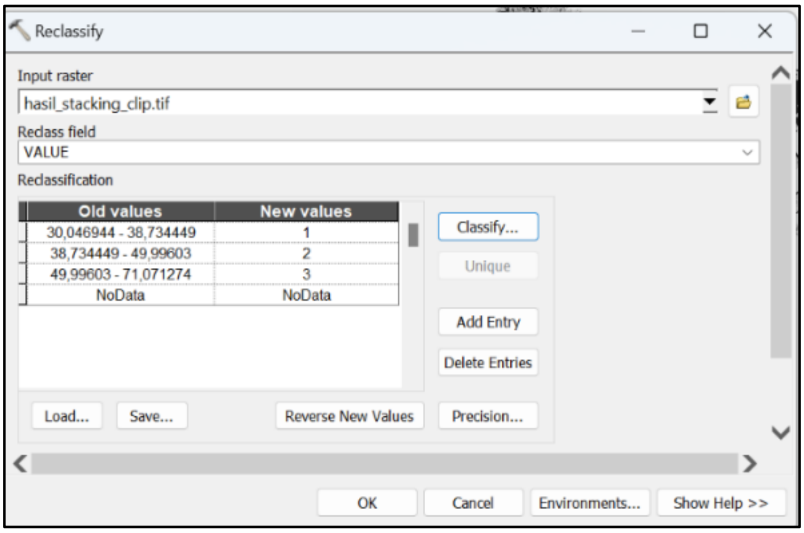
Accuracy Results

The accuracy results obtained from analyzing the prediction of Mangrove Health using the MHI value and vegetation index which resulted in an R-Squared (R2) value of 0.757. The R2 value serves as a means of measuring how much influence independent variables have on dependent variables. And there is a result of the accuracy of the calculation of Error (Root Mean Squared Error) RMSE = 8.515.

Mapping and Classification

Three vegetation indices in the form of NDVI, NDRE, and ARVI, were combined data processing to see the results of the distribution. After the prediction map is seen, the next stage is to reclassify the data, where the prediction raster values are reclassified into three categories of mangrove health (Unhealthy, Moderate, Healthy). The process of reclassifying this data is shown in Figure 10.

Figure 10: Reclassify Process



After reclassifying, the next step is to calculate the area and percentage using the *tools* in *Arcgis*, namely "Zonal Statistic as Table". These tools are used to help analyze the statistics of a raster by zone. By calculating the area and percentage, it can be determined which mangrove health category is the most dominant in the research area. The process of calculating the area and percentage is shown in Figure 11.

Figure 11: Calculating Area and Percentage

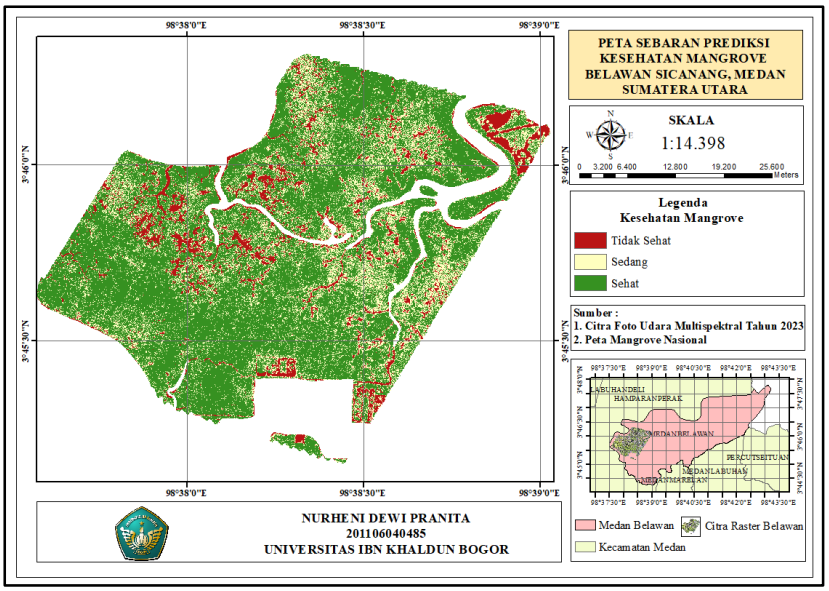
zona_reclass

OID	Value	COUNT	AREA	SUM
0	1	60688059	252086,211194	1865323860,12
1	2	153141542	636119,720013	7173604929,83
2	3	472878948	1964245,7564	25210028106,400002

Once the results of the area and percentage calculations are known, the results will be integrated with the mangrove health prediction distribution map. The formula for finding the percentage of mangrove health can be seen in Equation (1) and the final results of the mangrove health prediction distribution map are shown in Figure 12.

$$\text{Percentage} = \left(\frac{AREA}{Total\ AREA}\right) x100 \quad (1)$$

Figure 12: Mangrove Health Prediction Distribution Map



The image above shows that the results of the mangrove health prediction map have been *reclassified* with 3 classes, namely Unhealthy, Moderate and Healthy. Each class has a different percentage value, namely, the Unhealthy class with a percentage of 8.8%, Medium 22.3%, and Healthy 68.9%. With a total area of 2,852,451.68 m2.

CONCLUSION

Based totally at the studies consequences on the layout of the Random wooded area Regression version for Mangrove fitness, the following conclusions can be drawn:

1. The anticipated mangrove health results in Belawan Sicanang, Medan, North Sumatra, the use of the Random woodland Regression technique have been acquired with a expected cost of ± 29.90-seventy two.sixty seven. the overall location of 2,852,451.sixty eight m2 became divided into three training with various percentages: dangerous (8.eight%), mild (22.3%), and healthy (sixty eight.9%).
2. The accuracy effects were received with an R-Squared (R2) value of zero.757 and a Root suggest Squared blunders (RMSE) of eight.515. This indicates that the accuracy check results significantly give an explanation for the evolved version.

Primarily based on the studies results above, there are still many shortcomings inside the utility of the Random wooded area Regression method to mangrove health. consequently, the author offers guidelines and input to in addition increase this research, consisting of growing the sample length in the course of discipline facts series to gain greater accurate and strong effects.

ACKNOWLEDGEMENT

On this occasion, the author would like to express sincere gratitude to all parties who have provided support, prayers, motivation, and assistance throughout the preparation of this final project. Special thanks are extended to the author's parents, faculty and program leaders, supervisors, lecturers, and staff of the Faculty of Engineering & Science, as well as the 2020 Geospatial Information Technology colleagues for their continuous encouragement and help. The author also thanks all friends who contributed to the completion of this final project. The author realizes that this report is still far from perfect; therefore, constructive suggestions and feedback are highly appreciated.

REFERENCES

Aspan, H., Rian Putra, R., Sri Wahyuni, E., Pembangunan Panca Budi, U., & Batam, U. (2023). Sosialisasi Pengembangan Pariwisata Hutan Mangrove Belawan Sicanang Dengan Menerapkan Sistem Informasi Dala. *ARSY : Aplikasi Riset Kepada Masyarakat*, *4*(1), 2023–2024.

http://journal.al-matani.com/index.php/arsy

Doni Nurdiansah, & I Wayan Eka Dharmawan. (2021). Struktur Dan Kondisi Kesehatan Komunitas Mangrove Di Pulau Middleburg-Miossu, Papua Barat. *Jurnal Ilmu Dan Teknologi Kelautan Tropis*, *13*(1), 81–96. https://doi.org/10.29244/jitkt.v13i1.34484

Liu, Y., Liu, S., Li, J., Guo, X., Wang, S., & Lu, J. (2019). Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images. *Computers and Electronics in Agriculture*, *166*(July), 105026. https://doi.org/10.1016/j.compag.2019.105026

Minh, P., Id, H., Tinh, P. H., Son, N. P., & Thuy, T. Van. (2022). Mangrove Health Assessment Using Spatial Metrics And Multi-temporal Remote Sensing Data. *Plos One*, *17*, 1–20. https://doi.org/10.1371/journal.pone.0275928

Özen, F. (2024). Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey. *Heliyon*, *10*(4), 1–19. https://doi.org/10.1016/j.heliyon.2024.e25746

Singgalen, Y. A. (2023). Analisis Model Pengembangan Kawasan Ekowisata Mangrove Potensial Berbasis Hyper Spectral of Remote Sensing dan Analytical Hierarchy Process. *Journal of Information System Research (JOSH)*, *4*(3), 969–979. https://doi.org/10.47065/josh.v4i3.3385

Suyadi, S., Naroli, I., & Sitepu, A. B. (2021). Karakteristik Vegetasi Mangrove dan Pemanfaatannya: Studi Kasus di Seram Timur, Maluku. *Jurnal Kelautan Dan Perikanan Terapan (JKPT)*, *4*(1), 37. https://doi.org/10.15578/jkpt.v4i1.9502

Suyuti, H. (2019). Analisis Kesehatan Hutan Mangrove Dengan Menggunakan Metode Hemispherical Photography Di Kawasan Pesisir Kabupaten Tapanuli Tengah Provinsi Sumatera Utara. *Jurnal Online Mahasiswa*, 5–10.

Tridawati, A., Armijon, A., Yanto, F., & Novianti, T. C. (2023). Pemetaan Distribusi Hutan Mangrove Menggunakan Algoritma Machine Learning di Kawasan Hutan Mangrove Petengoran. *Jurnal Tekno Insentif*, *17*(2), 84–98. https://doi.org/10.36787/jti.v17i2.1101

Zhu, Y., Liu, K., Liu, L., Myint, S. W., Wang, S., Liu, H., & He, Z. (2017). Exploring the potential of world view-2 red-edge band-based vegetation indices for estimation of mangrove leaf area index with machine learning algorithms. *Remote Sensing*, *9*(10). https://doi.org/10.3390/rs9101060