https://doi.org/10.31436/iiumej.v25i2.3129

# BUILDING A GIS MAP FOR FORECASTING THE MIR INDEX IN AN GIANG

**THANH-NGHI DOAN**[1,3*], **KHANH TRAN THIEN NGUYEN**[2,3]

[1]*Faculty of Information Technology, An Giang University, An Giang, Vietnam*
[2]*Department of Science and Postgraduate Management, An Giang University, An Giang, Vietnam*
[3]*Vietnam National University, Ho Chi Minh City*

**Corresponding author: dtnghi@agu.edu.vn*

*(Received: 29 November 2023; Accepted: 8 February 2024; Published online: 15 July 2024)*

***ABSTRACT:*** The MIR aquatic plant signal is capable of predicting specific pollution sources of water, contributing significantly to the effective management of surface water resources in An Giang province. The use of aquatic plants in water pollution treatment brings about positive effects through natural self-purification processes as they consume organic and nutrient substances such as N and P. Therefore, it is crucial to develop a tool for monitoring and supervising aquatic plant species. This paper investigates the application of GIS technology to build a GIS map representing the current status of vegetation cover in An Giang province. The background layers of the GIS database, along with detailed attribute layers regarding species composition, dominant species, and vegetation area, will serve as the basis for managing, utilizing, conserving, and restoring vegetation cover in the research area. Additionally, a predictive model for MIR indices has been constructed using machine learning methods. The results indicate that the model has a coefficient of determination (R2) of 91.7% for the dependent variable MIR compared to the independent variables. Subsequently, these results are visually displayed on a GIS map at 18 monitoring points within An Giang province, enabling users to easily observe, compare, evaluate, and propose suitable solutions for surface water quality management.

***ABSTRAK:*** Isyarat tumbuhan akuatik MIR mampu meramalkan sumber pencemaran air secara spesifik, iaitu penyumbang penting kepada pengurusan berkesan permukaan sumber air di wilayah An Giang. Penggunaan tumbuhan akuatik dalam rawatan pencemaran membawa kepada kesan positif melalui proses rawatan kendiri secara semula jadi kerana ia mengandungi bahan organik dan nutrien seperti N dan P. Oleh itu, sangat penting bagi membangunkan alat pemantauan dan pengawasan spesies tumbuhan akuatik. Kajian ini mengkaji aplikasi teknologi GIS bagi membangunkan peta GIS mewakili status terkini keseluruhan tumbuhan di wilayah An Giang. Lapisan latar belakang pangkalan data GIS bersama lapisan sifat-sifat terperinci berkenaan spesies komposit, spesies dominan, dan kawasan tumbuh-tumbuhan, dapat menyediakan asas kepada pengurusan, penggunaan, pemuliharaan, dan pemulihan tumbuh-tumbuhan meliputi kawasan kajian. Tambahan, model ramalan MIR dibangunkan menggunakan kaedah pembelajaran mesin. Dapatan kajian menunjukkan model ini mempunyai pekali penentu (R2) sebanyak 91.7% bagi pembolehubah MIR bersandar berbanding pembolehubah tak bersandar. Menyebabkan dapatan ini secara visual dapat dilihat pada peta GIS menggunakan 18 titik pantauan dalam wilayah An Giang province, membolehkan pengguna mudah melihat, membandingkan, menilai, dan mencadangkan solusi sesuai bagi pengurusan kualiti permukaan air.

***KEYWORDS:*** *An Giang Province, GIS map, MIR index, Machine learning methods*

https://doi.org/10.31436/iiumej.v25i2.3129

# 1. INTRODUCTION

An Giang, situated within the southwest region of Vietnam's Mekong Delta, is a province characterized by an intricate network of rivers, canals, and channels, ensuring a consistent supply of surface water throughout the year. This hydrological setup confers numerous advantages in terms of biodiversity and sustenance for aquaculture endeavors [1]. The province's aquaculture sector has demonstrated substantial advancements in both quantity and quality, playing a pivotal role in propelling the province's socioeconomic progress. Nonetheless, the rapid expansion of aquaculture has triggered a range of concerning environmental repercussions, notably impacting the water bodies that receive aquaculture-related wastewater. Presently, the majority of wastewater stemming from fish farms across An Giang province is being discharged directly into the natural surroundings without undergoing any form of treatment. Only a small fraction of business-operated farming areas have invested in wastewater treatment systems, constituting roughly 10% of the total farming zones [2], [3], and [4]. In light of this situation, it is imperative to undertake an evaluation of the extent of pollution caused by aquaculture wastewater within the province and to devise suitable strategies to address this concern. Furthermore, in the context of integration and economic development, An Giang is facing increasing environmental pressure, particularly water pollution, which has affected the supply of fresh water for agricultural irrigation, daily life, and industrial purposes [5].

The signal of aquatic plant development, predicting the specific pollution source of water, contributes to effective and cost-efficient solutions for surface water management in An Giang province. Furthermore, the use of various plant species for water pollution treatment in aquatic ecosystems yields positive results through natural self-cleaning processes and the cooperative consumption of organic matter, nutrients, nitrogen (N), and phosphorus (P) in the water for growth, thereby purifying the water [6]. The plant biomass after harvesting can be utilized as animal feed or organic fertilizer for fields, closing the loop in the treatment process. Therefore, implementing surface water treatment technologies in natural or ecological conditions is one suitable solution in the context of climate change and the economy in An Giang province. In [7], the authors demonstrated the possibility of distinguishing macrophytes based on their ecological preferences for water nutrient concentrations, offering a potential foundation for an efficient freshwater assessment system. The Macrophyte Nutrient Index emerged as the indicator displaying the strongest correlation with diverse nutrient forms. Notably, calibrating the macrophyte method to align with local biogeographical conditions was found to significantly enhance the effectiveness of the assessment approach.

Environmental management plays a crucial role in protecting our environment and shaping strategic approaches for its preservation. In today's society, addressing issues related to the management of land, water, and air quality is an unavoidable reality. To assist decision-makers in developing effective environmental management strategies, it is essential to understand the intricacies and characteristics of environmental challenges. Spatial information and data are the fundamental building blocks of environmental management knowledge. Among the many technological solutions available, Geographic Information Systems (GIS) stands out as a leading tool for storing, analyzing, and visualizing geographically referenced information, providing a highly efficient and accurate way to tackle these challenges [8]. Currently, in the Mekong Delta region, there are some surveys on biodiversity and the distribution of various biological groups, primarily vegetation maps [9]. GIS technology has been applied to create GIS maps that depict the current state of vegetation cover in the surveyed areas. This includes background information layers in the GIS database and detailed attribute layers containing data on species composition, dominant species, and vegetation area. These maps serve as a foundation for the management, utilization, conservation, and restoration of vegetation in the research area [10]. Therefore, the development of GIS maps and predictive models is essential to contribute to the management and treatment of clean water resources and the conservation of aquatic vegetation in the canals and waterways of An Giang province, especially in the context of climate change, using information technology databases.

Within the analysis of aquatic systems, modeling water quality parameters holds significant importance. Traditional modeling methods rely on datasets with substantial amounts of unknown or unspecified input data, often involving time-consuming processes. The integration of artificial intelligence introduces a flexible mathematical framework capable of discerning non-linear and intricate relationships between input and output data. In [11], the authors suggested different Artificial Neural Network approaches, such as Multi-Layer Perceptron Neural Network and Radial Basis Function Neural Network, to validate the efficacy of these methods in estimating water quality parameters. In [12], the authors demonstrated the use of grid search for optimizing and fine-tuning parameters in four classification models for predicting Water Quality Class and four regression models for predicting Water Quality Index. In [13], the authors leveraged machine learning algorithms to construct a model capable of forecasting both the water quality index and its corresponding quality class. Their proposed method relies on four key water parameters: temperature, pH, turbidity, and coliforms. The application of multiple regression algorithms has proven to be crucial and effective in predicting the water quality index. Additionally, the adoption of artificial neural networks represents the most efficient means for classifying water quality.

This paper presents the construction of a dataset of water surface macrophyte biological signals from the canals, rivers, and streams of An Giang province, as well as a machine learning model to predict the environmental pollution level based on the distribution and frequency of aquatic plants. Several machine learning models have developed, compared, and evaluated the predictive model against various machine learning methods, including Artificial Neural Networks (ANN) [14], k-nearest neighbors algorithm (kNN) [15], Support vector machines (SVMs) [16], Stochastic Gradient Descent [17], Random Forest [18], Linear regression, Gradient Boosting [19]. Experimental results show that the Gradient Boosting model successfully simulated the surface water pollution data, achieving the highest predictive accuracy with a coefficient of determination (R2) of 91.7% for the dependent variable MIR compared to the independent variables.

## 2. MATERIALS AND METHODS

### 2.1. Data collection

Surface water samples were collected at 18 sampling locations affected by aquaculture areas. The sample collection schedule was in March and June 2021. Fig. 1 shows the location map of surface water sampling. Table 1 shows the code, coordinates, and features of sampling locations.

Table 1. Location of surface water sampling.

| No. | Code | Coordinate | Location | Feature |
|---|---|---|---|---|
| 1 | AQ1 | 559.880, 1.188.560 | Long Hoa commune, Phu Tan district | The end of wastewater source of the Long Hoa raft floating fish farming |
| 2 | AQ2 | 570.753 1.168.426 | Kien An commune, Cho Moi district | The end of wastewater source of pen culture and pond fish farming area in Kien An |
| 3 | AQ3 | 540.104 1.184.937 | Chau Phu A ward, Chau Doc city | The end of wastewater source of the Vinh Nguon raft floating fish farming |
| 4 | AQ4 | 540.708 1.184.988 | Da Phuoc commune, An Phu district | The end of wastewater source of the Da Phuoc raft floating fish farming |
| 5 | AQ5 | 548.722 1.172.710 | Vinh Thanh Trung commune, Chau Phu district | Impact of wastewater from fish farming ponds on Xang Vinh Tre canal |
| 6 | AQ6 | 567.496 1.156.042 | Binh Thanh commune, Chau Thanh district | The end of wastewater source of pond fish farming area in Binh Thanh |
| 7 | AQ8 | 576.289 1.151.944 | My Hoa Hung commune, Long Xuyen city | The end of wastewater source of pen culture and pond fish farming area in My Hoa Hung |
| 8 | AQ11 | 574.103 1.139.929 | My Thoi ward, Long Xuyen city | Impact of wastewater from fish farming ponds |
| 9 | AQ13 | 552.267 1.153.651 | Vinh Thanh commune, Chau Thanh district | Impact of wastewater from rice-shrimp farming area |
| 10 | AQ14 | 572.673 1.139.250 | Phu Thuan commnue, Thoai Son district | The beginning of the Don Dong canal, adjacent to the Moi canal |
| 11 | AQ15 | 565.734 1.136.318 | Vinh Khanh commune, Thoai Son district | The beginning of the Don Dong channel, adjacent to the Ong Co canal |
| 12 | AQ19 | 584.521 1.157.065 | My An commune, Cho Moi district | Impact of wastewater from raft floating fish farming |
| 13 | AQ20 | 548.134 1.201.285 | Vinh Hoa commune, Tan Chau town | Impact of wastewater from pond fish farming area |
| 14 | AQ21 | 546.162 1.169.627 | Phu My commune, Chau Phu district | Impact of wastewater from Loc Kim Chi fish farming area, the confluence between Hao De Lon canal and Xang Vinh Tre canal |
| 15 | AQ22 | 548.706 1.158.636 | Binh Phu commune, Chau Phu district | Impact from wastewater from Nam Viet Binh Phu fish farming area, the confluence between 13 canal and Xang Cay Duong canal |
| 16 | AQ23 | 552.578 1.171.504 | Phu Binh commune, Phu Tan district | Impact of wastewater from Pangasius farming area in Phu Binh commune |
| 17 | AQ24 | 561.948 1.145.178 | Vinh Trach commune, Thoai Son district | Impacts from aquatic discharge source |
| 18 | AQ25 | 550.885 1.177.362 | Hoa Lac commune, Phu Tan district | Impact of wastewater from pond fish farming area of Hoa Lac |

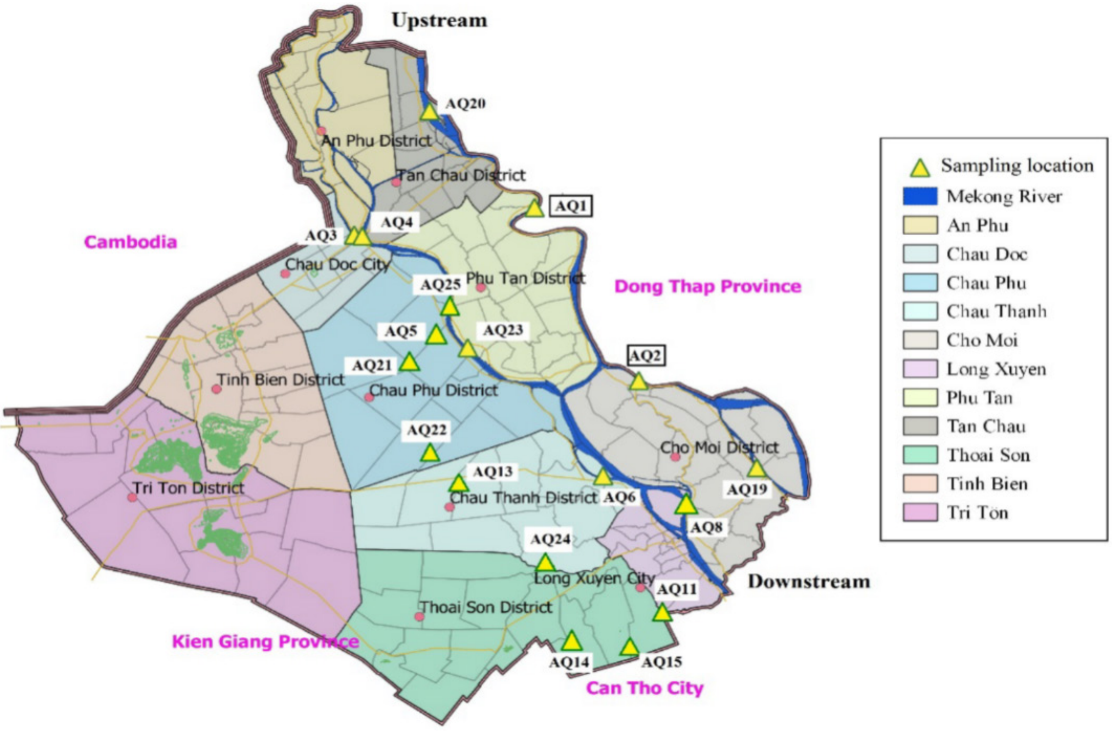



Figure 1. The location map of surface water sampling.

### 2.2. Methods of Sampling, Preserving, and Analyzing Water Samples

The assessment of 18 water samples involved the evaluation of various parameters, including Temperature, pH, Dissolved Oxygen (DO), Chemical oxygen demand (COD), Biological oxygen demand (BOD5), Total Suspended Solids (TSS), Ammonium (N-NH4+), Nitrate (N-NO3-), Phosphate (P-PO43-), Total Nitrogen (TN), Total Phosphorus (TP), and Coliform. The procedures for environmental monitoring followed the guidelines set forth in Circular 24/2017/TT-BTNMT, which outlines technical regulations for environmental monitoring and adhered to Vietnamese standards and regulations established by the Vietnam Ministry of Natural Resources and Environment in 2017.

To collect and preserve surface water and wastewater samples, the methods specified in TCVN 6663-1:2011 and TCVN 6663:3:2016 were employed, as outlined by the Vietnam Ministry of Science and Technology in 2011 and 2016, respectively. Sample analysis was conducted in accordance with the procedures described in Standard Methods [20] and the Vietnam National Standard on Water Quality, as specified by the Vietnam Ministry of Science and Technology in 1995, 1996, 2000, 2011, and 2016 (Table 2 and Table 3). For the measurement of temperature, DO, and pH, a PCD 650/pH 600 Eutech measuring probe was utilized.

The study assessed the plant growth performance within the wastewater treatment tank by tallying the branches of each plant during various experimental stages and gauging the length of the plant's stem both before and after a 28-day experiment. This research was conducted at two chosen agricultural households in Chau Phu district, An Giang province. The experiment was organized in the following manner:

Bamboo poles enclose the wastewater tank and feature a waterproof rubber lining, measuring 1.5 meters by 1.5 meters by 1 meter, resulting in a total water storage capacity of 1.8 cubic meters. Aquatic plants floating on the water surface were gathered from the wastewater discharge area in accordance with the density observed in our survey. The preparation proceeded as follows:

Table 2. Method of analyzing/measuring water samples

| No. | Parameter | Method of analyzing/measuring |
|---|---|---|
| 1 | Chemical oxygen demand (COD) | SMEWW 5220C:2017 |
| 2 | Biological oxygen demand ($BOD_5$) | SMEWW 5210B:2017 |
| 3 | Total Suspended Solids (TSS) | SMEWW 2540D:2017 |
| 4 | Ammonium ($N\text{-}NH_4^+$) | TCVN 5988:1995 |
| 5 | Total Nitrogen (TN) | TCVN 6638:2000 |
| 6 | Total Phosphorus (TP) | SMEWW 4500.P.B&E:2017 |
| 7 | Nitrate ($N\text{-}NO_3^-$) | SMEWW 4500-$NO_3^-$-E:2017 |
| 8 | Phosphate ($P\text{-}PO_4^{3-}$), | SMEWW 4500-P.E:2017 |
| 9 | Coliform | TCVN 6187-2:1996 |
| 10 | pH | TCVN 6492:2011 |
| 11 | Temperature | SMEWW 2550B:2012 |
| 12 | DO | TCVN 7325:2016 |

Water hyacinth: Select plants with a length of 20 cm and 4-5 leaves. Remove the roots, damaged stems, and leaves from the water hyacinth and place them in clean water for a period of 7 days prior to the experiment [21]. Water spinach and Climbing dayflower: Select young, healthy plants that are uniform in color and height, measuring 20 cm in height. Prior to the experiment, these plants should be washed with distilled water, as outlined in the study by [22].

The framework for the aquatic floating raft was constructed using D90mm PVC pipes, with the raft's dimensions measuring 1 meter by 1 meter. The mesh supporting the groups of aquatic plants is made of polyethylene with 2 cm-sized openings. As a result, the total area of the raft structure is 1 square meter. At each stage, wastewater was gathered from the discharge pipes of two fish ponds within the chosen farming households.

Aquatic plants were positioned on floating rafts, with a spacing of 30 cm between individual plants and 30 cm between rows, spanning across the wastewater tank. Within each floating raft, a total of 3 plant types were incorporated, resulting in 9 plant groups, each comprising 2-3 plant types. The plant density within the rafts was arranged based on the observed densities within the surveyed area. Subsequently, the rafts were secured within the wastewater tank using ropes. Fig. 2 illustrates the wastewater tank and the floating raft, while Fig. 3 provides an overview of the experimental design. Table 3 displays the water quality rating based on Water Quality Index ranges and recommendations for surface water usage.

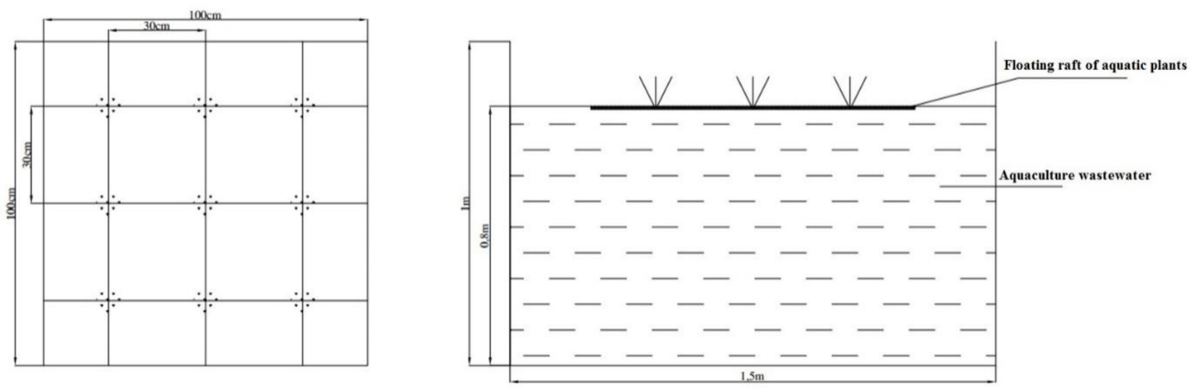



Figure 2. The wastewater tank and the floating raft of aquatic plants

https://doi.org/10.31436/iiumej.v25i2.3129

Figure 3. Design of Experiment.

Table 3. Water Quality Rating and recommendation of usage

| Water Quality Index Range | Water Quality Rating | Color | Intended use |
|---|---|---|---|
| 91-100 | Excellent | | Good for water supply |
| 76-90 | Good | | For water supply but requires appropriate treatment measures |
| 51-75 | Medium | | For irrigation and other similar purposes |
| 26-50 | Poor | | For water transport and other similar purposes |
| 10-25 | Polluted | | Water is heavily polluted, requires future treatment |
| <10 | Seriously polluted | | Water is poisoned, requires treatment |

*Note.* From "Decision 1460/QD-TCMT on promulgating technical guidelines for calculation and Vietnam water quality index (VN_WQI)" by Vietnam Environment Administration, 2019.

## 2.3. Macrophyte Index for Rivers

Macrophytes are widely used to assess the environmental conditions of various river ecosystems [23]. To obtain a more objective assessment of river water quality in addition to using hydrochemical indicators, the Macrophyte Index for Rivers (MIR) has been implemented [24]. The research method utilizes floating plants as a biological indicator to predict the surface water quality in An Giang province. Based on the collected macrophyte data, the Macrophyte Index for Rivers (MIR) was adapted by WFD to suit the water environment of the Mekong Delta region. In this method, the value of the MIR index depends on the presence of identified species in the river and their indicator values. According to the study conducted by the authors [7], the MIR (Macrophyte Index for Rivers) is calculated from field records using Eq. (1).

$$IR = \frac{\sum_{i=1}^{N} l_i.w_i.p_i}{\sum_{i=1}^{N} w_i.p_i} \times 10 \qquad (1)$$

where $MIR$ is the Macrophyte Index for the river at the sampling point, $N$ is the number of species at the sampling point, $l_i$ is the indicator value for classification unit $i$, $w_i$ is the weighting factor of classification unit i, and $p_i$ is the coverage density of classification unit $i$ on a 9-point scale. Specifically: 1 for 0.1%, 2 for 0.1–1%, 3 for 1–2.5%, 4 for 2.5–5%, 5 for 5–10%, 6 for 10–25%, 7 for 25–50%, 8 for 50–75%, and 9 for 75%. The classification of water quality in the research areas is based on the MIR index, as studied in the paper by [24], and is presented in Table 4.

Table 4. Water quality classification based on the MIR index.

| No. | Ecological state | MIR | Water quality |
|---|---|---|---|
| 1 | Very good | >=44.7 | I |
| 2 | Good | (44.7 – 36.5] | II |
| 3 | Medium | (36.5 – 28.2] | III |
| 4 | Bad | (28.2 – 20.0] | IV |
| 5 | Very bad | <20.0 | V |

### 2.4. Creating GIS map and database

The establishment of the current status map of aquatic plants using traditional methods has been in existence for a long time but comes with numerous limitations in implementation, requiring significant investments of time and effort in data collection and consolidation at various local levels. The lengthy process of aggregating and constructing the current status map of aquatic plants for a region leads to outdated and inaccurate information on the map. Therefore, there is a need for an alternative method to overcome the shortcomings of the traditional approach in investigating and researching the current status of aquatic plants, analyzing geographically relevant information, and meeting the urgent demands of practical production and scientific research. The choice of how to represent data on a map is a crucial step in geographic analysis. In our study, we utilized Geographic Information Systems (GIS) to create a digital map illustrating the aquatic plants found in An Giang province. GIS is a computerized database designed for managing and utilizing spatial data effectively. We carefully determined the structure of the main folders and subfolders within the GIS database. The plant data collected during field surveys was gathered using GPS and subsequently integrated into the GIS system. A thorough validation process was conducted using field forms to ensure data accuracy. We employed a georelational data model within the GIS to store attribute data alongside spatial data. Additionally, we adopted an object-based data model that consolidates spatial data with other attributes within a single system. This approach simplifies the management process by eliminating the complexity of coordinating and synchronizing two separate data sets, as is often required in a split system. The initial plant layer in our current database lacked sufficient information. To address this issue, we added columns to the attribute table of the existing plant layers to input relevant information, as described in studies by [25] and [8].

In the context of the Mekong Delta region, several studies have been conducted to survey the biodiversity and map the distribution of various biological groups, predominantly focusing on aquatic plant maps [26], [27]. GIS technology has been applied to create GIS-based maps illustrating the current status of aquatic plant coverage in areas under investigation. Within these maps, foundational information layers in the GIS database, along with detailed attribute layers containing species composition, dominant species, and area coverage of aquatic plants, serve as the basis for managing, utilizing, conserving, and restoring aquatic plant communities within the study region [10]. This paper presents the process of using remote sensing analysis to establish the current status map of aquatic plants, specifically in An Giang province. The current status map of aquatic plants in An Giang province has been developed with the following content:

- Determining the presence of floating aquatic plant families along river and stream segments flowing through An Giang province at 18 observation points:
- Designing the standard Quadrat frame and conducting practical surveys at 18 observation points.

- Determine the number of species and calculate the species density.
- Research on using floating plants as a biological indicator to predict surface water quality in An Giang province.
- Building an online GIS database and map of aquatic plant species data. Building an online aquatic GIS map system using open-source WebGIS technology. The general model of the system is presented in Fig. 5. The components of the model include:
- MySQL is a place to store spatial and non-spatial data. This data is stored in the form of tables. Therefore, this is the place to build a database of aquatic information at all survey points of the project in An Giang province.
- Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps, used to overlay maps, calculate spatial information, and receive and display information from WMS services and WebService, combining with HTML and JavaScript in designing interfaces and displaying spatial databases on the Web.
- QGIS is used to edit, edit, map, and serve information management of land plots in raw material production areas.

### 2.5. Method for building an online GIS database and map of aquatic plant species data

Surveying is the professional process of collecting information and data on aquatic life and water environment monitoring by research staff and specialized units at the Department of Natural Resources and Environment of An Giang Province. Research the structure of relevant forms and reports specified in documents on the water environment and aquatic life management. Object-Oriented Analysis and Design (OOAD) system analysis, from focusing on user needs expressed through building use case diagrams to database design data and processing, creating software, using, testing, and maintaining software.

#### *2.5.1. Predicting Model Based on Machine Learning Methods*

Building a predictive model for environmental pollution indicators based on the distribution and occurrence frequency of invasive alien species in An Giang. The predictive model is constructed using state-of-the-art advanced machine learning techniques such as the Adaptive Neuro-Fuzzy Inference System (ANFIS), Radial Basis Function Neural Networks (RBF-ANN), Multi-Layer Perceptron Neural Networks (MLP-ANN), and deep learning techniques.

#### *2.5.2. Data preprocessing, removing noisy data during the data collection process*

Constructing a predictive model for environmental pollution indicators is based on the best techniques and methods in the fields of machine learning and artificial intelligence, such as neural networks and deep learning. The predictive model is researched and built and can be described in a general way, as shown in Fig. 4. Evaluating the predictive model, comparing it with other traditional methods, and validating the predictive model by consulting with experts' opinions.

The predictive model for the MIR biological index has been researched and constructed and can be described in a general manner, as shown in Fig. 4. In this model, the input values consist of data on the aquatic plant density of four species: Eichornia crassipes, Polygonum hydropiper, Ipomoea aquatica, and Centrostachys aquatica, with a total of 68 samples as presented in Table 11. We have divided this dataset into two parts: one for training, comprising 80% of the samples, and the remaining 20% for testing and assessing the accuracy of the machine learning models.

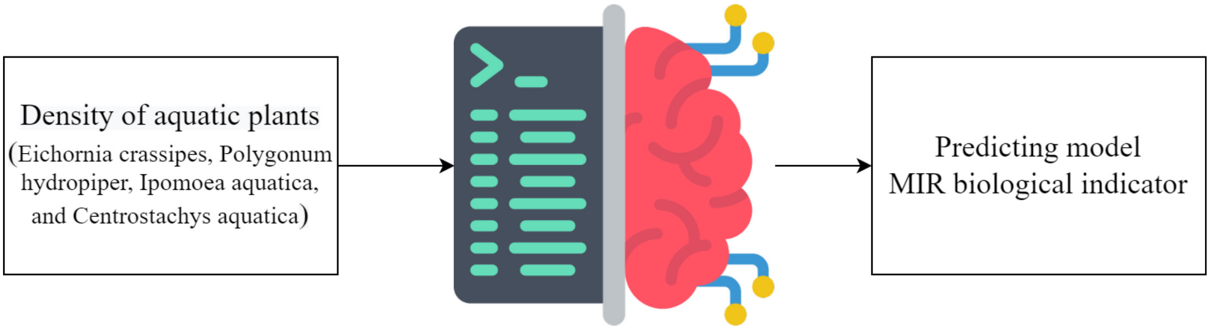
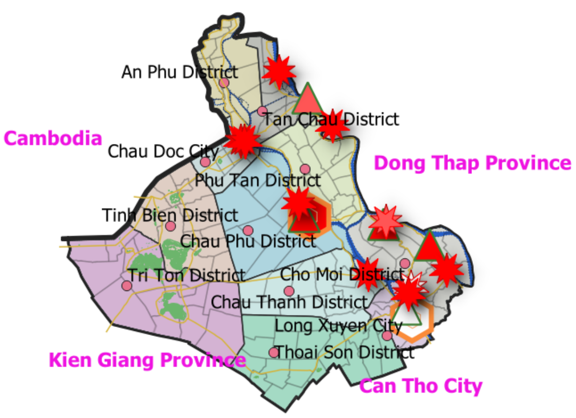



Figure 4. Machine learning model predicts MIR index

## 3. RESULTS AND DISCUSSION

### 3.1. MIR GIS Map in An Giang Province

Fig. 5 describes a GIS map constructed with information layers, including the locations of observation points, MIR index collection, and aquatic plant species. QGIS software was used to edit the map and display the layers, while VN2000 coordinates were used for map positioning. The content of both general maps and aquatic plant distribution maps consists of two fundamental parts closely related to each other: geographical foundation elements and specialized content elements. The geographical foundation elements are considered the background of the thematic map and are seen as the framework for determining the thematic elements. In the case of aquatic plant distribution maps, the geographical foundation elements serve as the basis for orientation and clarifying the characteristics and patterns of distribution of the thematic map's content. Based on available 1:50,000 topographic base map data and thematic maps at the same scale, as well as acquired remote sensing imagery, the current state map of vegetation in An Giang Province was established at a scale of 1:50,000 using the UTM WGS 84 coordinate system. The thematic content elements represent the spatial distribution of corresponding types of vegetation cover for each plot. These plots accurately depict the location, area, and shape of each vegetation cover type in proportion to the map's scale.




| Feature | Value |
| --- | --- |
| Fishsery | |
| Title | |
| (Derived) | |
| (Actions) | |
| Code | TS1 (TĐ)-PT |
| x | 10,747287999999999 |
| y | 105,299304000000006 |
| Address | Long Hoa Commune, Phu Tan District |
| Feature | At the end of the waste stream of Long Hoa raf... |
| Intended_use | Monitoring the impact of aquaculture activities ... |
| Map_code | FS1 |
| Eichhornia_crassipes | 433 |
| Polygonum_tomentosum | 0 |
| Ipomoea_aquatica | 54 |
| Centrostachys_aquatica | 0 |
| MIR | 20 |

Figure 5. The map is constructed with information layers pertaining to the Fishery aquaculture region.

### 3.2. Survey of the Frequency of Occurrence of Aquatic Plant Species and MIR Index in An Giang

Field surveys at 18 locations along rivers and waterways within the jurisdiction of An Giang Province revealed that aquatic plant species such as water hyacinth, water spinach, morning

https://doi.org/10.31436/iiumej.v25i2.3129

glory, and water mimosa were found. However, there was no presence of lotus, red water lily, or water lily at any of the 18 survey locations. These aquatic plant species are all listed in the book by [28], and some of them were also studied by [29] in Can Tho City. It can be observed that, although there is not a great variety of species at the surveyed locations, these species are valuable for water pollution treatment, as demonstrated in numerous studies. Among them, water hyacinth is the most predominant and is present throughout the year at all survey locations. Actual photos of these aquatic plant species in the surveyed area are presented in Fig. 6.

Figure 6. Photos of aquatic plant species in the surveyed area

**The area affected by urbanization:** The results of the survey on the frequency of occurrence and density of aquatic plant species in the urban-affected area in 2021 are presented in Table 5. Water hyacinth was present in all seasons throughout the year. Water hyacinth predominated over other species with an occurrence frequency of over 85%. It had a 100% occurrence rate in all four seasons, with a density ranging from 10.25 to 19.3 plants/$m^2$ in the urban-affected areas. Water spinach was relatively abundant, with an occurrence frequency ranging from 20% to 80%. It appeared in most locations during all four seasons; some locations had it in only three or one season, with a density ranging from 0.55 to 5.45 plants/$m^2$. On the other hand, water mimosa were relatively scarce. Water mimosa mainly appeared in March, June, and September but was not observed in November, with an occurrence frequency ranging from 15% to 95% and a density of 0.4 to 6.8 plants/$m^2$. Some water mimosa appeared in only one to two seasons, mainly in September and November, with an occurrence frequency ranging from 15% to 90% and a density ranging from 0.4 to 6.1 plants/$m^2$.

Table 5. The frequency of occurrence and species density
in the urban-affected area in 2021.

| No | Location code | Total | Frequency of appearance (%) | | | | | | | Density (plant/m$^2$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Eichornia crassipes | Polygonum hydropiper | Ipomoea aquatica | Centrostachys aquatica | Nulumbo nucifera | Nelumbonaceae | Eichhornia crassipes | Eichornia crassipes | Polygonum hydropiper | Ipomoea aquatica | Centrostachys aquatica | Nulumbo nucifera | Nelumbonaceae | Eichhornia crassipes |
| 1 | MT1 (TĐ-ĐT)-TC | T3 | 100 | 25 | 35 | - | - | - | - | 14,8 | 0,6 | 3,3 | - | - | - | - |
| | | T6 | 100 | - | 60 | - | - | - | - | 18,45 | - | 5,45 | - | - | - | - |
| | | T9 | 100 | 15 | 80 | 25 | - | - | - | 15,4 | 0,4 | 4,0 | 0,6 | - | - | - |
| | | T11 | 100 | - | 80 | 25 | - | - | - | 14,2 | - | 4,4 | 0,7 | - | - | - |
| 2 | MT2 (TĐ-ĐT)-CM | T3 | 100 | - | 20 | - | - | - | - | 17,2 | - | 0,65 | - | - | - | - |
| | | T6 | 90 | - | 20 | - | - | - | - | 14,35 | - | 1,3 | - | - | - | - |
| | | T9 | 100 | - | 60 | 45 | - | - | - | 12,95 | - | 1,15 | 4,75 | - | - | - |
| | | T11 | 100 | - | 20 | 90 | - | - | - | 13,8 | - | 0,55 | 6,1 | - | - | - |
| 3 | MT3 (TĐ-ĐT)-CM | T3 | 100 | - | 35 | - | - | - | - | 19,3 | - | 2,9 | - | - | - | - |
| | | T6 | 100 | - | 45 | - | - | - | - | 17,75 | - | 3,55 | - | - | - | - |
| | | T9 | 100 | 40 | 40 | - | - | - | - | 14,3 | 1,4 | 0,6 | - | - | - | - |
| | | T11 | 90 | - | - | 55 | - | - | - | 13,4 | - | - | 2,35 | - | - | - |
| 4 | MH3 (TĐ-ĐT)-LX | T3 | 100 | - | 55 | - | - | - | - | 10,25 | - | 2,85 | - | - | - | - |
| | | T6 | 90 | - | 30 | - | - | - | - | 12,45 | - | 1,8 | - | - | - | - |
| | | T9 | 100 | - | 35 | 15 | - | - | - | 14,6 | - | 1,3 | 0,4 | - | - | - |
| | | T11 | 100 | - | 55 | - | - | - | - | 15,6 | - | 4,05 | - | - | - | - |
| 5 | MH2 (TĐ-ĐT)-CP | T3 | 100 | 75 | - | - | - | - | - | 10,85 | 6,8 | - | - | - | - | - |
| | | T6 | 85 | 95 | - | - | - | - | - | 15,05 | 6,5 | - | - | - | - | - |
| | | T9 | 100 | 75 | - | - | - | - | - | 14,55 | 4,8 | - | - | - | - | - |
| | | T11 | 100 | - | 30 | 80 | - | - | - | 14,65 | - | 0,8 | 2,55 | - | - | - |

**The area affected by industrial zones/clusters:** Table 6 presents the frequency of occurrence and species density in the area affected by industrial zones in 2021. In this table, both Water Hyacinth and Water Spinach were present in all seasons. Water Hyacinth predominated over other species, with an occurrence frequency of 80-100% in the locations, with densities ranging from 6.2 to 20.15 plants/m$^2$. Water Spinach was present with an occurrence frequency ranging from 15% to 100% and densities ranging from 0.5 to 8.45 plants/m$^2$. Water Mimosa was present at 2 locations during 3 seasons of the year, appearing at both locations in September and November, with occurrence frequencies ranging from 20% to 60% and densities of 0.7 to 4.05 plants/m$^2$. Nymphoides were present at one location in March and June, with an occurrence frequency ranging from 60% to 65% and densities of 3.8 to 4.05 plants/m$^2$, while it was not observed at the remaining locations.

Table 6. The frequency of occurrence and species density in the area affected by
industrial zones in 2021.

| No | Location code | Total | Frequency of appearance (%) | | | | | | | Density (plant/m$^2$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Eichornia crassipes | Polygonum hydropiper | Ipomoea aquatica | Centrostachys aquatica | Nulumbo nucifera | Nelumbonaceae | Eichhornia crassipes | Eichhornia crassipes | Polygonum hydropiper | Ipomoea aquatica | Centrostachys aquatica | Nulumbo nucifera | Nelumbonaceae | Eichhornia crassipes |
| 6 | MH1 (TĐ-CN)-CP | T3 | 100 | - | 40 | 20 | - | - | - | 20,15 | - | 2,4 | 0,95 | - | - | - |
| | | T6 | 100 | - | 15 | - | - | - | - | 24,1 | - | 0,5 | - | - | - | - |
| | | T9 | 100 | - | 70 | 30 | - | - | - | 15,6 | - | 6,05 | 0,7 | - | - | - |
| | | T11 | 100 | - | 80 | 45 | - | - | - | 16,95 | - | 5,7 | 3,75 | - | - | - |
| 7 | MH2 (TĐ-CN)-LX | T3 | 100 | 65 | 55 | - | - | - | - | 17,8 | 3,8 | 4,25 | - | - | - | - |
| | | T6 | 100 | 60 | 100 | 60 | - | - | - | 6,2 | 4,05 | 8,45 | 4,05 | - | - | - |
| | | T9 | 80 | - | 90 | 50 | - | - | - | 8,85 | - | 6,6 | 3,8 | - | - | - |
| | | T11 | 90 | - | 80 | 35 | - | - | - | 11,85 | - | 7,3 | 2,8 | - | - | - |

**The area affected by aquaculture:** Table 7 presents the frequency of occurrence and species density in the area affected by aquaculture in 2021. Only water hyacinth was present

https://doi.org/10.31436/iiumej.v25i2.3129

in all seasons, with occurrence frequencies ranging from 35% to 100%. It had a 100% occurrence rate in all four seasons, with densities ranging from 5.7 to 28.05 plants/$m^2$. Water spinach was present with occurrence frequencies ranging from 10% to 100% and densities ranging from 0.4 to 9.95 plants/$m^2$ at locations where it was present, but it had a lower occurrence compared to water hyacinth. Only two locations had water spinach in all four seasons, and one location showed no presence. Meanwhile, Water Mimosa were present with relatively low occurrence frequencies at a few locations, with some locations showing no presence. Water mimosa appeared at only one location in three seasons, while the remaining locations mostly appeared in one to two seasons, with occurrence frequencies ranging from 10% to 90% and densities from 0.15 to 4.45 plants/$m^2$. Additionally, Water Mimosa was present at only 2 locations and appeared in only one season throughout the year.

Table 7. The frequency of occurrence and species density in the area affected by aquaculture in 2021

| No | Location code | Total | Frequency of appearance (%) | | | | | | | Density (plant/$m^2$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Eichornia crassipes | Polygonum hydropiper | Ipomoea aquatica | Centrostachys aquatica | Nulumbo nucifera | Nelumbonaceae | Eichhornia crassipes | Eichornia crassipes | Polygonum hydropiper | Ipomoea aquatica | Centrostachys aquatica | Nulumbo nucifera | Nelumbonaceae | Eichhornia crassipes |
| 1 | TS1(TĐ)-PT | T3 | 80 | - | 60 | - | - | - | - | 21,65 | - | 2,7 | - | - | - | - |
| | | T6 | 95 | - | 80 | 10 | - | - | - | 14,15 | - | 4,1 | 0,15 | - | - | - |
| | | T9 | 100 | - | 40 | - | - | - | - | 20,4 | - | 1,65 | - | - | - | - |
| | | T11 | 100 | - | - | - | - | - | - | 23,45 | - | - | - | - | - | - |
| 2 | TS2(TĐ)-CM | T3 | 100 | - | - | 45 | - | - | - | 8,4 | - | - | 0,15 | - | - | - |
| | | T6 | 100 | - | - | 40 | - | - | - | 9,1 | - | - | 0,75 | - | - | - |
| | | T9 | 100 | - | - | 30 | - | - | - | 8,85 | - | - | 0,6 | - | - | - |
| | | T11 | 100 | - | - | - | - | - | - | 8,45 | - | - | - | - | - | - |
| 3 | TS3(TĐ)-CĐ | T3 | 90 | - | 35 | - | - | - | - | 11,65 | - | 1,35 | - | - | - | - |
| | | T6 | 100 | - | - | - | - | - | - | 25,35 | - | - | - | - | - | - |
| | | T9 | 100 | - | - | - | - | - | - | 28,05 | - | - | - | - | - | - |
| | | T11 | 100 | - | - | - | - | - | - | 24,9 | - | - | - | - | - | - |
| 4 | TS4(TĐ)-AP | T3 | 100 | - | 70 | - | - | - | - | 16,85 | - | 5,1 | - | - | - | - |
| | | T6 | 85 | - | 65 | - | - | - | - | 14,35 | - | 4,0 | - | - | - | - |
| | | T9 | 90 | - | 55 | 10 | - | - | - | 15,8 | - | 3,45 | 0,35 | - | - | - |
| | | T11 | 35 | - | 65 | 70 | - | - | - | 2,95 | - | 4,45 | 6,95 | - | - | - |
| 5 | TS6(TĐ)-CT | T3 | 100 | - | 40 | - | - | - | - | 21,05 | - | 1,05 | - | - | - | - |
| | | T6 | 100 | - | 30 | - | - | - | - | 15,2 | - | 1,0 | - | - | - | - |
| | | T9 | 100 | - | - | - | - | - | - | 22,95 | - | - | - | - | - | - |
| | | T11 | 80 | - | - | - | - | - | - | 12,6 | - | - | - | - | - | - |
| 6 | TS7(TĐ)-LX | T3 | 100 | 25 | | 35 | - | - | - | 23,4 | 1,55 | - | 2,1 | - | - | - |
| | | T6 | 100 | - | 65 | - | - | - | - | 12,35 | - | 2,1 | - | - | - | - |
| | | T9 | 100 | - | 55 | - | - | - | - | 15,9 | - | 1,3 | - | - | - | - |
| | | T11 | 70 | - | 45 | 75 | - | - | - | 9,35 | - | 2,95 | 7,95 | - | - | - |
| 7 | TS8(TĐ)-LX | T3 | 100 | - | 40 | - | - | - | - | 22,95 | - | 3,1 | - | - | - | - |
| | | T6 | 100 | - | 25 | - | - | - | - | 29,3 | - | 0,4 | - | - | - | - |
| | | T9 | 100 | - | 20 | - | - | - | - | 28,5 | - | 0,9 | - | - | - | - |
| | | T11 | 45 | 70 | 10 | 90 | - | - | - | 5,7 | 6,8 | 0,75 | 8,95 | - | - | - |
| 8 | TS19(TĐ)-CM | T3 | 100 | - | - | - | - | - | - | 10,25 | - | - | - | - | - | - |
| | | T6 | 100 | - | 30 | - | - | - | - | 12,25 | - | 0,5 | - | - | - | - |
| | | T9 | 100 | - | 25 | - | - | - | - | 13,05 | - | 0,4 | - | - | - | - |
| | | T11 | 100 | - | - | - | - | - | - | 12,3 | - | - | - | - | - | - |
| 9 | TS20(TĐ)-TC | T3 | 100 | - | 25 | - | - | - | - | 18,45 | - | 0,6 | - | - | - | - |
| | | T6 | 100 | - | - | - | - | - | - | 19,75 | - | - | - | - | - | - |
| | | T9 | 100 | - | - | - | - | - | - | 20,05 | - | - | - | - | - | - |
| | | T11 | 100 | - | - | - | - | - | - | 18,8 | - | - | - | - | - | - |
| 10 | TS23(TĐ)-PT | T3 | 100 | - | 70 | - | - | - | - | 17,15 | - | 4,9 | - | - | - | - |
| | | T6 | 90 | - | 85 | - | - | - | - | 12,95 | - | 9,95 | - | - | - | - |
| | | T9 | 100 | - | - | - | - | - | - | 24,0 | - | - | - | - | - | - |
| | | T11 | 100 | - | 35 | 85 | - | - | - | 14,9 | - | 1,0 | 6,25 | - | - | - |

The MIR index values for the points affected by urban areas, industrial zones/clusters, and aquaculture areas have been determined. The average MIR values for each season and year indicate that the water quality at the observation points is very poor. This is specifically shown in Table 8, Fig. 7 and Fig. 8. The MIR index has a correlation with the TSS, PO43-, and NH4+-N parameters at significance levels of 1%, 5%, and 5%, respectively. This finding is consistent with studies by [30] and [31], where these variables can play an important role as determinants of the presence of macrophyte species in human-impacted water sources [32].

Table 8. The average MIR values for each season and year at the affected points.

| Time \ Impact point | Urban | Industry | Fisher area |
|---|---|---|---|
| **1 st** | 19,12 | 18,19 | 19,52 |
| **2 sd** | 19,39 | 18,30 | 19,75 |
| **3 rd** | 17,97 | 18,36 | 19,71 |
| **4 th** | 18,31 | 18,09 | 18,84 |
| **Average year** | **18,7** | **18,24** | **19,46** |

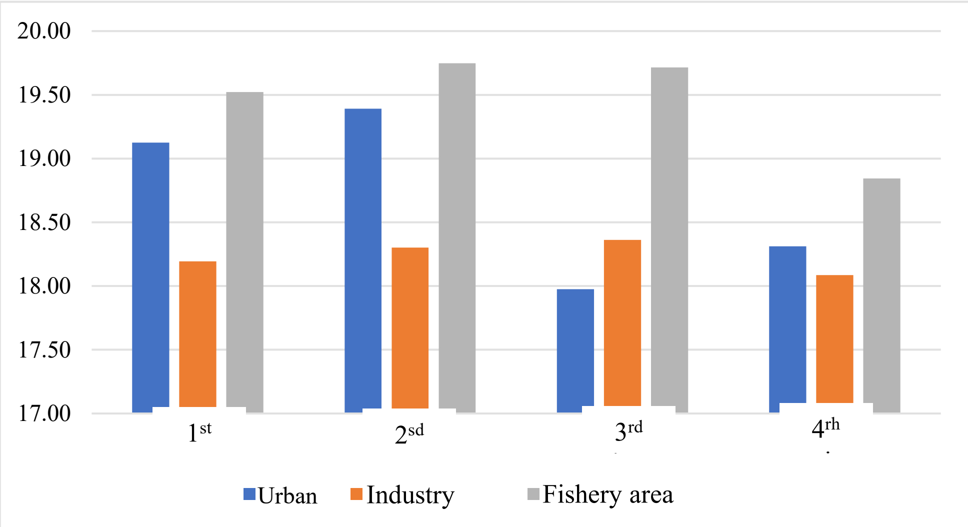



Figure 7. Average MIR value of impact points across each phase

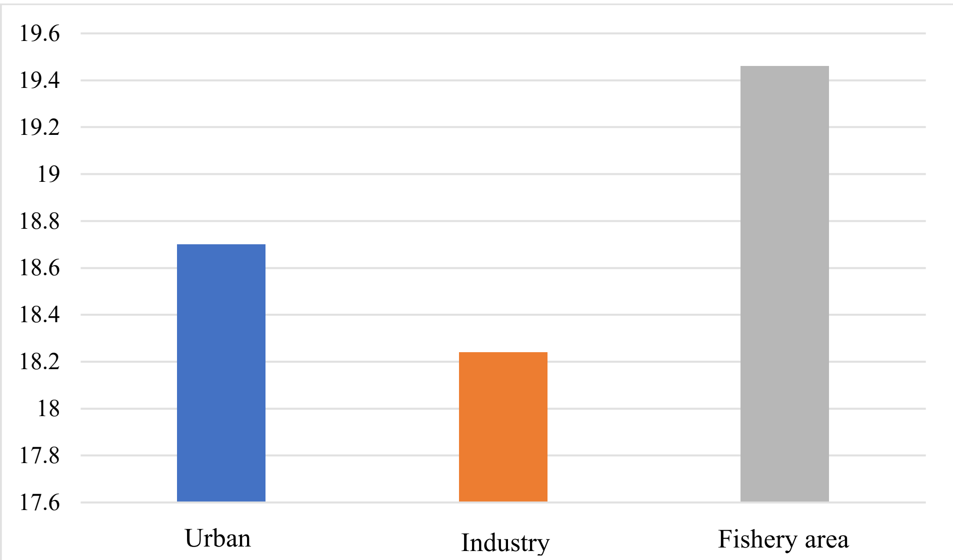



Figure 8. Average MIR value of impact points during the year

### 3.3 Build a predicting model

In this study, the Orange Data Mining Toolbox [33] was used to build and evaluate predictive models for the MIR biological index, as illustrated in Fig. 9. This is an open-source platform with machine-learning algorithms and data visualization capabilities. This system can help create visual data analysis workflows with numerous tools and utilities for users and can handle large datasets.

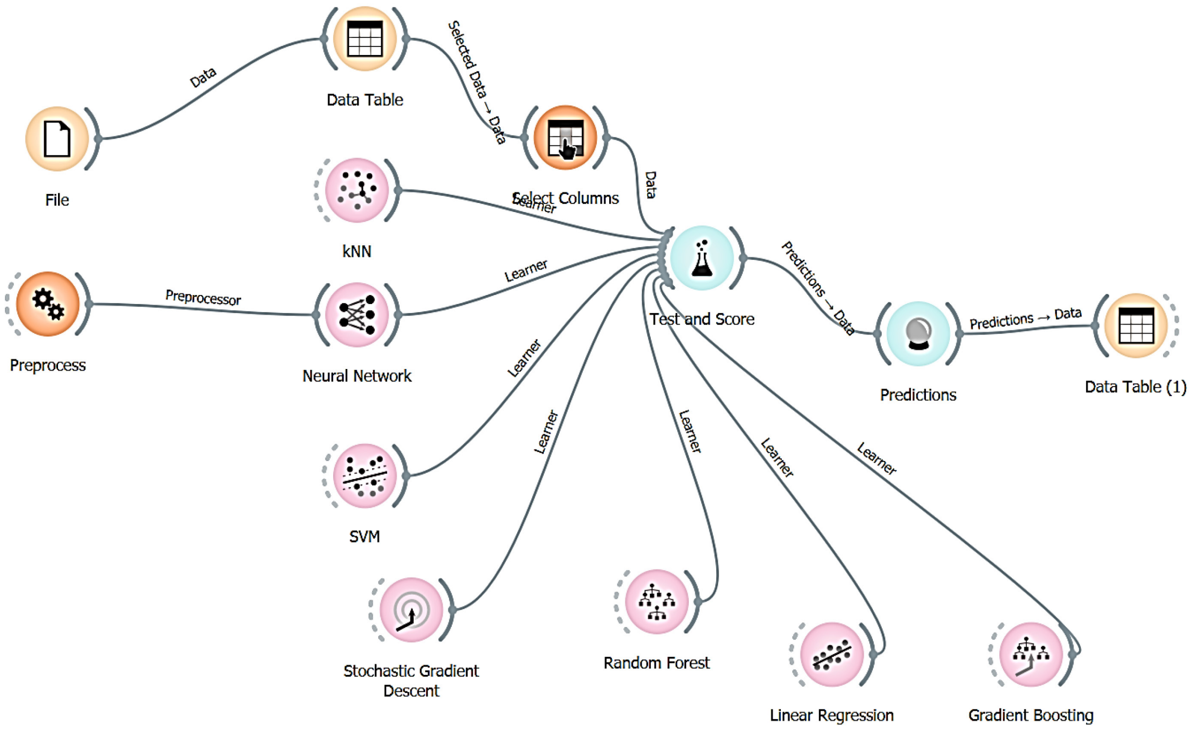



Figure 9. MIR biological indicator prediction model using Orange Data Mining Toolbox

The configuration used to train the machine learning models is as follows: 1) Sampling type: 3-fold cross-validation, 2) Data instances: 68, 3) Features: 4, 4) Meta attributes: 8, 5) Target: Numeric variable 'MIR'. The machine learning models used to predict the MIR index include kNN, Random Forest, Linear Regression, and Gradient Boosting, with the following specific configurations (Table 9):

Table 9. Specific configurations of machine learning models

| Methods | Configuration |
|---|---|
| kNN | Number of neighbor = 5, Distance metric: Euclidean, Weight: Uniform |
| SVM | C=1, epsilon = 0.1, RBF Kernel, Numerical tolerance = 0.001, Iteration limit = 100 |
| SGD | Classification: Hinge, Regression: Squared Loss, Regularization: Ridge (L2), Strength (a) = 0.00001, Learning rate: Constant, Initial learning rate = 0.01, Number of iterations = 1000, Tolerance = 0.001 |
| Random Forest | Number of trees = 10, Limit depth of individual trees = 3, Do not split subsets smaller than 5 |
| Neural Network | Neurons in hidden layers = 500, Activation = ReLu, Solver: Adam, Maximal number of iterations = 1000 |
| Linear Regression | Fit intercept, No regularization, Alpha = 0.0001 |
| Gradient Boosting | Number of trees = 100, Learning rate = 0.1, Limit depth of individual trees = 3, Do not split subsets smaller than 2, Fraction of training instances = 1 |

To evaluate and compare the effectiveness of the machine learning models on this dataset, considering the number of species and the calculated MIR index, we used the following model evaluation metrics: MSE, RMS, MAE, and R2. MSE measures the average of the squared errors or deviations (the difference between the estimated value and the actual value). RMSE is the square root of the average of the squares of a set of numbers (a measure of the degree of imperfection of the fit of the estimator to the data). MAE is used to measure the level of closeness of a prediction to the actual result. R2 is the proportion of the variance in the dependent variable that can be predicted from the independent variable.

Table 10 shows the results of the MSE, RMSE, MAE, and R2 scores for the prediction model with Gradient Boosting, which is the best when compared to other machine learning models. This algorithm typically demands less training time and computational resources compared to deep learning algorithms, rendering them particularly suitable for smaller datasets or scenarios requiring swift model development. Additionally, gradient boosted tree models often exhibit superior accuracy compared to neural networks and greater interpretability than linear models [34]. The R2 ratio of variance in the dependent variable MIR compared to the independent variables is 91.7%. Therefore, the prediction results of this model are highly reliable and can be used to forecast the MIR biological index in practice in An Giang Province. The table provides detailed forecasting figures from the machine learning models using the independent variables: density of Eichornia crassipes, Polygonum hydropiper, Ipomoea aquatica, and Centrostachys Aquatica.

Table 10. Compare the prediction results of machine learning models on data sets at 18 monitoring points in An Giang

| Model | MSE | RMSE | MAE | R2 |
|---|---|---|---|---|
| kNN | 1.105 | 1.051 | 0.752 | 0.210 |
| SVM | 0.572 | 0.756 | 0.515 | 0.590 |
| SGD | 0.505 | 0.711 | 0.571 | 0.638 |
| Random Forest | 0.129 | 0.359 | 0.174 | 0.9075 |
| Neural Network | 5.131 | 2.265 | 1.112 | -2.667 |
| Linear Regression | 0.506 | 0.711 | 0.565 | 0.637 |
| Gradient Boosting | 0.115 | 0.340 | 0.146 | 0.917 |

Table 11 presents the predicted values from the machine learning models at the 18 observation points in An Giang Province. The results show that the machine learning models have successfully learned from the training data and can predict the MIR index. This demonstrates the effectiveness of the machine learning models when applied to real-world prediction tasks at the observation points.

The graphs of the machine learning models trained to predict the MIR biological index are presented in Fig. 10. The results show that the derived curves of the machine learning model closely match the MIR graph. This indicates that these models have effectively captured the data presented in the table with the number of species and the calculated MIR index.

Table 11. Performance of machine learning models on monitoring dataset at 18 points in An Giang Province

| No | MIR | Neural Network | SVM | SGD | kNN | Random Forest | Linear Regression | Gradient Boosting | Eichhornia crassipes | Polygonum tomentosum | Ipomoea aquatica | Centrostachys aquatica |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 18.6667 | 18.5922 | 19.4028 | 19.2884 | 18.9303 | 19.6408 | 19.2911 | 19.8184 | 296 | 12 | 66 | 0 |
| **2** | 20 | 20.1572 | 19.885 | 19.5363 | 19.7831 | 20 | 19.5525 | 19.9982 | 344 | 0 | 13 | 0 |
| **3** | 20 | 19.9862 | 19.7539 | 19.4665 | 19.4302 | 19.9226 | 19.4811 | 19.9997 | 287 | 0 | 26 | 0 |
| **4** | 17.5294 | 18.562 | 17.7207 | 17.411 | 17.8566 | 17.6805 | 17.3726 | 17.5879 | 276 | 0 | 11 | 122 |
| **5** | 20 | 19.7507 | 19.7959 | 19.5496 | 19.5758 | 20 | 19.558 | 19.9982 | 355 | 0 | 71 | 0 |
| **6** | 20 | 19.86 | 19.9293 | 19.6695 | 19.6774 | 20 | 19.6796 | 19.9997 | 453 | 0 | 57 | 0 |
| **7** | 16.9565 | 22.044 | 18.1176 | 17.7168 | 19.2414 | 17.5261 | 17.6618 | 17.1922 | 608 | 136 | 0 | 0 |
| **8** | 17.9545 | 15.7738 | 17.9873 | 18.6202 | 18.4583 | 17.819 | 18.6131 | 17.7938 | 293 | 0 | 16 | 51 |
| **9** | 16.6038 | 14.0625 | 18.2146 | 16.635 | 17.8691 | 17.3972 | 16.5506 | 17.2708 | 124 | 81 | 169 | 81 |
| **10** | 20 | 19.7415 | 19.9495 | 19.6451 | 19.6774 | 20 | 19.6556 | 19.9997 | 433 | 0 | 54 | 0 |
| **11** | 20 | 20.0947 | 19.9457 | 19.6892 | 19.2414 | 20 | 19.707 | 19.9988 | 469 | 0 | 0 | 0 |
| **12** | 20 | 19.7394 | 19.5298 | 19.4004 | 19.7831 | 19.9226 | 19.415 | 20 | 233 | 0 | 27 | 0 |
| **13** | 20 | 19.6406 | 19.4947 | 19.4664 | 18.8745 | 19.9226 | 19.4737 | 19.9997 | 287 | 0 | 80 | 0 |
| **14** | 18.9655 | 18.6997 | 19.4867 | 19.3847 | 19.1472 | 19.1122 | 19.3903 | 18.9256 | 316 | 0 | 69 | 7 |
| **15** | 20 | 20.0618 | 19.8061 | 19.4873 | 19.4302 | 20 | 19.5027 | 19.997 | 304 | 0 | 20 | 0 |
| **16** | 20 | 20.1023 | 19.9418 | 19.677 | 20 | 20 | 19.6948 | 19.9988 | 459 | 0 | 0 | 0 |
| **17** | 20 | 19.9572 | 19.5679 | 19.4237 | 19.6471 | 19.9226 | 19.4419 | 19.9978 | 252 | 0 | 0 | 0 |
| **18** | 20 | 19.8161 | 19.5493 | 19.4175 | 19.6403 | 19.9226 | 19.43 | 20 | 247 | 0 | 42 | 0 |
| **19** | 20 | 20.0827 | 19.91 | 19.8127 | 20 | 20 | 19.8279 | 19.9998 | 570 | 0 | 18 | 0 |
| **20** | 15.8 | 13.4451 | 18.2523 | 14.1162 | 17.7802 | 17.2127 | 13.9792 | 16.7172 | 114 | 136 | 15 | 179 |
| **21** | 20 | 19.9067 | 19.6363 | 19.4347 | 19.4302 | 19.9226 | 19.4518 | 20 | 261 | 0 | 8 | 0 |
| **22** | 20 | 20.1969 | 19.8939 | 19.5987 | 20 | 20 | 19.6166 | 19.9991 | 395 | 0 | 0 | 0 |
| **23** | 20 | 27.7457 | 18.4205 | 19.4319 | 18.9157 | 19.9226 | 19.4231 | 20.0072 | 259 | 0 | 199 | 0 |
| **24** | 18.5714 | 18.4532 | 18.9162 | 19.307 | 18.9593 | 18.3467 | 19.3224 | 17.9571 | 284 | 0 | 88 | 14 |
| **25** | 17.6404 | 15.8079 | 17.6843 | 18.4737 | 18.8799 | 17.8711 | 18.4274 | 17.5786 | 259 | 0 | 23 | 95 |
| **26** | 20 | 19.9687 | 19.8423 | 19.6184 | 19.6774 | 20 | 19.6462 | 20 | 421 | 0 | 58 | 0 |
| **27** | 17.5 | 17.6556 | 18.7199 | 19.1562 | 19.374 | 17.9602 | 19.1415 | 17.691 | 268 | 0 | 0 | 47 |
| **28** | 20 | 19.8443 | 19.7763 | 19.635 | 20 | 20 | 19.6546 | 19.9998 | 249 | 0 | 36 | 0 |
| **29** | 20 | 19.7411 | 19.3252 | 19.5146 | 18.9593 | 20 | 19.5401 | 19.9996 | 312 | 0 | 81 | 0 |
| **30** | 16.9565 | 27.6958 | 18.0199 | 15.723 | 18.6035 | 17.5188 | 15.6783 | 17.1392 | 301 | 213 | 0 | 0 |
| **31** | 18 | 13.6666 | 17.705 | 18.0742 | 19.1022 | 17.5188 | 18.079 | 17.7949 | 356 | 76 | 85 | 0 |
| **32** | 18.0851 | 17.4166 | 17.7899 | 18.3526 | 18.6951 | 17.8711 | 18.3248 | 17.838 | 177 | 0 | 132 | 76 |
| **33** | 18.0851 | 18.156 | 18.1337 | 18.5808 | 19.1602 | 17.8711 | 18.5701 | 17.9475 | 237 | 0 | 146 | 56 |
| **34** | 20 | 19.931 | 19.9713 | 19.6919 | 19.6774 | 20 | 19.7173 | 20 | 408 | 0 | 33 | 0 |
| **35** | 19 | 20.4839 | 19.8023 | 19.6842 | 20 | 19.4254 | 19.6959 | 19.286 | 168 | 0 | 0 | 3 |
| **36** | 18.1818 | 19.6098 | 19.6083 | 19.5363 | 20 | 18.3467 | 19.5402 | 17.957 | 182 | 0 | 0 | 15 |
| **37** | 18.1818 | 19.8335 | 19.664 | 19.5728 | 20 | 18.3467 | 19.5786 | 17.948 | 177 | 0 | 0 | 12 |
| **38** | 20 | 19.7877 | 19.1695 | 19.457 | 19.0961 | 20 | 19.4851 | 19.9996 | 337 | 0 | 102 | 0 |
| **39** | 20 | 19.9586 | 20.0059 | 19.7329 | 19.6774 | 20 | 19.7579 | 20 | 421 | 0 | 21 | 0 |
| **40** | 16.2069 | 12.5687 | 18.5693 | 18.6916 | 20 | 17.9249 | 18.6786 | 18.133 | 468 | 31 | 0 | 42 |
| **41** | 20 | 20.1003 | 19.8844 | 19.6866 | 19.5762 | 20 | 19.7081 | 20 | 318 | 0 | 26 | 0 |
| **42** | 20 | 20.4068 | 19.89 | 19.7333 | 20 | 20 | 19.7485 | 19.9998 | 205 | 0 | 0 | 0 |
| **43** | 20 | 20.1649 | 19.8888 | 19.7845 | 19.6774 | 20 | 19.8061 | 20 | 376 | 0 | 0 | 0 |
| **44** | 20 | 19.8155 | 19.2232 | 19.4712 | 19.0961 | 20 | 19.4992 | 19.9996 | 343 | 0 | 98 | 0 |
| **45** | 20 | 20.0201 | 19.9597 | 19.8156 | 20 | 20 | 19.8412 | 20 | 480 | 0 | 0 | 0 |
| **46** | 17.6404 | 17.0757 | 17.6139 | 18.1142 | 18.0142 | 17.6844 | 18.0486 | 17.7948 | 298 | 0 | 20 | 125 |
| **47** | 20 | 19.8254 | 19.6561 | 19.6676 | 19.7931 | 20 | 19.6839 | 20.0002 | 369 | 0 | 109 | 0 |
| **48** | 17.8788 | 17.9322 | 19.4881 | 19.3085 | 19.2407 | 18.2561 | 19.3161 | 18.3355 | 308 | 8 | 80 | 12 |
| **49** | 18.2353 | 15.7899 | 19.4891 | 19.2135 | 20 | 18.0966 | 19.2196 | 18.1419 | 286 | 28 | 12 | 0 |
| **50** | 18.9157 | 19.4316 | 19.7674 | 19.4632 | 19.7333 | 18.8093 | 19.4727 | 18.9704 | 292 | 0 | 26 | 8 |
| **51** | 17.2 | 11.5502 | 17.9209 | 18.2739 | 20 | 17.4975 | 18.2609 | 17.3525 | 291 | 96 | 0 | 0 |
| **52** | 18.3871 | 18.4138 | 19.4915 | 19.3141 | 20 | 18.2313 | 19.3146 | 18.1762 | 403 | 0 | 48 | 19 |
| **53** | 20 | 19.8625 | 19.938 | 19.6684 | 19.2414 | 20 | 19.675 | 20.0002 | 482 | 0 | 10 | 0 |
| **54** | 18.6364 | 18.6427 | 19.3101 | 19.4005 | 19.7143 | 18.2561 | 19.4114 | 18.5599 | 312 | 0 | 121 | 14 |
| **55** | 18.0851 | 15.7789 | 18.0273 | 18.3117 | 19.5074 | 17.9114 | 18.2857 | 17.9832 | 339 | 0 | 114 | 75 |
| **56** | 19.2857 | 19.6641 | 19.7722 | 19.5722 | 19.2407 | 19.2409 | 19.5882 | 19.077 | 283 | 0 | 82 | 3 |
| **57** | 20 | 19.6687 | 19.1712 | 19.5526 | 19.0727 | 20 | 19.5695 | 19.9221 | 169 | 0 | 0 | 0 |
| **58** | 20 | 20.1972 | 19.8288 | 19.6733 | 19.2414 | 20 | 19.6785 | 19.9993 | 507 | 0 | 0 | 0 |
| **59** | 20 | 20.9454 | 19.6528 | 19.6926 | 19.2414 | 20 | 19.6959 | 19.9993 | 561 | 0 | 0 | 0 |
| **60** | 20 | 20.1372 | 19.8491 | 19.6701 | 19.2414 | 20 | 19.6756 | 19.9993 | 498 | 0 | 0 | 0 |
| **61** | 17.4074 | 22.3486 | 17.9996 | 17.0522 | 17.5341 | 17.7819 | 16.9971 | 17.8931 | 59 | 0 | 89 | 139 |
| **62** | 17.5862 | 21.5141 | 17.9737 | 16.7268 | 17.7961 | 17.7819 | 16.6538 | 17.8908 | 187 | 0 | 59 | 159 |
| **63** | 20 | 19.9472 | 19.8507 | 19.6809 | 20 | 20 | 19.6914 | 20.0002 | 459 | 0 | 62 | 0 |
| **64** | 20 | 21.0837 | 19.5714 | 19.7048 | 19.2414 | 20 | 19.7076 | 20.0002 | 586 | 0 | 8 | 0 |
| **65** | 20 | 19.8574 | 19.7528 | 19.5837 | 20 | 20 | 19.5986 | 19.9998 | 245 | 0 | 10 | 0 |
| **66** | 20 | 19.869 | 19.6595 | 19.5801 | 20 | 20 | 19.5943 | 19.9989 | 246 | 0 | 0 | 0 |
| **67** | 20 | 20.0412 | 20.0446 | 19.6288 | 20 | 20 | 19.6395 | 20.0002 | 369 | 0 | 12 | 0 |
| **68** | 20 | 20.1205 | 19.9502 | 19.6355 | 20 | 20 | 19.6443 | 19.9993 | 401 | 0 | 0 | 0 |

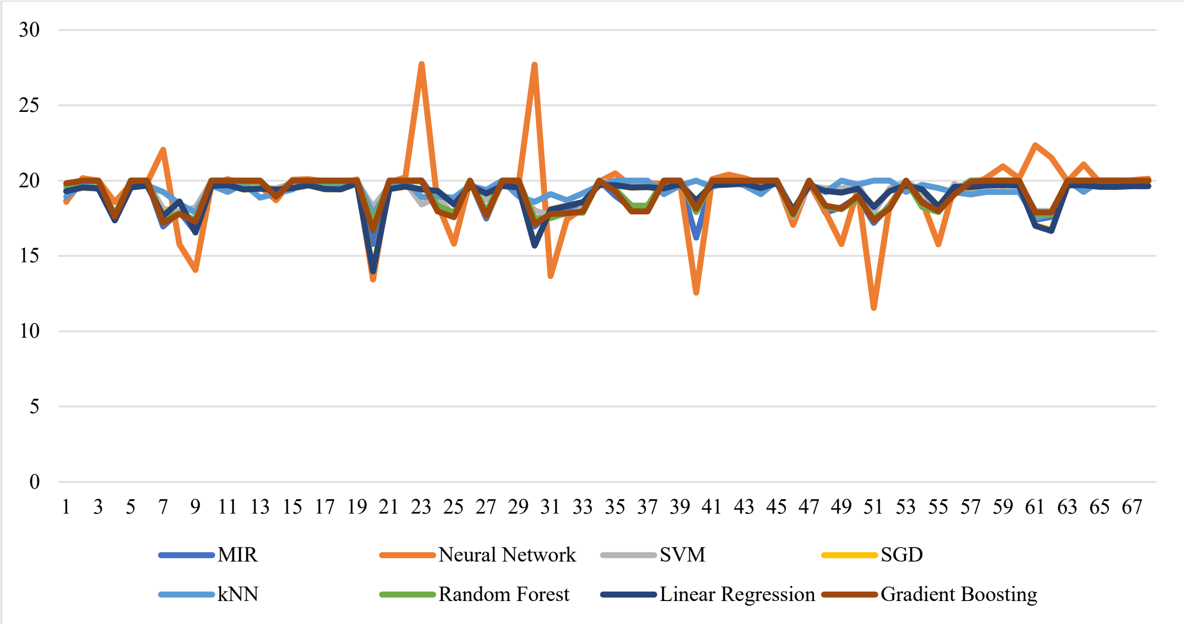



Fig. 10. Graph of MIR index prediction results of machine learning models

Fig. 11 presents a visual map of the predicted biological index at 18 observation points in An Giang Province. This helps users easily observe, compare, evaluate, and suggest appropriate solutions. The star symbols indicate the locations of the MIR pollution index, with darker red colors indicating higher levels of pollution, corresponding to lower water quality.

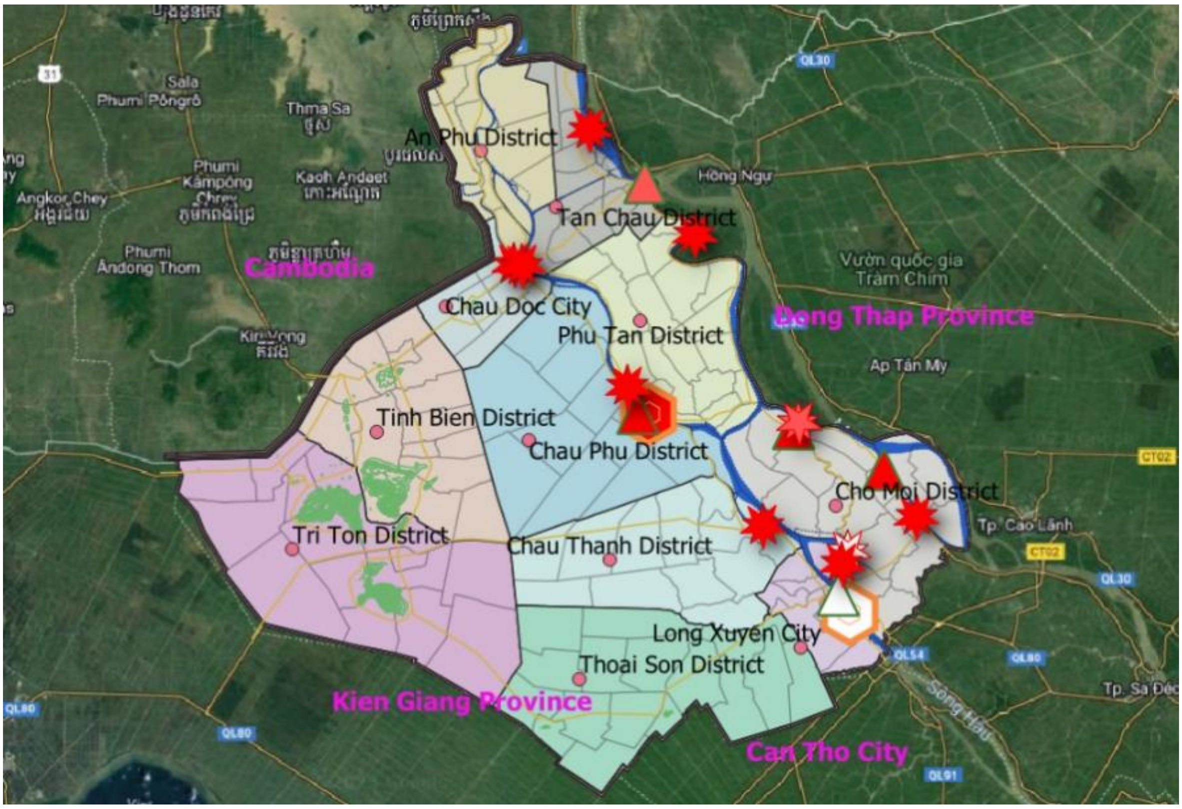



Figure 11. Map showing the results of MIR biological indicator forecasting in An Giang province.

## 4. CONCLUSION

This research collected and established a surface water quality monitoring database at 18 observation points. The results of surface water quality monitoring affected by urban areas, industrial clusters, and aquaculture areas in 2021 showed that 7 out of 14 parameters exceeded the allowable standards, including DO, TSS, COD, BOD5, PO43--P, NH4+-N, and Coliform; while the remaining parameters were within limits specified in the QCVN 08-MT:2015/BTNMT, column A1 - National technical standards for surface water quality. Four floating aquatic plant species were identified: Water Hyacinth, Water Mimosa, Water Spinach, and Water Mimosa. There was no presence of lotus, water lily, or water hyacinth. Among them, water hyacinth was the most dominant, present at all times of the year at all observation locations, while water spinach and water mimosa appeared relatively infrequently and mainly in scattered locations. The MIR index at points affected by urban areas, industrial clusters, and aquaculture areas was determined, and the average MIR values for each period and year indicated that water quality at the monitoring points was very poor. The results also showed that the calculated MIR index had a significant correlation with TSS, PO43-, and NH4+-N at significance levels of 1%, 5%, and 5%, respectively. As a result, this study constructed a GIS map of surface water monitoring, the number of plant species, the MIR index, and a prediction model for the MIR biological index with an accuracy of over 91.7%. This helps users easily observe, compare, evaluate, predict, and propose suitable solutions for managing the quality of surface water in An Giang Province.

## ACKNOWLEDGEMENT

The authors would like to thank An Giang University and Vietnam National University in Ho Chi Minh City, Vietnam.

## REFERENCES

[1] An Giang Center for Environmental Monitoring and Techniques Resources. (2019) Report on results of environmental monitoring in An Giang province November, 2019, People's Committee of An Giang province

[2] Nguyen, K.; Yoshiaki, K.; Xiao, L.; Endo, R.; Shibuya, T. (2015) Microalgae Culture with Digestate from Methane Fermentation, *Eco-Engineering*, Vol. 27, No. 1, 7–11. doi:10.11450/seitaikogaku.27.7

[3] People's Committee of An Giang province. (2019). An Giang Department of Natural Resources and Environment. (2020). Report on environmental status of An Giang province in the period 2016-2020.

[4] Nguyen, V. P.; Nguyen, K. T. T.; Ton, L. T.; Nguyen, D. T.; Nguyen, K. Q.; Vu, M. T.; Tran, H. N. (2020) Dual-Electronic Nanomaterial (Synthetic Clay) for Effective Removal of Toxic Cationic and Oxyanionic Metal Ions from Water, *Journal of Nanomaterials*, Vol. 2020, 1783749. doi:10.1155/2020/1783749

[5] An Giang province monitoring center. (2019) Report on environmental monitoring results of An Giang province

[6] Demars, B. O. L.; Edwards, A. C. (2009) Distribution of aquatic macrophytes in contrasting river systems: A critique of compositional-based assessment of water quality, *Science of The Total Environment*, Vol. 407, No. 2, 975–990. doi:https://doi.org/10.1016/j.scitotenv.2008.09.012

[7] Szoszkiewicz, K.; Jusik, S.; Pietruczuk, K.; Gebler, D. (2020) The macrophyte index for rivers (MIR) as an advantageous approach to running water assessment in local geographical conditions, *Water (Switzerland)*, Vol. 12, No. 1. doi:10.3390/w12010108

[8] Le Hoang Tu, Phan Thi Ha, Vo Ngoc Quynh Tram, Nguyen Ngoc Thuy, Dang Nguyen Dong Phuong, Tran Thong Nhat, N. K. L. (2023) GIS Application in Environmental Management: A Review, *VNU Journal of Science: Earth and Environmental Sciences,* Vol. 39, No. 2

[9] Tran Triet, N. P. B. (2002) *Correlation between hydrophytes and water environment at Tram Chim National Park, Dong Thap*

[10] Dũng, P.; Tin, H.; Pháp, T. (2013) Species composition and distribution of mangrove plants in lap an swamp, phu loc district, Thua Thien Hue province, *Hue University Journal of Science: Agriculture and Rural Development*, Vol. 75. doi:10.26459/jard.v75i6.3157

[11] Najah Ahmed, A.; Binti Othman, F.; Abdulmohsin Afan, H.; Khaleel Ibrahim, R.; Ming Fai, C.; Shabbir Hossain, M.; Ehteram, M.; Elshafie, A. (2019) Machine learning methods for better water quality prediction, *Journal of Hydrology*, Vol. 578, 124084. doi:https://doi.org/10.1016/j.jhydrol.2019.124084

[12] Shams, M. Y.; Elshewey, A. M.; El-kenawy, E.-S. M.; Ibrahim, A.; Talaat, F. M.; Tarek, Z. (2023). Water quality prediction using machine learning models based on grid search method, *Multimedia Tools and Applications*. doi:10.1007/s11042-023-16737-4

[13] Azrour, M.; Mabrouki, J.; Fattah, G.; Guezzaz, A.; Aziz, F. (2022) Machine learning algorithms for efficient water quality prediction, *Modeling Earth Systems and Environment*, Vol. 8, 1–9. doi:10.1007/s40808-021-01266-6

[14] McCulloch, W. S.; Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, 115–133. doi:https://doi.org/10.1007/BF02478259

[15] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression, *The American Statistician*, Vol. 46, No. 3, 175–185

[16] Cortes, C.; Vapnik, V. (1995) Support-vector networks, *Machine Learning*, Vol. 20, No. 3, 273–297

[17] Ruder, S. (2016) An overview of gradient descent optimization algorithms, *ArXiv Preprint ArXiv:1609.04747*

[18] Ho, T. K. (1995) Random decision forests, *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 1), 278–282

[19] Friedman, J. H. (2001) Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 1189–1232

[20] Lipps, W. C., Baxter, T. E., & Braun-Howland, E. (n.d.). Standard methods for the examination of water and wastewater, *Standard Methods For the Examination of Water and Wastewater*. doi:10.2105/SMWW.2882.030

[21] Nguyen, L.; Linh, N.; Thịnh, Đ.; Hang, P.; Nguyen Vo Chau, N.; Thu, V. (2015) Evaluation of treatment efficiency of domestic wastewater by aquatic plants, *Can Tho University Journal of Science*

[22] Khôi, C. M.; Dũng, N. V. C.; Nhiên, C. T. (2012) Ability to treat dissolved nitrogen and organic phosphorus pollution in pangasius pond wastewater by Luc Binh (Eichhorinna crassipes) and Vertiver grass (Vertiver zizanioides), *Can Tho University Journal of Science*, 151–160

[23] Ceschin, S.; Ferrante, G.; Mariani, F.; Traversetti, L.; Ellwood, N. T. W. (2020) Habitat change and alteration of plant and invertebrate communities in waterbodies dominated by the invasive alien macrophyte Lemna minuta Kunth, *Biological Invasions*, Vol. 22, No. 4, 1325–1337. doi:10.1007/s10530-019-02185-5

[24] Nekos, Alla and BoiarynMariia and Lugowska, Maria and Tsos, Oksana and Netrobchuk, I. (2021) Assessment of the ecological condition of the Western Bug river basin according to the macrophyte index for rivers (MIR), *Visnyk of V.N. Karazin Kharkiv National University, Series Geology. Geography. Ecology*

[25] Ismail, M. H. (2010) Determining and Mapping of Vegetation using GIS and Phytosociological Approach in Mount Tahan, Malaysia, *Journal of Agricultural Science*, Vol. 2, No. 2. doi:10.5539/jas.v2n2p80

[26] Nguyen Quang Tuan; Trần Văn No; Do Thi Viet Huong. (2010) Applying GIS and remote sensing in creating a map of the current state of vegetation in 2008 at a scale of 1:50,000 in Ky Anh district, Ha Tinh province, *Hue University Journal of Science*

[27] Bareth, G.; Waldhoff, G. (2017) GIS for Mapping Vegetation, *Reference Module in Earth Systems and Environmental Sciences*. doi:10.1016/B978-0-12-409548-9.09636-6

[28] Đỗ Huy Bích; Đặng Quang Chung; Bùi Xuân Chương; Nguyễn Thượng Dong; Đỗ Trung Đàm; Phạm Ngọc Hiển; Vũ Ngọc Lộ; Phạm Duy Mai; Phạm Kim Mãn; Đoàn Thị Nhu; Nguyễn Tập; Trần Toàn. (2006) *Medicinal Plants and Medicinal Animals in Vietnam (Volume 1)*, Nhà xuất bản Khoa học và Kỹ thuật, Hà Nội

[29] Đan, T. H.; Duy, N. P.; Thọ, B. T. (2012) The distribution of aquatic plants in the organic polluted canals in Can Tho City in rainy season, *Tạp Chí Khoa Học Đại Học Cần Thơ*, No. 23a, 283–293

[30] Bytyçi, P.; Shala-Abazi, A.; Zhushi-Etemi, F.; Bonifazi, G.; Hyseni-Spahiu, M.; Fetoshi, O.; Çadraku, H.; Feka, F.; Millaku, F. (2022) The Macrophyte Indices for Rivers to Assess the Ecological Conditions in the Klina River in the Republic of Kosovo, *Plants*. doi:10.3390/plants11111469

[31] Tarkowska-Kukuryk, M.; Grzywna, A. (2022) Macrophyte communities as indicators of the ecological status of drainage canals and regulated rivers (Eastern Poland), *Environmental Monitoring and Assessment*, Vol. 194, No. 3, 210. doi:10.1007/s10661-022-09777-0

[32] Bytyçi, P.; Shala-Abazi, A.; Hyseni Spahiu, M.; Feka, F.; Fetoshi, O.; Bonifazi, G.; Millaku, F. (2021) *The Macrophyte Index for Rivers (MIR) to Running Water Assessment in River Klina to the Republic of Kosovo*

[33] Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; Štajdohar, M.; Umek, L.; Žagar, L.; Žbontar, J.; Žitnik, M.; Zupan, B. (2013) Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research*, Vol. 14, 2349–2353

[34] Lundberg, S.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. (2020) From Local Explanations to Global Understanding with Explainable AI for Trees, *Nature Machine Intelligence*, Vol. 2. doi:10.1038/s42256-019-0138-9