Ontology-assisted Expert System for Algae Identification with Certainty Factor

Foni A. Setiawan1,2, Reny Puspasari3, Lindung P. Manik4, Zaenal Akbar1, Yulia A. Kartika1, Ika A. Satya1, Dadan R. Saleh3, Ariani Indrawati1, Keiji Suzuki4, Hatim Albasri3, Masaaki Wada4

1Research Center for Informatics, National Research and Innovation Agency, Bandung 40135, Indonesia
2Department of Informatics Engineering, Universitas Ibn Khaldun, Bogor 16164, Indonesia
3Center for Fisheries Research, Ministry of Marine Affairs and Fisheries, Jakarta 10041, Indonesia
4Future University Hakodate, Hakodate, Japan

Corresponding author: Foni A. Setiawan (e-mail: masagus@masagus.id).

This work was supported in part by the Japan International Cooperation Agency, Japan Science Technology Agency, and the Indonesian Ministry of Marine Affairs and Fisheries under SATREPS Mariculture Project Grant “Optimizing Mariculture based on Big Data with Decision Support System.”

ABSTRACT Harmful Algal Bloom (HAB) is one of nature’s responses to nutrient enrichment in aquatic systems and increasingly occurs in coastal waters, such as in Lampung Bay and Jakarta Bay, Indonesia. HABs present environmental and fisheries management challenges due to their unpredictability, spatial coverage, and detrimental health effects to coastal organisms, including humans. Here, we propose an automated algae species identification system assisted and validated by expert judgment. The system uses ontology as guidance to determine the species of algae and certainty factors to indicate the level of confidence of the experts when providing a statement or judgment for a particular object or event under consideration. We tested the system to identify 60 sample data using 51 predetermined algal characteristics. The tests were narrowed down to the 20 most common HAB-causing algae types found in the study sites and compared with identification by the experts. The results showed that the system has successfully identified the test data with an accuracy of 73.33%. The system also has a high agreement (above 79.75%) with the identification performed by the experts on six algae species. Further improvement on the system’s accuracy could facilitate its use as an alternative tool in rapid algal identification or part of an early warning system for HABs.

INDEX TERMS Harmful algal bloom, algae identification, ontology, certainty factor, expert system.

I. INTRODUCTION

Algal blooms and red tides have become common in many saline or freshwater systems worldwide, including Indonesia. Both are considered a phenomenon in which one or more algae/phytoplanktons increase rapidly in the water column. The bloom takes on a red or brown color depending on the pigment contained in the phytoplankton [1]. Some causative red tide species are harmful and produce toxins that can directly kill the organism affected or be transferred through the food web chain from which they retain the name as harmful algal blooms or HABs [2]. Besides toxic, HABs can also cause respiratory damage to aquatic organisms due to the build-up of algal or its remains in the respiratory system and suffocating hypoxic conditions.

HABs have become a crucial environmental issue in the last decade, particularly for aquaculture industries, due to their impact on fish culture. Between 1990–2015, at least 23 cases of HABs were recorded in Indonesia waters distributed from Java Sea, Jakarta Bay, Ambon Bay, Lampung Bay, and West Sumatera [3]. However, despite HABs being frequently reported by the media, research institutions, and government, HABs prevention or efficient mitigation to avoid or reduce losses to fish farmers or mass fish mortality has yet to be seen. The outbreak of HABs is usually related to changes in environmental conditions. Some physical and chemical water parameters induce the rapid growth of HABs species. Some scientists argued that the apparent increase in harmful blooms is strongly linked to eutrophication in aquatic systems [4]. In addition, the enrichment provides a vast
quantity of nutrients needed by phytoplankton to grow. Therefore, detecting the changes in environmental conditions could be used as an early warning system for HABs prevention or mitigation.

Furthermore, identifying the harmful impacts of HABs is also another mitigation approach. For example, various studies reported the detrimental effects of cyanobacteria (blue-green algae), predominantly found in eutrophic fresh and brackish waters, to aquatic organisms. In addition, some cyanobacteria species produce microcystins, a group of chemically stable cyclic peptide toxins responsible for poisoning cases of animals and humans [5]. Thus, knowledge on the characteristics of abundance and poisoning potential of algae to other living organisms will feed critical information to prevent HABs or mitigate their effects successfully.

Various species of algae have been identified to cause HABs. At least 32 [6], 37 [7], and 28 [8] causative species of algae were reported to cause HABs events in the Yellow Sea, South Chinese Sea (SCS), and Southern SCS, respectively. However, some causative species of HABs are still unidentified or at least debated their identification accuracies [8]. As such, studies on increasing the accuracy and rapid identification of HABs-caused species have been proposed. For example, [9] developed a framework for automatic identification for one species of algae, namely *Neomeries vanbosseae M. A. Howe* (NVH), which belongs to the Chlorophyta division. In another study, [10] proposed morphological processing and fluorescence imaging for phytoplankton identification. The study utilized a database of the shape of each algal. Similarly, [11] developed a novel multi-target deep learning framework for algal detection and classification. The study collected a large-scale colored microscopic algae dataset and applied extensive identification experiments based on the dataset. Lastly, [12] explained a critical development direction for the algae classification and determination based on image analysis of biological-morphological differences. However, this study had to deal with some challenges, such as the high degree of sample imbalance and the difficulty of formalizing the description of local physiological features.

These previous studies had relied on identifying algae species based on image analysis and required large datasets processed using machine learning (ML) techniques. An ML system focuses on statistical modeling, which entails creating a recognition model using the existing training data and then using that model to infer conclusions from new observations. For example, in a support vector machine (SVM) or neural network (NN), the modeling might be made up of differentiable parameters like weights and biases. In order to minimize an error/cost function, systems are usually optimized using gradient descent-based optimization methods. This optimization technique achieves minimal error surface. Nevertheless, the result might not be easy to explain, particularly in artificial neural network-based systems. The final result is a functional system for inference with no explanation, which is problematic in some circumstances [13]. For example, a study by [14] demonstrated that the SVM performed poor classification when inferencing outside its training set. Furthermore, SVM and NN require large datasets of hundreds or thousands of images, which bears a high collection cost.

A type of artificial intelligence (AI), the knowledge-based system (KBS), offers a better solution which SVM and NN lack thereof. KBS uses rules to do induction and draws a judgment based on facts or knowledge and rules. The drawback of KBS is in generating a knowledge base (KB), which may be time-consuming [15]. However, the model’s advantage is that it provides more information on the reason. The KB serves as ground truth, which can be used for various purposes, such as building a faceted search [16] or even a knowledge-based chatbot [17]. Furthermore, the system can explain its logical “thinking” process and why it got to that particular conclusion. To the best of our knowledge, this study is the first of its kind in identifying algal species based on their morphology using KBS in the form of an ontology-assisted expert system coupled with certainty factors.

Currently, most laboratories use labor-intensive manual identification techniques by comparing the observed algal cell morphology using a light microscope with the closest resembling images in plankton printed books or electronic databases. The accuracy of this manual identification process varies depending on the analyst’s level of expertise and experience. The method is also a tedious job and time-consuming due to one by one comparison of morphological characteristics.

An algal bloom outbreak does not suddenly occur. When algae bloom does occur, the growth rate of algae increases rapidly. However, observable signs of environmental precursors and algal bloom developmental stages can be easily recognized, such as gradually changing watercolor. By determining the causative algal bloom species as early as possible before reaching the population peak, the negative impacts of algal blooms, mainly harmful ones, can be prevented or mitigated. The development of a rapid identification system is primarily needed to shorten the algae identification process in the laboratory. Rapid species identification results can allow managers to make the right and on-time decisions, particularly in the mariculture industry. Therefore, we propose an expert system to assist researchers and laboratory technicians in rapidly identifying algae based on their morphological characteristics. The system uses ontology to guide the identification process in determining the type of algae coupled with a certainty factor indicating its confidence level.

The rest of the paper is organized as follows. First, in Section II, we describe the data and the research methodology. Then, we explain our findings, including the results and the discussions in Section III and Section IV, respectively. Finally, the conclusion and the future works are given in Section V.
II. MATERIALS AND METHODS

This section describes materials and methods used in this research, including study sites and data, knowledge base, inference engine, dialog interface, and the design of the experiment.

A. STUDY SITES AND DATA

This study took place in two locations, namely Lampung Bay and Jakarta Bay. Lampung Bay is located in Lampung province and adjacent to the Sunda Strait, Indonesia. A shallow water area characterizes the bay with an average depth of 20 meters, strongly influenced by anthropogenic activities from the surrounding dense settlements. Tidal currents dominate its water exchange where water mass from the Indian Ocean enters the bay through the east coast then directly turns to the west side at the tip of the bay. Current velocity decreases on the west coast of the bay due to coastal morphological conditions, where there are some small bays and small islands along the west coast [18]. As a result, it accumulates nutrient concentrations and generates frequent algae blooms during certain environmental conditions [19].

Jakarta Bay is located in the north of the Jakarta province, Indonesia. It is shallow water with an average depth of 12.5 m, affected by the water mass from the Java Sea, where the monsoon strongly influences the movement of the water mass. Similar to Lampung Bay, Jakarta Bay is also strongly influenced by anthropogenic activities.

Lampung Bay and Jakarta Bay received high input of nutrients from land activities that generate frequent algae blooms [20]. As a result, the trend of algal blooms in Jakarta Bay has increased since 2000. There are 16 species of algae identified as blooming causative [21], where most of the abundant species found are Skeletonema and Chaetoceros [22]. On the other side, since 2012, there has been an increase in the frequency of red tides in Lampung Bay caused by the blooming of Chaetoceros polykrikoides algae [23]. The data used in this study is algae data obtained from the two bays’ water taken from 2017 to 2019. Fig. 1 shows a map of the research sites located in Lampung Bay and Jakarta Bay.

![Map of research location: Lampung Bay and Jakarta Bay.](image)

B. KNOWLEDGE BASE

We used ontology as the knowledge base of the expert system. Apart from being a taxonomic reference of the species and genera understudy, ontology provides axioms as boundaries to give the characteristics of the object. For example, *Alexandrium* is a genus of algae with a flagella length ranging between 16 and 55 µm. Therefore, when the identification process is carried out, if there are algae that do not have flagella or have a length outside this range, the system will refuse to classify the algae into the genus *Alexandrium*. Fig. 2 shows the ontology design process using the Protégé editor.

Another benefit of using an ontology in this expert system is to assist in the “pruning” process. Ontology is used to guide the dialogue so that there is no waste of questions given to users. It is done by narrowing down the candidate questions that users ask based on the previous questions. This pruning technique will be explained later in Section II.C.

This study proposed a certainty factor (CF) to determine how confident an expert is in the assertion that an algal under consideration has specific characteristics. CF is used when expert judgment is needed on a statement in the form of object/event features: how sure an object/event has specific features. Two variables must be obtained from the experts, namely the measure of belief (MB) and the measure of disbelief (MD). The function to calculate the certainty factor, \( CF \), of the hypothesis \( H \) given evidence \( E \), is defined in the equation (1), where:

\[
MB(H, E) = \text{Measure of belief to the hypothesis } H \text{ given evidence } E \text{ (between 0 and 1); and}
\]

\[
MD(H, E) = \text{Measure of disbelief to the hypothesis } H \text{ given evidence } E \text{ (between 0 and 1).}
\]

\[
CF(H, E) = MB(H, E) - MD(H, E)
\]  

(1)
Certainty factor, \( CF \), not only uses values to assume an expert’s level of confidence but can also use values from initial facts found by users [24]. For this case, the \( CF(H, E) \) is defined in equation (2), where:

\[
CF(E) = \text{the CF value of evidence } E \text{ determined by the user; and}
\]

\[
CF(Rule) = \text{the CF rule value determined by the expert.}
\]

\[
CF(H, E) = CF(E) \times CF(Rule)
\]

Equation (3) combines two or more CFs into one CF that represents these CF values, where:

\[
CF_{comb}(CF1, CF2) = \text{Combined certainty factor, also called as the confidence level (CL);}
\]

\[
CF1 = \text{Certainty factor 1; and}
\]

\[
CF2 = \text{Certainty factor 2.}
\]

\[
CF_{comb}(CF1, CF2) = CF1 + CF2 \times (1 - CF1)
\]

We used a relational database to accommodate the need for certainty factor calculations in the developed expert system, with a structure shown in Fig. 3. The \textit{object_features} table connects the \textit{objects} table with the \textit{features} table, which contains the object relations with its features along with its MB and MD values.
the screenshot of the object_features table with the processed MB and MD values.

FIGURE 4. Screenshot of the object_features table content.

C. INFEREN CE ENGINE

The application begins by displaying a list of algae images. The user is welcomed to choose an image of the algae that closely resembles the algal to be identified. Next, the system asks the user a question about the characteristics of the algal to be identified. The user only needs to answer ‘Yes’ or ‘No’ whether the characteristic is present in the algal in question. The system processes the answer from the user and considers it as input for the question to be asked next. The given answers by the user become provisions for the system to select the candidate objects. The application continues in this loop until one of three conditions is met: (1) the object in question is identified; (2) the object in question cannot be identified; or (3) the user cancels the consultation session. The flowchart of the inference engine is presented in Fig. 5.

It can be seen from the flowchart that three critical processes guide the application in identifying objects. The first is finding the next feature asked in a consultation session with the user (fetchNextQuestion). The second is getting the selected candidate object(s) based on the questions asked by the system and the answers given by the user (getObjectCandidates). The third is calculating the CF value of the selected object given the characteristics that the user determines. The calculation of this CF value is carried out using equation (1), described in Section II.B.

The fetchNextQuestion function takes three parameters: the candidate objects for the algae to be identified, the features asked by the user, and the features selected as the characteristics of the object by the user. The main principle of this function is retrieving the next feature to be asked and getting the minimum degree of feature similarity of the candidate objects. It is done so that when the user answers ‘Yes’ to a feature asked for, the system can eliminate candidate objects that do not have that feature as quickly as possible.

FIGURE 5. Flowchart of the inference engine.

Let say there is an object set \( O \) and a feature set \( F \) as follows:

\[ O = \{A, B, C\} \]

\[ F = \{F_1, F_2, F_3, F_4, F_5, F_6\} \]

\( F_N \) represents the set of features that object \( N \) has, for example:

\[ F_A = \{F_1, F_2, F_3, F_4, F_5, F_6\} \]

\[ F_B = \{F_2, F_3, F_4, F_5, F_6\} \]

\[ F_C = \{F_3, F_4, F_5, F_6\} \]
Then, to determine the next feature asked of the user, we have to search for the features that have the slightest similarity between the three objects. If $\varsigma(F_1)$ is the set of feature similarity of $F_1$ (the set of objects that have the $F_1$ feature) and $\delta(F_1)$ is its degree (the number of objects that have this feature), we obtained that the $F_1$ feature was chosen for the next question. The algorithm for the $\text{fetchNextQuestion}$ function is as follows.

$$\varsigma(F_1) = \{F_A\} \Rightarrow \delta(F_1) = 1$$

$$\varsigma(F_2) = \{F_A, F_B\} \Rightarrow \delta(F_2) = 2$$

$$\varsigma(F_3) = \{F_A, F_B, F_C\} \Rightarrow \delta(F_3) = 3$$

$$\delta(F_4) = \delta(F_5) = \delta(F_6) = 3$$

### Algorithm 1 fetchNextQuestion

**Input**

- $\text{ObjectCandidates}$: candidate objects for the algae to be identified
- $\text{AskedFeatures}$: features that have been asked of the user
- $\text{SelectedFeatures}$: features that have been selected as the characteristics of the object. Initial value: all features in the database

**Output**

- $\text{FeatureCandidate}$: feature candidate that will be asked of the user for the next question

**Description**

This function determines the next feature that will be asked of the user for the next question in a dialogue session.

1. begin
2. $\text{FeatureGroups} \leftarrow \text{enumerate feature groups of } \text{SelectedFeatures}$
3. $\text{FeatureCandidate} \leftarrow \emptyset$
4. $\text{Candidates} \leftarrow \text{select features of objects in } \text{ObjectCandidates} \text{ that are not listed in } \text{AskedFeatures} \text{ and do not have feature group listed in } \text{FeatureGroups}$
5. if $\text{Candidates} \neq \emptyset$ then
6. // get the minimum degree of feature similarity
7. if number of elements in $\text{Candidates} = 1$ then
8. $\text{FeatureCandidate} \leftarrow \text{Candidates}[0]$
9. else
10. $\text{FeaturesCount} \leftarrow \emptyset$
11. foreach $\text{obj} \in \text{ObjectCandidates}$ do
12. $\text{Count} \leftarrow 0$
13. foreach $\text{feat} \in \text{Candidates}$ do
14. if object $\text{obj}$ has feature $\text{feat}$ then $\text{Count} \leftarrow \text{Count} + 1$
15. end foreach
16. $\text{FeaturesCount}[\text{feat}] \leftarrow \text{Count}$
17. end foreach
18. sort array $\text{FeaturesCount}$ in ascending order
19. $\text{FeatureCandidate} \leftarrow \text{index of first element in } \text{FeaturesCount}$
20. end if
21. // ---
22. else
23. $\text{FeatureCandidate} \leftarrow \text{null}$
24. end if
25. return $\text{FeatureCandidate}$
26. end

The $\text{getObjectCandidates}$ function retrieves all objects that have the features selected by the user. It filters the current object candidates based on the new set of selected features and an unselected feature (if any) to generate new object candidates. Objects that do not have all the features listed in the selected features will be excluded from the object candidates. Moreover, when the user answers ‘No’ for a dominant feature possessed by an object, the object is removed from the list of object candidates. The algorithm of this function is as follows.
Algorithm 2 getObjectCandidates

Input
SelectedObjects: current selected objects
SelectedFeatures: selected features
UnselectedFeature: an unselected feature

Output
Candidates: object candidates

Description
This function filters objects that have all the features listed in SelectedFeatures and do not have a dominant UnselectedFeature from current SelectedObjects.

begin
Candidates []
Matrix []
if SelectedObjects = [] then SelectedObjects All objects in the objects table
if SelectedFeatures ≠ [] then
foreach (obj in SelectedObjects) do
foreach (feat in SelectedFeatures) do
Check whether object obj has feature feat
if obj has feature feat then
Matrix[obj][feat] 1;
else
Matrix[obj][feat] 0;
end if
end foreach
end foreach
foreach (obj in SelectedObjects) do
if sum of Matrix[obj][feat] = number of feat then
push obj into Candidates
end if
end foreach
else
Candidates SelectedObjects
end if
if UnselectedFeature ≠ <empty> and UnselectedFeature not in SelectedFeatures then
foreach (cand in Candidates) do
if cand has feature UnselectedFeature and UnselectedFeature is a dominant feature then
remove cand from Candidates
end if
end foreach
end if
return Candidates
end

D. DIALOG INTERFACE
The expert system was implemented as a web-based application called Algies (Algae Identification Expert System). It can be accessed via a web browser at the URL address http://www.masagus.id/algies/. First entering the application, the user will be shown a welcome screen that contains how many species/genera are included in the database. Then, there is a button to start the consultation and to select the language to be used.

During the consultation session, users are presented with questions generated by the system according to the fetchNextQuestion algorithm described in Section II.C. When the user answers ‘Yes’ that the character is present in the observed algae, or ‘No’ if not, the system selects or excludes object candidates according to the getObjectCandidates algorithm. The consultation continues until the algal in the database or the object can no longer be identified. When the algal is identified, the system then calculates the confidence level of this identification process. At any time, the user can cancel the ongoing consultation session. In addition, we added a facility to evaluate system performance by displaying a feedback form at the end of the consultation session. Fig. 6 shows a screenshot when the user is in consultation with Algies: (a) the system asks about the characteristics of the algae, and (b) the system provides identification results based on the characteristics given.
EXPERIMENTAL DESIGN

Once the expert system is developed, the system is tested to see its performance, i.e., how accurately the expert system provides the correct answer from the characteristics described by the user. We took 60 samples of algae which were distributed evenly among three experts. Algies then tried to identify them one by one. Finally, algae experts verified the identification results. Algae identification accuracy is calculated as the ratio of correct answers (correctly identified objects) to the number of experiments, as defined in equation (4).

\[
\text{accuracy} = \frac{\sum \text{correct}}{\sum \text{experiment}} \quad (4)
\]

Equation (5) is used to calculate the overall confidence level. It is the average of all confidence levels for correctly identified objects, where \( CL \) is the confidence level calculated using equation (3) and \( n \) is the number of correctly identified objects.

\[
CL = \frac{\sum CL}{n} \quad (5)
\]

III. RESULTS

This section describes the results obtained in the research that has been carried out.

A. STUDY SITES AND DATA

Based on the HABs events that occurred in Lampung Bay and Jakarta Bay over the last ten years, the types of algae (species/genera) reported include [21-23]: (1) Nitzschia, (2) Pseudo-nitzschia, (3) Skeletonema costatum, (4) Thalassiosira, (5) Chaetoceros socialis, (6) Noctiluca scintillans, (7) Protoperidinium claudicans, (8) Proorocentrum micans, (9) Gymnodinium catenatum, (10) Gymnodinium sanguineum, (11) Gyrodinium lachryma, (12) Pyrodinium bahamense, (13) Gambierdiscus toxicus, (14) Alexandrium, (15) Ceratium furca, (16) Dinophysis caudata, (17) Protoperidinium, (18) Trichodesmium, (19) Cochlodinium polykrikoides, and (20) Dinophysis acuminata. Therefore, in this study, we focused on the data collected for these 20 species/genera of algae.

| No | Characteristic / Value |
|----|------------------------|
| 1  | Shape                  |
|    | (a) Stretched straight  |
|    | (b) Oval               |
|    | (c) Cylindrical        |
|    | (d) Disc               |
|    | (e) Rounded            |
|    | (f) Cordate            |
|    | (g) Rounded oval       |
|    | (h) Polyhedral         |
|    | (i) Isodiametric       |
|    | (j) Filament           |
|    | (k) Branched           |
|    | (l) Vase               |
|    | (m) Chained like a train |
| 2  | Length                 |
|    | (a) \( \leq 15 \mu m \) |
|    | (b) \( \leq 55 \mu m \) |
|    | (c) \( \leq 80 \mu m \) |
|    | (d) \( \leq 95 \mu m \) |
|    | (e) \( \leq 120 \mu m \) |
|    | (f) \( \leq 230 \mu m \) |
|    | (g) \( > 230 \mu m \)  |
| 3  | Color                  |
|    | (a) No Color (Transparent) |
|    | (b) Tawny              |
|    | (c) Green              |
|    | (d) Brownish Red       |
|    | (e) Brownish Orange    |
|    | (f) Yellowish Green    |
| 4  | Number of Chloroplasts |
|    | (a) 1                 |
|    | (b) 2 or 3            |
|    | (c) > 3               |
| 5  | Cell type              |
|    | (a) Unicellular        |
|    | (b) Multicellular      |
| 6  | has Fluorescent body   |
| 7  | has Raphe             |
| 8  | has Valve             |
| 9  | has Flagella          |
| 10 | has Prickle           |
| 11 | has Arm               |
| 12 | Theca                 |
|    | (a) Theca with cellulose walls (plates) |
|    | (b) No Theca or without cellulose walls (plates) |
| 13 | has Nucleus           |
| 14 | Cingulum              |
|    | (a) Cingulum in the Anterior |
|    | (b) Cingulum around cells |
| 15 | has Horn              |
| 16 | Vacuole               |
|    | (a) One vacuole       |
|    | (b) Two vacuole       |
| 17 | Striae                |
|    | (a) Radial            |
|    | (b) Non Radial        |
| 18 | has Tooth             |
| 19 | has Tentacle          |
| 20 | has Cyst              |
| 21 | has Setae             |
TABLE 2. The completed expert assessment form.

| No | Char | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    | 14    | 15    | 16    | 17    | 18    | 19    | 20    |
|----|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1  | 1(a) | 4.375 | 4.375 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 2  | 1(b) | 3.875 | 8     |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 3  | 1(c) | 5.5   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 4  | 1(d) |       | 5.375 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 5  | 1(e) |       |       | 5.25  |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 6  | 1(f) |       |       |       | 7.625 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 7  | 1(g) |       |       |       |       | 5.75  | 4     | 4.25  | 5.125 | 4.75  |       |       |       |       |       |       |       |       |       |       |       |
| 8  | 1(h) |       |       |       |       |       | 4.875 | 5.75  | 7.25  |       |       |       |       |       |       |       |       |       |       |       |       |
| 9  | 1(i) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 5.5   |
| 10 | 1(j) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 7.5   |
| 11 | 1(k) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 8     |
| 12 | 1(l) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 4     |
| 13 | 1(m) |       |       |       | 8     |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 8     |
| 14 | 2(a) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 6.5   |
| 15 | 2(b) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 16 | 2(c) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 17 | 2(d) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 18 | 2(e) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 19 | 2(f) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 20 | 2(g) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 21 | 3(a) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 22 | 3(b) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 23 | 3(c) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 24 | 3(d) |       |       |       |       | 8     |       |       |       |       |       |       |       |       |       |       |       |       |       |       | 6.5   |
| 25 | 3(e) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 26 | 3(f) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 27 | 4(a) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 28 | 4(b) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 29 | 4(c) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 30 | 5(a) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 31 | 5(b) |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 32 | 6    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 33 | 7    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 34 | 8    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 35 | 9    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 36 | 10   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 37 | 11   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 38 | 12(a) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 39 | 12(b) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 40 | 13   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 41 | 14(a) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 42 | 14(b) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 43 | 15   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 44 | 16(a) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 45 | 16(b) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 46 | 17(a) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 47 | 17(b) |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 48 | 18   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 49 | 19   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 50 | 20   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 51 | 21   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
The algal characteristics that we collected from these algae consist of 21 features, namely: (1) shape, (2) length, (3) color, (4) number of chloroplast, (5) cell type, (6) fluorescent body, (7) raphe, (8) valve, (9) flagella, (10) prickle, (11) arm, (12) theca, (13) nucleus, (14) cingulum, (15) horn, (16) vacuole, (17) striae, (18) tooth, (19) tentacle, (20) cyst, and (21) setae. These characteristics are then broken down to values that the algae may possess under study. Determination of characteristics also considers the dominant features possessed by an object to make it easier to distinguish it from other objects. The results are shown in Table 1.

**B. KNOWLEDGE BASE**

The collected algal data and their characteristics were then compiled into a table of species/genera ↔ feature relationships. There were a total of 51 features possessed by the 20 algae. The table was then given to five algae experts for assessment and validation. The experts have at least ten years of experience in algal identification and analysis. The assessment results are shown in Table 2.

The Char column contains the feature numbers as depicted in Table 1. Columns 1 to 20 represent species/genera codes as listed in Section III.A. A number in the cell represents a dominance level of the feature possessed by the species/genera. For example, in the first row, columns 1, 2, 15, and 18 contain the numbers 4.375, 4.375, 4.5, and 7, respectively. It means that the Stretched straight shape (feature no. 1(a)) is owned by species/genera Nitzschia, Pseudo-nitzschia, Ceratium furca, and Trichodesmium with dominance level of characteristic as mentioned above. This value is the average given by the experts according to their beliefs about the feature dominance level possessed by a species/genus. This value is given based on the criterion range of values as listed in Table 3.

| Table 3. The criterion range of values for the level of dominance of the feature possessed by a species/genus. |
| --- |
| Range | Description |
| 2 — 4 | The feature is not dominant (many found in other species/genera) |
| 4 — 6 | The feature is quite dominant |
| 6 — 8 | The feature is dominant |
| 8 — 10 | The feature is very dominant (the main characteristic of the species/genus) |

**Table 4. Application testing results by evaluators.**

| No | Species/Genus | Expert #1 | Expert #2 | Expert #3 | CL Avg(%) | Incorrect System Identification |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Nitzschia | 1 | 100 | 1 | 100 | 1 | 100 | 100 |
| 2 | Pseudo-nitzschia | 1 | 85.5 | 1 | 100 | 1 | 85.5 | 90.33 |
| 3 | Skeletonema costatum | 1 | 72.67 | 1 | 66.67 | 1 | 100 | 79.78 |
| 4 | Thalassiosira | 0 | 55 | 1 | 56 | 0 | 55 | - |
| 5 | Chaetoceros socialis | 1 | 65.67 | 1 | 72.67 | 0 | 72.67 | - |
| 6 | Noctiluca scintillans | 1 | 94 | 1 | 94 | 1 | 94 | 94.00 |
| 7 | Protoperidinium claudicans | 1 | 48 | 1 | 55 | 0 | 96 | - Ceratium furca |
| 8 | Pronerocentrum micans | 1 | 83 | 1 | 83 | 1 | 83 | 83.00 |
| 9 | Gymnodinium catenatum | 1 | 51 | 1 | 51 | 0 | 48 | - Protoperidinium claudicans |
| 10 | Gymnodinium sanguineum | 0 | 51 | 0 | 51 | 0 | 51 | - |
| 11 | Gyrodinium lachryma | 1 | 94.75 | 1 | 94.75 | 0 | 100 | - |
| 12 | Pyrodinium bahamense | 1 | 88.67 | 1 | 88.67 | 1 | 88.67 | 88.67 |
| 13 | Gambierdiscus toxicus | 0 | 51 | 1 | 55 | 0 | 51 | - Alexandrium |
| 14 | Alexandrium | 1 | 47 | 1 | 60 | 1 | 51 | 52.67 |
| 15 | Ceratium furca | 1 | 45 | 0 | 55 | 0 | 55 | - Protoperidinium claudicans |
| 16 | Dinophysis caudata | 1 | 79.75 | 1 | 79.75 | 1 | 79.75 | 79.75 |
| 17 | Protoperidinium | 0 | 55 | 1 | 91 | 0 | 55 | - |
| 18 | Trichodesmium | 1 | 90.67 | 1 | 90.67 | 1 | 90.67 | 90.67 |
| 19 | Cochlodinium polykrikoides | 1 | 56.67 | 1 | 70.67 | 1 | 70.67 | 66.00 |
| 20 | Dinophysis acuminata | 1 | 93.5 | 1 | 93.5 | 1 | 94.5 | 93.83 |

| Num of Correct | 16 | 17 | 11 | Avg of Accuracy = (80 + 85 + 55) / 3 = 73.33% |
| --- | --- | --- | --- | --- |
| Accuracy | 80% | 85% | 55% | --- |
C. EXPERIMENT

Tests were carried out to see the system’s performance in identifying algae types based on their characteristics. First, we trained three algae experts to use and evaluate Algies. One evaluator was from the Ministry of Marine Affairs and Fisheries, and the other two were university academics. Two evaluators hold doctorate degrees and one has a master’s degree in related algal science. We then provided each evaluator with 20 samples of algae to be identified using Algies. The results of Algies identification were then compared with the results of evaluator identification, i.e., correct or incorrect. The results of testing the Algies system by the three evaluators are presented in Table 4. The accuracy is calculated using equation (4). We also calculated the average value of confidence level and accuracy to see the overall system performance. Moreover, we captured the incorrect species/genus names suggested by the system in the last column for analysis purposes.

IV. DISCUSSION

An expert system simulates the judgment of humans by using the encoded knowledge. In general, the quality of a judgment relies on the techniques used to represent knowledge and perform reasoning. For example, correctly identifying HABs-caused species or genera of algae requires precise knowledge. Combining the automated expert system and the quality of encoded knowledge could theoretically overpower the contemporary methods of algal identification in providing rapid results for decision-making. Unfortunately, the lack of automatic identification tools has restricted researchers and laboratory technicians from delivering fast results of HABs causative species.

In this work, a rapid identification system is proposed based on an expert system with certainty factors. First, the morphological characteristics of targeted algae were formalized as a knowledge base (KB). After that, an inference engine was developed to drive a dialogue interface between users and the KB. The interface facilitates a structured exploration of the KB through a series of questioning and answering. As a result, the species or genera of a specimen will be identified.

We incorporated 51 algae characteristics as features into our expert system. We tested the system on 20 types of algae, where 73.33% of accuracy was obtained. From the correctly identified species/genera, the evaluators had high confidence in the results; the confidence levels are between 52.67% to 100%, with an average of 83.51%. The confidence levels are lower from the incorrectly identified species/genera, between 50% to 96.2% and 62.28% on average. This finding indicates that the developed expert system has worked as expected where the certainty factor has a linear impact on the results. When experts are confident with their judgments, it has a high chance of producing a correct result.

In Table 4, the genus Nitzschia is the most easily identified object by Algies with the highest confidence level (100%) in each test carried out by the evaluators. This result indicates that the description of the genus Nitzschia is clear and significantly different from the other objects. The other objects consistently identified by the system are Noctiluca scintillans, Trichodesmium, Pyrodinium bahamense, Procorcentrum micans and Dinophysis caudata. The system identification results for these species/genera were in line with the identification results of all evaluators with a confidence level above 79.75%.

Algies had consistently misidentified Gymnodinium sanguineum. The species was incorrectly identified as Gymnodinium catenatum. The result underlines a critical finding that the description of Gymnodinium sanguineum has to be improved to differentiate it from Gymnodinium catenatum and vice versa. The result also outlines the strength of KBS in feeding its “logical thinking” and “conclusion” for further improvement, as discussed at the beginning of this paper.

One thing that should be noted for further inspection is the identification result of Gyrodinium lachryma. There is an inconsistency in the species identification process. Two of the three evaluators showed correct results with a high confidence level (94.75%). However, the other evaluator showed that the result was incorrect with a high confidence level, i.e., 96.5%. This finding indicates that there might be inconsistencies in the weighting of the characteristics dominance level of this species.

Further, we compared the disagreement between evaluators by computing how many evaluators disagreed on a result from the developed expert system. We obtained that one evaluator disagreed on 33.33% cases, two and three evaluators on 55.56% cases, and 11.11% cases, respectively. As depicted by this result, more than half of disagreement cases (66.67%) were performed by two or more evaluators. It indicates that the judgment of the majority of experts has a high impact on the produced results.

V. CONCLUSION

A novel expert system for identifying algae is proposed in this paper. The proposed method used algae’s morphology knowledge in the form of ontology as the system’s guide. The knowledge of 51 algae characteristics was acquired from five algae experts on 20 HABs causative species/genera. Furthermore, certainty factors were used to indicate the level of confidence. The experimental results showed that the system’s accuracy and confidence level of correctly identified algae yielded 73.33% and 83.51%, respectively.

There are rooms for improvement to increase the system’s accuracy, such as attributing different weights for each characteristic or adjusting certainty factors. For example, more weights could be given to the more essential determinants (based on experts’ knowledge) of
an algae type. Furthermore, the certainty factors could be tuned automatically using machine learning techniques. Moreover, a hybrid approach, combining the system with computer vision, is worth trying in the future.

ACKNOWLEDGMENT
The authors gratefully thank five anonymous experts who provided critical evaluation judgment in the study. The other members of the SATREPS Mariculture Project from Japan and Indonesia are also thanked for their indirect contributions to the study.

REFERENCES

[1] H. Thoha et al., “Resting cyst distribution and molecular identification of the harmful dinoflagellate Margaletidinium polykrikoides (Gymnodiniales, dinophyceae) in Lampung Bay, Sumatra, Indonesia,” Front. Microbiol., vol. 10, no. FEB, pp. 1–12, Feb. 2019. DOI:10.3389/fmicro.2019.00306.

[2] E. Berdalet et al., “Marine harmful algal blooms, human health and wellbeing: Challenges and opportunities in the 21st century,” J. Mar. Biol. Assoc. United Kingdom, vol. 96, no. 1, pp. 61–91, Nov. 2016. DOI:10.1017/s0025315415001733.

[3] V. Aditya, A. Koswara, N. Fitriya, A. Rachman, T. Sidabutar, and H. Thoha, “Public Awareness on Harmful Algal Bloom (Hab) in Lampung Bay,” Mar. Res. Indones., vol. 38, no. 2, pp. 71–75, Apr. 2015. DOI:10.14203/mri.v38i2.58.

[4] P. M. Gilbert, “Eutrophication, harmful algae and biodiversity — Challenging paradigms in a world of complex nutrient changes,” Mar. Pollut. Bull., vol. 124, no. 2, pp. 59–606, Nov. 2017. DOI:10.1016/j.marpolbul.2017.04.027.

[5] W. W. Carmichael, “Health Effects of Toxin-Producing Cyanobacteria: The CyanoHABs” Hum. Ecol. Risk Assess. An Int. J. Hum. Ecol. Risk Assess., vol. 7, no. 5, pp. 1393–1407, Sep. 2001. DOI:10.1080/20018091095087.

[6] D. L. Tang, B. P. Di, G. Wei, I. H. Ni, S. O. Im, and S. F. Wang, “Spatial, seasonal and species variations of harmful algal blooms in the South Yellow Sea and East China Sea,” Hydrobiologia, vol. 568, no. 1, pp. 245–253, Jul. 2006. DOI:10.1007/s10750-006-0108-1.

[7] S. F. Wang, D. L. Tang, F. L. He, Y. Fukuyo, and R. V. Azanza, “Occurrences of harmful algal blooms (HABs) associated with ocean environments in the South China Sea,” Hydrobiologia, vol. 596, no. 1, pp. 79–93, Jul. 2008. DOI:10.1007/s10750-007-9059-4.

[8] A. T. Yhiguez et al., “Over 30 years of HABs in the Philippines and Malaysia: What have we learned?,” Harmful Algae, vol. 102, no. February, 2021. DOI:10.1016/j.hal.2020.101776.

[9] C. S. Tan, P. Y. Lau, S.-M. Phang, and T. J. Low, “A framework for the automatic identification of algae,” in Proc. Int. Conf. Comput. Inf. Sci. (ICCOINS), Malaysia, 2014, pp. 1–6. DOI:10.1109/iccoins.2014.6868360.

[10] M. Lauffer, F. Genty, J. L. Collette, and S. Margueroen, “Phytoplankton identification by combined methods of morphological processing and fluorescence imaging,” in Proc. 2015 IEEE Work. Environ. Energy, Struct. Monit. Syst. EESMS, 2015, pp. 131–135. DOI:10.1109/eems.2015.7175865.

[11] P. Qian et al., “Multi-Target Deep Learning for Algal Detection and Classification,” in Proc. Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, 2020-July, pp. 1954–1957. DOI:10.1109/emcb44109.2020.9176204.

[12] Q. Xiaoyan, “Research on Imbalanced Microscopic Image Classification of Harmful Algae,” IEEE Access, vol. 8, pp. 125438–125446, 2020. DOI: 10.1109/access.2020.3007646.

[13] K. P. Sudheer, “Knowledge Extraction from Trained Neural Network River Flow Models,” J. Hydrol. Eng., vol. 10, no. 4, pp. 264–269, 2005. DOI:10.1061/(asce)1084-0699(2005)10:4(264).

[14] E. De Maat, K. Krabben, and R. Winkels, “Machine learning versus knowledge based classification of legal texts,” Front. Artif. Intell. Appl., vol. 223, no. June 2014, pp. 87–96, 2010. DOI:10.3233/978-1-60750-682-9-87.

[15] J. H. Gennari et al., “The evolution of Protégé: An environment for knowledge-based systems development,” Int. J. Hum. Comput. Stud., vol. 58, no. 1, pp. 89–123, Jan. 2003. DOI:10.1016/s1071-5819(02)00127-1.

[16] Z. Akbar et al., “An ontology-driven personalized faceted search for exploring knowledge bases of capsicum,” Futur. Internet, vol. 13, no. 7, pp. 1–17, Jun. 2021. DOI: 10.3390/fi13070712.

[17] L. Manik, “Out-of-Scope Intent Detection on A Knowledge-Based Chatbot,” Int. J. Intell. Eng. Syst., vol. 14, no. 5, pp. 446–457, Oct. 2021. DOI: 10.22266/ijies2021.1031.39.

[18] J. Pariwono, “Oceanographic Condition of Lampung Coastal Waters in Coastal Resource Management Program Publication,” Coastal Resource Center, University of Rhode Island, Jakarta, Indonesia, Technical Report (TE-99/12-1), pp. 28, 1998.

[19] R. Puspasari, Y. Sugianti, A. Rustam, R. A. Adi, S. S. Sagala, and W. S. Pranowo, “The outbreak of Cochlodinium sp.: The red tide maker in the coastal of Lampung Bay,” IOP Conf. Ser. Earth Environ. Sci., vol. 176, no. 1, Jul. 2018. DOI: 10.1088/1755-1315/176/1/012021.

[20] A. Damar, F. Colijn, K. J. Hesse, and Y. Wardiatno, “The eutrophication states of Jakarta, Lampung and Semangka Bays: Nutrient and phytoplankton dynamics in Indonesian tropical waters,” Journal of Tropical Biology & Conservation (JTBC), vol. 9, no. 1, pp. 61–81, 2012.

[21] G. R. Barokah, A. K. Putri, and G. Gunawan, “The Abundance of Phytoplankton Causing HAB (Harmful Algal Bloom) in Lampung Bay during West and East Monsoon (Indonesia Language),” JPB Kelautan dan Perikanan, vol. 11, no. 2, pp. 115-126, 2012. DOI: http://dx.doi.org/10.1557/pjbp.k.v11i2.302.

[22] R. Rositasari et al., “Five decades of LIPI Contribution in Jakarta Bay (5 Dekade LIPI di Teluk Jakarta),” Research Center for Oceanography LIPI, pp. 128, 2017.

[23] T. Sidiabutar, “Mass Mortality of Fish in Lampung Bay, Indonesia,” omni.akua, vol. 12, no. 1, May 2016. DOI: 10.20884/1.oa.2016.12.1.27.

[24] J. A. Widians, N. Puspitasari, and A. Febriansyah, “Disease Diagnosis System Using Certainty Factor,” in Proc the 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Oct. 2019. DOI: 10.1109/icceieee47180.2019.8981421.