Freshwater Microalage Image Identification and Classification Based on Machine Learning Technique

A. Victoria Anand Mary[1], G. Prabakaran[2] and S. Mohan[3]
[1]Research Scholar, [2&3]Assistant Professor
[1, 2&3]Department of Computer Science and Engineering, Annamalai Unviersity, Chidambaram, Tamilnadu, India
E-Mail: victoriaphd16@gmail.com, gpaucse@yahoo.com, mohanaucs@gmail.com

Abstract - **Algae is an aquatic organism of an enormous and diverse group, which has the ability to conduct photosynthesis. The various sorts of microalgae play trivial roles in marine and fresh water environment. Microalgae are of various sizes and shapes, ranging from unicellular to multicellular forms**. **These algae were from the division of Anabaena, Oscillatoria, Microcystis Scenedesmus, Pediastrum and Cosmarium found in fresh water lake. In very high density these microalgae may discolor the water, outcompete, and become poisonous to other life forms. This is technically termed as harmful algal blooms. It is one of the most serious water pollution problems. Today, humans in many ways to use microalgae's for example, as fertilizers, soil conditioners, and livestock feed. A hybrid method is apply to automatic detection and recognition of some selected freshwater algae genera by combining the image processing technique with ANN approaches. Thus, analysis and prediction of algae is significant, which can be achieved using machine learning processing.**
Keywords: **Photosynthesis, Microscopic, Chlorella, Harmful Algal Blooms, Microalgae**

I. INTRODUCTION

The aquatic microscopic species are culture in tanks or ponds to harvest it or use it to treat effluents pumped in the ponds or rivers. This paper presents a survey on various algae image analysis techniques by means of Machine Learning Processing (MLP). These days, algae are well thought-out as a resource for the production of renewable energy. The concentration of carbon dioxide is considerably raised in the atmosphere, which leads to the climatic changes, globally. Fuel costs are increase due to the shortage of fossil fuels. Therefore, we are in the position to finding alternate solutions in this problem. Algae are produced rich in fats, carbohydrates, and proteins. Algae can grow well in marine and freshwater environment; it also can grow up in wastewater. It gives fresh oxygen to the environment. Algae oil is use to make biodiesel, and the algae protein can be used as feedstock and food supplement. The most important living things on our green planet are single cell algae. A few benefits of algae are listed below:

1. Algae's produces oxygen which other aquatic life uses.
2. It is significant role to humans in the form of food and medicine.
3. It is very important in many food chains acting as the primary producer of organic matter.

4. It plays a significant role in treating the wastewater, reduce the effects of harmful chemicals and to decrease carbon- dioxide emitted from some of the power plants.

Manual and practical analysis and prediction are tedious process together with loads of human errors. This process can be made effortless with the help of a various ML techniques.

II. RELATED WORKS

Review of related works has been conducted on fresh water and marine microalgae images. The related work carried out associated to most recent techniques of predicting and investigate the algae through its images.

Ching Soon Tan *et al.*, [1] have facilitated marine biologists a framework for being an assistive tool. The framework consists of (i) pre-processing the image, (ii) segmenting region of interest, (iii) extracting features, (iv) evaluating and combining those features based on the given decision criterions, and (v) the quantification. This work is the developed and deployed for automatic recognition system for biodiversity studies in large quantity of algae to identify algae species, it complements the existing intensive procedures of an operator. Akhlaqur Rahman *et al.,* [2] have offered the applications for aquaculture of different machine learning algorithms. The historical data for aquaculture is achieved from farm practices, yields, and environmental data sources. The relations among these different variables can be obtained by applying machine-learning algorithms.

Sansoen Promdaen *et al.*, [3] have planned a new method for automated recognition of 12 microalgae that are most common in the water resources of Thailand. To handle difficulties such as unclear algae boundary and noisy background in the method, a new method is proposed for segmenting algae bodies from an image background and for computing texture descriptors from a blurry texture object. The Feature combination approach is useful to handle a variation of algae shapes of the same genus. Andrew Lensen *et al.*, [4] have established a new biological application of detecting Phormidium algae in rivers of New Zealand using raw images captured from the air. This method is great task to detect the algae. The method proposed synthesizes a set

of image operators and adopts a local thresholding approach to segmenting an image into algae and non-algae regions.

Wisnu Jatmiko *et al*., [5] have facilitated on the modeling of the algal growth, which is one of the parameter that defines the amount of C02, which can be fixated by algal. From the data observation, the growth behavior is modeled by regression method and Multi Layer Perceptron (MLP) algorithm. To achieve optimization, the MLP algorithm is combined with Canonical Particle Swarm Optimization (CPSO) resulting more accurate MLP-CPSO model. Devesh Prakash *et al*., [6] have designed a comprehensive mathematical model describing algal dynamics in batch reactors has been attempted.

Purnomol *et al*., [7] have proposed the growth behavior is modeled by regression graphs using Single Hidden Layer Feed-forward Network (SLFN). Extreme Learning Machine (ELM) algorithm is applied to train and test the data. The ELM algorithm is accepted to be the fastest algorithm to learn an SLFN for regression. To optimize the algorithm, ELM is also combined with GA. Kyle Dannemiller *et al*., [8] have presented a novel method to develop the quality of the algae micro-image and segmenting it. It involves two steps to automatic recognize and classify the algae in an images. The Retinex enhancement technique is applied to improve the algae image quality. Then, Support Vector Machine (SVM) is used to segment the algae from the improved quality image background.

Ethan T. Daniels *et al*., [9] have presented a handheld, low-cost in-situ device employing spectro photometry and optical filtering. For automatic concentration detection a pattern recognition method was formed. This method utilizes a binary classification to make a distinction between low and high concentrations of algae. Features for classification were defined by the spectral peaks evaluated, these include: RMS value, distance between edges, variance, and energy. Thu-Hien Tran *et al*., [10] have proposed a ML method based on the Least Squares Support Vector Regression (LS-SVR) for modeling the growth time of the green alga Klebsormidium flaccidum on mortar surfaces. And to use optimization technique to identify an appropriate set of the LS-SVR hyper- parameters, the Flower Pollination Algorithm (FPA) is employed.

Xiaogang Ouyang *et al*., [11] have proposed an image processing method based on color segmentation to discriminate the distribution of algae on the insulator surface. Two main procedures are applied to an un-ambiguous insulator images using a digital camera pre-supposed with a white balance recognition and extraction of green algae pixels from the insulator images through a specific program, which is capable of discriminating the algae distribution and calculating the algae coverage rate and characteristic green value.

Iago Correa *et al*., [12] have proposed a deep learning technique to solve the classification of microalgae by

creating a dataset that allow us to adopt this technique. This present work is the first one to apply this kind of technique on the microalgae classification task. The obtained results show the capabilities of the method to properly classify the data by using as input the low-resolution images acquired by a particle analyzer instead of pre-processed features. Lakshmi *et al*., [13] have proposed a classification technique. Due to various factors such as change in size and shape with climatic changes, various growth periods, and the presence of other microbes it is very difficult to automatic identification and classification of algal community. An elaborate analysis of Artificial Neural Network (ANN) concepts and Convolution Neural Network (CNN) of deep learning technique are proposed. This is automatically measure the algae growth through the image classification techniques from algae digital images.

From the related work of the available sources, following open issues are identified that need further investigations.
1. No appropriate method of classification and prediction, when there are more noises, weak edge in the image.
2. Poor lighting and occultation could degrade the overall Performance.
3. Application of optimization algorithms to enhance the performance of image analysis.

Therefore, this experiment is an attempt to develop an automatic recognition and classification system of some selected freshwater algae genera by a hybrid system having the combination of image processing with ANN approaches.

III. METHODOLOGY

We implement the automated freshwater microalgae identification and classification model by using MATLAB R2015a. It is a multi paradigm numerical computing programming language, which has built-in functions to support image processing and classification methods. The prototype for developing automated system has the following phases 1. Image preprocessing, 2. Segmentation, 3. Feature extraction, and 4. Classification. The proposed system architecture model is shown in Fig.1.
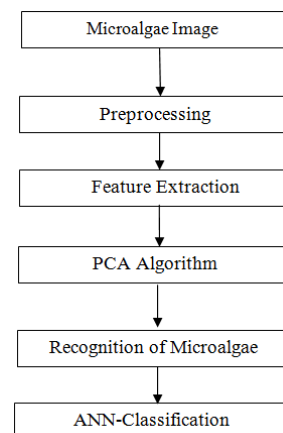

Fig.1 Proposed System Architecture Model

The various microalgae image identification and classification methodology is given below.

A. Microalgae Image Preprocessing: The Microalgae images are captured from the microscope, which has some noise, having low contrast quality; it can contain some open and close gap, small objects or unwanted area. It eradicates noise to develop intelligibility and appearances of images.

B. Microalgae Image Segmentation: An algae image encompasses of foreign objects and some other microorganisms. Image segmentation is used to detect the number of object in binary image, which has been preprocessed earlier. The second order canny edge detector algorithm is applied to perform image segmentation in this experiment.

C. Feature Extraction: The Feature extraction employed to transform binary and color image from the preprocessed stage into a set of parameters that described the microalgae features. Feature extracted from the preprocessed microalgae image using both binary and color image include shape, area, minor and major axis, perimeter and Fourier spectrum with Principal Component Analysis (PCA). The detailed method of feature extraction is described as follow.

1. Text Features

Text property is an important and powerful property for retrieval information. These features are statistical features such as energy, entropy, homogeneity, contrast, etc,. The various phases are discussed here: To smooth the image of the image Gaussian filter is applied with the specified standard deviation 's' which is reduce the noise. Next, the local gradient and the edge direction were calculated at each point. The G_x and G_y were assessed from the first derivation for intensity pixels. In the direction of the gradient, the edge point is recognized to be a point of locally maximum. Then the non-maximal control in the gradient magnitude image was cast-off to provide a thin line, which is present, the ridge of the edge point is resolute in threshold.

2. Shape Features

Shape is visual feature and essential for usage of content description. This description can be divided into two main categories such as region-based and boundary descriptors. Region properties are used to measure the connected components in the binary image. These features are area of the object, centroid, perimeter, diameter [structural features based on region properties].

a. Perimeter of Object
   The perimeter of an object is the sum of all the distance among each of the adjacent pair of pixels in its border; which is presented in red pixel in the microalgae image. It is incorporated the features since it provides an indication about the image object size.

b. Angel of Inclination
   A simple function is developed to find the angle of inclination of the image by obtaining the longest path that existed between each two point on algae boundary, which is done automatically without any human intervention. This angle of inclination is used to rotate the algae image to be aligned horizontally. The angle of inclination is calculated by using the Equation 1:
   $$\Theta (x, y) = \tan^{-1} (s1-s2)/(1+s1*s2) \qquad (1)$$

   The slope of lines denoted by s1 slope value of points (x1,y1) to (x2,y2) and s2 slope value of points (x3,y3) to (x1,y1) that form the angle.

c. Width Factor
   To differentiate amongst algae having similar shape, the object width factor was calculated by slicing across the major axis and parallel to minor axis. Likewise, the feature points become stabled into a number of vertical strips. For each strip, the ratio of strip length to the object width is calculated by using the following Equation 2.
   $$I_c = (\text{Width of Image column }/\text{object length}) \qquad (2)$$

   where I_c is the ratio of column c. This result is normalized to obtain only five features. The object area was calculated by counting the number of white pixels inside the object boundary. The factor of area is represented by the actual number of white pixels in object region, which is selected in image.

3. Fourier Spectrum with PCA Feature Extraction

To extract texture feature from an algae image, the Fourier spectrum method is used, which produced an increased accuracy while image detection. Fourier spectrum is perfectly used for relating the directionality of periodic two-dimensional patterns. The spectrum features are expressed in polar coordinates to produce a function P (r, θ). By annularity sampling the function S(r,θ) with radius function $R_1(r)$ and angle function $R_2(\o)$, a one-dimension functions are obtained. The extracted feature sets from an algae object may contain some redundant feature. The PCA algorithm is applied to reduce and summarize the extracted features of Fourier spectrum method by eliminating redundancies. Eight Eigen value extracted which are included in feature extraction process.

4. Combinational Features

Feature vector consist of the combination of structural, textural and statistical data. The vision data consists of the algae image shape characteristics, which are experimented in the classifiers and recognizers. As a final point, the algorithm succeeds in the edge linking by integrating the weak pixels that stayed connected to the strong pixels. The needed morphological processes are accomplished on binary images such that the filling of boundary area, image

border removal and elimination of any unimportant region. The proposed system algae image features are listed below. F1- Shape Index, F2-Minor axis, F3-Major axis, F4&F5-Minor,Major axis ratio, F6-Area, F7-Perimeter, F8-Entropy, F9-Homogenetity, F10-Entropy, F11-Contrast, F12-Centroids, F13-F16-object width factors, F17-F24- Fourier spectrum normalized by PCA.

D. MLP ANN for Classification

To perform the classification on extracted feature vectors, the Multilayer Perceptron Network trained with back propagation error algorithm ANN is used. This method is broadly used for pattern recognition and classification. To approximate any continuous function, hidden layer feed forward neural network is used in this work.

The ANN architecture be made of three typical layers, which includes the input layer holds 24 input nodes, hidden layer holds 8 nodes and output layer holds 6 nodes. To access the network performance, the standard Root Mean Squared Error (RMSE) function was used based on trial and error to assess network performance and a momentum value of 0.05 is set. When the above parameters are fixed with the optimal step sizes taken in weight space, where a function of the learning rate 0.05. A network that classified the training data perfectly but unable to generalize and classify new 'unseen' data. To get enhanced generalization, 10% of the input dataset kept aside for cross validation. Testing dataset was then used to avoid biasness in result. There was a set of images that are not used for training the ANN.

IV. MICROALGAE IMAGE DATA SET

The microalgae images have been collected from Culture Collection of Algae and Protozoa (CCAP) dataset. From water samples collected from different ponds and lakes the images of freshwater microalgae have been captured.
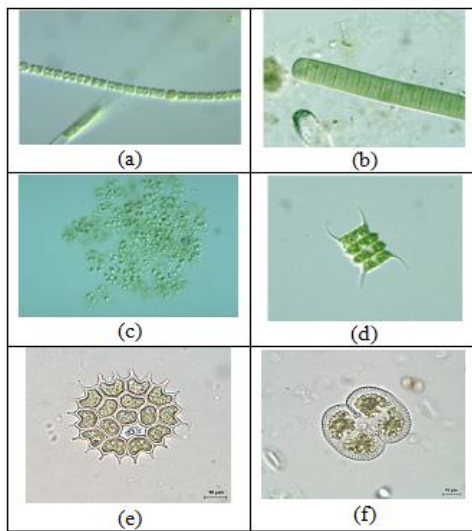

Fig. 2 shows Genera of Microalgae of fresh wate a). Anabaena b). Oscillatoria c).Microcystis d).Scenedesmus e). Pediastrum f).Cosmarium

Using the electronic microscope water samples (Fig.2) were studied and the images of freshwater algae were moved to digital storage devices(a high resolution camera model) which is attached to the microscope lens, and connected with personal computer for image acquisition. The two types of toxin toxic and non- toxic with two subdivision color blue-green and green algaes (Table I). It is found in different shapes such as filament, colony and unicell. Further, it is analyzed in inside of circle, ellipse, spiral and cylinder shapes.

TABLE I THE MICROALGAE GENESES NAME AND SHAPE

Toxin Type	Division	Genus	Shape
Toxic	Blue-green algae	Anabaena	Filament
		Oscillatoria	Filament
		Microcystis	Colony
Non-toxic	Green algae	Scenedesmus	Colony
		Pediastrum	Colony
		Cosmarium	Unicells

V. EXPERIMENTS AND RESULTS

The proposed model implements image preprocessing and classification of algae images automatically. The training was carried out for 300 images and testing was carried out for 240 images that have not been used for the training of MLP in order to test the accuracy of the system. Forty images of each genus have been selected and were tested in this experiment. The recall is measure to evaluate the proposed method. The attributes are consider for the evaluation metrics such as True Positive (TP), True Negative (TN), False Positive (TP) and False Negative (FN). It can be measures true positive values, which are correctly indentifying the image. The recall is estimated by using the Equation 3.

Recall=(TP/(TP+FN))*100   (3)

Another parameter of precision is also estimated by using Equation 4.

Precision=(TP/(TP+FN)) * 100   (4)

The total number of correctly find out image over the total number of images. The accuracy is estimated by using Equation 5.

Accuracy= ((TN+TP)/(TN+TP+FN+FP))*100   (5)

These results are used for computer-based classification of selected microalgae from the available image dataset samples. The method for separating the objects found in the image lead to a short irregularly shaped image regions enclosing the algae.

Table II shows the proposed system performance, where the overall accuracy is 96.25 %.The recall and precision of

proposed system perform 98% and 95% respectively. By the observed system the accuracy rate for Anabaena and Oscillatoria can be misclassified with each other as their spiral shape seems similar for the classifier and extracted feature. MLP was applied in this experiment instead of SVM or RBF because the data utilized in this experiment are limited with a small number of algae and numbers of extracted features. The accuracy rate achieved in this experiment is acceptable and consider higher rate when it is compared with other similar studies.

TABLE II THE PROPOSED SYSTEM CONFUSION MATRIX WITH ACCURACY LEVEL

Name of Micro algae	No. of Test samples	System Recognition Result						Unknown	Accuracy in Percentage (%)
		A	B	C	D	E	F		
A	40	37	2	0	1	0	0	0	92.5
B	40	2	38	0	0	0	0	0	95.0
C	40	0	0	39	1	0	0	0	97.5
D	40	0	0	0	40	0	0	0	100
E	40	0	0	0	0	38	2	0	95
F	40	0	0	0	0	1	39	0	97.5

*A-Anabaena, B-Oscillatoria, C-Microcystis, D-Scenedesmus, E- Pediastrum F-Cosmarium

VI. CONCLUSION

In this work, an image processing techniques with ANN approach is presented to identify and classify selected genus of freshwater algae from two different divisions having different size and shape. This experiment revealed that computerized recognition approach is essential for freshwater algae, and prove that the classifying procedure is practicable for identification of the selected freshwater algae. The enhanced accuracy resulted from the better techniques used for preprocessing and due to the definite features selected during feature extraction process. Additionally, the system reliability was reliant on the combination of approaches used for image preprocessing, segmentation and the training of data set. Testing results also exposed that the developed system was reliable to be used for monitoring water quality of ponds and lakes. Further, this experiment should be extended to analysis on wavelet transform techniques in future work.

REFERENCES

[1] Ching Soon Tan, P.Y. Lau, S.M. Phang, and T.J. Low, "A framework for the automatic identification of algae", *Int.Conf. Comput. Inf. Sc*i., pp. 1–6, 2014.

[2] Akhlaqur Rahman and S. Tasnim, "Application of Machine Learning Techniques in Aquaculture", *Int. J. Comput. Trends Technol*., Vol.10, No. 4, pp. 214–215, 2014.

[3] Sansoen Promdaen, P. Wattuya, and N. Sanevas, "Automated microalgae image classification", *Procedia Comput. Sci*., Vol. 29, pp. 1981–1992, 2014.

[4] Andrew Lensen, H. Al-Sahaf, M. Zhang, and B. Verma, "Genetic Programming for algae detection in river images", *IEEE Congr. Evol. Comput. CEC 2015 - Proc.*, pp. 2468–2475, 2015.

[5] Wisnu Jatmiko, "Algal growth rate modeling and prediction optimization using incorporation of MLP and CPSO algorithm", *Int. Symp. Micro-NanoMechatronics Hum. Sci. MHS 2015, 2016.*

[6] Prakash, A. Ratn, S. Kumar, and P. Bose, "Calibration of Algal Growth Model Using Multi-objective Genetic Algorithm", *Int. J. Environ. Sci. Dev.,* Vol. 6, No. 12, pp. 901–907, 2015.

[7] Purnomo "Genetic algorithm optimization for extreme learning machine based micro algal growth forecasting of Chlamydomonas sp", *ICACSIS 2015 - 2015 Int. Conf. Adv. Comput. Sci. Inf. Syst. Proc., no. October 2015*, pp. 243–248, 2016.

[8] Dannemiller, K. Ahmadi, and E. Salari, "A New Method for the Segmentation of Algae Images Using Retinex and Support Vector Machine", pp. 361–364, 2015.

[9] Ethan T. Daniels Benjamin D. McPheron, "A machine learning Approach to Classifying Algae Concentrations", 2017.

[10] Tran and Hoang, "Estimation of algal colonization growth on mortar surface using a hybridization of machine learning and metaheuristic optimization", *Sadhana - Acad. Proc. Eng. Sci*., Vol. 42, No. 6, pp. 929–939, 2017.

[11] Ouyang, Z. Jia, S. Yang, H. Chen, and C. Wei, "Research of image processing method on alage conolized on the insulator surface", *IEEE Electr. Insul. Conf. EIC 2017*, , pp. 26–29, June 2017.

[12] Correa, P. Drews, S. Botelho, M. S. De Souza, and V. M. Tavano, "Deep learning for microalgae classification", *Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017*, January-December, pp. 20–25, 2018.

[13] Lakshmi and Sivakumar, "Biologically Rationalized Computing Techniques for Image Processing Applications", Vol. 25, 2018.

[14] Deglint, C. Jin, A. Chao, and A. Wong, "The feasibility of automated identification of six algae types using neural networks and fluorescence-based spectral-morphological features", pp. 1–13, 2018.