

Received November 22, 2018, accepted December 11, 2018, date of publication December 20, 2018, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889017

The Feasibility of Automated Identification of Six Algae Types Using Feed-Forward Neural Networks and Fluorescence-Based Spectral-Morphological Features

JASON L. DEGLINT¹, CHAO JIN^{1,2}, ANGELA CHAO¹, AND ALEXANDER WONG¹, (Senior Member, IEEE)

¹Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

²School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China

Corresponding authors: Jason L. Deglint (jdeglint@uwaterloo.ca) and Chao Jin (j3chao@uwaterloo.ca)

This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Canada Research Chairs program. This was selected when the article was submitted for review via "IEEE ScholarOne Manuscripts."

ABSTRACT Harmful algae blooms are a growing global concern since they negatively affect the quality of drinking water. The gold-standard process to identify and enumerate algae requires highly trained professionals to manually observe algae under a microscope. Therefore, an automated approach to identify and enumerate these micro-organisms is needed. This research investigates the feasibility of leveraging machine learning and fluorescence-based spectral-morphological features to enable the identification of six different algae types in an automated fashion. A custom multi-band fluorescence imaging microscope is used to capture fluorescence data of water samples at six different excitation wavelengths ranging from 405 to 530 nm. Automated data processing and segmentation were performed on the captured data to isolate different micro-organisms from the water sample. Different morphological and spectral fluorescence features are then extracted from the isolated micro-organism imaging data and is used to train neural network classification models. The experimental results using three different neural network classification models (one trained on morphological features, one trained on fluorescence-based spectral features, and one trained on fluorescence-based spectral-morphological features) showed that the use of either fluorescence-based spectral features or fluorescence-based spectral-morphological features to train neural network classification models led to statistically significant improvements in identification accuracy when compared to the use of morphological features (with average identification accuracies of $95.7\% \pm 3.5\%$ and $96.1\% \pm 1.5\%$, respectively). These preliminary results are promising and illustrate the feasibility of leveraging machine learning and fluorescence-based spectral-morphological features as a viable method for automated identification of different algae types.

INDEX TERMS Artificial neural networks, feature extraction, fluorescence, image classification, machine learning, multispectral imaging, optical microscopy, supervised learning, water conservation.

I. INTRODUCTION

Harmful algae blooms (HABs) are increasingly becoming a major threat to our water bodies. An illustrative example of the threat of HABs is an incident in the summer of 2011, where Lake Erie experienced the largest harmful algae bloom in recorded history [1] (see Fig. 1 for a Moderate Resolution Imaging Spectroradiometer (MODIS) image captured by the Aqua satellite of the incident). This bloom was primarily *Microcystis aeruginosa*, one of the most lethal cyanobacteria

genera according to the Great Lakes Environmental Research Laboratory [2].

For example, many types of cyanobacteria (a type of algae) can be extremely dangerous for humans and animals. Swallowing *Microcystis aeruginosa* cyanobacteria can have serious side effects such as abdominal pain, diarrhea, vomiting, blistered mouths, dry coughs, and headaches [3]. In addition, *Anabaena sp.*, another type of cyanobacteria, can produce lethal neurotoxins called anatoxin-a which



FIGURE 1. A MODIS image showing Lake Erie on October 9, 2011. The bloom was primarily *Microcystis Aeruginosa*, according to the Great Lakes Environmental Research Laboratory [2].

has been shown to cause death by progressive respiratory paralysis [3].

In another illustrative example, a toxin produced by *Microcystis*, called Microcystin-LR (MC-LR), is strictly regulated by the World Health Organization (WHO) since it can be lethal for humans [4]. In addition, the maximum acceptable concentration (MAC) for the cyanobacterial toxin Microcystin-LR (MC-LR) in drinking water is 0.0015 mg/L, based on guidelines from the Government of Canada [5]. To better prevent toxin exposure during a bloom event, active monitoring of water quality is critical as it enables the collection of both temporal and spatial trends of bloom activity. These spatial-temporal trends can then be inspected and analyzed via predictive analytic strategies, thus providing key stakeholders with early warning signs that a bloom may occur.

Commonly, the task of identification and enumeration of algae in natural water is conducted at certificated laboratory facilities, where highly-trained taxonomists perform manual analysis on the water samples [6]. This time-consuming process requires samples to be preserved, shipped, and then inspected at the laboratory facilities with expensive laboratory equipment. As an alternative, in-situ monitoring devices such as fluorometric meters have been used to detect specific pigment levels, which can act as an indirect indicator for the purpose of monitoring in practice. While this alternative approach can be useful for rough assessments of algae distributions, it still requires further manual confirmation due to the level of granularity in order to inform decision-makers to take the necessary actions needed to minimize exposure risk. Furthermore, a study presented by Culverhouse *et al.* [7]

show that human taxonomists have an identification accuracy between 67% and 83%, depending on the taxonomist. It is their conclusion is that the experts in the study are not unanimous in their identification, even when inspecting microorganisms with very distinct morphology [7]–[10]. Therefore, a method that could not only directly identify algae types in an automated and cost effective manner is highly desired for water industry.

In this study, an investigation is conducted on the feasibility of leveraging machine learning and fluorescence-based spectral-morphological features to enable the identification of six different algae types in an automated fashion. More specifically, this paper explores and investigates the efficacy of a number of different morphological and spectral fluorescence features extracted from multi-band fluorescence imaging data when used to train neural network classification models designed for the purpose of identification of six algae types in an automated manner. Artificial neural networks have been shown to be an effective machine learning tool to determine non-linear mappings from measured inputs to a classification output [11]. Therefore, in this paper a feed-forward neural network is used to explore whether it can be used as a potential predictor of algae when trained using fluorescence-based spectral-morphological features.

The paper is organized as follows. First, related work is discussed in Section II. Second, the proposed methodology for investigating the feasibility of automated identification of algae types is presented in Section III. Third, the experimental setup used in this study is presented in Section IV. Fourth, experimental results are presented and discussed in Section V. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

A number of computational techniques and methods have been proposed for the purpose of automatically identification of different algae types using imaging data. In the following section, in chronological order, a summary of the major contributions in the recent past in this area is presented.

In 2002, Walker *et al.* [12] leveraged fluorescence excitation in their imaging protocol for automated identification of *Anabaena sp.* and *Microcystis sp.*. More specifically, their protocol was able to achieve over 97% identification accuracy when looking for *Anabaena sp.* and *Microcystis sp.* in natural populations found in Lake Biwa, Japan, by capturing a single fluorescence image and a single brightfield image. The authors claim that without the use of the fluorescence component, the automated identification of microalgae in the sediment saturated samples would be nearly impossible.

Similar results were found by Hence *et al.* [13] in 2008 when they showed that by using epifluorescence microscopy in combination with brightfield microscopy they could correctly identify between 13 different phytoplankton samples as either algae or non-algae. They accomplished this by using three different filter sets to capture the fluorescence data and built a hand-tuned classifier based on empirically derived thresholds. However, the main drawback of both

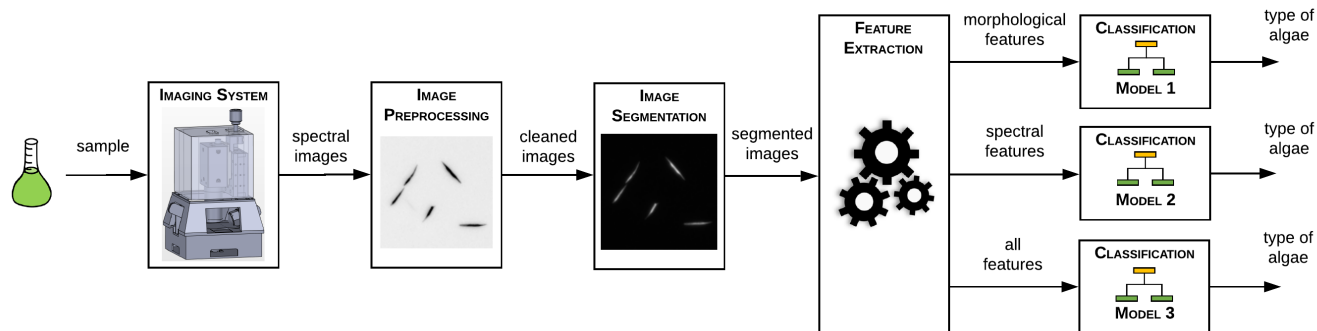


FIGURE 2. The proposed methodology of automated identification of different algae types can be broken into five main steps: First, a water sample containing algae was imaged using a custom multi-band fluorescence imaging microscope (Section III-A). Next, this data is then processed using data processing algorithms for background subtraction (Section III-B). Next, the processed data is segmented into background objects and micro-organism objects, and then the individual algae are isolated (Section III-C). A number of morphological and spectral fluorescence features are then extracted from the isolated micro-organism imaging data (Section III-D). Finally, these extracted features are used to train three different neural network models designed for the purpose of identification of algae types given an isolated micro-organism.

these methods is they only tested classifiers with two classes, either *Anabaena sp.* and *Microcystis sp.* [12] or algae or non-algae [13]. Furthermore, both of these methods leveraged only a single fluorescence wavelength in combination with the brightfield image.

In 2010, Hu *et al.* [14] utilized the fact that different algae species have different ratios of antenna pigments, which results in different fluorescence emission spectra. More specifically, Hu *et al.* illuminated twenty different algae from six algae divisions (Dinophyta, Bacillariophyta, Chrysophyta, Cyanophyta, Cryptophyta, and Chlorophyta) at four different wavelengths (440 nm, 470 nm, 530 nm, and 580 nm), and then measured the emission spectra from 600 nm - 750 nm with a 5 nm resolution. By concatenating these four emission spectra together and conducting a multivariate linear regression and weighted least-squares it was found that each of the feature vectors from each phylum was independent from the others. Although Hu *et al.* showed that these feature vectors were independent, one major drawback of the method Hu *et al.* proposed is that the relative ratios of different algae can only be achieved at the phyla level when mixing two species from different phyla.

More recently, Deglint *et al.* [15] conducted a comprehensive spectral analysis of the fluorescence characteristics of three algae species when excited at twelve discrete spectral wavelengths. Their findings was that the fluorescence spectra of the three algae species appear quite distinctive, and thus the use of multi-band fluorescence imaging microscopy could be a promising direction to explore. However, that study is highly preliminary as the number of algae species studied was very limited and a more comprehensive quantitative investigation on how best to leverage such spectral information was not well explored for the purpose of automated identification of algae types.

Motivated by the findings of Deglint *et al.* [15], this research aims to go a major step further by investigating and exploring the utilization of machine learning and fluorescence-based spectral-morphological features derived

from multi-band fluorescence imaging microscopy data at different excitation wavelengths (between 405 nm - 530 nm) and a larger number of algae types (six different algae types in total).

III. METHODOLOGY

The proposed methodology used in this study to explore the feasibility of automated identification of different algae types using machine learning and fluorescence-based spectral-morphological features can be broken into five main steps (see Fig. 2). First, a water sample containing algae was imaged using a custom multi-band fluorescence imaging microscope to capture fluorescence imaging data at a number of different excitation wavelengths (Section III-A). The captured fluorescence imaging data are then processed using data processing algorithms for background subtraction (Section III-B). Next, the processed fluorescence imaging data is first segmented into background objects and micro-organism objects, and the individual algae are isolated and extracted to produce isolated micro-organism imaging data (Section III-C). A number of morphological and spectral fluorescence features are then extracted from the isolated micro-organism imaging data, and used to train three different neural network models designed for the purpose of identification of algae types given an isolated micro-organism. (Section III-D). More specifically, three different neural network models trained: i) using only the fluorescence-based morphological features (Model 1), ii) using only the fluorescence-based spectral features (Model 2), and iii) using fluorescence-based spectral-morphological features (Model 3).

A. DATA ACQUISITION

The custom multi-band fluorescence imaging microscope used in this study for capturing fluorescence imaging data is composed of five main elements, as seen in Fig. 3. First, a light source at a particular excitation wavelength (Fig. 3A) (blue arrow) is used to illuminate a water sample placed on a blank slide (Fig. 3B), thus effectively exciting the algae in the

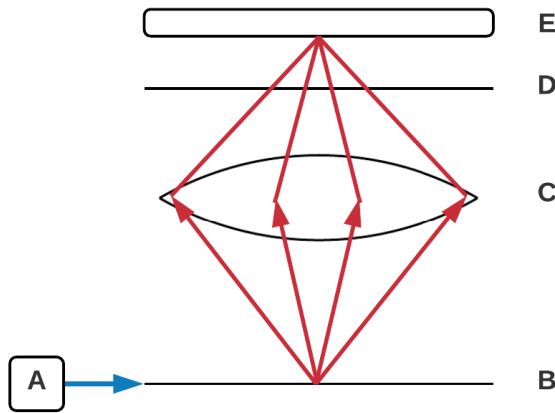


FIGURE 3. Image acquisition is performed using a custom multi-band fluorescence imaging microscope, which is composed of five main elements: a light source at a particular excitation wavelength (A), a water sample placed on a blank slide (B), a magnification lens (C), a highpass filter (D), and a monochromatic sensor (E).

water sample. The algae in the water sample then fluoresce and emit light at a lower energy (red arrows) and (Fig. 3C) the emitted light is focused using an magnification lens. This focused light passes through (Fig. 3D) a highpass filter before hitting (Fig. 3E) a monochromatic sensor.

Given that multiple wavelengths can be used to excite the algae, let $\lambda_1, \lambda_2, \dots, \lambda_m$ be the individual fluorescence images captured using different excitation wavelengths of a given water sample. Therefore, we define the entire multi-band fluorescence image Λ_{raw} as

$$\Lambda_{raw} = [\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_i \quad \dots \quad \lambda_m]. \quad (1)$$

B. DATA PROCESSING

To improve the quality of the captured fluorescence imaging data obtained from the custom multi-band fluorescence imaging microscope for subsequent micro-organism isolation and neural network classification modeling steps, a set of automated data processing algorithms are first performed to compensate for some of the issues associated with the captured data. First, to compensate for the fact that the illumination source in the imaging microscope does not illuminate the imaging field of view in a perfectly homogenous manner, iterative background subtraction is performed on each λ_i in Λ_{raw} . This background subtraction can be expressed as,

$$\Lambda_{corrected} = \Lambda_{raw} - \Lambda_{background} \quad (2)$$

where $\Lambda_{corrected}$ and $\Lambda_{background}$ denotes the illumination corrected imaging data and the background image, respectively. To approximate the background image $\Lambda_{background}$, a Gaussian low-pass filter was first applied to Λ_{raw} to perform noise suppression. Next, to suppress features in Λ_{raw} at different scales to better approximate $\Lambda_{background}$, an iterative multi-scale morphological opening was performed on the Gaussian low-pass filtered image, where the size of a disk structuring element is changed at each iteration.

C. IMAGE SEGMENTATION AND MICRO-ORGANISM EXTRACTION

Given the corrected multi-band fluorescence imaging data, $\Lambda_{corrected}$, the next step is to now segment the background from the micro-organism objects, as well as isolate each micro-organism in the captured data. This will allow features to be extracted from each segmented organism, which is vital when training and testing a given classification model. To achieve segmentation of background from the micro-organism objects in the captured data, a binary background-foreground classifier was used to classify each pixel as either the foreground C_f or the background C_b . The advantage of using fluorescent images when segmenting is that there is a large contrast between the foreground and the background, allowing a single decision boundary to be found that can separate the organisms from the background. To learn the decision boundary, θ , of this binary background-foreground classifier, the within class variance was minimized, as originally proposed by Otsu *et al.* [16]. Therefore, the binary background-foreground classifier can be expressed as,

$$L_i(\underline{x}) = \begin{cases} C_f & \text{if } f_i(\underline{x}) > \theta \\ C_b & \text{otherwise} \end{cases} \quad (3)$$

where $f_i(\underline{x})$ is the pixel intensity at pixel \underline{x} for a given wavelength image λ_i , where $i \in [1, m]$.

Given the segmented foreground-background information, a connected-components analysis strategy was used to group neighboring pixels in the foreground class together to isolate individual micro-organisms. Each isolated micro-organism in the water sample can be defined as ϕ_j , where $j \in [1, n]$, where n is the total number of micro-organisms in the segmented image set Φ .

D. FLUORESCENCE-DRIVEN SPECTRAL-MORPHOLOGICAL FEATURE EXTRACTION AND NEURAL NETWORK CLASSIFICATION MODELING

Having isolated the micro-organisms in the imaging data, the goal is then to learn a classification model for identifying between the different types of algae in an automated manner. To achieve this, a set of fluorescence-based morphological and spectral features for characterizing individual micro-organisms were first extracted. The motivation for the proposed fluorescence-based morphological and spectral feature set is that, by leveraging not only features for characterizing the morphology of a micro-organism but also a greater number of spectral features gained from the custom-built fluorescence imaging microscope, a more complete profile can be constructed around the micro-organism to enable better discrimination between different types of algae than can be achieved using previous approaches. In the proposed fluorescence-based spectral-morphological feature set, five morphological features were chosen, as proposed in [17], which can be described as follows:

- 1) **Area:** The total number of pixels in an isolated micro-organism.

- 2) **Convex Area:** The total number of pixels of the convex hull of an isolated micro-organism.
- 3) **Eccentricity:** The eccentricity of an isolated micro-organism is the ratio of the distance between the foci of the ellipse and its major axis, and is bounded between zero and one.
- 4) **Equivalent Diameter:** The diameter of a circle with the same area as an isolated micro-organism.
- 5) **Extent:** The ratio of pixels of an isolated micro-organism to the bounding box that contains that isolated micro-organism.

These five spatial features were chosen as a base representative set to describe the general shape and physical characteristics of a specimen and are commonly used to describe a given shape of an object in an image [18]. The area and equivalent diameter both embed the relative scale of different organisms since the relative scale between different species can be useful when classifying different types of algae. The convex area and extent are two measures how compact a given organism is. The eccentricity encapsulates both the major and minor axis length and gives an idea of how elongated an organism is, which is known to vary among different algae types. The pose and orientation of a given algae were not measured since the exact location and position of a given sample does not affect its classification.

Furthermore, a set of spectral features for characterizing the mean of the fluorescence intensities in an isolated micro-organism are incorporated at each of the different captured excitation wavelengths. In this study, since six different excitation wavelengths are captured, the set of spectral features can be described as follows:

- 1) **Emission Signal 1:** The mean fluorescence intensity for an isolated micro-organism when excited at 405 nm.
- 2) **Emission Signal 2:** The mean fluorescence intensity for an isolated micro-organism when excited at 420 nm.
- 3) **Emission Signal 3:** The mean fluorescence intensity for an isolated micro-organism when excited at 450 nm.
- 4) **Emission Signal 4:** The mean fluorescence intensity for an isolated micro-organism when excited at 470 nm.
- 5) **Emission Signal 5:** The mean fluorescence intensity for an isolated micro-organism when excited at 500 nm.
- 6) **Emission Signal 6:** The mean fluorescence intensity for an isolated micro-organism when excited at 530 nm.

These six different excitation wavelengths were chosen to cover the part of the electro-magnetic visible spectrum which consists of higher energy, since fluorescence is caused when an electron is excited by high energy light and is then later emitted by a lower energy photon. The six chosen high-power LEDs cover the excitation wavelength of the major pigments found in algae such as Chlorophyll *a* and Chlorophyll *b* as well as β -Carotene, phycoerythrin and phycocyanin [6], [19].

Finally, given the fluorescence-based morphological and spectral feature vectors extracted from the isolated microorganisms, a classification model must be learned to predict the associated output class (algae type) given on these input feature vectors for the purpose of automated identification

of algae types. A number of machine learning approaches can be leveraged to learn the relationship between the input fluorescence-based morphological and spectral and the associated algae type, ranging from support vector machines [20] to decision trees [21] and Naive Bayes [22].

In this study, the classification models used are feedforward neural networks, or also known as multi-layer perceptron networks, which is an artificial neural network where information moves from the input layer, through a given amount of hidden layers, and then to the output layer. An advantage in leveraging a feedforward neural network for the classification model is that a feedforward neural network is an universal approximator [23], [24]. As such, a feedforward neural network has the ability to approximate any continuous function with a finite number of neurons, and thus well-suited for learning a good approximation of a function that maps the fluorescence-based morphological and spectral features to the corresponding algae type. Each layer of the network consists of multiple neurons that take a weighted sum of the inputs, x_k and bias, b and transform them with a non-linear activation function, $f(z)$. This non-linear function takes as input

$$z = \sum_{k=1}^n x_k w_k + b \quad (4)$$

where w_k is a given weight for n inputs [25].

Using these extracted morphological and spectral features in tandem with a feedforward neural network, three different neural network classification models are trained and evaluated.

- **Model 1:** Uses only the morphological features.
- **Model 2:** Uses only the spectral features.
- **Model 3:** Uses spectral-morphological features.

Model 1 is trained and tested using only the five extracted morphological features. This model is used as a baseline to determine how well classification can be achieved when only looking at the size and shape characteristics of the algae. Next, Model 2 is trained and tested using only the six extracted spectral features. By only using these spectral features a comparison of the relative performance of Model 1 and Model 2 can be done. Finally, Model 3 is trained and tested using all the fluorescence-based spectral-morphological features. This model will closely mimic what human taxonomists use when classifying different types of algae since it incorporates both the color as well as the shape and size of the algae.

IV. EXPERIMENTAL SETUP

To investigate and explore the feasibility of leveraging machine learning and fluorescence-based spectral-morphological features to enable the identification of different algae types in an automated fashion, a number of experiments were designed and tested. First, the algae types selected for the proposed experiments are presented (Section IV-A) and then a discussion of the hardware implementation to

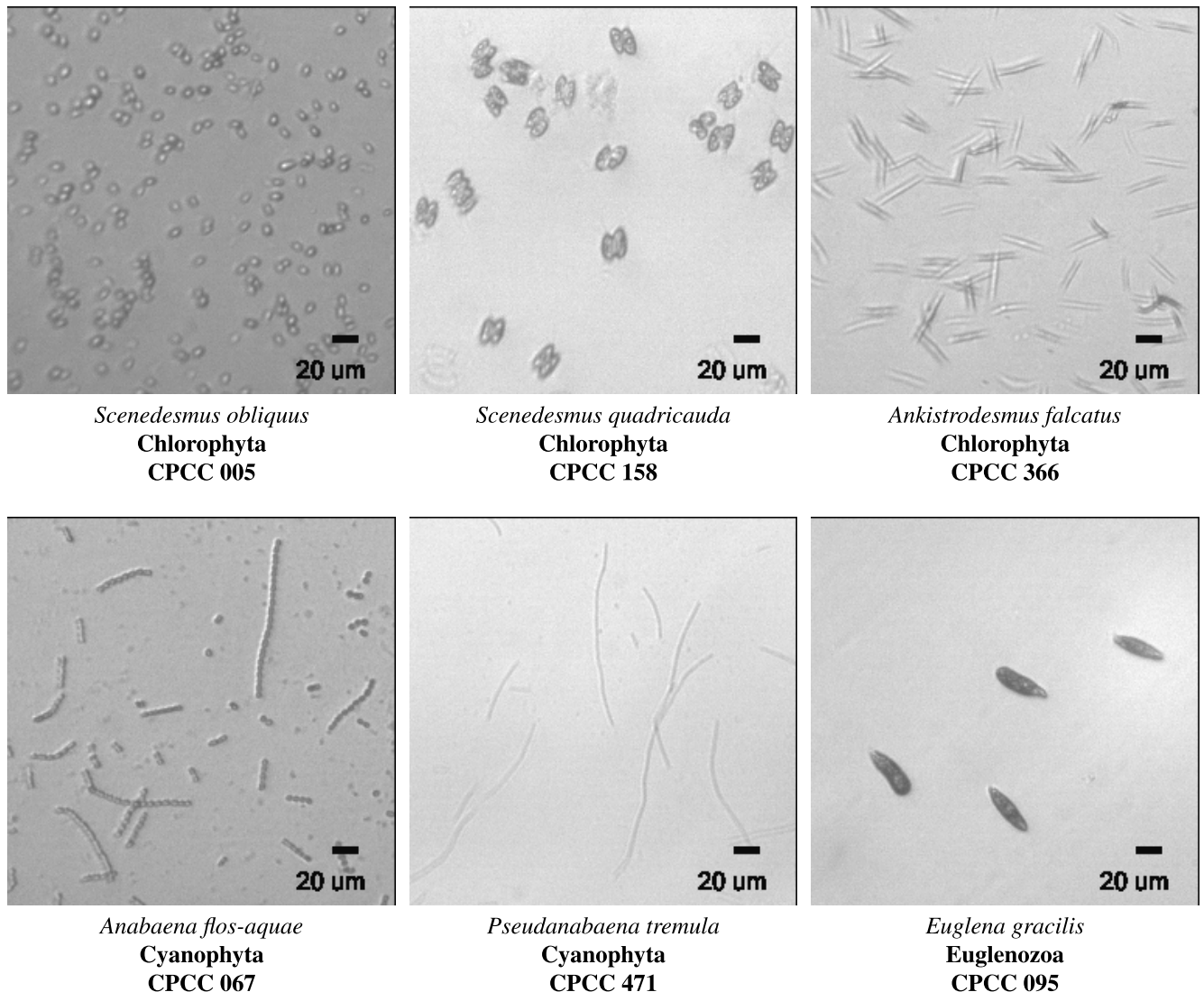


FIGURE 4. The six algae under investigation in this study are *Scenedesmus obliquus* (CPCC 005), *Scenedesmus quadricauda* (CPCC 158), *Ankistrodesmus falcatus* (CPCC 366), *Anabaena flos-aquae* (CPCC 067), *Pseudanabaena tremula* (CPCC 471), and *Euglena gracilis* (CPCC 095). These algae are from three different phyla classes (Chlorophyta, Cyanophyta, and Euglenozoa) and two samples are from the same genus (*Scenedesmus*).

collect data follows (Section IV-B). Finally, a full description of the three model architectures will be presented (Section IV-C).

A. TYPES OF ALGAE

Six different types of algae from the Canadian Phycological Culture Centre (CPCC) were chosen to explore the identification accuracy of the learned neural network models. As seen in Fig. 4, these six algae types, with their respective CPCC number, are broken into their respective phyla. These algae types along with the corresponding number of micro-organism samples for each type are described as follows:

I. Chlorophyta (green algae)

1. *Scenedesmus obliquus* (CPCC 005): 751 samples
2. *Scenedesmus quadricauda* (CPCC 158): 382 samples
3. *Ankistrodesmus falcatus* (CPCC 366): 500 samples

II. Cyanophyta (blue-green algae or cyanobacteria)

4. *Anabaena flos-aquae* (CPCC 067): 548 samples
5. *Pseudanabaena tremula* (CPCC 471): 299 samples

III. Euglenozoa

6. *Euglena gracilis* (CPCC 095): 131 samples

To build up this dataset, pure samples of each of the six types of algae were imaged with the custom-built fluorescence imaging microscope, which will be described in the next section. A cropped section of the brightfield images can be seen in Fig. 4. This was captured by placing a white light source under the microscope slide and capturing an image with the custom-built fluorescence imaging microscope.

These samples were strategically chosen to be a broad representation of algae, given that three different phyla classes are present. However, special attention was given to blue-green and green algae since they are the most common toxin producers in our waters. Furthermore, in the

Cyanophyta class, two very similar filamentous algae were chosen (*Anabaena flos-aquae* and *Pseudanabaena tremula*) to see how far the proposed classification could differentiate between two similar filamentous algae. Finally, two species from the *Scenedesmus* genus were also chosen to determine how well a neural network classifier model driven by fluorescence-based spectral-morphological features can identify between two algae with very similar morphology.

B. HARDWARE CONFIGURATION

The custom-built fluorescence imaging microscope contains six high power LEDs that emits light at six different spectral wavelengths (405 nm, 420 nm, 450 nm, 470 nm, 500 nm, and 530 nm). These high-powered LEDs are placed orthogonal to a 3" × 1" microscope slide which had a pure algae sample on it as well as a standard cover slip. The excited algae samples then emitted fluoresced light, which was focused by a 20×, passing through a 600 nm highpass filter and onto a 4.1 MP camera. The spatial resolution of this system is 1.2 μm / pixel, as determined using the 1951 US Air Force (USAF) glass slide resolution target. A sample image of each type of algae fluorescing at each excitation wavelength can be seen in Fig. 5. The value of having this spectral information comes into play especially when the morphology between certain algae types are very similar in nature, such as in *Anabaena flos-aquae* (CPCC 067) and *Pseudanabaena tremula* (CPCC 471).

C. MODEL ARCHITECTURES

As previously discussed in Section III-D, three neural network classification models will be trained and tested to explore their relative identification performance. The network architecture used for the three neural network classification models are based on a feedforward network architecture and can be seen in Fig. 6. The number of input features of this network is 5, 6, and 11, for Model 1, Model 2, and Model 3, respectively. Model 1 only uses 5 morphological features, while Model 2 uses 6 spectral features. Model three combines both these morphological and spectral features together to form a morphological-spectral feature set. The neural network classification models each contain three hidden layers, with the number of neurons at each layer being 12, 8, and 6, respectively. This decrease in the amount of neurons as the network continues to go deeper allows the input features to be transformed into a better representation for improve discrimination power between the different algae types. The same neural network architecture was used in this study to enable a consistent and fair comparison based purely on the use of different sets of features (morphological features only vs. spectral features only vs. combination of morphological and spectral features), which is the main contribution of this study. In the situation where different architectures are also used along with different sets of features, it would be quite difficult to identify whether changes in performance are due to the use of different sets/combinations of features, or due to the change in neural network architecture.

TABLE 1. The mean and standard deviation of the identification accuracy using the test data across 20 test runs of Monte Carlo Cross Validation. Model 1, trained using morphological features, had the lowest average accuracy and the highest standard deviation. Model 2, trained using spectral features, and Model 3 trained using morphological-spectral features, achieved significantly higher average accuracies and lower standard deviations.

	Accuracy	
Model 1	53.0%	± 3.5%
Model 2	95.7%	± 1.5%
Model 3	96.1%	± 0.8%

In the network architecture design, the rectified linear unit (ReLU) activation function was chosen for the network. Finally, the neural network classification models contain a softmax function, a normalized exponential function, which is used to transform the output to sum to one to mimic a probability mass function. Therefore, the features of a given model are directly fed into a given neural network model and are combined in a linear fashion and then passed through a non-linear activation function. The softmax layer normalises the output so that the sum of outputs is equal to one, mimicking a probability density function.

The neural network classification models were evaluated by using 20 runs of Monte Carlo Cross Validation (MCCV), where 70% of the data was randomly selected without replacement for a given run of the cross validation to be the training set. The remaining 30% of the input features were used as the test data-set.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section the experimental results will be presented (Section V-A) and discussed (Section V-B). Then future work (Section V-C) will be presented.

A. EXPERIMENTAL RESULTS

The mean and standard deviation of identification accuracies across the 20 test runs of Monte Carlo Cross Validation can be seen in Table 1. A number of observations can be made based on the identification accuracy results. First of all, it was observed that Model 1, which utilizes only the set of five morphological features had the lowest average identification accuracy at 53.0%. Model 2, which leverages the set of six fluorescence-based spectral features significantly outperformed Model 1, with an average identification accuracy of 95.7%. Finally, Model 3, which leverages the combined fluorescence-based morphological-spectral feature set also demonstrate a strong average identification accuracy at 96.1%. Therefore, it can be shown that the utilization of fluorescence-based spectral features is very important for the automated identification of different algae types, and further extends upon the observations made by Walker *et al.* [12] regarding the necessity of spectral information. It can also be observed that the standard deviation is significantly lower for Model 2 compared to Model 1, which indicates that the utilization of spectral features also provide more consistent

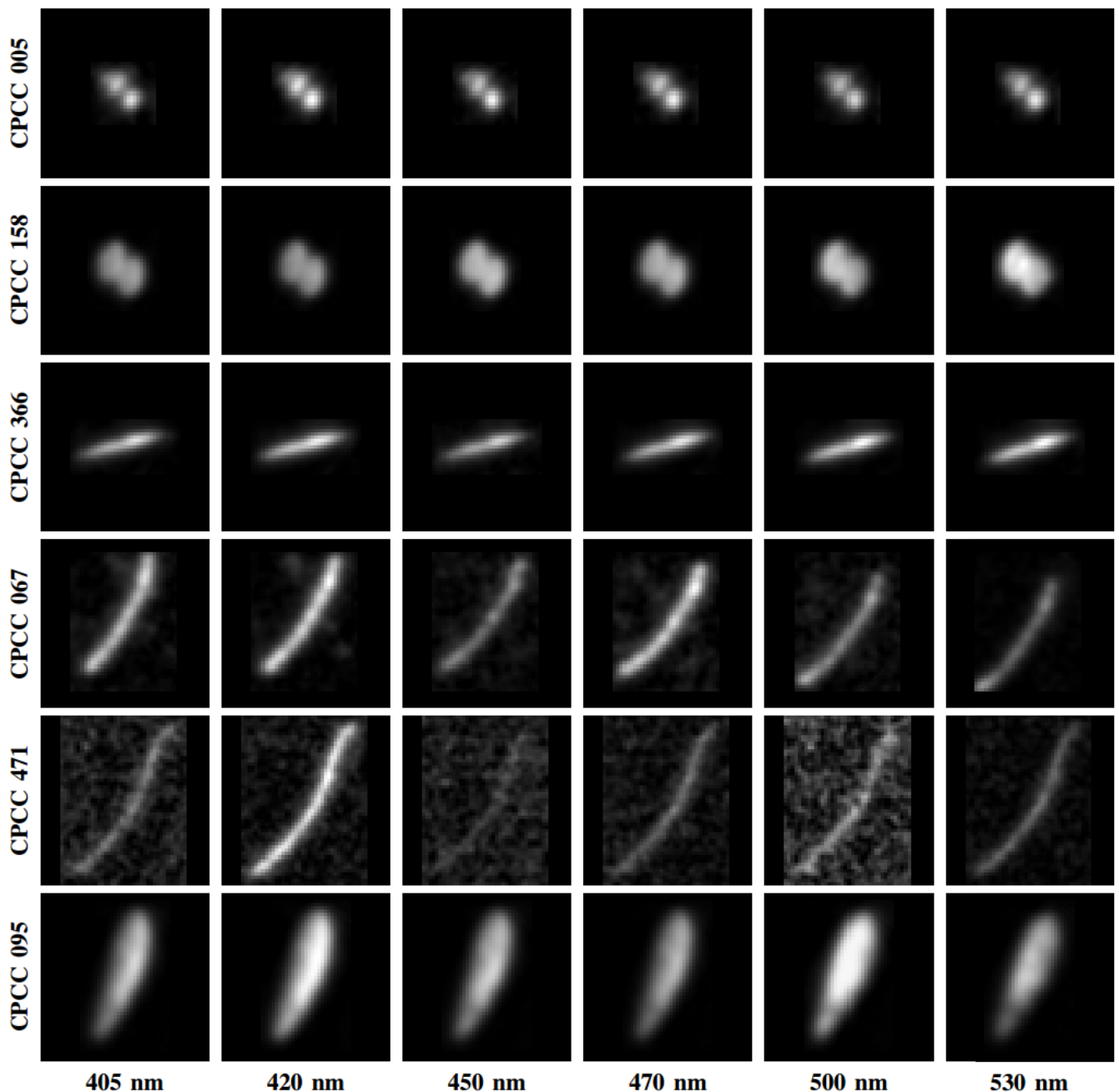


FIGURE 5. Six types of algae from three phyla classes were imaged: *Scenedesmus obliquus* (CPCC 005), *Scenedesmus quadricauda* (CPCC 158), *Ankistrodesmus falcatus* (CPCC 366), *Anabaena flos-aquae* (CPCC 067), *Pseudanabaena tremula* (CPCC 471), and *Euglena gracilis* (CPCC 095). Each algae was excited at six different wavelengths (405 nm, 420 nm, 450 nm, 470 nm, 500 nm, and 530 nm) and the fluorescent signal was captured with a monochrome sensor. Sample micro-organisms can be seen for each algae type at each excitation wavelength.

identification performance across different permutations of samples, which is important for generalizability in real-world scenarios. In addition, it can be observed that, when comparing Model 2 and Model 3, the average identification accuracy for Model 3 is increased compared to Model 2, as well as had a lower standard deviation. What this means is that by also leveraging morphological features, Model 3 can provide consistently improved identification performance.

Since the increase in average identification accuracy in Model 3 compared to Model 2 was relatively low, a pairwise t-test was run between all pairs of two models with a 1% significance level, as seen in Table 2. As expected, the pairwise t-test for Model 1 vs Model 2 as well as Model 1 vs Model 3 show that the improvements of Model 2 over Model 1, as well as the improvements of Model 3 over Model 1 are both statistically significant in terms of identification

TABLE 2. A pairwise t-test was run between all pairs of neural network classification models with a 1% significance level.

paired-sample t-test	Reject null hypothesis?	p-value
Model 1 vs Model 2	yes	7.35E-23
Model 1 vs Model 3	yes	6.04E-23
Model 2 vs Model 3	no	8.54E-01

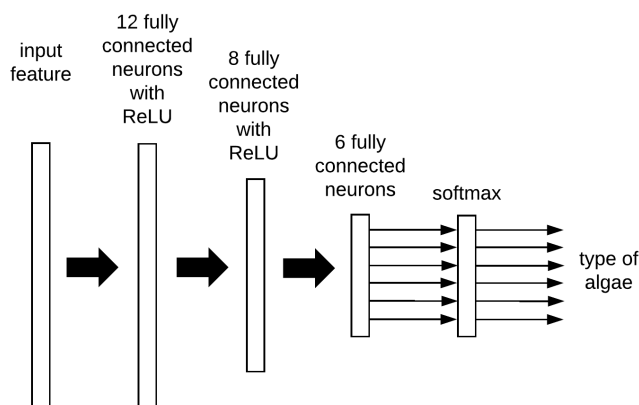


FIGURE 6. A fully-connected feed-forward neural network architecture was leveraged for all three neural network classification models. The number of input features of this network is 5, 6, and 11, for Model 1, Model 2, and Model 3, respectively. Model 1 only uses 5 morphological features, while Model 2 uses 6 spectral features. Model three combines both these morphological and spectral features together to form a morphological-spectral feature set.

accuracy. However, it was also found that Model 2 and Model 3 are do not show a statistically significant difference in their identification accuracy. Therefore, based on these results, it was observed that the set of fluorescence-based spectral features leveraged in the study is the major contributor to improved identification accuracy performance.

To gain a deeper understanding of the identification error associated with each neural network classification model, the confusion matrix of the best performing run from each 20 Monte Carlo Cross Validation runs for Model 1, Model 2, and Model 3 can be seen in Table 3, Table 4, and Table 5, respectively. The confusion matrix is an accepted and common method of reporting multi-class performance since a more detailed analysis of the classification error can be determined. In a confusion matrix any off-diagonal entries indicate a miss-classification from the model and therefore one can determine the precision and recall of a given classifier when inspecting a given class [26], [27]. In these tables, the major sources of identification error have been highlighted in red. Selected miss-classified samples from each of these highlighted red errors can be seen in Fig. 7, Fig. 8, and Fig. 9, for Model 1, Model 2, and Model 3, respectively. It is important to note that the values reported in Table 1 are the mean and standard deviation of the 20 Monte Carlo cross validation runs, while the confusion matrices in Table 3, Table 4, and Table 5 show the classification test accuracy of the best performing model from the 20 Monte Carlo cross validation runs.

B. DISCUSSION

In Table 3 and Fig. 7, it can be observed that for Model 1, which uses only morphological features, the greatest source of misidentifications came from *Anabaena flos-aquae* (CPCC 067) and *Pseudanabaena tremula* (CPCC 471) samples being mislabeled as *Ankistrodesmus falcatus* (CPCC 366). By inspecting both Fig. 4 and Fig. 5 it is clear that both of these species can have an elongated shape, which can cause problems with identification when leveraging only morphological features. The other most common misidentification error came from *Anabaena flos-aquae* (CPCC 067) being mislabeled as *Scenedesmus obliquus* (CPCC 005). Once again by looking at Fig. 4 and Fig. 5 smaller *Anabaena flos-aquae* (CPCC 067) could easily be confused with an *Scenedesmus obliquus* (CPCC 005) sample using just morphological features. Therefore it can be concluded, that for the extracted morphological features, and for these six types of algae, utilizing only the selected five morphological features is insufficient for differentiating between different samples. These results indicate the need to add additional features to increase the classification performance.

When inspecting Table 4 and Fig. 8, it can be observed that for Model 2, which uses only spectral features, the misidentification error was significantly less compared to that of Model 1, with the largest source of error being 4 samples of *Euglena gracilis* (CPCC 095) being mislabeled as *Scenedesmus quadricauda* (CPCC 158) and 3 samples of *Scenedesmus quadricauda* (CPCC 158) were mislabeled as *Euglena gracilis* (CPCC 095). These results are consistent with the observation that the spectral features between *Scenedesmus quadricauda* (CPCC 158) and *Euglena gracilis* (CPCC 095) is much closer compared to other samples. The other main source of error is when *Anabaena flos-aquae* (CPCC 067) is classified as *Pseudanabaena tremula* (CPCC 471). This miss-classification is intuitive as both of these species are both from the Cyanophyta phylum and are both filamentous types of algae.

Finally, in Table 5 and Fig. 9, it can be observed that for Model 3, which utilized fluorescence-based morphological-spectral features, the misidentification error was significantly less compared to that of Model 1, with sources of error being 6 samples of CPCC 067 being mislabeled as *Scenedesmus obliquus* (CPCC 005) and five samples of *Pseudanabaena tremula* (CPCC 471) being mislabeled as *Anabaena flos-aquae* (CPCC 067). This particular source of misidentification may be attributed to the fact that both are filamentous types with very similar shapes and characteristics. In addition, as seen in Fig. 8, certain *Anabaena flos-aquae* (CPCC 067) were classified as *Scenedesmus quadricauda* (CPCC 158) when the chain structure of the *Anabaena flos-aquae* broke apart, leaving one or two single-celled organisms on their own. When inspecting Fig. 4, a given *Anabaena flos-aquae* can easily be mistaken as a *Scenedesmus quadricauda*. When inspecting Fig. 9, it is observed that some images, such as *Pseudanabaena tremula* (CPCC 471) have not been

TABLE 3. The best performing Model 1 from 20 Monte Carlo cross validation runs.

Model 1		PREDICTED					
		CPCC 005	CPCC 158	CPCC 366	CPCC 067	CPCC 471	CPCC 095
TRUE	CPCC 005	189	9	31	12	0	2
	CPCC 158	4	70	8	7	0	10
	CPCC 366	10	3	119	11	3	7
	CPCC 067	65	10	50	21	11	9
	CPCC 471	14	2	34	8	20	7
	CPCC 095	2	6	5	2	1	21

TABLE 4. The best performing Model 2 from 20 Monte Carlo cross validation runs.

Model 2		PREDICTED					
		CPCC 005	CPCC 158	CPCC 366	CPCC 067	CPCC 471	CPCC 095
TRUE	CPCC 005	220	0	0	0	0	0
	CPCC 158	0	117	1	1	0	3
	CPCC 366	0	0	160	0	0	0
	CPCC 067	0	0	0	146	6	0
	CPCC 471	1	0	0	0	85	0
	CPCC 095	0	4	1	0	1	37

TABLE 5. The best performing Model 3 from 20 Monte Carlo cross validation runs.

Model 3		PREDICTED					
		CPCC 005	CPCC 158	CPCC 366	CPCC 067	CPCC 471	CPCC 095
TRUE	CPCC 005	205	0	0	0	0	0
	CPCC 158	0	114	0	0	0	3
	CPCC 366	0	0	165	0	0	0
	CPCC 067	6	0	0	155	2	0
	CPCC 471	0	0	0	5	81	0
	CPCC 095	0	2	0	0	0	45

well segmented. This reveals that the classification error is primarily due to errors in the segmentation.

Comparing the preliminary results achieved using the neural network classification models to that achieved by human taxonomists, which have an identification accuracy between 67% and 83% [7], it can be observed that Model 1 dramatically under-performed compared to human taxonomists, demonstrating that the utilization of just morphological features is not a reliable way to identify between different types of algae via imaging. However, both Model 2 and Model 3 demonstrated very encouraging identification accuracies when compared to that of human taxonomists, and thus illustrate not only that fluorescence-based spectral features are highly effective for identifying between different types of algae, but also that such automated identification methods can be a very valuable tool for human taxonomists to leverage to reduce the time-consuming and tedious task of manually isolating and analyzing individual micro-organisms, and thus be able to spend more time on the more important judgment and assessment of water quality and determining the

appropriate course of action to take to mitigate the situation. Therefore, this approach of capturing data at multiple wavelengths and processing it with a neural network shows promise that on-site monitoring of algae types is potentially possible. Given that the identification accuracy of human taxonomists are typically between the range of 67% and 83%, this method illustrates the feasibility of leveraging machine learning and fluorescence-based spectral-morphological features as a viable method for automated identification of different algae types. While in this study it was shown that the performance of Model 2 and that of Model 3 are quite similar, which does in fact illustrate the effectiveness of using spectral features, and does appear to be the most cost-effective method for practical applications, it can be noted that as the number of algae type grows, the combination of both the spectral and morphological features may be more advantageous.

C. FUTURE WORK

As previously discussed in Section V-B, using only the selected five morphological features is insufficient for





Model 1				
TRUE	<i>Scenedesmus obliquus</i> CPCC 005	<i>Anabaena flos-aquae</i> CPCC 067	<i>Anabaena flos-aquae</i> CPCC 067	<i>Pseudanabaena tremula</i> CPCC 471
PREDICTED	<i>Ankistrodesmus falcatus</i> CPCC 366	<i>Scenedesmus obliquus</i> CPCC 005	<i>Ankistrodesmus falcatus</i> CPCC 366	<i>Ankistrodesmus falcatus</i> CPCC 366

FIGURE 7. Misidentified samples from Model 1 corresponding to errors highlighted in red from Table 3.





Model 2				
TRUE	<i>Scenedesmus quadricauda</i> CPCC 158	<i>Anabaena flos-aquae</i> CPCC 067	<i>Anabaena flos-aquae</i> CPCC 067	<i>Euglena gracilis</i> CPCC 095
PREDICTED	<i>Euglena gracilis</i> CPCC 095	<i>Pseudanabaena tremula</i> CPCC 471	<i>Pseudanabaena tremula</i> CPCC 471	<i>Scenedesmus quadricauda</i> CPCC 158

FIGURE 8. Misidentified samples from Model 2 corresponding to errors highlighted in red from Table 4.





Model 3				
TRUE	<i>Anabaena flos-aquae</i> CPCC 067	<i>Anabaena flos-aquae</i> CPCC 067	<i>Pseudanabaena tremula</i> CPCC 471	<i>Pseudanabaena tremula</i> CPCC 471
PREDICTED	<i>Scenedesmus obliquus</i> CPCC 005	<i>Scenedesmus obliquus</i> CPCC 005	<i>Anabaena flos-aquae</i> CPCC 067	<i>Anabaena flos-aquae</i> CPCC 067

FIGURE 9. Misidentified samples from Model 3 corresponding to errors highlighted in red from Table 5.

differentiating between the different six different samples in these presented experiments, as demonstrated in Model 1. Therefore work must be conducted to expand the list of

morphological features to contain a more comprehensive set of physical descriptors of algae. Future work also includes determining which spectral and morphological

features impact the models performance in order to find the optimal set of features for classification. Research will also be conducted to evaluate the performance differences between using morphological features, spectral features, and a combination of the two types of features for a larger scale study with more algae types. In this future study investigation will be done around exploring additional morphological features that include both low-level features (edge and curvature based) as well as high-level features (statistical and shape based).

Research will be conducted to explore different neural network architectures. Other types of neural network structures, such as convolutional neural networks [28], [29], will also be explored as they learn the optimal set of features in an end-to-end manner. This involves exploring the design of more advanced neural network classification architectures for handling scenarios characterized by the need to automatically identify a greater number of algae types in contaminated and mixed water samples. Future work also includes the exploration of neural network ensembles composed of networks that perform well at identifying different groups of algae types to complement the performance of each other. Such a system will allow for near real-time analysis of a water sample to determine which types of algae are present as well as their relative concentrations. This will give water treatment plants and other organizations the ability to build up a database of algae activity over time allowing them an early warning sign that bloom might occur.

VI. CONCLUSIONS

In this paper, the feasibility of leveraging machine learning and fluorescence-based spectral-morphological features for automated identification of algae type was explored. In particular, neural network classification models were trained to identify different algae types using fluorescence-based spectral features and morphological features extracted from imaging data captured using a custom multi-band fluorescence imaging microscope at six different excitation wavelengths (405 nm, 420 nm, 450 nm, 470 nm, 500 nm, and 530 nm). Experimental results using three different neural network models (one trained on morphological features, one trained on spectral fluorescence features, and one trained on spectral-morphological fluorescence features) on six different algae types (*Scenedesmus obliquus* (CPCC 005), *Scenedesmus quadricauda* (CPCC 158), *Ankistrodesmus falcatus* (CPCC 366), *Anabaena flos-aquae* (CPCC 067), *Pseudanabaena tremula* (CPCC 471), and *Euglena gracilis* (CPCC 095)) demonstrated that neural network classification models trained using either fluorescence-based spectral features or fluorescence-based spectral-morphological features resulted in average identification accuracies of 95.7% and 96.1%, respectively. As such, the results of this study illustrate that leveraging machine learning and fluorescence-based spectral-morphological features can be a feasible direction for further exploration for the purpose of automated identification of different algae types.

CONTRIBUTIONS

JLD, CJ, and AW designed the functionality of the hardware system. AC and JLD designed the form factor of the system and AC created the 3D model. JLD collected the data used in the experiments. JLD and AW designed the data pipeline and architecture of the different machine learning classification models. JLD, CJ, and AW conducted the analysis and contributed to the writing of the manuscript.

REFERENCES

- [1] A. M. Michalak *et al.*, "Record-setting algal Bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 16, pp. 6448–6452, 2013.
- [2] NASA. (Oct. 2011). *Toxic Algae Bloom in Lake Erie*. [Online]. Available: <https://earthobservatory.nasa.gov/IOTD/view.php?id=76127>
- [3] I. R. Falconer, "Potential impact on human health of toxic cyanobacteria," *Phycologia*, vol. 35, no. 6S, pp. 6–11, 1996.
- [4] "Cyanobacterial toxins: Microcystin-LR," in *Guidelines for Drinking-Water Quality*, vol. 2. Geneva, Switzerland: World Health Organization, 1998. [Online]. Available: <http://apps.who.int/iris/handle/10665/63844>
- [5] *Guidelines for Canadian Drinking Water Quality: Guideline Technical Document—Cyanobacterial Toxins*, Health Canada, Ottawa, ON, Canada, 2018. [Online]. Available: <https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-guideline-technical-document-cyanobacterial-toxins-document.html>
- [6] L. Barsanti and P. Gualtieri, *Algae: Anatomy, Biochemistry, and Biotechnology*. Boca Raton, FL, USA: CRC Press, 2014.
- [7] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry, and S. González-Gil, "Do experts make mistakes? A comparison of human and machine identification of dinoflagellates," *Mar. Ecol. Prog. Ser.*, vol. 247, pp. 17–25, Feb. 2003.
- [8] M. E. Sieracki *et al.*, "Optical plankton imaging and analysis systems for ocean observation," in *Proc. Ocean Obs.*, vol. 9, 2010, pp. 21–25.
- [9] R. G. Colares, P. Machado, M. de Faria, A. Detoni, and V. Tavano, "Microalgae classification using semi-supervised and active learning based on Gaussian mixture models," *J. Brazilian Comput. Soc.*, vol. 19, no. 4, pp. 411–422, 2013.
- [10] I. Corrêa, P. Drews, M. S. de Souza, and V. M. Tavano, "Supervised microalgae classification in imbalanced dataset," in *Proc. 5th Brazilian Conf. Intell. Syst. (BRACIS)*, Oct. 2016, pp. 49–54.
- [11] I. N. da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, Eds., *Artificial Neural Networks: A Practical Course*. Cham, Switzerland: Springer, 2017.
- [12] R. F. Walker, K. Ishikawa, and M. Kumagai, "Fluorescence-assisted image analysis of freshwater microalgae," *J. Microbiol. Methods*, vol. 51, no. 2, pp. 149–162, 2002.
- [13] B. A. Hense, P. Gais, U. Jütting, H. Scherb, and K. Rodenacker, "Use of fluorescence information for automated phytoplankton investigation by image analysis," *J. Plankton Res.*, vol. 30, no. 5, pp. 587–606, 2008.
- [14] X. Hu, R. Su, F. Zhang, X. Wang, H. Wang, and Z. Zheng, "Multiple excitation wavelength fluorescence emission spectra technique for discrimination of phytoplankton," *J. Ocean Univ. China*, vol. 9, no. 1, pp. 16–24, 2010.
- [15] J. L. Deglint, J. Chao, and A. Wong, "A comprehensive spectral analysis of the auto-fluorescence characteristics of three algae species at twelve discrete excitation wavelengths," *J. Comput. Vis. Imag. Syst.*, vol. 3, no. 1, 2017. [Online]. Available: <http://www.jcvvis.net/>
- [16] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, nos. 285–296, pp. 23–27, 1975.
- [17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2012.
- [18] M. A. Wirth, "Shape analysis and measurement," *School Comput. Sci., Univ. Guelph, Guelph, ON, Canada, Tech. Rep. CIS 6320*, 2001.
- [19] J. D. Wehr, R. G. Sheath, and J. P. Kociolek, Eds., *Freshwater Algae of North America: Ecology and Classification*. Amsterdam, The Netherlands: Elsevier, 2015.
- [20] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

- [21] L. Breiman, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, 2017.
- [22] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Kuala Lumpur, Malaysia: Pearson, 2016.
- [23] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.
- [24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [25] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
- [26] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, 2009.
- [27] A. Géron, *Hands-on Machine Learning With Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Newton, MA, USA: O'Reilly Media, 2017.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [29] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer. (2014). "DenseNet: Implementing efficient ConvNet descriptor pyramids." [Online]. Available: <https://arxiv.org/abs/1404.1869>



ANGELA CHAO received the B.A.Sc. degree (Hons.) in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2018.

She is currently pursuing the M.S. degree in robotic systems development with Carnegie Mellon University, Pittsburgh, PA, USA, and hopes to pursue a career in robotic surgical devices. She is passionate about medical applications of engineering and has previously worked as a Research and

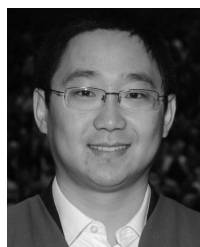
Development Intern with Baylis Medical Company. She received the Baylis Medical Award, in 2018, for her undergraduate capstone project in the design of a novel surgical instrument for cardiac valve repair.



JASON L. DEGLINT received the bachelor's degree (Hons.) in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2014, and the M.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016, where he is currently pursuing the Ph.D. degree in systems design engineering. He is currently being supervised by Dr. A. Wong and Dr. C. Jin. His research interests include computational imaging and using machine learning and multispectral imaging for computational biology and water conservation.

During his undergraduate studies, he worked in a variety of different companies as a co-op. First, in 2011, he was with Research In Motion (BlackBerry). Then, in 2012, he was with the National Aerospace Laboratory, Amsterdam, The Netherlands. Finally, in 2013, he was with MacDonald, Dettwiler and Associates Ltd. (MDA). During his M.A.Sc. degree, he was with Christie Digital Systems Inc., from 2014 to 2015. During his graduate studies, he founded Hedgehog Medical Inc., in 2016, which provides software to find and track artery walls from ultrasound images and videos.

His awards include the co-op Student of the Year Award, in 2012, from the University of Victoria, the Talbot Memorial Fund Award, in 2013, and the Andreas Antoniou Medal for Digital Signal Processing, in 2015. He also received the Distinguished Paper Award from the Society of Information Display, in 2015, and the Best Paper Award for the Conference of Computer Vision and Imaging Systems, in 2015. Finally, he received the Sandford Fleming Foundation Teaching Assistant Award, in 2016, from the University of Waterloo and the AquaHacking Challenge First Place Prize, in 2017.



CHAO JIN received the Ph.D. degree in civil engineering from the University of Waterloo, Waterloo, ON, Canada, in 2014. He currently holds a Research Assistant Professor position with the Department of Systems Design Engineering, University of Waterloo, working on several projects related to colloidal science and engineering, innovative device for water quality assessment, sustainability analysis for wastewater treatment, and sludge handling, and artificial intelligent applications for water and wastewater treatment processes.



ALEXANDER WONG (M'05–SM'16) received the B.A.Sc. degree in computer engineering, the M.A.Sc. degree in electrical and computer engineering, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005, 2007, and 2010, respectively. He is currently the Canada Research Chair of Artificial Intelligence and Medical Imaging, the Co-Director of the Vision and Image Processing Research Group, and an

Associate Professor with the Department of Systems Design Engineering, University of Waterloo. He has authored over 450 refereed journal and conference papers and patents, in various fields, such as computational imaging, artificial intelligence, computer vision, graphics, image processing, and multimedia systems. His research interests include integrative biomedical imaging systems design, operational artificial intelligence, and scalable and explainable deep learning.

He has received a number of awards, including two outstanding performance awards, the Distinguished Performance Award, an Engineering Research Excellence Award, the Sandford Fleming Teaching Excellence Award, an Early Researcher Award from the Ministry of Economic Development and Innovation, the Best Paper Award with the NIPS Workshop on NIPS Workshop on Transparent and Interpretable Machine Learning, in 2017, the Best Paper Award with the NIPS Workshop on Efficient Methods for Deep Neural Networks, in 2016, two Best Paper Awards by the Canadian Image Processing and Pattern Recognition Society, in 2009 and 2014, the Distinguished Paper Award by the Society of Information Display, in 2015, two Best Paper Awards for the Conference of Computer Vision and Imaging Systems, in 2015 and 2017, the Synaptive Best Medical Imaging Paper Award, in 2016, two Magna Cum Laude Awards and one Cum Laude Award from the Annual Meeting of the Imaging Network of Ontario, AquaHacking Challenge First Prize, in 2017, the Best Student Paper with the Ottawa Hockey Analytics Conference, in 2017, and the Alumni Gold Medal.

...