Research Article
Microalgae Classification Using Improved Metaheuristic Algorithm

Liu Wei, 1 Su XiaoPan, 2 and Faezeh Heydari 3
1 Gannan University of Science & Technology, Ganzhou, Jiangxi 341000, China
2 Ganzhou 851, Ganzhou, Jiangxi 341000, China
3 Faculty of Engineering, Saveh Branch Islamic Azad University, Saveh, Iran

Correspondence should be addressed to Faezeh Heydari; faezeh.heydari@t-mco.ir

Received 16 March 2022; Revised 5 April 2022; Accepted 21 April 2022; Published 11 May 2022

Academic Editor: Ramin Ranjbarzadeh

Copyright © 2022 Liu Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Microalgae are present at all levels of nutrients and food networks, so in aquatic environments they are an important part of the food chain of aquatic organisms, which also play an important role as biological purifiers of water resources and regulation. They also affect the pH of the environment; also plants are the only organisms capable of synthesizing long-chain fatty acids. Therefore, microalgae are the supplier and primary source of unsaturated fatty acids (PUFA) for all organisms present in the food chain of the aquatic environment. It should be noted that many microalgae are also biological indicators of water and reflect the ecological status of the environment. Precise classification of microalgae is related to the human observation capability. The present study proposes a new optimized classification technique with higher accuracy to provide a computer-aided classification of the microalgae. The method begins with an image segmentation to determine the region of interest. Then, the features have been extracted and fed to a Support Vector Machine (SVM) for final classification. The comparison results of the proposed method with some other methods show that the proposed method with 0.828 Kappa, and 0.342 and 0.855 min and max value of F1, provides the highest accuracy compared to the others.

1. Introduction

Long-chain polyunsaturated fatty acids (LC-PUFAs) such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) are among the essential components of various metabolic processes in human life [1]. There are significant quantities of PUFAs in oily fish products that are used to prevent heart disease and cancer [2]. Some efforts have been made to decrease the expense of these products by replacing fish oil ingredients with microalgae products. In the C. Cohn microalgae life cycle, two distinct types are considered: motile and nonmotile cells (cysts). Furthermore, the varying sizes of the microalgae signify different phases of the cell life cycle. Small cells, for example, represent a young offspring and cysts represent an adult cell. Motile cells typically emerge in an oval form, while nonmotile cells are usually circular and appear larger than motile cells. Since microalgae are the main manufacturer in the aquatic food chain and oxygen maker in the aquatic environment; it is a substantial microscopic aquatic life form [3]. Algae are considered as a biological index in water resource management to show water quality due to their sensitivity to environmental variations. This turns microalgae detection into a significant subject in water resource management. However, this purpose takes a lot of time due to the need for expert biologists to finish it. One simpler technique for reaching this aim is to use image processing [4]. This study presents a new method for automatic segmentation of microalgae in microscopic images. However, microalgae have a high significance; there are a small number of research works about them. For example, OSA et al. [5] proposed a method for the segmentation of green microalgae images based on region growth. The method was started by a morphological feature extraction followed by a taxonomical classification of the species. The method is based on the seeded region growing principle and a fine-tuned filtering preprocessing stage,
which are used for smoothing the input image. Simulation results indicated that the proposed method provides proper precision for green microalgae image segmentation compared with some state-of-the-art methods.

Giraldo-Zuluaga et al. [6] proposed an automatic detection system for Scenedesmus polymorphic microalgae in microscopic images. In the study, the image was first prepossessed by contrast equalization and then the image was segmented by the active contours. The statistical features evaluation with texture features was presented for algae description.

Wei et al. [7] proposed a method for improving the hyperspectral microscopic imaging system to provide hyperspectral images of microalgae samples over pretreated steps. Afterward, the Fisher algorithm was utilized for microalgae species identification. For analyzing the ability of the method, it is assessed in terms of sensitivity and specificity and the results showed high resolution for these two measurement indicators.

Final results indicated that using the proposed method provides fast and suitable results.

Liao et al. [8] presented a method based on deep learning to identify marine microalgae. The study used a convolutional neural network (CNN) to classify the marine microalgae. The study addressed two collections of classification experiments, where one group categorized the algae into 5 classes based on the family category, and the other one categorized the algae into 11 classes based on species category. Simulation results indicated that the polarization information developed using the high and robust classification precision of the proposed method for low-resolution microagal images.

Xu et al. [9] proposed an identification methodology for microalgae images based on transmission hyperspectral microscopic imaging and transmission hyperspectral microscopic imager (THMI). Hyperspectral imaging (HSI) of three species of microalgae was proposed to confirm their absorption features. Principal Component Analysis (PCA) and peak ratio algorithms are used to analyze the transmission spectra for dimensionality reduction and feature extraction, and finally, the Support Vector Machine (SVM) model was employed for classification. Simulation results indicated that the proposed method has a good potential for the classification of microalgae.

As can be observed from the literature, lots of works were performed to provide a proper identifier for the microalgae. However, several works were done for this purpose; the recent works showed that using optimized techniques, especially metaheuristic algorithms, provides higher efficiency for microalgae classification. This motivates us to provide a new modified metaheuristic to use as an efficient tool for the classification. Therefore, in this research, we worked on a high-ability classifier for the identification of the microalgae in the images. The main contributions of this study can be highlighted as follows: (1) Providing a new technique for better segmentation of microalgae. (2) Designing a new improved version of the metaheuristic technique, called Quantum Thermal Exchange Optimization (QTEO) algorithm for this study. (3) Using the designed technique to optimize Kapur’s entropy thresholding method. (4) Validation of the proposed method with some state-of-the-art methods. In the next sections, the method for this purpose has been described and validated in detail.

2. Histogram Equalization

Image histogram equalization is one of the image processing methods that is used to better display the output image or for further analysis [10–16]. A simple example of histogram equalization for microalgae images is shown in Figure 1.

As can be observed from Figure 1 that the probability density value for the equalized image is more distributed than the original image. This can be so helpful for image segmentation and extracting helpful features from the images.

3. The Quantum-Based Thermal Exchange Optimization Algorithm

3.1. Conception. Optimization in engineering and science refers to selecting the best member from a set of achievable members [17, 18]. In the simplest form, an attempt is made to systematically select data from an achievable and computable set to obtain the optimal value (minimum or maximum) of a real function [19]. Many optimization problems in engineering are naturally more complex and difficult than can be solved by conventional optimization methods such as dynamic programming methods and the like [20]. Several optimization problems today, including nonpolynomial (NP-Hard) problems, are approximately solvable with existing computers [21]. One of the solutions to deal with such problems is the use of approximate and inexact algorithms [22]. These algorithms do not guarantee that the solution obtained is optimal and only with a lot of time can a relatively accurate solution be obtained, and, depending on the time spent, the accuracy of the solution changes [23]. One way to solve these kinds of problems is to consider all possible solutions and calculate the objective functions associated with them and, finally, select the best solution [24]. It is clear that the method of complete counting ultimately leads to an accurate solution to the problem, but in practice, it is impossible to use it due to a large number of possible solutions [25]. Due to the problems related to the complete counting method, there has always been an emphasis on creating more effective and efficient methods [26]. In this regard, various algorithms have been developed, for example, Genetic Algorithm (GA) [27], Ant Lion Optimizer (ALO) algorithm [28], equilibrium optimiser [29], Mayfly Optimization Algorithm (MOA) [30], and World Cup Optimization (WCO) algorithm [31].

Recently, a new metaheuristic, called Thermal Exchange Optimization (TEO) algorithm, has been proposed, which is derived by the temperature behavior of the objects and their positions and their variations between cold and warm parts [32]. In this study, a new improved version of the TEO algorithm, called the Quantum-based TEO (QTEO) algorithm, has been proposed to improve the precision and consistency of the original method. The main
target is to utilize this new algorithm for microalgae segmentation. The main conception of the method is based on the Newton law of cooling. This law explains that the body temperature ratio varies proportionally around the body temperature. If the temperature variation in the body and its neighborhood is small, the mean value of the heat exchanged between the body and its neighborhood is almost proportional to the difference in temperature of the body and the environment which is due to the convection, infrared radiation, and conduction. The law of cooling is achieved by the following equation [33]:

\[ \frac{dQ}{dt} = \alpha \times A \times (T_s - T_a), \] (1)

where \( \alpha \) describes the coefficient of the heat transfer, \( Q \) signifies the heat, \( A \) states the area surface which transmits heat, object geometry, and surface state, and \( T_a \) and \( T_b \) represent the ambient and body temperatures.

The losing time is \( \alpha \times A \times (T_a - T)dt \) which determines the temperature variation in the reserved heat as the temperature falls \( dT \), i.e.,

\[ V \times c \times \rho \times dT = -\alpha \times A \times (T - T_b)dt, \] (2)

where \( \rho \) signifies the density \((kg/m^3)\), \( V \) describes the volume \((m^3)\), and \( c \) defines the specific heat \((J/kg/K)\).

Henceforth [33],

\[ \frac{T - T_b}{T_M - T_b} = \exp \left( -\frac{\alpha \times A}{\rho \times V \times c} \right), \] (3)

where \( T_M \) represents the early high temperature. The above equation is correct when \( \alpha \times A \times t/V \times \rho \times c \) does not depend on \( T \):

\[ \zeta = \frac{\alpha \times A}{V \times \rho \times c}. \] (4)

Therefore, by considering \( \zeta \) as a constant,

\[ \frac{T - T_b}{T_M - T_b} = \exp (-\zeta t). \] (5)

Consequently,

\[ T = (T_M - T_b) \times \exp (-\gamma t) + T_b. \] (6)

3.2. The Algorithm. Based on the Thermal Exchange Optimization (TEO) algorithm, some solution candidates are assumed as cooling ingredients and the remaining candidates are considered as the environment. Afterward, this process is established reversely.
The TEO algorithm starts with initializing a certain number of randomly distributed candidates as the solution candidates. This is done by the following equation [34]:

\[ T_i^0 = T_{\min} + \delta \times (T_{\max} - T_{\min}) \quad i = 1, 2, \ldots, n, \]  

where \( \delta \) describes a random value in the range [0, 1], \( T_i^0 \) represents the algorithm primary population for the \( i^{th} \) object, and \( T_{\min} \) and \( T_{\max} \) represent the minimum and maximum limitations. The candidates are then evaluated on the cost function to verify their ability. Afterward, \( T \) number of best candidate vectors are saved in Thermal Memory (TM) to provide higher efficiency of the algorithm with lower complexity. Some of these candidates are then added up to the main candidates and the same number of them in the individuals has been eliminated. The candidates contain two equal types, including environment and the heat and cooling transfer objects. Figure 2 shows this state.

As can be seen from Figure 2, the candidates contain two equal types, including environment and the heat and cooling transfer objects. Assume \( T_j \) as environment object for \( T_{\omega(t+1)} \) cooling object and vise versa. If \( \zeta \) is greater than the object, the temperature exchanges steadily. \( \zeta \) is mathematically achieved by the following:

\[ y = \frac{\cos(\text{object})}{\cos(\text{worst object})} \]  

To model the effect of time in the algorithm, the following formula has been considered:

\[ t = \frac{\text{iteration}}{\text{Max.iteration}} \]  

For establishing the global searching, the environmental temperature variation is considered that can be assumed by the following equation:

\[ T_i^n = (1 - (m_1 + m_2 \times (1 - t) \times \text{rand})) \times T_{i}^{\ast}, \]  

where \( T_i^n \) defines the preceding object temperature modified by \( T_i \), and \( m_1 \) and \( m_2 \) describe the control variables, respectively. By considering the previous models, the new object temperature is mathematically updated as follows:

\[ T_i^+ = T_i^n + (T_i^{\text{old}} - T_i^n) \exp(-\zeta t). \]

Finally, the term \( Pr \) is a component to show the cooling of the objects.

The \( Pr \) candidates are compared with randomly distributed values, \( R(i) \) in the range [0, 1].

In this situation, if \( R(i) < Pr \), one dimension of the \( i^{th} \) the candidate is randomly chosen and the value is rephrased as follows:

\[ T_{i,j} = T_{j}^{\text{old}} + \text{rand}(T_{j}^{\max} - T_{j}^{\min}) \exp(-\zeta t), \]  

where \( T_{i,j} \) represents the \( j^{th} \) variable of the candidate number \( i \) and \( T_{j}^{\text{min}} \) and \( T_{j}^{\max} \) describe the lower and the upper limitations of the \( j^{th} \) variable, respectively.

The algorithm is stopped if the termination condition has been reached.

3.3. The Quantum Thermal Exchange Optimization (QTEO) Algorithm. However, Thermal Exchange Optimization (TEO) algorithm is a well-organized optimization technique; it is sometimes stuck in the local optimum solution. This study uses a quantum theory-based modification to resolve this issue as is possible. Since the quantum position is not instantaneously definite, it should be determined by a wave function \( \phi(X,t) \), where the candidates’ position is known by \( X \). This states that the square mode defines the density probability of the individual in space at position \( X \). The equation for this case can be formulated as follows:

\[ |\phi|^2 dx dy dz = Q \text{ dxdydz}, \]  

where \( Q \) signifies the probability density function to describe the following conditions:

\[ \int_{-\infty}^{\infty} |\phi|^2 dx dy dz = \int_{-\infty}^{\infty} Q \text{ dxdydz} = 1. \]

The formulation for the position of the candidates is achieved by the following:

\[ P(t) = \delta \times P_g(t) + (1 - \delta) \times P_i(t), \]

\[ m_i(t) = \frac{1}{M} \sum_{j=1}^{M} P_i(t), \]

\[ D(t + 1) = 2\beta \times |m_i(t) - X(t)|, \]

\[ \beta = \frac{\gamma}{\zeta} (\sqrt{\beta} - \frac{\beta}{\sqrt{\beta}}) \times \left( \frac{t}{G_{\text{max}}} \right), \]

\[ X(t) = P(t) \pm \left( \frac{D}{2} \right) \times \ln\left( \frac{1}{y} \right). \]

\( M \) defines the population size, \( P(t) \) indicates the local attraction area for the \( i^{th} \) iteration, \( m_i(t) \) signifies mean value of the optimal situation, \( D \) determines the weighted distance between the individuals and the mean optimal situation of the population, \( G_{\text{max}} \) describes the maximum number of iterations, \( P_g \) and \( P_i \) represent the global optimal positions and the candidate position, respectively. \( \delta \) and \( \gamma \) provide two stochastic values in the range [0, 1], and \( \beta \) signifies the shrinkage-expansion coefficient in the range \( \beta \) to \( \bar{\beta} \). Here, \( \beta \) and \( \bar{\beta} \) are 1 and 0.5, respectively. Based on the above explanations, the initial position can be updated as follows:

\[ T_i^0 = P(t) + \beta \times |m_i - x_k| \times \ln\left( \frac{1}{y} \right). \]

4. Segmentation of Microalgae

4.1. Concept of Entropy Criterion (Kapur’s Method). Image segmentation is one of the most important basic operations for image analysis. The purpose of segmenting images is to divide the image into unequal areas so that these areas are not shared. Image segmentation is used in many applications of image processing and computer vision, especially medical imaging. In recent years, many methods have been proposed for image segmentation. These methods can be divided into
two general categories of supervised and unsupervised methods. The unsupervised method is preferred in real-time applications because it does not require manual segmentation. Thresholding is one of the most important methods in image segmentation. The main goal in threshold-based methods is to find a threshold value for two-level thresholds and a few threshold values for multilevel thresholds. In two-level thresholding, only one threshold value is selected and the image pixels are divided into two groups. A popular method for thresholding is Kapur thresholding. The Kapur thresholding is a nonparametric technique to provide optimal threshold values based on image entropy and the histogram of the considered image. The key idea is to maximize the image entropy to get the optimal threshold value. To get this end, it assumes TH as a vector of the image threshold, i.e. \[ \text{TH} = \{th_1, th_2, \ldots, th_{m-1}\} \] (17)

In this situation, the Kapur entropy method is defined by the following equation:

\[ J_{\text{max}}(\text{TH}) = \sum_{i=1}^{m} H_i^c, \]
\[ c = \begin{cases} 
1, & \text{if RGB scale image,} \\
2, & \text{if Gray scale image.}
\end{cases} \] (18)

\( H_i^c \) describes the \( i^{th} \) entropy of the image and is obtained as follows:

\[ H_i^c = \frac{1}{m_i} \sum_{c=1}^{m_i} \ln \left(\frac{H_i^c}{\omega_i^c}\right), \]
\[ H_1^c = \frac{1}{m_1} \sum_{c=1}^{m_1} \ln \left(\frac{H_1^c}{\omega_1^c}\right), \]
\[ H_2^c = \frac{1}{m_2} \sum_{c=1}^{m_2} \ln \left(\frac{H_2^c}{\omega_2^c}\right), \]
where \( \omega_i^c \) describes the \( i^{th} \) histogram for \( c_i \), and \( H_i^c \) signifies the histogram of the image intensity levels. \( \ln(.) \) defines the natural logarithm.

This study uses the formerly explained new designed Quantum-based Thermal Exchange Optimization Algorithm to achieve the optimal threshold values by threshold points tuning until \( J_{\text{max}} \) is met.

4.2. Image Thresholding Using Quantum-Based Thermal Exchange Optimization Algorithm. Each one of the candidates in the proposed algorithm shows a decision variable for the threshold values and image segmentation. The population can be indicated as follows:

\[ AT = [x_1^c, x_2^c, \ldots, x_{\text{candidate}}^c]^T, \]
\[ x_i^c = [th_1^c, th_2^c, \ldots, th_{m_i}^c], \]
\[ c = 1, 2, 3, \]

where \( AT \) describes the population size, \( x_i \) signifies the \( i^{th} \) element of \( AT \), \( T \) indicates the transpose sign, and \( c \) defines set for RGB images while \( c = 1 \) is selected for grayscale images. The boundaries of the search space in the present study are between 0 and 255, which point to the image intensity levels.

4.3. Image Segmentation Evaluation. The proposed QTEO-based segmentation method is applied on MATLAB R2017b platform and based on a laptop with Intel\textsuperscript® Core\textsuperscript™ i7, and 16 GB RAM. The parameters of the proposed QTEO-based segmentation method are initiated with following values: the initial numbers of continents are selected as 5 and the number of individuals is set as 100. The value of the \( m_1 \) and \( m_2 \) is set to be 2, the dimension of the search space is \( Th \) (chosen threshold), and the maximum number of iterations is set 200. The algorithm will be terminated after maximizing the objective function \( J_{\text{max}} \). To obtain a fair analysis, the method has been performed 30 times iteratively and the mean value is considered as the optimal threshold. In the present study, several databases from the Internet are tested and assessed to determine the performance of the suggested segmentation methodology. Figure 3 and 4 show the image segmentation based on QTEO-Kapur.

As can be observed from figures, the method can provide useful results for the segmentation of the microalgae.

5. Feature Extraction

The next step is to use a feature extraction to extract the segmented image features for classification of the image with higher accuracy and less time. In this study, the Gray-Level Cooccurrence Matrix (GLCM) is utilized as the feature extraction method. Tissue characteristics were evaluated by tissue statistical analyses over the statistical distributions of grayscale compounds observed in a specific position about each other. The GLCM is a matrix with identical numbers of rows and columns that shows the grayscale levels of the image. In other words, if an image gray number is up to \( P \),
the dimensions of the simultaneous event matrix will be $P \times P$ matrix. The current study assessed the features including entropy, contrast, homogeneity, energy, and correlation. The formulation of these features is given below. The correlation feature is achieved as follows:

$$\text{corr} = \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j}$$  \hspace{1cm} (21)

The correlation feature is achieved as follows:
The entropy of the features is achieved by the following:

\[
\text{Entropy} = - \sum_{i=1}^{M} \sum_{j=1}^{N} p(i, j) \log p(i, j).
\]  

The energy of the features is obtained as follows:

\[
\text{Cont} = \sum_{i=1}^{M} \sum_{j=1}^{N} p^2(i, j).
\]
and finally, the homogeneity is achieved by the following equations:

\[ H_{gy} = \sum_{i=1}^{M} \sum_{j=1}^{N} p(i, j) \frac{1 + |i - j|}{\sigma(i, j)^2} \]

(25)

where \( \mu \) signifies the mean value, \( \sigma \) describes the standard deviation, \( p(i, j) \) states the intensity values for pixels at position \( (i, j) \), and \( M \) and \( N \) represent the image size.

6. Simulation Results

Various databases have been proposed for assessing the microalgae diagnosis systems. In this study, we use the AlgaeVision database as a popular one for the analysis. AlgaeVision database contains a virtual image collection of freshwater and terrestrial algae that are collected from Britain and Ireland. AlgaeVision includes 2,300 images of nearly 250 genera and 680 species of British and Irish algae [36]. The data includes 19 different classes. The minimum number of the elements of the class is 10. The classes that are considered in this study are Bacillariophyceae-1, Spirotrichea, Peridiniales, Nanoplankton <5 \( \mu \)m, Nanoplankton 5–10 \( \mu \)m, Gymnodiniales, Prorocentrales, Cochlodinium, Ellipsoidal Microplankton, Ellipsoidal Nanoplankton 10–20 \( \mu \)m, Spherical Nanoplankton 10–20 \( \mu \)m, Spherical Microplankton, Pseudonitzschia, Chaetocerotales, Coscinodiscophyceae, Bacillariophyceae-2, Naviculales, Corophthron, and Rhizosoleniales.

For recognizing algae, the extracted features must be classified. In this study, the Support Vector Machine (SVM) algorithm is employed as a fast and simple classifier to classify positive and negative examples with a maximum margin. The SVM is an N-dimensional set of points that indicates the boundaries of the categories. The SVM algorithm finds the best limit in the data and makes the highest distance as much as possible from all categories. The idea is to recognize the best decision surface as follows:

\[ y = \text{sgn} \left( \sum_{i=1}^{N} y_i a_i K(x, x_i) + b \right) \]

(26)

where \( x \) defines a \( d \)-dimensional test set vector, \( x_i \) describes the \( i \)th training set vector, \( y \) represents a class label in the range \(-1 \text{ and } 1\), \( N \) is the training set numbers, \( a = [a_1 \ldots a_N] \) and \( b \) represent the model parameters, respectively, and \( K(x, x_i) \) states a kernel function. Here, the Sequential Minimal Optimization (SMO) is used for solving the quadratic programming (QP) problem of the SVM to remove the need for an extra storage matrix. To authenticate the efficiency of the suggested approach, different measurement indicators have been adopted. Due to the presence of multiple classes, we used specific features including Kappa coefficient, macro-, and microaverage. Kappa metric describes the agreement level of two or more observers in the same class.

### Table 1: The classification results in the microalgae dataset.

<table>
<thead>
<tr>
<th>Method/Indicator</th>
<th>Kappa</th>
<th>Max F1</th>
<th>Min F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMO</td>
<td>0.828</td>
<td>0.342</td>
<td>0.855</td>
</tr>
<tr>
<td>Region growing [37]</td>
<td>0.746</td>
<td>0.329</td>
<td>0.749</td>
</tr>
<tr>
<td>NAABB [38]</td>
<td>0.543</td>
<td>0.284</td>
<td>0.768</td>
</tr>
<tr>
<td>Label-free [39]</td>
<td>0.679</td>
<td>0.348</td>
<td>0.739</td>
</tr>
<tr>
<td>Neural network [40]</td>
<td>0.637</td>
<td>0.304</td>
<td>0.849</td>
</tr>
</tbody>
</table>

### Table 2: The resampling results were performed with several sample sizes from 25% to 200%.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Kappa</th>
<th>Max F1</th>
<th>Min F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.703</td>
<td>0.329</td>
<td>0.851</td>
</tr>
<tr>
<td>50</td>
<td>0.725</td>
<td>0.337</td>
<td>0.846</td>
</tr>
<tr>
<td>100</td>
<td>0.732</td>
<td>0.343</td>
<td>0.831</td>
</tr>
<tr>
<td>200</td>
<td>0.733</td>
<td>0.353</td>
<td>0.826</td>
</tr>
<tr>
<td>Neural network [40]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.708</td>
<td>0.506</td>
<td>0.843</td>
</tr>
<tr>
<td>50</td>
<td>0.725</td>
<td>0.627</td>
<td>0.915</td>
</tr>
<tr>
<td>100</td>
<td>0.764</td>
<td>0.711</td>
<td>0.978</td>
</tr>
<tr>
<td>200</td>
<td>0.777</td>
<td>0.724</td>
<td>0.988</td>
</tr>
<tr>
<td>Region growing [37]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.620</td>
<td>0.296</td>
<td>0.843</td>
</tr>
<tr>
<td>50</td>
<td>0.625</td>
<td>0.305</td>
<td>0.831</td>
</tr>
<tr>
<td>100</td>
<td>0.613</td>
<td>0.325</td>
<td>0.827</td>
</tr>
<tr>
<td>200</td>
<td>0.634</td>
<td>0.364</td>
<td>0.796</td>
</tr>
<tr>
<td>NAABB [38]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.649</td>
<td>0.316</td>
<td>0.758</td>
</tr>
<tr>
<td>50</td>
<td>0.658</td>
<td>0.328</td>
<td>0.746</td>
</tr>
<tr>
<td>100</td>
<td>0.676</td>
<td>0.346</td>
<td>0.723</td>
</tr>
<tr>
<td>200</td>
<td>0.681</td>
<td>0.358</td>
<td>0.698</td>
</tr>
<tr>
<td>Label-free [39]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.538</td>
<td>0.324</td>
<td>0.712</td>
</tr>
<tr>
<td>50</td>
<td>0.554</td>
<td>0.349</td>
<td>0.708</td>
</tr>
<tr>
<td>100</td>
<td>0.562</td>
<td>0.382</td>
<td>0.692</td>
</tr>
<tr>
<td>200</td>
<td>0.591</td>
<td>0.416</td>
<td>0.697</td>
</tr>
</tbody>
</table>

Kappa's value is between \(-1 \text{ and } 1\), where 1 describes the full agreement and 0 defines that the achieved results are too less than the expected value for agreement. F-Score metric indicates the harmonic mean value between the recall and precision. The MaxF1 and MinF1 describe the maximum value and the minimum value of the F-Score. The term MinF1 has more effect on the classifier efficiency on populations with more classes that represents the imbalanced dataset limitation. This metric is one of the widely used indicators to assess classifiers. Furthermore, the term MaxF1 has been affected by classes with small number of elements, which has an effective impact on the results evaluation. A 10-fold cross-validation method is used for evaluating the models. Table 1 illustrates the classification results in the microalgae dataset.

For analysis of the robustness of the algorithm and to achieve the oversampling strategy, a random resampling method has been used. This method can over/undersample the dataset based on a percentage through some original number of instances. This technique generates a random set of sample data by sampling with (or without) replacement. The uniformity of the resampled dataset in this technique is defined by a parameter called bias factor. It changes in the range \([0, 1]\) where the values near 1 make a uniform dataset. The presented study evaluates different bias factors. This
method allows the classifier to retain smaller classes while optimizing accuracy. As can be observed from Table 1, the value of MinF1 is larger than the MaxF1 values, which are due to the high efficiency of the algorithm to correctly organize the majority classes. The term MaxF1 has been more affected by the smaller classes. Table 2 indicates the resampling results which are performed with several sample sizes from 25% to 200% to deal with the imbalance issue.

Table 2 illustrates the improvement of the results. The method has a satisfactory improvement in terms of MaxF1; therefore, it can be observed that better results are obtained by the SMO. Moreover, the Kappa value achieved by the proposed SMO algorithm provides a larger growth slope which is predicted due to its better efficiency for Gaussian distributed spaces.

As mentioned in this paper, the proposed method uses an optimal technique using the Kapur method for segmentation of the microalgae. The main advantage of this technique is that using metaheuristic-based techniques can help to run from the local minimization of the problem. However, this can make the problem about more complexity of the proposed method toward the classical one.

7. Conclusions

The present study proposed a new optimized and pipelined method for in-line detection of microorganisms. The method was first used as histogram equalization to improve the quality of the image quality. Then an image segmentation based on an optimized Kapur method was designed and used for separating the microalgae. The optimization of the Kapur in the study is based on a new improved version of the Quantum-based Thermal Exchange Optimization Algorithm. The new algorithm was designed for improving the segmentation accuracy. The Gray-Level Cooccurrence Matrix (GLCM) was then used for extracting the main features of the lesion area. A 10-fold classification was then performed to classify the Sequential Minimal Optimization- (SMO-) based Support Vector Machine (SVM). The suggested method was compared with some state-of-the-art methods by three different metrics. The final results showed that the suggested technique with 0.828 Kappa, and 0.342 and 0.855 min and max value of F1, provides the best efficiency against the other compared methods.

Data Availability

The data are available at https://algaevision.myspecies.info/

Conflicts of Interest

The authors announce that no conflicts of interest exist in this research.

Acknowledgments

This research was supported by the Educational Science Foundation of Jiangxi Province (# 41562019).

References

10 Mathematical Problems in Engineering


