British Journal of Computer, Networking and Information Technology
ISSN: 2689-5315
Volume 2, Issue 1, 2019 (pp. 34-44)
www.abjournals.org

SPIRULINA DETECTION USING DEEP LEARNING APPROACH

Othman Siddik[1] and Atila Bostan[2]

[1]Department of Modelling and Design of Engineering Systems, Faculty of Graduate School of Natural and Applied Sciences, University of Atilim.
[2]Department of Computer Engineering, Faculty of Graduate School of Natural and Applied Sciences, University of Atilim.

ABSTRACT: *Spirulina is an algal microorganism that comes in four species. It is a green growth microorganism are widely used in water quality determination and monitoring. Profound learning and convolutional neural systems are turning into a broadly utilized strategy for picture classification in a number of domains. Moreover, Deep-Learning and Convolutional Neural Networks are yielding better outcomes and utilize strategy in a variety of picture classification problems. Computerized discovery of Spirulina is one of the most top research topics in the field. For this study, a broad spirulina picture dataset was specifically built. This paper presents the outcomes of a study which utilised Convolutional Neural Network technology in the spirulina discovery problem in order to ascertain whether it is appropriate to identify Spirulina. A comprehensive spirulina picture dataset was specifically devised utilizing an artificial picture augmentation technique on real pictures of spirulina gathered from waterways in Turkey. The dataset covers distinctive light conditions and it was computationally expanded from 60 original sample pictures to 1000 pictures. In this study, a customized Convolutional Neural Network configuration that settles the 4-class spirulina detection and identification problem from an extensive set of microscopic pictures are proposed. Outcomes are examined and compared with those of past investigations. We present a preliminary investigation that utilizes the Convolutional Neural Network in the problem of discovering spirulina.*

KEYWORDS: Deep Learning, Convolutional Neural Networks, Classification, Image Acquisition, Spirulina Detection.

INTRODUCTION

As can be seen in the European Parliament global reference and the Council of the European Union's Water Policy Directive Framework, a few classes of water assets serve as bio indicators for quality and other factor [1]. Diatom detection indicators commonly require a particular examination which requires expert to be able to classify diatoms in water. Moreover, the depiction of morphological spirulina and frustules (the arrangement of interfacing, siliceous connections with a pipe) can vary. Some varieties of spirulina, viewed as customary classes for a long time, are at present isolated into totally separate species and the advancement of refined classes persists [2]. The improvement of computerized microalgae location devices that measure structural information would be significant within a full application scope of for every fully trained specialist and non-specialist. Accomplishments are found in related work for the discovery of microalgae [3]. Additionally, there were no research found still on Spirulina all related works will be from different types of diatoms in water, for detection spirulina CNN are used which Convolutional Neural Network (CNN) is a model for all in-depth learning, Neural networks with deep classes are called convolutional neural network CNN and it's with more layers for detection features and in creating implementing the neurons this is essentially new behaviors in machine learning theories. CNN contains number of layers the importance to control tuning for statistical parameters needed for testing and the way of training model to achieve good results.

LITERATURE REVIEW

Diatom automated detection continues to be an associate level public challenge. Performances have been described in the related work for the detection of microalgae, however, all of them have done based on general image features detection that is not enough, and as a result, a lower rate of success and lower results showed were there are large numbers of diatom types for detection, as mentioned in the related work of this paper.

Bueno, G. and Deniz, O used feature detection to detect diatom in water with an entire data of 24,000 segmented units which has 80 classes, and the system reached 98% efficiency of successful detection rate [4]. In 2012 a paper submitted by Dimitrovski et al his paper is with 1093 image data set which has 55 classes the efficiency rate was 96.17% [5]. In 2002 DuBuf H used Bagging Tree classifier over 781 samples with 96.9% accuracy [6]. In 2003 Pappas J used Multiple Discriminant Analysis classifier over 66 samples with 80.3 accuracy [7]. Du Buf and Bayer proposed Diatom Identification they used RGB level formated images in extracting diatom from the image. They report an accuracy rate of 83.4% [8]. In this study we have studied on 1000 microscopic images which included 4 varieties of Spirulina. We utilized deep learning technology in detection of Spirulina. To the best of our knowledge and previous studies we have encountered in the literature, this is the preliminary employing CNN in Spirulina detection.

METHODOLOGY

In this part, the methodology followed in this study is presented. The study is divided into two distinct processes, namely data preparation and machine learning phases. In the first step row images of spirulina and non-spirulina were standardized and augmented to aid the machine learning process. Later, CNN was utilized as the spirulina detection technology in the machine learning phase. In this section, initially the formation and augmentation of the experimental picture data set is explained, followed by description of the technologies used in the experimentation (classification) and comparison. The first step includes building a dataset of microscopic spirulina images. Following a series of preprocessing operations on the original spirulina images, such as image augmentation, labeling and recovery, an experimental set of 1000 images was obtained. Details of the image augmentation methodology are similar to those used in [9]. The following sections present further details on the experimental dataset, classification practice and comparison methodology.

Data Labelling

All original images were collected from rivers and lakes in Turkey and they had been labeled by two experts who are studying in the Biology Department at Selçuk University of Turkey. The images were taken using a Sony DSC-W570 16.1-megapixel camera. Spirulina images were labeled into 4 classes, namely Laxa, Major, Nordst and Princeps, Class-labelling of data were performed manually using optical microscopes. Total number of original images is 90 where 60 spirulina and 30 non-Spirulina. This section presents a deeper understanding on image modification and augmentation used in the study.

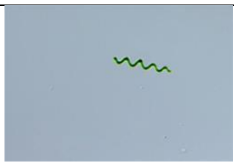
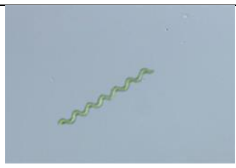
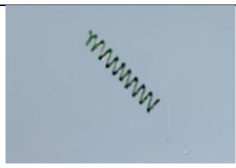
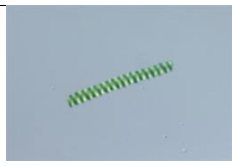
Image Processing

Following the image labelling process image pre-processing operation were applied on the image dataset. The original images obtained from the microscope varied in size from 120 × 120 to 500 × 500 pixels. For uniformity of size, each image was resized to 270 x 270 pixels using the Nearest-Neighbor interpolation method and by supplanting each pixel with the closest pixel in the output.

Image Dataset Composition

A total of 60 original single-cell spirulina images were collected out of four different types of spirulina in various rotations and sizes and 30 non spirulina images. However, because the number of original images were evaluated to be unsatisfactory for a machine learning process, it is deemed necessary to augment the image dataset. In order to generate close-original images out of actual ones, we rotated the objects in original images with random degrees and concocted various possible combination of them in to an artificial image. No filter was applied on the images. Computational delays were avoided by fixing every image to 270× 270 pixels, the artificial image generation process yielded 300 in total including original images that are 60 spirulina and 30 non-spirulina images. So that 210 images were artificially formed. However, this data set was not suitable to run a deep learning training model as most of images were composed of multi object spirulina. Therefore, it became necessary to crop each single object in the image discretely. Finally, 500 spirulina images and 500 non-spirulina object images became available in our dataset for the 4 species of spirulina. Tables 1 and 2 show the original and modified image counts along with sample images in each class. Here we used type A for (Laxa) and B for (Major) and C for (Magnifica) and D for (Princeps) as original image input.

Table 1a: Spirulina Image Composition

Varieties of Spirulina	(A)	(B)	(C)	(D)
Original Image Sample & Count (60)	 15	 22	 13	 10

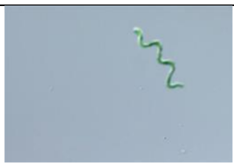
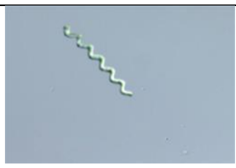
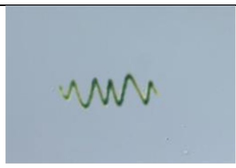
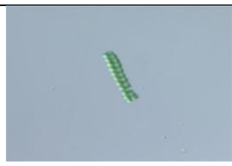
Modified Image (Rotation-Resizing) Sample & Count (500)	 121	 184	 115-	 80

Table 2b: Non-Spirulina Image Composition

Non-Spirulina Images Sample & Count (500)	 **500**

Deep Learning

In this study CNN technology was used in the classification process. In the following sections, a comprehensive summary of CNN technology is presented, after which special aspects of validation, testing and preparation are described.

Convolutional Neural Networks

Convolutional Neural Networks (ConvNets or CNN) is a category of Neural Networks that have proven to be very effective in areas such as image recognition and classification. CNN have been successful in identifying faces, objects and traffic signs as well as powering vision in robots and self-driving cars. ANN with deep classes are called convolutional neural networks due to there being a higher number of layers for the detection of features and in creating and implementing neurons. There are essentially new behaviors in machine learning theories. A CNN contains a number of layers which are important in controlling tuning for statistical parameters needed for testing and training models to achieve good results. Every single image sample is fed as an input to train the model by calculating its features. With this repetition, the net will be enhanced by pushing it closer to a solution which is required to reduce the loss function. Moreover, there are specific scientific tools to resolve this problem, including the Nesterov Accelerated Gradient (NAG) presented in paper [11], the Stochastic Gradient Descent, and the Adaptive Gradient explained in paper [12]. In this study Alex Net is used in Spirulina detection. Details for Alex Net is presented in section 3.5.

Tuning

AlexNet has been used in several areas in diatom detection. AlexNet was used in object detection from water in the study conducted in 2019 [13], the other study in 2017 utilized AlexNet in diatom classification on a 60,000 image date set which covered 80-class. They have reported 93% accuracy in classification [14], as per the AlexNet, in another current study [15] Krizhevsky et al. preferred this tool because of its GPU support. Readily available Alex net is advantageous to the other ones in field as per the performance. With these current trends and wide usage, we have opted for AlexNet as the tool in this study as well. Figure 1 shows the design of the network. We can see in detail below the Alex Net layers from one to five that are the convolution layers and the last three layers being fully connected layers:

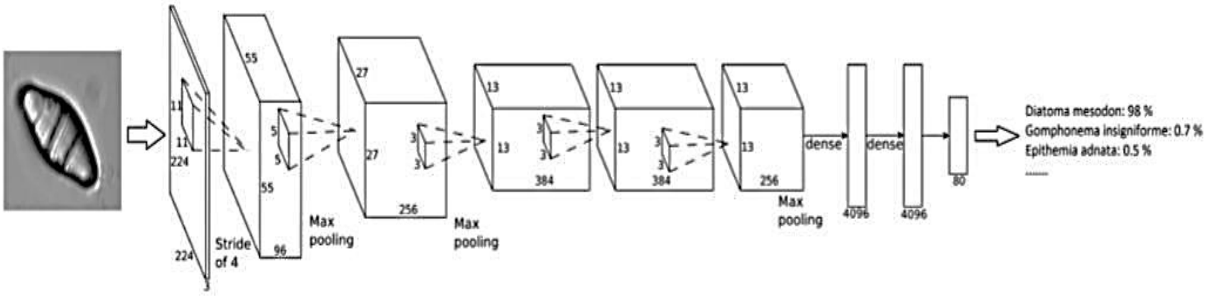


Figure 1: Alex Net Network architecture [13].

In the training part, the rate of initial learning and the period of the regularization technique for neural networks for back propagation is separate from the main tuning part. SGD (the Stochastic Gradient Descent) was used to compute the next update to the parameters and produce the best result. The deep neural network had two main problems, as explained and reported in the structure paper [6]. The first problem was data development as the model is able to fit with several images in the same direction and rotations. Every feature would have been predicted only in the direction in which it was trained. It was extremely important to find any duplicate images in the data set and ensure that different versions of the same samples were provided. The second main problem was in the architectures of the dropout layers such that some of the layers that detected features were turned off. Here, the idea was to "turn off" randomly some of the neurons where the network avoided local dependencies, making us aware that this was a robust design.

Validation

Confusion matrix was utilized for performance quantification. Confusion matrix is a tabular description which declared the connection between the predicted case as spirulina or non-spirulina. In the performance measurement the results of the classification were divided in to four classes of which the definitions are given below.

- True Positive (TP): count of spirulina labeled objects while the object is actually spirulina.
- False Positive (FP): count of spirulina labeled objects while the object is actually non-spirulina.
- True Negative (TN): count of non-spirulina labeled objects while the object is actually non-spirulina.

- False Negative (FN): count of non-spirulina labeled objects while the actual object is spirulina.

We calculated the sensitivity in order to quantify the likelihood of study-model to label an object as spirulina while the object is actually a spirulina. The formula to compute the sensitivity is given below.

Sensitivity = TP / (TP+FN)

= X (True Positive)/Y+Z (False Positive + True Positive)

= Likelihood of the existing test being positive when an object is detected.

We calculated the specificity in order to quantify the likelihood of study-model to label an object as non-spirulina while the object is actually a non-spirulina. The formula to compute the specificity is given below.

Specificity = TN / (TN+FP)

IMPLEMENTATION AND RESULTS

Total number of images was 1000 of which 20 percent were identified for testing for each test run. This 20 percent was equally divided into spirulina and non-spirulina images (10 percent each) that would be used in testing the efficiency of the classification. All images were chosen randomly from the dataset which contains 500 spirulina and 500 non-spirulina images. The selection for the first run was randomly select from the top 100 spirulina images from database and the same method was applied for the non-spirulina images first 100 images from the database. Each test was run on the same hardware and it took approximately 0.006 s to detect a picture. Using the datasets and the method introduced in the previous section, experiments were performed changing the images of the input image as randomly select the second 100 image which is taken from the database for training and applying validation theory to check the performance of the convolutional neural networks. Hence, the training was based on using the same number of samples per species throughout the four classes. The experiments were therefore balanced. The results illustrated for each run using validation tables with calculated sensitivity and specificity in following sections.

First Run Results

First run result are in table 2 and specificity and sensitivity are in table 3 Specificity and sensitivity measures for the first run on the CNN theory with the data set were prepared The selection for the first run was randomly select from the top 100 spirulina images from database and the same method was applied for the non-spirulina images first 100 image from the database.

British Journal of Computer, Networking and Information Technology
ISSN: 2689-5315
Volume 2, Issue 1, 2019 (pp. 34-44)
www.abjournals.org

Table 2: Result of Spirulina Detection using Deep Learning (first run)

TP 100	FP 0	Total Test Positive 100
FN 9	TN 91	Total Test Negative 100
Total 109	Total 91	Total 200

Table 3: Sensitivity and Specificity of Spirulina Detection Results in the first run,

Sensitivity	Specificity
1.00	0.91

As presented in table 3 the Specificity is 0.91 only 9 negative false result, while all the spirulina objects detected correctly100% successfully.

Deep Learning Second Run

Second run validation results are in table 4 and specificity and sensitivity are in table 5 Specificity and sensitivity measures for the first run on the CNN theory with the data set were prepared The selection for the second run was randomly select from the second top 100 spirulina images from database and the same method was applied for the non-spirulina images second 100 image from the database.

Table 4: Result of Spirulina Detection using Deep Learning (second run)

TP 99	FP 1	Total Test Positive 100
FN 0	TN 100	Total Test Negative 100
Total 99	Total 101	Total 200

Table 5: Sensitivity and Specificity of Spirulina Detection Results in the second run,

Sensitivity	Specificity
0.99	1.00

As presented in table 5 the Sensitivity is 0.99 with only 1 False Positive result, while all the spirulina objects detected correctly 99% successfully.

Deep Learning Third Run

Third run validation results are in table 6 and specificity and sensitivity are in table 7 Specificity and sensitivity measures for the first run on the CNN theory with the data set were prepared The selection for the third run was randomly select from the third top 100 spirulina images from database and the same method was applied for the non-spirulina images third 100 image from the database.

Table 6: Result of Spirulina Detection using Deep Learning (third run)

TP 100	FP 0	Total Test Positive 100
FN 0	TN 100	Total Test Negative 100
Total 100	Total 100	Total 200

Table 7: Sensitivity and Specificity of Spirulina Detection Results in the third run,

Sensitivity	Specificity
1.00	1.00

As presented in table 7 the Sensitivity is 100% with only 0 False Positive result, while all the spirulina objects detected correctly 100% successfully.

Deep Learning Fourth Run

Fourth run validation results are in table 8 and specificity and sensitivity are in table 9 Specificity and sensitivity measures for the first run on the CNN theory with the data set were prepared The selection for the fourth run was randomly select from the fourth top 100 spirulina images from database and the same method was applied for the non-spirulina images fourth 100 image from the database.

Table 8: Result of Spirulina Detection using Deep Learning (fourth run)

TP 100	FP 0	Total Test Positive 100
FN 0	TN 100	Total Test Negative 100
Total 100	Total 100	Total 200

Table 9: Sensitivity and Specificity of Spirulina Detection Results in the fourth run,

Sensitivity	Specificity
1.00	1.00

As presented in table 9 the Sensitivity is 100% with only 0 False Positive result, while all the spirulina objects detected correctly 100% successfully.

DISCUSSION

This paper presents results of four runs, in each run input images were randomly selected, and ensured that the input images for each run be differed from one to another.

As presented in table 3 the Specificity is 0.91 only 9 negative false result, while all the spirulina objects detected correctly100% successfully. As presented in table 5 the Sensitivity is 0.99 with only 1 False Positive result, while all the spirulina objects detected correctly 99% successfully. As presented in table 7 the Sensitivity is 100% with only 0 False Positive result, while all the spirulina objects detected correctly 100% successfully. As presented in table 9 the Sensitivity is 100% with only 0 False Positive result, while all the spirulina objects detected correctly 100% successfully. The results from the first run of our system are shown in Table 2 and the sensitivity and specificity results are shown in Table 3, in which we can see 9 error detections as true negatives where any spirulina in the image could not be detected by the system. In the second run (Table 4), the system received only one false positive wherein there was no spirulina in the images and the system resulted in an output as spirulina.

The point of these result compared with related work on Diatom detection in water as Du Buf and Bayer [8] proposed Diatom Identification they used RGB level formatted images in extracting diatom from the image. They report an accuracy rate of 83.4% the reason it's with low rate than deep learning approach results are because of the environment of image if image is noise with environment objects the colour detection method false rate will be high. Another comparison with related work on Diatoms detection as in 2002 DuBuf H used Bagging Tree classifier over 781 samples with 96.9 accuracy [6] but this paper was only on circular diatoms where texture features not considered as our spirulina species are with zigzag shape. Spirulina detection based on colour with the ratio being 80% successful detection, as mention in the paper [17] The results of this paper to a successful detection rate where only 15 matches detected out of 39 images, for both methods SIFT and FAST which means successful detection rate are less than 50% of effectiveness. The results obtained in this study indicate Deep learning-based detection as successful method to detect spirulina in water, since deep learning calculations as talked about before are that they attempt to take in abnormal features highlights from information in a steady way. This takes out the need of speciality and hard-core feature descent.

CONCLUSION

In this paper, the entire work process of a spirulina images detection process with CNN has been covered from dataset building to framework identification. The application of CNN to the issue of Spirulina order has produced a number of intriguing results, including invariance in the picture due to various preparation strategies, The results in the third and fourth runs were the with no false output detection with 100% sensitivity and specificity, The point of these result compared with related work on Spirulina detection, as we see in results this is a condition for handcrafted feature approaches. Adequately, automated detection should be done accurately, although our purpose in the current study shows how successful deep learning approach is in Spirulina detection.

FUTURE RESEARCH

This study may be considered as the early beginning of automation that can distinguish and detect diatom pollution in green growth. Another enhancement yet to be studied would include calculating the number of diatoms to comprehend extent of pollution.

This can be accomplished by utilizing Region-based Convolutional Neural Networks (CNNs). Another enhancement yet to be added would include more original image rather than artificial images as input, also gathering these spirulina image samples from around the world not only from Turkey Rivers and leaks.

REFERENCES

[1] Wenke, H., Nicole, K., (2003). The European Parliament and the Council of the European Union. Directive 2000/60/EC. Establishing a Framework for Community Action in the Field of Water Policy, Official Journal of the European Community: Maastricht, the Netherlands.
[2] Smol, J., Stoermer, E., (2010). Applications for the Environmental and Earth Sciences; Cambridge University Press: Cambridge, UK.
[3] Blanco, S., Becares, E., (2010). Are biotic indices sensitive to river toxicants? A comparison of metrics based on diatoms and macro-invertebrates. Chemosphere, 79, 18–25.
[4] Bueno, G., Deniz, O., Pedraza, A., Salido, J., (2017). Automated Diatom Classification (Part A): Handcrafted feature approaches. Appl. Sci., in press.
[5] Dimitrovski, I., Kocev, D., Loskovska, S., Dzeroski, S., (2012). Hierarchical classification of diatom images using ensembles of predictive clustering trees. Ecol. Inform., 7, 19–29.
[6] DuBuf, H (2002). M.Automatic Diatom Identification; Seriesin Machine Perception and Artificial Intelligence; World Scientific Publishing Co. Munich, Germany.
[7] Pappas, J., Stoermer, E, (2003). Legendre shape descriptors and shape group determination of specimens in the Cymbella cistula species complex. Phycologia 42, 90–97.
[8] Hans, D., Micha M., (2002). Automatic Diatom Identification. New Jersey; London: World Scientific, ISBN 981-02-4886-5

[9] Lai, Q., Lee, K., Tang, A., Wong, K., So, H., Tsia, K., (2016). High-throughput time-stretch imaging flow cytometry for multi-class classification of phytoplankton. Opt. Express, 24, 28170–28184.
[10] Gonzalez, R., Woods, R., (2008). Digital Image Processing, Pearson Prentice Hall: Upper Saddle River, NJ, USA.
[11] Nesterov,Y., (2007). Gradient Methods for Minimizing Composite Objective Function; Technical Report; University College London:London, UK.
[12] Duchi,J., Hazan, E., Singer, Y., (2011). Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res, 12, 2121–2159.
[13] Jose, L., Gloria, B., Maria, T., Maria, O., (2019). Automated Identification and Classification of Diatoms from Water Resources, university del valle Colombia, springer nature swizerland AG: CIARP, LNCS 11401, app. 496-503.
[14] Pedraza, A., Bueno, G., Deniz, O., Cristóbal, G., Blanco, S., Borrego, M., (2017). Automated Diatom Classification Part B, A Deep Learning Approach, Applied Sciences (2017) 7(5) 460, DOI: 10.3390/app7050460.
[15] Krizhevsky, A., Sutskever, I., Hinton, G., (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM. 60 (6): 84–90. doi:10.1145/3065386. ISSN 0001-0782.
[16] Krizhevsky, A., Sutskever, I., Hinton, G., (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems; Neural Information Processing Systems Foundation, Inc.: La Jolla, CA, USA; pp. 1097–1105.
[17] Othman, S., Atila B., (2018). Automated System for Spirulina Detection in Water, International Eurasian Conference on Science, Engineering and Technology EurasianSciEnTech, 274, November 22-23, 2018.