ISSN 1870-4069

Microalgae Cell Counting and Identification via Artificial Intelligence Techniques: An Interdisciplinary Approach

V. A. González-Huitrón[2], A.E. Rodríguez-Mata[1], L. Miranda [2], D.M. Arias[3], Lisbel Bárzaga-Martell[4], H. Rodríguez-Rangel[2,*]

[1] Tecnológico Nacional de México/IT de Chihuahua, Mexico

[2] Tecnológico Nacional de México/IT de Culiacan, Mexico

[3] Universidad Nacional Autónoma de México, Instituto de Energías Renovables Mexico

[4] Universidad de Chile, Departamento de Ingeniería Eléctrica, Chile

[*]Corresponding author: hrodriguez@itculiacan.edu.mx

Abstract. Artificial intelligence is currently being applied to many industrial processes. However, there is little work on the relationship between artificial intelligence and biotechnology areas oriented towards the cultivation of seaweeds, especially in the production of *Chlorella Vulgaris*. This microalgae is a genus of unicellular green algae of the phylum Chlorophyta. It is spherical in shape, measuring 2 to 10 micrometers in diameter, and has no flagellum, which has many applications in the food and pharmaceutical industry. This is why it is essential to experiment with these cells and at the same time develop technologies or tools that allow us to carry out projects or experiments in an easier way in order to speed up the results. This work proposes the development of software capable of speeding up and optimizing one of the many tasks carried out with these cells through artificial intelligence and its pattern detection methods based on autoencoders to count the cells present in a certain number of samples. An identification process of more than 95 % more effective in prediction and identification is achieved.

Keywords: Autoencoder, graph theory, pattern detection, microalgae.

1 Introduction

Microalgae are autotrophic photosynthetic microorganisms with a high diversity group. More than 40,000 species of eukaryotic microalgae have been explored for their potential in food processing as they are a source of fatty acids and vitamins, in wastewater treatment, biofuels, and the production of high-end pharmacological products. *Chlorella Vulgaris* is a single-celled green microalgae present in some freshwater rivers of Mexico; with an important photosynthetic contribution in the waterways of Sinaloa [13]. This microalga is a source of proteins, omega-3 polyunsaturated fatty acids, polysaccharides, vitamins, and minerals, which provide several biological, industrial and pharmacological properties, many of which have been shown to be crucial in supplying human dietary deficiencies and providing an alternative for the treatment of some diseases. In fact, medical studies have shown that its use improves hyperlipidemia and hyperglycemia and protects against oxidative stress. Therefore, its cultivation and large-scale production, as well as research on its optimization, continues in an improvement process [17, 19, 18, 14].

One of the main issues related to microalgae optimizing production is an efficient estimation of the algae population. Microalgae counting is crucial to maintain control of biomass growth, prevent culture contamination by other species and other unfavorable conditions, and determine the level of lipids, carotenoids, proteins, and other substances of interest. Moreover, microalgae are also generally used as a bioindicator that provides information on water quality, so cell count is of great importance in the mass cultivation of Chorella. Therefore, specialized monitoring of microalgae is essential to up-scale *Chlorella Vulgaris* cultivation. The concentration of microalgae cells is the main and most used parameter in the cultivation of microalgae and ecological monitoring since it allows us to measure the optimal conditions of growth and growth and increase productivity while predicting any cell damage in real-time [3, 20].

Off-line techniques of microalgae counting include manual counting using a Neubauer counting chamber and expert judgment, adding high uncertainty to the measurement. The traditional detection process is cumbersome, laborious, and time-consuming, and it is impossible to analyze the concentration of algal cells quickly [7]. The automatic identification of microalgae samples is, therefore, a technical need to be solved since this would reduce decision-making times, in addition to the problem of reducing the impact on water quality. The identification and quantification of microalgae in water samples is made manually and at intervals in the Neabauer chambers, which is a slow and complex process.

The cell measurement and decision-making process can be optimized, leading to more reliable measurements using artificial intelligence techniques. Computer vision is a broad area of science that relies on techniques from various fields, such as pattern recognition, artificial intelligence, statistics, and machine learning. Automatic visual classification of objects is based on the geometric similarity of objects in several objects. It is the technical basis for implementing intelligent data and object recognition, including the classification and counting microalgae cells. The classifiers are designed manually under the knowledge and adherence to the experience of the expert trainer.

To improve cell counting, several remarkable works have studied the classification and cell quantification through microscope images, that is, from a visual sample of the amplification of a photograph obtained from a microscope, automatic quantification can be made using as reference the Neubahuer counting chamber, and with this infer the total concentration in growth phase in a river or bioreactor [15]. Based on general handcrafted features, some classifiers have given good results in the implementation of computer vision techniques using neural networks for the classification of marine and freshwater microalgae and diatom cells [16, 2, 12, 1].

For instance, Luo and Gao [19] used artificial neural networks (ANN) with Fourier spectrum to identify diatoms, achieving 94% accuracy. Mosleh et al. (Ref 5.) used Fourier spectrum with principal component analysis (PCA) and ANN to classify microalgae *Navicula*, *Scenedesmus* and cyanobacteria *Microcystis Oscillatoria* and *Chroococcus* in river water, reaching 93% accuracy. Other work includes Cerbin et al. [3], who employed ANN to identify *Scenedesmus obliquus*, achieving 90% accuracy. Furthermore, recently, Giraldo-Zuluaga et al. ([20]) achieved accuracies of 98.63% and 97.32% with support vector machine and ANN, respectively, alternative automatic algae counting of *Scenedesmus* sp. However, there are not a study of quantification and classification of cells of the microalgae *Chlorella Vulgaris* through these models to the best of our knowledge [8, 9].

Autoencoders are a sort of artificial intelligence that functions as a filter, learning to rebuild the picture to be classed to calculate a percentage of similarity with the item to be discovered [21]. This type of artificial vision technique are implemented to identify cells need to collect data and identify patterns with the help of scientific data obtained by colloquial scientific methods. With this, the expert knowledge, to provide training of autoencoders [6].

In this work, we prepared an experimental technical methodology where we propose a system of counting and classification of cells of the microalgae *Chlorella Vulgaris* present in the rivers of Sinaloa, Mexico. We have made the extraction and isolation of the microalga *Chlorella Vulgaris*, as well as its production of controlled and automated rectangular photobioreactors in the laboratory, to implement an artificial vision and deep learning techniques for quantification and classification of *Chlorella Vulgaris* cells giving with this the design of a biomass sensor of the microalga through the use of the images obtained in a microscope.

For this, It was necessary to develop software that can detect the area of interest and, in turn, can perform the detection of the cells to count them subsequently. For this purpose, it was necessary to use image processing techniques, such as the Hough transform to detect straight lines and detect circumferences, edge detection, and artificial intelligence, such as autoencoders. The development of the project consists of 3 main parts, detection of cells or possible cells in the image, development and training of the auto-encoder that processes the possible cells found to determine whether or not it is a cell and the development of the user interface.

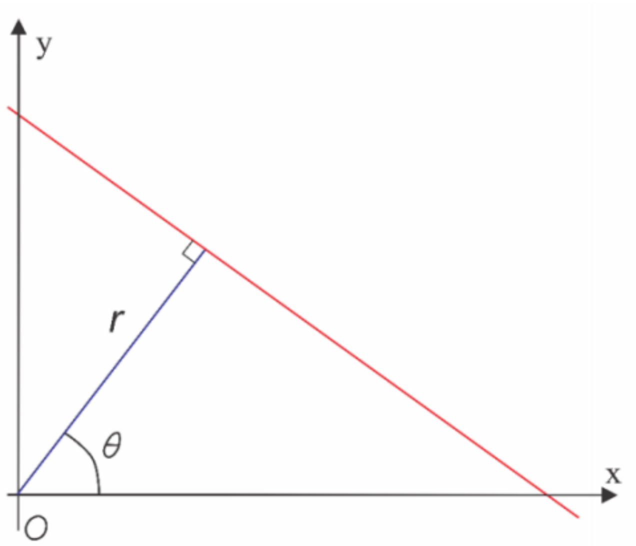

Fig. 1. Hough's T. Line Detection Graph.

2 Mathematical Background

2.1 Hough Transform

The Hough transform is an algorithm used in pattern recognition of an image, which allows finding shapes such as circles, lines, among others, within the image. The simplest version consists of finding lines, but depending on the image and the problem, it can be modified to see other types of shapes. The operation mode is statistical, and according to the points you have, you must find out the possible lines in which the issue can be achieved employing an operation applied to each line in a particular range.

Line detection. The Hough transform uses in its operation a parametric representation of geometric form, i.e., if it has a straight line, it would be represented as a line geometric parametric representation, i.e., if we have a straight line, this would be represented with the parameters rho and $theta$ (Eq. (1)), where rho is the distance between the line and the origin, and $theta$ is the angle of the vector from the origin to the nearest point, see Fig. 1. Employing the parameterization, the equation of the line could be written as follows:

$$\rho = x \cos\theta + y \sin\theta, \tag{1}$$

One of the characteristics that the Hough transform has is that if they are represented in a Cartesian plane, the line would be represented by a Cartesian plane, the line would be represented by the coordinates of the (ρ,θ), and the point

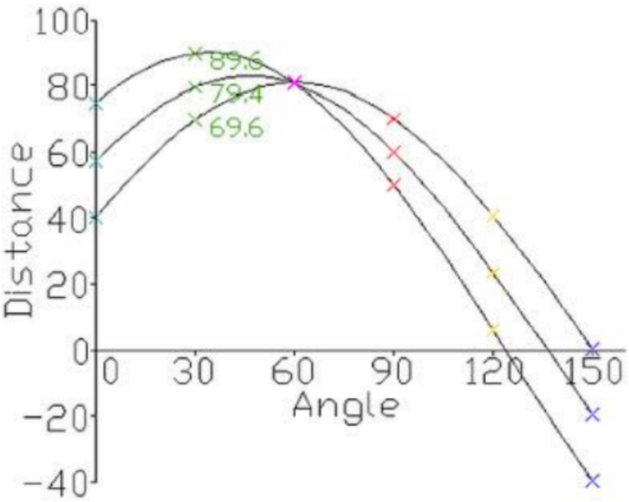

Fig. 2. Point intersection.

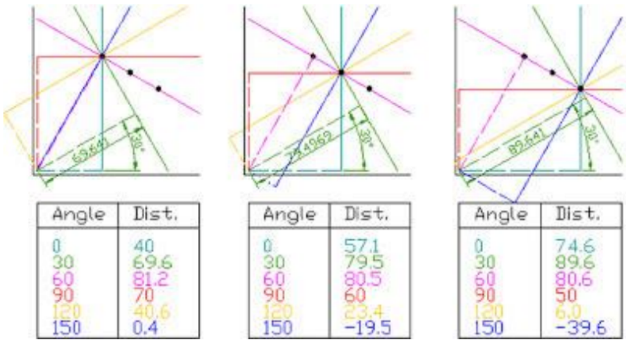

Fig. 3. Angles and distances of each line from origin.

would be represented as a sine function. Therefore, if you have two points, they would be symbolized by two sinusoidal symbols by means of two sinusoidal out of phase alpha degrees according to the coordinates of the points. Coordinates of the issues, but if the two points share the same straight line, the two sinusoids will end up crossing every 180 degrees since the sinusoidal function represents the set of infinite consecutive lines that pass through the point, as shown in Figs 2 and 3.

Circumference detection As previously mentioned, the Hough transform is not restricted only to line detection, although it is commonly used for that purpose. Circumferences can be detected by applying the equation (2):

$$(x-a)^2 + (y-b)^2 = r^2. \tag{2}$$

Three parameters are necessary to describe a circumference:

- Axes of the center of the circumference (a,b).
- Radio (r).

To find circles using the Hough transform, an accumulator with three dimensions (a, b, r) is needed. Once this procedure is completed, the highest values in the accumulator are searched for and the radius and center of the circle are obtained. If the radius were known beforehand, only a two-dimensional accumulator would be needed.

2.2 Edge Detection by Canny's Method

The Canny algorithm is an operator developed by John F. Canny in 1986 that uses a multi-stage algorithm to highly detect the edges of objects contained in images. Canny's goal is to find the optimal edge detection algorithm. An optimal edge detector fulfills the following:

- Good detection: the algorithm should mark as many real numbers on the edges of the image as possible.
- Good location - the marker edges should be as close as possible to the edge of the actual image.
- Minimal response - The edge of an image should only be marked once, and whenever possible, image noise should not create false edges.

Canny use the calculus of variations to satisfy these requirements - a technique that finds the function that optimizes a given process. The optimal operation in Canny's algorithm is described by the sum of four exponential terms, but can be approximated by the first derivative of a Gaussian to satisfy these requirements; Canny uses the calculus of variations - a technique that finds the function that optimizes a given functional. The optimal operation in Canny's algorithm is described by the sum of four exponential terms but can be approximated by the first derivative of a Gaussian.

Noise reduction Canny's edge detection algorithm uses a filter based on the first derivative of a Gaussian. Since it is susceptible to noise present in raw image data, the original image is transformed with a Gaussian filter. The result is an image that is slightly blurred concerning the original version. This new image is not affected by a single pixel of noise to any significant degree.

Example of a 5x5 Gaussian filter:

$$B = \frac{1}{159}\left(\begin{bmatrix} 2&4&8&2&3\\ 4&9&7&7&9\\ 5&5&6&5&3\\ 4&1&6&1&7\\ 2&3&8&4&7 \end{bmatrix}\right) * A. \tag{3}$$

Finding the gradient intensity of the image The edge of an image can point in different directions, so Canny's algorithm uses four filters to detect horizontal, vertical, and diagonal blurred image edges. The edge detection operator returns a value for the first derivative in the horizontal direction (Gy) and the vertical direction (Gx). From this, the edge gradient and direction can be determined:

$$G = \sqrt{G_x^2 + G_y^2}, \tag{4}$$

$$\theta = \arctan \frac{G_y}{G_x}. \tag{5}$$

3 Methodology

Microalgae were grown at the laboratory level under controlled conditions. Samples were taken from the lower basin of the Culiacán river since it is the area with the most downslope and the lowest flow velocity. Samples were taken in a $3L$ Vandorn bottle, and the samples were preserved at 4ºC until they were analyzed in the laboratory *Chlorella Vulgaris* was identified, and by means of a Pasteur pipette, it was manually isolated in the culture medium. This operation was repeated until the algal biomass that reproduced corresponded only to *Chlorella Vulgaris*. The culture was grown in Suoeka medium [11] in a cylindrical photobioreactor with a 3-liter Airlift configuration (see figure 4). biomass, pH, ORP, dissolved solids, conductivity, and culture parameters were measured with LabGenius multiparametric equipment. After 15 days, it was possible to take samples in the laboratory and obtain Neubauer chamber photographs through microscope photography.

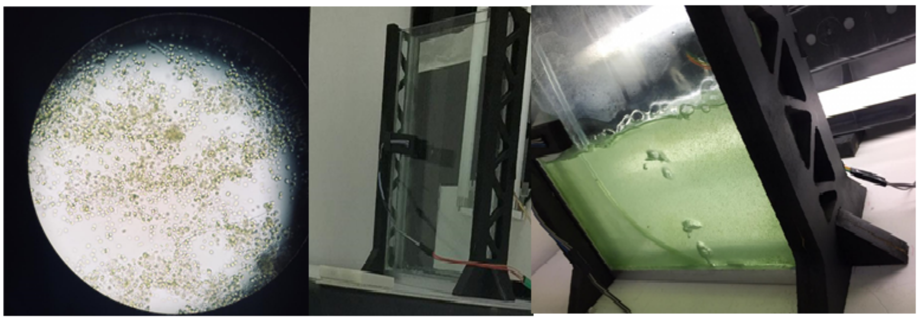

Fig. 4. Isolet *Chlorella Vulgaris.* and the photobioreactor used

Autoencoders are neural networks that aim to generate new data by first compressing the input into a latent variable space and then reconstructing the output based on the acquired information. This type of network consists of two parts: encoders and decoders.

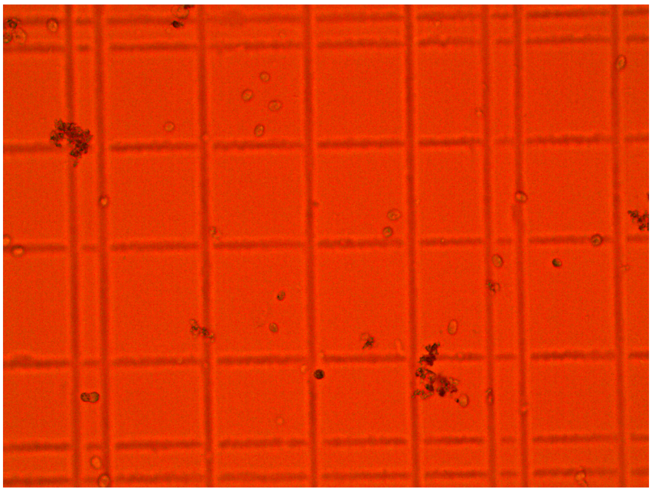

Fig. 5. Image to be processed through a microscope, for cell counting with a Neubauer camera.

4 Results

4.1 Image Segmentation and Detection of Possible Cells

Remarkable results were obtained in the detection of live *Chlorella Vulgaris* cells. It is essential to know the type of image to be processed in order to be able to make the detections correctly. In this case, we have the images as in Figure 5. To perform the analysis, we isolated and delimited the image to our area of interest, using the image processing methods of Canny and edge detection and the Hough Transform with the help of artificial vision techniques [5].

It is worth mentioning that for the Hough transform to work, it is necessary to pre-process the images to detect the edges of the objects so that the Hough transform function can perform its straight line detection. Figure 6 shows an example of how edge detection is done.

Figure 7 shows the example of line detection which helps us to determine the area of interest, as well as to do the cropping and segmentation of the image. To eliminate these parts, we generate four binary masks made from the detection of the lines at the boundary of the area of interest.

Once the multiplication of our binary mask by the original image is done, we will obtain something like in Figure 8 and then proceed to detect our viable cells employing the Hough Transform again, repeating all the previous process, but, now to notice circumferences. [10].

We are applying the Canny edge detection method as shown in Figure 9.To finish this process, we use the Hough transform for circumferences to detect viable cells within our image, cut out small images of these feasible cells, and then pass them to our neural network to determine if it is a cell or not. Figure 9

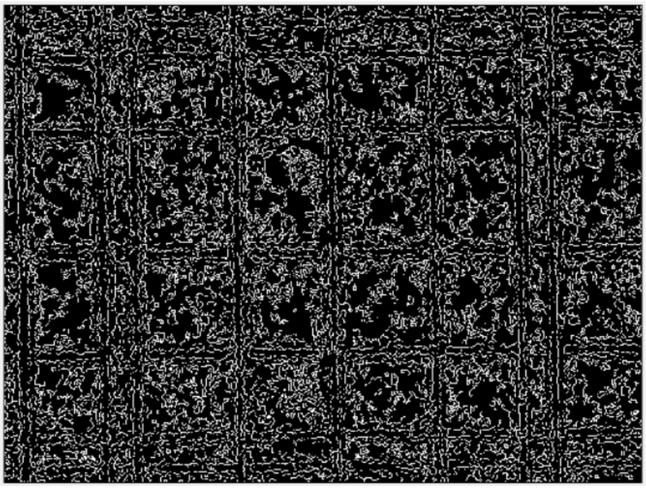

Fig. 6. Image with detected edges.

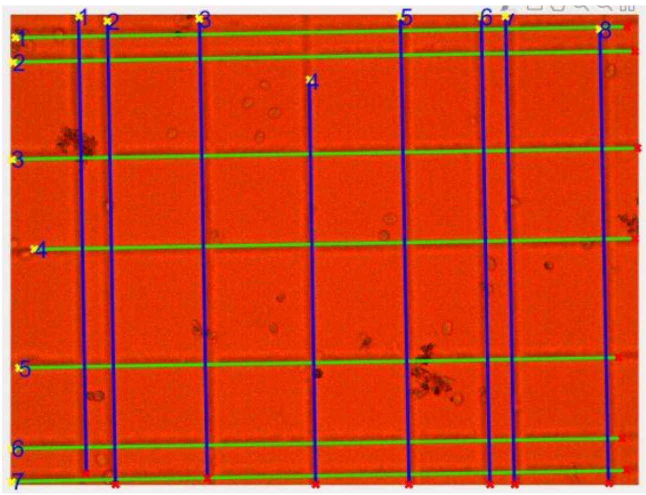

Fig. 7. Image with detected lines.

shows the result of the possible circumferences found in the image, and Figure 10 shows the cut out of an objective cell that will be sent to the neural network to be processed and analyzed.

4.2 Autoencoder development and training

The development of our auto-encoder was based on one made by the developers of Keras [4]; it was only modified and adjusted in terms of information retention

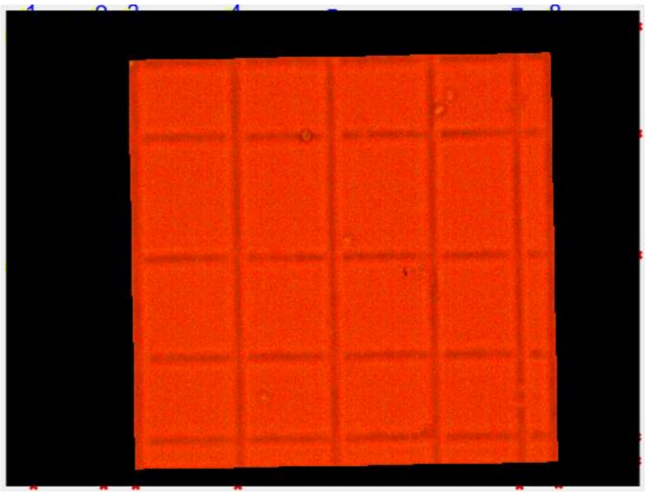

Fig. 8. Image with region of interest detected.

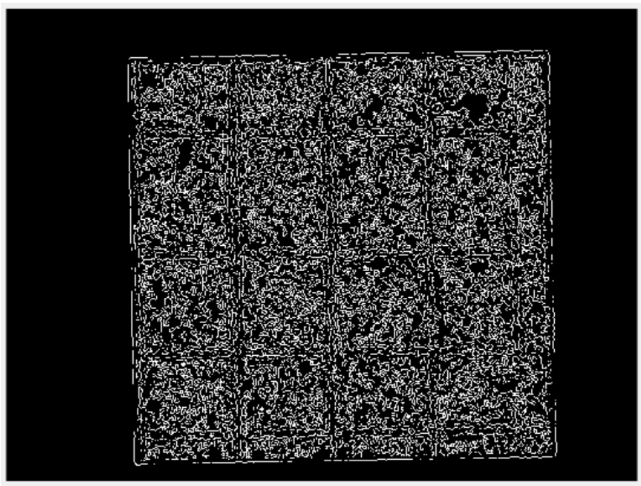

Fig. 9. Image of the area of interest with Canny.

parameters, number of training epochs, etc.

To train the neural network, it was necessary to rely on expert knowledge in cell detection to provide the auto-encoder with accurate information for training. Figure 9 is the type of image that was provided to the network with a cell for training.

It was necessary to have a database of authentic positive and accurate negative images so that the auto-encoder could learn to differentiate between the two. The true-positive images were used for training, while the true-negative

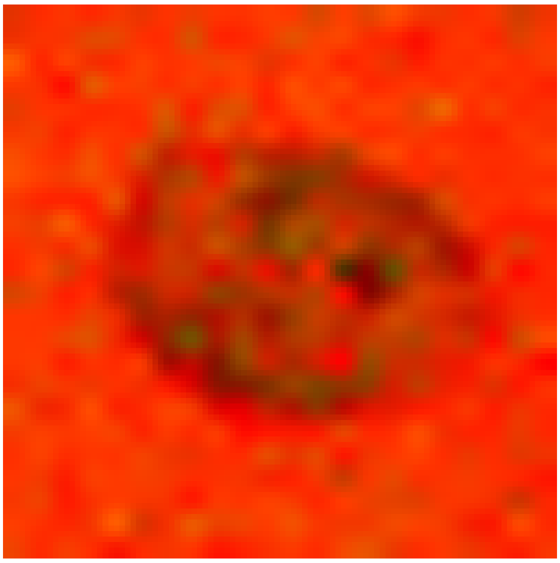

Fig. 10. Image of a real cell with digital zoom.

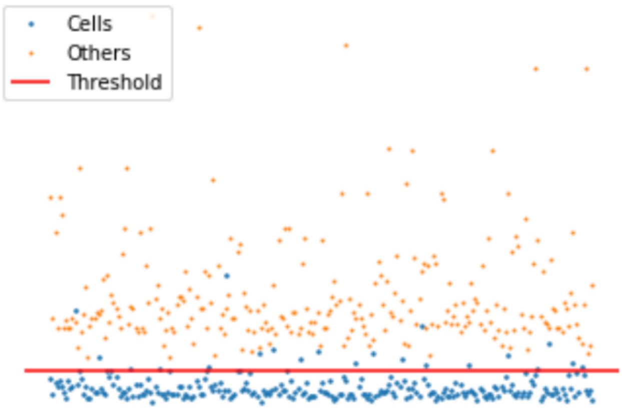

Fig. 11. Neural network prediction results.

images were used to test the network to see if it could detect. These tests yielded very favorable results.Considering that out of a database of 499 images (with 251 negative and 248 positive images), the neural network could recognise 100% of negative images as true negative and 95.76% of positive images as true positive say that the network is highly effective in detection.

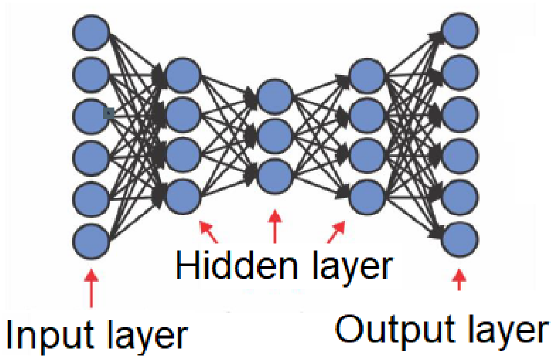

Fig. 12. Autoencoder Conceptual structure.

5 Conclusions

Based on the results obtained, it can be shown that the application of artificial intelligence for various tasks such as optimising time and processes, as well as automating tasks in which only repetitive and little changing work is needed. In this case the development of this platform makes automation possible, which in turn also improves the working conditions for researchers in the field of biology (microalgae research). It could be quantitatively demonstrated that the autocoder can estimate and detect *Chlorella Vulgaris* cells with high accuracy.

6 Futures Works

We are currently developing and designing the user interface, and the expert can give us the guidelines on how it should be managed and planned. This interface is being developed with the MATLAB AppDesigner application.

Acknowledgments. This project belongs to a project of the Tecnologico Nacional de Mexico *"Estimacion en la recuperacion de nutrientes para produccion de microalgas monitoreadas a traves de observadores de alta ganancia. Un enfoque multidiclinario"* with code 1229221-P, in the year 2021-2022.

References

1. Baek, S.-S., Pyo, J., Pachepsky, Y., Park, Y., Ligaray, M., Ahn, C.-Y., Kim, Y.-H., Chun, J. A., Cho, K. H.: Identification and enumeration of cyanobacteria species using a deep neural network. Ecological Indicators, vol. 115, pp. 106395 (2020)
2. Bueno, G., Deniz, O., Pedraza, A., Ruiz-Santaquiteria, J., Salido, J., Cristóbal, G., Borrego-Ramos, M., Blanco, S.: Automated diatom classification (part a): handcrafted feature approaches. Applied Sciences, vol. 7, no. 8, pp. 753 (2017)

3. Cerbin, S., Nowakowski, K., Dach, J., Pilarski, K., Boniecki, P., Przybyl, J., Lewicki, A.: Possibilities of neural image analysis implementation in monitoring of microalgae production as a substrate for biogas plant. In: Fourth International Conference on Digital Image Processing (ICDIP 2012). vol. 8334, pp. 83342A. International Society for Optics and Photonics (2012)
4. Chollet, F.: Building autoencoders in keras. The Keras Blog, vol. 14 (2016)
5. Fokkinga, M.: The hough transform. Journal of functional programming, vol. 21, no. 2, pp. 129 (2011)
6. Hatipoglu, N., Bilgin, G.: Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships. Medical & biological engineering & computing, vol. 55, no. 10, pp. 1829–1848 (2017)
7. Liu, J.-Y., Zeng, L.-H., Ren, Z.-H.: The application of spectroscopy technology in the monitoring of microalgae cells concentration. Applied Spectroscopy Reviews, pp. 1–22 (2020)
8. Luo, Q., Gao, Y., Luo, J., Chen, C., Liang, J., Yang, C.: Automatic identification of diatoms with circular shape using texture analysis, (2011)
9. Mosleh, M. A., Manssor, H., Malek, S., Milow, P., Salleh, A.: A preliminary study on automated freshwater algae recognition and classification system. In: BMC bioinformatics. vol. 13, pp. 1–13. BioMed Central (2012)
10. Olijve, L. L., Oude Vrielink, A. S., Voets, I. K.: A simple and quantitative method to evaluate ice recrystallization kinetics using the circle hough transform algorithm. Crystal Growth & Design, vol. 16, no. 8, pp. 4190–4195 (2016)
11. Ortiz-Moreno, M. L., Cortés-Castillo, C. E., Sánchez-Villarraga, J., Padilla, J., Otero-Paternina, A. M.: Evaluación del crecimiento de la microalga chlorella sorokiniana en diferentes medios de cultivo en condiciones autotróficas y mixotróficas. Orinoquia, vol. 16, no. 1, pp. 11–20 (2012)
12. Pedraza, A., Bueno, G., Deniz, O., Cristóbal, G., Blanco, S., Borrego-Ramos, M.: Automated diatom classification (part b): a deep learning approach. Applied Sciences, vol. 7, no. 5, pp. 460 (2017)
13. PÉREZ-BRAVO, S. G., MENDOZA-MARTÍNEZ, A. M., CASTAÑEDA-CHÁVEZ, M. d. R., AGUILERA-VÁZQUEZ, L.: Bioenergía a partir de microalgas en méxico,
14. Ru, I. T. K., Sung, Y. Y., Jusoh, M., Wahid, M. E. A., Nagappan, T.: Chlorella vulgaris: A perspective on its potential for combining high biomass with high value bioproducts. Applied Phycology, vol. 1, no. 1, pp. 2–11 (2020)
15. Ruiz-Santaquiteria, J., Bueno, G., Deniz, O., Vallez, N., Cristobal, G.: Semantic versus instance segmentation in microscopic algae detection. Engineering Applications of Artificial Intelligence, vol. 87, pp. 103271 (2020)
16. Schulze, K., Tillich, U. M., Dandekar, T., Frohme, M.: Planktovision-an automated analysis system for the identification of phytoplankton. BMC bioinformatics, vol. 14, no. 1, pp. 1–10 (2013)
17. Shah, M., Mahfuzur, R., Liang, Y., Cheng, J. J., Daroch, M.: Astaxanthin-producing green microalga haematococcus pluvialis: from single cell to high value commercial products. Frontiers in plant science, vol. 7, pp. 531 (2016)
18. Sharma, Y. C., Singh, B., Korstad, J.: A critical review on recent methods used for economically viable and eco-friendly development of microalgae as a potential feedstock for synthesis of biodiesel. Green chemistry, vol. 13, no. 11, pp. 2993–3006 (2011)

19. Sudhakar, M., Kumar, B. R., Mathimani, T., Arunkumar, K.: A review on bioenergy and bioactive compounds from microalgae and macroalgae-sustainable energy perspective. Journal of Cleaner Production, vol. 228, pp. 1320–1333 (2019)
20. Tarno, H., Qi, H., Endoh, R., Kobayashi, M., Goto, H., Futai, K.: Types of frass produced by the ambrosia beetle platypus quercivorus during gallery construction, and host suitability of five tree species for the beetle. Journal of Forest Research, vol. 16, no. 1, pp. 68–75 (2011)
21. Wajeed, M. A., Sreenivasulu, V.: Image based tumor cells identification using convolutional neural network and auto encoders. Traitement du Signal, vol. 36, no. 5, pp. 445–453 (2019)