Estimating Microbial Diversity via Morphological Based Microscopic Image Analysis: Methods and Metrics

Anaahat Dhindsa1,2*, Sanjay Bhatia3, Sunil Agrawal2 and B.S. Sohi4

1ECE Department, Chandigarh University, Gharuan, Punjab, India.
2ECE Department, UIET, Panjab University, Chandigarh, India.
3PG Department of Zoology, University of Jammu, Gujranwala, Jammu, Jammu and Kashmir, India.
4Chandigarh University, Gharuan, Punjab, India.

Abstract

To accelerate the monitoring and counting of biodiversity of various species, there is a need for automating the process of computing biodiversity. The calculations of the alpha and beta biodiversity indexes are fundamental for the analysis of ecological and biodiversity studies. Sukhna and Dhanas lakes, India are critical for the maintenance of the health of the residents and aquatic life thriving in them. Both lakes are prone to pollution. Due to these factors, there is a need for building digitized infrastructure for monitoring the health of these lakes. Hence in this research work, an automated algorithm has been devised for the computation of biodiversity of microorganisms. The work focuses on the surface water of both these lakes. The automation of biodiversity computation is done with image processing algorithms and is applied to the primary data collected. From this study, it is apparent that the counting of microorganisms using image processing algorithms is an easier and efficient way for biodiversity studies as compared to the manual process of estimating the population of microbes. The results show that the species richness of Dhanas Lake is more as compared to Sukhna Lake. The dissimilarity between the two lakes is five species as per the primary data collected. This shows that the biodiversity of Dhanas Lake is better than the Sukhna Lake but it is prone to harmful algal blooms. This may be attributed to the fact that Dhanas Lake may have multiple sources of pollution that need to be identified.

Keywords: Biodiversity Index, Biodiversity Monitoring, Image Processing, Computational Environment Sciences, Algae, Fungi

*Correspondence: anaahat.dhindsa85@gmail.com

(Received: November 05, 2020; accepted: December 24, 2020)
INTRODUCTION

For biodiversity, the present forecast seems bleak. In a recent United Nations study, an international alliance of scientists claimed that we are on track to lose over one million identified organisms in the next eight decades, i.e. one species out of eight. Meanwhile, the number of individual organisms have plummeted. Migratory birds are estimated to have lost about 70 percent of their population, and tigers have lost 97 percent of their population. The combined biomass of humans and domestic animals has reached 24 times that of all other wild mammals, over the course of only a few decades. Almost all organisms including birds and animals depend on the biodiversity of microbes. Microbes that are not visible to the naked eye, actively play an important role in maintaining and conserving life on this planet. Most importantly, many organisms are part of our medical and healing system. Due to this reason, the motivation of contemporary researchers working in this context is to develop modules, procedures, and methods that assist in gathering environmental data, used for enhanced understanding and analysis. Tasks such as mapping of major and minor species of an area, error checking of archives, and monitoring biodiversity changes are already in the process of automation. The information on the existence of species is derived with the help of two methods. The first method is collection and analysis of the primary data and the second method is studying the archives of repositories. With time, the repositories are being digitalized and there are multiple formats to hold information on the species data. Global Biodiversity Information Facility (GBIF) is an organization that provides access to biodiversity data using Internet-based interfaces. GBIF also helps people to become citizen scientists and conservationists. Besides, there are many other organizations such as Greenpeace, Tropicos, Birdlife international, worldwide fund for Nature (WWF), Global environment facility, and United Nations Environment that are collecting ecological and biodiversity data.

In this domain, limited research work in regards to the calculation of biodiversity indexes of microbes found in Sukhna and Dhanas lakes is reported. Work related to the quality of water, catchment characteristics, consequences of climatic changes of these water bodies have few citations in contemporary literature.

This may be attributed to the fact that the collection of biodiversity data is a challenging task wherein each type of species/taxa requires a different sampling process and counting protocol. Due to this fact, standard steps are followed for the collection of different sample species. For example, the collection of samples of mosses and lichens requires microscopes and the area under observation has to be divided into grids so that the diversity can be computed at the fine-grain level. On the other hand, ferns are sampled using transects that are parallel in nature, since it is easy for the sampler to count in such a manner. Likewise, macro-fungi are counted using transects/ paths that have quadrant shapes. This requires instruments such as magnifying glasses and microscopes. Arthropods are sampled using light, pitfall, entomological umbrella, mesh, etc. For identification and counting of arthropods, microscopes and magnification equipment are used. The biodiversity studies of the mammals and birds are done using photo trapping, night vision videography, and other indirect methods such as droppings counting, footprints etc. In the case of fish and water-borne microorganisms (micro-fungi, algae, bacteria, etc.), the use of mesh, nets, and microscopes are required. Many researchers have built toolkits and manuals that help people interested in recording and maintaining the count of all kind of species. But most of these toolkits and plans are restricted to either manual recording or tools such as Microsoft Excel. In recent years, computer assisted photo-identifier applications are increasingly being used. But hardly any of these have features to count biodiversity indexes automatically for a particular region. Few programming (python and java) libraries are however available for computing biodiversity indexes. Using such libraries, automated tools can be constructed. All these developments can help in building local biodiversity strategy and action plans (LBSAP). In India, it is the mandate of Section 41 of Biological Diversity Act 2002 for all the local bodies and agencies to implement appropriate steps to maintain a healthy level of biodiversity along with checking of pollution sources. Hence, building automated algorithms for computing biodiversity is important.
Studies related to the water bodies and other objects associated with natural habitats give classification and taxonomy of various regions of the water bodies\textsuperscript{18}. Water bodies (brooks, streams, creeks, gulf, river, lake, pond, sea, and ocean) are typically classified based on the use of water and other resources that can be obtained from the water body. It is the responsibility of the local governments and national agencies to monitor and maintain the quality of the various water bodies as per their respective category. For monitoring, many observatories and sampling stations have been made and a wide spectrum of technologies are used to maintain the record and conduct an analysis of the data collected\textsuperscript{19}. The current ecological studies show extensive literature on the use of cameras, cloud, satellite imagery, drones and deep penetration probes for data collection related to ecological indicators. The objective of these advancements is to contribute to the field of computational biodiversity, which is statistically correlated with the health of humankind.

Researchers, biologists and environmentalists have developed many computational methods for understanding the impact of ecological/biodiversity changes on earth as a habitat\textsuperscript{20}. The computation methods begin with calculations of the number of species (richness) in a particular region and offer different ways and methods to understand the status of biodiversity spots.

Diversity calculations are done using different kinds of metrics\textsuperscript{21,22}. The most prominent indexes currently used by researchers are discussed here. Shannon-diversity index is commonly used for computing species diversity in terms of rarity and commonness of the species\textsuperscript{23}. Mathematically, it requires three variables: the total number of species in the community (richness), the proportion of species and equability (evenness). This index is similar to the Simpson index\textsuperscript{24}, which also considers abundance and evenness as a parameter for computing the diversity index. The values of the diversity are computed on a scale of zero and one. The value zero signifies very high diversity and one signifies no diversity. Dominance Index\textsuperscript{25} is a metric that gives us the measure of diversity that attempts to find information on the species that are conspicuous in nature at a particular site. Due to differential in the ratio of predators, nutrition, pollutants, and light conditions at different levels of the habitat, some species may dominate at a particular region. Dominance Index computation helps to get information on this aspect. Berger-Parker index\textsuperscript{26} is another way to identify the dominant species of an area. The reciprocal of the value of this index can be inferred as a reduced level of dominance. Equability Index computes the evenness ratio of the species. An index by the name Hill Number\textsuperscript{27} is typically used for diversity characterization of the taxonomic survey data and phylogenetic records. This metric also considers the sampling ratio used in the computations. A higher number of samples lead to a higher degree of diversity discovery. Renyi Index\textsuperscript{28} is also used in computing the diversity of evolutionary data that details the relationship between the species. Margalef Richness Index\textsuperscript{29} is computed using (S-1) ln (n) formula, where S is the number of taxa, and ‘n’ is the number of individuals found in the site under observation. When the ratio of the square root of a sample with the number of taxa is considered, the index is called the Menhinick Index. Gini Diversity Index helps us to get deeper insights into the characterization of the species, for e.g. finding areas that have fish having a similar length. The Whittaker Index\textsuperscript{30} uses gamma diversity index in its computation for the characterization of the species. It is computed as the ratio of the gamma index and average sample diversity. Here, each sample size has to be standardized and diversity is measured as species richness. The mathematics of this index gives values from 0 to 2. Here, zero means minimum beta diversity, 1 means completely similar species are there and the value of 2 means that there is no overlap. Many indexes compute the similarity and complementarity of a site from which the data has been computed. Such indexes describe the characteristics of different sites and are comparative. Marczewski Steinhaus dissimilarity index\textsuperscript{31} is a complement of the Jaccard similarity index\textsuperscript{32}. Preston Diagrams are graphical presentation tools (a Kind of bar Graph) that helps to find species abundance distribution (SAD). Rarefaction and Lorenz are other types of graphical representations from which interpretation of the biodiversity of an area can be done.

This literature survey gives copious evidence to support the fact that there are
multiple ways in which the indexes are computed. Depending on the scope of the study, a particular index may be chosen. However, the current literature shows that the most commonly used index is the Shannon index. Automated microscopy image processing is the latest technique by which faster computations of the diversity indexes can be computed. Image processing has many advantages, especially when it comes to isolating microbes from images and counting microorganisms’ species richness. After collection of the primary data (water samples), the microscopic images or slides are made and automatic, counting algorithms are applied to obtain the results. Current research focuses on counting different kinds of algae and even microfungal colonies. In this context, the contemporary laboratory manuals show that there are primarily two ways by which the counting is done viz., direct and indirect methods. In indirect methods, counting is done with the help of electronic microscope based images/slides, electronic counter, and membrane filter, etc. Direct methods include the use of optical density properties of the solution containing microorganisms, weight of the culture, and by observing chemical/metabolic activity that changes due to factors such as light for example. In both methods, dye/strainers are extensively used so that the boundaries of the microorganisms become more readable. In the context of our research work, we are using indirect methods for counting microbes.

**MATERIAL AND METHODS**

This section gives comprehensive information on all the steps taken to achieve the previously mentioned purpose of the study, i.e. the automating the biodiversity counting tasks. This section covers the design of the methods that help to compute and validate the process of computing biodiversity automatically. Fig. 1 gives the functional flow of the proposed work.

**Collection of Primary Data**

Since there was no publicly available data/benchmark record on Sukhna and Dhanas lakes, therefore one of the major objectives of this research work was to collect primary data. The sample collection was done from the surface zone (littoral). On visual inspection, it was observed that Sukhna Lake has a very large catchment area (25 square kilometers) and it has lower levels of nutrition/pollutants as compared to Dhanas Lake. Dhanas lake is a second artificial man made water body in Chandigarh, built to catch adjacent rainwater. Both the lakes are important water bodies for the city dwellers and scientific community. The communities want to maintain the good health of these lakes and preserve aquatic life. This is important as the lakes are prone to the growth of weed and algal blooms, and are

![Fig. 1. Flow of proposed work](image-url)
also surrounded by human activities. For our study, around ten random [Fig. 2] sample collection points were selected for both the lakes. Each sample station was assigned a unique identifier number.

**Calculation Using Microscopic Images**

Around three thousand slides were made and these slides were captured using a microscope having magnification 4 x. For machine learning algorithms, there was a need to grade the images and eliminate the images that did not hold useful information. The grading and marking of images were then performed manually with the help of a professional, and reference manuals were consulted. The final collection of 600 (whole mounted) images was selected for the final catalog. In this section, automated processes for counting the number of species from the images for computing the alpha and beta biodiversity indexes are given. The mathematical formula for computing these indexes is given in Table 1.

By definition, the Alpha biodiversity index is the average of species found in a particular lake. In the context of our research work, each substation is considered as a sampling area for calculating the alpha index. For comparative analysis between the two lakes, beta indexes need to be computed. Using these metrics, an automated algorithm was devised that works iteratively on the images and counts the entire index metric.

Following are the steps involved in automatic biodiversity computations from the slides:

1. **Initialize counting arrays for each microorganism to be cataloged:** These arrays store values of species count and in the end; they help in the aggregation of these counts.
2. **Read each microorganism slide/Image:** For each group/species, a separate folder is created and a file-based dataset is created. Each slide is then processed using a loop for reading its pixels values and positions.
3. **Segment each image using Segmentation Algorithm:** The input matrix is processed using a generalized segmentation algorithm.

**Table 1. Formulae for computing Biodiversity Indexes**

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Biodiversity Index</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Simpson Index</td>
<td>$\sum \frac{n(n-1)}{N(N-1)}$</td>
<td>$n$ = the total number of organisms of a particular species</td>
</tr>
<tr>
<td>2</td>
<td>Dominance Index</td>
<td>$1 - \sum \frac{n(n-1)}{N(N-1)}$</td>
<td>$N$ = the total number of organisms of all species</td>
</tr>
<tr>
<td>3</td>
<td>Reciprocal Simpson Index</td>
<td>$\frac{1}{\sum \frac{n(n-1)}{N(N-1)}}$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Shannon Index</td>
<td>$\sum n_i \cdot \ln \frac{n_i}{N}$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Common Species Index</td>
<td>$A \cap B$</td>
<td>Number of species found</td>
</tr>
<tr>
<td>6</td>
<td>Dissimilarity Index</td>
<td>$A' \cap B'$</td>
<td>Number of species found both in A and B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$A \in U$, $B \in U$, $A \cap B$</td>
<td>Number of species found in Sukhna lake and B found in Dhanas lake</td>
</tr>
</tbody>
</table>

Number of species found in A but not in B or Number of species found in B but not in A.
that uses pixel-clustering (Otsu) for gathering similar data points and extracts boundaries of microbes using kirsch convolution filter. The output (segmented images) of these algorithms are stored in a public repository; Mendeley\textsuperscript{35}.

4. Run Biodiversity Index Function on each slide: For each image (segmented), this step invokes the function for counting each microbe. This is done with the help of region props method. The region props method first gives labels/indexes to each region it

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Microorganism</th>
<th>Category</th>
<th>No. of Instances</th>
<th>Sukhna (S)</th>
<th>Dhanas (D)</th>
<th>Lake</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yeast</td>
<td>Fungi</td>
<td>3456</td>
<td>1335</td>
<td>2121</td>
<td>S, D</td>
</tr>
<tr>
<td>2</td>
<td>Paramecium</td>
<td>Protozoa</td>
<td>4668</td>
<td>2236</td>
<td>2432</td>
<td>S, D</td>
</tr>
<tr>
<td>3</td>
<td>Volvox</td>
<td>Algae</td>
<td>7136</td>
<td>3246</td>
<td>3890</td>
<td>S, D</td>
</tr>
<tr>
<td>4</td>
<td>Spirogyra</td>
<td>Algae</td>
<td>4365</td>
<td>2247</td>
<td>2118</td>
<td>S, D</td>
</tr>
<tr>
<td>5</td>
<td>Amoeba</td>
<td>Bacteria</td>
<td>6327</td>
<td>3026</td>
<td>3301</td>
<td>S, D</td>
</tr>
<tr>
<td>6</td>
<td>Pseudomonas</td>
<td>Fungi</td>
<td>595</td>
<td>346</td>
<td>249</td>
<td>S, D</td>
</tr>
<tr>
<td>7</td>
<td>Mycelium</td>
<td>Fungi</td>
<td>1246</td>
<td>347</td>
<td>899</td>
<td>S, D</td>
</tr>
<tr>
<td>8</td>
<td>Nostoc</td>
<td>Algae</td>
<td>2856</td>
<td>1346</td>
<td>1510</td>
<td>S, D</td>
</tr>
<tr>
<td>9</td>
<td>Chlorella</td>
<td>Algae</td>
<td>213</td>
<td>164</td>
<td>49</td>
<td>S, D</td>
</tr>
<tr>
<td>10</td>
<td>Coccus</td>
<td>Bacteria</td>
<td>101</td>
<td>0</td>
<td>101</td>
<td>D</td>
</tr>
<tr>
<td>11</td>
<td>Cosmarium</td>
<td>Algae</td>
<td>262</td>
<td>191</td>
<td>71</td>
<td>S, D</td>
</tr>
<tr>
<td>12</td>
<td>Ulothrix</td>
<td>Algae</td>
<td>4154</td>
<td>2352</td>
<td>1802</td>
<td>S, D</td>
</tr>
<tr>
<td>13</td>
<td>Oedogonium</td>
<td>Algae</td>
<td>3814</td>
<td>2025</td>
<td>1789</td>
<td>S, D</td>
</tr>
<tr>
<td>14</td>
<td>Cocconeis cistula</td>
<td>Algae</td>
<td>186</td>
<td>147</td>
<td>39</td>
<td>S, D</td>
</tr>
<tr>
<td>15</td>
<td>Aspergillus</td>
<td>Fungi</td>
<td>3290</td>
<td>0</td>
<td>3290</td>
<td>D</td>
</tr>
<tr>
<td>16</td>
<td>Pencicillus</td>
<td>Fungi</td>
<td>1611</td>
<td>0</td>
<td>1611</td>
<td>D</td>
</tr>
<tr>
<td>17</td>
<td>Spirulina</td>
<td>Algae</td>
<td>887</td>
<td>352</td>
<td>535</td>
<td>S, D</td>
</tr>
<tr>
<td>18</td>
<td>Oscillatoria</td>
<td>Algae</td>
<td>3001</td>
<td>2323</td>
<td>678</td>
<td>S, D</td>
</tr>
<tr>
<td>19</td>
<td>Anabaena Azolla</td>
<td>Algae</td>
<td>7114</td>
<td>0</td>
<td>7114</td>
<td>D</td>
</tr>
<tr>
<td>20</td>
<td>Euglena</td>
<td>Algae</td>
<td>1390</td>
<td>778</td>
<td>612</td>
<td>S, D</td>
</tr>
<tr>
<td>21</td>
<td>Brown Algae</td>
<td>Algae</td>
<td>5687</td>
<td>2346</td>
<td>3341</td>
<td>S, D</td>
</tr>
<tr>
<td>22</td>
<td>Rhizopus</td>
<td>Fungi</td>
<td>1264</td>
<td>569</td>
<td>695</td>
<td>S, D</td>
</tr>
<tr>
<td>23</td>
<td>Trichoderma</td>
<td>Fungi</td>
<td>1248</td>
<td>345</td>
<td>903</td>
<td>S, D</td>
</tr>
<tr>
<td>24</td>
<td>Microcystis</td>
<td>Algae</td>
<td>412</td>
<td>0</td>
<td>412</td>
<td>D</td>
</tr>
<tr>
<td>25</td>
<td>Vaucheria</td>
<td>Algae</td>
<td>2670</td>
<td>1346</td>
<td>1324</td>
<td>S, D</td>
</tr>
<tr>
<td>26</td>
<td>Pithophora</td>
<td>Algae</td>
<td>1619</td>
<td>807</td>
<td>812</td>
<td>S, D</td>
</tr>
<tr>
<td>27</td>
<td>Chlamydomonas</td>
<td>Algae</td>
<td>1020</td>
<td>591</td>
<td>429</td>
<td>S, D</td>
</tr>
<tr>
<td>28</td>
<td>Cladophora</td>
<td>Algae</td>
<td>4645</td>
<td>2334</td>
<td>2311</td>
<td>S, D</td>
</tr>
<tr>
<td>29</td>
<td>Others</td>
<td>Unknown</td>
<td>3358</td>
<td>1472</td>
<td>1886</td>
<td>S, D</td>
</tr>
</tbody>
</table>

Fig. 2. Sample Collection Points

Table 2. Input data for Computing Indexes
has found. By definition, a minimum of four connected pixels (that have similar properties) is considered as a valid region. The region may or may not represent a microbe or microbe colony. Hence for each microbe, a threshold of pixel numbers was computed using a visual inspection tool for ensuring that each region represents a colony of respective microbes.

5. Aggregate the Biodiversity Indexes: The last step includes the aggregation of all counts from each slide/image for computing biodiversity indexes (Simpson Index, Dominance Index, Shannon Index).

Fig. 3. Population Count

Fig. 4. Biodiversity Indexes
Reciprocal Simpson, Shannon, Common Species Index, Dissimilarity Index) as per the definition.

Table 2 and Fig. 3 gives information on population statistics obtained with the use of image processing segmentation algorithm. Fig. 4 and Fig. 5 gives information on the values of alpha and beta indexes respectively, obtained after running the automated biodiversity counting algorithm.

RESULTS AND DISCUSSION

In terms of population, it can thus be observed that Dhanas Lake has a higher number of species and a correspondingly higher count of microorganisms. The average population is also higher. This can be attributed to the fact that spatially data is denser in the case of the Dhanas dataset. Also, the nutrition level at most of the data collection points is higher in the case of Dhanas Lake. Visually, it can be observed that this lake has higher levels of turbidity and muddiness. Statistically, it can further be observed that a fair amount (3358) of microorganisms were not identifiable. Volvox, Brown algae, and Anabaena have a maximum population in Dhanas Lake. The Cocconeis cistula has the lowest count in both the lakes. In the case of Sukhna Lake, it can be observed that Ulothrix, Volvox and Amoeba have maximum populations. The Sukhna Lake does not have any identifiable instance of Coccus, Aspergillus, Pencicllus, Microcystis and Anabaena Azolla. With help of these statistics, the biodiversity of both lakes was computed automatically.

The balance of the ecology in an area is directly related to the count and type of microorganisms found. The Simpson and Shannon Indexes help to track this aspect. It can be observed that the total number of microorganisms found and identified in the case of Dhanas Lake are higher than Sukhna Lake. This implies that the Simpson index of Dhanas is lower as compared to Sukhna Lake as shown in Fig. 4. Consequently, it can be inferred that Dhanas Lake has a higher richness and higher level of diversity as compared to Sukhna Lake.

Longitudinal studies show that the main reason for disturbance in ecological balance is due to the reduction of species population or dominance of particular species. The dominance of a particular species has two possible consequences. The first one is that the count of many species may
get reduced due to the predatory nature of one of the species or it may be just dominant due to the abundance of particular types of food and is not impacting the population cycle of the other species. For further insight, the dominance index was computed and observed statistically. Ideally, the dominance index value should be one and it can be seen from the above table that Sukhna has a dominance index value of 0.93 and Dhanas has 0.941. From this, it can be safely concluded that Dhanas has marginally a higher level of biodiversity. This argument is reinforced when the value of the Reciprocal Simpson index is considered, as it can be seen that the Dhanas Lake’s Reciprocal Simpson index value is higher as compared to Sukhna Lake. Clearly, the relative abundance of different species in the case of Dhanas is also higher than Sukhna. The most abundant species are Volvox and Amoeba in the case of Sukhna. The metrics used here for understanding the biodiversity of the Sukhna and Dhanas Lakes are based on the random sampling technique. The species are randomly sampled from each dataset of both lakes so that bias is reduced and the effect of the sample size is accounted for. The metrics consider abundance as well as evenness for giving an insight on diversity for the two sites.

When a comparison is done between the two sites for computing Beta Diversity [Fig. 5], it can be observed that the similarity index (Common Species by Intersection) is 24 and the Dissimilarity index is five.

**CONCLUSION**

To the best of our knowledge, it can be said with confidence that limited work has been done in the context of computing Sukhna and Dhanas biodiversity indexes using microscopic imaging. The research work gives details on the composition of species in each sample collection station. This biodiversity study is an attempt to automate the computations of biodiversity indexes by using laboratory developed microscopic slides/images. To achieve the previously mentioned goals of this study, an automatic algorithm to segment and count the population of each microorganism was created. The experimentation shows that our novel hybrid segmentation algorithm is quite useful for calculations of biodiversity indexes accurately.

The goal of the research is to establish the facts from the biodiversity study of Sukhna and Dhanas Lake. The first fact that can be established from this study is that biodiversity is not the same in both the lakes and there are significant differences in terms of characteristics of the water bodies. This may be attributed to the different compositions of pollution and nutrition at surface level and the exposure to sunlight. The statistical analysis of the biodiversity values gives us a hint on the state of ecology and the health of microorganisms in the lake. Clearly, from the metric values, we can say that there is a need to improve the biodiversity of both lakes and these water bodies are candidates for ecological sensitivity studies. It is apparent the Dhanas Lake has better biodiversity than Sukhna but it seems to be prone to harmful algal blooms. This may be because higher levels of nutrition/contamination that promotes the growth of harmful algae are entering the Lake. The sources of pollution need to be plugged so that ecological balance can be made.

This research work is an attempt to demonstrate the role of automated computer algorithms for supporting biodiversity studies. For further directions, it is suggested that other regions of water bodies must be included and leveraging of machine and deep learning algorithms should be done to automate the identification process of microorganisms.

**ACKNOWLEDGMENTS**

Authors are highly thankful to Dr. Sanjay Bhatia, P.G. Department of Zoology, University of Jammu, India for helping us in doing the tedious work of manual identification of microorganisms and for providing constant guidance and support.

**CONFLICT OF INTEREST**

All authors declare that there is no conflict of interest.

**AUTHORS’ CONTRIBUTION**

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

**FUNDING**

None.
DATA AVAILABILITY

The dataset and code used during the current study are available in the Mendeley repository, at Doi: 10.17632/f9m85ptmwc.2, 2020.

ETHICS STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

2. Il’in VP. The conception, requirements and structure of the integrated computational environment. In: Communications in Computer and Information Science. 2019. doi: 10.1007/978-3-030-05807-4_56
33. Guisan C, Heine J, Garcia-Rosello E, Gonzalez-Dacosta J, Vilas LG, Perez-Schofield B. DER: An algorithm for comparing species diversity between...


