Vol. 20, No. 1 March 2023 | DOI https://doi.org/10.33480/techno.v20i1.4028

DIGITAL IMAGE IDENTIFICATION OF PLANKTON USING REGIONPROPS AND BAGGING DECISION TREE ALGORITHM

Hikmatulloh[1*]; Yan Rianto[2]; Anton[3]

[1,3]Informatika, [2]Ilmu Komputer
Universitas Nusa Mandiri
http://www.nusamandiri.ac.id/
hikmatulloh.hkl@nusamandiri.ac.id[1*]; yan.rianto@lipi.go.id[2]; anton@nusamandiri.ac.id[3]

Abstract - *The role of plankton is very important for the life of the surrounding organisms, so research on plankton is needed because of its relationship with the survival of other living things. The obstacles that are often encountered in plankton research, especially in the identification of plankton, are the inefficiency in the aspect of time and these organisms have a very small average size. In this case, a better alternative is needed in identifying this type of plankton by digitally processing the image on the plankton image or commonly referred to as "Digital Image Processing". This study aims to process 144 digital plankton images which were acquired to be 75% as training data and 25% as testing data, and these images were obtained from research at the Kanopi Indonesia foundation. In the process, the shape of this image is analyzed using the Regionprops function to obtain the distinguishing features of each type of plankton. After the feature extracted image, the data processing is then carried out by classifying each type of plankton. To produce a better data classification, this study uses the Bagging Decision Tree algorithm in processing the data and produces an accuracy of 92.59%. The Bagging Decision Tree algorithm is quite good and easy to implement into a plankton type identification program, as evidenced by testing the test image data resulting in 33 correctly identified images from a total of 36 tests.*

Keyword : *Plankton, Regionprops, Bagging, Decision Tree.*

Intisari - Peranan plankton sangat penting bagi kehidupan organisme disekitarnya, sehingga penelitian prihal plankton sangatlah dibutuhkan karena kaitannya dengan kelangsungan kehidupan mahluk hidup lainnya. Kendala yang sering didapatkan dalam hal penelitian plankton khususnya dalam hal pengidentifikasian plankton yaitu tidak efisiennya dalam aspek waktu dan organisme ini memiliki ukuran rata-rata yang sangat kecil. Dalam hal ini diperlukan alternatif yang lebih baik dalam pengidentifikasian jenis plankton ini dengan cara pemrosesan gambar pada citra plankton secara digital atau biasa disebut dengan istilah *"Digital Image Processing"*. Penelitian ini bertujuan untuk melakukan pengolahan citra digital plankton sebanyak 144 citra yang yang dibagi menjadi 75% sebagai data pelatihan dan 25% sebagai data pengujian, dan citra tersebut didapatkan dari riset pada yayasan Kanopi Indonesia. Dalam prosesnya citra ini dianalisa bentuk menggunakan fungsi *Regionprops* sehingga didapatkan fitur pembeda dari masing-masing jenis plankton. Setelah citra terekstraksi fitur nya selanjutnya dilakukan pengolahan data dengan mengklasifikasikan setiap jenis plankton tersebut. Untuk menghasilkan sebuah klasifikasi data yang lebih baik, dalam penelitian ini menggunakan algoritma *Bagging Decision Tree* dalam pengolahan data nya dan menghasilkan akurasi sebesar 92.59%. Algoritma *Bagging Decision Tree* ini cukup baik dan mudah untuk di implemntasikan kedalam sebuah program identifikasi jenis plankton, terbukti dengan pengujian pada data citra pengujian menghasilkan 33 citra teridentifikasi dengan benar dari total pengujian sebanyak 36 citra.

Keyword : Plankton, *Regionprops*, *Bagging*, *Decision Tree*.

INTRODUCTION

Plankton are micro-organisms that are found living in the waters of rivers, lakes, reservoirs, brackish and sea (Saputra et al. 2017). The movement of plankton is strongly influenced by water currents because these organisms live by floating or floating in the water. In the waters, the existence of plankton has a very important role and is needed by other living things. Because considering this organism has an important role as a primary producer in the waters, knowing the level of fertility of the waters and as an indicator in determining a fishing area (Eramma et al. 2022).

Basically, plankton is divided into two types, namely phytoplankton and zooplankton. Phytoplankton is a group of plankton that has chlorophyll in its body (Pommerening, Zhang, and Zhang 2021). So that this type of plankton can produce its own food through the process of photosynthesis. While zooplankton are herbivores, i.e. eat producers (phytoplankton) (Noviani, Rahman, and Sofarini 2021). In this case the existence of zooplankton becomes a controller for phytoplankton as producers in a waters, resulting in an abundance relationship between these two types of organisms, where phytoplankton will develop rapidly when the zooplankton population decreases.

The development of research on plankton is very important because of its relation to the survival of other living things. However, the problem that occurs in conducting plankton research, especially to identify plankton species, is inefficiency from a time perspective (Bandara et al. 2021). In research (Aytan, Esensoy, and Senturk 2022) mentioned that the identification of these organisms is often hampered because these organisms have a very small size (mostly <20 mm for their total length), so that in every study of this plankton, accuracy is needed in analyzing its shape.

Based on these problems, a better alternative is needed in identifying plankton types, so that the authors in this study have the goal of segmenting digital plankton images so that the shape of the plankton can be known so that shape analysis can be carried out, looking for image shape characteristic values of plankton so that it can be analyzed using the regionprops function, classifying data on the value of plankton shape analysis using the Bagging Decision Tree method, and implementing a program to identify the type of plankton.

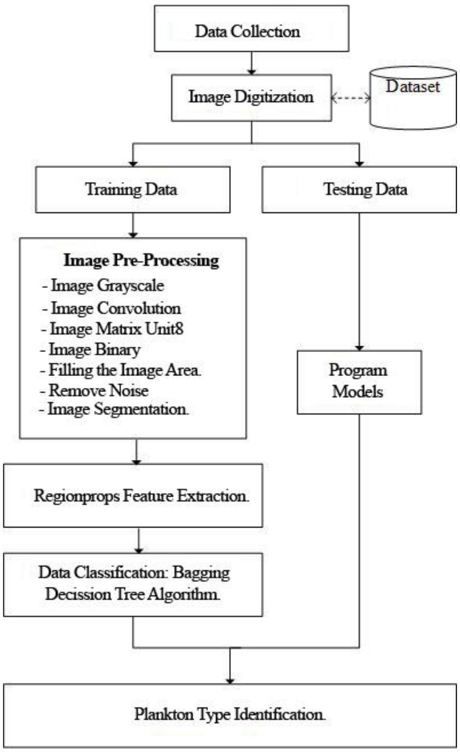
This research will propose a way that will certainly be more effective and efficient in identifying these types of plankton, namely by processing images on plankton images digitally or commonly referred to as "Digital Image Processing". Digital Image Processing is a field of science that studies how an image is formed, processed, and analyzed so as to produce information that can be understood by humans (Rafiq, Yusuf, and Pujono 2019). The processing results of the plankton image are in the form of feature extraction output results which will later be processed to produce a data classification model.

Image processing is carried out to extract these plankton features based on the different shapes each type of plankton has, so the function of regionprops will be very suitable in analyzing each type of plankton. Region properties (regionprops) is a function that belongs to matlab to measure a set of properties of each region that has been labeled in the label matrix L (Sariningsih et al. 2017). This property will be used between *Area, ConvexArea, MajorAxisLength, MinorAxisLength, EquivDiameter, Perimeter, Eccentricity, Solidity,* dan *Extent.*

After all the image features have been extracted, then the image feature data will be classified, and in classifying this study using plankton imagery generated from research conducted at the Kanopi Indonesia Yogyakarta foundation, and the targeted variables have been obtained from each group of images. This research will be implemented into a program that can be used especially by ecologists in analyzing plankton types, so the Bagging Decision Tree algorithm will be used as a method for classifying the plankton images. Bagging Decision Tree is a method in Machine Learning that can be visualized and easy to understand and implement (Pamungkas, Adiwijaya, and Utama 2020). This method is expected to produce a more accurate analysis of plankton species.

MATERIALS AND METHODS

The model in this study in general can be described as follows.



Source: research results (2020).

Figure 1. Research Model

The source of the data was obtained from survey activities to collect data on the diversity of species and abundance of plankton in the Opak Sunyai Watershed area which includes the upper reaches of the Kuning River in Cangkringan and Wedomartani, Sleman Regency, the middle part in Wukirsari Village, Imogiri District and the confluence between the Opak River and the Oyo River District. Pundong, and the downstream part, namely the estuary in the Baros Mangrove area, Tirtohargo Village, Kretek District, Bantul Regency. The easiest technique for taking plankton samples from waters can generally be done by filtering a large mass of water with a fine net (plankton net). Plankton sampling was carried out by filtering water near the surface using 10 liters of plankton net water, then the sample was treated with 4% formalin preservative. Plankton sampling was carried out in 3 repetitions at each sampling point at each research station.

The data analysis process is assisted by several tools including electron microscopes, counting cells, dropper pipettes, and cameras. The water sample is taken using a dropper pipette of 1 ml and input into the counting cell and then observed under an electron microscope. After getting a fairly clear picture, it is then analyzed based on the name of the plankton species that is matched to the plankton identification book and recorded on the observation table sheet. After the process of observation and analysis of plankton species then enters the image digitization process, by means of plankton images from the electron microscope documented using a camera and each image is cropped image with the same dimensions 459x459 pixel format (*.jpg), so that an image is obtained. ready to be used as a dataset.

In this process, 144 images of plankton were obtained from 6 (six) types of plankton, including 24 images of the appendicularia, 24 images brachiolaria, 24 images chaetognath, 24 images copepod, 24 images detritus, and 24 images of the diatoms. The 144 plankton image data will be acquired, 75% will be used as training image data and 25% will be used as test image data, of which 108 images will be used as training data and 36 images will be used as test data. And the following shows the types of several plankton images resulting from the image digitization process, the images displayed are 6 (six) images used for training data.

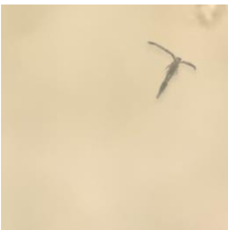
Source: research results (2020).

Figure 2. *Appendicularia* type plankton

Source: research results (2020).

Figure 3. *Brachiolaria* type plankton

Source: research results (2020).
Figure 4. *Chaetognath* type plankton



Source: research results (2020).
Figure 5. *Copepod* type plankton

Source: research results (2020).
Figure 6. *Detritus* type plankton

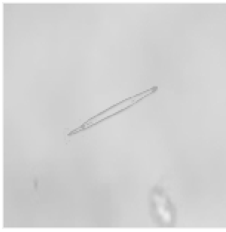
Source: research results (2020).
Figure 7. *Diatom* type plankton

Figure 2 shows the type of Appendicularia, Appendicularia or known as tadpole larvae because they look like frog tadpoles (Matsuo et al. 2020). Figure 3 is the type of Brachiolaria, Brachiolaria lives as plankton, depends on seawater and moves with coordinated movements of external cilia. The larvae are bilaterally symmetrical and have distinct left and right sides (Middleton et al. 2021). Figure 4 is the type of Chaetognath, Chaetognaths are an exclusive marine phylum known as arrow worms. The bear's body is transparent, with a maximum body length of: 105 mm; tail: 10-18% of body length (Vinther and Parry 2019). Figure 5 is the type of Copepod, Copepod is the largest and most diverse group of plankton, they are everywhere and are very abundant in waters, ecologically copepods have the most important relationship with phytoplankton and are higher in food webs in waters (Pitois et al. 2021). Figure 6, namely the type of Detritus, Detritus consists mostly of silt particles and fine sand around it and is an organic material that attaches and forms aggregates (Siano et al. 2021). And figure 7, namely the type of Diatom, Diatom is a unicellular phytoplankton, but is often found in the form of colonies. Diatom in terms means 2 parts that can not be divided again which reflects the structure of the diatom cell (Behrenfeld et al. 2021).

RESULTS AND DISCUSSION

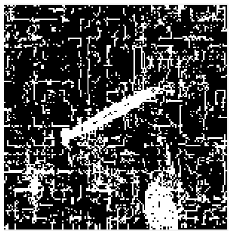
In this study, image segmentation was carried out. Image segmentation is an image processing that aims to separate the object region from the background area so that objects are easily analyzed in order to recognize objects that involve a lot of visual perception (Huang et al. 2020). The method used in segmenting images using several operators in MatLab includes: (taken as an example of an image of the chaetognath type) (Vinther and Parry 2019).

a. Changing the original RGB type image *(Red, Green, Blue)* which has 3 (three) dimensions into a 2 (two) dimensional image that has a gray color scale.



Source: research results (2020).
Figure 8. The Result of Changing the RGB Image to a Gray Image

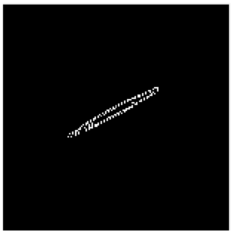
b. Then apply a convolutional filter with input horizontally and vertically on the image that has been converted to gray scale.



Source: research results (2020).
Figure 9. The Result of Changing the Gray Image to a Convolutional Image

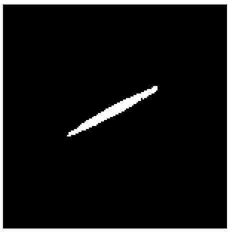
Vol. 20, No. 1 March 2023 | DOI https://doi.org/10.33480/techno.v20i1.4028

c. After converting the object array into a uint8 matrix, the image is then converted into a binary image with a scale of 0 and 1 according to the specified threshold so that the object's edges are obtained.



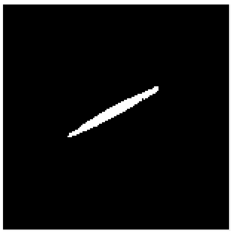
Source: research results (2020).
Figure 10. The results of changing the convolutional image into a binary image

d. After the edge of the image object is selected, fill in the image hole region.



Source: research results (2020).
Figure 11. Results of Binary Image Changes by Filling in Hole Regions

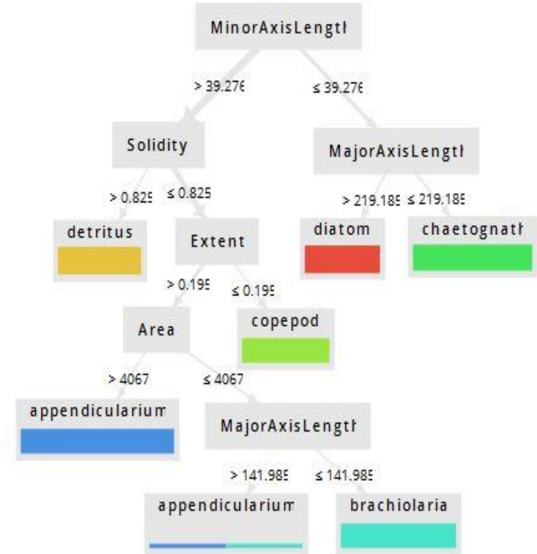
e. The last step to produce better image segmentation is to remove small pixels with a certain scale in the form of noise contained in the image.



Source: research results (2020).
Figure 12. Result of Image Change After Removing Noise

Measurement of some of the properties of each object region that has been segmented in this case using a Matlab function, namely regionprops. Some of the properties used in analyzing the features of the image in this case include: *Area, ConvexArea, MajorAxisLength, MinorAxisLength, EquivDiameter, Perimeter, Eccentricity, Solidity,* dan *Extent* (MathWorks 2020).

The results of the classification model on plankton data were carried out using the Bagging Decision Tree algorithm to produce several decision tree models, the resulting decision trees will be used to determine an implementation of an identification program. The following are the results of the decision tree from the research conducted.



Source: research results (2020).
Figure 13. Decision Tree

After the research rules from the classification model are obtained, then the model will be used as an algorithm for implementation into an identification program. The resulting program algorithm that races on a decision tree is as follows.

```
MinorAxisLength > 39.276
|   Solidity > 0.825: detritus
|   Solidity ≤ 0.825
|   |   Extent > 0.195
|   |   |   Area > 4067:
appendicularium
|   |   |   Area ≤ 4067:
brachiolaria
|   |   Extent ≤ 0.195: copepod
MinorAxisLength ≤ 39.276
|   MajorAxisLength > 219.185:
diatom
|      MajorAxisLength  ≤  219.185:
chaetognath
```

The quality of a data classification model in plankton imagery can be calculated through the resulting confusion matrix table. The following is the confusion matrix table generated in the Bagging Decision Tree algorithm.

Table 1. *Confusion Matrix* Algoritma *Bagging Decision Tree*

	true app	True bra	True cha	True cop	True det	True dia	Class precision
pred. app	17	2	0	0	0	0	89.74%
pred. bra	1	14	0	2	0	0	82.35%
pred. cha	0	0	18	0	0	1	94.74%
pred. cop	0	2	0	16	0	0	88.89%
pred. det	0	0	0	0	18	0	100.00%
pred. dia	0	0	0	0	0	17	100.00%
Class recall	94.44%	77.78%	100.00%	88.89%	100.00%	94.44%	

From table 1 above, it is known:

$$\textit{Accuracy} = \frac{\sum TP}{Total\ Data} \times 100\%$$

$$= \frac{TP(\text{app})+TP(\text{bra})+\text{TP(cha)}+\text{TP(cop)}+\text{TP(det)}+\text{TP(dia)}}{Total\ Data} \times 100\%$$

$$= \frac{17+14+18+16+18+17}{108} \times 100\%$$

$$= \frac{100}{108} \times 100\%$$

$$= 92{,}59\%$$

$$All\,Precision = \sum \frac{TP}{(TP+FP)}$$

$$= \frac{TP(app)}{(TP(app)+FP(app))} + \frac{TP(bra)}{(TP(bra)+FP(bra))} + \frac{TP(cha)}{(TP(cha)+FP(cha))} + \frac{TP(cop)}{(TP(cop)+FP(cop))} + \frac{TP(det)}{(TP(det)+FP(det))} + \frac{TP(dia)}{(TP(dia)+FP(dia))}$$

$$= \frac{17}{(17+1)} + \frac{14}{(14+4)} + \frac{18}{(18+0)} + \frac{16}{(16+2)} + \frac{18}{(18+0)} + \frac{17}{(17+1)}$$

$$= 0.94 + 0.78 + +0.89 + 1 + 0.94$$

$$= 4{,}55$$

$$\textit{Precision} = \frac{all\ precision}{total\ class} \times 100\%$$

$$= \frac{4.55}{6} \times 100\%$$

$$= 75.83\%$$

$$All\ Recall = \sum \frac{TP}{(TP+FN)}$$

$$= \frac{TP(app)}{(TP(app)+FN(app))} + \frac{TP(bra)}{(TP(bra)+FN(bra))} + \frac{TP(cha)}{(TP(cha)+FN(cha))} + \frac{TP(cop)}{(TP(cop)+FN(cop))} + \frac{TP(det)}{(TP(det)+FN(det))} + \frac{TP(dia)}{(TP(dia)+FN(dia))}$$

$$= \frac{17}{(17+2)} + \frac{14}{(14+3)} + \frac{18}{(18+1)} + \frac{16}{(16+2)} + \frac{18}{(18+0)} + \frac{17}{(17+0)}$$

$$= 0.89 + 0.82 + 0.95 + 0.89 + 1 + 1$$

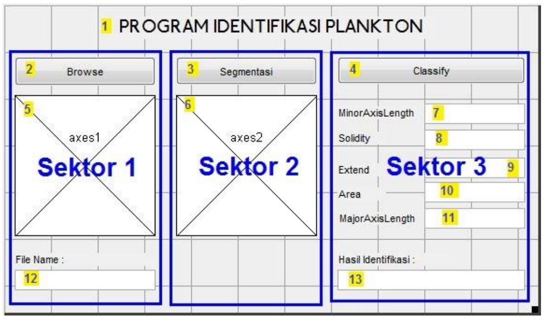
$$= 5.55$$

$$\textit{Recall} = \frac{all\ precision}{total\ class} \times 100\%$$

$$= \frac{5.55}{6} \times 100\%$$

$$= 92.50\%$$

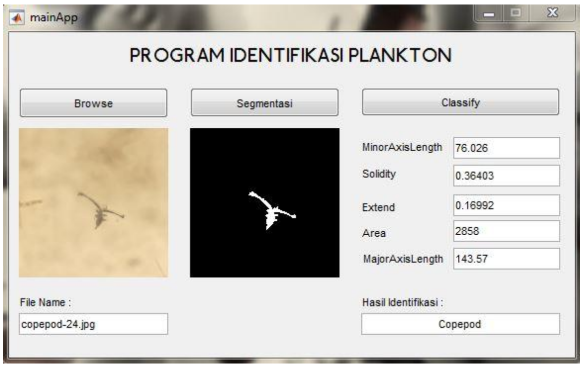
The design of the application built is divided into process sectors, the first sector is the process of selecting and displaying digital plankton images to be processed, the second sector is the process of segmenting digital plankton images and displaying the results of image segmentation, and the third sector is the feature extraction process regionprops to image segmentation results and displays the results of the digital image identification of the plankton. The following shows the program design along with an explanation of each component.



Source: research results (2020).
Figure 14. Program Design

Based on Figure 14, the resulting application program is executed sequentially according to the sector sequence that has been described, regardless of the procedure for using the application and displaying the results of the application made are as follows.

1. The user presses the "Browse" button to retrieve image data which will be displayed on the axes component below and the file name of the captured image will also be displayed in the edit text column below.
2. After the image data image appears, the user presses the "Segmentation" button to carry out the segmentation process and the results of the segmentation will be displayed on the axes component below.
3. The final step is that the user presses the "Classify" button to find out the results of the identification after the feature value appears in the text edit for each column of the regionprops property, and the results of the identification will appear in the text edit component below.



Source: research results (2020).
Figure 15. Display of Program Implementation Results

From the program implementation results, a testing process was carried out on the available test data, from the 36 test data it produced test results with 33 correctly identified data and 3 incorrectly identified data.

CONCLUSION

The conclusions obtained from the results of research on the classification of plankton species with image processing techniques on digital plankton images are 144 images, 108 images as training data and 36 image data as test data.

The plankton image segmentation process was successfully carried out with several functional stages from Matlab including, changing the RGB image to greyscale, then converting it into a convolutional image, then making the uint8 matrix image, then converting it into a binary image, after the image becomes binary then fill in the image hole area and the last step to perfect the segmentation is done by removing the noise.

The regionprops function can be used to analyze the shape of each type of plankton, the regionprops properties include Area, ConvexArea, MajorAxisLength, MinorAxisLength, EquivDiameter, Perimeter, Eccentricity, Solidity, and Extent. But in this case only a few properties that influence classifying this plankton include *MinorAxisLength, MajorAxisLength, Solidity, Extent,* dan *Area*.

The results of this plankton type classification using the Bagging Decision Tree algorithm yield an accuracy of 92.59%, a precision of 75.83%, and a recall of 92.50%. The implementation of the plankton identification program is carried out by referring to the resulting decision tree model algorithm. The program was tested with test data of 36 images where 33 images were identified correctly and 3 images were identified incorrectly.

REFERENCE

Aytan, Ulgen, F. Başak Esensoy, and Yasemen Senturk. 2022. “Microplastic Ingestion and Egestion by Copepods in the Black Sea.” *Science of The Total Environment*. https://www.sciencedirect.com/science/article/abs/pii/S0048969721059994.

Bandara, Kanchana et al. 2021. “Two Hundred Years of Zooplankton Verticalmigration Research.” *Biological Reviews*. https://onlinelibrary.wiley.com/doi/epdf/10.1111/brv.12715.

Behrenfeld, Michael J. et al. 2021. “Thoughts on the Evolution and Ecological Niche of Diatoms.” *Ecological Monographs* 91(3): 1–25.

Eramma, N. et al. 2022. “Zooplankton Productivity Evaluation of Lentic and Lotic Ecosystem.” *Limnology*. https://www.intechopen.com/chapters/83428.

Huang, Li et al. 2020. “Jointly Network Image Processing: Multi-Task Image Semantic Segmentation of Indoor Scene Based on CNN.” *IET Image Processing* 14(15): 3689–97.

MathWorks. 2020. “MathWork.” https://www.mathworks.com/.

Matsuo, Masaki, Takeshi A. Onuma, Tatsuya Omotezako, and Hiroki Nishida. 2020. “Protein Phosphatase 2A Is Essential to Maintain Meiotic Arrest, and to Prevent Ca2+ Burst at Spawning and Eventual Parthenogenesis in the Larvacean

Oikopleura Dioica." *Developmental Biology* 460(2): 155–63. https://doi.org/10.1016/j.ydbio.2019.12.005.

Middleton, J, C Bursch, J Maurer, and R Masui. 2021. "Marine Phytoplankton of South Central Alaska." : 1–21.

Noviani, Eka, Abdur Rahman, and Dini Sofarini. 2021. "Struktur Komunitas Plankton Dan Perubahan Kebiasaan Makan Ikan Gabus (*Channa Striata*, Bloch.) Dan Ikan Sepat Siam (*Trichogaster Pectoralis* , Regan.) Di Rawa Danau Bangkau, Kalimantan Selatan." *Aquatic* 4(2): 1–117. http://103.81.100.242/index.php/aquatic/article/view/1410%0Ahttp://103.81.100.242/index.php/aquatic/article/download/1410/755.

Pamungkas, Yoga Widi, Adiwijaya Adiwijaya, and Dody Qori Utama. 2020. "Klasifikasi Gambar Gigitan Ular Menggunakan Regionprops Dan Algoritma Decision Tree." *Jurnal Sistem Komputer dan Informatika (JSON)* 1(2): 69.

Pitois, Sophie G. et al. 2021. "A First Approach to Build and Test the Copepod Mean Size and Total Abundance (CMSTA) Ecological Indicator Using in-Situ Size Measurements from the Plankton Imager (PI)." *Ecological Indicators* 123: 107307. https://doi.org/10.1016/j.ecolind.2020.107307.

Pommerening, Arne, Gongqiao Zhang, and Xiaohong Zhang. 2021. "Unravelling the Mechanisms of Spatial Correlation between Species and Size Diversity in Forest Ecosystems." *Ecological Indicators* 121(September 2020): 106995. https://doi.org/10.1016/j.ecolind.2020.106995.

Rafiq, Arif Ainur, Muhamad Yusuf, and Pujono Pujono. 2019. "Digital Image Processing Menggunakan Perangkat Lunak Ni Vision Dan Ip Kamera Dengan Rover Bogie Robot." *Jurnal Ecotipe (Electronic, Control, Telecommunication, Information, and Power Engineering)* 6(1): 1–11.

Saputra, Adang, Kukuh Nirmala, Tri Heru Prihadi, and Joni Haryadi. 2017. "STUDI KESUBURAN PERAIRAN DANAU BERATAN UNTUK BUDIDAYA PERIKANAN : Prosiding Forum Inovasi Teknologi Akuakultur 2009." *Prosiding Forum Inovasi Teknologi Akuakultu*: 553–59.

Sariningsih, Neina Oktavia, Fiky Yosef S S T, Ig Prasetya Dwi W, and M Eng. 2017. "PERANCANGAN DAN IMPLEMENTASI TRACKING OBYEK MANUSIA UNTUK AUTONOMOUS CAR DESIGN AND IMPLEMENTATION OF HUMAN OBJECT TRACKING FOR." 4(1): 71–78.

Siano, Raffaele et al. 2021. "Sediment Archives Reveal Irreversible Shifts in Plankton Communities after World War II and Agricultural Pollution." *Current Biology* 31(12): 2682-2689.e7.

Vinther, Jakob, and Luke A. Parry. 2019. "Bilateral Jaw Elements in Amiskwia Sagittiformis Bridge the Morphological Gap between Gnathiferans and Chaetognaths." *Current Biology* 29(5): 881-888.e1. https://doi.org/10.1016/j.cub.2019.01.052.