Identification of monogeneans parasite using gray level co-occurrence matrix and artificial neural network

Hikmatulloh\textsuperscript{1*}, Dwiza Riana\textsuperscript{2}, Jamal Maulana Hudin\textsuperscript{3}, Susilawati\textsuperscript{4}, Dede Wintana\textsuperscript{5}, Sri Hadianti\textsuperscript{6}

\textsuperscript{1}Teknologi Komputer, \textsuperscript{2}Ilmu Komputer, \textsuperscript{3}Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, ID
\textsuperscript{4}Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, ID
\textsuperscript{5}Sistem Informasi, Sekolah Tinggi Manajemen Ilmu Komputer Nusa Mandiri, Jakarta, ID
\textsuperscript{6}Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, ID

E-mail: hikmatul0808@bsi.ac.id

Abstract. Flatworms that are parasitic and attack the skin and gills of fish are called monogenean parasites. In this study 3 monogenic parasites were used, namely grandis, liewi, and jokorensis. GLCM (Grey Level Co-occurrence Matrix) used in digital image processing to obtain feature extraction on images. ANN (Artificial Neural Network) used to classification mongenean parasites based on features extracted. This research resulted overall accuracy of 86.67%. Studying monogenean parasites is needed because it is related to the health of fish that we need in our daily lives, so that this research can ease and assist the work of ecologists in classifying monogenean parasites more quickly and efficiently.

1. Introduction
The existence of fish decreases, if there is no anticipation and action it will have an impact on human life, because humans are one of the higher food chain positions. This was revealed based on data collected during 1930-2010. In that time, fish populations in the world’s seas declined by 4.1% \cite{1}. One of the causes of this decline in fish population is due to the poor health of fish due to parasitic attacks or other things. One parasite that can attack fish is the monogenean parasite. Monogeneans are the flatworms (Platyhelminthes) that are primarily found on gills and skin of fishes. Monogenean parasites have attachment appendages at their haptoral regions that help them to move about the body surface and feed on skin and gill debris. Haptoral attachment organs consist of sclerotized hard parts such as hooks, anchors and marginal hooks. Monogenean species are differentiated based on their haptoral bars, anchors, marginal hooks, reproductive parts’ (male and female copulatory organs) morphological characters and soft anatomical parts \cite{2}.

Monogeneans are important pathogenic parasites of captured captive fishes and exhibit a single-host life cycle, which makes them amenable to in vivo culture. Continuous cultures of oviparous monogenean parasites provide a valuable resource of eggs, oncomiracidia (larvae) and adult parasites for use in varied ecological and applied scientific research \cite{3}. Monogen identification is not easy because it requires expertise in the field of ecology. Therefore, this
research will introduce a method for classifying monogenous parasites by applying digital image processing and machine learning methods.

Feature extraction is obtained using GLCM (Grey Level Co-occurrence Matrix). A GLCM is generated by counting the co-occurring grey level intensity values found in an image given a linear spatial relationship between two pixels [4]. Gray-Level Co-occurrence Matrix is a statistical tool which allows considering the spatial relations between the neighboring pixels, especially useful for texture analysis purposes. Assuming the maximum dynamic range of the image luminance equal to e.g. $K = 256$ levels, the $K \times K$ elements GLCM matrix can be obtained. The value of each element located at the i-th row and j-th column informs how many times pixels having the luminance level i are located in the specified neighborhood respectively to the pixels of the luminance level j [5]. Data mining is called knowledge discovery because it is a field that seeks to find information that has meaning and is useful from large amounts of data. Data mining is an interactive or automated process to find patterns in the data and predict future behavior based on the data patterns [7]. Data mining that will be used in this study is the ANN (Artificial Neural Network) Algorithm.

2. Methods
An explanation of the research methodology in this research is presented in the flowchart in Figure 1.

![System Flowchart](image)

**Figure 1.** System Flowchart.

The image was obtained from a previous study entitled “Data from: Monogenean anchor morphometry: systematic value, phylogenetic signal, and evolution” by Khang, Tsung Fei; Soo, Oi Yoon Michelle; Tan, Wooi Boon; dan Lim, Lee Hong Susan 2017 [6]. Three monogeneans were taken for this study including grandis, liewi, and johorensis. Dataset used in this study
amounted to 60 monogeneans parasitic images, 45 images were used as training data and 15 images were used as testing data.

To get the image segmentation and extraction features from the image, image Processing is performed using the R2015a matlab installed on Intel Atom Inside with 2GB RAM Windows 7 (32-bit). After the image features are obtained, the image identification is performed using the artificial neural network algorithm.

3. Result
3.1. Image Processing

![Image Processing Stages](image.png)

**Figure 2.** Process in Image Processing.

The following is an explanation in the image processing stages for segmenting monogeneans parasitic images:

- From the original image which is an RGB image converted to a greyscale image.
- Image convolution stage is carried out which aims to produce edge detection, operator roberts used to convolution this image vertically and horizontally.
- Image thresholding is performed by changing the data type to uint8 so that the image can be binary.
- Operation Morphologi to fill the edges of the image and eliminate components that are less than n pixels.

The following describes one of the monogenean images after feature extraction in table 1 below.
Table 1. Average Value Fitur Extraction.

| No | Image Type | Contrast | Correlation | Energy | Homogenety |
|----|-------------|----------|-------------|--------|------------|
| 1  | kedahensis  | 0.42873  | 0.74787     | 0.1819 | 0.82779    |
| 2  | kederai     | 0.59074  | 0.79869     | 0.13079| 0.80084    |
| 3  | navjotsodhii| 0.4493   | 0.81128     | 0.14881| 0.8256     |

The GLCM (gray co-occurrence matrix) method is used as a way to classify images using second order statistical measurements, GLCM is an image processing methodology used to describe the spatial relationship between gray values in two-dimensional images [8]. The results of GLCM feature extraction for the three types of monogenean are not fully varied so these results indicate that this feature is not entirely good in identifying the image of this monogenic parasite.

3.2. Image Identification

ANN (Artificial Neural Network) is used as a method in classifying the three types of parasitic monogeneans based on the extracted feature values. A total of 45 image data were imported as training inputs and training targets, while 15 other image data were imported as test inputs and testing targets. In this training Backpropagation Feed-forward is chosen to determine the type of network using the TRAINCGB training function. Next is the design of artificial neural network algorithms in training to classify monogenean parasites using 4 number of layers.

This network in its training is carried out several times to produce a good classification to get a network trained with the best performance. The following results of classification in the best training are presented in the figure below. Artificial Neural Network in its best training produces MSE (mean square error) $= 0.30413$ at epoch 191 (Figure 4).

![Artificial Neural Network Diagram](image)

Figure 3. Artificial Neural Network Diagram

![Best Validation Performance](image)

Figure 4. Best Validation Performance

Table 2. Confusion matrix of data testing.

|           | true randis | true liewi | true johorensis |
|-----------|-------------|------------|-----------------|
| pred. grandis | 5           | 0          | 0               |
| pred. liewi  | 0           | 4          | 1               |
| pred. johorensis | 0          | 1          | 4               |
The artificial neural network approach can classify with an overall accuracy of 86.67%, grandis can be identified as many 5 correctly from 5 testing data, liewi can be identified as many 4 species correctly from 7 testing data, 1 was identified as a monogenean johorensis. And monogenean johorensis parasites can be identified as many as 4 species correctly from 5 testing data and 1 identified as grandis and 2 was identified as liewi. The following confusion matrix table is presented in table 2 below.

\[
\text{True Positif (TP)} = 5 + 4 + 4 = 13 \\
\text{Total Data} = 15 \\
\text{Accuracy} = \frac{TP}{\text{Total Data}} \times 100\% \\
\text{Accuracy} = \frac{13}{15} \times 100\% \\
\text{Accuracy} = 86.67\%
\]

4. Conclusions

The identification of monogenean grandis, liewi, and johorensis parasites using the GLCM (gray level co-occurrence matrix) feature extraction and ANN (artificial neural network) algorithm achieved an overall accuracy of 86.67%. This research resulted in the identification of monogenic parasites that can be used to assist the work of ecologists in classifying monogenic species. In future studies it should include more datasets of the species used and choose better features so as to improve accuracy.

References

[1] Kompas.com, “Studi Baru, Populasi Ikan di Lautan Dunia Menurun Drastis,” Kompas.com, 2019. [Online]. Available: https://sains.kompas.com/read/2019/03/13/193200223/studi-baru-populasi-ikan-di-lautan-dunia-menurun-drastis.

[2] @articlekalafi2016automated, title=Automated identification of Monogeneans using digital image processing and K-nearest neighbour approaches, author=Kalafi, Elham Yousef and Tan, Wooi Boon and Town, Christopher and Dhillon, Sarinder Kaur, journal=BMC bioinformatics, volume=17, number=19, pages=511, year=2016, publisher=Springer

[3] @incollectionhutson2018monogenean, title=Monogenean parasite cultures: current techniques and recent advances, author=Hutson, Kate Suzanne and Brazenor, Alexander Karlis and Vaughan, David Brendan and Trujillo-González, Alejandro, booktitle=Advances in parasitology, volume=99, pages=61–91, year=2018, publisher=Elsevier

[4] @incollectionhutson2018monogenean, title=Monogenean parasite cultures: current techniques and recent advances, author=Hutson, Kate Suzanne and Brazenor, Alexander Karlis and Vaughan, David Brendan and Trujillo-González, Alejandro, booktitle=Advances in parasitology, volume=99, pages=61–91, year=2018, publisher=Elsevier

[5] @inproceedingsokarma2016no, title=No-reference quality assessment of 3D prints based on the GLCM analysis, author=Okarma, Krzysztof and Fastowicz, Jarosław, booktitle=2016 21st International Conference on Methods and Models in Automation and Robotics (MMAR), pages=788–793, year=2016, organization=IEEE

[6] @articlekurniawan2018analisa, title=Analisa Tekstur Kulit Wajah Menggunakan Fitur Gray Level Co-Occurrence Matrix, author=Kurniawan, Ilham and Riana, Dwiza, journal=SNIT 2018, volume=1, number=1, pages=187–192, year=2018