

## Identification of selected monogeneans using image processing, artificial neural network and K-nearest neighbor

Yousef Kalafi E.<sup>1</sup>; Wooi Boon T.<sup>1</sup>; Town C.<sup>2</sup>; Kaur Dhillon S.<sup>1\*</sup>

Received: June 2017

Accepted: August 2017

### Abstract

Over the last two decades, improvements in developing computational tools have made significant contributions to the classification of images of biological specimens to their corresponding species. These days, identification of biological species is much easier for taxonomists and even non-taxonomists due to the development of automated computer techniques and systems. In this study, we developed a fully automated identification model for monogenean images based on the shape characters of the haptor organs of eight species: *Sinodiplectanotrema malayanum*, *Diplectanum jaculator*, *Trianchoratus pahangensis*, *Trianchoratus lonianchoratus*, *Trianchoratus malayensis*, *Metahaliotrema ypsilocleithru*, *Metahaliotrema mizellei* and *Metahaliotrema similis*. Linear Discriminant Analysis (LDA) method was used to reduce the dimension of extracted feature vectors which were then used in the classification with K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN) classifiers for the identification of monogenean specimens of eight species. The need for the discovery of new characters for identification of species has been acknowledged for long by systematic parasitology. Using the overall form of anchors and bars for extraction of features led to acceptable results in automated classification of monogeneans. To date, this is the first fully automated identification model for monogeneans with an accuracy of 86.25% using KNN and 93.1% using ANN.

**Keywords:** Monogenean, Morphology, Fish parasite, Artificial neural networks, K-nearest neighbor

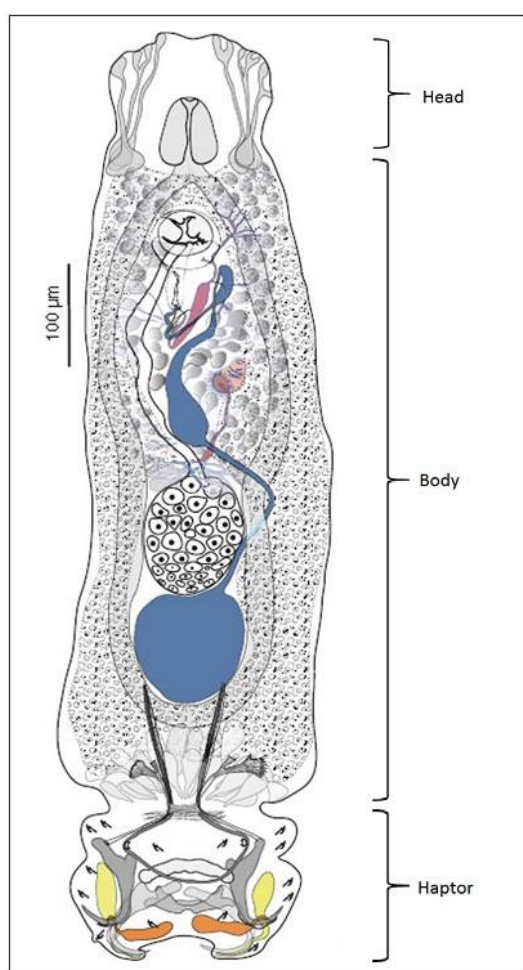
1-Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

2-Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK

\*Corresponding author's Email: sarinder@um.edu.my

## Introduction

Monogeneans are platyhelminthes which are characterized by having a proper body and haptor with sizes ranging from 0.5mm to 1-2cm in length live on lower aquatic invertebrates or the gills, skin or fins of fishes as hosts. Their appendage attachments in their anterior and posterior (haptoral) regions are used to prevent physical dislodgement from the host (Fig. 1).



**Figure 1: Illustration of a monogenean worm consisting of three main parts: head, body and haptor.**

Monogeneans are a diverse group, with several thousand species described in the world (Poulin, 2002). The diversity of monogeneans is not only in terms of numbers but also in terms of their

morphology and ecology and with respect to the variation of structural designs in the attachment organs (Kearn, 1994), which are usually used for species identification. The haptoral attachment organs of monogeneans are sclerotized structures of anchors, bars and marginal hooks. In particular, the morphology of each of these organs is unique to monogenean species (Boeger and Kritsky, 1993) and is used as a diagnostic feature in their taxonomical classification (Vignon, 2011).

Earlier, Active Shape Models (ASM) (Ali *et al.*, 2012) were used to classify several *Gyrodactylus* species according to attachment hooks. ASM were applied to extract diagnostic information from hook images as features. Extracted features were used as input data to Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP) and Support Vector Machine (SVM) classifiers. According to Khang *et al.* (2016), data from size and shape of anchors were generated using geometric morphometrics. They used principal components and cluster analysis to classify 13 species of *Ligophorus*.

Innovations in the area of computer vision have significantly contributed to the development of automated taxonomic identification systems such as an automated identification system which estimates densities of whiteflies, aphids and thrips in a greenhouse (CHO *et al.*, 2008), automatic image recognition and diagnosis of protozoan parasites (Castañón *et al.*, 2007), automatic recognition of biological particles in microscopic images

(Ranzato *et al.*, 2007), automatic detection of malaria parasites for estimating parasitemia (Savkare and Narote, 2011), automated identification of copepods using digital image processing and artificial neural networks (Leow *et al.*, 2015), automated identification of fish species based on otolith contour using short-time Fourier transform and discriminant analysis (STFT-DA) (Salimi *et al.*, 2016), and other systems (Larios *et al.*, 2008; Vogt *et al.*, 2009; Mansoor *et al.*, 2011; Feng *et al.*, 2016; Perre *et al.*, 2016).

Many classification methods such as Artificial Neural Network (ANN) (Yang *et al.*, 2001; Cho *et al.*, 2008; Mansoor *et al.*, 2011), KNN (Keller *et al.*, 1985; Parisi-Baradad *et al.*, 2010), SVM (Thiel *et al.*, 1996; Pronobis *et al.*, 2010), Discriminant Analysis (DA) (Thiel *et al.*, 1996; Salimi *et al.*, 2016), Decision trees (Jalba *et al.*, 2005), Semantically-Related Visual (SRV) (Feng and Bhanu, 2013), and Convolutional Neural Networks (Gomez *et al.*, 2016), etc. have been utilized for developing automated identification systems.

Automated classification of images of specimens requires development of models and methods that are able to characterize species images based on the texture or shape of objects to extract important visual information for classification. Current approaches in monogenean identification rely heavily on manual input during image processing and feature extraction such as specifying morphological landmark features. These manual identification

methods are performed on every single image (specimen) which substantially slows down the process of identification and classification. Hence, we propose a fully automated identification model for monogeneans which is robust with respect to variable imaging conditions and damaged specimens.

### Materials and methods

Recognition of monogeneans is based on morphometric features of their hard parts (Lim and Gibson, 2010). For this study, images of the hard haptor organs such as anchors and bars were captured using a Leica digital camera DFC 320 attached to Leica DMRB microscope at 40× magnification. The resolution of the images was 1044×772 pixels and they were saved in Tagged Image File format (TIF).

Our database consists of 160 images from 8 species (20 images of each species):

*Sinodiplectanotrema malayanum*, *Diplectanum jaculator*, *Trianchoratus pahangensis*, *Trianchoratus lonianchoratus*, *Trianchoratus malayensis*, *Metahaliotrema ypsilocleithru*, *Metahaliotrema mizellei* and *Metahaliotrema similis*. Fig. 2

illustrates the flowchart for the development of automated identification for monogeneans.

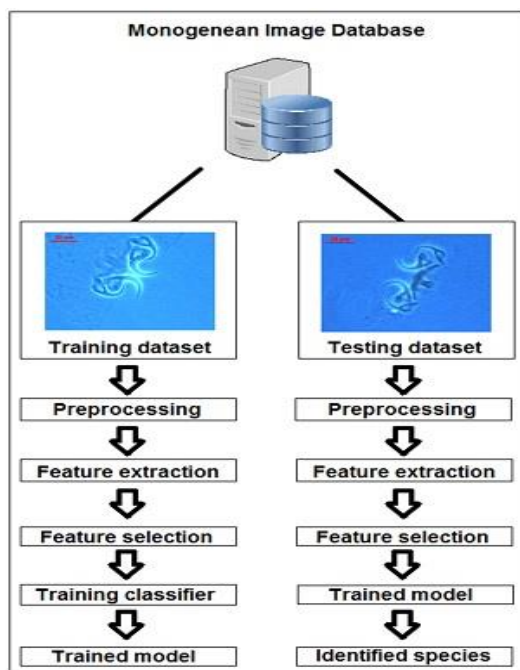


Figure 2: Flowchart for development of proposed model for monogenean identification.

### Preprocessing

Monogenean specimen images are very complex due to their messy background and overlapping of anchors and bars. Despite consistent efforts to acquire clear images, some overlapping and clutter were unavoidable (Fig. 3).

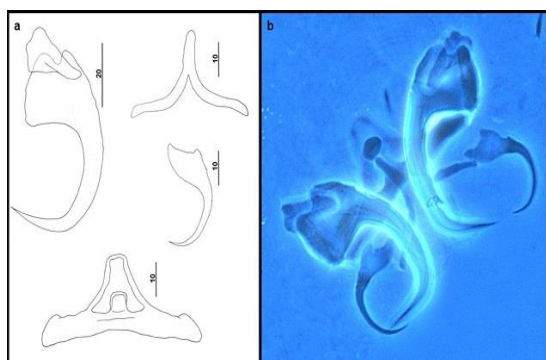


Figure 3: The illustration of anchors and bars of *Metahaliotrema ypsilocleithrum*. a) The illustration of dorsal and ventral anchors and bars. b) The microscopic image of anchors and bars and their overlapping.

Hence, preprocessing played an important role to omit redundant information and to highlight reliable

features for the next process in feature extraction. Preprocessing started with converting RGB images to intensity images. Then intensity images were filtered and edges of anchors and bars were detected (Fig. 4).

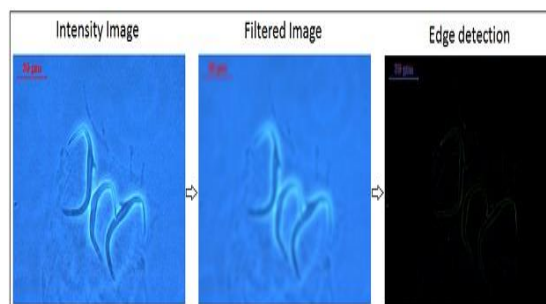


Figure 4: The process of detecting edges from intensity image.

Since the segmented images contained negative and positive values, the images were binarized with a threshold of zero. Then the borders were cleared and objects smaller than 1000 pixels were removed (Fig. 5). Coordinates of contour pixels for species' anchors were calculated. Features were extracted either from all anchors and bars as a consolidated object or from individual anchors.

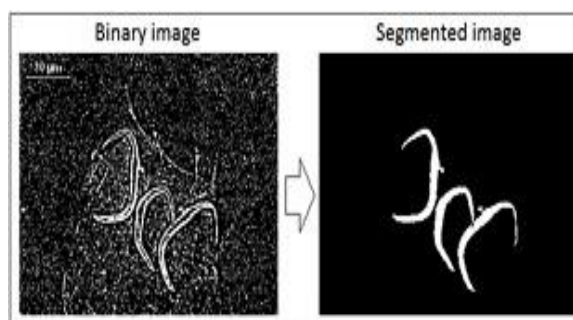


Figure 5: The process of converting binary image to segmented image.

### Feature extraction

Binary images were used two times for feature extraction; (i) firstly using all

anchors and bars as a consolidated object and (ii) secondly by calculating coordinates of one anchor and then extracting the features from that anchor. Features extracted consisted of: length of bounding box, width of bounding box, center of bounding box, orientation of bounding box, perimeter, perimeter density, area, area density, Euler number, entropy, major axis length and number of white pixels in white area. A feature vector with 24 elements was computed from these features.

#### *Feature selection*

Feature selection is a technique for reducing the dimensionality of feature vectors. In this study, the informative and independent feature vectors were transformed to smaller dimensions by using the LDA (Park and Park, 2008) method (Song *et al.*, 2010). The goal of LDA is to distinguish multiple classes by maximizing the inter-class variance and minimizing the intra-class variance. Therefore, besides projecting a feature space to a smaller subspace, the class-discriminatory information is also maintained. In this approach, firstly, 24 dimensional mean vectors for each of the 8 classes were calculated. After computing the in-between class and within-class scatter matrix, the corresponding eigenvalues and eigenvectors were calculated. Then eigenvectors were sorted according to descending eigenvalues and seven eigenvectors with largest eigenvalues were selected to form  $24 \times 7$  dimensional matrix  $W$ . The matrix  $W$  will be used to transform the samples to new subspace

of  $160 \times 7$  feature vector.

#### *Classification*

The features that were extracted and selected in the previous stage were used as inputs to KNN and ANN classifiers to train the system based on a training set and evaluate performance of each trained model on a testing dataset.

Commonly, practical and theoretical data do not follow the same assumptions and since our dataset was from the real world, it was an advantage to use KNN. Therefore, no hypothesis was made on the fundamental data distribution. In addition, based on previous approaches in the classification of monogenean samples (McHugh *et al.*, 2000; Ali *et al.*, 2011; Ali *et al.*, 2012), the performance of KNN was as reliable to be used in our investigation. In this study, the trained model from the KNN classifier was constructed using 80 images (10 for each species) and tested with 80 images of monogeneans. According to successful experiments in (Sang-Hee, 2010; Jin *et al.*, 2015), we had decided to use half (10 images) of each species images to train the classifier and the other half as a testing set and the best results were achieved with 9 nearest neighbors. The ANN classifier structure was a two layer feed-forward network with ten sigmoid hidden nodes and eight output neurons (Fig. 6). The number of nodes in the hidden layer is important since they may cause overfitting (if there are so many neurons in hidden layer) or underfitting (if the number of neurons are few), therefore, we used trial and error

method to select the number of hidden neurons. This approach began by selecting two to fifteen nodes. The best result was achieved with ten nodes. We divided the entire dataset (160 images) into three random subsets: training (70%), testing (15%) and validation (15%) set. The training dataset was used for training ANN, the testing dataset for performance measurement of the networks and the validation set to measure generalization of network and terminate training before overfitting.

For evaluating the trained network we used confusion matrices and Mean Square Error (MSE).

In both KNN and ANN the best results were achieved after 20 iterations. The best results in KNN classification were obtained using 50% of samples in testing and 50% in training data. On the other hand the best results in ANN were accomplished by using 70% of samples in training, 15% in testing and 15% in validation set.

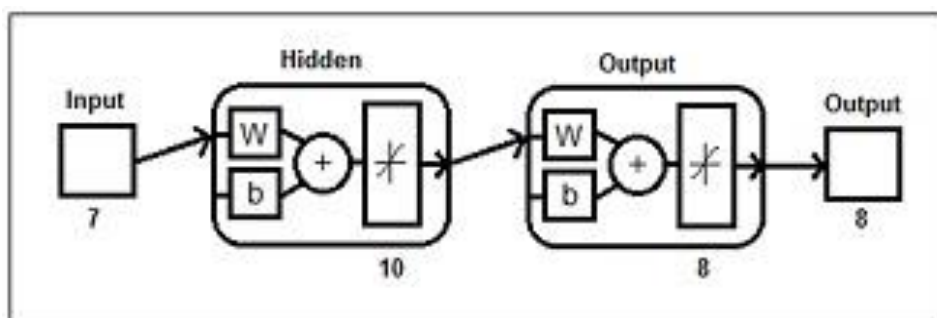


Figure 6: Neural Network with 10 sigmoid hidden nodes and eight output neurons.

## Results

The results of this study were achieved by using QWin Plus software package as imaging modular and MATLAB R2013a [28] as simulation and image processing and classification tool, installed on Intel(R) Xeon (R) CPU E5-1620 v2 @ 3.70GHz, 16 GB RAM, Windows 7 Professional (64-bit).

### Feature selection

The feature space was defined according to shape characteristics of anchors and bars of monogeneans. A total of 24 feature elements were initially extracted (Table 1). By adopting LDA (dimensionality reduction method), the feature vector

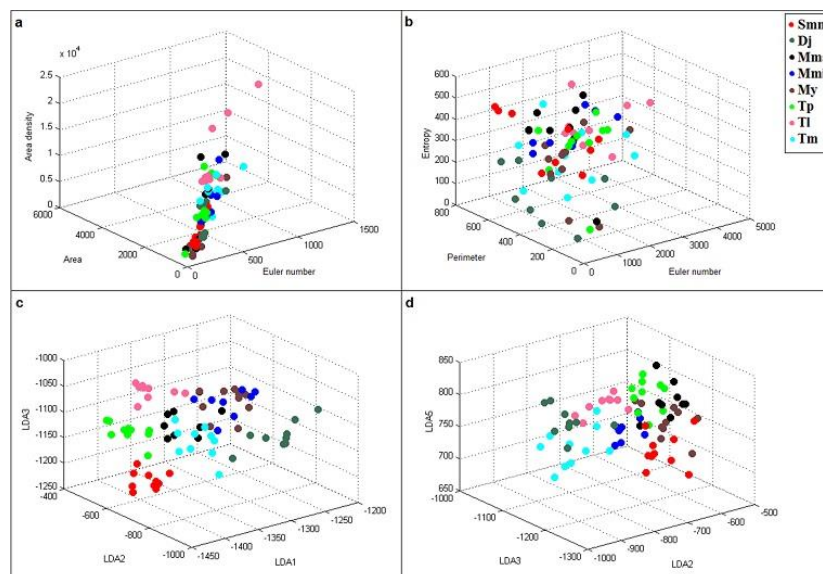
with 24 elements was transformed to a lower dimensional feature space with seven distinct elements. First, 24 dimensional mean vectors for each of the 8 classes were calculated. After computing the in-between class and within-class scatter matrix, the corresponding eigenvalues and eigenvectors were calculated. Then eigenvectors were sorted according to descending eigenvalues and seven eigenvectors with largest eigenvalues were selected to form  $24 \times 7$  dimensional matrix  $W$ . The matrix  $W$  will be used to transform the samples to a new subspace of  $160 \times 7$  feature vector.

**Table 1: Extracted features from shape characteristics of monogenean.**

Anchor	Anchors and bars together	Features
×	×	Area
×	×	Area density
×	×	Perimeter
×	×	Perimeter density
×	×	Length of bounding box
×	×	Width of bounding box
×	×	Center of bounding box
×	×	Orientation of bounding box
×	×	Euler number
×	×	Entropy
×	×	Major axis length
		Number of white pixels in white area

The distribution of some of the feature values are illustrated in Fig. 7 as a 3D scatter plot. It is notable that in Fig. 7(a,b), the features are not distinct enough to separate and classify eight classes. Instead, in Fig. 7 (c,d), after

LDA feature transformation, the features are transformed to a discrete feature space in which the eight species are more easily separable for classification.



**Figure 7: Distribution of feature values in 3D scatter plot. a) Illustration of distribution of features values which were extracted from area density, area and Euler number of anchors and bars of monogenean. b) Illustration of distribution of features values which were extracted from entropy, perimeter and Euler number of anchors and bars of monogenean. c & d) Illustration of distribution of features values which were selected by LDA.**

*Classification*

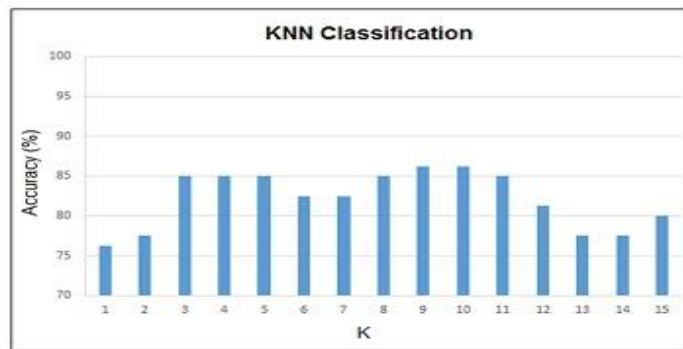
The experiment was conducted on eight species of four monogenean families,

classified using KNN and ANN. In KNN we used 80 images for training and 80 images for testing the trained

model. In ANN, 112 images were used for training, 24 images for testing the network and 24 images for system validation. We found that ANN accuracy of 93.1% on the test set outperformed KNN classifier accuracy of 86.25%.

#### *K-Nearest Neighbor (KNN)*

In KNN classification, we tried the classification in 15 iterations and the best score was achieved with 9 nearest neighbors (Fig. 8).



**Figure 8: Accuracy of KNN with different K values. It is notable that when k is 9 and 10 KNN classifier yielded the best performance.**

According to the confusion matrix (Table 2), the overall classification score for 8 species was 86.25%. In Table 2, *S. malayanum* and *M. similis* were identified correctly. *T. pahangensis* has one misclassification with *M. similis*. *M. mizellei* has one misclassification with *T. pahangensis*. *T. lonianchoratus* has one misclassification with *M. mizellei* and

one misclassification with *M. ypsilocleithru*. *T. malayensis* has one misclassification with *T. pahangensis*. *M. ypsilocleithru* has one misclassification with samples of *T. pahangensis*, *T. malayensis*, *M. similis* and two misclassifications with *M. mizellei*. Finally, *D. jaculator* has only one misclassification with *M. mizellei*.

**Table 2: Confusion matrix of KNN classification for 8 species of *Sinodiplectanotrema malayanum* (Smm), *Diplectanum jaculator* (Dj), *Trianchoratus pahangensis* (Tp), *Trianchoratus lonianchoratus* (Tl), *Trianchoratus malayensis* (Tm), *Metahaliotrema ypsilocleithru* (My), *Metahaliotrema mizellei* (Mmi) and *Metahaliotrema similis* (Mma).**

		Predicted								Accuracy (%)
Species		Smm	Tp	Mmi	Mma	Tl	Tm	My	Dj	
Actual	Smm	10	0	0	0	0	0	0	0	100
	Tp	0	9	0	1	0	0	0	0	90
	Mmi	0	1	9	0	0	0	0	0	90
	Mma	0	0	0	10	0	0	0	0	100
	Tl	0	0	1	0	8	0	1	0	80
	Tm	0	1	0	0	0	9	0	0	90
	My	0	1	2	1	0	1	5	0	50
	Dj	0	0	1	0	0	0	0	9	90
	Overall									



*Artificial Neural Network (ANN)*

The ANN classification structure was a two layer feed-forward network which was trained with back propagation and with respect to ten hidden neurons in the hidden layer and eight neurons in the output layer. After 46 iterations, the best trained network was selected with

MSE of 0.026168 on the validated set at epoch 40 (Fig. 9). According to confusion matrix in Table 3, it is notable that the best overall accomplished classification was 93.1% of all 160 images in the training, validation and testing sets.

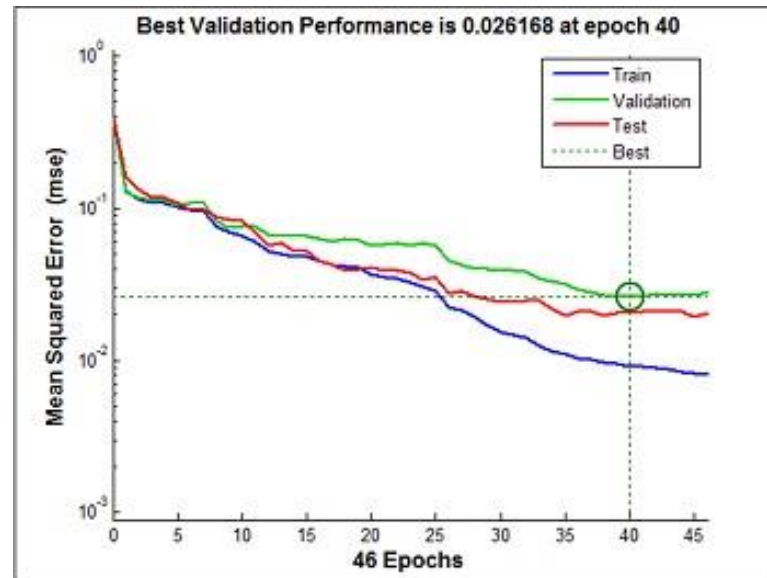


Figure 9: Illustration of performance evaluation of trained network by MSE. Best trained network was constructed at epoch 40.

Table 3: Confusion matrix of overall performance of ANN classification for 8 species of *Sinodiplectanotrema malayanum* (Smm), *Diplectanum jaculator* (Dj), *Trianchoratus pahangensis* (Tp), *Trianchoratus lonianchoratus* (Tl), *Trianchoratus malayensis* (Tm), *Metahaliotrema ypsilocleithru* (My), *Metahaliotrema mizellei* (Mmi) and *Metahaliotrema similis* (Mma).

		Predicted							Accuracy (%)	
Species		Smm	Tp	Mmi	Mma	Tl	Tm	My	Dj	
Actual	Smm	20	0	0	0	0	0	0	0	100
	Tp	0	19	0	0	0	1	0	0	95
	Mmi	0	0	19	1	1	0	0	1	86.4
	Mma	0	1	0	19	0	0	0	0	95
	Tl	0	0	0	0	18	0	0	0	100
	Tm	0	0	0	0	1	18	1	0	90
	My	1	0	1	0	0	0	17	0	89.5
	Dj	1	0	0	0	0	1	0	19	90.5
Overall									93.1	

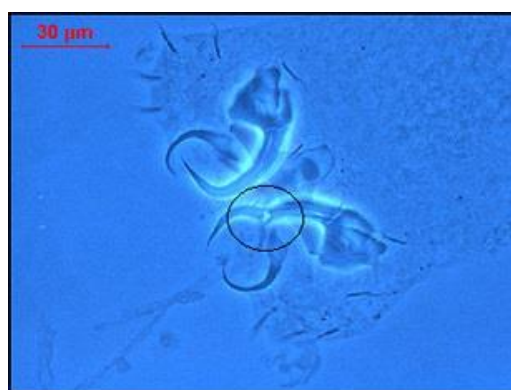
## Discussion

The proposed automated identification method in this study was able to classify monogeneans at the species level with an overall accuracy of 86.25% using K Nearest Neighbor classification and 93.1% using Artificial Neural Network classification for eight species of monogenean. The images focused on anchors and bars of specimens since these organs contain diagnostic shape features which are used for classification of monogenean species.

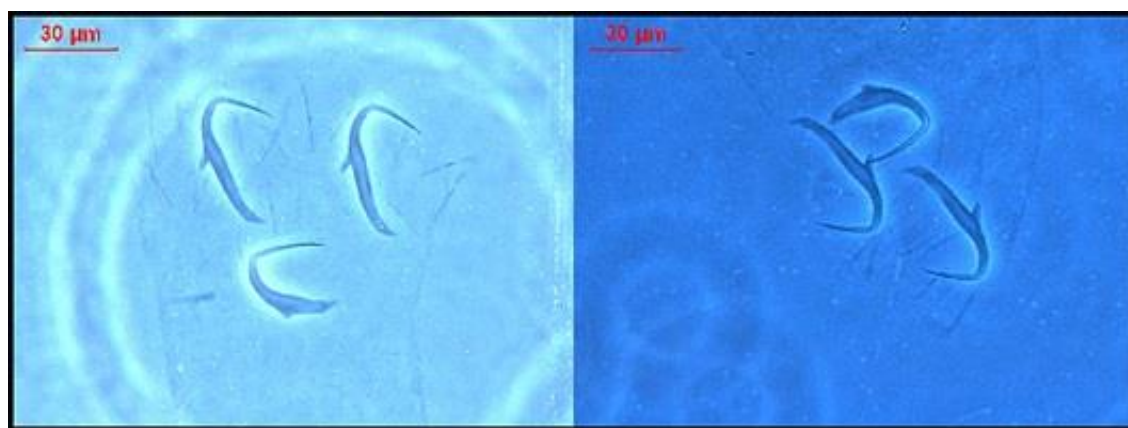
### *Image processing*

In this experiment the database also contained some poor quality images which we had to use in this study as they were rare collections which were only available in small numbers. Despite that, the automated

identification technique for monogeneans developed in this study could identify the species with an acceptable accuracy. This shows the technique is robust with respect to variable imaging conditions and damaged specimens (Figs. 10 and 11).



**Figure 10:** The broken tail of anchor in *Metahaliotrema ypsilocleithru*



**Figure 11:** Variable imaging condition in terms of lighting source.

Additionally, extracting the anchors and bars as a single organ in poor quality images was also challenging since the bars and anchors were overlapping. Such samples are difficult even for experts to distinguish. Increasing the quality of images in the future will lead

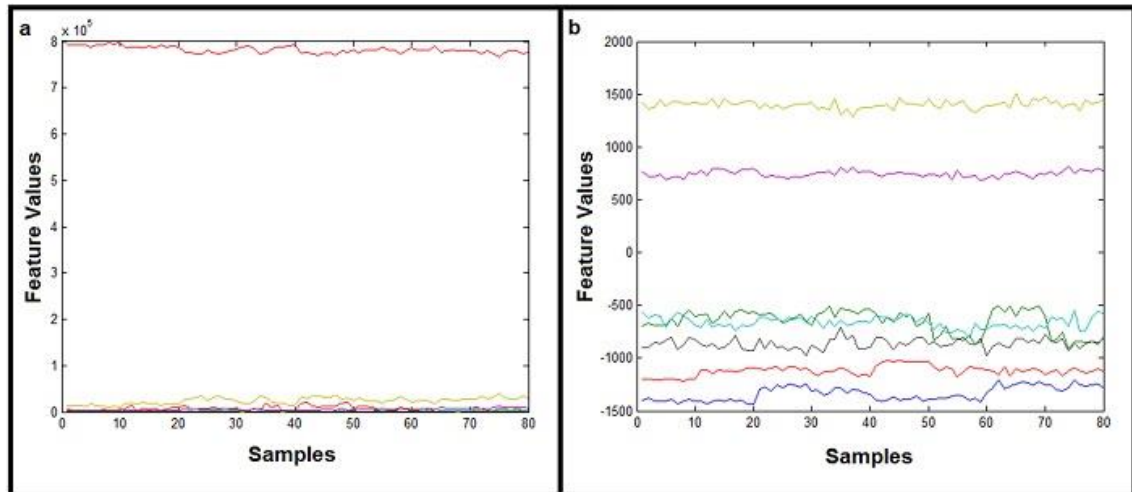
to improved identification accuracy.

### *Feature selection*

The extracted features from anchors and bars were affected by microscopic clutters, some broken specimens and overlapping of anchors and bars in

images. LDA was used for transforming feature vectors to distinct feature space of seven elements. As in Fig. 12 (a), the 24 features were close in values and using them as input to classifiers, made many misclassifications as a result, but

after using LDA (Fig. 12 (b)) the dimensionality of the feature vector decreased and feature vectors also transformed into a separable feature space.



**Figure 12: Feature vector comparison after and before feature transformation. a) Illustration of 24 dimensional extracted feature vector for 80 samples. Except one of the features, the rest contain close values. b) Illustration of 7 dimensional feature vector which is the result of LDA feature transformation.**

### Classification

In both KNN and ANN, *S. malayanum* was classified correctly due to its distinct shape and size of anchors and bars in this species. There was one misclassification of *T. pahangensis* with *M. similis* in KNN mainly because the shapes of their tails were similar and one misclassification with *T. malayensis* in ANN since both of them have three anchors. There was one misclassification of *M. mizellei* with *M. similis* in KNN since both are from the same genus, as such the overall shape of all anchors and bars as an object is similar. In KNN, the classification of *M. similis* was 100% correct while in ANN there was one misclassification with *T. pahangensis*. This could be due

to the similar shape of their tails. The classification of *T. lonianchoratus* in ANN was 100% correct while in KNN there were two misclassifications with *M. mizellei* and *M. ypsilocleithru*.

Comparing the classification accuracy in ANN (93.1%) and the accuracy in KNN (86.26%), we can declare that ANN classifiers were more powerful in classifying samples than KNN classifiers. This means in ANN the features were more distinct for training the network but the distance distinction in KNN was not sufficient for classification as much as ANN.

### Future works

One of the critical issues affecting performance of identification systems is

the quality of the images. Improving the quality of the data would lead to an increase in identification accuracy.

The acquired images were two dimensional (2D) and due to the loss of some information in 2D imaging, it is suggested that in future, the model be based on three dimensional (3D) images. As the solution to loss of information in 2D imaging, in the study by Leow *et al.* (2015), they used a built in function in their imaging software, called Extended Depth of Focus Imaging (EFI) to create a single plane image with increased in-focus details.

In this study, two classification techniques were adopted for developing an automated identification system for monogeneans. However, other classification techniques such as SVM, DA, and decision tree may improve the performance of the system. The classification performance in some of these methods is also dependent on the size of the database. The dataset size can be increased with new collections of monogenean specimens through field work which we plan to conduct in the future. The specimens used in this study were archive specimens from Lim 1998; Lim, 2006 and Lim *et al.*, 2010.

In this paper, we presented a fully automated identification technique that classifies monogeneans at the species level based on microscopic images of haptor anchors and bars with an overall accuracy of 86.25% using K Nearest Neighbors and 93.1% using Artificial Neural Network. The need for the discovery of new characters for identification of species has been acknowledged for a long time by

systematic parasitology (Vignon, 2011) and because of the lack of discrimination of traditional methods, several researchers have used additional points to take into account the maximum amount of shape information (Murith and Beverley-Burton, 1985; Řehulková and Gelnar, 2005). Using overall form of anchors and bars for extraction of features led us to create new characters in the morphological classification of monogeneans which had never been used before.

This study proposes a model for automated identification of images of eight selected monogenean species and it works by running the commands in the MATLAB workspace which means there is no dedicated user interface. As future work, a stand-alone Graphical User Interface (GUI) could be deployed as an executable application for ease of use by taxonomists. To increase the number of species in the proposed model, further enhancements are required.

### Acknowledgments

The authors are thankful to late Prof. Susan Lim Lee Hong from whom all the specimens were obtained. This project was supported by University of Malaya Postgraduate Research Fund (PG092-2013B) to the first author and the University of Malaya Research Grant (UMRG) Program Based Grant (RP008-2012A) and the University of Malaya's Living Lab Grant Program – Sustainability Science (LL020-16SUS) to the fourth author.

## References

- Ali, R., Hussain, A., Bron, J.E. and Shinn, A.P., 2011.** Multi-stage classification of *Gyrodactylus* species using machine learning and feature selection techniques. In 2011 11th International Conference on Intelligent Systems Design and Applications. 2011 11th International Conference on Intelligent Systems Design and Applications. pp. 457–462. <http://ieeexplore.ieee.org/document/6121698/> 30 September 2016.
- Ali, R., Hussain, A., Bron, J.E. and Shinn, A.P., 2012.** The use of ASM feature extraction and machine learning for the discrimination of members of the fish ectoparasite genus *Gyrodactylus*. In T. Huang, Z. Zeng, C. Li, & C. S. Leung, eds. Neural Information Processing. Lecture Notes in Computer Science. Springer Berlin Heidelberg. pp. 256–263. [http://link.springer.com/chapter/10.1007/978-3-642-34478-7\\_32](http://link.springer.com/chapter/10.1007/978-3-642-34478-7_32) 30 September 2016.
- Boeger, W.A. and Kritsky, D.C., 1993.** Phylogeny and a revised classification of the Monogenoidea Bychowsky, 1937 (Platyhelminthes). *Systematic Parasitology*, 26(1), 1–32.
- Castañón, C.A.B., Fraga, J.S., Fernandez, S., Gruber, A. and da F. Costa, L., 2007.** Biological shape characterization for automatic image recognition and diagnosis of protozoan parasites of the genus *Eimeria*. *Pattern Recognition*, 40(7), 1899–1910.
- Cho, J., Choi, J., Qiao, M., Ji, C.W. and Kim, H.Y., 2008.** Automatic identification of tobacco whiteflies, aphids and thrips in greenhouse using image processing techniques. In 4th WSEAS International Conference on MATHEMATICAL BIOLOGY and ECOLOGY. Mexico. pp. 46–53. [https://www.researchgate.net/profile/Chang\\_Woo\\_Ji2/publication/264853945\\_Automatic\\_Identification\\_of\\_Tobacco\\_Whiteflies\\_Aphids\\_and\\_Thrips\\_in\\_Greenhouse\\_Using\\_Image\\_Processing\\_Techniques/links/55ca8f7108aebc967dfbe60e.pdf](https://www.researchgate.net/profile/Chang_Woo_Ji2/publication/264853945_Automatic_Identification_of_Tobacco_Whiteflies_Aphids_and_Thrips_in_Greenhouse_Using_Image_Processing_Techniques/links/55ca8f7108aebc967dfbe60e.pdf) 30 September 2016.
- Feng, L. and Bhanu, B., 2013.** Automated identification and retrieval of moth images with semantically related visual attributes on the wings. In 20th IEEE International Conference on Image Processing (ICIP). pp. 2577–2581. <http://ieeexplore.ieee.org/document/6738531/> 30 September 2016.
- Feng, L., Bhanu, B. and Heraty, J., 2016.** A software system for automated identification and retrieval of moth images based on wing attributes. *Pattern Recognition*, 51, 225–241.
- Gomez, A., Salazar, A. and Vargas, F., 2016.** Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. arXiv:1603.06169 [cs]. <http://arxiv.org/abs/1603.06169> 28 December 2016.
- Jalba, A.C., Wilkinson, M.H.F.,**

- Roerdink, J.B.T.M., Bayer, M.M. and Juggins, S., 2005.** Automatic diatom identification using contour analysis by morphological curvature scale spaces. *Machine Vision and Applications*, 16(4), 217–228.
- Jin, T., Hou, X., Li, P. and Zhou, F., 2015.** A novel method of automatic plant species identification using sparse representation of leaf tooth features. *PLoS ONE*, 10(10), e0139482.
- Kearn, G.C., 1994.** Evolutionary expansion of the monogenea. *International Journal for Parasitology*, 24(8), 1227–1271.
- Keller, J.M., Gray, M.R. and Givens, J.A., 1985.** A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 4, 580–585.
- Khang, T.F., Soo, O.Y.M., Tan, W.B. and Lim, L.H.S., 2016.** Monogenean anchor morphometry: systematic value, phylogenetic signal, and evolution. *PeerJ*, 4, e1668.
- Larios, N., Deng, H., Zhang, W., Sarpola, M., Yuen, J., Paasch, R., Moldenke, A., Lytle, D.A., Correa, S.R., Mortensen, E.N., Shapiro, L.G. and Dietterich, T.G., 2008.** Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. *Machine Vision and Applications*, 19(2), 105–123.
- Leow, L.K., Chew, L.L., Chong, V.C. and Dhillon, S.K., 2015.** Automated identification of copepods using digital image processing and artificial neural network. *BMC Bioinformatics*, 16(18), 1.
- Lim, L.H.S., 2006.** Diplectanids (Monogenea) on the archerfish *Toxotes jaculatrix* (Pallas) (Toxotidae) off Peninsular Malaysia. *Systematic Parasitology*, 64(1), 13–25.
- Lim, L.H.S., 1998.** Diversity of monogeneans in Southeast Asia. *International Journal for Parasitology*, 28(10), 1495–1515.
- Lim, L.H.S. and Gibson, D.I., 2010.** Taxonomy, taxonomists and biodiversity. *Sarawak Biodiversity Centre*, pp. 33–43.
- Lim, L.H.S., Tan, W.B. and Gibson, D.I., 2010.** Description of *Sinodiplectanotre mamalayanum* n. sp. (Monogenea: Diplectanidae), with comments on the taxonomic position of the genus. *Systematic Parasitology*, 76(2), 145–157.
- Mansoor, H., Sorayya, M., Aishah, S., Mogeheb, A. and Mosleh, A., 2011.** Automatic recognition system for some cyanobacteria using image processing techniques and ANN approach. In Int Conf on Envir and Comp Science IPCBEE. pp. 73–78.
- McHugh, E.S., Shinn, A.P. and Kay, J.W., 2000.** Discrimination of the notifiable pathogen *Gyrodactylus salaris* from *G. thymalli* (Monogenea) using statistical classifiers applied to morphometric data. *Parasitology*, 121(Pt 3), 315–323.
- Murith, D. and Beverley-Burton, M., 1985.** *Salsuginus Beverley-Burton*, 1984 (Monogenea:

- Ancyrocephalidae) from Cyprinodontoidei (Atheriniformes) in North America with descriptions of *Salsuginus angularis* (Mueller, 1934) Beverley-Burton, 1984 from *Fundulus diaphanus* and *Salsuginus heteroclitii* n. sp. from *F. heteroclitus*. *Canadian Journal of Zoology*, 63(3), 703–714.
- Parisi-Baradad, V., Manjabacas, A., Lombarte, A., Olivella, R., Chicò, J. and García-Ladona, E., 2010.** Automated taxon identification of teleost fishes using an otolith online database—AFORO. *Fisheries Research*, 105(1), 13–20.
- Park, C.H. and Park, H., 2008.** A comparison of generalized linear discriminant analysis algorithms. *Pattern Recognition*, 41(3), 1083–1097.
- Perre, P., Faria, F.A., Jorge, L.R., Rocha, A., Torres, R.S., Souza-Filho, M.F., Lewinsohn, T.M. and Zucchi, R.A., 2016.** Toward an automated identification of anastrepha fruit flies in the fraterculus group (Diptera, Tephritidae). *Neotropical Entomology*. pp. 1–5.
- Poulin, R., 2002.** The evolution of monogenean diversity. *International Journal for Parasitology*, 32(3): 245–254.
- Pronobis, A., Jie, L. and Caputo, B., 2010.** The more you learn, the less you store: Memory-controlled incremental SVM for visual place recognition. *Image and Vision Computing*, 28(7), 1080–1097.
- Ranzato, M., Taylor, P.E., House, J.M., Flagan, R.C., LeCun, Y. and Perona, P., 2007.** Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters*, 28(1), 31–39.
- Řehulková, E. and Gelnar, M., 2005.** A revised diagnosis of thylacicleidus (monogenea: dactylogyridae) with a redescription of the type species, *Thylacicleidus serendipitus*, and descriptions of two new species from Southeast Asian pufferfishes (tetraodontiformes: tetraodontidae). *Journal of Parasitology*, 91(4), 794–807.
- Salimi, N., Loh, K.H., Dhillon, S.K. and Chong, V.C., 2016.** Fully-automated identification of fish species based on otolith contour: using short-time Fourier transform and discriminant analysis (STFT-DA). *PeerJ*, 4, e1664.
- Sang-Hee, L., 2010.** A novel approach to shape recognition using shape outline. *Journal of the Korean Physical Society*, 56(31), 1016.
- Savkare, S.S. and Narote, S.P., 2011.** Automatic detection of malaria parasites for estimating parasitemia. *International Journal of Computer Science and Security (IJCSS)*, 5(3): 310.
- Song, F., Mei, D. and Li, H., 2010.** Feature selection based on linear discriminant analysis. In *Intelligent System Design and Engineering Application (ISDEA)*, 2010 International Conference on. pp. 746–749. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5743287](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5743287) 10 November 2016.
- Thiel, S.U., Wiltshire, R.J. and**

- Davies, L.J., 1996.** Automated object recognition of blue-green algae for measuring water quality-a preliminary study. *Oceanographic Literature Review*, 1(43), 85.
- Vignon, M., 2011.** Putting in shape – towards a unified approach for the taxonomic description of monogenean haptor hard parts. *Systematic Parasitology*, 79(3), 161–174.
- Vogt, A., Cholewinski, A., Shen, X., Nelson, S.G., Lazo, J.S., Tsang, M. and Hukriede, N.A., 2009.** Automated image-based phenotypic analysis in zebrafish embryos. *Developmental Dynamics*, 238(3), 656–663.
- Yang, Y.S., Park, D.K., Kim, H.C., Choi, M.H. and Chai, J.Y., 2001.** Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE Transactions on Biomedical Engineering*, 48(6), 718–730.