

Original Article

Comparison of Fully Automated and Semi-Automated Methods for Species Identification

(monogenean / automated species identification / image processing / classification / artificial neural networks / k-nearest neighbour)

E. Y. KALAFI¹, M. K. ANUAR¹, M. K. SAKHARKAR², S. K. DHILLON¹

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

²Drug Discovery and Development Research Group, College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, Canada

Abstract. The process of manual species identification is a daunting task, so much so that the number of taxonomists is seen to be declining. In order to assist taxonomists, many methods and algorithms have been proposed to develop semi-automated and fully automated systems for species identification. While semi-automated tools would require manual intervention by a domain expert, fully automated tools are assumed to be not as reliable as manual or semi-automated identification tools. Hence, in this study we investigate the accuracy of fully automated and semi-automated models for species identification. We have built fully automated and semi-automated species classification models using the monogenean species image dataset. With respect to monogeneans' morphology, they are differentiated based on the morphological characteristics of haptor bars, anchors, marginal hooks and reproductive organs (male and female copulatory organs). Landmarks (in the semi-automated model) and shape morphometric features (in the fully automated model) were extracted from four monogenean species images, which were then classified using k-nearest neighbour and

artificial neural network. In semi-automated models, a classification accuracy of 96.67 % was obtained using the k-nearest neighbour and 97.5 % using the artificial neural network, whereas in fully automated models, a classification accuracy of 90 % was obtained using the k-nearest neighbour and 98.8 % using the artificial neural network. As for the cross-validation, semi-automated models performed at 91.2 %, whereas fully automated models performed slightly higher at 93.75 %.

Introduction

Monogeneans are a group of ectoparasites with categorical homogeneous but diverse morphology and advance adaptive radiation in the flatworm clade (Brooks and McLennan, 1993). The variation of structural designs in the attachment organs are usually used as discriminant characters in taxonomy (Kearn, 2013). In particular, monogeneans can be classified based on haptor attachment organs such as anchors, bars and hooks. There are about 25,000 species of monogeneans (Whittington, 1998).

Taxonomists' eagerness to reduce the time consumed for analysing samples (Benfield et al., 2007) and to significantly cut down the costs (Kalafi et al., 2018) were the main reasons that influenced the development of image-based identification systems. Culverhouse et al. (2003) have shown that categorizing specimens from species that have significant variations in their morphology is difficult. They also reported that the returned accuracy by the trained personnel and experts for distinguishing and labelling specimens is expected to be in the range of 64 % to 95 %, which is within the performance range of automated methods. Classification of specimens' images to their corresponding species requires development of models and methods that are able to characterize a species' morphology and apply this knowledge to their recognition (Wong et al., 2016; Ali et al., 2017).

Received June 8, 2018. Accepted September 10, 2018.

This project was supported by the University of Malaya Research Program Grant (PRGS 2017-1) to the fourth author.

Corresponding author: Sarinder Kaur A/p Kashmir Singh, Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia. Phone: (+603) 79676741; Fax: (+603) 79674178; e-mail: sarinder@um.edu.my

Abbreviations: ANN – artificial neural network, ASM – active shape models, KNN – k-nearest neighbours, LDA – linear discriminant analysis, MHBI – monogenean haptor bar image, MLP – multilayer perceptron, MSE – mean square error, NB – naive Bayes, PC – principal component, SBS – sequential backward selection, SFFS – sequential forward floating selection, SFS – sequential forward selection, SVM – support vector machine.

Morphometric analysis plays an important role in biological taxonomy since it hands out quantitative description of the organism (See et al., 2016). Morphometric analysis is carried out by measuring linear inter-landmark distances such as length, width, and height and by use of multivariate statistical tools describe the patterns of shape variation among classes (Adams et al., 2004). However, there was a problem in producing similar results for two different specimens based on data without geometrical information by using distance measurements. To overcome this problem, a method called landmark-based morphometry was applied whereby a set of used landmarks are described by a tightly defined set of rules.

In semi-automated identification of monogenean species (Ali et al., 2012), features were extracted using active shape models (ASM) and classification was performed using two linear classifiers: linear discriminant analysis (LDA) and k-nearest neighbours (KNN) (Duda et al., 2012), and two non-linear classifiers: multilayer perceptron (MLP) and support vector machine (SVM). The study reported an accuracy of 85.71 % using LDA, 95.59 % using MLP and 98.75 % using KNN (Ali et al., 2012). KNN outperformed other classifiers due to the small sample size, as KNN is known for classifying with a limited number of datasets (Jin et al., 2015). In the study by Ververidis and Kotropoulos (2008), sequential backward selection (SBS), sequential forward selection (SFS) and sequential forward floating selection (SFFS) were used for selecting proper features for monogenean classification, and the results indicated that of the 25 features, 21 were ideal in classification of monogenean species. In multi-stage classification of monogenean species using 21 features, LDA outperformed KNN and naive Bayes (NB) (Ali et al., 2011). Morphological data from anchor and body size of 13 ligophorus (Monogenea: Ancyrocephalidae) species from Malaysia were used in a semi-automated classification model developed by Khang et al. (2016). Khang et al. (2016) employed morphological characters in anchor morphometry, adopting phylogenetic regression of anchor shape against body size and anchor size, and discovered two new morphological characters according to the length between the inner root point and the dent point, and the length between the inner root length and the outer root length. In the image retrieval framework for monogeneans (Abu et al., 2013), a monogenean haptoral bar image (MHBI) architecture was developed to demonstrate the integration of textual and content-based information in constructing a viable image retrieval system for the monogenean species.

Fully automated identification systems may significantly improve species recognition by facilitating reliable identification of any specimen in a population. Automated methods that rely on pattern recognition and image analysis have been widely applied to recognition and categorization of biological images in the field of biodiversity (Mosleh et al., 2012; Loos and Ernst, 2013; Yu et al., 2013; Leow et al., 2015; Zhan et al., 2015;

Kalafi et al., 2016; Salimi et al., 2016; Mehdipour Ghazi et al., 2017). In the study by Kalafi et al. (2016), a fully automated identification model was developed using KNN and artificial neural network (ANN) classifiers and an overall accuracy of 90 % was achieved. Kalafi et al. (2016) performed automated selection of features based on the morphology of anchors and bars used in classifiers.

Contrary to previous studies, in the current study, we compared the performance of semi-automated and fully automated models for the classification of monogeneans.

Material and Methods

Recognition of monogeneans is usually done based on the shape and size of their hard parts, which are dorsal and ventral anchors, bars, as well as their male and female copulatory organs (Kalafi et al., 2016). Thus, accurate recognition of monogeneans is very much dependent on the features that are extracted from these parts. In this study, digital images of anchors from monogenean species, namely, *Sinodiplectanotrema malayanus* (Sm), *Trianchoratus pahangensis* (Tp), *Metahaliotrema mizellei* (Mm) and *Metahaliotrema similis* (Ms) were taken using a Leica digital camera attached to a Leica microscope (Wetzlar, Germany) at 40× magnification (Fig. 1).

Semi-automated classification

In the semi-automated approach, all computations were performed using Acer Aspire ES1-432-P8HQ hardware and R program (Version 3.3.2).

Landmark coordinates

Landmarks coordinates of the monogenean anchors were plotted manually using the R program. We used monogeneaGM R package (Khang et al., 2016) to obtain the geometric morphometric data of monogenean species. In this study, five landmarks were taken: inner

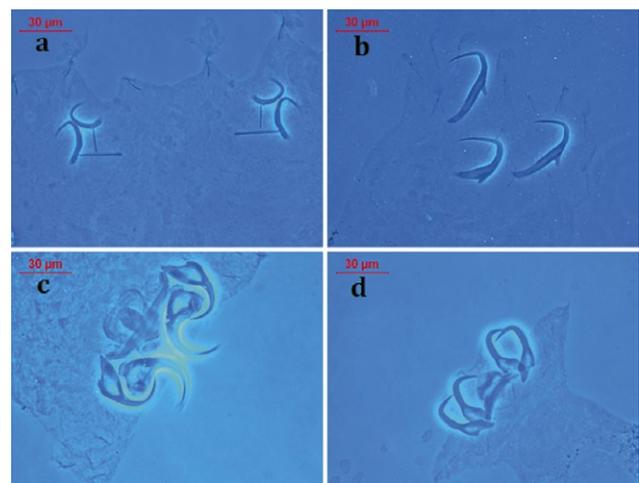


Fig. 1. Illustration of anchors and bars of monogeneans (a) *Sinodiplectanotrema malayanus*, (b) *Trianchoratus pahangensis*, (c) *Metahaliotrema mizellei*, (d) *Metahaliotrema similis*

Table 1. Features of inter-landmark distances extracted in the semi-automated identification model

Left Dorsal Inner Root (LDIR)	Left Ventral Inner Root (LVIR)
Left Dorsal Outer Root (LDOR)	Left Ventral Outer Root (LVOR)
Left Dorsal Inner Length (LDIL)	Left Ventral Inner Length (LVIL)
Left Dorsal Outer Length (LDOL)	Left Ventral Outer Length (LVOL)
Left Dorsal Point (LDPt)	Left Ventral Point (LVPt)
Left Dorsal projection from LM2 to LM4 (LD24)	Left Ventral projection from LM2 to LM4 (LV24)
Left Dorsal projection from LM1 to LM3 (LD13)	Left Ventral projection from LM1 to LM3 (LV13)
Right Dorsal Inner Root (RDIR)	Right Ventral Inner Root (RVIR)
Right Dorsal Outer Root (RDOR)	Right Ventral Outer Root (RVOR)
Right Dorsal Inner Length (RDIL)	Right Ventral Inner Length (RVIL)
Right Dorsal Outer Length (RDOL)	Right Ventral Outer Length (RVOL)
Right Dorsal Point (RDPt)	Right Ventral Point (RVPt)
Right Dorsal projection from LM2 to LM4 (RD24)	Right Ventral projection from LM2 to LM4 (RV24)
Right Dorsal projection from LM1 to LM3 (RD13)	Right Ventral projection from LM1 to LM3 (RV13)

root point, groove point, outer root point, curve point, and tip point, as in Fig. 1a. Landmark selection (“point-and-click” at the locations of the landmarks) was performed in every image in the database. Finally, a total of 20 landmark coordinates were selected for every image.

Inter-landmark distances

In the next step, inter-landmark distances between two points of the landmark coordinates were calculated. The guideline for the measurements is shown in Fig. 1b. Each of the seven inter-landmark distances contains four variables: R/L and V/D prefixes of a variable name, which indicates right or left forms and ventral or dorsal anchor, respectively. This makes up 28 features for every specimen, which are then used to represent each of the images.

Data analysis

The list of 28 features of inter-landmark distances used in the data analysis is presented in Table 1. The features were extracted from landmarks and inter-landmark distances in each anchor (Fig. 2).

These features represent each individual, and a total of 80 individuals were analysed using principal component analysis (PCA). Scatterplot was used to aid interpretation of the principal component (PC) axes.

In addition, KNN and ANN were used as classification methods to demonstrate their relative performance of classification at the species level. KNN classification is a non-parametric classification method in which pixels are assigned to classes based on the number of neighbour pixels that belong to the classes (Lamourette, 2000). Due to the small sample size used in this study, KNN with 5-fold cross-validation was applied to assess the results generalization of the system to an independent dataset. As for ANN, the multi-layered perceptron (MLP) with feed forward neural network of the ANN architecture maps input data into a set of appropriate outputs with 50 hidden layers.

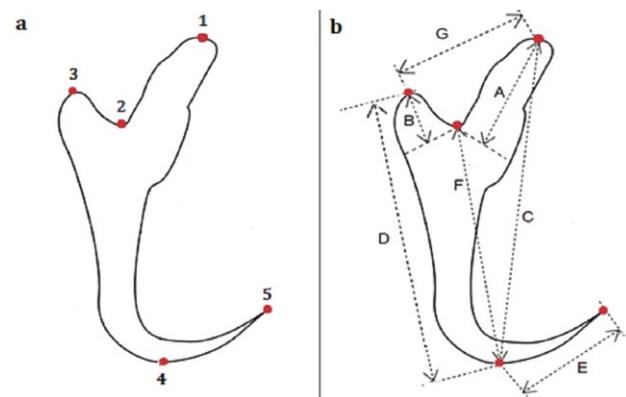


Fig. 2. Anchor morphometry

(a) Landmarks include: 1. inner root point, 2. groove point, 3. outer root point, 4. curve point, 5. tip point. (b) Inter-landmark distances include: A. inner root (IR), B. outer root (OR), C. inner length (IL), D. outer length (OL), E. point (Pt), F. projection from LM2 to LM4 (24), G. projection from LM1 to LM3 (13).

Fully automated classification

Matlab R2013a (“Image Processing Toolbox – MATLAB,” n.d.) was used as the Image Processing Toolbox, installed on Intel(R) Xeon (R) CPU E5-1620 v2 @ 3.70GHz, 16.00GB RAM, Windows 7 Professional (64-bit) to develop a fully automated identification model in this study.

Image processing

Background feature minimization is an important pre-processing step in monogenean classification (Fig. 3) to distinguish the soft part features of monogeneans from the hard parts, particularly to avoid unreliable results during texture analysis. In this study, image processing was performed using the following essential steps:

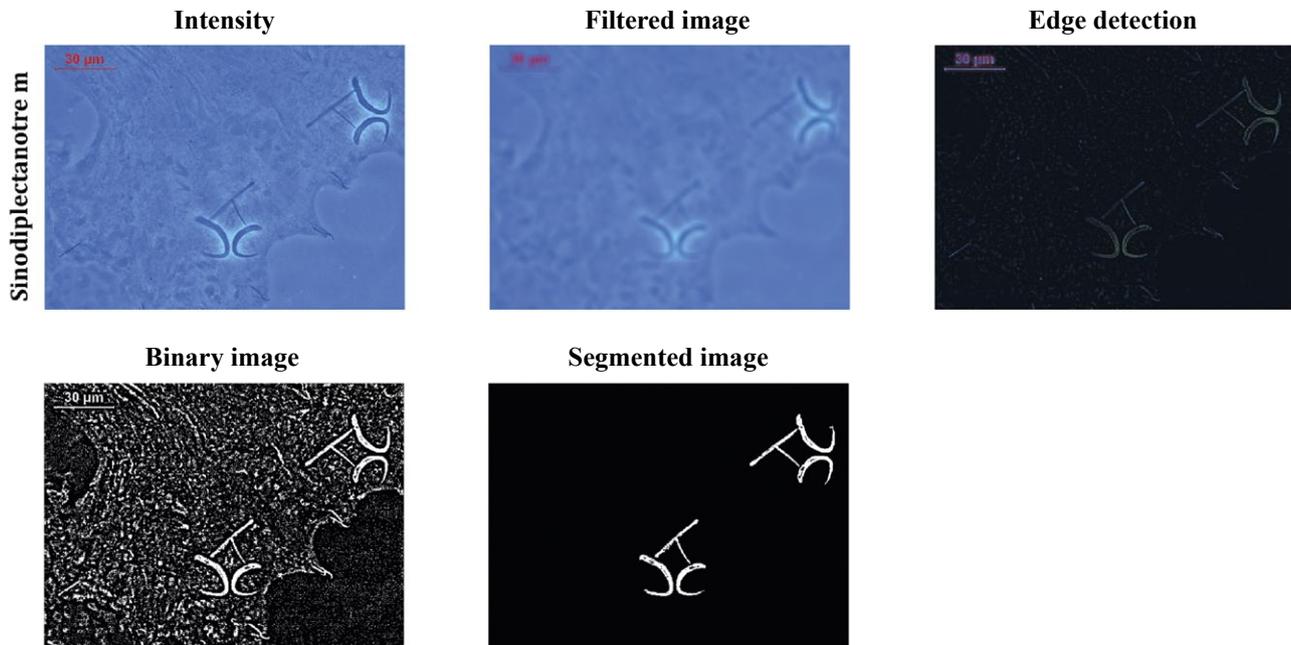


Fig. 3. Image processing steps in the fully automated identification model

- Conversion of RGB images to intensity images;
- Filtering of intensity images;
- Detection of anchor and bar edges of monogeneans;
- Conversion of images to binary images with a threshold of zero. The intensity image was deducted from the filtered image and the result was an intensity image that contained negative and positive values. Therefore, pixels greater than 0 were labelled as 1 (white) and other pixels were labelled as 0 (black).
- Particles smaller than 1000 pixels were removed and only the segmented bars and anchors were used in feature extraction.

Feature extraction

The shape descriptors were used to extract features from bars and anchors in the segmented images by using appropriate functions in Matlab. The extracted features are shown in Table 2.

Feature selection

In order to decrease the number of unnecessary features and to increase the performance of classifiers, the attributes were selected automatically using LDA. LDA was used as a feature dimensionality reduction technique as well as a pre-step for a typical classification task. After performing LDA, the feature vector of 10 elements was reduced to a feature vector of three elements.

Classification

In this study, classification was performed using two classifiers, which are KNN and ANN. KNN identifies the test sample by a majority vote of its neighbours, which are assigned to the class that is most common among its nearest neighbours. KNN was performed in many iterations and the best performance was achieved at the third nearest neighbour in the semi-automated

Table 2. Extracted features in the fully automated model

Features	Description
Euler number	Number of objects in the region minus the number of holes in these objects
Perimeter	Distance around the boundary of the region
Area	Actual number of pixels in the region of a particular object
Area density	Mass of a substance covering a unit of area
Perimeter density	Measure of length of the perimeter of a set in free boundary
Centre of bounding box	Central point of the smallest rectangle containing the region
Length of bounding box	Length of the smallest rectangle containing the region
Width of bounding box	Width of the smallest rectangle containing the region
Orientation of bounding box	Angle between the x-axis and the major axis of the ellipse that has the same second-moments as the smallest rectangle containing the region ("Measure properties of image regions – MATLAB regionprops," n.d. 2018, https://www.mathworks.com/help/images/ref/regionprops.html)

model and at the 10th nearest neighbour in the fully automated identification model. Five-fold cross-validation was used to assess how the results of our model generalize to an independent dataset. The ANN classification structure in the fully automated identification model was a two-layer feed-forward network, which was trained with back propagation and with ten hidden neurons in the hidden layer and eight neurons in the output layer.

Results

In the semi-automated model, the value of *k* was set to 3 and data was randomly selected for training (60 %) and testing (40 %). The result of the classification achieved an overall accuracy of 96.67 % using KNN and 97.5 % using ANN.

In the fully automated model, 10 images of each monogenean species were used to train KNN. We performed KNN in many iterations to find the nearest neighbour and to do a majority voting. The best result achieved was with ten nearest neighbours with an overall accuracy of 90 %.

In the ANN classification, the overall accuracy achieved was 97.50 % and only one species was misclassified in the ANN classification. After 62 iterations, the best trained network was selected with mean square error (MSE) of 0.003972 for the validated set at epoch 56 (Fig. 4). The overall accuracy of ANN classification was 98.8 % of all 80 images in the training, validation and testing set.

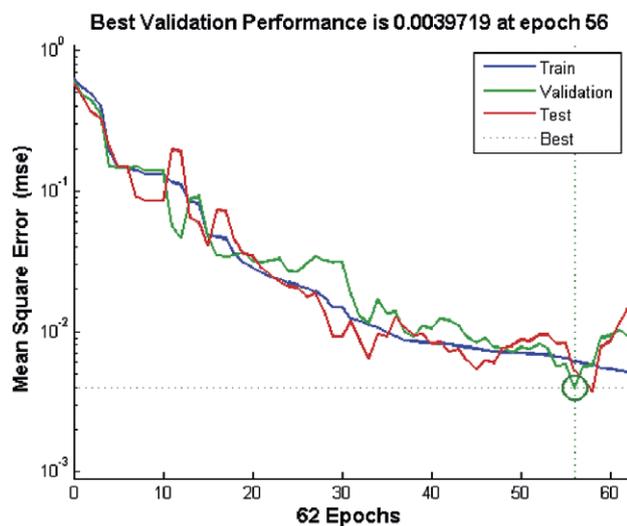


Fig. 4. Illustration of performance evaluation of the trained network by MSE

System evaluation

The performance of both semi and fully automated models were evaluated by comparing the output from the trained models with the predicted result using the testing dataset as the input.

Cross-validation is a method that is able to provide an unbiased estimation, which presents high variance with small samples. It is an efficient method to estimate the class distribution of two or more groups when the number of samples is rather small for effective sample splitting (Fu and al., 2005). In this study, 5-fold cross-validation was implemented to evaluate the performance of both models. The average accuracy of the correct classification in the semi-automated model was 91.25 % and 93.75 % in the fully automated model.

Table 3 presents the accuracy of different classification methods for four monogenean species in fully and semi-automated identification models. Overall, ANN outperforms KNN; however, KNN's accuracy was close to ANN, especially in the semi-automated model.

Discussion

Morphometric approaches are commonly built according to distance measurements. However, results of morphometric analyses can depend upon the acquired images (Kalafi et al., 2015, 2016) and the particular set of measurements chosen (Strauss and Bookstein, 1982; James Rohlf and Marcus, 1993). According to some studies, most morphological features that are extracted from haptor hard parts are highly correlated (Shinn et al., 1996; Preez and Maritz, 2006) and automatic classification of monogenean species requires improved discriminant methods for such multicollinearity, especially for small sample sizes where several morphological measurements are used to classify a few individuals (Vignon, 2011).

According to KNN and ANN classification results in both semi and fully automated identification models, it is notable that ANN outperformed KNN in both classification models. However, more importantly, the overall performance in both classification models was quite close. This demonstrates that features which were extracted automatically from shape descriptors of anchors and bars are as reliable as the features that were extracted manually from the anchors' landmarks.

In KNN classification of the semi-automated model, one sample of *Tp* was misclassified with *Mm*. This misclassification was caused by the same inter-landmark distances between both species, even though they have

Table 3. Summary of KNN, ANN and cross-validation classifications in fully and semi-automated classification

	KNN	ANN	Cross-validation
Semi-automated model	96.7	97.5	91.2
Fully automated model	90.0	98.8	93.7

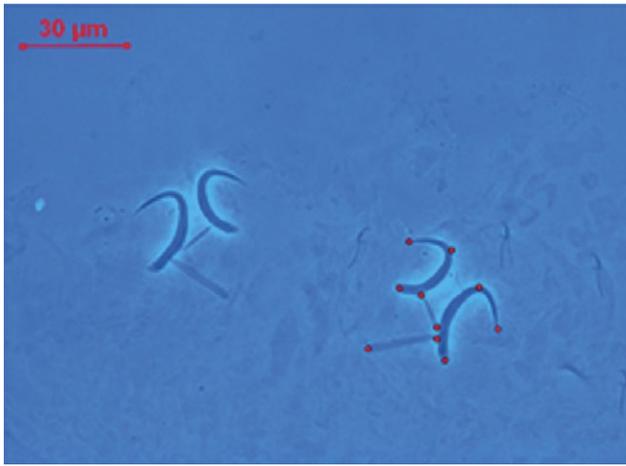


Fig. 5. Determining landmarks of *Sinodiplectanotrema malayanus* anchors

different shapes. This misclassification can be sorted if shape analysis is included to further differentiate the species.

One of the challenges in morphometric measurements is the dependency of the generated results on the quality of landmarks. In the semi-automated identification model, much more effort is required in choosing landmarks that would have high diagnostic property. Each landmark has to be present on every studied specimen (Fig. 5), and if the landmark is not present on at least one of the studied specimens, it either has to be marked approximately or it cannot be used at all (Webster and Sheets, 2010).

In the fully automated identification model, although measurements are assessed automatically, they are still highly dependent on the quality of images (Kalafi et al., 2016).

In semi-automated identification of monogeneans, features were extracted from one anchor, but in the fully automated identification model, features were extracted from all anchors and bars. Hence, more detailed image processing can be performed using the automated approach rather than manual, which is not only more tedious, but time consuming. Finally, in line with the advancement of technology and image processing approaches, it would be ideal to shift from using manual or semi-automated techniques in the species recognition to fully automated systems. It is hoped that in the era of data science, species recognition in a large amount of datasets can be done using the computing power, which is quick and easy to use.

Conclusions

In this study, ANN and KNN classifiers were used in both semi and fully automated identification models for monogenean images. Both models may assist taxonomists and monogenean experts to analyse monogenean morphology; however, preference should be given to a fully automated identification model since less effort is required to handle the morphological measurements.

Acknowledgement

Authors are thankful to late Prof. Susan Lim Lee Hong, who had contributed actively to global monogenean systematics for decades.

Competing interests

The authors declare that they have no competing interest.

References

- Abu, A., Lim, S. L. H., Sidhu, A. S., Dhillon, S. K. (2013) Biodiversity image retrieval framework for monogeneans. *Syst. Biodivers.* **11**, 19-33.
- Adams, D. C., Rohlf, F. J., Slice, D. E. (2004) Geometric morphometrics: ten years of progress following the "revolution". *Ital. J. Zool.* **71**, 5-16.
- Ali, R., Hussain, A., Bron, J. E., Shinn, A. P. (2011) Multi-stage classification of Gyrodactylus species using machine learning and feature selection techniques. In: *11th International Conference on Intelligence Systems Design and Application*, pp. 457-462.
- Ali, R., Hussain, A., Bron, J. E., 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: *ICONIP 2012. Neural Information Processing. Lecture Notes in Computer Science* **7666**, pp. 256-263, Springer, Berlin.
- Ali, N. M., Khan, H. A., Then, A. Y.-H., Ching, C. V., Gaur, M., Dhillon, S. K. (2017) Fish ontology framework for taxonomy-based fish recognition. *PeerJ.* **5**, e3811.
- Benfield, M., Grosjean, P., Culverhouse, P., Irigolen, X., Sieracki, M., Lopez-Urrutia, A., Dam, H., Hu, Q., Davis, C., Hanson, A., Pilskaln, C., Riseman, E., Schulz, H., Utgoff, P., Gorsky, G. (2007) RAPID: Research on automated plankton identification. *Oceanography (Wash. D.C.)* **20**, 172-187.
- Brooks, D. R., McLennan, D. A. (1993) Comparative study of adaptive radiations with an example using parasitic flatworms (Platyhelminthes: Cercomeria). *Am. Nat.* **142**, 755-778.
- Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., Gonzalez-Gil, S. (2003) Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Mar. Ecol. Prog. Ser.* **247**, 17-25.
- Duda, R. O., Hart, P. E., Stork, D. G. (2012) *Pattern Classification*. John Wiley & Sons Inc.
- Fu, W. J., Carroll, R. J., Wang, S. (2005) Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics* **21**, 1979-1986.
- James Rohlf, F., Marcus, L. F. (1993) A revolution morphometrics. *Trends Ecol. Evol.* **8**, 129-132.
- Jin, T., Hou, X., Li, P., Zhou, F. (2015) A novel method of automatic plant species identification using sparse representation of leaf tooth features. *PLoS One* **10**, e0139482.
- Kalafi, E. Y., Tan, W. B., Town, C., Dhillon, S. K. (2016) Automated identification of monogeneans using digital image processing and k-nearest neighbour approaches. *BMC Bioinformatics* **17**, 511.

- Kalafi, E. Y., Town, C., Dhillon, S. K. (2018) How automated image analysis techniques help scientists in species identification and classification? *Folia Morphol. (Warsz)* **77**, 179-193.
- Kearn, G. (2013) Some aspects of the biology of monogenean (Platyhelminth) parasites of marine and freshwater fishes. *Oceanography* **2**, 1-7.
- Khang, T. F., Soo, O. Y. M., Tan, W. B., Lim, L. H. S. (2016) Monogenean anchor morphometry: systematic value, phylogenetic signal, and evolution. *PeerJ*, **4**, e1668.
- Latourrette, M. (2000) Toward an explanatory similarity measure for Nearest-Neighbor Classification. In: *ECML 2000. Proceedings of the 11th European Conference on Machine Learning*, pp. 238-245, London, UK.
- Leow, L. K., Chew, L.-L., Chong, V. C., Dhillon, S. K. (2015) Automated identification of copepods using digital image processing and artificial neural network. *BMC Bioinformatics* **16**, S4.
- Loos, A., Ernst, A. (2013) An automated chimpanzee identification system using face detection and recognition. *EURASIP J. Image Video Process* **2013**, 1-17.
- Mehdipour Ghazi, M., Yanikoglu, B., Aptoula, E. (2017) Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* **235**, 228-235.
- Mosleh, M. A., Manssor, H., Malek, S., Milow, P., Salleh, A. (2012) A preliminary study on automated freshwater algae recognition and classification system. *BMC Bioinformatics* **13**, S25.
- Preez, L. H. D., Maritz, M. F. (2006) Demonstrating morphometric protocols using polystome marginal hooklet measurements. *Syst. Parasitol.* **63**, 1-15.
- Salimi, N., Loh, K. H., Dhillon, S. K., 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.
- See, M., Marsham, S., Chang, C. W., Chong, V. C., A. Sasekumar, Dhillon, S. K., Loh, K. H. (2016) The use of otolith morphometrics in determining the size and species identification of eight mullets (Mugiliformes: Mugilidae) from Malaysia. *Sains Malaysiana* **45**, 735-743.
- Shinn, A. P., Clers, S. des, Gibson, D. I., Sommerville, C. (1996) Multivariate analyses of morphometrical features from Gyrodactylus spp. (Monogenea) parasitising British salmonids: light microscope based studies. *Syst. Parasitol.* **33**, 115-125.
- Strauss, R. E., Bookstein, F. L. (1982) The truss: body form reconstructions in morphometrics. *Syst. Zool.* **31**, 113-135.
- Ververidis, D., Kotropoulos, C. (2008) Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition. *Signal Process.* **88**, 2956-2970.
- Vignon, M. (2011) Putting in shape – towards a unified approach for the taxonomic description of monogenean hap-toral hard parts. *Syst. Parasitol.* **79**, 161-174.
- Webster, M., Sheets, H. D. (2010) A practical introduction to landmark-based geometric morphometrics. *Quant. Methods Paleobiol.* **16**, 163-188.
- Whittington, I. D. (1998) Diversity “down under”: monogeneans in the Antipodes (Australia) with a prediction of monogenean biodiversity worldwide. *Int. J. Parasit.* **28**, 1481-1493.
- Wong, J. Y., Chu, C., Chong, V. C., Dhillon, S. K., Loh, K. H. (2016) Automated otolith image classification with multiple views: an evaluation on Sciaenidae. *J. Fish Biol.* **89**, 1324-1344.
- Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., Huang, T. (2013) Automated identification of animal species in camera trap images. *EURASIP J. Image Video Process* **2013**, 52-53.
- Zhan, M., Crane, M. M., Entchev, E. V., Caballero, A., de Abreu, D. A. F., Ch'ng, Q., Lu, H. (2015) Automated processing of imaging data through multi-tiered classification of biological structures illustrated using *Caenorhabditis elegans*. *PLoS Comput. Biol.* **11**, e1004194.