Classification of Fish Ectoparasite Genus *Gyrodactylus* SEM Images Using ASM and Complex Network Model

Rozniza Ali[1,2], Bo Jiang[3], Mustafa Man[1], Amir Hussain[2], and Bin Luo[3]

[1] School of Informatic and Applied Mathematics, Universiti Malaysia Terengganu, Malaysia
{rozniza,mustafaman}@umt.edu.my
[2] Institute of Computing Science and Mathematics, University of Stirling, UK
{ali,ahu}@cs.stir.ac.uk
[3] Computer Science Department, Anhui University, China
{jiangbo,luobin}@ahu.edu.cn

Abstract. Active Shape Models and Complex Network method are applied to the attachment hooks of several species of *Gyrodactylus*, including the notifiable pathogen *G. salaris*, to classify each species to their true species type. ASM is used as a feature extraction tool to select information from hook images that can be used as input data into trained classifiers. Linear (*i.e.* LDA and K-NN) and non-linear (*i.e.* MLP and SVM) models are used to classify *Gyrodactylus* species. Species of *Gyrodactylus*, ectoparasitic monogenetic flukes of fish, are difficult to discriminate and identify on morphology alone and their speciation currently requires taxonomic expertise. The current exercise sets out to confidently classify species, which in this example includes a species which is notifiable pathogen of Atlantic salmon, to their true class with a high degree of accuracy. The results show that Multi-Layer Perceptron (MLP) is the best classifier for performing the initial classification of *Gyrodactylus* species, with an average of 98.36%. Using MLP classifier, only one species has been misallocated. It is essential, therefore, to employ a method that does not generate type I or type II misclassifications where *G. salaris* is concerned. In comparison, only K-NN classifier has managed to to achieve full classification on the *G. salaris*.

Keywords: *Gyrodactylus*, classification, Active Shape Model, Complex Network.

1 Introduction

There are over 440 described species of *Gyrodactylus* which are typically small ($<$1mm), ectoparasitic monogenetic flukes of fish [14]. While most species of *Gyrodactylus* are non-pathogenic, causing little harm to their hosts, other species like *Gyrodactylus salaris* Malmberg, 1957, which is an OIE (Office International des Epizooties) - listed pathogen of Atlantic salmon, has led to a catastrophic decimation in the size of the juvenile salmon population in over 40 Norwegian rivers [7].

Uncontrolled increases in the size of the parasite population on resident salmon populations have necessitated extreme measures such as the use of the biocide rotenone to kill-out entire river systems, to remove the entire fish population within a river and the parasite [7]. Given the impact that *G. salaris* has had in Norway and elsewhere in Scandinavia [1], many European states including the UK now have mandatory surveillance

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 103–110, 2014.

programmes screening wild salmonid populations (*i.e.* brown trout, charr, grayling, Atlantic salmon etc) for the presence of notifiable pathogens including *G. salaris*. Current OIE methodologies for the identification of *G. salaris* from other species of *Gyrodactylus* that occur on salmonids require confirmation from both morphological and molecular approaches, which can be time consuming. If *G. salaris* specimens, however, are overlooked in a diagnostic sample or misclassified, the environmental and economic implications can be severe [20]. For this reason and because of the widely varying pathogenicity seen between closely related species, accurate pathogen identification is of paramount importance.

The discrimination of species from their congeners, however, is compounded by a limited number of morphological discrete characteristics which makes identification difficult. The task of morphological identification is, therefore, currently heavily reliant upon a limited number of domain experts available to analyse and determine species groups. This time can be dramatically reduced if the initial identification of *G. salaris* or *G. salaris*-like specimens by the morphology step can be improved and accelerated. In the event of a suspected outbreak, the demand for identification may significant exceed the available supply of suitable expertise and facilities. There is, therefore, a real need for the development of rapid, accurate, semi-automatic / automatic diagnostic tools that are able to confidently identify *G. salaris* in any population of specimens.

The aims of the current study were to explore the potential use of an Active Shape Model (ASM) combining with Complex Network method to extract features information from the attachment hooks of each species of *Gyrodactylus*. Given the small size of the marginal hook sickles (*i.e.* $<$7m), which are regarded as the most taxonomically informative morphological structure, this study will begin with an assessment of scanning electron microscope (SEM) images which give the best quality images. Given the subtle differences in the hook shape of each species, it is hoped that this approach moves towards the rapid automated classification of species with improved rates of correct classification over existing methods and negates the current laborious process of taking manual measurements which are used to assist experts in identifying species.

2 Specimen Preparation

Specimens of *Gyrodactylus* (*G. derjavinoides* n = 25; *G. salaris* n = 34; *G. truttae* n = 9) were removed from their respective salmonid hosts and fixed in 80% ethanol. Subsequently specimens were prepared for scanning electron microscopy (SEM) by transferring individual, distilled water rinsed, specimens onto 13 mm diameter round glass coverslips, where they had their posterior attachment organ excised using a scalpel and the attachment hooks released using a proteinase-K based digestion fluid (*i.e.* 100 μg/ml proteinase K, 75 mM Tris-HCl, pH 8, 10 mM EDTA, 5% SDS). Once the hooks were freed from enclosing tissue, the preparations were flushed with distilled water, air-dried, sputter-coated with gold and then examined and photographed using a JEOL JSM5200 scanning electron microscope operating at an accelerating voltage of 10 kV.

3 Segmentation and Feature Extraction

The application of the ASM method as a segmentation (landmark points) of tool and extracting features using Complex Network approach to the analysis of *Gyrodactylus* attachment hooks is presented in Fig. 2. Specimens of *Gyrodactylus* were picked from the skin and fins of salmonids and their attachment hooks released by proteolytic digestion. Images of the smallest hook structures, the marginal hook sickles which are the key to separating species and typically measure less than 0.007 mm in length, were captured using a scanning electron microscope. The images were pre-processed before being subjected to an Active Shape Model and Complex Network feature extraction step to define 110 landmarks and to fit the model to the training set of hook images. A Complex Network reduced the data to 49 variables which were used to train 4 classifiers (K-NN, LDA, MLP, SVM) and separate the three species of *Gyrodactylus* which includes the notifiable pathogen, *G. salaris*. Abbreviations: K-NN, K Nearest Neighbors; LDA, Linear Discriminant Analysis; MLP, Multi-Layer Perceptron; SVM, Support Vector Machine.

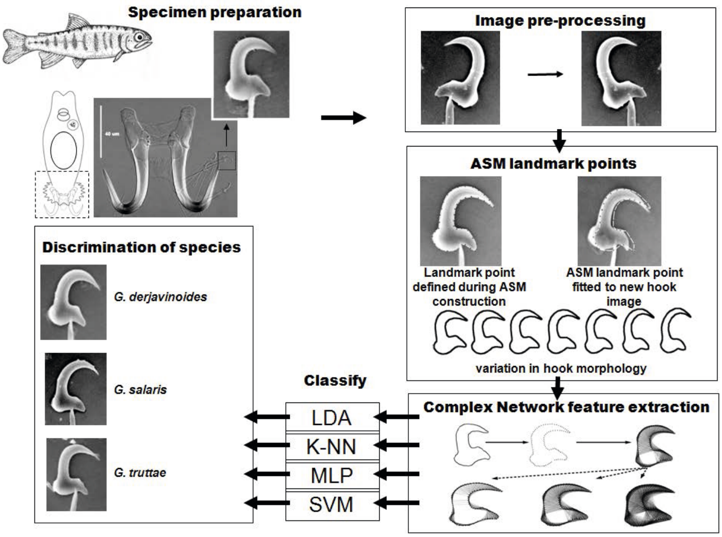

Fig. 1. The methodological approach used in the current study, the ASM were used as landmark points to segment the focus object, while the Complex Network were used to extract the informative features. Four classifiers were accessed and compared in species identification.

3.1 Existing Methods of Parasite Identification and Classification

A number of statistical classification based approaches applied to morphological data [20], [21], and molecular-based techniques targeting specific genomic regions [13], [19], have been developed to discriminate the pathogenic species, *G. salaris*, from other non-pathogenic species of *Gyrodactylus* that co-occur on salmonid hosts. While each technique is able to detect *G. salaris* within a population of specimens and to discriminate it from its congeners with high levels of correct classification, the techniques can be time consuming [20]. If image recognition software could be developed to extract key discriminatory features from the attachment hooks of each species, then it is anticipated that the identification process could be accelerated with equivalent or better rates of correct identification.

3.2 Landmark Points Using ASM

ASM is a feature extraction based technique that has been successfully applied in human face [9] and leaf [12] recognition, the screening of skin cancers [8], and, in the segmentation of lung radiographs [17] and of protozoan parasites from images captured with the light microscope [16], among a range of other studies. The ASM technique permits users to construct a general shape model which is subsequently applied to all images in order to landmark the image area for every given image, providing a pattern that encapsulates the variation seen across the range of shape images. The subsequent ability (classification rate) of the developed model to separate ”image classes” is in part based on the number of images used in the training set - in theory, the greater the number of images that are used in training and constructing the models, the better the classification ability of the resultant model. Given the success of ASM in resolving image-based, shape recognition problems within the biomedical sphere, the current study set out to determine its utility when applied to SEM images of *Gyrodactylus* hooks.

ASM were originally developed for the recognition of landmarks on medical x-rays. Landmark points can be acquired by applying a sample template to a ”problem area”, which appears to represent a better strategy over edge-based detection approaches [18], as any noise or unwanted objects within the image can be ignored in the selection of the shape contour. In the current study, the shape of each attachment hook image is presented by a vector of the position of each landmark, $D = (d_1, e_1, ..., d_n, e_n)$, where $(d_i e_i)$ denotes the 2D image coordinate of the i^{th} landmark point. The shape vector of the hook is then normalised into a common coordinate system. Procrustes analysis is then applied in aligning the training set of images. This aligns each shape so that the sum of distances of each shape to the mean $F = \sum |D_i - \bar{D}|^2$ is minimised. For this purpose, one hook image is selected as an example initial estimate of the mean shape and scaled so that $|\bar{D}| = 1$, which minimises the F.

Assuming s sets of landmark points D_i which are aligned into a common shape pattern for each species, if this distribution can be modelled, then new examples can be generated similar to those in the original training set s, and then these new shapes can be examined to decide whether they represent reasonable examples. In particular, $D = M(b)$ is used to generate new vectors, where b is a vector of parameters of the model. If the distribution parameters can be modelled, $p(b)$, these can then be limited such that the generated D's are similar to those in the training set. Similarly it should be possible to estimate $p(D)$ using the model.

Once the ASM model has been constructed, it is important to fit the defined model to a series of new input images to determine the parameters of the model that are the best descriptors of hook shape. ASM finds the most accurate parameters of the defined model for the new hook images. The ASM fitting attempts to ”best fit” the defined model parameter to each image. Cootes *et al.* [10] explained that by adjusting each model parameter from the defined model will permit an extraction pattern of the image series to be created. During the model fitting, it measures newly introduced images and uses this model to correct the values of current parameters, leading to a better fit.

Once the shape of the images available, then the next step will be the feature extraction. The landmark point need to perform fist, where the SEM images of *Gyrodactylus* specimens contain tissue that difficult to distinguish from actual shape. The Complex Network perform feature extraction using the landmark points of information.

3.3 Extraction Features Using Complex Network

Recently, complex network based shape representation has been shown effectively and widely used in shape and image recognition and retrieval [5,4,6]. In general, this method consists of the following two steps.

(1) Shape representation with complex network model. First, N landmark (key) points should be extracted from the shape contour. Then, with these landmark points, we can construct a complex network $G =< V, E >$ as follows. Each landmark point is represented as a vertex in the network. For each pair of vertices, there is an edge with the corresponding weight w_ij representing the Euclidean distance between them. Therefore, the network can be represented by a $N \times N$ weight matrix W, normalized into interval $[0, 1]$ [5,4].

(2) Feature extraction. There are two main kinds of characteristic (measurements) that can be used to characterize topological connectivity of the complex network. One is the static statistic measurements, and the other is dynamic evolution [5,4]. The five static measurements used in this paper are the maximum degree, average degree, average joint degree, average shortest path length and entropy. Dynamic evolution is also an important characteristic for complex networks. In this paper, we use the evolution process proposed in the work [5,4]. Figure 1 shows the complex network representation and its dynamic evolution process.

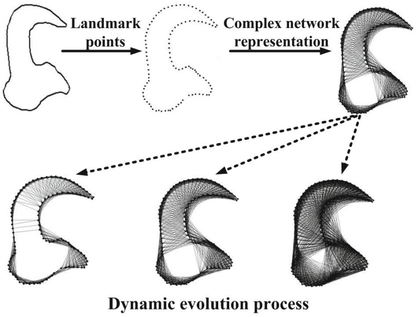

Fig. 2. Shape representation and the dynamic evolution process of complex network

4 Experimental Results

Although the attachment apparatus of *Gyrodactylus* consists of three main elements (*i.e* two larger centrally positioned anchors or hamuli; two connecting bars between the hamuli; and, 16 peripherally distributed marginal hooks), this study sets out to classify species based on features extracted from the sickles of the marginal hooks only. As the study is based on the analysis of biological structures, these require processing subsequent to capture in order to standardised the position and format of the image. Processing to standardise the orientation of the image is applied to reduce processing time and complexity during the training and construction of the ASM model. Then, the data were assessed using four methods of machine learning classifiers, namely are Linear Discriminant Analysia (LDA), K Nearest Neighbor (K-NN), Multi-layer Perceptron (MLP) and Support Vector Machine (SVM). For each approach, a 10-fold cross validation was used *i.e.* the data were divided into k (10) subsets, where k-1 subsets were used for training and the remaining subset used as the test set. This process was repeated 10 times using a different test set on each run and the average classification performance computed.

The K-NN classifier improved upon the classification of *G. salaris* specimens with all being correctly classified (Table 2), while two more species remain miclassified; such as *G. derjavinoides* specimens was misallocated as *G. salaris* and *G. truttae*. Also some of *G. truttae* that has been misclassified as *G. salaris*. Other classifier model LDA (Table 1) and SVM (Table 4) were also expremented. Among these two models, SVM has perform better than LDA, where using SVM classifier, *G. derjavinoides* has managed to have full classification. The MLP classifier, was able to correctly classify all specimens of *Gyrodactylus* to their true class, except for one specimens of *G. salaris* which were classified as *G. truttae* (Table 3). Comaparing to the other models, MLP has achieved highest classification rate at 98.38%. This is not surprising, since MLP is a well performance classifier in many field [11], [15].

Table 1. A confusion matrix showing the classification of *Gyrodactylus* specimen using an LDA classifier

	G. der	*G. sal*	*G. tru*	Sum
G. der	**24**	0	1	25
G. sal	1	**28**	5	34
G. tru	0	2	**7**	9
Sum	25	30	13	68

Table 2. Using the K-NN classifier, *G. salaris* (*G. sal*) is manage to have full classification, while other species remain misclassified

	G. der	*G. sal*	*G. tru*	Sum
G. der	**23**	1	1	25
G. sal	0	**34**	0	34
G. tru	2	1	**6**	9
Sum	25	36	7	68

This achievement is same as performance using ASM-PCA [3], and this performance is better than 25 point-to-point measurements manually extracted from light micrographs of 557 specimens (*i.e.* 92.59%) [2], this approach appears promising and now will be applied to hooks prepared for light microscopy hopefully with equal or better

Table 3. MLP classifer performs well with the correct classification *G. derjavinoides* (*G. der*) and *G. truttae* (*G. tru*)

	G. der	***G. sal***	***G. tru***	**Sum**
G. der	**25**	0	0	25
G. sal	0	**33**	1	34
G. tru	0	0	**9**	9
Sum	25	33	10	68

Table 4. Two specimens (*G. salaris* (*G. sal*) and *G. trutte* (*G. tru*)) are unable to achieve full classification using SVM classifier

	G. der	***G. sal***	***G. tru***	**Sum**
G. der	**25**	0	0	25
G. sal	0	**32**	2	34
G. tru	0	1	**8**	9
Sum	25	33	7	68

rates of correct classification. The ASM and Complex Network based approach applied to SEM images of the hook sickles of *Gyrodactylus* appears to out perform or equal other methods that have been tested to identify and discriminate this species with confidence. This study will continue and will explore the potential of using the ASM and Complex Network method in combination with multi-stage or ensemble classification techniques to improve upon the classification accuracy of each species using image taken with light microscope.

5 Conclusion

The current study set out to explore the utility of a novel ASM and Complex Network based approach in extracting and thus classifying species of *Gyrodactylus* which are ectoparasites of fish. ASM and Complex Network applied to 68 SEM images of the marginal hook sickle was able to overcome the limitation and difficulties in extracting feature information from the hooks. The best approach, which used a MLP method of classification, where only one species remain misclassified.

This work continues, exploring the more pertinent and realistic research problem of classifying specimens based on light microscope images which necessitates image preprocessing. In addition, this work will assess the performance of this method on larger datasets and will explore new methods based on an ensemble of classifiers, which have shown promising results, with the aims of providing a reliable model for the identification of species, including the pathogen *G. salaris*, by non-experts and fish health researchers.

Acknowledgements. Gratefully acknowledges the colloboration work between Stirling University, United Kingdom and Anhui University, China (Sino-UK Higher Education Research Partnership Funding Call). To the team of parasitology, Stirling University, thank you for allowing the use of *Gyrodactylus* dataset in this study.

References

1. Alenäs, I.: Gyrodactylus salaris på lax i svenska vattendrag och laxproblematiken på svenska västkysten. Vann 1, 135–142 (1998) (In Swedish)

2. Ali, R., Hussain, A., Bron, J.E., Shinn, A.P.: Multi-stage classification of Gyrodactylus species using machine learning and feature selection techniques. Intelligent Systems Design and Applications, 457–462 (2011)
3. Ali, R., Hussain, A., Bron, J.E., Shinn, A.P.: The use of asm feature extraction and machine learning for the discrimination of members of the fish ectoparasite genus Gyrodactylus. Neural Information Processing 7666, 457–462 (2011)
4. Backes, A.R., Bruno, O.M.: Shape classification using complex network and multi-scale fractal dimension. Pattern Recognition Letters 31(1), 45–51 (2010)
5. Backes, A.R., Casanova, D., Bruno, O.M.: A complex network-based approach for boundary shape analysis. Pattern Recognition 42(8), 54–67 (2009)
6. Backes, A.R., Martinez, A.S., Bruno, O.M.: Texture analysis using graphs generated by deterministic partially self-avoiding walks. Pattern Recognition 44(8), 1684–1689 (2011)
7. Bakke, T.A., Cable, J., Harris, P.D.: The biology of gyrodactylid monogeneans: the "Russian-doll killers". Advances in Parasitology 64, 161–376 (2007)
8. Blackledge, J.M., Dubovitskiy, A.: Object detection and classification with applications to skin cancer screening. Intelligent Systems 1(1), 34–45 (2008)
9. Choi, J., Chung, Y., Kim, K., Yoo, J.: Face recognition using energy probability in DCT domain, pp. 1549–1552. IEEE (2006)
10. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)
11. Cordella, L.P., Limongiello, A., Sansone, C.: Network intrusion detection by a multi-stage classification system. In: Roli, F., Kittler, J., Windeatt, T. (eds.) MCS 2004. LNCS, vol. 3077, pp. 324–333. Springer, Heidelberg (2004)
12. Du, J.X., Wang, X.F., Zhang, G.J.: Leaf shape based plant species recognition. Applied Mathematics and Computation 185(2), 883–893 (2007)
13. Hansen, H., Bachmann, L., Bakke, T.: Mitochondrial DNA variation of Gyrodactylus spp. (Monogenea, Gyrodactylidae) populations infecting Atlantic salmon, grayling, and rainbow trout in Norway and Sweden. Parasitology 33, 1471–1478 (2003)
14. Harris, P.D., Shinn, A.P., Cable, J., Bakke, T.A.: Nominal species of the genus Gyrodactylus v. Nordmann 1832 (Monogenea: Gyrodactylidae), with a list of principal host species. Systematic Parasitology 59, 1–27 (2004)
15. Kabir, M.F., Schmoldt, D.L., Araman, P.A., Schafer, M.E., Lee, S.M.: Classifying defects in pallet stringers by ultrasonic scanning. Wood and Fiber Science 34 (2003)
16. Lai, C.H., Yu, S.S., Tseng, H.Y., Tsai, M.H.: A protozoan parasite extraction scheme for digital microscope images. Computerized Medical Imaging and Graphics 34, 122–130 (2010)
17. Lee, J.S., Wu, H.H., Yuan, M.Z.: Lung segmentation for chest radiograph by using adaptive active shape models. In: Int. Conf. on Information Assurance and Security, pp. 383–386 (2009)
18. Maini, R., Aggarwal, H.: Study and comparison of various image edge detection techniques. Image Processing 3(1), 1–60 (2009)
19. Meinila, M., Kuusela, J., Zietara, M.: Brief report: Primers for amplifying 820 bp of highly polymorphic mitochondrial COI gene of Gyrodactylus salaris. Hereditas 137, 72–74 (2002)
20. Shinn, A.P., Collins, C., García-Vásquez, A., Snow, M., Paladini, G., Lindenstrøm, T., Longshaw, M., Matějusová, I., Stone, D.M., Turnbull, J.F., Picon-Camacho, S.M., Vázquez Rivera, C., Duguid, R.A., Mo, T.A., Hansen, H., Olstad, K., Cable, J., Harris, P.D., Kerr, R., Graham, D., Yoon, G.H., Buchmann, K., Raynard, R., Irving, S., Bron, J.E.: Multi-centre testing and validation of current protocols for Gyrodactylus salaris (Monogenea) identification. International Journal of Parasitology 40, 1455–1467 (2010)
21. Shinn, A.P., Hansen, H., Bachmann, L., Bakke, T.A.: The use of morphometric characters to discriminate specimens of laboratory-reared and wild populations of Gyrodactylus salaris and G. thymalli (monogenea). Folia Parasitologica 51, 239–252 (2004)