Diagnosis of Fish Diseases Using Artificial Neural Networks

J.N.S. Lopes[1], A.N.A. Gonçalves[2], R.Y. Fujimoto[1,3] and J.C.C. Carvalho[3]

**[1] Programa de Pós-graduação em Biologia Ambiental, Universidade Federal do Pará
Bragança, Pará, Brazil**

**[2] Programa de Pós-graduação em Genética e Biologia Molecular – com Ênfase em Bioinformática, Universidade Federal do Pará
Belém, Pará, Brazil**

**[3] Docente da Faculdade de Engenharia de Pesca, Universidade Federal do Pará
Bragança, Pará, Brazil**

Abstract

Artificial neural networks (ANNs) are computational intelligence techniques, which are used in many applications, such as disease diagnosis. The objective of this study was to evaluate two artificial neural networks created for the diagnosis of diseases in fish caused by protozoa and bacteria. As a classification system, ANNs are an important tool for decision-making in disease diagnosis. A back-propagation feed-forward was selected, with two layers, sigmoid and linear activation functions, and the Levenberg -Marquardat algorithm, for the training of the ANNs. The results of the application of these neural networks for the diagnosis of fish diseases based on test cases indicated a 97% success rate for the classification of both bacterial and protozoan diseases.

Keywords: *Artificial Neural Networks, Fish Disease Diagnosis, Feed-forward back-propagation network, Artificial Intelligence, and Decision Support Systems.*

1. Introduction

Artificial Neural Networks (ANNs) are adaptive models inspired by the organization of neurons in the human brain and learning of novel patterns from an initial detail [1]. These networks have been used successfully in areas ranging from engineering to medicine [2], [3], [4], [5], [6], [7]. In the medical field, ANNs are used to assist the specialist in the analysis, diagnosis and treatment of diseases. Most medical applications of artificial neural networks are classification problems, in other words, the task is based on the classification of measured parameters, to assign the patient to a small set of classes, which are the diseases [8], [9]. The back-propagation model with two layers (hidden and output), with the sigmoid and linear activation functions, has been used for the solution of some of these problems [6], [9].

In contrast with the field of human medicine, few studies based on neural networks have been developed for the diagnosis of diseases in fish. This is partly because fish diseases are complex phenomena, the diagnosis of which demands considerable expertise, but also because infected fish tend to die quickly without adequate treatment [10]. Economically, the most important diseases include those caused by bacteria and protozoa. These diseases are especially difficult to identify because their clinical signs are similar, and differences may only arise during the acute or chronic phase, and in many cases, transmission patterns are unknown [11], [12]. Given this, there is a clear need for the development of new and more effective approaches to the diagnosis of bacterial and protozoan diseases in these animals.

Zeldis and Prescott [13] discussed problems and solutions for the development of a program for the diagnosis of diseases in fishes. They reviewed the different techniques employed by the experts in the field, emphasizing the considerable difficulties of diagnosing fish diseases, but concluded that the use of artificial neural networks would not be feasible due to the lack of an adequate database of fish diseases.

However, while no central agency store these data, a number of university laboratories have accumulated a large quantity of data, which permits a more systematic evaluation of the phenomenon. In the present study, data from such a source are used to evaluate the applicability of ANNs to the diagnosis of fish diseases caused by bacteria and protozoa. The aim is to provide a reliable system for the rapid and accurate diagnosis of diseases in fishes.

2. Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are computational systems that simulate biological neural networks, which can also be defined as a specific type of parallel processing system, based on distributional or connectionist methods [14]. Their internal structure can also be modified in accordance with a specific function [1]. The structure of a network of this type is characterized by a number of interconnected elements (neurons) that learn by modifying themselves. As in nature, the function of the network is determined by the connections between the elements [2].

In this configuration of neural networks, a subset of processing elements can be added to the network (layers). This configuration is referred to as a neural network multi-layer perceptron (MLP-ANN). This MLP-ANN is widely- used in applications such as approximation functions, feature extraction, optimization, classification and ease-of- use [15].

In this configuration, the first layer is the input layer and the last, the output layer, between which there may be one or more extra layers, known as hidden layers (see Figure 1). Within the context of a given learning algorithm, this configuration enables neural networks to achieve a specific function, as well as allowing adjustments in the value of the connections (weights) between elements [9], reducing the mean square error.

The back-propagation approach and its variants are widely used as learning algorithms in neural networks. The procedure is based on the calculation of the gradient vector error, with the error gradually decreasing until all the expected results are displayed [2].

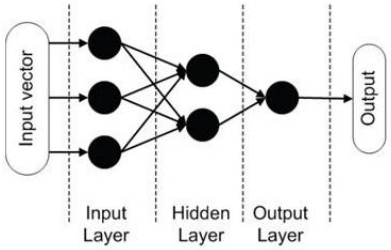

Figure 1 Structure of the multi-layer neural network perceptron.

SOURCE: Krenker, Bester and Kos, 2011 [16].

Artificial neural networks can be divided into two categories - supervised and the unsupervised – based on the learning process. In supervised learning, inputs and outputs are presented to the network, which will adopt the patterns that provide the desired outputs [17]. In order to ensure that the output (system response) achieves a satisfactory result, the neural network adjusts the relative weights of the connections, using an interactive process. Following unsupervised learning, the network develops its own representation of the input stimuli in order to calculate the weights of acceptable connections until finding the answer to the problem. This type of network creates a map of self-organization, which has only inputs and no known responses. An ANN thus becomes a powerful and versatile tool, due to its considerable capacity for learning and, theoretically, that it can provide continuous mapping of any database with arbitrary accuracy [9].

3. The Proposed Method Based on Neural Network

Two back-propagation feed-forward neural networks were constructed, one for bacterial and the other for protozoan diseases. The neural networks were derived from data sets provided by the Ichthyoparasitology and Fisheries Laboratory at the Federal University of Pará (Brazil), which provided information on the clinical signs of diseased fishes and their diagnosis.

The data of the network were divided into inputs and outputs. The input data were the clinical signs, with the presence of signs being scored as 1 (present) or 0 (absent). The output data for each disease group is the diagnosis of the disease.

The network for the diagnosis of bacterial diseases was composed of 43 inputs, 20 neurons in the hidden layer and 12 neurons in the output layer (Figure 2), while that for protozoan diseases had 28 inputs, 22 neurons in the hidden layer and eight neurons in output layer (Figure 3). The structures proposed for the neural networks are presented in Figures 2 e 3.

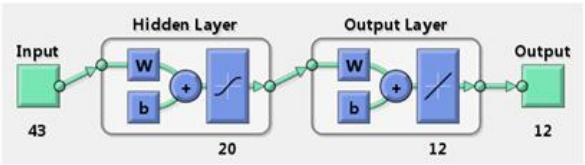

Figure 2 The proposed neural network for the diagnosis of bacterial diseases.

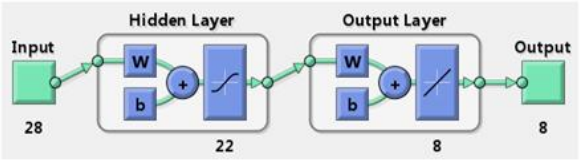

Figure 3 The proposed neural network for the diagnosis of protozoan diseases.

3.1 Feed-forward architecture

This feed-forward neural network model was selected for this project because this approach has been used successfully in other contexts for classification, prediction and troubleshooting. In this network model, information moves in only one direction, always forward from the input nodes, spreading to the hidden nodes and the on to the output nodes, where the output is compared with the desired value, resulting in an error for each element of the output [17].

In this system of feed-forward neural network, the hidden "neurons" are able to learn data patterns during the training phase and are then able to map the relationship between input/output pairs. In the hidden layer, each neuron uses a transfer function to process the data it receives from the input layer and then transfers this processed information on to the neurons of the output layer. The output of the hidden layer can be represented by the following function:

$$YNx1 = f(WNxM\ XM, 1 + bN, 1) \quad (1)$$

where Y is a vector containing the output of each neuron (N) in a layer, W is a matrix containing the weights of each input (M) for all the neurons, X is a vector containing the inputs, b is a vector containing the bias, and f (.) is the activation function [18].

In these types of network, data input and output are automatically divided into training, validation, and test sets. The training data are used for network learning. Training upholds the parameters set in the network structure, and then a set of validation data is used to minimize overfitting. These validation data are used to check the increase in accuracy in comparison with the training data, and are not shown in the final network. If accuracy increases during training, but then validation remains constant or decreases, network training is terminated.

3.2 Fish Disease Diagnosis Data (Protozoan diseases)

A database provided by a specialist in fish diseases was transcribed into the form used for the construction, validation, and testing of the network. The data were then analyzed for the diagnosis of potential diseases caused by protozoa pathogens.

Thirty records of diagnosed diseases caused by protozoa were used, based on the set of clinical signs shown in table 1. Of the 30 samples in the data set, 80% were used to train the neural network, while the remaining 20% were used in the test network.

3.3 Fish Disease Diagnosis Data (Bacterial diseases)

The data on the diseases caused by bacteria were also converted into a form appropriate for analysis in the neural network. Thirty-one records were analyzed based on the clinical signs outlined in table 2. As for the previous procedure, 80% of the samples were used to train the neural network, while the other 20% were used in the test network.

3.4 Performance Evaluation

The tool used for the construction, running and evaluation of the proposed neural networks was Matlab Toolbox 7.10. In both networks, the feed-forward model with sigmoid activation function in the hidden layer and a linear output layer was adopted because it is the most frequently-used procedure for function fitting (or nonlinear regression) problems. The Levenberg-Manquardt training algorithm was also used for the back-propagation network. This type of algorithm is faster for standard and feed-forward networks, and performs better for function fitting (nonlinear regression) than for pattern recognition problems. The network produced in this study can be expected to successfully diagnose eight types of disease caused by protozoa, and 12 by bacteria (Table 3).

4. Results of the Experiment

The result obtained from the artificial neural network approach to the diagnosis of diseases, based on reported clinical signs, demonstrated that the network was able to learn the patterns corresponding to the clinical signs of specific fish diseases. The networks were also subjected to the respective test sets (unknown cases), which again produced satisfactory results, as described below (tables 3 and 4).

4.1 Artificial Neural network 1: Protozoan diseases

The network classified 97% of the cases in the protozoan test set. The validation vectors used to stop the training network at the point set by the training algorithm are shown in figure 4. Validation ceased when the GRADIENT performance decreased, the performance adaptive variable (MU) was reduced, and the validation performance (VAL FAIL) increased.

The best performance validation score (0.01088) was recorded at time 4 (figure 5). The mean square error (MSE) is the mean square of the differences between actual and desired outputs. Lower values indicate better performance, and zero is equal to no error. The validation and test curves were very similar. The percentage accuracy in the sample simulation of the feed-

forward back-propagation network was 97%. Overall MSE was 5.44087e-3 and regression (R) was 9.83867e.-1.

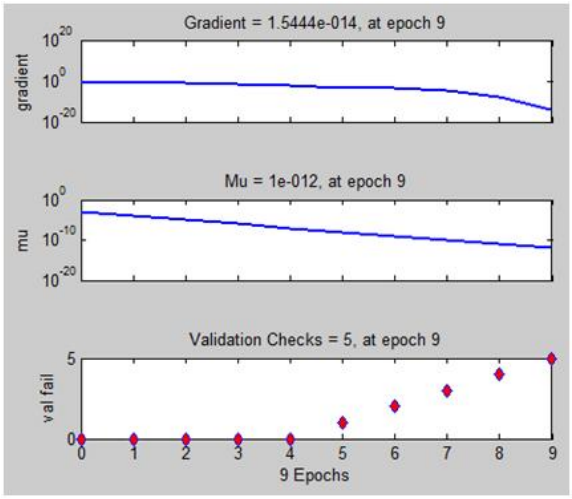

Figure 4: Training state values

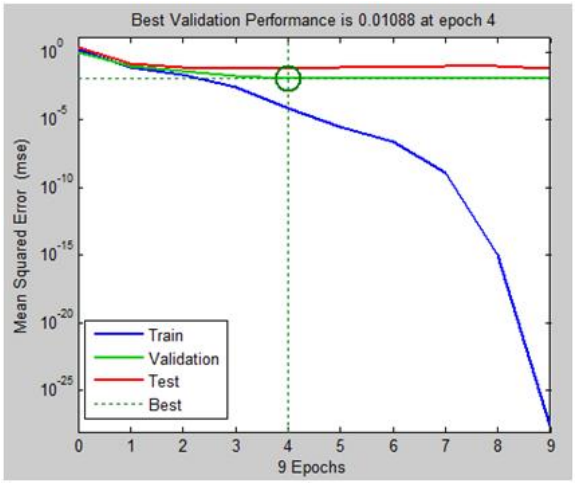

Figure 5: Network error values plot.

Table 3: The Mean Square Error (MSE) and regression values (R) for training, validation and testing.

	MSE	R
Training	7.08966e-5	9.99771e-1
Validation	1.08804e-2	9.99501e-1
Testing	6.44062e-2	9.97507e-1

4.2 Artificial Neural network 2: Bacterial diseases

The second network also classified 97% of the cases in the bacterial test set. The validation vectors used to stop the training network at the point set by training algorithm are shown in Figure 6. Once again, validation ceased when the GRADIENT performance decreased, the performance adaptive variable (MU) was reduced, and the validation performance (VAL FAIL) increased. The best performance validation score (0.056193) was recorded at epoch 7 (Figure 7). The percentage accuracy in the sample simulation of the feed-forward back-propagation network 97%, MSE was 2.28988e-2 and R was 9.54099e-1.

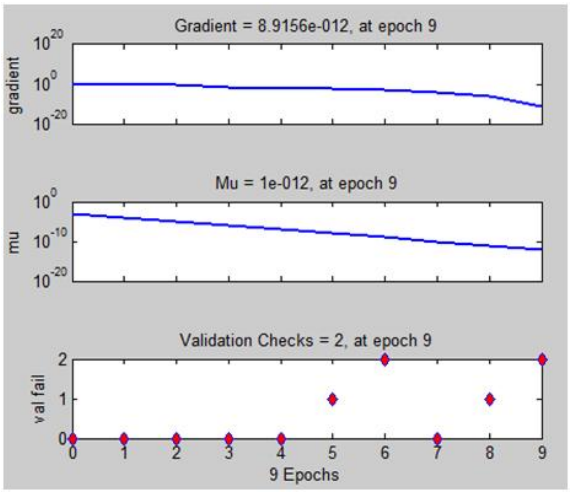

Figure 6: Training state values

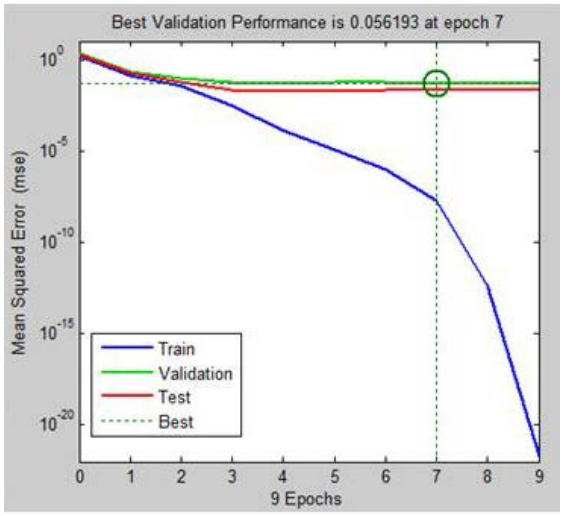

Figure 7: Network error values plot.

Table 4: The Mean Square Error (MSE) and regression values (R) for training, validation and testing.

	MSE	R
Training	1.92955e-8	9.99999e-1
Validation	5.61926e-2	9.14124e-1
Testing	7.65403e-3	9.65968e-1

5. Conclusions and Future Work

This article presents the construction and testing of two feed-forward back-propagation neural networks for the diagnosis of protozoan and bacterial diseases in fishes. The artificial neural networks had satisfactory outcomes for both data sets. The results indicate that artificial neural networks provide a viable approach for the diagnosis of diseases in fish, and may be further enhanced to aid in the treatment of these diseases, as well as the diagnosis of diseases caused by other vectors, such as parasites or fungi, and well as disorders related to environmental changes.

Acknowledgments

"The author J. N. S. Lopes thanks Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-PROCAD) for financial support".

References

[1] E. Grossi, "Artificial Neural Networks and Predictive Medicine: a Revolutionary Paradigm Shift" - Artificial Neural Networks - Methodological Advances and Biomedical Applications, India: INTECHWEB.ORG, 2011.
[2] G. Caocci, R. Baccoli and G. L. Nasa, "The Usefulness of Artificial Neural Networks in Predicting the Outcome of Hematopoietic Stem Cell Transplantation" - Artificial Neural Networks - Methodological Advances and Biomedical Applications, India: INTECHWEB.ORG, 2011.
[3] G. Zini, "Artificial intelligence in hematology". Hematology, Vol. 10, No 5, 2005, pp. 393-400.
[4] M. Carter, Minds and computers: An introduction to the philosophy of artificial intelligence, Edinburgh: Edinburgh University Press, ISBN 9780748620999, 2007.
[5] M. Suka, S. Oeda, T. Ichimura, K. Yoshida, and J. Takezawa, "Neural Networks Applied to Medical Data for Prediction of Patient Outcome", Trends in Intelligent Systems and Computer Engineering. O. Castillo, L. Xu, and Sio-Iong Ao. Vol. 6, Springer. ISBN 978-0-387-74934-1, 2008.
[6] Q. K. Ai-Shayea, and I. S. H. Bahia, "Urinay System Diseases Diagnosis Using Artificial Neural Networks", International Journal of Computer Science and Network Security (IJCSNS), vol. 10, Nº 7, 2010, pp. 118-122.
[7] R. Linder, I. R. König, C. Weimar, H. C. Diener, S. J. Pöppl, A. Ziegler, "Two models for outcome prediction - a comparison of logistic regression and neural networks", Methods of Information in Medicine, Vol. 45, No. 5, 2006, pp. 536-540.
[8] R. Dybowski, and V. Gant, Clinical Applications of Artificial Neural Networks, Cambridge University Press, 2007.
[9] Q. K. Ai-Shayea, "Artificial Neural Networks in Medical Diagnosis", International Journal of Computer Science Issues (IJCSI), Vol. 8, No 2, 2011, pp. 150-154.
[10] D. Li, Z. Fu and Y. Duan, "Fish-Expert: a web-based expert system for fish diseases diagnosis", Expert Systems with Applications, Vol. 23, 2002, pp. 311-320.
[11] G.C. Pavanelli, J.C. Eiras and R.M. Takemoto, Doenças de Peixes – profilaxia, diagnóstico e tratamento, Maringá: Eduem, 2008.
[12] M.P. Georgiadis, I.A. Gardner, R.P. Hedrick, "The role of epidemiology in the prevention, diagnosis, and control of infectious diseases of fish", Preventive Veterinary Medicine, Vol. 48, 2001, pp.287-302.
[13] D. Zeldis, and S. Prescott, "Fish disease diagnosis program – problems and some solutions", Aquacultural Engineering, Vol. 23, 2000, pp. 3-11.
[14] A. A. Hopgood, Intelligent systems for engineers and scientists, Florida: CRC Press, 2000.
[15] C. M. Bishop, Pattern Recognition and Machine Learning, Cambridge: Springer, 2006.
[16] A. Krenker, J. Bester and A. Kos, "Introduction to the Artificial Neural Networks" - Artificial Neural Networks - Methodological Advances and Biomedical Applications, India: INTECHWEB.ORG, 2011.
[17] S. Sumathi and S. Paneerselvam, Computational intelligence paradigms: theory & applications using MATLAB, CRC Press, 2010.
[18] J. A. Freeman, and D. M. Skapura, Neural networks: algorithms, applications and programming techniques, Addison-Wesley, 1991.

J. N.S. Lopes received her undergraduate degree in Information Systems from the Amazon Higher Education Institute (IESAM), Brazil, in 2009, and will receive her masters degree in Environmental Biology from the Federal University of Pará (UFPA), Brazil, in 2011. Her current research interests include neural networks and fuzzy logic.

A. N. A. Gonçalves, received his undergraduate degree in Computer Engineering from IESAM in 2009 and will receive his masters in Genetic and Molecular Biology with emphasis on Bioinformatics, from UFPA in 2012. He is interested in Artificial Intelligence, programming languages, and databases and algorithms for Bioinformatics.

R. Y. Fujimoto received his undergraduate degree (1998) and masters (2001) in Animal Sciences, and doctorate (2004) in aquaculture from the Julio de Mesquita Filho State University of Sao Paulo. He is currently an adjunct professor at UFPA, where he coordinates the fish farming and ichthyoparasitology laboratory. He has research experience in fish diseases.

J. C. C. Carvalho, received his undergraduate degree (1996), masters (1999) and doctorate (2006) in Electrical Engineering from UFPA. He is currently associate professor of Information Systems I at the UFPA - Bragança campus.

Table 1: Clinical Signs variables used to analyze the data set of protozoan diseases.

Clinical sign of disease	
Nº	**Diagnostic Variable**
1	Abscess {Yes, no}
2	Anorexia {Yes, no}
3	Apathy {Yes, no}
4	Ascites {Yes, no}
5	Cotton-like appearance {Yes, no}
6	Gills with excess mucus {Yes, no}
7	Gills with blood {Yes, no}
8	Blindness {Yes, no}
9	Pale coloration {Yes, no}
10	Dark coloration {Yes, no}
11	Dyspnea {Yes, no}
12	Swimming disorders {Yes, no}
13	Exophthalmos {Yes, no}
14	Injuries to the body {Yes, no}
15	Differentiated feces {Yes, no}
16	Hypertrophy of organs {Yes, no}
17	Bleeding in external organs{Yes, no}
18	Organs with lesions {Yes, no}
19	Ulcerative lesions {Yes, no}
20	White blemishes{Yes, no}
21	Disjointed movements {Yes, no}
22	Fins destroyed {Yes, no}
23	Fins closed {Yes, no}
24	Nodules {Yes, no}
25	White spots {Yes, no}
26	Skin mucus{Yes, no}
27	Rash{Yes, no}
28	Abnormal tegument {Yes, no}

Table 2: Clinical Signs variables used to analyze the data set of bacterial diseases.

Clinical sign of disease	
Nº	**Diagnostic Variable**
1	Abscess {Yes, no}
2	Anorexia {Yes, no}
3	Apathy {Yes, no}
4	Ascites {Yes, no}
5	Swollen anus {Yes, no}
6	Hemorrhagic anus {Yes, no}
7	Hemorrhagic areas{Yes, no}
8	Gills affected {Yes, no}
9	Gills pale {Yes, no}
10	Blindness {Yes, no}
11	Red coloration {Yes, no}
12	Pale coloration {Yes, no}
13	Dark coloration {Yes, no}
14	Abnormal growth {Yes, no}
15	Dyspnea {Yes, no}
16	Fin disorders {Yes, no}
17	Edema {Yes, no}
18	Abnormal scales {Yes, no}
19	Exophthalmos {Yes, no}
20	Furuncle {Yes, no}
21	Hypertrophy in organs {Yes, no}
22	Bleeding {Yes, no}
23	Bleeding in the eyes {Yes, no}
24	Bleeding in the external organs {Yes, no}
25	Bleeding in the internal organs {Yes, no}
26	Minor hemorrhage {Yes, no}
27	White lesions {Yes, no}
28	Dark lesions {Yes, no}
29	Hemorrhagic lesions {Yes, no}
30	Lesions in organs {Yes, no}
31	Minor lesions {Yes, no}
32	Ulcerative lesions {Yes, no}
33	White blemishes{Yes, no}
34	Intense bruising {Yes, no}
35	Membrane surrounding the organs {Yes, no}
36	Fins destroyed {Yes, no}
37	Fins closed {Yes, no}
38	Nodules {Yes, no}
39	Red spots {Yes, no}
40	Delay of sexual maturation{Yes, no}
41	Ulcers {Yes, no}
42	Abnormal tegument {Yes, no}
43	Disease spreading in hours {Yes, no}

Table 3: Types of disease diagnosed for the two groups.

Group	Type of disease
Bacterial	Mycobacteriosis
Bacterial	*Streptococcus* infection
Bacterial	Peduncle disease
Bacterial	"Spinal column" disease
Bacterial	Bacterial gill disease
Bacterial	Septicemia provoked by *Edwardsiella*
Bacterial	Red mouth disease
Bacterial	Furunculosis
Bacterial	Septicemia caused by mobile *Aeromonas*
Bacterial	Septicemia caused by *Pseudomonas*
Bacterial	Bacterial kidney disease
Bacterial	Pseudo-renal disease
Protozoan	Velvet disease
Protozoan	Ichthyobodosis
Protozoan	Disease caused by rhizopods (amoebae)
Protozoan	Disease caused by flagelates (*Hexamita* spp.)
Protozoan	Disease caused by ciliates I (*Trichodina* spp.)
Protozoan	Disease caused by ciliates II (*Chilodonella* spp.)
Protozoan	Disease caused by ciliates III (*Sessilina*)
Protozoan	White spot disease