Detection Of Freshwater Fish White Spots Disease Using The Machine Learning LR Classifier And ACO

Ammar Adl
Faculty of Computers and Information Systems, Beni-Suef University
ammar@fcis.bsu.edu.eg

Neama Younan
Faculty of Computers and Information Systems, Beni-Suef University
Beni-Suef, Egypt
neamayounan_sd@fcis.bsu.edu.eg

***Abstract*— Fish disease diagnosis is a difficult process and needs high level of expertise to identify it, so we proposed a fish diseases diagnose system with high accuracy by using Image Processing, Machine Learning and Swarm Optimization to recognize and identify the Ich (ichthyophthiriasis or white spots) disease which is caused by Ichthyophthirius multifiliis through microscopic slides for the Ichthyophthirius multifiliis, through three steps, first preprocessing input image like Removing the noise and background .After that extracted the features from microscopic images by using ORB (Oriented Fast and rotated Brief) .Finally, classify the Ich infected and Non-Ich infected fish image through Logistic Regression Machine Learning Algorithm combined to Ant Colony Optimization (ACO) and find the classification accuracy through it, the proposed solution achieved accuracy of 92.8%. The Experimentation has been done by PYTHON environment on microscopic samples for the causative.**

Keywords— white spots, Machine Learning, LR Classifier, ORB, ACO.

I. INTRODUCTION

Fish disease is a serious problem because of its rapid spread through the water to neighboring aqua-farms so rapid and accurate diagnosis is required to control the disease and to prevent the spread of that disease [1], [2]. Fish disease diagnoses were made by using the accumulated experiences of fish-farmer and veterinarian, but this is a traditional way therefore the accuracy of the final diagnosis ultimately depends on the skill and experience of the person and the time spent studying each disease. In order to solve this problem, computer science methodologies such as machine learning, image processing techniques such as feature extraction techniques have been used to help to diagnose fish diseases. The goal of this research is to detect and diagnose one of the most harmful diseases. It is ichthyophthiriasis or white spots (Ich). Ich is a protozoan disease, often called 'white spots disease' .Its scientific name is Ichthyophthiriasis and the

causative agent is Ichthyophthirius multifiliis. It is wide spread in all freshwater fish. Mature parasites are round or oval in shape and measure up to 1.5(mm) across. The surface of the parasite's body is covered in small hair-like structures known as cilia. When viewed under the microscope, a characteristic horseshoe-shaped nucleus may be seen. It is very damaging to the gills and skin [3], [4]. Infected fish have small white spots on the skin and gills and produce excess mucus, due to irritation [5]. It is still misidentified by the people. To identify the features of Ich disease infected fish and find the region of interest for the accurate and fast diagnose, so in this paper, we proposed an Expert System to identify the features of Ich disease and can detect it, that System had been implemented in three steps .First, pre-processing techniques like enhancement (remove noise and remove background using GrabCut algorithm), it has been done on various microscope samples to get the Ichthyophthirius multifiliis parasite only. After the first part of the paper, features are extracted through the ORB feature extraction technique to get values of the features of that disease (Ich) then classify these features by using Logistic Regression Classifier combined to Ant Colony Optimization (ACO) [6] diseased and non-diseased fish, classifier has been used to find the accuracy, The proposed solution achieved accuracy of 92.8%.

II. RELATED WORK

There are some researchers who worked on this problem to solve it with different ways. We will discuss some of these ways. We started our research with the previous work of other researchers like Lyubchenko, V., Matarneh, R. and Koblyin, O., in their paper (Digital Image Processing Techniques For Detection And Diagnosis Of Fish Diseases) [7] they said that the whole image processing methodology can be summarized in the following steps, 1: selection of markers for individual objects. 2: detection of objects in the image. 3: calculating the proportion of an object in the image. 4: calculation the proportion of infected area to the fish area. Also Shaveta Malik, Tapas Kumar and A.K Sahoo used the Machine Learning and HOG feature extraction in their paper (A Novel Approach to Fish Disease Diagnostic System based on Machine Learning) [8] to identify Epizootic Ulcerative syndrome disease and the accuracy was 87.0% . But we combined the Image Processing with Machine Learning and Swarm optimization to get highest accuracy and we have already achieved it (92.8%)

III. MATERIALS AND METHODS

The monitoring and examination of fish disease were found manually by more searches in multiple media sources and engines. This needs an immense amount of work, effort and redundant Processing time. The system that we developed is an automated system for recognizing fish diseases based on microscopic images. In this paper, we focus on white spots disease on freshwater fish to detect this disease from it. Firstly, applying Gaussian filter on images to reduce noise, then we separated background to get object only (Ichthyophthirius multifiliis parasite) by GrabCut algorithm after that we used ORB algorithm in feature extraction which extracted some features, we extracted 4.0 features

for Ich disease to make the system learn which is infected or not, then we used classification techniques for training and testing the data and we got high accuracy which is 92.8% with Logistic Regression Classifier. The result will help fish farms to know the infected fish earlier. Where if the automated detection technique is utilized it will take fewer efforts, less time and more accurately so that idea is important for the fish farms to detect the white spots disease easily and accurately.

- STEPS OF THE PROPOSED METHODOLOGY:

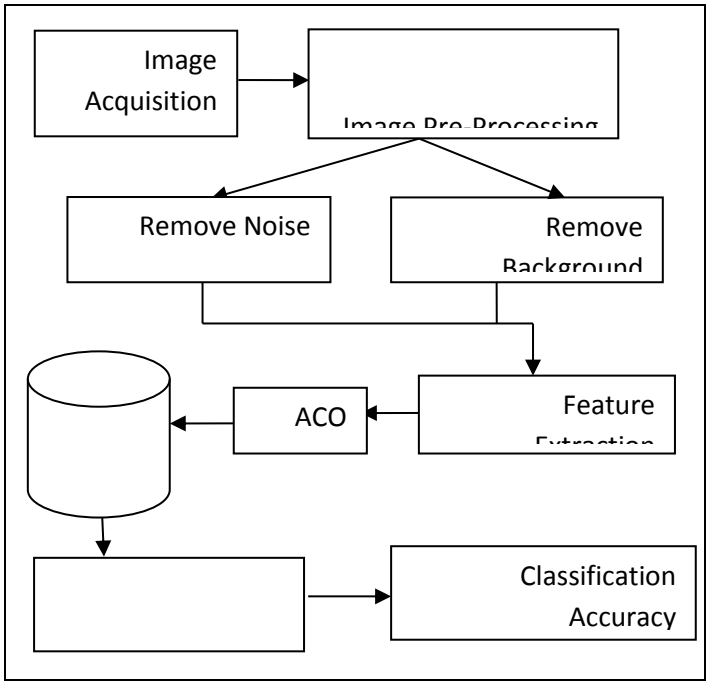


Fig. 1. Workflow of the proposed system that shows the total processes of that system in steps.

A. Image acquisition:

The first stage and process of storage a digitalization image. Here the Ichthyophthirius multifiliis microscopic image. Acquire RGB colors for the image. With JPG or PNG formatted to save the image. This is the example of two images that we work on.

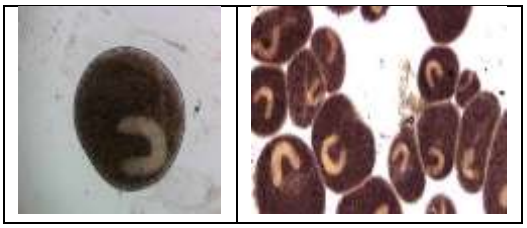


Fig. 2. Ichthyophthirius multifiliis parasite under the microscope.

B. Image Pre-processing:

First, Remove Noise: This stage is used to improve the quality of the image and removing the noise from it by using the Gaussian filter [9]. Convert the RGB image to gray level. Second, Remove Background: to get the Ichthyophthirius multifiliis parasite only to improve the quality to extract features accurately only for the parasite. We used the GrabCut algorithm [10] to help us to remove the background.

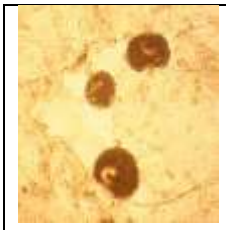
		
Original image	After Removing Background	After Applying Gaussian Filter

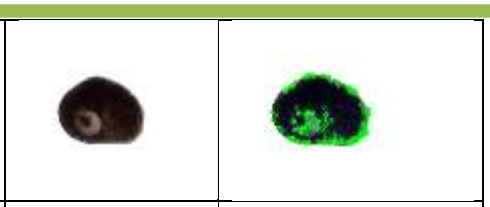
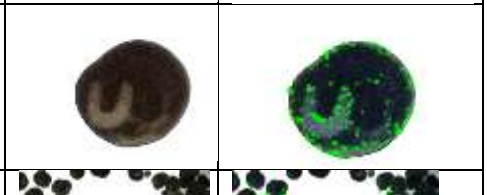
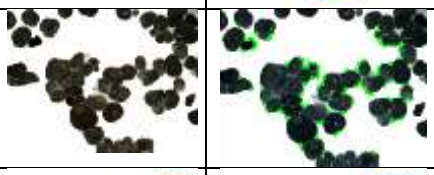
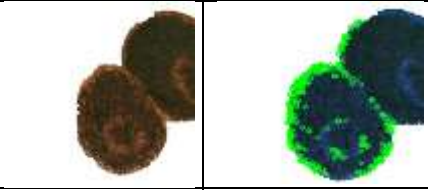
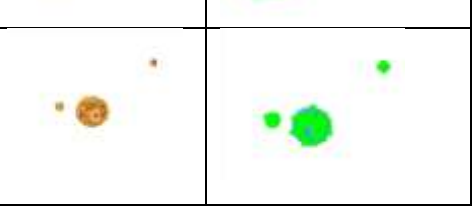
Fig. 3. Samples of Ichthyophthirius multifiliis parasite separated by using GrabCut Algorithm.

C. Features Extraction with ACO:

We usually used feature extraction phase to identify the important features that describe the disease. So we decided to use ORB feature extraction technique. ORB technique is better than SIFT, Fast and HOG feature extraction [11], [12], [13], [14]. We evaluated features for the Ichthyophthirius multifiliis parasite and got four features (Mean, Median, Standard Deviation and Variance), then we used the ACO to select the optimum features of the given dataset, it is used for reducing non-useful information from the dataset and it is frequently used in both image processing and machine learning technique. It chooses the best principle component for training data for each class or chooses first few components so that the transformed data are reduced [15], [16]. This paper presents an improvement over the standard Ant Colony Optimization which is a wrapper based approach to improve the Classification accuracy with reduced features of the data after extracting the features through ORB (Oriented FAST and Rotated BRIEF).

TABLE I. SHOWS THE FEATURES OF THE DISEASE WHICH EXTRACTED BY USING ORB TECHNIQUE.

	Before Applying ORB	After Applying ORB

1.		
2.		
3.		
4.		
5.		
6.		

The following table represents the database that contains the extracted features by ORB technique that features are extracted from the result of preprocessing step which is for infected and healthy samples.

TABLE II. REPRESENTS SAMPLES OF EXTRACTED FEATURES USING ORB WHERE LABEL (1) FOR THE INFECTED FISH AND LABEL (0) FOR THE HEALTHY.

Label	Main	Median	Variance	STD
1	156.26803	164.0	5568.82432	74.6245579
1	134.68974	133.0	5768.9315	75.9538254
1	158.61764	170.0	5454.30016	73.8533394
1	146.16359	147.0	5521.94069	74.3096179
0	144.51203	149.0	5414.12228	73.587195
0	149.0134	156.0	5695.71212	75.4694185
0	145.56168	150.0	5682.20694	75.381486

D. Training and Classification:

Train the machine to make it be able to detect or classify unknown input image for that disease, so we constructed our dataset for two classes and divided it into two parts. About 80.0% of the dataset that has been used in the training stage to know which the infected and non-infected images and the remained 20.0% of the dataset has been used in the testing stage to test the efficiency and the performance of the LR model in recognition that disease. After we extracted the features, we have used these features to make the training and classification between these microscopic images that are infected or not by using Logistic Regression Classifier [17], [18]. Images are classified by comparing to the features that are extracted from the enhanced images and saved into CSV file.

IV. RESULTS

In this paper we cover the way that is faster in time to detect the Ich (white spots diseases) which got 92.8% accurately. So enhanced the image by removing the noise using Gaussian filter and got foreground by using GrabCut algorithm, then used (ORB) feature extraction algorithm which is faster to extract the important features from the image. When we compared the six different algorithms models in classification we got that "LogisticRegressionClassifer" is the best model that achieved 92.8% accuracy and the second model is Gaussian NB which achieved 85%. Finally we decided to use" Logistic Regression Model" which is the best model that scored the highest accuracy [19].

TABLE III. REPRESENTS THE PERCENTAGES OF THE SIX TESTED MODELS.

Model	Accuracy
LR	92.8%
KNN	73%
Decision Tree	75%
SVM	83%
NB	77%
Gaussian NB	85%

A. Algorithms Comparison:

According to these results we found that Logistic Regression model is the best in classification so that we used it on our database.

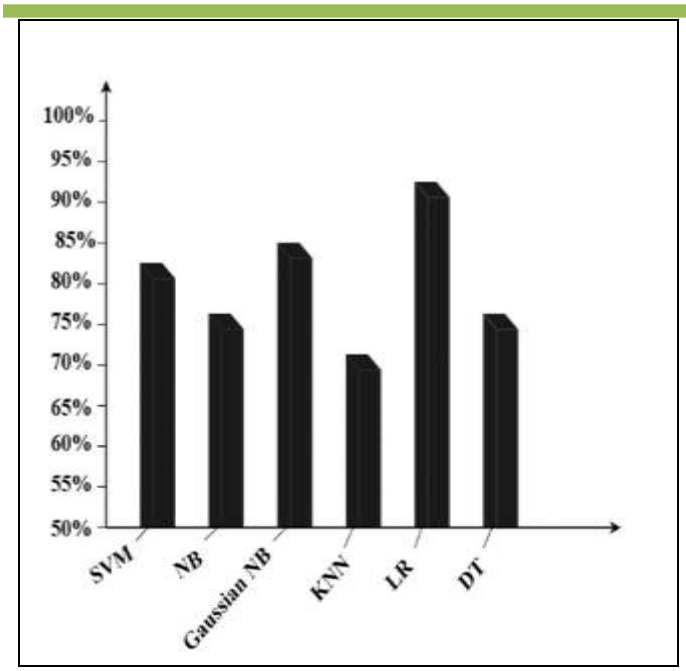


Fig. 4. Shows the results of the accuracy comparison among the algorithms (LR, Gaussian NB, KNN, CART, NB and SVM).

The Confusion Matrix is a handy presentation for the final accuracy of LR model [20]. The algorithm recorded a final score 92.8% in the next confusion matrix, the final result will be clear and understood.

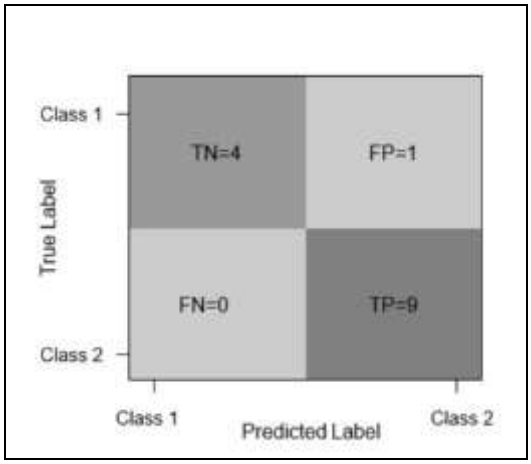


Fig. 5. The Confusing matrix that illustrates the final accuracy of the trained algorithm.

B. Classification Report:

After we have trained and fitted our machine learning classification model, we should have evaluated the model's performance in the condition of unbalanced classes. Therefor

Classification Report is the best choice and more preferred in that situation to show the model accuracy in detailed report [19].

TABLE IV.CLASSIFICATION REPORT THAT SHOWS THE FINAL ACCURACY OF THE LR MODEL.

Class	Precision	Recall	F1_Score	Support
Infected	0.90	1.00	0.95	9.0
Non_infected	1.00	0.80	0.89	5.0
Avg/total	0.94	0.93	0.93	14.0

V. DISCUSSION

We collected the dataset from multiple media sources and engines. It contains 200.0 images which are classified in 113.0 infected images and 87.0 healthy images. All this images are captured from real Labs under the microscope. We removed the noise by Gaussian filter and used the GrabCut algorithm to remove background, then we extracted features by using ORB feature extraction technique and selected the efficient features by ACO finally, we used the Logistic regression classifier as a classification model to detect the ichthyophthiriasis or white spots fish disease after applying five models before, by doing the comparison; the LR model scored the high result after basing the features vector list and the labels list to fit them and do the learning process. The final score of the total algorithm is 92.8% and it is the highest one on the comparison and the confusing matrix in (Fig. 5) show the result of the algorithm, in this figure what happens is that the algorithm answered 4.0 imaged from class one true and 1.0 image false and also answered no images from class two false and 9.0 images true. Knowing that the dataset have been constructed and built from the two classes (1.0 for infected images and 0.0 for non-infected images) and saved into a CSV excel file.

VI. CONCLUSIONS

By using machine learning techniques with swarm optimization and image processing techniques we proposed an accurate system that can identify white spots disease with an accuracy of 92.8%. We used filters to remove the noise, ones to enhance the contrast of the image and GrabCut algorithm to get parasite only, ORB algorithm to extract relevant features of the image and use these features to train our proposed (Logistic Regression Classifier), the result is 92.8%. Also, we propose a powerful dataset with 200.0 images.

VII. FUTURE WORK

We aim to increase our dataset from other fish farms with other environment to raise the accuracy more than 92.8%, update Machine Learning algorithms to Deep Learning and detect other diseases to wild the system.

VIII. REFERENCES

[1] MALIK, Shaveta; KUMAR, Tapas; SAHOO, A. K. A Novel Approach to Fish Disease Diagnostic System based on Machine Learning. Advances in Image and Video Processing, 2017, 5.1: 49.

[2] MALIK, Shaveta; KUMAR, Tapas; SAHOO, A. K. A Novel Approach to Fish Disease Diagnostic System based on Machine Learning. Advances in Image and Video Processing, 2017, 5.1: 49.

[3] DE CASTRO NIZIO, Daniela Aparecida, et al. Essential oils of Varronia curassavica accessions have different activity against white spot disease in freshwater fish. Parasitology research, 2018, 117.1: 97-105.

[4] Wen, J., et al., "A New Method for Fish-Disease Diagnostic Problem Solving Based on Parsimonious Covering Theory and Fuzzy Inference Model," International Federation for Information Processing, Vol. 217, Artificial Intelligence in Theory and Practice, ed. M. Bramer, (Boston:Springer), pp.455-464, 2006.

[5] AL-JUBURY, Azmi, et al. Impact of Pseudomonas H6 surfactant on all external life cycle stages of the fish parasitic ciliate Ichthyophthirius multifiliis. Journal of fish diseases, 2018, 41.7: 1147-1152.

[6] JAFFEL, Zaineb; FARAH, Mohamed. A symbiotic organisms search algorithm for feature selection in satellite image classification. In: Advanced Technologies for Signal and Image Processing (ATSIP), 2018 4th International Conference on. IEEE, 2018. p. 1-5.

[7] Lyubchenko, V., Matarneh, R., Kobylin, O., & Lyashenko, V. (2016). Digital Image Processing Techniques for Detection and Diagnosis of Fish Diseases.

[8] Malik, S., Kumar, T., & Sahoo, A. K. (2017). A Novel Approach to Fish Disease Diagnostic System based on Machine Learning. Advances in Image and Video Processing, 5(1), 49.

[8] Ravisankar, P., Sharmila, T. S., & Rajendran, V. (2018). Acoustic image enhancement using Gaussian and laplacian pyramid–a multiresolution based technique. Multimedia Tools and Applications, 77(5), 5547-5561.

[9] Dung, L. R., Yang, Y. M., & Wu, Y. Y. (2018). A Hierarchical Grab Cut Image Segmentation Algorithm. Journal of Computer and Communications, 6(02), 48.

[10] Eken, S., Aydin, E., & Sayar, A. (2018). Difet: Distributed Feature Extraction Tool for High Spatial Resolution Remote Sensing Images. arXiv preprint arXiv:1808.08521.

[11] Liu, J., Yu, J., Wang, C., & Zhang, X. (2018, February). An ORB Feature Matching Algorithm for Mobile Devices. In Proceedings of the 2018 10th International Conference on Computer and Automation Engineering (pp. 116-119). ACM.

[12] Chen, J., Luo, L., Wang, S., & Wu, H. (2018, July). Automatic Panoramic UAV Image Mosaic Using ORB Features and Robust Transformation Estimation. In 2018 37th Chinese Control Conference (CCC) (pp. 4265-4270). IEEE.
[13] Karami, E., Prasad, S., & Shehata, M. (2017). Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images. arXiv preprint arXiv:1710.02726.
[14] Fernández, S. Álvarez, D. Díaz, M. Iglesias, B. Ena, Scheduling a galvanizing line by ant colony optimization, in Swarm Intelligence, 9th International Conference, ANTS 2014, ed. By M. Dorigo, M. Birattari, S. Garnier, H. Hamann, M. Montes de Oca, C. Solnon, T. Stützle. Lecture Notes in Computer Science,vol. 8667 (Springer, Heidelberg, 2014), pp. 146–157.
[15] Hilbe, Joseph M. Logistic regression models. CRC press, 2017.
[16] Kuo, C. L., Duan, Y., & Grady, J. (2018). Unconditional or conditional logistic regression Model for age-Matched case–control Data?. Frontiers in public health, 6, 57.
[17] Arora, A., Lin, J. J., Gasperian, A., Stein, J., Maldjian, J., Kahana, M. J., & Lega, B. (2018). Comparison of logistic regression, support vector machines, and deep learning classifiers for predicting memory encoding success using human intracranial EEG recordings. Journal of neural engineering.
[18] Cattelan, M., & Varin, C. (2018). Marginal logistic regression for spatially clustered binary data. Journal of the Royal Statistical Society: Series C (Applied Statistics).
[19] García-Balboa, J., Alba-Fernández, M., Ariza-López, F., & Rodríguez-Avi, J. (2018). Analysis of Thematic Similarity Using Confusion Matrices. ISPRS International Journal of Geo-Information, 7(6), 233.
[20] Li, N., & Jimenez, R. (2018). A logistic regression classifier for long-term probabilistic prediction of rock burst hazard. Natural Hazards, 90(1), 197-215.