Diagnosis System For Fish Diseases Trichodina And Gyrodactylus Under Microscope Based On Decision Tree And Lda.

Eman Mohamed
Information Technology Department, Information Systems, Beni-Suef University , emanfathy_sd@fcis.bsu.edu.eg.

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

ABSTRACT:
This paper presents an application for rapid disease diagnosis from parasites' microscopic images of diseased fishes **Trichodina and Gyrodactylus**. A Neural Network-based fish disease diagnosis system was developed, allowing for earlier treatment of infected fish to prevent the spread of disease**.** The system is composed of some steps**:** Frist, segment pathogen area from the microscopic images of infected fish tissue by applying various kinds of image processing such as noise reduction, background extraction, and object detection, second Feature Extraction using ORB, statistical pattern matching LDA compared to the registered pathogen dataset. the last the machine learning DecisionTree Classifier that was used to get the best results in classifying
the fish diseases 87.5%.

Keywords: Trichodina, **Gyrodactylus**, Decision Tree, Oriented FAST and Rotated BRIEF (ORB), Linear discriminant analysis (LDA), Speeded up robust features (SURF).

1 Introduction:

Traditionally, fish disease diagnoses were made using the accumulated experiences of fish-farmer or fish-veterinarian. But the final Neural Network has been used to help to diagnose fish diseases.Fish disease is a serious problem because of its rapid spread through the water to neighboring aqua-farms. Therefore rapid and accurate diagnosis is required to control the disease and to prevent the spread of disease. Gyrodactylus is an interesting little parasite; species of which can be found infesting many different types of freshwater [3]. It is a so-called monogenean trematodes, it generally infests the skin and fins of the fish [4], and Trichodina is one of the parasites that you can't see without a microscope. This parasite is usually found on the gills, skin or fins of the fish and both diseases Spread in tilapia fish[1][2]. This paper proposed a combination of technique which gives better accuracy and compared with the existing combination. In proposed combination techniques which combine after collecting parasites' microscopic images of diseased fishes proposed a system in which the image of the diseased fish recognizes by

using LDA and preprocessing images by convert image from RGB to Gray level to avoid any appearance of false points. Feature extraction will be done through the ORB feature extraction technique from all microscope images of fish to make the dataset [7]. Classify this dataset by using Decision Tree Classifier and LDA make predictions by estimating the probability that a new set of inputs belongs to each disease [6]. The dataset to be divided into training validation set and testing validation. All algorithms are implemented in Dataset and display the best results for 87.5%.

2 Material and methods:

After collecting images of both diseases is created by extracting microscope image features of the fish and storing them in an excel file. Then Training and testing sample has to be taken and the output will be identified images which is present in the dataset.

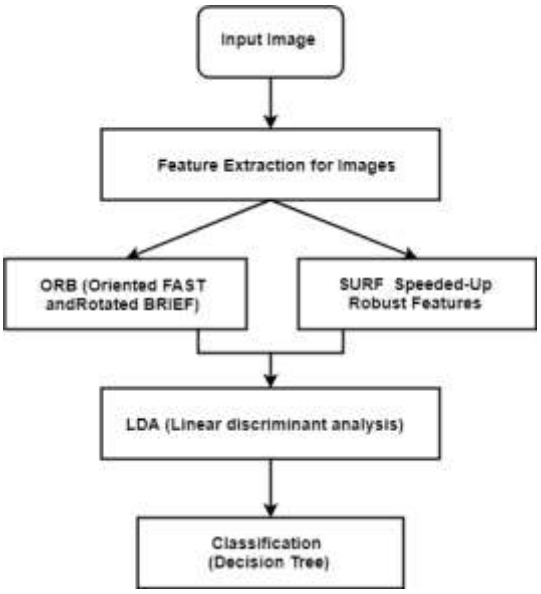


Figure: 1Flow chart of the system.

2.1Oriented FAST and Rotated BRIEF (ORB):

ORB is basically a fusion of FAST key point detector and BRIEF descriptor with many modifications to enhance the performance. First, it uses FAST to find key points, and then apply the Harris corner measure to find top N points among them [7][10].

2.2 Speeded up robust features (SURF):

It is a patented local feature detector and descriptor. It can be used for tasks such as object recognition, image registration. To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. Its feature descriptor is based

on the sum of the Haar wavelet response around the point of interest. These can also be computed with the aid of the integral image[8][9].

2.3 Fisher's Linear Discriminant Analysis (LDA):

LDA makes predictions by estimating the probability that a new set of inputs belongs to each class. The class that gets the highest probability is the output class and a prediction is made. The model uses Bayes Theorem to estimate the probabilities. Briefly, Bayes' Theorem can be used to estimate the probability of the output class (k) given the input (x) using the probability of each class and the probability of the data belonging to each class[13][14].

2.4 Decision Tree in Machine Learning:

Decision Trees classify instances by sorting them based on feature values. Each node in a decision tree represents aFeature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. The feature that best divides the training data would be the root node of the tree[6][15].

3 Proposed Methodology:

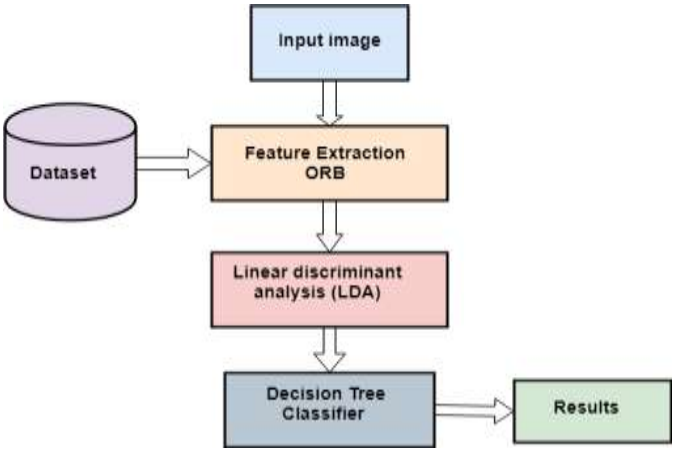
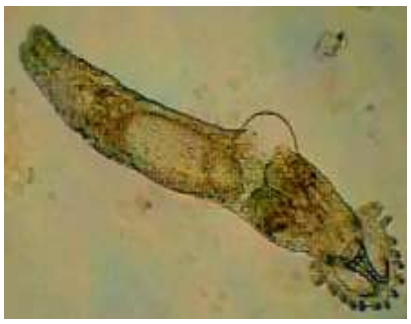


Figure: 2proposed method

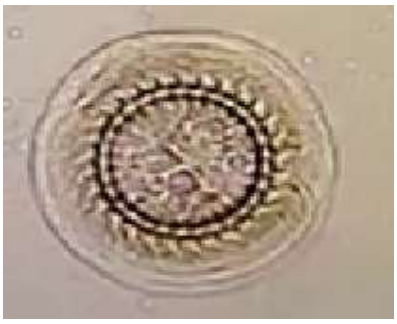
3.1 Steps as shown in (figure: 2)

- **Input images:**

Those collected from different resources for two diseases under a microscope and Converted to the desired output format according to the system.



(**Gyrodactylus**)

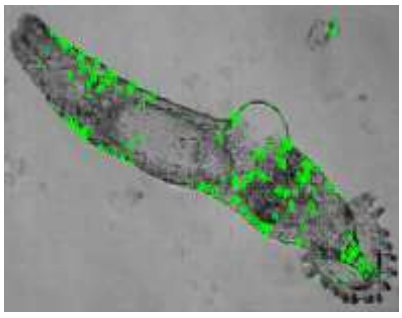


(**Trichodina**)

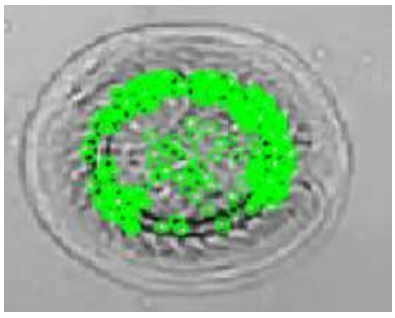
Table: 1both diseases.

- **Feature Extraction and detection:**

Extract the feature from the feature extraction algorithm which is ORB (Oriented FAST and Rotated BRIEF) for recognition. To diagnose **Gyrodactylus and Trichodina**, infected images are loaded into the database and features are extracted.



(1)



(2)

Table: 2Extract images.

- **Dataset:**

After the feature extraction dataset creates by storing interest points for every image in an excel file.

	mean	median	stander	var	target
count	262.000000	262.000000	262.000000	262.000000	262.000000
mean	133.799604	135.874046	74.407862	5538.005335	1.629771
std	6.516057	8.652142	1.216991	184.682151	0.483790
min	98.045608	96.000000	70.938491	5032.269528	1.000000
25%	129.818424	130.250000	73.837406	5451.962499	1.000000
50%	134.074875	136.000000	74.266483	5515.510532	2.000000
75%	138.583188	142.750000	74.731698	5584.826612	2.000000
max	148.722055	158.000000	82.123353	6744.245132	2.000000

Figure: 3Dataset

- **Fisher'sLinear discriminant analysis (LDA):**

Linear discriminant analysis (LDA) is one of the most popular supervised dimensionality reduction methods for the dataset[13].

```
# Applying Fisher's Linear Discriminant Analysis
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
lda = LinearDiscriminantAnalysis(n_components = 2)
X_train = lda.fit_transform(X_train, y_train)
X_test = lda.transform(X_test)
```

Figure: 4LDA Algorithm.

- **Classification (Decision Tree):**

After the feature selection and created dataset classification will do by a classifier which classifies the diseased fish using Decision Tree Classifier and LDA makes predictions by estimating the probability that a new set of inputs belongs to each disease. The dataset to be divided into training validation set and testing validation. Then a classifier detects input image which diseases or not[15][16].

Algorithm	Accuracy
ORB-LDA-DT	87.5%
SURF-LDA-DT	85%

Table: 3Accuracy of algorithms.

4 Experimentation & Results:

After applying a comparison between algorithms we get these results. Extract the Features from ORB (Oriented FAST and Rotated BRIEF) as it is a better feature extractor as compared to the SURF and it is corner detector and find the interesting point. After Feature Extractor applied the LDA as it reduces the dimensions and preserves the useful information. Machine Learning Algorithm has been applied (Decision Tree) to get better accuracy as it gives better results after train the system.

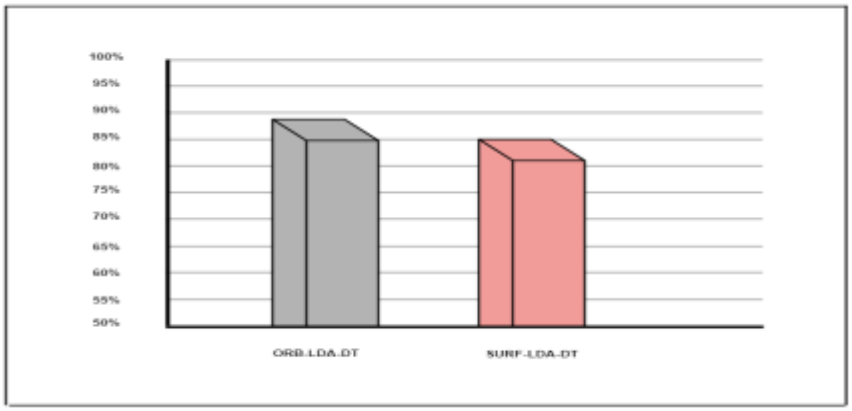


Figure: 5Comparision between algorithms.

The showing the comparison between the classification accuracy of SURF-LDA-DT and proposed combination which is ORB-LDA-DT.

4.1 Confusion Matrix of SURF-LDA-DT:

As shows in the confusion Matrix shows the 85% accuracy through the classifier

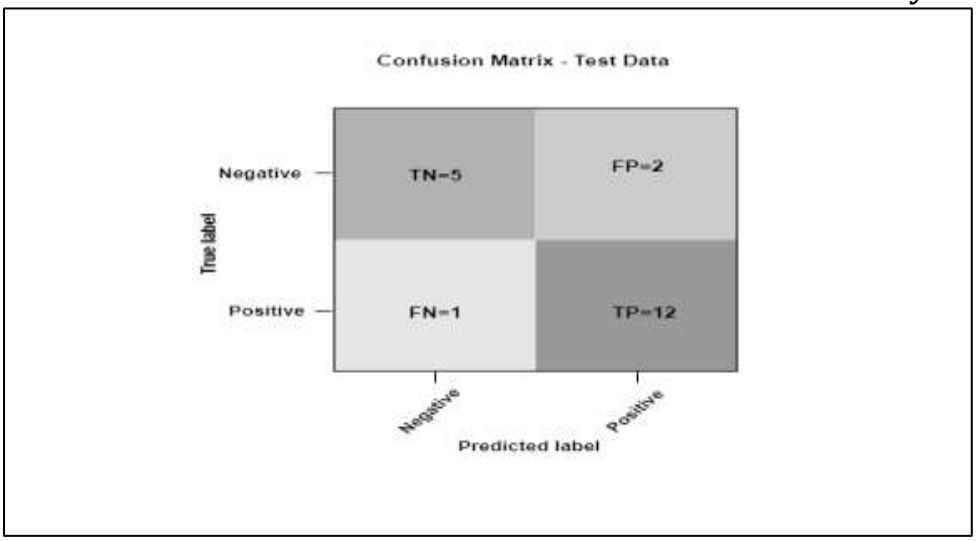


Figure: 6Confusion Matrix of (SURF-LDA-DT).

The above figure 6 shows a confusion matrix of SURF-LDA-DT which gives 85% accuracy. It is not better than the proposed combination[12][17].

4.2 Classification Report SURF-LDA-DT:

Support refers to the number of observations in each class [18].

	precision	recall	f1-score	Support
Trichodina	0.83	0.71	0.77	7
Gyrodactylus	0.86	0.92	0.89	12
avg / total	0.85	0.85	0.85	20

Table: 4classification report SURF-LDA-DT

4.2 Confusion Matrix of ORB-LDA-DT:

As shown in figure 7, the confusion Matrix shows the 87.5% accuracy through the classifier[17] **.**

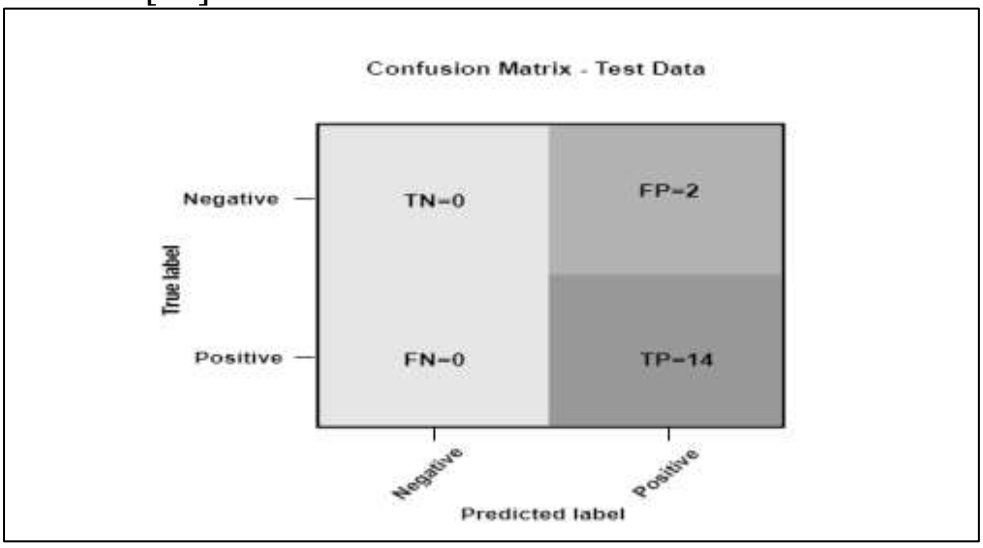


Figure: 7Confusion Matrix of (ORB-LDA-DT).

4.4 Classification Report ORB-LDA-DT:

	precision	recall	f1-score	Support
Trichodina	0.00	0.00	0.00	2
Gyrodactylus	0.88	1.00	0.93	14
avg / total	0.77	0.88	0.82	16

Table: 5classification report ORB-LDA-DT.

4.5 Final Accuracy:

The proposed combination gives better accuracy of 87.5% as compared to (SURF-LDA-DT).

Algorithm	Accuracy
ORB-LDA-DT	87.5%
SURF-LDA-DT	85%

Table: 5validation results.

5 Discussion:

5.1 Using ORB (Oriented FAST and Rotated BRIEF) for recognition to diagnose **Gyrodactylusand and Trichodina. Using ORB** after applying SURF model before, by doing the benchmark; ORB scored the high result.
5.2 The validation process gets the results in (Table 5) and also the ORB scored the high result than the SURF model.

5.3 Using Fisher's Linear Discriminant Analysis (LDA) after feature extraction. LDA is one of the most popular supervised dimensionality reduction methods for the dataset. Discriminant analysis is used when groups are known a priori (**Gyrodactylus and Trichodina**). Each case must have a score on one or more quantitative predictor measures, and a score on a group measure. In simple terms, discriminant function analysis is classification. the act of distributing things into groups, classes or categories of the same type.

5.4 The final score of the total algorithm is 87.5% and it's the highest one on the benchmark and the confusing matrix in (Figure: 7) show the result of the algorithm, in this figure what happens is that the algorithm answered no images from class one false and 2 images true and also answered no images from class two false and 14 images true.

5 Conclusions

The proposed combination of ORB-LDA-DT gives better classification accuracy than the existing combined technique which is SURF-LDA-DT to recognize the **Trichodina and Gyrodactylus**infected disease fish as it is misidentified to detect. It is a fast and efficient method to recognize and detect the **Trichodina and Gyrodactylus** diseases as compared to the Traditional Method. The Decision Tree results based on the real images of infected Fish disease dataset. The proposed combination (ORB-LDA-DT) gives 87.5% accuracy.

6 Future works:

We aim to increase our dataset with more samples and update machine learning algorithms to deep learning techniques to increase the accuracy of the system by up to 98%. Also, we aim to identify other diseases in fish rather than only **Gyrodactylus and Trichodina**diseases; we plan to detect the disease based on the kinds of pathogens such as parasites, bacteria, fungal, and the virus for more convenient and useful aid to the diagnosis system. Also, we will add Microscope slides for the disease. Also, we will make machine learning algorithms can apply to different feature Descriptors.

8 References:

[1] Abdelkhalek, N. K., Mohamed, A., Salama, M. F., Elmishmishy, B., Ali, M. O., El-Ashram, A., ...& Al-Araby, M. A. (2018). Molecular identification of Trichodinacompacta Van As and Basson, 1989 (Ciliophora: Peritrichia) from cultured Oreochromisniloticus in Egypt and its impact on immune responses and tissue pathology. Parasitology research, 117(6), 1907-1914.

[2] Haque, M. A., Kibria, M. M., &Asmat, G. S. (2018). Trichodinaamblypharyngodoni sp. n. and Trichodinahoffmani Wellborn, 1967 (Ciliophora: Trichodinidae) from the freshwater fishes in the BaikkaBeel of Moulvibazar district in Sylhet division, Bangladesh. Annals of parasitology, (2), 101-107.

[3] Barzegar, M., Mousavi, H. E., Rahmati-holasoo, H., Mirghaed, A. T., &Bozorgnia, A. (2018). Gyrodactylus (Monogenea, Gyrodactylidae) parasite fauna of fishes in some rivers of the southern Caspian Sea basin in Mazandaran province. Iranian Journal of Veterinary Medicine, 12(1), 35-44.

[4] Mieszkowska, A., Górniak, M., Jurczak-Kurek, A., &Ziętara, M. S. (2018). Revision of Gyrodactylussalaris phylogeny inspired by new evidence for Eemian crossing between lineages living on grayling in Baltic and White sea basins. PeerJ, 6, e5167.

[5] Brunello, A., Marzano, E., Montanari, A., &Sciavicco, G. (2017). Decision Tree Pruning via Multi-Objective Evolutionary Computation. INTERNATIONAL JOURNAL OF MACHINE LEARNING AND COMPUTING, 7(6), 167-175.

[6] 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.

[7] 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.

[8] 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.

[9] Chen, S. J., Zheng, S. Z., Xu, Z. G., Guo, C. C., & Ma, X. L. (2018). AN IMPROVED IMAGE MATCHING METHOD BASED ON SURF ALGORITHM. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42(3).

[10] Yu, L., Yu, Z., & Gong, Y. (2015). An improved ORB algorithm of extracting and matching features. International Journal of Signal Processing, Image Processing, and Pattern Recognition, 8(5), 117-126.

[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).

[12] Haghighi, S., Jasemi, M., Hessabi, S., & Zolanvari, A. (2018). PyCM: Multiclass confusion matrix library in Python. Journal of Open Source Software, 3(25), 729.

[13] Zhao, H., Wang, Z., &Nie, F. (2018). A New Formulation of Linear Discriminant Analysis for Robust Dimensionality Reduction. IEEE Transactions on Knowledge and Data Engineering.

[14] Cai, W., Guan, G., Pan, R., Zhu, X., & Wang, H. (2018). Network linear discriminant analysis. Computational Statistics & Data Analysis, 117, 32-44.

[15] Loussaief, S., &Abdelkrim, A. (2016, December). Machine learning framework for image classification. In 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)(pp. 58-61). IEEE.

[16] Chandralekha, M., &Shenbagavadivu, N. (2018). Performance Analysis Of Various Machine Learning Techniques To Predict Cardiovascular Disease: An Emprical Study. Appl. Math, 12(1), 217-226.

[17] Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011, April). Confusion Matrix-based Feature Selection.In MAICS (pp. 120-127).

[18] https://botfactory.in/2018/08/30/classification-report-sklearn/.