Fish diseases identification and classification using Machine Learning

Divinely S. J, Sivakami K, Dr. V. Jayaraj
PG Scholar, Dept of ECE, Nehru INSTITUTE of Engineering& Technology,
Assistant Professor, Dept of ECE, Nehru Institute of Engineering & Technology,
Professor and Head, Dept of ECE, Nehru Institute of Engineering &Technology

Abstract— **In the process of aquaculture, farmers usually cannot make accurate diagnosis for the sudden fish disease. And they can hardly take treatment measures timely and effectively. They need to get professional advice and guidance from the fish disease experts. This often results to miss the best treatment time. The fish disease usually spreads rapidly and the fish may dead massively which causes a loss to the farmers. Even under the condition of relatively rich communication ways, it is difficult for experts to observed the field aquaculture environment intuitively in real-time, the dynamic condition and anatomical details of the fish. All this information are the key elements for the fish disease diagnosis. To help farmers of aquaculture fish diseases is detected and classified using Probabilistic Neural Network (PNN).**

Index Terms—**ammonia poisoning, Camallanus Worms,Dropsy**

I. INTRODUCTION

The disease possibility for the aquatic fish increases under the high density cultivation mode [1, 2]. With varieties of aquaculture, the fish disease also has a new development trend [3]. Many incurable diseases need professional diagnosis experts to diagnosis and give a reasonable treatment [4, 5]. Some fish diseases break out and spread very quickly [6]. The aquatic fish will be infected and dead in a short period of time if the diagnosis is delayed and the effective treatment is not taken timely [7]. This will bring great economic losses to farmers [8]. In order to achieve the fish disease diagnosis and treatment in time, then to reduce the risk of aquaculture, a real-time and remote diagnosis expert system for fish disease is established based on the modern internet communication technology. The on-site real-time video could be gathered while the fish disease breaks out. The micro-dissection image could also be sent to the remote expert in real-time by multimedia micro-dissection table. The expert can give a professional diagnosis and treatment to famer, so as to reduce the losses of breeding maximally. Fish like snakeheads and barbs in fresh water and mullets in brackish water are severely affected. However, the susceptibility of the Indian major carps, the major carps being affected have come from India, but the diagnostic features have not been demonstrated in the report.[10] Correct diagnosis of EUS is important to avoid confusion with other ulcerative conditions. It is important to note that ulcer is a non-specific clinical lesion which may be caused by many different agents [11]. So, all ulcers are not EUS ulcers because they are not in epizootic proportions or not seasonal in nature. Manually diagnosis is not corrected somehow and more chances to give error. To recognize and identification of the EUS disease for the ease of people as compare to manual detection as Fish is a food of millions of people. Segmentation has done on various fish images to enhance the image, various edge detection techniques have been applied on the fish image and filter out the noise, it has been found that salt and pepper noise find in water. After the first part of paper, feature extraction has been done through various descriptors and to classify the disease and non-disease fish, classifier has been used to find the accuracy. If PCA is applied after feature extraction through various descriptor then features have been reduced because it reduce the dimensionality of the features ,only the non-important content will loss. Many researchers have done lot of work. Dalal et al.[14] extracted the HOG features from all locations of a dense grid on a image region and the combined features are classified by using linear SVM. Experiments are conducted on KTH database and gives better performance in terms of 100% recognition rate for training set and 89.8% accuracy for test set.Valentin Lyubchenko et al. [16] in the paper choose or select the color markers to detect the infected area or to distinguish the normal and infected area[17]. The main disadvantage of this methodology is the appearing of false point marked as infected areas due to the automatic allocation of the selected color which may affect the calculation. On the contrary, this methodology has many advantages such as: the ability to change the size of the marker when selecting the colors in the image segmentation to avoid any appearance of false points, the ability to analyze fish skin areas that are difficult to identify visually and the ability to isolate areas of the skin with varying levels of destruction. Hitesh Chakravorty et al. [18] proposed a system in which the image of the diseased fish recognizes by using PCA method and diseased area segmentation of fish image based on color features with K means clustering.

II. PREPROCESSING

The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features important for further processing. Position dependent brightness correction.

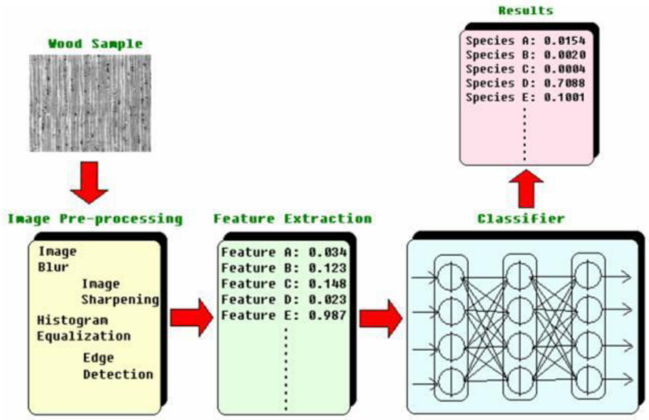

Figure 1 Preprocessing of images

Ideally, the sensitivity of image acquisition and digitization devices should not depend on position in the image, but this assumption is not valid in many practical cases.o Sources of degradation. Uneven sensitivity of light sensors v Uneven object illuminationo Systematic degradation can be suppressed by brightness correction. Pixel Brightness Transformationso Brightness transformations modify pixel brightness -- the transformation depends on the properties of a pixel itself.o Brightness correctionso Gray scale transformations.o Brightness correction v considers original brightness v pixel position in the image.o Gray scale transformations v change brightness without regard to position in the image.

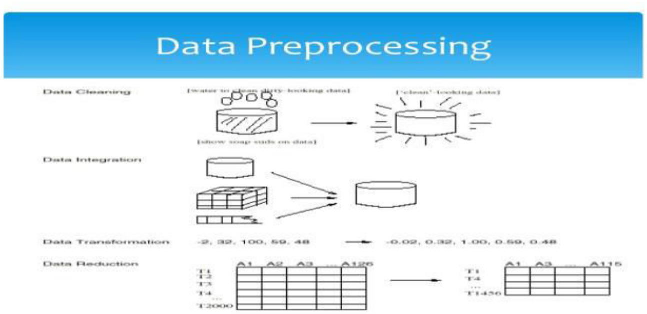

Figure 2 Data Preprocessing

III. CURVELET TRANSFORM

Curvelet transform is a higher dimensional generalization of the wavelet transform designed to represent images at different scales and different angles. it actually overcomes the missing directional selectivity of wavelet transforms in images. curvelets are designed to handle curves using only a small number of coefficients. hence the curvelet handles curve discontinuities well. The 2-D complex wavelets are essentially constructed by using tensor-product one-dimensional (1-D) wavelets. The directional selectivity provided by complex wavelets (six directions) is much better than that obtained by the classical DWTs (three directions), but is still limited.

The second-generation curvelet transform has been shown to be a very efficient tool for many different applications in image processing, seismic data exploration, fluid mechanics, and solving partial different equations (PDEs).Following the considerations in the 1-D case, the elements of the curvelet family should now provide a tiling of the 2-D frequency space. Therefore the curvelet construction is now based on the following two main ideas. Consider polar coordinates in frequency domain. 1) Construct curvelet elements being locally supported near 2) wedges according to Figure 3, where the number of wedges is Nj 54 # 2<j/2= at the scale 22j

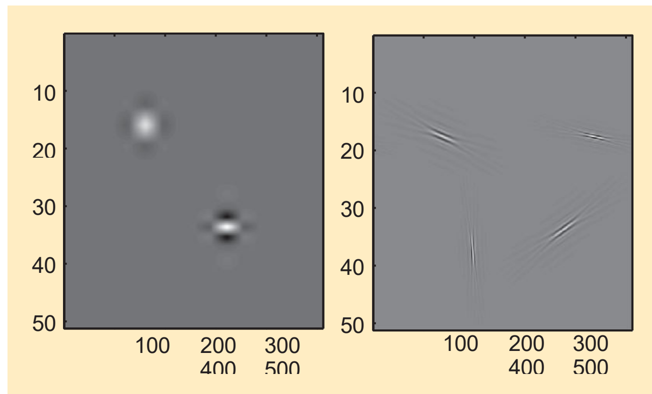

Figure 3 Polar coordinates of curvelet transform

The idea of the 3-D curvelet transform on Cartesian arrays can be carried out analogously as done in the section "Transition to Cartesian Arrays" for the 2-D case. This time, one considers curvelet functions being supported on sheared truncated pyramids instead of sheared trapezoids. The aim of local feature representation is to represent the image which is based on some salient regions. The image is depends on its local structures with a set of local feature descriptors and which is obtain from a set of image regions called interest regions. HOG feature descriptor is for the object detection. It uses gradient orientation in the localized potions of an image and It divides the image into blocks and further blocks divided into cells then calculate the histogram of each cell after that normalize the histograms of the cells. Overlapping of the blocks has done in it. Timeliness of traditional fish disease diagnosis is not good enough. Some expert system based on artificial intelligence could diagnose preliminarily through image or verbal description. But it can't show the true status of the fish disease in real time. Without the expert help online, the accuracy could not be guaranteed. The misdiagnosis risk is higher.

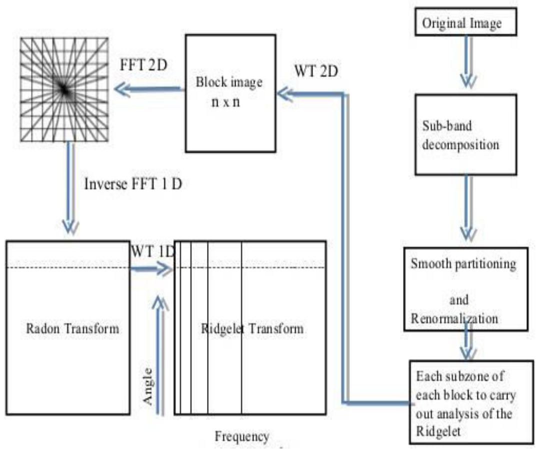

Figure 4 Block diagram of curvelet transform

The noise is exceed when the signal is processed by the wavelet transform from high frequencies (low dilatations) to low (high dilations). In addition to its property to detect singularities. The curvelet transform allows to extend this application to detection of extended object. The 3-D curvelet functions then depend on four indices instead of three; the scale, the position and two angles; and for the shearing process, one can introduce 3-D shear matrices.

a b

Figure 5. a)before preprocessing b)after preprocessing

III METHODOLOGY

Figure shows the methodology of classification

Step1: Image Acquisition: -The collection or acquiring of the images from some source. The database collected from various internet resources.

Step 2: Image Pre-Processing: - The removal of the noise, applied the filter, normalization of the intensity image and various morphological operations applied on the image.

Step-3:-Image Segmentation:-Various segmentation algorithms are used to segment and enhance the image descriptors.

Step-4:-Feature Extraction:-Various feature Extraction methods have been applied for efficiency of detection

Step 5:-Image Classification:-Fishes with different diseases as ammonia poisoning, camallanus worms,dropsy is classified. And also uninfected fishes are noticed and separated.

Step 6:-Image Evaluation & Accuracy: Evaluate the methods and techniques. Appropriate evaluation strategy for segmentation and classification of image is determined. Various classification algorithms have been applied to calculate classification accuracy.

Machine learning algorithm is applied to the pattern recognition system and to get the good results after train the system by dataset. Divide the dataset into testing phase and Training phase to find the hidden neurons in pattern recognition system . K-NN (Nearest Neighbour):- It is a supervised learning algorithm and it commonly used machine learning methods . It classifying objects based on closest training examples in the feature vector. It is an important approach to nonparametric classification and is quite easy and efficient. In which classification assigns a class label for unknown samples by estimating its k-nearest neighbours based on known sample..

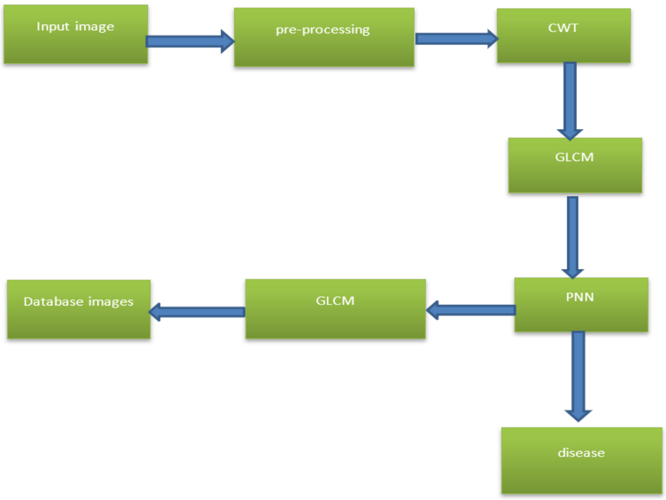

Figure 6. Block diagram of proposed methodology

IV GRAY LEVEL CO OCCURRENCE MATRIX

Also referred as **co-occurrence distribution.**It is the most classical second-order statistical method for texture analysis. An image is composed of **pixels** each with an intensity (a specific gray level), the GLCM is a tabulation of how often different combinations of gray levels co-occur in an image or image section. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity at the pixel of interest.GLCM texture feature operator produces a virtual variable which represents a specified texture calculation onasinglebeamechogram.

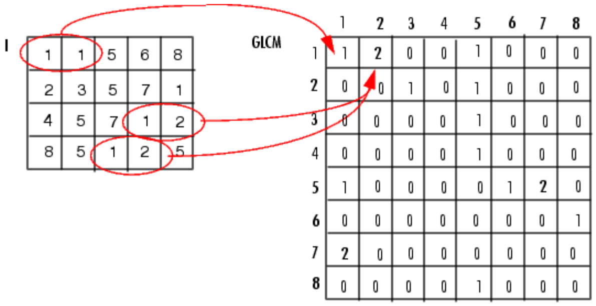

Figure 6. GLCM

Quantize the image data: Each sample on the echogram is treated as a single image pixel and its value is the intensity of that pixel. These intensities are then further quantized into a specified number of discrete gray levels, known as Quantization.

Create the GLCM*:* It will be a square matrix N x N in size where N is the Number of levels specified under Quantization.

Let s be the sample under consideration for the calculation.

Let W be the set of samples surrounding sample s which fall within a window centered upon sample s of the size specified under Window Size. Define each element i, j of the GLCM of sample present in set W, as the number of times two samples of intensities i and j occur in specified Spatial relationship. The sum of all the elements i, j of the GLCM will be the total number of times the specified spatial relationship occurs in W.

Make the GLCM symmetric: Make a transposed copy of the GLCM. Add this copy to the GLCM itself.This produces a symmetric matrix in which the relationship i to j is indistinguishable for the relationship j to i.. Due to summation of all the elements i, j of the GLCM will now be twice the total number of times the specified spatial relationship occurs in W.

Normalize the GLCM: Divide each element by the sum of all elements.The elements of the GLCM may now be considered probabilities of finding the relationship *i, j (or j, i)* in W.

Define each element i, j of the GLCM of sample present in set W, as the number of times two samples of intensities i and j occur in specified Spatial relationship. The sum of all the elements i, j of the GLCM will be the total number of times the specified spatial relationship occurs in W.

- Make the GLCM symmetric:
- Make a transposed copy of the GLCM.
- Add this copy to the GLCM itself.
- This produces a symmetric matrix in which the relationship i to j is indistinguishable for the relationship j to i.
- Due to summation of all the elements i, j of the GLCM will now be twice the total number of times the specified spatial relationship occurs in W.
- Normalize the GLCM:
- Divide each element by the sum of all elements.

The elements of the GLCM may now be considered probabilities of finding the relationship *i, j (or j, i)* in W.

Calculate the selected Feature. This calculation uses only the values in the GLCM. For e.g. Energy, Entropy, Contrast, Homogeneity, Correlation,

The sample s in the resulting virtual variable is replaced by the value of this calculated feature.

IV PROBABILISTIC NEURAL NETWORK

A **probabilistic neural network (PNN)** is a feed forward neural network which is widely used in classification and pattern recognition problems. In the PNN algorithm, the parent probability distribution function (PDF) of each class is approximated by a Parzen window and a non-parametric function. Then, using PDF of each class, the class probability of a new input data is estimated and Bayes' rule is then employed to allocate the class with highest posterior probability to new input data. By this method, the probability of mis-classification is minimized.. In a PNN, the operations are organized into a multilayered feed forward network with four layers:Input layer,Hidden layer and Pattern layer/Summation layer/Output layer

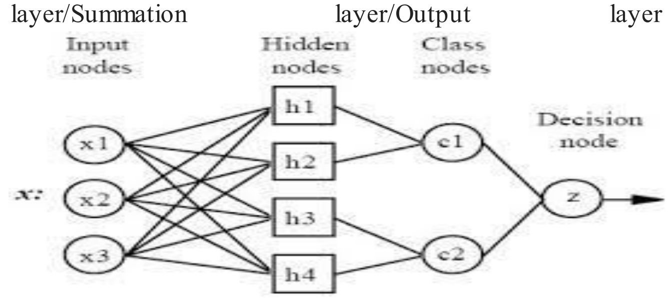

Figure 7 Nodes in GLCM

PNN is often used in classification problems-When an input is present, the first layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 (positive identification) for that class and a 0 (negative identification) for non-targeted classes.

Input layer-Each neuron in the input layer represents a predictor variable. In categorical variables, *N-1* neurons are used when there are *N* number of categories. It standardizes the range of the values by subtracting the median and dividing by the interquartile range Then the input neurons feed the values to each of the neurons in the hidden layer.

Hidden layer-This layer contains one neuron for each case in the training data set. It stores the values of the predictor variables for the case along with the target value. A hidden neuron computes the Euclidean distance of the test case from the neuron's center point and then applies the kernel function using the sigma values.

Summation layer-For PNN networks there is one pattern neuron for each category of the target variable. The actual target category of each training case is stored with each hidden neuron; the weighted value coming out of a hidden neuron is fed only to the pattern neuron that corresponds to the hidden neuron's category. The pattern neurons add the values for the class they represent.

Output layer-The output layer compares the weighted votes for each target category accumulated in the pattern layer and uses the largest vote to predict the target category.

- PNNs are much faster than multilayer perceptron networks.
- PNNs can be more accurate than multilayer perceptron networks.
- PNN networks are relatively insensitive to outliers.
- PNN networks generate accurate predicted target probability scores.
- PNNs approach Bayes optimal classification.

V RESULTS AND ANALYSIS

A. *Proposed methodology*

Step 1 :-Preprocessing i.e. RGB to gray scale conversion has been applied for enhancing the fish images

Step 2:-Extract the Features from CWT (Curvelet Wavelet Transform) as it is a better feature extractor as compared to the others and it is corner detector and find the interest point.

Step 3 :-After CWT applied the GLCM (Gray level co-occurrence Matrix) as it reduce the dimensions and preserve useful information.

Step 4:- Machine Learning Algorithm has been applied (Probabilistic Neural Network) to get the better accuracy as it gives better results after train the system.

B. *Performance Measure of PNN*

Proposed combination of CWT-GLCM-PNN technique compare with the other Feature descriptor technique with Machine learning algorithm and find the accuracy, among all these the proposed combination feature discriptor gives better accuracy as shows in Figure 8.

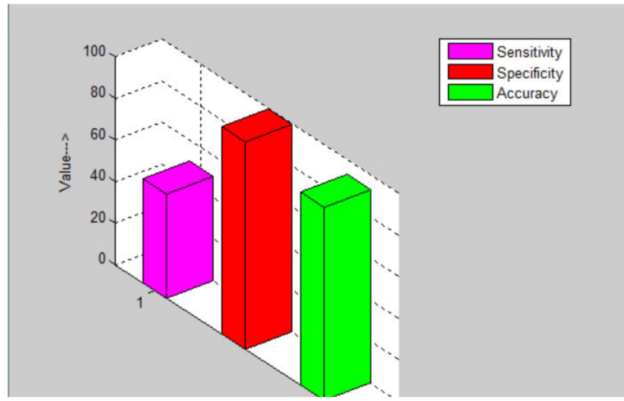

Figure 8. Performance of PNN

VI CONCLUSION

The paper concludes that the proposed combination gave better accuracy after applied the Machine learning algorithm. PNN helped in increase the accuracy. The Experimentation has been applied on the real images of the infected fish images dataset. The implementation has been done in MATLAB software. It automatically detects or diagnoses the Fish EUS disease. In Future scope Machine learning algorithms can apply on different feature Descriptors.

ACKNOWLEDGEMENT

The infected Fish images have been collected from various sources of databases and preprocessed to get better results.

REFERENCES

[1] Miaojun Xu, Jianke Zhang, Xiaoqiu Tan, "Intelligent fish disease diagnostic system based on SMS platform", Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications (ISDEA2013), 2013, pp. 897-900

[2] Han Chang-Min, Yang Soo-Yung, Ceong Heeteak, Park Jeong-Seon, "Clinical sign-based fish disease diagnosis aid system", Lecture Notes in Electrical Engineering, 2011, v98, pp. 10391046

[3] Lee Sang-Woong, Park Jeong-Seon, "Microscopic image recognitionbased fish disease diagnosis system", Journal of Convergence Information Technology, 2011, v6, n10, pp. 355-364

[4] Song Jian, Sun Xueyan, "Fish disease image retrieval method based on color and shape features", International Review on Computers and Software, 2012, v7, n2, pp.885-890

[5] Li Daoliang, Fu Zetian, Duan Yanqing, "Fish-Expert: A webbased expert system for fish disease diagnosis", Expert Systems with Applications, 2002, v23, n3, pp. 311-320

[6] Zeldis Daniel, Prescott, Shawn, “Fish disease diagnosis program - problems and some solutions”, Aquacultural Engineering, 2000, v23, n1, pp. 3-11

[7] Zhu Wei, Li, Daoliang, “A CBR system for fish disease diagnosis”, IFIP Advances in Information and Communication Technology, 2007, v259, pp. 1453-1457

[8] Zhang Jie, Li Daoliang, “A call center oriented consultant system for fish disease diagnosis in China”, IFIP Advances in Information and Communication Technology, 2008, v259, pp. 1447-1451

[9] Nayak, K. K., Pradhan, J., Das, B. K. (2014). Original Research Characterization, pathogenicity, antibiotic sensitivity and immune response of Flavobacterium columnare isolated from Cirrhinus mrigala and Carassius auratus. Int. J. Curr. Microbiol. App. Sci, 3(11), 273-287.

[10] Jeong-Seon Park, Myung-Joo Oh, and Soonhee Han, Fish Disease Diagnosis System Based on Image Processing of Pathogens Microscopic Images, Frontiers in the Convergence of Bioscience and Information Technologies 2007

[11] Burge, C. A., Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Ford, S. E. (2014). Burge, C. A., Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Ford, S. E. (2014). Climate change influences on marine infectious diseases: implications for management and society. Annual review of marine science, 6, 249-277.

[12] Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., Saksida, S. M. (2015). Infectious diseases affect marine fisheries and aquaculture economics. Annual review of marine science, 7, 471-496.

[13] Afferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., Saksida, S. M. (2015). Infectious diseases affect marine fisheries and aquaculture economics.,Annual review of marine science, 7, 471-496

[14] Navneet Dalal, Bill Triggs,”Histograms of Oriented Gradients for Human Detection”International Conference on Computer Vision & Pattern Recognition (CVPR ’05), Jun 2005, San Diego, United States.

[15] Valentin Lyubchenko,Rami Matarneh,Oleg Kobylin,Vyacheslav Lyashenko,”Digital Image Processing Techniques for Detection and Diagnosis of Fish Diseases”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 7, July 2016

[16] A. Jeyanthi Suresh , P.Asha ,Human Action Recognition in Video using Histogram of Oriented Gradient (HOG) Features and Probabilistic Neural Network (PNN), International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 7, July 2016

[17] Daoliang Lia, Zetian Fua, Yanqing Duanb,2002.Fish-Expert: a webbased expert system for fish disease diagnosis, Expert Systems with Applications 23 ,311–320.

[18] Abdi. H., & Williams, L.J. (2010). "Principal component analysis". Wiley Interdisciplinary Reviews: Computational Statistics. 2 (4): 433–459. doi:10.1002/wics.101