

MATLAB and Performance Metrics for Fish Disease Detection

Posted on October 11, 2026

View this essay online:

<https://academic-master.com/matlab-performance-metrics-fish-disease-detection/>

Introduction

Fisheries and aquaculture play an important role in social and economic development [1]. They contribute to the world's protein supply, food security and economy. Demand for fish is rising as the population grows and the benefits of fish as a source of animal protein become more widely recognized. Disease can limit production and threaten the continued development of the fisheries sector [2]. Because disease can spread rapidly among fish, efficient, accurate and timely identification is needed to support prevention. Delayed detection and prevention can increase costs and may have negative effects on human health [3].

This manuscript examines the application of MATLAB across three fish-disease detection studies. It explains the performance metrics used to evaluate their models, including MSE, accuracy, precision, F1-score, MCC, sensitivity, specificity, NPV, FNR and FPR. It also describes the datasets, compares the proposed models with existing approaches and reviews the training and testing procedures.

MATLAB for Technical Computing

MATLAB is a language that has been specifically created for technical computing and is well-known for its superior performance capabilities. It integrates computation, visualization, and programming in a user-friendly environment where problems and solutions are expressed in familiar mathematical notation. MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This enables you to solve numerous technical computing issues, particularly those involving matrix and vector formulations, in a fraction of the time it would take to write a program in a noninteractive scalar language such as C or Fortran. The name MATLAB stands for matrix laboratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together represent the state-of-the-art in software for matrix computation. MATLAB has evolved over a period of years with input from many users. In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering, and science. In the field of technical computing, MATLAB is the preferred tool for efficient research, development, and analysis. MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to learn and

apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include [image processing](#), signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others. This paper focuses on image processing and the assessment of performance using evaluation metrics [4].

Performance Metrics

Performance metrics are numerical indicators used to assess the effectiveness and quality of predictions. Evaluation metrics are closely connected to machine learning tasks and provide insights into a model's strengths and weaknesses. They can be grouped into three categories: classification metrics, such as accuracy, precision, F-measure, sensitivity, specificity, Matthews Correlation Coefficient (MCC), Negative Predictive Value (NPV), False Positive Rate (FPR) and False Negative Rate (FNR); regression metrics, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE); and clustering metrics, such as silhouette score and the Davies-Bouldin index [5].

In the first study, the proposed model's performance was measured using MSE. Backpropagation was used to reduce errors; genetic algorithms and [optimization algorithms](#) can also serve this purpose. In the second and third studies, performance was evaluated using classification metrics, including accuracy, precision, F-measure, sensitivity, specificity, MCC, NPV, FNR and FPR.

Mean Squared Error

MSE measures errors in a statistical model by calculating the average squared difference between observed and predicted values. It is also known as Mean Squared Deviation (MSD). If the model has no errors, MSE is zero; its value increases as the magnitude of the errors increases. Equation (1) shows its mathematical representation.

$$(MSE) = \frac{\sum_{i=1}^N ((A_i - \hat{A}_i))^2}{N} \quad (1)$$

Where (A_i) is the i th observed value, $(\hat{A}_i)$ is the corresponding predicted value and N is the number of observations.

Accuracy

Accuracy is the classification metric that is most easily comprehensible. It measures the accuracy of estimated instances by comparing them with the entire quantity of instances. In spite of its simplicity, accuracy may have the potential to be misleading, particularly in imbalanced datasets where one class outnumbers the others. Eq (2) represents the formula of accuracy.

$$(Accuracy) = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

Precision

Precision is centered on identifying the accurate positives within the projected positives. It quantifies the ratio of accurate positive forecasts in relation to all positive forecasts. Eq (3) shows the formula of precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

Recall or Sensitivity

The measure of recall evaluates the model's capacity to identify all the genuine positive instances, quantifying the ratio between the number of accurate positive predictions and the overall number of real positive instances. Eq (4) represents the mathematical representation of Recall

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

F1-Score

The F1-Score is determined by finding the harmonic mean of precision and recall. This particular metric ensures a balanced assessment by considering both precision and recall. By amalgamating these two measures, a singular metric is acquired to evaluate the overall performance of a classifier. Notably, the harmonic mean assigns a greater emphasis on lower values, enabling the F1-Score to account for the impact of false positives and false negatives equally. Eq (5) shows the formula of F1-Score

$$\text{F-measure} = \frac{2 * TP}{2 * TP + FP + FN} \quad (5)$$

Specificity

The metric of specificity quantifies the ratio of accurately identified true negatives by the model. This signifies that there exists an additional ratio of true negatives that are incorrectly predicted as positives and could be denoted as false positives. This proportion may also be known as the True Negative Rate (TNR). Eq (6) illustrates the formula of specificity

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

Matthews Correlation Coefficient

MCC is a highly regarded classification metric that effectively condenses the information presented by a confusion matrix or an error matrix. It incorporates four quantities: true positives (TP), true

negatives (TN), false positives (FP) and false negatives (FN). If the forecast yields favourable rates for all four of these entities, it is referred to as a dependable metric generating elevated scores. Moreover, it is tailored to accommodate the majority of correlation coefficients. A value of +1 signifies the ideal concurrence between the projected and factual values, whereas MCC denotes a spectrum spanning from +1 to -1. A value of 0 denotes an absence of concordance, suggesting that the prognostications are indiscriminate in relation to the genuine results. Eq (7) shows the formula of MCC

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (7)$$

Negative Predictive Value

NPV is a classification metric that is routinely used in medical diagnostics and other binary classification tasks. It assesses the probability that a negative prediction or test outcome is accurate. Eq (8) presents the formula of NPV

$$NPV = \frac{TN}{TN + FN} \quad (8)$$

False-Negative Rate

The probability that the test fails to identify a true positive result, thereby leading to its omission, is widely recognized in the scientific community as the false-negative rate. This particular metric, alternatively designated as the "miss rate," serves as a measure of how often a test may mistakenly categorize an actual positive outcome as negative. Eq (9) illustrates the formula of FNR

$$FNR = \frac{FN}{FN + TP} \quad (9)$$

False-Positive Rate

The false-positive rate is the proportion of actual negative cases incorrectly classified as positive. It is calculated by dividing the number of false positives by the total number of actual negative cases. Equation (10) gives the formula.

$$FPR = \frac{FP}{FP + TN} \quad (10)$$

Database Details

The first study reviewed here relies on a pre-existing database obtained through an information-fusion investigation of integrated navigation with GPS/INS [11].

In the second study, images were collected from the Fish-Pak dataset [6]. The Fish-Pak dataset contains a compilation of images representing six distinct fish species, namely Catla (Thala), Hypophthalmichthys molitrix (Silver carp), Labeo rohita (Rohu), Cirrhinus mrigala (Mori), Cyprinus carpio (Common carp), and Ctenopharyngodon idella (Grass carp). These images, totaling 915 in number, were captured using a Canon EOS 1300D camera from three different locations, namely Head Qadirabad, Head Marala, and the Chenab River in Gujrat, Punjab, Pakistan. The Fish-Pak data set proves useful for image classification tasks that rely on visual characteristics. It should be noted that this data set is particularly suited for multiclass problems due to its incorporation of three distinct visual features, namely Scale (color), Body, and Head.

In the third study, data were collected from simulation reports generated using the researchers' software [7].

Comparison of Fish Disease Detection Methods

Table 1: Comparison table

Work	Dataset used	Techniques used	Accuracy
Jhansi and Sujatha 2023 [6]	Fish-Pak Dataset	ANN	96.439
Jhansi and Sujatha 2023 [7]	Open-source Dataset	DCNN	97.8
Ahmed <i>et al.</i> 2022 [8]	Internet and aquaculture firms	SVM	94.12 with augmentation, 91.42 without augmentation.
Malik <i>et al.</i> 2017 [9]	-	Neural network and K-NN	86
Gupta et al. 2022 [10]	Images captured from the 2-D welfare camera	CNN	96

Table 1 reports higher fish-disease detection accuracy for the second and third studies than for the other listed methods. The second study used a machine learning-based ANN for early fish-disease detection and proposed a hybrid BWO-MA algorithm to address global optimization problems [6]. The third study employed a DCNN and a hybrid optimization approach combining the Whale

Optimization Algorithm with a Genetic Algorithm (WOA-GA) and Ant Colony Optimization [7]. In the study by Ahmed et al. [8], image preprocessing and segmentation were used to reduce noise and enhance the image. Relevant features were then extracted, and a [Support Vector Machine \(SVM\)](#) with a kernel function was used to classify diseases.

In the study by Malik et al. [9], image enhancement techniques, including segmentation, edge detection and morphological operations, were first applied to an image of a fish affected by Epizootic Ulcerative Syndrome (EUS). Feature descriptors, including Histogram of Oriented Gradients (HOG) and Features from Accelerated Segment Test (FAST), were then used to extract features. Images of EUS-infected and uninfected fish were classified using machine learning algorithms, and classification accuracy was evaluated. Gupta et al. [10] used an image-based machine learning method to detect wounds and lice in farmed salmon. Their proposed convolutional neural network contained 15 convolutional and five dense layers. Its reported test accuracy was 96.7%, compared with 91.2% and 92.81% for VGG-16 and VGG-19, respectively.

Testing and Training

In the first study, 233 images of EUS-infected and uninfected fish were collected. Of these, 173 were used for training and the remaining 60 for testing [11]. An Unscented Kalman Filter (UKF) was used to remove noise. Its performance was assessed through the standard deviation of the error, reported as ranging from 900×10^{-4} to 1800×10^{-4} . The study also described an Extended Kalman Filter (EKF) as improving classification efficiency through noise removal. MSE was used to evaluate model error, and the suggested combination achieved a reported classification efficiency of 98%.

In the second study, data were collected from the Fish-Pak dataset [6]. It contains 915 images of six fish species: Catla (Thala), Hypophthalmichthys molitrix (silver carp), Labeo rohita (rohu), Cirrhinus mrigala (mori), Cyprinus carpio (common carp) and Ctenopharyngodon idella (grass carp). The dataset includes three visual features: scale colour, body and head. Twenty percent of the images were used for testing and the remaining 80% for training. The proposed model was evaluated using accuracy, precision, recall, F1-score and specificity. The reported values were 96.439, 95.341, 97.421, 94.264 and 95.243, respectively, exceeding those of the comparison methods.

In the third study, images were collected from an open-source database, with 80% used for training and 20% for testing [7]. The model was evaluated using accuracy, precision, sensitivity, specificity, F-measure, MCC, recall, NPV, FNR and FPR. The proposed H-RF model achieved a reported accuracy of 97%, higher than the comparison SVM, Naïve Bayes and Random Forest models. It also recorded lower FPR and FNR values of 1.21 and 8.47, respectively.

Summary

This paper has examined MATLAB as a tool for fish-disease detection and explained the performance metrics used to evaluate the reviewed models. It has described the datasets, compared the proposed approaches with existing methods and examined the training and testing procedures. Together, these elements provide a basis for understanding how the studies evaluated their detection systems.

References

- [1] Sikder, J., Sarek, K.I. and Das, U.K., 2021. Fish disease detection system: a case study of freshwater fishes of Bangladesh. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, 12(6), pp.867-871.
- [2] Hasan, N., Ibrahim, S. and Aqilah Azlan, A., 2022. Fish diseases detection using convolutional neural network (CNN). *International Journal of Nonlinear Analysis and Applications*, 13(1), pp.1977-1984.
- [3] Lyubchenko, V., Matarneh, R., Kobylin, O. and Lyashenko, V., 2016. Digital image processing techniques for detection and diagnosis of fish diseases.
- [4] Moler, C. and Little, J., 2020. A history of MATLAB. *Proceedings of the ACM on Programming Languages*, 4(HOPL), pp.1-67.
- [5] Yao, J. and Shepperd, M., 2021. The impact of using biased performance metrics on software defect prediction research. *Information and Software Technology*, 139, p.106664.
- [6] Jhansi, G. and Sujatha, K., 2023. An Enhanced Machine Learning Technique with a Hybrid Metaheuristic Algorithm to Identify the Fish Disease. *Wireless Personal Communications*, pp.1-21.
- [7] Jhansi, G. and Sujatha, K., 2023. HRFSVM: identification of fish disease using hybrid Random Forest and Support Vector Machine. *Environmental Monitoring and Assessment*, 195(8), p.918.
- [8] Ahmed, M.S., Aurpa, T.T. and Azad, M.A.K., 2022. Fish disease detection using image-based machine learning technique in aquaculture. *Journal of King Saud University-Computer and Information Sciences*, 34(8), pp.5170-5182.
- [9] Malik, S., Kumar, T. and Sahoo, A.K., 2017, August. Image processing techniques for identification of fish disease. In *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)* (pp. 55-59). IEEE.
- [10] Gupta, A., Bringsdal, E., Knausgård, K.M. and Goodwin, M., 2022. Accurate wound and lice detection in Atlantic salmon fish using a convolutional neural network. *Fishes*, 7(6), p.345.

[11] Jhansi, G., Sujatha, K., Ponmagal, R.S., Anand, M. and Srividhya, V., 2019. Diagnosis of Fish Disease using UKF and Elman Neural Networks. International Journal of Engineering and Advanced Technology, 8(3S), pp.438-441.