Detection of Fish Disease Using Machine Learning Techniques

Mr.K.Saidi Reddy[1] P.Swapna[2]
[1][2]Assistant Professor, Department of Information Technology, MREC (A), Hyderabad-500100

Abstract

In tank farming, fish infections pose a serious threat to the safety of the food source. It is still hard to spot sick fish in aquaculture at an early stage due to a lack of appropriate infrastructure. To halt the spread of disease, it is critical to spot diseased fish as soon as possible. Our objective in this study is to learn more about the salmon fish sickness because salmon aquaculture accounts for 70% (2.5 million lots) of the market and is the fastest-growing food manufacturing system globally. Fish that have been exposed to various viruses can be identified with the aid of an artificial intelligence tool and excellent picture processing. This task consists of two separate components. In the main portion, noise reduction and picture enhancement have been associated with photo pre-processing and segmentation, respectively. After deleting the implicated features in the second phase, we classify the diseases using the machine learning Support Vector Machine (SVM) technique using a kernel function. The processed photos from the first section have been run through this (SVM) model. After that, we balance a thorough test of the suggested approach combination using the collection of salmon fish photographs used to evaluate the condition of the fish. The dataset used for this work is entirely fresh and includes and does not include photo enhancement. Based on the results, it has been determined that our SVM performs admirably with 91.42 and also 94.12 percent accuracy, respectively, with and without improvement.

Keywords: CF, CBF, RS, hybrid RS, MICRO BLOGGING DATA.

1. Introduction

The practice of breeding, raising, and collecting fish, water plants, crabs, mollusks, and other aquatic animals is referred to as aquaculture. Figure 1 illustrates how it entails raising both freshwater and saltwater animals in a controlled setting for the production of both food and commercial commodities. There are essentially two types of tank farming. The first is agriculture, which is the practice of raising aquatic animals for food production, cosmetics, pharmaceuticals, artificial flavorings, and jeweler (such as cultivated pearls). In addition to their natural aquatic habitat, marine microorganisms are raised in cages, ponds, or other land- or sea-based constructions. Algae, mollusks, shrimp, aquatic fish, and a variety of other tiny species, such as sea cucumbers and sea horses, are among the various microorganisms that are being raised on shorelines all over the world. It helps the development of sustainable foods in addition to fostering local economies. However, a variety of marine firming practices can occasionally pose risks to marine and coastal environments, including the degradation of natural habitats, the release of nutrients and waste, the unintentional release of alien microorganisms, the transmission of diseases to wild resources, and changes to local and indigenous communities.

The second method is fish farming, which involves growing fish for sale in artificial tanks and other locations. Common fish species like catfish, tilapia, salmon, carp, cod, and trout are frequently raised in these cages. To supply the market with fish items the fish farming industry has grown in the contemporary era [34]. This kind of tank farming is well-liked since it is supposed to generate a cheap source of lean protein.

Aquaculture will produce more than 53% of all fish, invertebrate, and algae produced globally by 2020. Additionally, it would produce 97 percent of all food. The expected market global production of farmed salmon rose to slightly over 2.6 million tons in 2019, an increase of 7%. [12] Several variables that have the potential to fully wipe out global salmon farming are putting the traditional salmon industrial sector in peril.

Fish can be severely harmed by infections both in the wild and in tank farms. According to universal consensus, sickness is one of the major red flags for the financial viability of aquaculture. Fish can become ill from a wide range of infectious agents, including bacteria, viruses, protozoa, and metazoan bloodsuckers. Microbes are the primary source of the infectious sickness that frequently affects limited fish [26]. Infectious diseases are one of the major threats to productive tank farming. Because there are so many different kinds of fish in a short area, an ecological community is formed that is favorable to development and quickly spreads infectious diseases. In this crowded habitat, which is largely artificial, fish experience anxiety and respond accordingly? The water community and poor water circulation can significantly aid in the transmission of viruses in densely populated areas [29]. With the aid of specific image processing, disease identification can be aided by the elimination of desirable qualities.

Photo division becomes crucial for a variety of study subject disciplines, including computer vision, artificial intelligence, etc. The term "k implies division" refers to a typical technique for breaking apart distinct portions of a photograph without losing information. The authors of [18] used the k indicates division approach for visual inspection. K proposes that segmentation is employed in another application, as detailed at [11], to detect transcribed Hindi personalities.

One of the most well-known tracked artificial intelligence techniques, support vector maker (SVM), has actually provided practical solutions for many classification issues across several fields. It is an effective classification tool that has a focus on accurate predictions for unlabeled data. Three kernel features were used by the authors of [19] to create an SVM design that could distinguish between healthy and dengue-human blood serums. It has already been suggested that SVMs should be used for the photo category in [3], where they replicate the architecture by combining SVM and convolution semantic networks (CNN). SVM delivers very accurate outcomes in many scenarios.

2. Literature Survey

1. The extended cubic b-spine interpolation technique was used to solve a two-point, straight-line border worth problem.

The authors are NUR NADIAH ABD HAMID *, AHMAD ABD. MAJID and AHMAD IZANI MD.ISMAIL to solve issues involving second order linear two-point limit values, extensive cubic B-spine interpolation was applied. The extended cubic B-spine, a modification to the cubic B-spine, has one form specification. The estimated analytical service that would be generated for the problems would depend on to obtain the optimal value those results in the best fit; the differential equations in the issues were optimized. This method closely approximated the answers to the issues compared to alternatives based on finite distinction, finite aspect, limited volume, and cubic B-spine interpolation.

2. An 8-bit Bayer pattern rib shade area is interpolated into a 24-bit ice xyz shade area using a median computation-based method. The author is T. Chary. What is described is an integrated method and tool for shade area conversion and interpolation. With this technique, an XYZ room image may be created directly from a raw image with a Bayer pattern where each pixel only has one of the color components needed to create a

full shade resolution pixel, without the need for any additional conversion or interpolation procedures. In one specific scenario, a raw 8-bit Bayer pattern picture might be instantaneously transformed to a 24-bit XYZ space using a single pass. A technique that entails combining the methods for color interpolation and color space conversion; choosing a missing color component for a location in a pixel by calculating the mean of pixels close by that have the same shade as the missing color component; using the claimed technique on the pixels in a raw image; and finally, changing the pixels' color space.

3. A method of categorizing images that blends convolution neural networks with assistance vector machines (SVM) (CNN).

A. F. AGARAP IS THE AUTHOR.

Convolution neural networks (CNNs) are comparable to "ordinary" neural networks in that they are composed of hidden layers made up of nerve cells with "learnable" parameters. These nerve cells take in information, produce a dot product, and then respond non-linearly. The complete network expresses the mapping between raw image pixels and their course grades. The Soft ax feature is frequently utilized as the classifier at the network's lowest layer. Studies have been conducted to evaluate this standard, nevertheless (Alalshekmubarak and Smith, 2013; Agora, 2017; Flavor, 2013). Simple support vector machines (SVM) were first used in the mentioned research' synthetic semantic network architecture. Another assignment dealing with the topic is this one, which is driven by (Flavor, 2013). Using the MNIST dataset, empirical results show that the CNN-SVM version was able to obtain a test precision of less than 99.04 percent (Lacuna, Cortes, as well as Burges, 2010). However, the CNN-Soft ax was able to achieve an examination precision of 99.23% using the same dataset. Both models were also tested on the Fashion-MNIST dataset (Xiao, Result, and Volga, 2017), which is a more difficult picture category dataset than MNIST (Zalandoresearch, 2017). This was proved by the difference in test precision between CNN-SVM and CNN-Soft ax, which was 90.72 percent for CNN-SVM and 91.86 percent for CNN-Soft ax. If datasets were subjected to information preprocessing techniques and a basic CNN architecture that was significantly more advanced than the one utilized in this work, the results indicated above might be enhanced.

4. A user-friendly summary of support vector devices.

The authors are J. Weston and A. Ben-Hurl. The Support Vector Machine (SVM) is a well-liked classifier in bioinformatics. Gaining the most from SVMs requires an understanding of how they operate and how users can affect their accuracy. We emphasize the practical use of SVMs while simultaneously giving the user a fundamental understanding of the theory that underpins them. We discuss the effects of the SVM parameters on the final classifier, how to determine optimal values for those parameters, how to normalize data, what affects training time, and how to employ SVM training software.

5. An individual's first exposure to support vector machines.

A novel technique for improving rib to xyz based on enhanced pattern search the writers are S. Bianca, F. Aspirin, A. Russo, and R. Stettin. It is advised to employ a cunning strategy when inspecting the color of biscuit products. Under-baked, decently baked, over-baked, and severely over-baked biscuits were divided into these four groups in this method, which made use of sophisticated categorizing techniques based on Walk’s evaluation and Assistance Vector Machines (SVM). Both straight and multi-step categories were used to evaluate the system's accuracy against conventional discriminate evaluation. The Walk’s radial basis SVM was shown to be significantly more accurate than other classifiers. Real-time execution was made possible by multi-core CPUs with sophisticated multiple-buffering and multithreading formulas.

At a rate of more than 96%, the system properly identified both stationary and moving biscuits travelling at a speed of 9 m/min. It was discovered that whether or not biscuits were touched had no discernible effect on the accuracy of baking evaluation. Nevertheless, photo handling for cookies that were touched took far longer than it did for cookies that were not touched, taking an average of 36.3 ms and 9.0 ms, respectively. The rate reduction was brought on by the convoluted watershed-based formula used to split touched cookies. This image computing technology might be able to enable the high-volume biscuit manufacturing. Research conclusions An SVM-based algorithm with direct and non-linear kernels has been developed for real-time analysis of cookies on a moving conveyor belt. The contacting and overlapping items in pictures are handled using Otsu's formula and watershed enhancement. Parallelism is made possible through the use of multi-threading methods and multi-core CPUs. For fixed and moving biscuits travelling at 9 m/min, the system's right category price is greater than 96%.

3. Methodology

A few jobs were completely devoted to a few standard photo-editing procedures for assessing fish health. Image-based detection was proposed by shaved et al. [22] using picture division as a side detection using Smart, Prewitt, and Sober. However, they did not go into great detail on how exactly they got rid of the trait. Using a Pie chart of Slope (HOG) and Features from Accelerated Section Examination, they merged both classification strategies for attribute reduction (QUICK). They attempted to blend categories to create a far better method rather than employing a specific one with less accuracy. Another method suggested by Lyubchenko et al. [21] is the grouping of the objects in the image. Different photo segmentation tasks based on a scale of many collections were required for this framework. Here, they selected pens for both personal items and items that a specific pen might come into contact with. They ultimately estimated the proportion of an object in the image as well as the proportion of an infected area to the fish body in order to identify fish disease. However, private marking a product is time-consuming and ineffective.

Some techniques concentrate on fusing image processing and artificial intelligence. Mali et al. [3] recommended using a unique fish disease called Epizootic Ulcerative Syndrome (EUS) discovery approach. The fungus Aphanomyces invades is the cause of this illness. In the architecture shown below, they attempted to combine the Attributes from Accelerated Section Examination (QUICKLY) feature detector with the Principal Component Analysis (PCA) and Histogram of Oriented Gradients (HOG) feature detectors (semantic network). FAST-PCANN and HOG-PCA-NN work together to produce a classifier accuracy of 86 percent, as opposed to the previous mix's precision of 65.8 percent.

3.1. The Dies-Benefits

1. The security of the food supply is seriously threatened by the fish conditions in tank farming.

2. Swift identification of diseased fish is necessary to stop the spread of the illness. Salmon tank farming makes up 70% (2.5 million tons) of the market and is the fastest-growing technique of food production in the world. We are therefore curious to find out more about the salmon fish situation in aquaculture.

3. Fish that have been polluted with numerous germs are identified using a machine learning method in conjunction with flawless image processing.

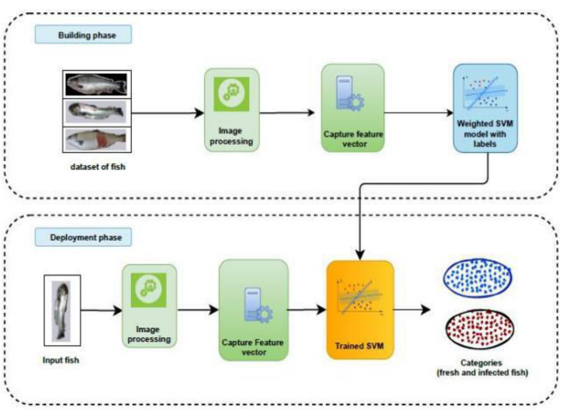
Support vector machine (SVM), one of the most popular artificial intelligence approaches, has made it simple to resolve a lot of classification problems in a variety of sectors. It is a potent classification method that generates superb predictions for unlabeled data. Authors in [19] developed an SVM architecture based on 3-bit features to differentiate between balanced and healthy lotions and blood products tainted with the dengue virus. [3] Suggests another SVM technique for categorizing photos, mimicking the approach by fusing SVM with convolution neural networks (CNN). In many different scenarios, SVM provides astonishingly accurate results.

3.2. Benefits:

1. The security of the food supply is seriously threatened by the fish conditions in tank farming.

2. It is still challenging to learn how to recognize diseased fish in aquaculture at an early stage due to a lack of appropriate equipment.

3. Finding contaminated fish as soon as feasible is crucial to halt the spread of a problem. Our objective in this study is to comprehend the salmon fish condition in tank farming since salmon aquaculture, which accounts for 70% (2.5 million tons) of the market, is the food manufacturing system with the quickest rate of growth internationally.



4. Result Analysis

1) **Submit the Fish Dataset**: We'll use this module to do so.

2) Perform CLAHE, LABORATORY, as well as Interpolation:
With the aid of this component, we will read every image, perform interpolation, process it using CLAHE and LAB, normalize it, and then divide the dataset into train and also test sets.
Start the decision tree.

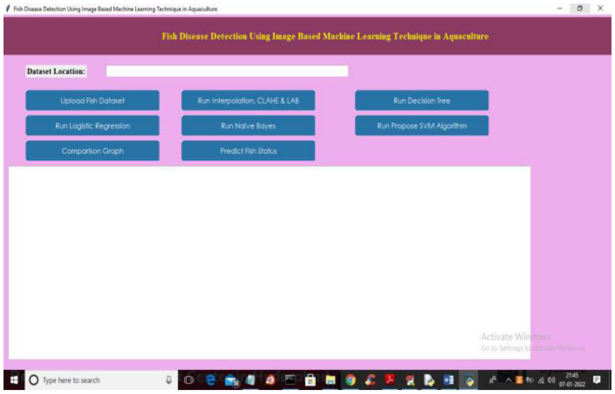
A decision tree will be trained using processed training data, and it will then be used to evaluate data in order to determine prediction precision and other metrics.

Step 4 of running logistic regression:
Images will be examined using the advanced version of logistic regression to determine prediction accuracy and other parameters. Certainly, the input will be refined train photos.

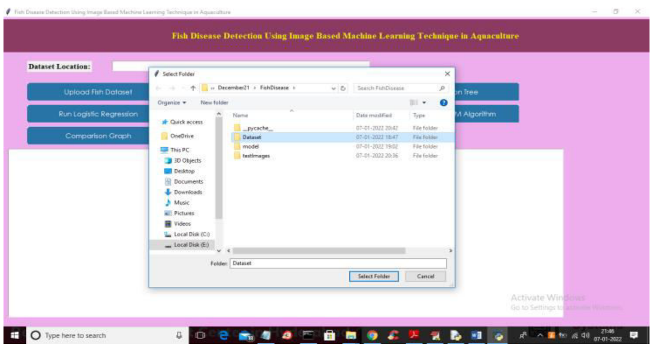
Run Negligible Bays (5):
In order to train a model, processed train images are input into an ignorant bays algorithm, which is subsequently applied to inspect images to calculate metrics such as forecast accuracy.

The suggested SVM formula should be run:
The SVM method will obtain the enhanced train images in order to train a model, which will then be used to the examined train photos in order to calculate forecast precision and numerous other metrics.
7) A graph with comparisons
With this module, we will generate charts for precision and a number of other data.

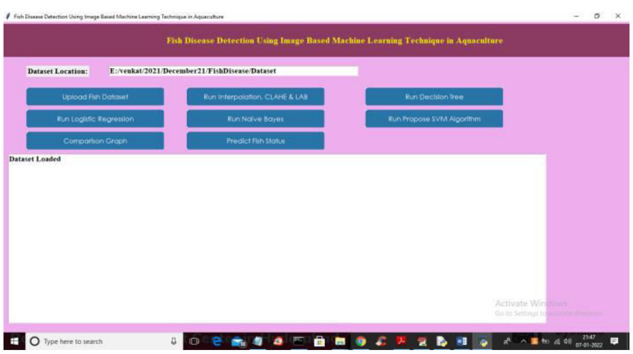
8) Determine Fish Position:
With the help of this module, we may upload a test image and the SVM algorithm will absolutely determine whether the image contains fresh fish.



In above screen click on 'Upload Fish Dataset' button to upload dataset and to get below screen

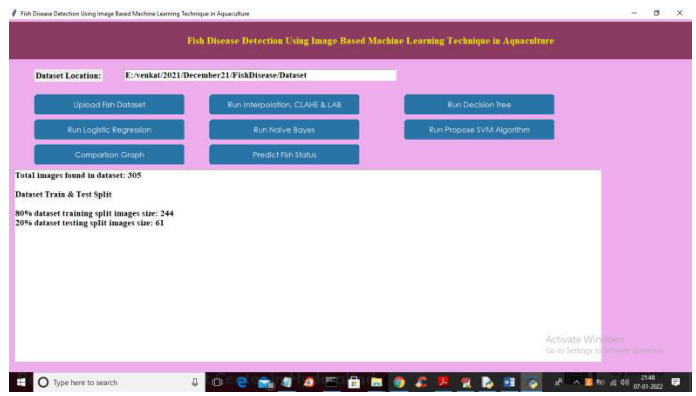


In above screen selecting and uploading 'Dataset' folder and then click on 'Select Folder' button to load dataset and to get below screen

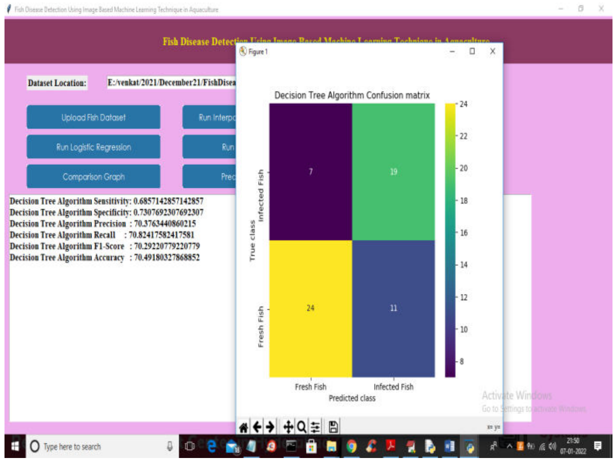


In above screen dataset loaded and now click on 'Run Interpolation, CLAHE & LAB' button to apply all 3 techniques and to process images and then split dataset into train and test

In above screen we can see all images are process using 3 techniques and after processing image will be converted to above format and now close above image to get below screen

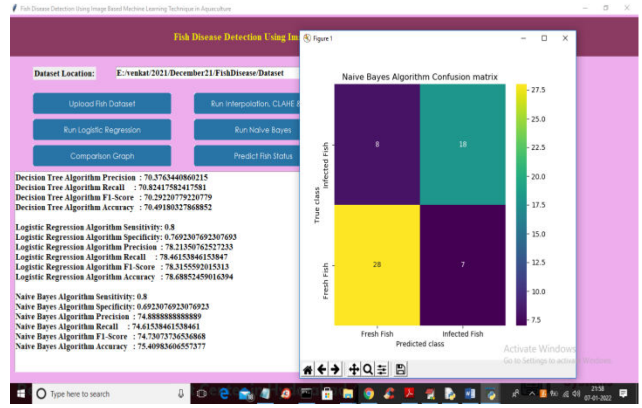


The dataset, which is represented in the graphic above, initially contained 305 images that were later divided into training and testing sets, each consisting of 244 and 61 images. After gathering the process images and test data, click "Run Decision Tree" to train the decision tree and display the screen.

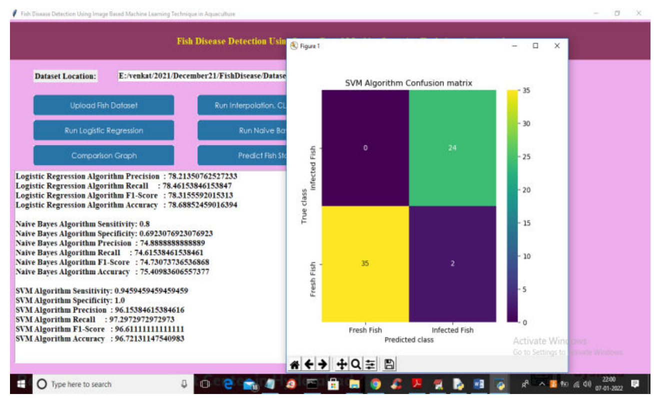


With the decision tree in the previous screen, we were able to achieve a 70% accuracy rate. We can also see various statistical values and a confusion matrix chart where we can see that 24 of the photographs were correctly predicted as fresh fish and 7 were mistakenly predicted as contaminated fish. Close the previous graph now, and then click the "Run Logistic Regression" button to teach the formula and obtain the results shown below.

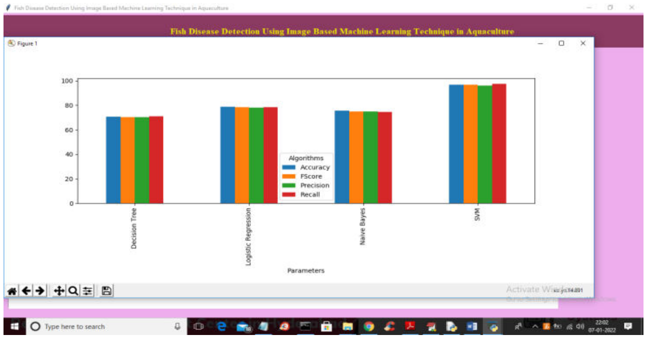
In the screen above, logistic regression offered us a 78 percent accuracy rate. In the confusion matrix, we can see that 28 predictions for the fresh fish picture were correct while only 6 were incorrect. Shut the above graph now, and then click "Run Nave Bays" to educate Nave Bays as well as bring up the below-screen home window.



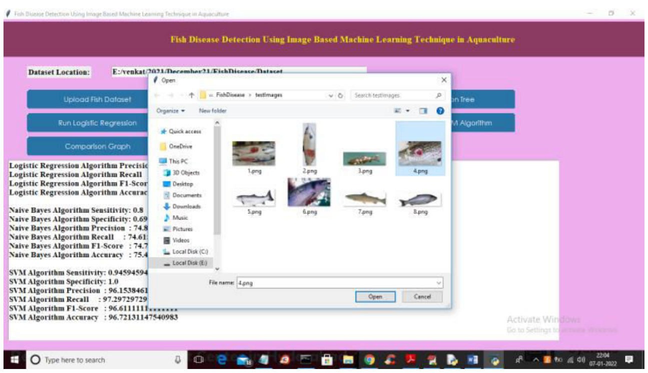
Close the graph in the above page, and then click the "Run Propose SVM Formula" button to train SVM and acquire the results shown below. In the confusion matrix in the above screen, 28 photographs are correctly anticipated as fresh fish, while 8 pictures are mistakenly predicted.



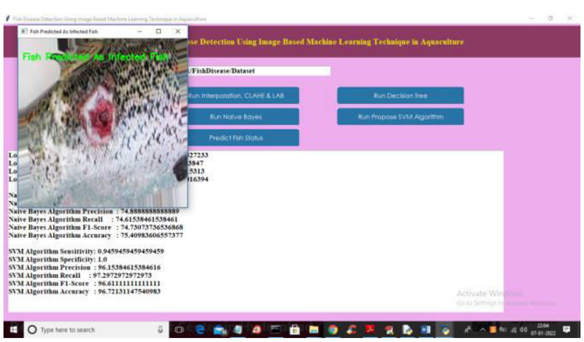
In the image above, we achieved 96 percent precision with SVM, and the complication matrix graph reveals that 35 of the predictions were true and 0 were mistaken. SVM is therefore superior to a number of other strategies. Click the "Contrast Chart" option to display the comparison graph below.



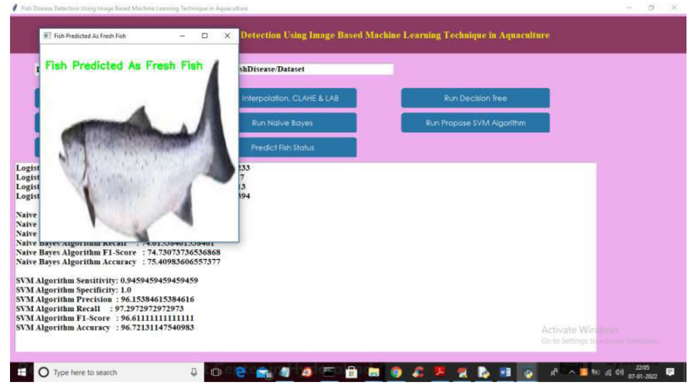
The method names on the x-axis of the above graph are represented, while the accuracy, precision, and other metrics are displayed on the y-axis as different colored bars. SVM is performing well in the above chart, which is currently closed. After that, click the "Predict Fish Standing" button to submit a test image and to receive a forecast.



In above screen selecting and uploading '4.png' file and then click on 'Open' button to load image and to get below prediction



In above screen fish predicted as INFECTED FISH and now test other images



In above screen uploaded fish predicted as FRESH similarly you can upload other images and test.

5. Conclusion

In this study work, we present a substantial machine learning-based classification design (SVM) for identifying sick fish. Our model is trained to be fresh and also unique using the real-world sans enriched dataset (163 polluted and 68 fresh) (785 contaminated and 320 fresh). The two primary categories of fish are fresh fish and another kind known as contaminated fish. A variety of metrics are used to test our model, and the classified results of those evaluations are additionally shown with visual interactivity. We also enhanced photo processing techniques including flexible histogram equalization, cubic spine interpolation, and k-means segmentation to make our input image more adaptive for our classifier. Additionally, we contrast the outcomes of our design with three different category versions and find that our proposed classifier performs better in this circumstance.

This study contributes to the development of an automated fish detection system that outperforms previous techniques that rely on image processing or are less accurate. We combine the most up-to-date photo processing methodology with the suitable supervised learning approaches. For our brand-new real-world dataset, we successfully create a classifier that can forecast sick fish with greater accuracy than rival methods.

We intend to employ many convolution neural networks in the future to more thoroughly and accurately recognize fish condition (CNN). We will undoubtedly concentrate on creating a genuine Iota device using the provided framework. Following this approach allows aquaculture farmers to identify unwell salmon fish and take the appropriate action before suffering an unanticipated loss in their farming. To greatly expand the applicability of our technique in several additional tank farming businesses, we will work with numerous fish datasets. Because salmon fish is one of the most popular products worldwide, we will also focus on expanding our current dataset.

References

[1]. A. A. M. Bad Humid, N. and A. Isanti. Extended cubic b-spine interpolation method applied to linear two-point boundary value problem. World Academy of Science, 62, 2010.

[2]. T. Acharya. Median computation-based integrated color interpolation and color space conversion methodology from 8-bit bayer pattern rgb color space to 24-bit cie xyz color space, 2002. US Patent 6,366,692.

[3]. A. F. Agarap. An architecture combining convolutional neural network (cnn) and support vector machine (svm) for image classification. arXiv preprint arXiv:1712.03541, 2017.

[4]. A. Ben-Hur and J. Weston. A user's guide to support vector machines. In Data mining techniques for the life sciences, pages 223–239. Springer, 2010.

[5]. S. Bianco, F. Gasparini, A. Russo, and R. Schettini. A new method for rgb to xyz transformation based on pattern search optimization. IEEE Transactions on Consumer Electronics, 53(3):1020–1028, 2007.

[6]. E. Bisong. Google colaboratory. In Building Machine Learning and Deep Learning Models on Google Cloud Platform, pages 59–64. Springer, 2019.

[7]. A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7):1145– 1159, 1997.

[8]. S. A. Burney and H. Tariq. K-means cluster analysis for image segmentation. International Journal of Computer Applications, 96(4), 2014.

[9]. M. A. Chandra and S. Bedi. Survey on svm and their application in image classification. International Journal of Information Technology, pages 1–11, 2018.

[10]. L. de Oliveira Martins, G. B. Junior, A. C. Silva, A. C. de Paiva, and M. Gattass. Detection of masses in digital mammograms using kmeans and support vector machine. ELCVIA Electronic Letters on Computer Vision and Image Analysis, 8(2):39–50, 2009.