DISCOVERY OF FISH INFECTION USING ML METHOD

[1]Mr.G.Gopala Krishna,[2]Natte Namratha, [3]Maale Arpitha, [4]Palle Maheshwari, [5]Gantlapu Ramya

[1]Assistant Professor,[2 3 4 5]B.Tech.Scholar

Dept of Computer Science and Engineering

J.B.Institute of Engineering and Technology Moinabad, Hyderabad, Telangana, India

ABSTRACT

Fish diseases in aquaculture represent a major threat to food security. Identification of infected fish in aquaculture remains difficult to detect at an early stage due to the scarcity of the necessary infrastructure. Timely identification of infected fish is a mandatory step to prevent the spread of the disease. In this work, we want to detect salmon disease in aquaculture, since salmon farming is the fastest growing food production system globally and represents 70 percent (2.5 million tons) of the market . Thanks to the alliance of impeccable image processing and a machine learning mechanism, we were able to identify infected fish caused by various pathogens. This job is divided in two parts. In the primitive part, image preprocessing and segmentation were applied for noise reduction and image amplification, respectively. In the second part, we extract the respective features for disease classification with the help of machine learning support vector machine (SVM) algorithm with kernel function. The processed images from the first part went through this SVM model. We then adapted a comprehensive experiment to the set of techniques proposed on a salmon image dataset used for fish disease detection. We have moved this work to a new data set that compromises both image augmentation and without. The results have given a verdict on the performance of our implemented SVM. Significantly with an accuracy of 91.42 and 94.12 percent, respectively, with and without magnification.

I. INTRODUCTION

The word aquaculture is related to firming, including breed- ing, raising, and harvesting fishes, aquatic plants, crustaceans, mollusks, and aquatic organisms. It involves the cultivation of both freshwater and saltwater creatures under a controlled condition and is used to produce food and commercial prod- ucts as shown in Figure 1. There are mainly two types of aquaculture. The first one is Mariculture which is the farm- ing of marine organisms for food and other products such as pharmaceuticals, food additives, jewelry (e.g., cultured pearls), nutraceuticals, and cosmetics. Marine organisms are farmed either in the natural marine environment or in the land- or sea-based enclosures, such as cages, ponds, or raceways. Seaweeds, mollusks, shrimps, marine fish, and a wide range of other minor species such as sea cucumbers and sea horses are among the wide range of organisms presently farmed around the world's coastlines. It contributes to sus- tainable food production and the economic development of local communities. However, sometimes at a large scale of marine firming become a threat to marine and coastal envi- ronments like degradation of natural habitats, nutrients, and waste discharge, accidental release of alien organisms, the transmission of diseases to wild stocks, and displacement of local and indigenous communities.

The second one is Fish farming which is the cultivation

of fish for commercial purposes in human-made tanks and other enclosures. Usually, some common types of fish like catfish, tilapia, salmon, carp, cod, and trout are firmed in grown to meet the demand for fish products This form of aquaculture is widespread for a long time as it is said to produce a cheap source of protein.

productions, accounting for almost 53% of all fish and in- vertebrate production and 97% of the total seaweed manu- facture as of 2020. Estimated global production of farmed salmon stepped up by 7 percent in 2019, to just over 2.6 mil- lion tonnes of the market [12]. Global aquaculture of salmon has a threat of various diseases that can devastate the conventional production of salmon.

Diseases have a dangerous impact on fishes in both the natural environment and in aquaculture. Diseases are glob- ally admitted as one of the most severe warnings to the economic success of aquaculture. Diseases of fishes are provoked by a spacious range of contagious organisms such as bacteria, viruses, protozoan, and metazoan parasites. Bacteria are accountable for the preponderance of the contagious diseases in confined fish Infectious diseases create one in every foremost vital threat to victorious aquaculture. The massive numbers of fishes gathered in a tiny region gives an ecosystem favorable for development and quickly spreads contagious diseases. In this jam-packed situation, a comparatively fabricated environment, fishes are stressed and also respond to disease. Furthermore, the water ecosystem and insufficient water flow make it easier for the spread of pathogens in gathered populations. Detection of dis- ease with the cooperation of some image processing can help to extract good features.

Image segmentation becomes indispensable for various

Research fields like computer vision, artificial intelligence, etc. The k means segmentation is a popular image processing

Figure 1: Aquaculture

Technique that mainly partitions different regions in an image without loss of information. In [18], authors applied k means segmentation for authentication of images. Another application of k means segmentation shown at [11] where they use this technique to recognize handwritten Hindi char- acters.

One of the most popular supervised machine learning techniques, support vector machine (SVM), has brought convenient solutions for many classification problems in various fields. It is a powerful classification tool that brings out quality predictions for unlabeled data. In Authors built an SVM model based on three kernel functions to differentiate dengue human infected blood sera and healthy sera. For im- age classification, another SVM architecture has been pro- posed in [3] where they emulate the architecture by combining convolutional neural network (CNN) with SVM. SVM provides remarkable accuracy in many contexts we conduct our research on the salmon fish disease classification, either the fish has an infection or not, with a machine vision-based technique. A feature set is a trade-off for the classification of the disease. Image processing techniques are used to extort the features from the images, then a support vector machine (SVM) is employed for the successful classification of infectious disease. Hither, we summarize the entire concept of this work's contribution given below:

- Propose a groundbreaking framework for fish disease detection based on the machine learning model (SVM).
- Appraising and analyzing the performance of our pro- posed model both with and without image augmentation.
- Juxtaposing our proposed model with a good perform- ing model by some evaluation metrics.

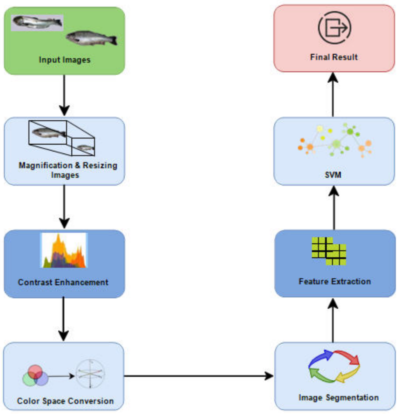
1.1BACKGROUND

Some works focused on only some basic image processing techniques for the identification of fish disease. proposed an image-based detection technique where firstly applies image segmentation as an edge detection with Canny, Prewitt, and Sobel. However, they did not specify the exact technique that engrossed for feature extraction. In feature extraction, they applied Histogram of Gradient (HOG) and Features from Accelerated Segment Test (FAST) for classification with a combination of both techniques. They tried to discover a better classification with a combination instead of applying a specific method with less exactness. Another technique Lyubchenko et al. proposed a structure called the clustering of objects in the image that obliged diverse image segmentation actions based on a scale of various clusters. Here, they chose markers for individual objects and objects encountered with a specific marker. Finally, they calculated the proportion of an object in the image and the proportion of infected area to the fish body to identify fish disease. However, individual marking of an object is time- consuming and not effective.There are some approaches focused on the consolidation of image processing and machine learning. Malik et al.proposed a specific fish disease called Epizootic Ulcerative Syndrome (EUS) detection approach. Aphanomyces invadans, a fungal pathogen, cause this disease. Here, they approached combination styles that combine the Principal Component Analysis (PCA) and Histogram of Oriented Gra- dients (HOG) with Features from Accelerated Segment Test (FAST) feature detector and then classify over machine learn- ing algorithm (neural network). The sequence of FAST-PCA- NN gives 86 percent accuracy through the classifier, and HOG-PCA-NN gives 65.8 percent accuracy that is less than the previous combination proposed a sensitive topic that is kidney stone detection. In this paper, the authors apply morphological operations and segmentation to determine ROI (region of interest) for the SVM classification technique. After ap- plying this technique, they investigated the kidney stone im- ages with some difficulties, such as the similarity of kidney stone and low image resolution. Zhou et al. introduced a device-free present detection and localization with SVM aid. Here, the detection algorithm can detect human pres- ence through the SVM classifier using CSI (channel state information) fingerprint. Trojans in hardware detection depend on SVM based approach. Here, the authors evaluated a trojans detection method with their designed hard- ware. For SVM analysis, their netlists consist of three types of hardware trojan with normal and abnormal behavior.

We can conclude that none has performed any depth re-search work on salmon fish disease classification regarding the research obligations described above. Furthermore, most of the research works involved typical fish disease classifica- tion but not in aquaculture. All those described techniques depend solely on image processing or a combination of im- age processing and machine learning technique but not up to the mark.

3. Preliminary and Proposed Framework

This section has several stages presented in Figure 2. Here we precisely present the appurtenant technologies and a solution framework of salmon fish disease classification.



. Figure 2: Proposed Framework (The overall anatomy of our proposed work gradually from input to result).

Cubic Splines Interpolation

Raw images appeared on the dataset in various sizes. If we do not resize these images before training the classifier, the classifier's efficiency may be decreased. As we collected these images from different sources, we reshape them beforeapplying them to the classifier.

For image magnification and fixed-size conversion, we

use an improved interpolation method called extended ***Cu-bic Splines interpolation*** For a finite interval [*a*, *b*], let

Given , the extended cubic B-spline function, is a linear com bination of the extended cubic B-spline basisfunction in,

$$S(x) = \sum_{i=-3}^{n-1} C_i EB_{3,i}(x), \quad x \; c \; x_0, x_n$$

Adaptive Histogram Equalization

Adaptive histogram equalization (AHE) is a vision pro-cessing approach used to enhance contrast in images. Here, we use an extension of AHE called contrast limited adaptive histogram equalization (CLAHE) . CLAHE diverges from conventional AHE in its contrast limiting. A user-defined value called clip limit where CLAHE limits the augmenta- tion by clipping the histogram. The amount of clamor in the histogram depends on the clipping level. Also, the smooth- ness and enhancement of contrast rely on this clipping level too. A modification of the limited contrast technique called adaptive histogram clip (AHC) can also be applied. AHC dynamically calibrates the clipping level and balanced over- enhancement of a background area of images. Here, we use one of the AHC called the Rayleigh distribution that forms a normal histogram.

Experimental Dataset

Since there is no accessible dataset of salmon fresh and infected fish, we prepared a shibboleth and novel dataset - some images from the internet and most of them from some aquaculture firms. The dataset contains images of fresh and infected salmon fish displays in Figure 5. We collect a total of 266 images that are used to train and validate our model. For training and testing, we split our dataset with a ratio of train and test data that depicts in. The total number of training and testing images are 231 and 35, respectively.Hence, our data acquisition is a complicated process. We apply the image augmentation technique in our dataset for expanding the dataset. Here we use the image_argumentor tool with some image augmentation operations such as Hor-izontal Fip (fliph), Vertical Flip (flipv), Rotates (rot), Pixel Shifting (trans), and Zooms (zoom). After the augmenta- tion operation, we perceive 1,105 training images depicted in Table 2. The total number of training and testing images is 1,105 and 221, respectively

Features Extraction

We consider two types of feature extraction techniques: one is statistical features, and another one is grey-level co- occurrence matrix (GLCM) features based on interpreting fish diseases. Statistical features.

Experimental Results

This section palpates our SVM model results to inspect our model's robustness and see the outcomes of our utilized techniques in both the regular and augmented datasets. Here, we present the actual upshots and comparisons with some graphical representations and fected fish class accuracy is 93.50% and 94.90%, higher than the fresh fish class. We also see the FPR and FNR in both classes, and the infected class shows slightly higher FNR and lower FPR than the fresh fish class. Here, the low percentage of FPR and FNR refers to our model is not underfit or overfit. So, as an individual class prediction, the infected fish class performs satisfactorily. We represent the two confusion matrix as a heat map for better graphical representation shown in Figure 7 for both with and without augmentation. This heat map can conve- niently display the classification and misclassification of our binary class. From this confusion matrix in we see that fresh fish is misclassified only two times with infected fishes, and infected fish is misclassified once with fresh fish. shows the misclassification of seven fresh fish with infected fish and six infected fish with fresh fish.

CONCLUSION

We present a machine learning-based classification model (SVM) to identify infected fish in this research work. The real world without enhanced data set (163 infected and 68 new) and the augmented data set (785 infected and 320 new) are used to train our new model. We basically classify fish into two individual categories: fresh fish and infected fish. We evaluate our model on different metrics and display the classified results through visual interaction of those classification results. In addition to developing our classifier, we apply updated image processing techniques such as k-means segmentation, cubic linear interpolation, and adaptive histogram equalization to transform our input image more adaptively to our classifier. We also compare the results of our model with three classification models and observe that our proposed classifier is the best solution in this case.

This work contributes to producing an automated fish detection system that is superior to existing systems that rely on image processing or lower resolution. We rely not only on cutting-edge image processing technology, but also on supervised and ordered learning techniques. We have successfully developed a classifier that predicts infected fish with the best accuracy rate compared to other systems for our new real-world data set.

In the future, we plan to use a different convolutional neural network (CNN) architecture to identify fish diseases with greater accuracy and precision. Furthermore, we will focus on the implementation of a real-world IoT device using the proposed system. Doing so can be a definitive solution for fish farmers to identify infected salmon and take appropriate measures before facing any unexpected loss in their culture. We will work with different fish data sets to make our system more usable in other sectors of aquaculture. We will also focus on increasing our existing data set, as salmon is a highly in-demand product around the world.

REFERENCES

[1] A. A. M. Abd Hamid, N. and A. Izani. Extended cubic b-spline inter- polation 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 ma- chines. 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 seg- mentation. 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 Technol- ogy, 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 k- means and support vector machine. ELCVIA Electronic Letters on Computer Vision and Image Analysis, 8(2):39–50, 2009.

[11] A. Gaur and S. Yadav. Handwritten hindi character recognition using k-means clustering and svm. In 2015 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services, pages 65–70. IEEE, 2015.