**A Machine Learning Approach for Early Detection of Fish Diseases by analyzing Water Quality**

**ABSTRACT:**

Fish diseases in aquaculture constitute a significant hazard to nutriment security. Identification of infected fishes in aquaculture remains challenging to find out at the early stage due to the dearth of necessary infrastructure. The identification of infected fish timely is an obligatory step to thwart from spreading disease. In this work, we want to find out the salmon fish disease in aquaculture, as salmon aquaculture is the fastest-growing food production system globally, accounting for 70 percent (2.5 million tons) of the market. In the alliance of flawless image processing and machine learning mechanism, we identify the infected fishes caused by the various pathogen. This work divides into two portions. In the rudimentary portion, image pre-processing and segmentation have been applied to reduce noise and exaggerate the image, respectively. In the second portion, we extract the involved features to classify the diseases with the help of the Support Vector Machine (SVM) algorithm of machine learning with a kernel function. The processed images of the first portion have passed through this (SVM) model. Then we harmonize a comprehensive experiment with the proposed combination of techniques on the salmon fish image dataset used to examine the fish disease. We have conveyed this work on a novel dataset compromising with and without image augmentation. The results have bought a judgment of our applied SVM performs notably with 91.42 and 94.12 percent of accuracy, respectively, with and without augmentation.

**EXISTING SYSTEM :**

The CNN system with perfect images without noise. Next step is to build our own dataset of fish in the ocean. Because it is difficult to obtain images from other kinds of objects such as the sea turtle, coral and so on. This part of the research, fish is the only object to be detected. For the collection of 410 images, many of them have multiple fish in one image, so the detection is challenging. The same method was chosen to create ground truth image. And all the parameters introduced before remain the same, only the class information is made up of a 1 × 1 × 18 vector instead of 1 × 1 × 30 because of reduce in the classes. Figure 11 illustrates one labelled image example. It is obvious that this data set is totally different from the ideal images from ImageNET.

**Disadvantages**

1. It takes more time
2. Less accuracy

**PROPOSED SYSTEM :**

The paper author is analysing water quality by applying machine learning algorithm to predict fish diseases. If water contains high toxins or viruses then it will affect fish lungs which causes disease inside fish or fish may die. So we can collect water samples and then apply on machine learning model to predict water quality and if quality is not good then we can say fish is not healthy. In propose paper author has used water quality dataset from KAGGLE website and then trained with Gradient Boosting algorithm. This algorithm giving more than 95% accurate prediction accuracy on test data.

**Advantages**

1.It takes less time

2.More prediction

**SYSTEM REQUIREMENT:**

**HARDWARE REQUIREMENTS:**

# Processor - I3(min)

* Speed - 1.1 GHz
* RAM - 4GB(min)
* Hard Disk - 500GB

**SOFTWARE REQUIREMENTS:**

* Operating System - Windows 10/above
* Programming Language - Python 3.7