ISSN [Online]: 2583-2654

An Automated Fish Illness Detection System from Under Water Using Machine Learning Technique Network

Mrs. SophiaS, Ptrajeeth S, Prathiksha B, Preethish R, S J Samskruthi
Department of CSE, Dr.T. Thimmaiah Institute of Technology, KGF, Karnataka, India

Abstract: The proposed Robotic Fish Illness Detection System (RFID) leverages machine learning to automate fish disease identification from images, operating through a concise five-stage process. Initially, raw images of fish are collected for analysis. These images then undergo pre-processing, where techniques like resizing and quality enhancement are applied to standardize and optimize them. In the feature extraction stage, pertinent features are extracted using methods such as edge and corner detection, pivotal for disease recognition. The system then undergoes rigorous training on labeled image data, associating features with known diseases to ensure accuracy. Utilizing CNN algorithms, the classification stage categorizes images based on extracted features, aided by the Hamming Distance algorithm for disease classification. Finally, the system determines whether the fish is diseased, offering a streamlined approach to fish health management in aquaculture.

Keywords: Robotic Fish Illness Detection System, preprocessing, feature extraction, machine learning training, CNN, classification, Hamming Distance.

I. INTRODUCTION

Aquaculture stands as a swiftly expanding sector in food production, entailing the controlled cultivation of both freshwater and saltwater creatures for commercial gain. Yet, accurate identification of fish species and prompt disease detection present formidable hurdles in this industry. Diseases, triggered by a spectrum of contagious organisms such as bacteria, viruses, and parasites, pose substantial threats to fish populations, whether in natural habitats or aquaculture environments. Among these pathogens, bacteria notably account for the majority of contagious diseases in confined fish.

In response to these pressing challenges, our project introduces an innovative system leveraging Convolutional Neural Network (CNN) algorithms for Fish Disease Detection. To underpin this endeavor, we meticulously assembled a comprehensive dataset comprising images of various fish species afflicted with diverse ailments. This dataset was meticulously crafted to encapsulate the real-world intricacies and variations encountered in aquaculture settings.

Our approach enhances disease detection through sophisticated image preprocessing techniques, including normalization, resizing, and noise reduction, which facilitate the extraction of salient features. The CNN model, rooted in deep learning principles, underwent rigorous training on this extensive dataset. During the testing phase, our system demonstrated remarkable accuracy in detecting fish diseases.

Furthermore, our project underscores the paramount importance of responsible data preprocessing and handling, emphasizing privacy and ethical considerations in the realm of advanced data-driven technologies. At the heart of our project lies the CNN model, serving as a cornerstone for accurately classifying whether a fish is diseased or non-diseased, thereby advancing disease management practices in aquaculture.

II RELATED WORK

Ahmed Waleed et al. [1] have developed a computer vision system for early detection of fish diseases in aquaculture. Their method analyzes abnormal fish behaviors via video trajectory analysis and employs various Convolutional Neural Network (CNN) architectures, with AlexNet showing promise

ISSN [Online]: 2583-2654

particularly in the XYZ color space. Future enhancements include integrating Raspberry hardware for temperature and pH monitoring, enabling GSM800 notifications, and expanding the dataset to include more fish diseases.

Yo-Ping Haung et al. [2] present an innovative hybrid model for automatic fish disease recognition from underwater images, achieving a remarkable accuracy of 94.28%. Their approach integrates multilayer fusion, attention mechanism, and online sequential extreme learning machine (OSELM) for improved feature extraction and faster learning. Future plans involve dataset expansion, comparative analysis with fuzzy logic, and refining the model for smart aquaculture applications, representing a significant step towards sustainable aquaculture practices.

Zhenxi Zhao,Yang Liu, et al. [3] implemented Composited FishNet framework, with its composite backbone (CBresnet) and enhanced path aggregation network (EPANet).By learning scene change information and reducing interference from underwater environmental factors, it strengthens the output of the main network for object information. The integration of high and low feature information through EPANet also enhances semantic information utilization.

Yinchi Ma et al. [4] created a real-time remote diagnosis expert system using modern internet technology. It allows for gathering on-site real-time videos during fish disease outbreaks and transmitting micro-dissection images to remote experts instantly. Following a B/S development model, users access it via the internet through a computer, featuring expert management, user management, and video transmission modules.

Shaveta Malik et al. [5] developed a comprehensive solution for early identification of EUS in fish caused by Aphanomyces invadans. Their approach combines image segmentation, edge detection, and morphological operations for initial processing, followed by HOG, FAST, and machine learning algorithms for classification, with PCA for dimensional reduction. Achieving superior accuracy with real EUS-infected fish images on MATLAB, future plans involve refining the methodology, exploring advanced techniques, incorporating diverse datasets, and collaborating with industry experts for real-world implementation and ongoing enhancement of fish disease identification in aquaculture.

Md Shoaib Ahmed et al. [6] developed a novel approach using image processing and machine learning to detect salmon fish diseases early in aquaculture. Achieving high accuracy rates of 91.42% and 94.12% on a unique dataset with and without image augmentation, their system shows promise for practical use. Future plans include refining the methodology, integrating diverse datasets, exploring advanced ML techniques, and collaborating with industry experts for real-world implementation, enhancing early disease identification in salmon aquaculture for food security.

S.N Pauzi et al. [7] conducted a systematic review of fish disease detection methods, exploring image processing techniques like rule-based expert systems, machine learning, and deep learning. Despite needing improvement, image processing offers a reliable alternative to manual detection, aiding in early disease identification and prevention. This review sheds light on the application of image processing in fish disease detection

.III METHODOLOGY

The proposed system begins by taking an input image set comprising both normal and affected fish images, which undergoes a series of preprocessing steps. This includes resizing, converting RGB to grayscale, and reducing noise to standardize and optimize the images. Subsequently, relevant features are extracted from the fish images, encompassing texture, color, shape, contour information, and other characteristics crucial for distinguishing between normal and infected cases. The model is then trained using a training dataset, wherein model parameters are adjusted to learn the patterns and features that differentiate healthy and diseased fish images. To ensure robust performance, the model's effectiveness is evaluated on a separate validation dataset to assess its generalization to new,

ISSN [Online]: 2583-2654

unseen data. Finally, the model's performance is rigorously assessed on the testing dataset, estimating metrics such as accuracy, precision, and recall, thereby validating its efficacy in accurately detecting fish diseases.

The automated fish disease detection system initiates with the input of fsh images.These images undergo preprocessing to enhance or modify images before feeding them into a machine learning algorithm. And the pre-processed images undergo feature extraction. Extracted features like shape, texture and color are then utilized to train a machine learning model, typically a Convolutional Neural Network. Trained on a labeled dataset, the model learns patterns and features. In application, the model classifies the fish images into diseased or non-diseased. The accuracy of this process relies on effective preprocessing, quality trainingdata, and the robustness of the trained model, collectively contributing to the system's efficacy

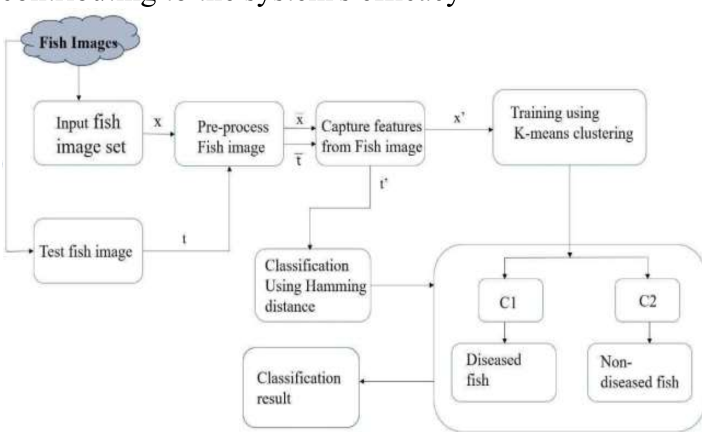

Fig.1. Proposed (RFID) System Architecture

.

A data flow diagram (DFD) is is realistic portrayal of the "stream" of information through a data framework. An information stream chart can likewise be utilized for the representation of information handling (organized plan). It is normal practice for a creator to draw a setting level DFD first which shows the cooperation between the framework and outside substances. DFD's show the progression of information from outer elements into the framework, how the information moves starting with one cycle then onto the next, as well as its sensible stockpiling. There are just four images: Squares addressing outside elements, which are sources and objections of data entering and leaving the framework. Adjusted square shapes addressing processes, in different strategies, might be called Exercises, Activities, Methodology, Subsystems and so forth which accept information as information, do handling to it, and result it. Bolts addressing the information streams, which is mathematical information

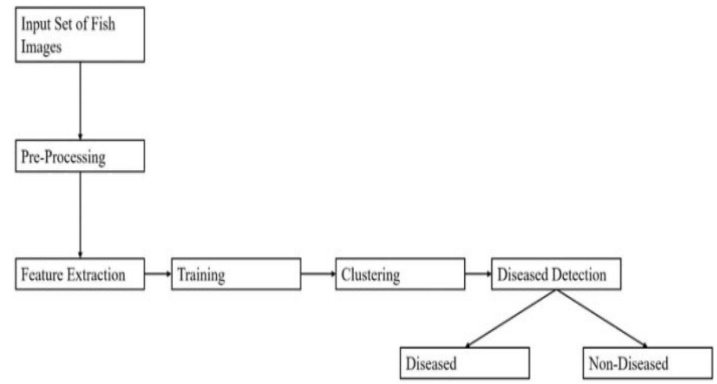

Fig. 2. Data Flow Diagram

A Data Collection

An Image of the form jpg will be taken as input. An input image set would refer to a collection of fish images of both diseased and non-diseased fish. These images are used as input to the data preprocessing phase, represented as *X*. The data would be preprocessed to remove noise, and other inconsistencies.

B Data Preprocessing

1. Data Cleaning

Data cleaning refers to the process of identifying and correcting errors, inconsistencies and inaccuracies within a dataset to improve its quality and reliability for analysis such as handling missing values that is dropping rows or columns with missing values, removing duplicates which means identifying and removing duplicate entries in dataset, consistency checks that is verifying data for consistency and rectifying any discrepancies.

2. Conversion to grayscale image

Converting an image to grayscale entails reducing it from RGB color, which includes RED, GREEN, and BLUE channels, to a single gray level. Grayscale images are simpler and require less memory compared

ISSN [Online]: 2583-2654

to RGB color images. This conversion task is common in image processing. Grayscale images utilize a single channel to represent intensity rather than separate color channels. Typically, this conversion involves taking a weighted average of the RGB channels to preserve luminance information. The resulting grayscale image represents the original in black and white, with pixel values indicating light intensity. This simplifies subsequent analysis and reduces computational complexity in certain applications.

$\bar{X}$ is the input given to convert into grayscale. The Technique used here is Luminosity
method. In this method, the grayscale intensity is calculated using weighted sum of its Red, Green and Blue values.
To convert $\bar{X}$ matrix into grayscale the mathematical equation is given by:

$$Xg = 0.2989\, R(\bar{X}) + 0.5870\, G(\bar{X}) + 0.1140\, B(\bar{X})$$

Where R, G and B are the red, green and blue channels in $\bar{X}$ matrix respectively.

3. *Reducing noise*

Noise is often considered as random or irrelevant information. Reducing noise in the context of image involves the process of minimizing or unwanted or irrelevant variations or disturbances in the image.

The technique used in reducing noise is Gaussian Blur. Gaussian blur is a popular image processing technique used for smoothing and noise reduction. It involves convolving the image with a Gaussian filter, which assigns weights to surrounding pixels based on a Gaussian distribution. This results in a softened appearance, reducing high-frequency noise and enhancing overall image quality. The mathematical equation of Gaussian Blur is given by:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

C Feature Extraction

In this stage, the proposed framework is at present centered around separating edge highlights from fish pictures determined to recognize remarkable examples that recognize ailing locales from non-sick regions utilizing convolution brain organization. In fish sickness recognition utilizing Convolutional Brain Organizations (CNNs), highlight extraction is a basic step. CNNs naturally gain applicable elements from input pictures. In this specific situation, the CNN layers recognize unmistakable examples or designs in fish pictures that are characteristic of sickness. These elements might incorporate shapes, surfaces, or varieties in power. The organization then utilizes these learned highlights to group districts of the picture as unhealthy or non-sick, making it an integral asset for robotizing the discovery interaction and further developing accuracy. Correlation Investigation.

$$z_i = \sum_{m=1}^{M} \sum_{n=1}^{N} x_i' . F + b$$

$$a_i = max(0, z_i)$$

$$\bar{X} = max_{(m,n) € \text{pooling window}} a_i$$

D Training

The training phase in machine learning involves teaching a model to make predictions or perform a task by exposing it to a labelled dataset. The goal is to train a model that can make accurate predictions on new, unseen data by learning patterns and relationships from the training set. In this stage, the proposed RFID system iteratively splits the input dataset into K dissimilar clusters based on K-Means technique and represents each individual cluster into separate data for analysis
process.

$$D(\bar{X}, \bar{V}) = \{\{d(\bar{X}_{ij}, \bar{V}_{rj})^{k-1, n-1, d-1}\} \forall v_r \in \bar{V}, \bar{V} \subseteq \bar{X}, \forall x_i \in \bar{X}, \forall x_{ij} \in x_i\}$$

$$d(x_{ij}, \bar{v}_{rj}) = \{[\sum^{d-0} (x_{ij} - \bar{v}_{rj})^2]^{1/2}\}$$

$$c_r = Min\{\{d(x_i, \bar{v}_r)^{k-1, n-1}\}, \forall c_r \subseteq \bar{X}\}$$

ISSN [Online]: 2583-2654

$$v_{\bar{r}} = \{\{\frac{1}{|C_r|} * \sum_{e=1}^{|C_r|} C_{re}\}, \forall C_{re} \in C_r\}$$

E Classification

As we delve into the robotic fish illness detection system, the Hamming distance emerges as a pivotal tool. This algorithm measures the similarity between feature vectors of fish images, enhancing the accuracy and robustness of the detection process. The Hamming distance calculation involves comparing binary strings bit by bit, tallying the differing positions. Mathematically, the Hamming distance (H) between two binary strings (x and y) of length n is computed as: : $H(x,y) = \sum_{i=1}^{n} |x_i - y_i|$ where x_i and y_i are the bits at the i-th position in the binary strings x and y, respectively. This calculated distance aids in identifying the individual fish. If the Hamming distance falls below a specific threshold, the fish are classified as non-diseased.

$$D = \sum_{i=1}^{n} diff(\bar{T}, \bar{X}_i)$$

IV RESULT AND DISCUSSION

1. Image Dataset

To create the image dataset for this project, images were collected from various sources, including online repositories, personal collections, and publicly available datasets. The dataset consists of 3500 images spanning different categories such Humans and Animals. The images were carefully curated to ensure diversity and represent a wide range of visual content.

Fig. 3. Image Dataset

2. Image Pre-Processing

During the dataset preparation phase, the images underwent several preprocessing steps. First, any images with dimensions significantly different from the required input size of 224x224 pixels were resized to maintain aspect ratio while fitting within the target dimensions.

Fig. 4. Image Pre-Processing

3. Output

This creates a web application using Flask, a Python web framework, to classify images uploaded by users. It initializes a Flask application and sets up a folder called "uploads" for user-uploaded images. The application loads a pre-trained deep learning model from a file named "model.h5", performs image preprocessing, and classifies the preprocessed image using the loaded model. The application uses routes such as the Home Page for user uploading images, the Prediction Page for classification, and file handling. The application also handles file uploads by checking if a file is present in the request and saving it to the "uploads" folder. Finally, the application runs in debug mode for development, providing error messages and automatically restarting the server when changes are made to the code.

ISSN [Online]: 2583-2654

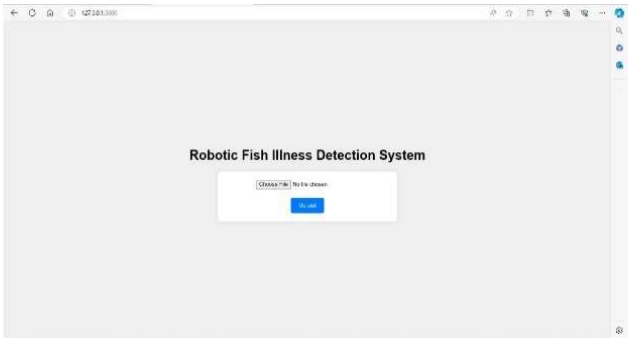

Fig. 5. Robotic Fish Illness Detection System

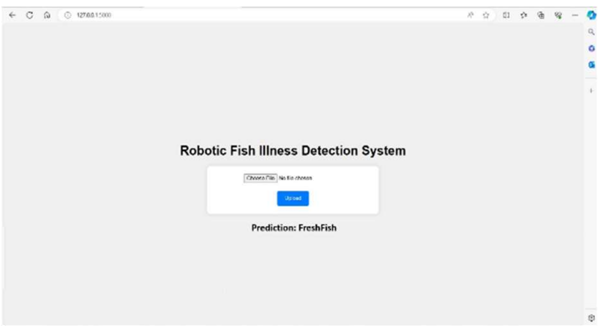

Fig. 6. Prediction 1(FreshFish)

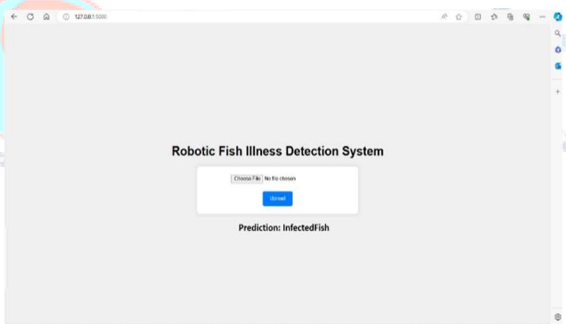

Fig.7. Prediction 2(InfectedFish)

V CONCLUSION

The fusion of Convolutional Neural Networks (CNNs) with the Hamming distance formula within the fish disease detection project represents a remarkable stride forward in the realm of automated disease surveillance within aquaculture settings. Leveraging the advanced capabilities of CNNs, the system adeptly extracts pertinent features from fish images, enabling the detection of subtle and intricate disease patterns that might otherwise elude human observation. Through the intricate layers of neural networks, CNNs can discern nuanced variations in fish health indicators, facilitating early disease detection and intervention.

Complementing the prowess of CNNs, the incorporation of the Hamming distance formula enriches the disease identification process by offering a quantitative assessment of the disparity between predicted disease patterns and their actual manifestations. By quantifying this discrepancy, the Hamming distance provides a tangible metric for gauging the accuracy and reliability of the detection system's outputs. This quantitative insight not only enhances the interpretability of the model's predictions but also serves as a valuable tool for refining and optimizing the system's performance over time.

One of the notable strengths of this integrated approach is its resilience to diverse and challenging image conditions commonly encountered in aquaculture environments. Whether contending with variations in lighting, water quality, or fish species, the CNN-Hamming distance framework demonstrates adaptability and robustness, ensuring consistent and accurate disease detection outcomes across a range of scenarios. Moreover, its potential for scalability suggests promising applications in large-scale aquaculture operations, where efficient disease monitoring is paramount for maintaining fish health and productivity.

Despite its strengths, the efficacy of the CNN-Hamming distance model is contingent upon the quality and diversity of the underlying training dataset. A comprehensive and representative dataset, encompassing a wide spectrum of fish species, disease states, and environmental conditions, is crucial for training a robust and generalizable model. Additionally, ongoing efforts to enhance the interpretability of CNNs and refine the model's performance through techniques like transfer learning and data augmentation hold promise for further optimizing its efficacy in real-world settings.

Looking ahead, continued research and development efforts aimed at refining and extending the capabilities of the CNN-Hamming distance framework are warranted. This could involve exploring novel architectures, integrating additional data sources (such as environmental monitoring data), and collaborating

ISSN [Online]: 2583-2654

with domain experts to validate and enhance the system's effectiveness in practical aquaculture settings. By laying the groundwork for efficient and data-driven fish disease detection systems, this project stands to make significant contributions to the sustainability and resilience of aquaculture practices, ultimately benefiting both industry stakeholders and global fish population health.

REFERENCES

[1] Ahemed Waleed, Hadeer Medhat, Mariam Esmail, Kareem Osama, Radwa Samy, Taraggy M Ghanim "Automatic Recognition of Fish Diseases in Fish Farms" 14th International Conference on Computer Engineering and Systems (ICCES) pp 1-6, December 2019

[2] Yo-Ping Haung and Simon Peter Khabusi "A CNN-OSELM Multi-Layer Fusion Network with attention Mechanism for fish disease recognition in aquaculture" IEEE systems, Man and Cybernetics Society Section vol. 11, pp 1-16, 29-May-2023

[3] Zhenxi Zhao, Yang Liu, Xudong Sun, Jinato Liu, Xinting Yang, and Chao Zhou "Composited FishNet: Fish Detection and Species Recognition From Low-Quality Underwater Videos" IEEE Transactions on Image Processing vol. 30, pp 1-16, 12-April-2021

[4] Yinchi Ma and Wen Ding "Design and Implementation of Remote Fish Disease Video Diagnosis Expert System" 9th IEEE International Conference on Communication Software and Networks, pp 1-4, May 2017

[5] Shaveta Malik, Tapas Kumar, and A. k. Sahoo "Image Processing Techniques for Identification of Fish Disease" IEEE 2nd International Conference on Signal and Image Processing, pp 1-5, August 2017

[6] Md Shoaib Ahmed, Tanjim Taharat Aurpa, and Md. Abdul Kalam Azad "Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture" Journal of King Saud University, pp 1-13, 9-May-2021

[7] S.N Pauzi, M G Hassan, N Yusoff, N H Harun, A H Abu Bakar and B C Kua "A Review On Image Processing For Fish Disease Detection" Asian Conference on Intelligent Computing and Data Sciences (ACIDS), pp 1-10, May-2021