Thesis Title: Artificial intelligence-based diseases detection of Shrimps

Final Year Project/Thesis/Internship
Course Code: SE 431

Supervisor

Mr. Md Rajib Mia

Lecturer (Senior Scale)
Department of Software Engineering
Daffodil International University

Submitted By

Name: Seamur Rahman
ID: 201-35-592

A thesis submitted as a partial requirement for the completion of the Bachelor of Science degree in Software Engineering.

Summer 2025

APPROVAL

This thesis titled on **"Artificial intelligence-based diseases detection of Shirmps"**, submitted by **Seamur Rahman (ID: 201-35-592)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

_______________________________ **Chairman**

Dr. Imran Mahmud
Professor & Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

_______________________________ **Internal Examiner 1**

Md Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

_______________________________ **Internal Examiner 2**

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

_______________________________ **External Examiner**

Md Habibur Rahman
Associate Professor
Department of Computer Science and Engineering
Islamic University, Bangladesh

DECLARATION

I hereby declare that the work presented in this thesis is my own original research work, carried out under the supervision of Mr. Md Rajib Mia, Lecturer(Senior Scale), Department of Software Engineering, Daffodil International University.

This work has not been submitted anywhere, either in whole or in part, for any degree, diploma, or publication in this or any other university. All sources of information used in this thesis have been duly acknowledged.

Supervised By

Mr. Md Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering

Submitted By

Siam

Seamur Rahman
ID: 201-35-592
Department of Software Engineering

Table of Contents

Abstract

Shrimp farming plays a crucial role in the global aquaculture industry. Shrimp aquaculture, involves the cultivation of shrimps in artificial ponds, tanks or other controlled environments. Shrimp farming is a significant source of income and employment for many countries, particularly in Southeast Asia. Shrimp farmers faces various diseases along with other challenges that can significantly impact the production and profitability of shrimp farming. The most devastating shrimp diseases results in production loss for the farmers, distributors and consumers. Frequent monitoring of shrimp health allows for early detection of any potential disease outbreaks, enabling prompt actions to mitigate risks. The existing researches on shrimp diseases, disease management and artificial intelligence-based disease classification are proposed in the literature review. We are proposing an artificial intelligence-based disease detection system for shrimps with the help of image processing, This will help with the monitoring and early detection of shrimp diseases. We approached and utilized some machine learning models to train and test on our dataset, then evaluated the results to progress towards our target of building a complete system or platform to successfully detect diseases at an early stage for the farmers.

Keywords: shrimp disease; machine learning models; feature extraction; shrimp aquaculture; disease outbreaks; fungal infections; artificial intelligence; disease detection; image processing; White Spot Syndrome Virus (WSSV); Early Mortality Syndrome (EMS)

---

1. Background

Aquaculture has become a reliable source for the uprising fish production in the subcontinent. Due to the suitable environmental condition of this region, the fish production industry is rapidly farming different breed of fishes. As a result, farmers are farming a huge number of freshwater species in their ponds, canals and agro-farms to meet the need of the fish consumers all over the world. Asia currently provides 40.1% contribution of the total fish production in the world [5]. It is believed that the need for more sustainable production in the aqua-culture industry is promptly needed as an estimation shows that by 2050, there will be a number of 9.7 billion people in need of huge food production. To address the issue, the current production has to be increased by at least 70% of the current production rate [12].

Sustainable Shrimp production can be a feasible and practical solution to address the need of the increasing demand for food production. The Global shrimp production is mainly dominated by India with the global shrimp trade worth of 17.7 B in the United States [3]. The other regions of Asia have also increased shrimp production in the last 15 years. The extended production of Shrimps not only contribute to the fish consumers of a particular region but also significantly helps the gross national production (GNP) of a country. It is important to ensure sustainable shrimp production for mitigating the financial and economic loss of the shrimp farmers. Sustainable Shrimp production depends on different aspects like the condition of the shrimp farms, the state of the water parameters like ph., dissolved oxygen, salinity in the shrimp habitats and the production rate of the shrimp farms [12].

Shrimp production farms typically rely on mundane and traditional cultivation methods. The production rate of industrial farms is generally higher than the local shrimp farms managed by the local farmers as they know very little about ideal farm management techniques. The water parameters juristically impact the production rate of shrimps as the shrimps are susceptible to different virus, bacterial, fungal infections and other diseases [10,12]. Some of the most notable and popular shrimp diseases are: White spot disease (WSD), White tail disease, Early Mortality Syndrome (EMS), Enterocytozoon Hepatopenaei (EHP), Yellow head disease (YHD) [14].

Due to the outbreak of several diseases in different shrimp farms, farmers constantly have to go through huge financial losses. They do not have the resource nor the remedy to come up with a counterplan to face this issue. The incorporation of machine learning to detect several shrimp diseases with the help of image processing can be immensely helpful for the aquaculture industry. Farmers do not have to rely on the specialists of the aquaculture sector to get details about the diseases, instead they can instantly know about the details of a particular disease infested in the shrimp farm. Based on the integration of the machine learning models in the shrimp farms we mainly see existing shrimp disease detection techniques can be divided into two categories: (A) traditional machine learning models [16,17,20,22] and (B) deep learning models [18,19,21,23] and (C) custom/hybrid machine learning models [19,23].

Finally, we introduce a shrimp disease detection system with the help of image processing using machine learning and deep learning models. We compared the performance and results (Training accuracy, test accuracy, validation) of these models and evaluated them to approach an accurate disease detection system for farmers.

2. Related Works

Shrimp disease detection is a broad and expansive study that cannot be confined on a particular machine learning algorithm. After revising and exploring different shrimp disease related studies and works, we can observe the domain of the related studies into three main categories: (A) Detection of shrimp diseases based on traditional machine learning models; (B) detection of shrimp diseases based on deep learning models and (C) detection of shrimp diseases with the help of custom-built machine learning models. This section will discuss about the performance of different machine learning models that were used to detect shrimp diseases.

2.1. Traditional machine learning models

Machine learning models are efficient at capturing complex patterns and features. Traditional machine learning models are extremely powerful at dealing with smaller datasets. In order to address the frequent infection of shrimp diseases traditional machine learning models were implemented in some shrimp farms in Vietnam [16]. Regular monitoring of shrimp farms for detecting disease outbreak can be quite bothersome. Quach et al. [16] came up with the idea of detecting shrimp diseases based on text-based description. The Vietnamese compound textual sources were transformed and labelled with different shrimp diseases. The dataset comprises of 1098 symptoms of 14 shrimp diseases based on the feedback of the experts in the Vietnamese fisheries sector. VnTokenizer was used to generate a final text size of 13,646 words. The labelled features are fed in the machine learning models to build a predictive system for shrimp disease detection. The evaluation of the models was done based on the accuracy and results observed from four different datasets. 5 models were performed and compared to evaluate the performance of the models to accurately classify shrimp diseases based on textual descriptions. Different parts of Vietnam are prominent for shrimp farming and some of the common shrimp diseases around this region are acute hepatopancreatic necrosis (AHPND), enterocytozoon hepatopenaei (EHP) and white spot syndrome virus (WSSV). Geographical information about the disease affected regions of Vietnam cannot be easily foretold. Geographical information systems (GIS) are one such tool that can be extremely beneficial to extract the geographical data from such regions. The tool integrates geographical features with attribute data to understand and analyze geospatial data. Khiem et al [17] used the technology of GIS to extract the geographical data for the shrimp disease affected regions in Vietnam. Later the features were used with the clinical signs and environmental factors as related parameters to investigate the disease outbreaks. Due to the occurrence of missing values in the environmental factors, data interpolation was used to predict missing values in the dataset. The supposition to predict the missing values were based on the environmental data of the nearby farms within 10km range. The traditional machine learning models were used to predict the disease affected areas for WSSV, EHP and AHPND diseases. Another task of the models was to predict the correlation of different environmental factors with shrimp diseases. Amongst all the environmental factors salinity and temperature were highly related with the prevalence of these shrimp diseases.

AHPND disease is a severe problem in the east coastal Region of Mekong Delta [20]. Unlike EMS and WSSV, AHPND cannot be easily recognized based on computer models as the symptoms of the disease shares similarities with other shrimp diseases. External environmental factors, shrimp seedstocks and poor nutritional management might contribute to the infection of AHPND in shrimps. In khiem et al. [20], several independent factors and parameters such as disease symptoms, visceral status, environmental factors, and general management of the shrimp farms were put into the machine learning models to detect which factor played a crucial role on the outbreak of the disease. According to result of the logistic regression model, it was suggested that salinity and water temperature highly contributed to the prevalence of AHPND in the Makong Delta region. A similar approach was taken in Makong Delta in [17] where salinity and temperatures were also the top factors to initiate a disease outbreak. Two other shrimp diseases like white spot disorder (WSD) and yellow head disease (YHD) are highly contagious and lethal to shrimp mortality. These diseases spread rapidly through water and poses huge threat to the farmers in India, Andhra Pradesh. In Sankar et al. [22] an automated approach using computer vision to identify shrimp infections by extracting features was described. The process involved the acquisition of shrimp images, image pre-processing with Gaussian mean to remove noises, gray scale conversion, binary conversion, segmentation, feature extraction, and classification using the KNN algorithm. The KNN classifier was used to predict the two diseases and assess their severity. The traditional models are suitable for small datasets but when it comes to larger datasets, more advanced approaches are taken.

2.2. Deep learning models

Deep learning models are advanced machine learning model that captures the function of a brain's neural setup and propels through its layers with the help of backwards propagation. Deep learning models have been integrated into shrimp farming to detect shrimp diseases based on image datasets. The contribution from shrimp farmers and distributors can be quite useful if recorded pictures of diseases shrimps are shared to the contribution hubs. A similar setup was established in Mekong delta to collect sample of shrimp diseases [18]. Shrimp diseases like AHPND, EMS, WSSV, white feces syndrome (WFS), and YHV are predominantly seen in this region. Modern and advanced disease detection are crucial for mitigating losses for the farmers but current practices of shrimp farming rely heavily on farmers' experiences with old and bruised up approaches. A study Trung et al [18], used deep learning models with transfer learning was to detect the six shrimp diseases in Mekong delta region in Vietnam. Machine learning models are reliant on training the models. With the help of transfer learning these models will not have to rely on training the models from scratch. Instead, it can use the feature of transferring the necessary domain from pre trained models and improve performance of a particular model. Furthermore, a well and knowledgeable domain can pass their ample and large learning data to less powerful and limited domains with this method. The deep learning models were used to detect shrimp diseases. Higher accuracy was observed after removing the backgrounds from the images. Some common shrimp diseases in China are Penaeus Vannamei: hepatopancreatic necrosis disease (HPND), red body disease (RBD), and whitish muscle disease (WMD). A rapid detection method based on convolutional neural network (CNN) for detecting these diseases were shown in Wang et al [21]. Firstly, the dataset was gone through the augmentation phase. The images were converted from rgb to average grey scale. The processed images were put into CNN algorithms as they have proven to show better results in Penaeus Vannamei detection. The CNN model was based on LeNet architecture with improved hyper-parameters. The dataset was captured in plain white background to capture high results for the model. The need for developing observation devices capable of capturing images in highly turbid water was addressed in [21]. Sometimes the need to tweak the deep learning models arises to generate better results. Combining the strength of two existing deep learning models in another model using transfer learning can result in better performance for different predictive purposes.

2.3. Custom/hybrid models

White spot disease is currently the most well-known shrimp disease in the world. The distinctive white spot-on different part of the shrimps makes it very familiar to distinguish from other diseases. In Edeh et al [19], it is suggested that environmental factors such as salinity, ph. and temperature may play a vital role in White spot disease outbreak as it was suggested in [17][20]. The environmental parameters and factors are taken from Bangladeshi shrimp farms. Data pre-processing was applied to fill the null and void values. Ensemble and adjustment of hyper parameters were applied so that the machine learning Model do not over-fit. This provided the models with a very high prediction rate and accuracy. Finally, they proposed a model named Ensemble of Random Forest and CHAID (ENRFCH) which benefits from the advantages of both machine learning models. Firstly, a subset is created from the original data and Random Forest classifier helps to capture complex pattern for the model. Then, the CHAID model is used to correct the errors made by the Random Forest model with the help of a loss function. Finally, ENRFCH addresses the limitations on the individual tree models. Implication of using deep learning models for intelligent detection systems for shrimps is discussed in [23]. The AlexNet model is used for feature extraction, extracting spatial texture, and color features from shrimp images. The employment of the IRFA algorithm for feature selection is enabled, determining the relationship between features for different shrimp classes. Finally, the custom model DLCNN model is used for classification, trained with the selected features using SoftMax classifier.

Table 1: Comparison Table

Ref.	Model	Accuracy/Result	Limitation
[16]	5 Machine Learning models were applied in this paper: Logistic Regression, Random Forest, Naïve Bayes, SVM, MLP	SVM had the highest accuracy of 83.64%	A suggestion for hyper-parameter tuning of the models can be made.
[17]	4 Machine learning models were applied in this paper: Logistic Regression, Neural Network, Random Forest, Gradient Boosting	AHPND had the highest accuracy of 91.89%	The research was meet with small datasets where missing values were incorporated with GIS technology.
[18]	2 Machine learning models were used: Inception-v3, MobileNet's 1.0-244.	Inception-V3 had the highest accuracy of 90.02%	The deployment of the models in real world scenario could be challenging
[19]	2 Machine learning models were used: Random forest, CHAID There was a proposed model named ENRFCH which took the advantages from both models.	ENRFCH had the highest accuracy of 98.28%	The external factors are highly correlated with white spot disease only. These factors can vary depending on other diseases.
[20]	The machine learning models used in the study are: Logistic Regression, K-nearest neighbor (KNN), Decision Tree, Artificial Neural Network (ANN).	Logistic Regression had the highest accuracy of 83.04%	Data was collected from several provinces which could contain biasness in the dataset as environment conditions vary across different regions.
[21]	The study mainly focused on CNN based model that was based on Li-Net model architecture	Normal shrimps: 96.7% Shrimps with HPND: 95.8% Shrimps with RBD: 96.7% Shrimps with WMD: 95.9%	*Penaeus Vannamei* has similarity with other diseases that can lead to overlapping with other disease characteristics.
[22]	The K-Nearest Neighbor algorithm was applied in this paper	KNN Model had the highest accuracy of 93.7%	KNN can be susceptible to datasets with limited and small size.

[23]	The machine learning models used in the simulations are: HML, AlexNet, CNN, ShrimpNet Proposed Model: TLOFS with DLCNN (Transfer Learning Optimized Feature Selection with Deep Learning Convolutional Neural Network)	DLCNN with TLOFS had the highest accuracy in this paper 99.98%.	Deep learning models are prone to over-fitting. Various conditions like environmental factors, dataset biasness, class imbalance and absence of parameter-tuning can be major hurdles to acquire better results.

Different approaches for the detection of shrimp related diseases were observed. Some methods tried to tune in the hyper parameters in an accurate and effective way for generating better result in shrimp disease detection. Whereas other approaches mediated their way towards building a customized machine learning model for detecting shrimp diseases. The incorporation of deep learning models along with the traditional models was another notable process as deep learning models tend to give better result compared to the traditional models. Finally, an overlook on external factors such as environmental factors, condition of shrimp habitats and the farming practices of farmers were also observed. The related literatures showed and suggested different ways to assess shrimp detection. We will address machine learning models suitable for shrimp disease detection in this study. Later, we will compare our results based on the accuracy and performance derived from our models with the models from the related literature section in a performance evaluation comparison.

3. Problem Statement

The mismanagement of shrimp farms in the Bagerhat and Satkhira regions of Bangladesh has resulted in frequent disease outbreaks, leading to significant financial losses for farmers. The inability to detect and address diseases in a timely manner forces farmers to sell infected shrimps in the local markets, further exacerbating their economic distress. The lack of effective disease detection methods adds to the challenges faced by shrimp farmers in maintaining healthy and productive farms.

Some of the diseases that arises in the Bagherhut Region are predominantly seen in the Monsoon period. The disease outbreak can last up to 2-3 months. Most of the diseases are generally bacterial, virus infected or fungal. Here is a detailed discussion about the predominant diseases seen in the Khulna Region:

White Spot Disease: Currently this is the biggest shrimp related disease in the world and the Bagerhut and Satkhira region has fallen prey to its claw. The trademark symptom of WSD is several white spots visible on the head portion of the shrimps and sometimes it can reach towards the tails as well. Several farms are being destroyed due to the predominant attack of this virus infected disease. Several corporations, institutions and companies are trying to find a remedy, yet its firm grip has not loosened a bit.

Early Mortality Syndrome: bacterial disease that often attacks the digestive system of the shrimps. The physical features are red spots all over the shrimp's body, it is another deadly shrimp disease as it bears high mortality rate and often a whole batch of shrimps can die within 2-3 hours. The discoloration of the body is another feature of the EMS disease.

White Tail Disease: Primary symptom of this disease is the discoloration in the tail of the shrimps. The disease is mostly prevalent in *Peneaus Monodon* (Bagda Shrimps) species. Sometimes white spots are visible with a discolored tail which becomes weak and brittle over time. It affects the mobility and swimming abilities of the shrimps. Although the exact cause of this disease is not found it can lead up to fungal infections and increase mortality.

White Muscle Disease: It is a muscular disorder that affects the shrimps, specifically due to inadequate levels of essential nutrients, particularly vitamin E and selenium. The disease targets the muscle tissues of the shrimp, resulting in various effects. A sign of WMD is weakness, where infected shrimp exhibit reduced muscle strength and overall vitality. The disease also induces muscle degeneration, causing the affected muscles to deteriorate and lose their normal function. Therefore, shrimp with WMD may experience difficulties with lethargic movement, disabling their ability to swim and perform regular activities.

EHP (Enterocytozoon hepatopenaei): It is caused by a parasite called *Enterocytozoon hepatopenaei*. The disease attacks the organ of the shrimp and as a result it might show pale color. Shrimps affected with this disease will lose appetite and their growth will reduce compared to other healthy shrimp specimens. It drastically drains up production rate causing huge economical loss for the shrimp farmers. The shrimps become easily susceptible to other diseases as their resistance is lowered.

These are some of the common disease outbreaks that are prevalent in Bagherhut and Satkhira Region. Additionally, the absence of a reliable system for monitoring water parameters in shrimp farming contributes to the problem. Fluctuations in water parameters can adversely affect shrimp health and make them more susceptible to diseases. Without timely information about water quality, farmers are unable to make necessary adjustments to create optimal conditions for shrimp farming.

Therefore, there is a critical need for a solution that can accurately detect shrimp diseases and provide real-time monitoring of water parameters. The proposed system, utilizing image processing and machine learning models, aims to address these challenges by enabling early disease detection and notifying farmers about unstable water parameters. Offering timely notifications and suggesting necessary changes for future farming practices, this system can contribute to the sustainable and profitable growth of the shrimp industry in Bangladesh.

4. Research Questions

- What are the existing technologies available in Bangladesh to assess and detect shrimp diseases?
- How to develop an image dataset of shrimp diseases for detection of shrimp diseases?
- How to develop a machine learning model for detecting shrimp diseases and come forth with an end user application to help the farmers?

5. Objectives

Assess the current shrimp disease detection technologies in Bangladesh.

The vast regions of Bangladesh are not suitable for shrimp farming. The coastal region of Bangladesh; mainly the Bagerhut and Shatkhira region are promising shrimp farming establishments right now. The ghers are infected with severe bacterial, fungal and virus affected diseases that pose a great threat to the shrimp farmers. The primary objective is to assess and explore different existing methods of shrimp disease detection in the coastal regions. This process involves literal consultation with shrimp farmers, documentation of the list of shrimp disease detection methods, reviewing the existing solutions and challenges, insights from aquaculture experts etc.

Create an image dataset for health and diseased shrimp

Shrimp datasets are the pathway to successfully detecting shrimp diseases with the integration of machine learning models. The key to take is to build an accurate and reliable model that will manifest and address the needs of the farmers. A dataset will include the diseased and healthy shrimp specimens of different shapes and varieties. A large dataset is typically a great initiator for successfully deploying a machine learning model with exact and accurate precision. With the collaboration of the shrimp farmers and aquaculture experts a large dataset is to be captured and documented that will include different shrimps with diverse diseases.

Develop an end user platform for farmers to detect shrimp diseases based on machine learning

The final step is to build a platform where the shrimp farmers can easily detect shrimp diseases using a product like a webpage or mobile application. The platform should integrate the machine learning models and successfully detect diseases with a simple capture from a camera or mobile phone. The outcome reflects on the ease of the farmers so that they avoid unnecessary and long procedures to consult with the experts to wait for confirmation of a particular disease. They can quickly and instantly recognize and assess the situation if a particular shrimp gher is affected by a disease.

6. Methodology

In this research, we used two machine learning methods to detect shrimp disease. In the first method, we used four deep learning models to classify shrimp images into five classes: EMS, EHP, white spot disease, white tail disease, and healthy shrimps. In the second method, we used two traditional machine learning models to classify shrimp images into two classes: healthy and diseased. We evaluated the performance of all four models using accuracy, precision, and recall metrics. We also deployed the model in a real-world application to test its effectiveness. The flowchart of the project is shown below. The project is divided into several parts, each with its own mechanics. First, we load the images in Joint Photographic Experts Group (JPEG) format and perform preprocessing. During preprocessing, we perform data augmentation, image resizing, and dataset splitting. Next, we fit the training data into the model. The deep learning models are pre-trained for image processing tasks. During training, it is important to set the hyperparameters. Once the images are trained by our classification models, we fit them with the testing dataset. In the final phase of the systemic structure, we perform and conduct accuracy, precision, validation and loss functions of our Models to see how well the models are performing. The goal of this project is to develop a machine learning model that can accurately detect shrimp disease. The model can be used by shrimp farmers to identify diseased shrimp early on and take steps to prevent the spread of disease.

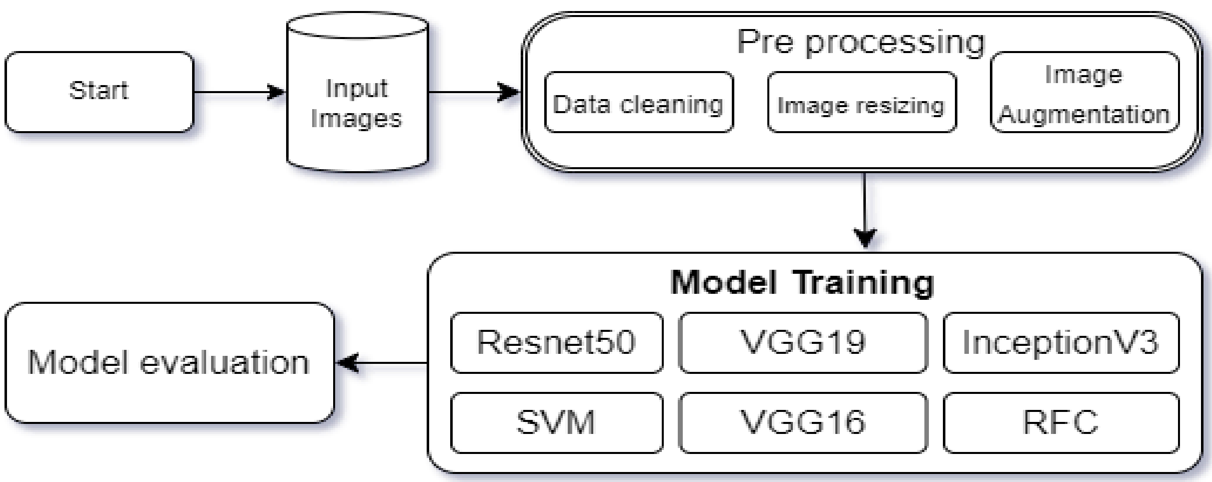

Fig 1: Flow-chart of The System Design.

6.1. Data collection

A number of 1,161 shrimp samples were collected for the machine learning project to detect shrimp diseases. The dataset consisted of both healthy shrimp samples and samples affected by different diseases prevalent in the shrimp farming industry. Specifically, 609 samples were collected from local fish markets, and these samples represented healthy shrimp specimens. On the other hand, 552 samples were collected from Bagherhut and these samples represented shrimp samples affected by various bacterial, fungal or virus diseases. The collected diseased samples consisted of four different shrimp diseases:

- **EMS (Early Mortality Syndrome or Red Spot):** A total of 247 samples were collected from shrimp exhibiting symptoms of EMS, characterized by red spots on different parts of the body.

Fig 2: Sample photos of EMS Disease

- **WSSV (White Spot Syndrome Virus):** The dataset included 185 samples of shrimp infected with WSSV. WSSV is a viral disease known to cause white spots on multiple parts of the shrimp's body and head.

Fig 3: Sample photos of WSSV Disease

- **EHP (Enterocytozoon Hepatopenaei):** A total number of 87 samples were collected from shrimp affected by EHP. The resemblance of EHP can be hard to spot on different breed of shrimps. Due to lethargic movement and reduced growth, shrimps affected with EHP can be confused as baby shrimps. The deciding factor to assess a shrimp affected with EHP is the discoloration of the body.

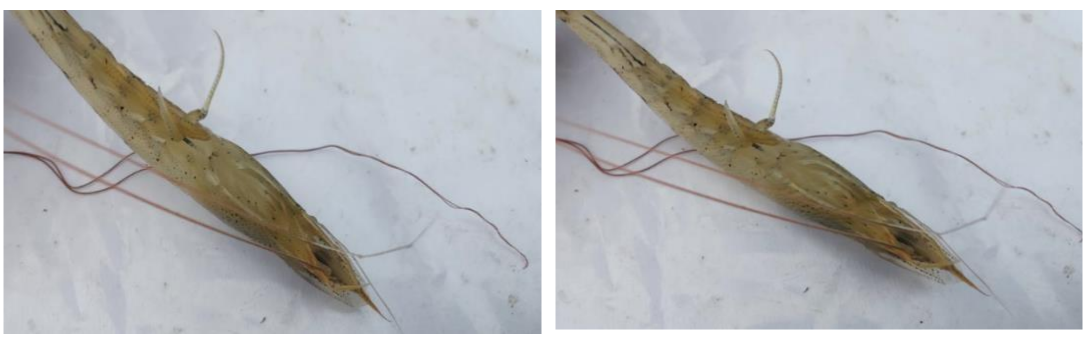

Fig 4: Sample photos of EHP Disease

- **White Tail Disease:** The dataset contained 33 samples from shrimp displaying symptoms of white tail disease. This disease is characterized by discoloration or whitening of the tail region.

Fig 5: Sample of white tail disease

All the images were captured in jpg format and the classification of different diseases along with healthy shrimp specimens laid the groundwork for the application of implementing machine learning models in them.

We have captured 4 different shrimp diseases; they are enlisted below:

Table 2: Shrimp classification

Name Of Disease	Number Of Images
• EHP	84
• EMS	268
• WSSV	183
• White Tail Disease	32
• Healthy	608

6.2. Data Preprocessing

To develop an accurate and robust shrimp disease detection machine learning model, it is necessary to apply effective data preprocessing techniques to ensure the quality, reliability, and suitability of the input data. The following procedures were done before training our machine learning models:

- Image Resizing (Cropping)
- Creating two separate class (“Disease”, “Healthy”)
- Image Augmentation
- Image Segmentation
- Data Labeling

We have manually resized all of the images to put the shrimps in the center of the image and remove all the unnecessary background information from the dataset. This also helped us to get better augmented images. By cropping the images, the size of the images was set to 244x244 pixels with 3 channels., the size of the dataset drastically reduced this helped to give more data to train on in a single batch. This is very important because of global maxima and global minima. If there is a small amount of data in a batch, then the models will not get to know if there is any other better point in the dataset to learn from. Thus, it will not improve the overall quality of our model.

We have put all the diseased shrimps (EHP, EMS, WSSV and White Tail Disease) in a class called “Disease” and the healthy shrimp images were transferred to class “Healthy”. It will help the model to train on these two classes for disease detection. We have augmented the images by rotating the images by 90° till it gets back to original position and used horizontal flip to flip the images. Comparison of before and after augmentation is shown below:

Table 3: Augmentation details

Class	Before augmentation	After Augmentation
Disease	566	2264
Healthy	608	2431

Data augmentation was performed with the help of TensorFlow API, and the extracted image amount was 14081. Finally, 80% data was put in to training dataset and the rest 10% and 10% images were put into the testing and validation folder.

6.3. Model Training and Implementation

Four deep learning models VGG16, VGG19, ResNet50 and InceptionV3 were used to train the image dataset. Additionally, 2 traditional models SVM and Random Forest classifier were used as well.

6.4. Hyper-parameter tuning:

These parameters were applied for the deep learning models where we used 5 classes:

Table 4: Hyperparameters used for VGG16, VGG19, ResNet50 (5 Classes)

Hyperparameters	**Value**
Random states	0
Weights	ImageNet
Include top	False
activation 1	Relu
activation 2	Softmax
Optimizer	Adam
Loss	Categorical_crossentropy

Hyper parameters are essential features that contribute to the learning algorithms with their settings and configurations. They help models with their preset values to adjust the complexity of the models for optimized learning patterns. Finally hyper parameters control the amount of regularization to prevent the model from over fitting. Another added feature of model regularization is to help the model perform well towards unknown data by avoiding excessive complexity.

In this case the random state was initialized with a specific value of 0 as it will ensure the training process is reproducible. Then the weights are added from imagenet dataset as it is a large dataset that was primarily built to advance the classification of deep learning models. The pre-trained weights were transferred from imagenet to our shrimp dataset to reduce computational time and train the models. We set include top to false as it excludes the final layer of the model for classification. It aids to customize the last layer according to the need of the disease detection task. The first layer uses ReLu activation function to capture complex patterns and increase performance. The final layer uses softmax activation function probability distribution over the different disease classes for a given input sample. For optimizer Adam was used as it is a popular model that reduces training and validation loss. The algorithm also helps to boost the performance of the model.

Now we converted the 5 classes into 2, naming them healthy and disease class. These are the following steps and tuning that was required for the machine learning models:

Finding the correct hyperparameter for a Machine Learning model is a difficult task. We have tuned our model's hyperparameters to get the most accurate result. Details of each model are given below:

- **SVM:** In SVM we have used a search algorithm called GridSearchCV to find the optimal hyperparameters for the SVM to train the data with. Using the GridSearchCV we found the following as the best parameters for our model:

Table 5: SVM parameters

Parameters	Value
• C (Regularization)	0.1
• gamma (Kernel Coefficient)	0.1
• kernel	linear

This helped us greatly improve the accuracy score of the model.

- **Random Forest Classifier:** In random forest we got very good results from training the model with the parameters we found best suited for our dataset. For the random forest we have used the following parameters for training:

Table 6: Random Forest parameters

Parameters	Value
• n_estimators	100
• criterion	gini
• min_weight_fraction_leaf	0.0
• bootstrap	True

- **VGG19:** In VGG19 we have added 13 additional layers where each layer tries to make the model as generalized as possible for creating a better model. By fine tuning we have found that instead of using 'adam' optimizer, using 'RMSprop' with learning rate set to 2^{-4} gives much better result. Our VGG19 model layers are shown below:

Layer (type)	Output Shape	Number of Parameters
vgg19 (Functional)	(None, 7, 7, 512)	20024384
dense	(None, 7, 7, 512)	262656
batch_normalization	(None, 7, 7, 512)	2048
activation	(None, 7, 7, 512)	0
dropout	(None, 7, 7, 512)	0
dense_1	(None, 7, 7, 256)	131328
batch_normalization_1	(None, 7, 7, 256)	1024
activation_1	(None, 7, 7, 256)	0
dropout_1	(None, 7, 7, 256)	0
dense_2	(None, 7, 7, 128)	32896
batch_normalization_2	(None, 7, 7, 128)	0
activation_2	(None, 7, 7, 128)	0
dropout_2	(None, 7, 7, 128)	0
global_average_pooling2d	(None, 128)	0
dense_3	(None, 2)	258

Fig 6: Additional VGG19 Layers

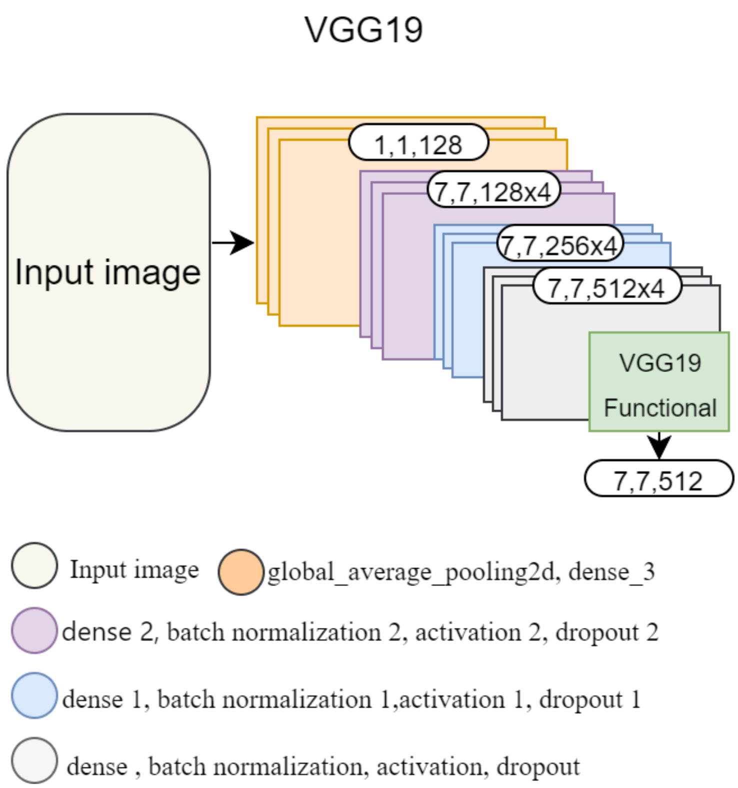

Fig 7: VGG19 Layers

- **ResNet50:** ResNet50 has similar layers as VGG19. But it has a different learning rate. Here the learning rate is set to 2^{-6}. Which resulted in better generalization of the model. Our ResNet50 model layers are shown below:

Layer (type)	Output Shape	Number of Parameters
resnet50 (Functional)	(None, 7, 7, 2048)	23587712
dense	(None, 7, 7, 512)	1049088
batch_normalization	(None, 7, 7, 512)	2048
activation	(None, 7, 7, 512)	0
dropout	(None, 7, 7, 512)	0
dense_1	(None, 7, 7, 256)	131328
batch_normalization_1	(None, 7, 7, 256)	1024
activation_1	(None, 7, 7, 256)	0
dropout_1	(None, 7, 7, 256)	0
dense_2	(None, 7, 7, 128)	32896
batch_normalization_2	(None, 7, 7, 128)	0
activation_2	(None, 7, 7, 128)	0
dropout_2	(None, 7, 7, 128)	0
global_average_pooling2d	(None, 128)	0
dense_3	(None, 2)	258

Fig 8: Additional ResNet50 Layers

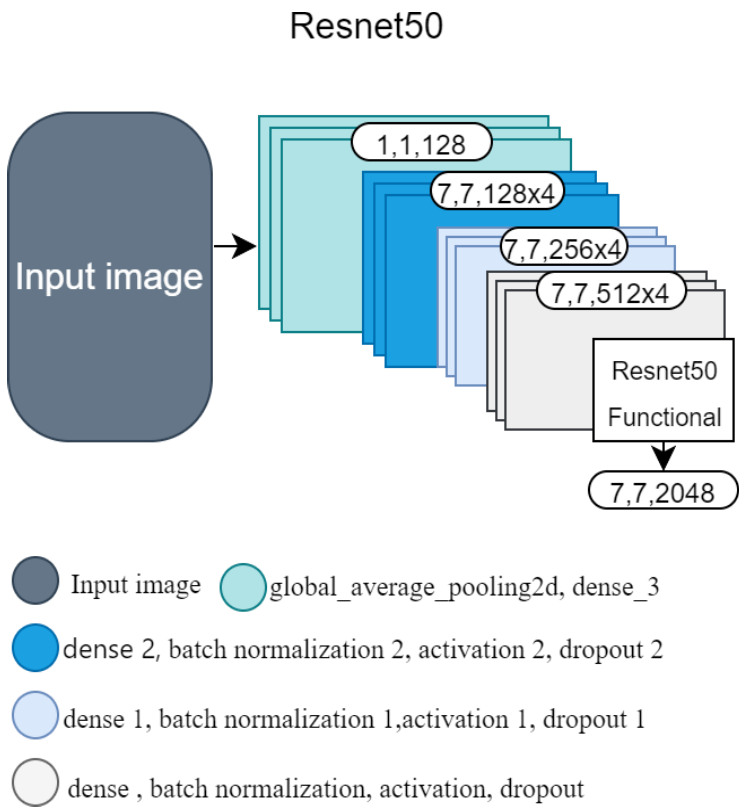

Fig 9: ResNet50 Layers

7. Model Evaluation and Result Analysis:

The experimental results will be displayed for the two phases of the project. Firstly, we will demonstrate the result obtained for the 5 classes with VGG16, VGG19, ResNet50 and inception V3. Later on, we will display the acquired results for the 2 classes with VGG19, SVM, Random Forest classifier and ResNet50.

Table 7: Result Analysis (5 classes)

	Model	Training Accuracy	Validation Accuracy	Testing Accuracy	Training loss	Validation loss
1	ResNet50	89.93%	89.67%	89.67%	28.82%	30.27%
2	VGG19	99.3%	99.33%	99.26%	2.19%	2.28%
3	VGG16	99.64%	99.26%	99.59%	1%	3.17%
4	InceptionV3	99.90%	98.39%	98.39%	0.42%	0.5%

The following graphs show the training and validation loss & accuracy of the 4 Machine Learning Models. We can see that inception v3 had the highest training accuracy out of all the models, but it is not a good indicator how the model will perform to new data. So, it is not a good parameter to assess the overall performance of the Model. VGG19 had the highest validation accuracy out of all the models. Validation accuracy is critical during the training stage of the model. Finally, VGG16 has the highest testing accuracy (99.59%) amongst all 4 of the models. Hence it is safe to assume that out of all the models VGG16 gave the highest accuracy. VGG19 and Inception V3 performed very well as well with 99.26% and 98.39% accuracy but accuracy is not enough to conclude that a model performed better than other ones. The consideration of other factors such as validation loss, training loss, f1-score, precision, recall are also important factors.

Amongst all 4 models Inception V3 had the lowest training and validation loss. Meaning it had better convergence. VGG19 also showed low loss values while ResNet50 had comparatively higher loss values. VGG16 was not far behind with validation loss and training loss as well.

Table 8: Result analysis (2 classes)

	Model	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Loss	Validation Loss
1	ResNet50	94.40%	96.59%	48.21%	15.08%	30.27%
2	ResNet50 (After tuning)	98.87%	99.15%	50.59%	10.91%	9.91%
3	VGG19	100.00%	99.36%	53.44%	0.05%	1.52%
4	VGG19 (After tuning)	99.87%	100.00%	58.19 %	1.98%	1.08%
5	SVM	100.00%	76.54%	96.06%	7.08%	52.49%
6	Random Forest	100.00%	68.86%	97.63%	13.90%	57.62%

From the results, we see that ResNet50 and VGG19 performed poorly at first but after hyper-parameter tuning the problems seem to be somewhat resolved because both models tend to generalize the data more but still the acquired accuracy is not satisfactory. SVM and Random Forest classifier had relatively low validation accuracy, but their testing accuracy improved a lot unlike the deep learning models because k-fold cross validation was implemented in SVM and Random Forest which distributes the data evenly in all instances. The model that performed the best in this case was Random Forest. It had the highest testing accuracy of 97.63%. SVM was a close second as it had 96.06% accuracy.

ResNet50 (5 classes)

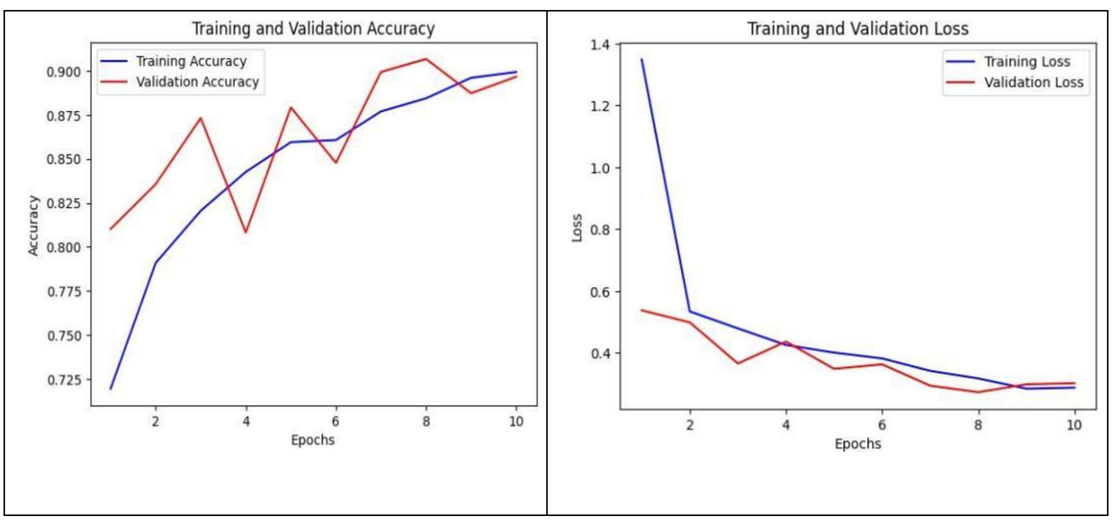

Fig. 10 Training and Validation Accuracy/Loss (ResNet50 5 classes)

Based on the graphs it is evident that data was not distributed properly in the ResNet50 model. From the training and validation graph it seems the validation curve goes upwards for the first 3 epochs. It means that the images were converged on the class type that had the higher amount of images compared to other classes. Later on, the curve takes a downward approach which implies that validation accuracy is decreasing. It means that the model could not detect the class types that had a relatively lower amount of images. Finally the whole graph shows that the dataset was not converged evenly in this model as classes with lower number images were not trained properly. We can see that validation loss was decreasing till two epochs which implies that the model was trying to converge the data but after a while it was not parallel with the training loss which indicates the model was not generalized properly.

ResNet50 (2 classes)

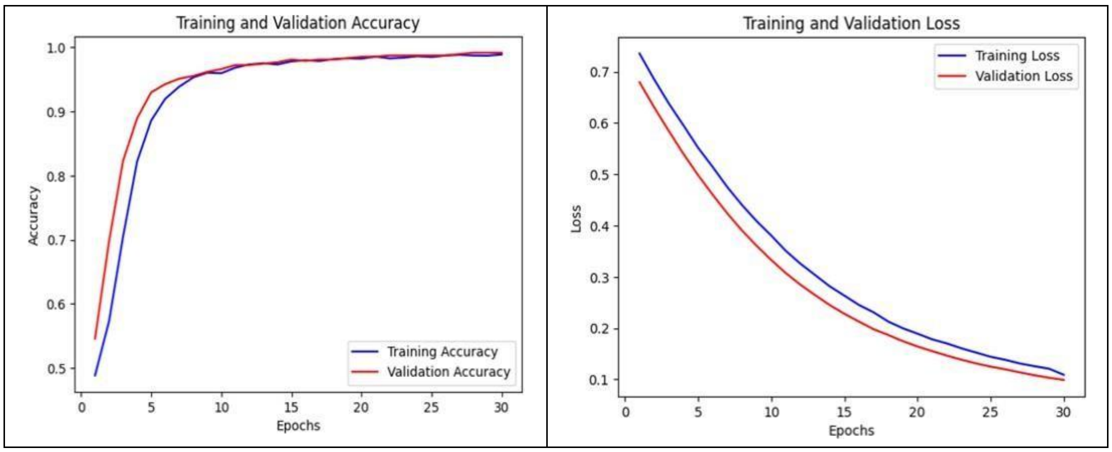

Fig. 11 Training and Validation Accuracy/Loss (ResNet50 2 classes)

From the graph we can see that ResNet50 runs parallel for both the training and testing accuracies. After tuning in the 5 classes into 2 classes, all the disease types were converted into the diseased class. The ResNet50 model failed to converge the dataset evenly as some of the diseases had comparatively higher number of images than others. As a result, Validation was not done properly and it resulted in an over fit model for two classes. The testing results implies that the model failed to validate the data properly.

VGG16 (5 classes)

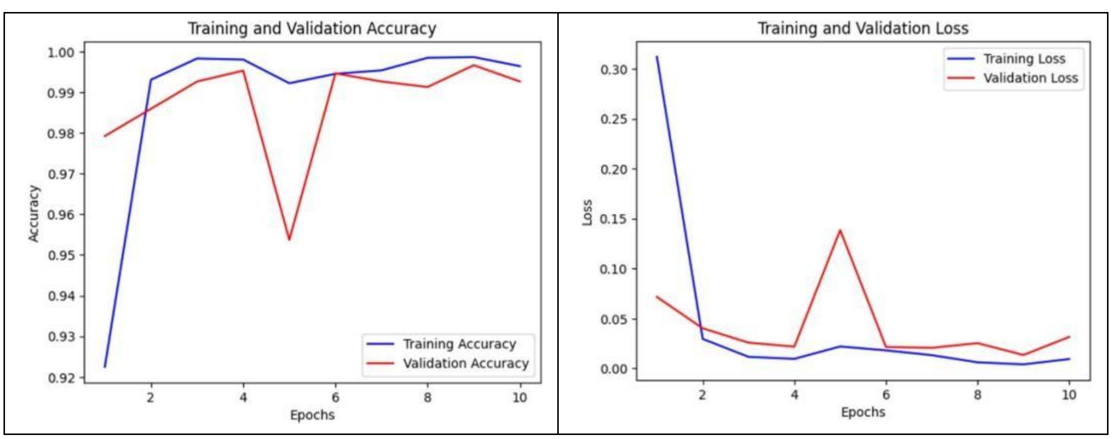

Fig. 12. Training and Validation Accuracy/Loss (VGG16 5 classes)

From the training and validation accuracy graph we see that VGG 16 performs over 95% accuracy. It is interesting to notice that after 5 epoch validation accuracy decreases as the sample size of a particular class type is not evenly generalized in this case. We expect the model to constantly distribute the data into an even format. The case for this graph represents that the model was not distributing the images properly and the classes with higher number of images were prioritized. Such is the case for validation loss in VGG 16. The validation loss increased gradually which was not expected from the model. It could not keep up with the training loss indicating the addition of further tweaking for the model.

VGG19 (5 classes)

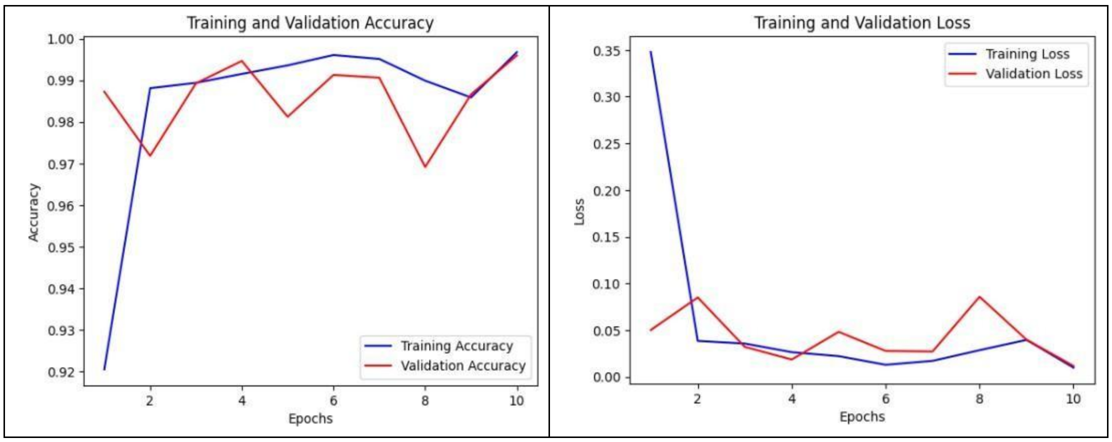

Fig. 13. Training and Validation Accuracy/Loss (VGG19 for 5 classes)

From the graph VGG19 goes through significant drop of validation accuracy. Validation accuracy generally increases over epochs but the model could not generalize all the classes. This is observed in the validation loss graph. Validation loss is expected to decrease over time but we see significant rise in validation loss. It implies that the model cannot identify the classes with their proper labels.

VGG19 (2 classes)

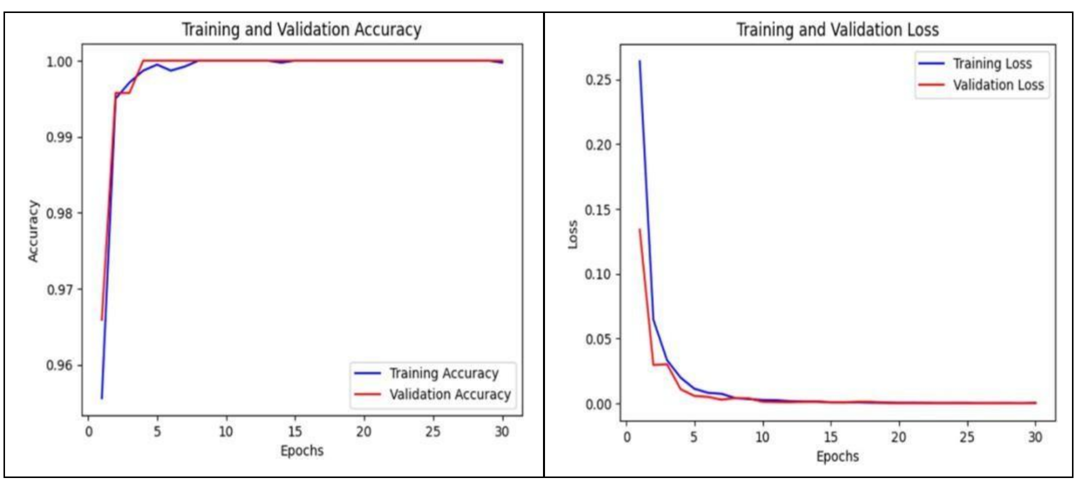

Fig. 14. Training and Validation Accuracy/Loss (VGG19 for 2 classes)

To address the issue of model generalization error the 5 classes were converted into 2 classes. It resulted in an over fit model. From the graph we can say the deep learning model could not identify and generalize the classes with proper distribution. In fact the 4 different diseases that different features that were risky to cumulate in a single class called diseased shrimps. Although the graph shows 100% validation accuracy it's relatively low testing accuracy proves that the model tends to be over fit as performs rather poorly to unseen and new data.

InceptionV3 (5 classes)

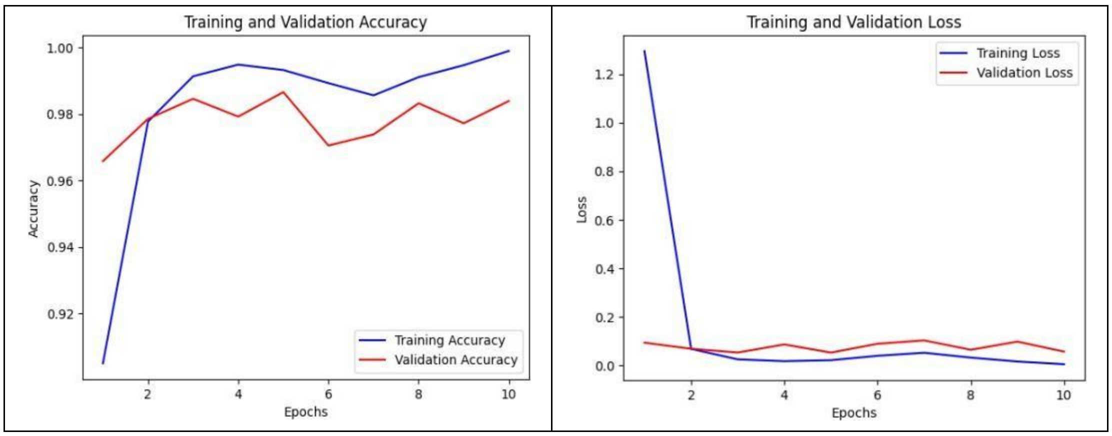

Fig. 15. Training and Validation Accuracy/Loss (InceptionV3 for 5 classes)

InceptionV3 was used only for 5 classes. The model shows signs of over fitting as the validation loss once again goes downwards with more epochs where the opposite should be expected. This once again shows that the different classes were not distributed evenly. Further tuning in the model could avoid the situation but the deep learning models are kind of like black boxes. It is not easy to tell the reason behind these occurrences.

SVM (for 2 classes)

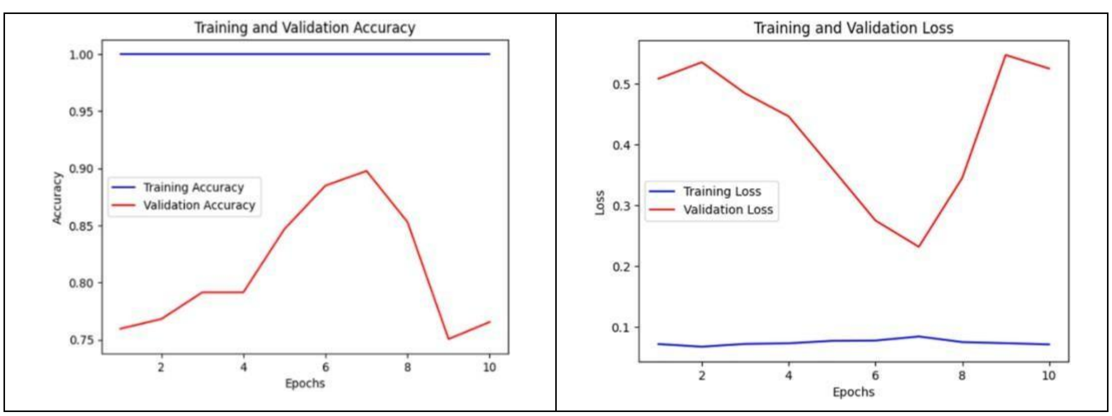

Fig. 16. Training and Validation Accuracy/Loss (SVM for 2 classes)

The highest observed validation accuracy is 90% after 6 epochs for this model. Validation accuracy falls after 8 epoch all the way to 75% validation accuracy. Training loss was observed from 1 to 4 epochs with almost 50% training loss. The training loss gradually decreased after 6 epochs. It means that SVM performed exceptionally well from 2 to 8 epochs as k-fold cross validation distributed the classes evenly and it was evident this generated high validation accuracy for the models. Still there is significant drop in validation accuracy compared with the training accuracy. Significant validation loss was also observed which tends to get minimized after 7 epochs. It implies the validation goes through some issues as it deviates from training loss again.

Random Forest Classifier (for 2 classes)

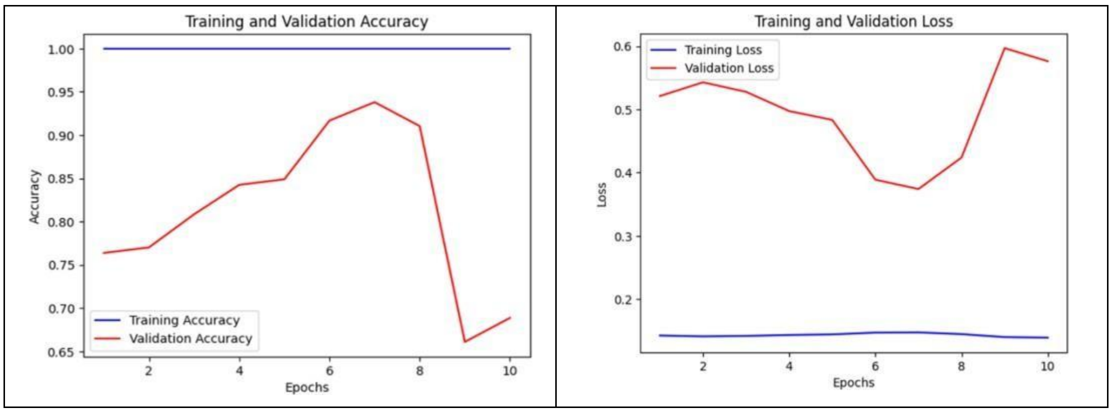

Fig. 17. Training and Validation Accuracy/Loss (Random Forest for 2 classes)

The graph shows the validation accuracy of Random Forest rising from the start till 8 epochs. The highest observed validation accuracy is near 95% and after 8 epochs it falls below 70% validation accuracy. The highest training loss was observed after 8 epochs where 60% image data were lost. The lowest validation loss was observed at 6 epochs where loss data percentage was below 40%. Much like SVM, the k-fold cross validation helps Random forest to generate and distribute the images evenly for a while. Although high difference is observed between training and validation loss the model performs relatively well in new testing sets.

We realize only validation and testing accuracy is not enough to determine how good a model performs. Other parameters such as precision, f1-score, recall etc. are also necessary to highlight the performance of a model. Here are the necessary notations for calculating these results.

$$Precision = true\ positive\ /\ (true\ positive + false\ positive)$$

$$Recall = true\ positive\ /\ (true\ positive + false\ negative)$$

$$\mathbf{F1}\ Score = \mathbf{2} \times (precision \times recall)\ /\ (presion + recall)$$

ResNet50 Confusion Matrix

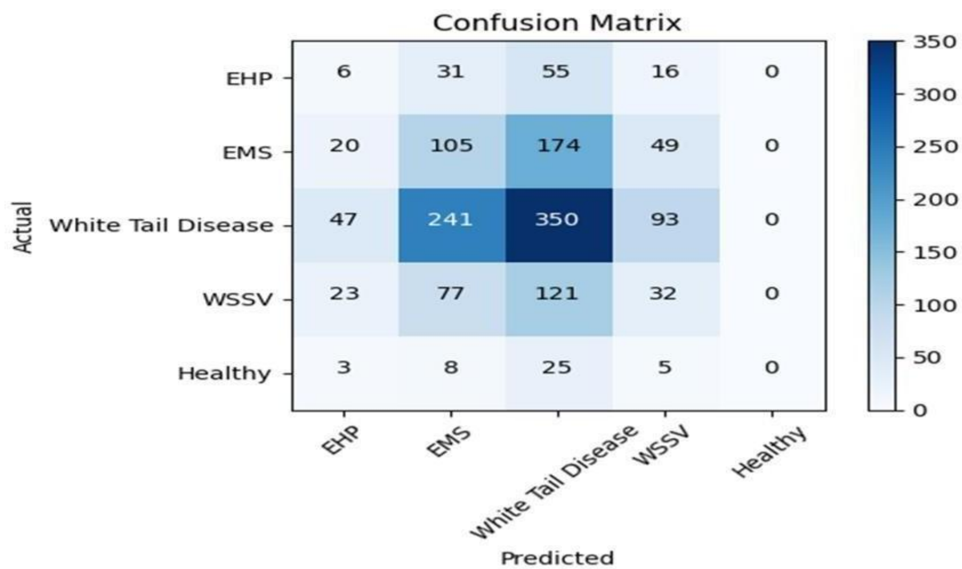

Fig.18. Confusion Matrix (ResNet50 for 5 classes)

We see that the confusion matrix with 5 classes shows how many times the model accurately detected the diseases and we see that white tail disease was accurately detected for most of the time. Whereas WSSV and EMS had larger number of images the model completely fails to identify almost half of the images accurately. Hence the classes were converted into two instead of 5 for better results. The model was tweaked with new parameters for even distribution of the images. The interpretation can be made that due to the relatively small size of the dataset the model was tend to over fit. Again, the improper distribution of the images could be another factor.

Table 9: Confusion Matrix (ResNet50 for 2 classes)

<table>
<tr><td rowspan="2">True Labels</td><td>Disease</td><td>101</td><td>102</td></tr>
<tr><td>Healthy</td><td>106</td><td>112</td></tr>
<tr><td></td><td></td><td>Disease</td><td>Healthy</td></tr>
<tr><td></td><td></td><td colspan="2">Predicted Labels</td></tr>
</table>

Table 10: Precision, Recall and F1-score for ResNet-50 (2 classes)

Class (for 2 classes)	Precision	recall	F1-score
0 (disease)	**0.4975**	**0.4883**	**0.4958**
1 (healthy)	**0.5138**	**0.5238**	**0.5187**

VGG-16 Confusion Matrix

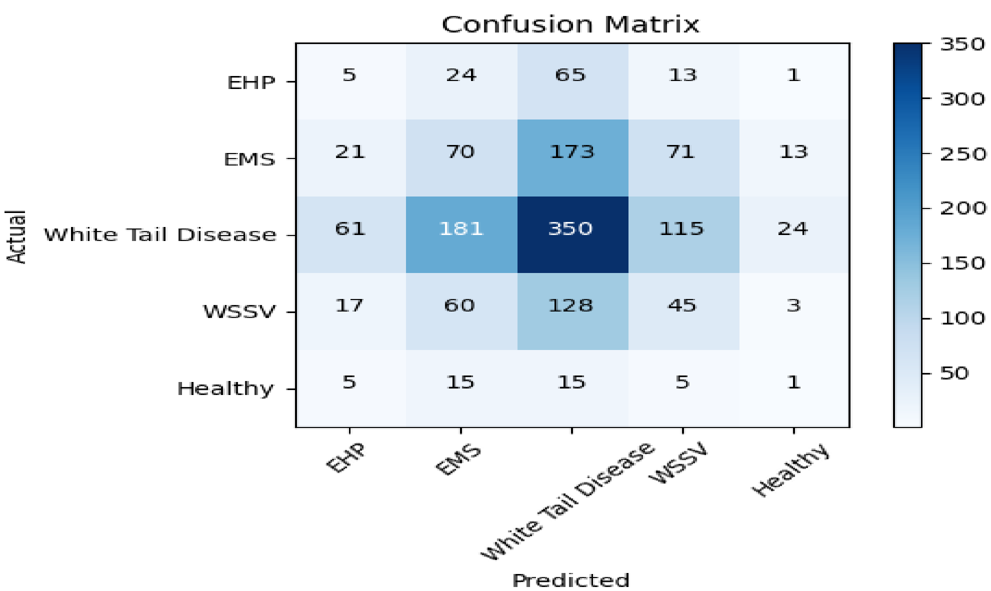

Fig. 19. Confusion Matrix (VGG 16 for 5 classes)

We observe similar case is evident in the VGG 16 as well where the model tends to generate lesser amount correctly predicted instances for shrimp disease detection. The model was not tweaked into two classes for better results unlike VGG 19 and ResNet50.

VGG-19 Confusion Matrix

Table: 11 Confusion Matrix (VGG19 for 2 classes)

True Labels		Predicted Labels: Disease	Predicted Labels: Healthy
	Disease	115	88
	Healthy	88	130

Table: 12 Precision, Recall and F1-score for VGG19

Class (for 2 classes)	Precision	recall	F1-score
0 (disease)	0.5662	0.5662	0.5662
1 (healthy)	0.5963	0.5963	0.5963

The model tends to achieve near 60% accuracy after tuning in the parameters and transferring the model into 2 classes. With further observation and tweaks the generated result can be even higher. The implication for further improvement can be made here.

Inception V3 Confusion Matrix

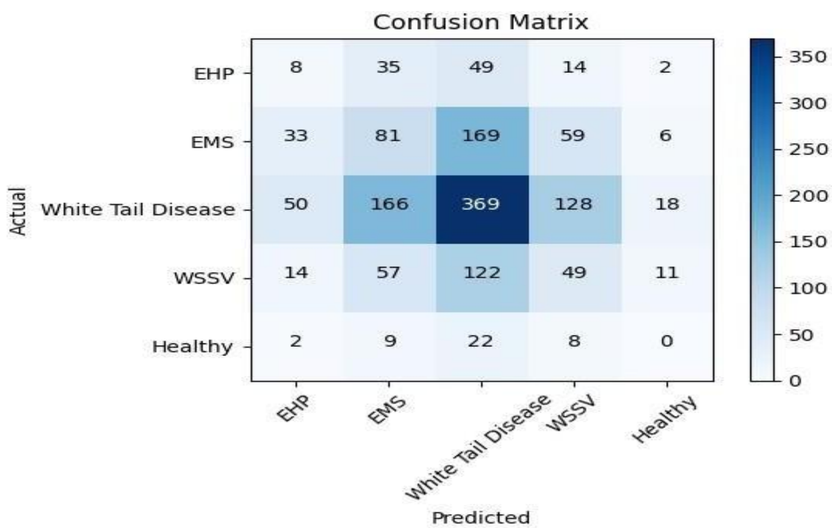

Fig. 20. Confusion Matrix (Inception V3)

Similarly to the VGG 16 this model was run only with 5 classes. White tail disease was once again accurately predicted most of the time. Although the main dataset had the most images in WSSV and EMS section it massively failed to predict both of these diseases implying that the model was faced with over fitting.

SVM Confusion Matrix

Table 13: Confusion Matrix (SVM for 2 classes)

True Labels	Disease	68	5
	Healthy	0	54
		Disease	Healthy
		Predicted Labels	

Table 14: Precision, Recall and F1-score for SVM

Class (for 2 classes)	precision	recall	F1-score
0 (disease)	0.9315	1.0000	0.9647
1 (healthy)	1.00	0.9153	0.9552

Random Forest Classifier Confusion Matrix

Table 15: Confusion Matrix (Random Forest Classifier for 2 classes)

True Labels	Disease	70	3
	Healthy	0	54
		Disease	Healthy
		Predicted Labels	

Table 16: Precision, Recall and F1-score for Random Forest Classifier

Class (for 2 classes)	precision	recall	F1-score
0 (disease)	0.9589	1.0000	0.9788
1 (healthy)	1.0000	0.9474	0.9726

From the confusion matrix the precision, recall and f1-score was calculated for the models used with 2 classes namely: ResNet50, VGG 19, SVM, Random Forest Classifier. Although the deep learning model excels the traditional models with higher accuracy and lesser loss functions the two deep learning models: VGG 19 and SVM drastically falls apart in precision, recall and f1 scores. It gives the impression that the deep learning models were not able to successfully differentiate diseased and healthy shrimps for most of the parts.

On the other hand, the two traditional models SVM and Random Forest classifier came up with perfect F1-score, precision and recall values. It means that they were able to successfully identify true positive values from false negatives and false positives. For example, the recall value for class 0 in random forest classifier was 1.00 meaning that it was able to successfully classify diseased shrimps all the time and never once did it identify a healthy shrimp as a diseased shrimp.

Based on the current observation between all 6 models, SVM and Random Forest Classifier are the models that performed the best. They both provided testing accuracy over 95% and performed wonderfully for detecting diseased and healthy shrimps separately.

Performance Evaluation

In this section we will compare our obtained result with the rest of the works that we have Used for our work. These related studies rendered our work in the correct direction and bestowed us to obtain better results for our models.

Table.15. Evaluation comparison with related works

Reference	Accuracy/Result
[16]	SVM accuracy was 83.64%
[17]	AHPND had the accuracy of 91.89%
[18]	Inception-V3 had the highest accuracy of 90.02%
[19]	ENRFCH had an accuracy of 98.28%
[20]	Logistic Regression had the highest accuracy of 83.04%
[21]	Normal shrimps: 96.7% Shrimps with HPND: 95.8% Shrimps with RBD: 96.7% Shrimps with WMD: 95.9%
[22]	KNN Model had the highest accuracy of 93.7%
[23]	DLCNN with TLOFS had the highest accuracy in this paper 99.98%.

Some of our have performed better than the models that have been mentioned in the related works section. Our SVM and Random Forest Classifier model have performed way better in the Test Accuracy section. Meaning it is better than other models in data that were not used in training. Our test Accuracy score is as follows for the two traditional models: SVM 96.06%, Random Forest 97.63%. Although the models of deep learning domain were rendered not as much accurate compared to the models in the tables a case can be made that the small number of dataset might be the reason behind the demise of the deep learning models. Furthermore, the deep learning models need a lot of tweaking and customization in the hyper parameter section. Further addition of better parameters can significantly improve the result of the VGG19 and SVM models.

Finally some of the models in the previous works established custom models based on neural network. So, there is a high implication that our work could be further nourished with the build and utilization of a custom model of our own. Until now, The SVM and Random Forest Classifier models hold their own against the models studied by us. The SVM and Random Forest with K-fold cross validation achieves significant improvements and excels the traditional models run by the other authors in the related field with higher precision, accuracy and performance.

8. Social and Environmental Impact of Engineering

Social Impacts: AI-based shrimp disease detection system improves disease management, economic stability, food security, environmental sustainability, and promotes knowledge transfer.

Disease Prevention and Management: Early detection of shrimp diseases enables prompt actions to prevent their spread, minimizing economic losses and improving farming practices.

Economic Stability: By mitigating disease outbreaks and improving shrimp health, the system maintains a stable production rate, ensuring a sustainable income for farmers and contributing to the local economy.

Food Security: Enhancing disease detection and management ensures a consistent supply of healthy shrimps, supporting local food security and the availability of a valuable protein source.

Environmental Sustainability: Real-time monitoring of water parameters prevents pollution and negative impacts on the ecosystem, promoting sustainable aquaculture and preserving the local environment and biodiversity.

Knowledge and Technology Transfer: Implementing advanced technology fosters collaborations, knowledge exchange, and adoption of innovative practices, enhancing efficiency and competitiveness in the shrimp farming industry.

Food Security: Shrimp is an important source of animal protein and a valuable export commodity. Disease outbreaks can disrupt the availability of healthy and disease-free shrimp for local consumption and international markets. By effectively detecting and managing diseases, this project promotes food security, ensuring a stable supply of safe and nutritious shrimp.

Public Health: Certain shrimp diseases, like EMS, can pose risks to human health if infected shrimp are consumed. By detecting and controlling diseases, this project reduces the likelihood of contaminated shrimp entering the food chain, protecting consumers from potential health hazards associated with diseased shrimp.

Environmental Impact: Improper shrimp farming practices, exacerbated by disease outbreaks, can have negative environmental consequences. Increased usage of antibiotics and chemicals to combat diseases can harm aquatic ecosystems. By enabling early disease detection, this project helps farmers adopt more sustainable practices, reducing reliance on harmful substances and minimizing the environmental impact of shrimp farming.

Farmer Well-being: Disease outbreaks can cause emotional distress and financial strain for shrimp farmers. By providing early detection and monitoring of diseases, this project supports farmers in managing their operations more effectively. Timely interventions and preventive measures recommended by the system can alleviate stress, improve overall well-being, and contribute to the resilience of shrimp farming communities.

Economic Impact: Shrimp diseases can result in significant financial losses for shrimp farmers and industry. By detecting diseases early and implementing appropriate measures, this project helps minimize economic losses, ensuring the financial stability of farmers and the shrimp industry in Bangladesh.

Legal Impacts: The project's legal impact encompasses intellectual property protection, data privacy compliance, ethical considerations, adherence to food safety standards, liability and responsibility concerns, and regulatory compliance. It is important to safeguard innovative developments, handle sensitive data appropriately, ensure animal welfare, meet food safety regulations, address liability issues, and adhere to relevant permits and regulations governing shrimp farming and technology deployment. Seeking legal guidance is advisable to navigate these legal considerations effectively.

Environmental Impacts: The environmental impacts of the project include water pollution from chemical discharge, habitat destruction through mangrove clearing, soil degradation from pond construction, chemical usage affecting water quality, loss of biodiversity due to habitat disruption, and greenhouse gas emissions contributing to climate change. Implementing sustainable practices such as responsible waste management, reduced chemical usage, habitat conservation, and exploring alternative energy sources can help mitigate these impacts.

9. Abbreviations

The following abbreviations are used in this article:

EMS	Early Mortality Syndrome
EHP	*Enterocytozoon Hepatopenaei*
WSSV	White Spot Syndrome Virus
WTD	White Tail Disease
WMD	White Muscle Disease
IoT	Internet of Things
SVM	Support Vector Machine
RFC	Random Forest Classifier
VGG	Visual Geometric Group
CNN	Convolutional Neural Network
RMSprop	Root Mean Squared Propagation
JPEG	Joint Photographic Experts Group
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

10. Reference

[1] FAO, “The State of World Fisheries and Aquaculture. Sustainability in action,” Rome, Italy: FAO, 2020, doi: 10.4060/ca9229en

[2] T. Islam, et al., “Status of shrimp diseases and their management practices at Satkhira in Bangladesh,” JOURNAL OF ENTOMOLOGY AND ZOOLOGY STUDIES, vol. 8, pp. 1017–1026, Jan. 2020.

[3] P. K. Patil et al., “Economic loss due to diseases in Indian shrimp farming with special reference to Enterocytozoon hepatopenaei (EHP) and white spot syndrome virus (WSSV),” Aquaculture,vol.533,p.736231,Feb.2021,doi: https://doi.org/10.1016/j.aquaculture.2020.736231.

[4] R. Adriman, M. Fitria, A. Afdhal and A. Y. Fernanda, “An IoT-Based System for Water Quality Monitoring and Notification System of Aquaculture Prawn Pond,” 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Solo, Indonesia, 2022, pp. 356-360, doi: 10.1109/COMNETSAT56033.2022.9994388.

[5] S. Karim, et al. “IoT Based Smart Fish Farming Aquaculture Monitoring System,” International Journal on Emerging Technologies. Apr 2021.

[6] Kun-Lin Tsai, Li-Woei Chen, Li-Jun Yang, Hung-Jr Shiu, Han-Wei Chen, “IoT based Smart Aquaculture System with Automatic Aerating and Water Quality Monitoring,” Journal of Internet Technology, vol. 23, no. 1, pp. 177-184, Jan. 2022.

[7] C. Dupont, P. Cousin and S. Dupont, “IoT for Aquaculture 4.0 Smart and easy-to-deploy real-time water monitoring with IoT,” 2018 Global Internet of Things Summit (GIoTS), Bilbao, Spain, 2018, pp. 1-5, doi: 10.1109/GIOTS.2018.8534581.

[8] C.-H. Chen, Y.-C. Wu, J.-X. Zhang, and Y.-H. Chen, “IoT-Based Fish Farm Water Quality Monitoring System,” Sensors, vol. 22, no. 17, p.6700, Sep. 2022, doi: 10.3390/s22176700.

[9] Y. Irawan, H. Fonda, E. Sabna, and A. Febriani, “Intelligent Quality Control of Shrimp Aquaculture Based on Real-Time System and IoT Using Mobile Device,” International Journal of Engineering Trends and Technology, vol. 69, no. 4, pp. 49–56, Apr. 2021, doi:10.14445/22315381/IJETT-V69I4P208

[10] M. Khan, S. I. Paul, Md. M. Rahman, and J. A. Lively, “Antimicrobial Resistant Bacteria in Shrimp and Shrimp Farms of Bangladesh,” Water, vol. 14, no. 19, p. 3172, Oct. 2022, doi:10.3390/w14193172.

[11] G. I. Jang et al., “Detection of Macrobrachium rosenbergii Nodavirus (MrNV) of White Tail Disease (WTD) in Apparently Healthy Giant Freshwater Prawn, Macrobrachium rosenbergii in Korea,” Fishes, vol.7, no. 5, p. 294, Oct. 2022, doi: 10.3390/fishes7050294.

[12] Rebecca S. Millard, Robert P. Ellis, Kelly S. Bateman, Lisa K. Bickley, Charles R. Tyler, Ronny van Aerle and Eduarda M. Santos,” How do abiotic environmental conditions influence shrimp susceptibility to disease? A critical analysis focussed on White Spot Disease,” Journal of Invertebrate Pathology, Volume 186, 22 November 2021, doi: 10.1016/j.jip.2020.107369.

[13] L. Liu, L.-P. Shan, Y. Zhou, and J. Chen, “Small Molecule Inhibitors of White Spot Syndrome Virus: Promise in Shrimp Seedling Culture,” International Journal of Molecular Sciences, vol. 22, no. 7, p. 3450, Mar.2021, doi: 10.3390/ijms22073450.

[14] D. Lee et al., “Viral Shrimp Diseases Listed by the OIE: A Review,” Viruses, vol. 14, no. 3, p. 585, Mar. 2022, doi: 10.3390/v14030585.

[15] Q. Wang, C. Qian, P. Nie, and M. Ye, “Rapid detection of Penaeus vannamei diseases via an improved LeNet,” Aquacultural Engineering, p. 102296, Oct. 2022, doi:10.1016/j.aquaeng.2022.102296.

[16] Hoang, Long & Quach, Luyl Da & Duong-Trung, Nghia & Nguyen, Chi-Ngon. (2019). TOWARDS MACHINE LEARNING APPROACHES TO IDENTIFY SHRIMP DISEASES BASED ON DESCRIPTION, doi:10.15625/vap.2019.00063.

[17] Khiem, N.M., Takahashi, Y., Yasuma, H. et al. Use of GIS and machine learning to predict disease in shrimp farmed on the east coast of the Mekong Delta, Vietnam. Fish Sci 88, 1–13 (2022). doi: https://doi.org/10.1007/s12562-021-01577-8.

[18] Duong-Trung, Nghia & Quach, Luyl Da & Nguyen, Chi-Ngon. (2020). Towards Classification of Shrimp Diseases Using Transferred Convolutional Neural Networks. Advances in Science Technology and Engineering Systems Journal. 5. 724-732. doi: 10.25046/aj050486.

[19] Onyema, Edeh & Dalal, Surjeet & Obagbuwa, Ibidun & Prasad, B V V S & Zanzote Ninoria, Shalini & Wajid, Mohd & Adesina, Ademola. (2022). Bootstrapping random forest and CHAID for prediction of white spot disease among shrimp farmers. Scientific Reports. 12. 20876. doi: 10.1038/s41598-022-25109-1.

[20] Khiem, N.M., Takahashi, Y., Oanh, D.T.H. et al. The use of machine learning to predict acute hepatopancreatic necrosis disease (AHPND) in shrimp farmed on the east coast of the Mekong Delta of Vietnam. Fish Sci 86, 673–683 (2020). doi: https://doi.org/10.1007/s12562-020-01427-z.

[21] N. Duong-Trung, L.-D. Quach, and C.-N. Nguyen, “Towards classification of shrimp diseases using transferred convolutional neural networks,” Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 4, pp. 724–732, 2020. doi:10.25046/aj050486.

[22] R. P. S. Sankar, V. Jenitta, B. Kannapiran and K. Valarmathi, "Diagnosis of shrimp condition using intelligent technique," 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), Karur, India, 2017, pp. 1-5, doi: 10.1109/ICEICE.2017.8191957.

[23] Varma & Krishna “Transfer learning-based optimal feature selection with DLCNN for shrimp recognition and classification,” International Journal of Intelligent Engineering and Systems, vol. 15, no. 5, pp. 91–102, 2022. doi:10.22266/ijies2022.1031.09.

Appendix

CO	Knowledge Profile	Engineering Problem
CO1	(i) Background [K1, K2, K3] K1: Detecting diseases in shrimp and water contamination falls under natural science of Biology. K2: Use of mathematical operation in disease detection using images in a popular machine learning model. K3: Use of Deep learning techniques.	(i) Background [EP1] Data modeling knowledge required. [K3] Different machine learning based model knowledge required. [K4] Web application and Mobile application design knowledge required. [K5] Image processing and shrimp disease research field. [K8] [EP6] Shrimp farmers, IoT service providers, IoT device manufacturers, Shrimp consumers, Shrimp researchers, Shrimp regulators will be the stakeholders for this project.
CO2	(i) Related Works [K8] We have studied papers related to shrimp disease, water contamination image processing and IoT-based water monitoring system to understand how it works and how we can build the system that can help farmers and other stakeholders.	(i) Related Works [EP1] Studied different papers on detecting diseases and water contamination. [K8] Knowledge of building web and mobile applications. [K4] Knowledge of building Arduino controllers for building IoT devices. [K4] Knowledge about popular machine learning models and their application. [K5, K6] (ii) Objectives [EP2, EP6, EP7] The system should be able to correctly identify anomalies in water and shrimp. It should also be user-friendly to the farmers. [EP2] Creating a business ecosystem for Shrimp farmers, IoT service providers, IoT device manufacturers, Shrimp consumers, Shrimp researchers and Shrimp regulators to work

		together to create a sustainable business. [EP6] Many different problems can occur. The sensors may fail or report false information. The internet connection may be unreliable. [EP7] (iii) Planned Methodology [EP2, EP6] Collecting image data and making a dataset to train the machine learning model. Testing different models to find the best result. Integrating the model to web server. [EP2] Contacting vendors to provide sensors and IoT controller. IoT service provider must provide the service. The farmers must adopt the new technology. [EP6]

Contribution

Name	Contribution
Anabil Sarker	100% Contribution on the following: • Image Preprocessing • Image Augmentation • Image Segmentation for SVM and Random Forest Classifier • Trained SVM, Random Forest Classifier, Inception V3, VGG 16, VGG 19 and ResNet 50 • Added all graphs of the model's performance and generated graphs • Hyperparameter Tuning • Plotting Graphs and Confusion Matrix • Performance Evaluation
Md Shorif Hossen	• Background • Objective discussion • Slide formation • Reference • Data collection • Related work
Ashiquzzaman Choudhury	• Data collection (diseased shrimps) • Related works (literature review) • Result discussion • Model evaluation • Research questions • Objective discussion.
Noortaz Ahmed	• Data collection (both diseased & healthy shrimps) • Relate works (literature review) • Model evaluation (diagrams) • Methodology (diagrams & discussion) • Social and environmental impact discussion • Reference & report formatting