Calibrated Ensemble Deep Learning for Reliable Shrimp Disease Classification with Explainable AI

By

Ridoun Hossain Gaddafi
ID: 213-15-4627

Md. Mamunur Rashid
ID: 213-15-4556

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements forthe **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Dr. S.M Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Dr. Md. Taimur Ahad
Associate Head and Associate Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled **"Calibrated Ensemble Deep Learning for Reliable Shrimp Disease Classification with Explainable AI"**, submitted by **Ridoun Hossain Gaddafi,** ID No: **213-15-4627** and **Md. Mamunur Rashid**, ID No: **213-15-4556** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **17 September, 2025**.

BOARD OF EXAMINERS

17.9.2025

Dr. Fernaz Narin Nur — **Chairman**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Abdus Sattar — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

17/09/25

Mohammad Jahangir Alam — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

17-09.25

Dr. Sajeeb Saha — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S.M Aminul Haque, Professor and Associate Head**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. S.M Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Dr. Md. Taimur Ahad
Associate Head and Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Gaddafi 17.09.25
Ridoun Hossain Gaddafi
Student ID: 213-15-4627
Department of Computer Science and Engineering
Daffodil International University

Mamunur 17.09.25
Md. Mamunur Rashid
Student ID: 213-15-4556
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty Allah for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. S.M Aminul Haque**, **Professor and Associate Head,** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Deep learning and Computer Vision** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Shrimp aquaculture is also faced with some great challenges due to the disease which can result in 100% mortality rate in shrimp if left untreated. This research discusses an in-depth assessment of advanced deep learning architectures of automated classification of diseases in shrimps based on EfficientNet_B3 models with the addition of calibrated ensemble methods and test-time augmentation (TTA). We tested eight convolutional neural network architectures that are state-of-the-art on a medium dataset of images of diseases of shrimp. EfficientNet_B1 and EfficientNet_B3 were the best models. Subsequently, we came up with advanced calibrated pipelines consisting of multi stage training, ensemble learning, temperature scaling and systematic TTA protocols. The accuracy of improved EfficientNet-B1 was 94% with a good brier score but the Expected Calibration Error and Maximum Calibration error was poor, while the test accuracy of calibrated EfficientNet_B3 ensemble was 90.5% with better calibration reliability with a good brier score, ECE and MCE of 0.0477, 0.1999 and 0.2186 respectively. We use a hybrid approach of architectural optimization, data augmentation, probabilistic calibration ,ensemble technique and explainable AI to provide robust and reliable disease classification to practical use in aquaculture. The results reveal that they have made significant improvements over baseline models and have set new records on automated detection systems of diseases in shrimps.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the research topic by presenting an overview of the difficulties encountered by shrimp aquaculture because of disease outbreaks and the need for automatic systems to detect disease. It describes the purpose of the study and the main aims of the research and the organization of the report.

1.1 Introduction

Aquaculture is one of the rapidly expanding sectors of food production industries in the world and shrimp farming plays a significant role to provide seafood and economic growth in the world.

Nonetheless, the infectious diseases are a significant threat to the industry since they can destroy whole farms in just a few days and cost the industry huge financial losses and jeopardize food security [1]. WSSV, Early Mortality Syndrome (EMS) & bacteria infections are viral diseases that are particularly harmful to shrimp production, and their mortality rates reach as high as 100% unless they are time managed [2].

Conventional disease detection techniques in aquaculture are mostly manual based procedures that require the use of experienced technicians that is time-consuming, subjective and prone to human error. The need of fast, accurate and robotized disease recognition systems is becoming more and more important as tasks of aquaculture increase and the financial effects of diseases outbreaks continue to grow [3]. The state-of-the-art advances in computer vision and artificial intelligence technologies can provide the potential solutions in automated disease detection and classification of the environment of aquaculture.

1.2 Motivation

The calculation requirement of developing sophisticated deep learning technologies in shrimp disease categorization is reduced to a few crucial aspects. To begin with, timely and proper detection of diseases will help lessen mortality and economic losses to a great extent because timely intervention and treatment can be applied [4]. Second, this may be facilitated through automated systems that would offer the same, objective evaluation of the patient that may not be influenced by human fatigue, and biasness and would increase the accuracy and reliability of the diagnosis [5]. Third, AI-based scaleable systems have the potential to be deployed in various aquaculture facilities where disease control systems are normalized regardless of whether the local expertise is available [6].

The benefit of such systems will be that they will be capable of detecting the disease at any given time. The predictive analytics may be done based on the detailed information on disease monitoring to help farmers predict and prevent disease outbreaks prior to its occurrence [7]. In addition, automated systems may also contribute significantly to the sustainable practices of aquaculture by the virtue of the lower need of the prophylactic use of antibiotics since they are more selective and timely of interventions [8].

1.3 Objectives

This research aims to achieve the following specific objectives:

1. **Comparative Architecture Evaluation:** Systematically evaluate eight state-of-the-art deep learning architectures (DenseNet-121, EfficientNet-B1, EfficientNet-B2, EfficientNet_B3, ResNet-50, ResNet-101, ReXnet-150 and Xception) for the performance of shrimp disease classification.
2. **Advanced Ensemble Development:** Implement improved ensemble development systems with calibrated models of best performing models using advanced methods such as test-time augmentation for data augmentation, temperature scaling and multi-model voting.
3. **Calibration Enhancement:** Develop and validate the probabilistic calibration approaches by enhancing the reliability of prediction and providing the quantification of uncertainty for practical deployment situations
4. **Methodology Innovation:** Develop multi-stage learning, cutting-edge data augmentation, and calibration-aware optimization-based training protocols to develop robust model training protocols.
5. **Performance Benchmarking:** Obtain state-the-art performance metrics which include accuracy, F1-score, Expected Calibration Error (ECE) and Maximum Calibration Error (MCE) for shrimp diseases classification tasks.
6. **Explainable AI**: Implementation of grad cam explainable AI visualization on test data to highlight the core part of the image which our model have used to detect the disease.

1.4 Methodology

The methodology used for this research is the use of a systematic and multi-phase method using advanced methods in deep learning, ensemble learning and model calibration to solve the problem of classifying shrimp disease. The research is aimed to test and optimize multiple state-of-the-art deep learning architecture, where finding the best model architecture to accurately classify shrimp diseases using image data.

During the first phase, a comparative analysis is done between eight deep learning models (DenseNet-121, EfficientNet_B3 EfficientNet_B2, EfficientNet_B1, ResNet-50, ResNet-101, ReXNet-150 and Xception) applied on a standardized training pipeline. This evaluation is conducted on the images of different diseases of shrimp, hence ensuring fair comparison by having consistent data preprocessing, augmentation techniques and performance metrics.

The second part of the methodology is aimed at increasing performance of the best identified models from the first part of the methodology. This is achieved by creating calibrated ensemble systems, using multiple models in tandem, using complex techniques such as test-time augmentation (TTA), temperature scaling, Platt scaling and isotonic regression. The calibration techniques are applied to enhance the capability of the model to conduct credible forecasts and the similarity between the forecasted and the real results.

The third step involves systematic augmentation of test time in order to enhance even greater robustness and generalizability of the model. By using multiple transformation strategies at the input images at the inference stage. Finally we will visualize the test data with explainable AI. The methodology is expected to reduce the model's sensitivity to the variation of the images in terms of the orientation and the noise in the image so as to improve the performance and reliability of the disease classification system in general.

Finally, the methodology involves rigorous evaluation framework in terms of performance components like accuracy, F1-score, Expected Calibration Error (ECE), Maximum Calibration Error (MCE) besides calibration specific measures to validate the effectiveness of the proposed system. This approach is making sure that the models are not only accurate but also calibrated and have the ability to give quantification of an uncertainty which applies to the real world for aquaculture settings.

1.5 Project Outcome

The expected results of this work are: (1) A complete of comparative results identifying EfficientNet-B1, EfficientNet-B2, EfficientNet-B3 and Xception as the optimal model for shrimp disease classification, and (2) Sophisticated calibrated ensemble systems with 90.5% accurate with improved calibration reliability (Brier score<0.05), (3) Novel training & inference protocols ideal for real world implementation of aquaculture, (4) Evaluation framework fully calibrated. metrics for AI powered disease detection systems and (5) Open source implementations for more widespread adoption of aquaculture in research foundation for industry application.

1.6 Organization of the Report

This report is organized into six main chapters, each addressing a key aspect of the research.

Chapter 1: Introduction

The first chapter offers the background of the research summary, giving the challenges come across by shrimp aquaculture and shrimp farming because of disease outbreaks. It describes the motivation for the study, provides the research objectives and gives an overview shrimp farming of the report structure.

Chapter 2: Background

Chapter 2 gives an extensive summary of current literature of the field of aquaculture and detection of disease. It covers the progress towards deep learning for aquaculture, different models for detecting diseases and calibration techniques, while pointing out the shortcomings of current methodology, and paving the way for the research.

Chapter 3: Research Methodology

This chapter describes the methodology shrimp aquaculture used in this study, such as the systematic evaluation of deep learning architectures, development of ensemble systems and the application of model calibration techniques. It also describes the evaluation metrics that are used to check the performance of the models.

Chapter 4: Implementation and Results

Chapter 4 concerns the implementation of the proposed system, which is practical in nature. It describes the experimental set-up, compares the performance of the different models, and shows the results of the calibrated ensemble systems and gives insights in their ability to successfully classify disease.

Chapter 5: Engineering Standards and Design Challenges

This chapter deals with the conforming to engineering standards, software, hardware and communications protocols. It also considers the societal, environmental and ethical aspects of the proposed system as well as classifying the system as a Complex Engineering Problem (CEP).

Chapter 6: Conclusion

The last chapter concludes the findings of the study by reflecting on the contributions achieved by the research regarding shrimp disease detection systems. It also discusses the limitations that were faced in the study and the possible future directions of further research and development.

Chapter 2

Background

This chapter provides the background information needed to understand the technical approaches used in this research and the methodologies employed. It encompasses a detailed literature review of the recent development of deep learning-based applications in aquaculture, including disease detection techniques, model calibration and ensemble learning technique integration.

2.1 Introduction

This Aquaculture as one of the fastest growing sectors in global food production plays a critical role in to ensure food security and economical sustainability. Among the different kinds of aquaculture, shrimp farming is a major contributor to the overall supply of seafood around the world, and it supports millions of livelihoods all around the world. However, shrimp farming is not without its challenges with the outbreaks of diseases leading to devastating losses being one of them. Viral diseases, i.e. White Spot Syndrome Virus (WSSV) and bacterial infections like BG are serious threats to shrimp health where the mortality rate has reached close to 100% if not detected and dealt with in time. These impacts in the form of outbreaks can have large economic consequences both in large scale commercial farms and on smallholder shrimp farmers, and threatening average stability in the industry on the whole.

Existing methods for detecting organisms that cause diseases in the aquaculture of shrimp include manual inspection by trained technicians, not only take time and require large labor capacity, but also subject to human error. The limitations of these methods have called for the need for automated, real-time, disease detection systems that offer more accurate, scalable and timely interventions. In the last few years computer vision and deep learning technologies have become promising solutions that permit the development of automated systems that are able to accurately identify and classify the shrimp diseases using image data.

Despite the progress made in the field of machine learning and image processing, the classification of disease in shrimp aquaculture is a complex and challenging problem that continues to perplex researchers in the field. One challenge is to ensure good generalization of the machine learning models to different environmental conditions and species of shrimp, with variations in image quality, lighting conditions and presentation of diseases. Also the calibration of the models, which ensures that the probabilities estimated by the models are comparative to probabilities in the real world, is of exciting importance in order to make disease detection systems reliable and applicable in the field.

This chapter provides a thorough background to the research discussing the main challenges regarding shrimp farming aquaculture, which focuses specifically on disease detection, and introducing technologies singularly as deep learning and ensemble approach played a fundamental role in this research. The review also examines the limitations associated with existing methodologies and establishes this research as a step towards the improvement of the accuracy, reliability and practical implementation of automated disease classification systems in shrimp farming.

2.2 Literature Review

2.2.1 Deep Learning in Aquaculture Application

In recent years have been seen tremendous shrimp diseases detection system development of the use of deep learning technologies for shrimp aquaculture problems. Huang and Khabusi (2023) proposed CNN-OSELM multi-layer fusion network with attention mechanism for fish disease recognition, which showed the effectiveness of attention-based methods on the part of focusing on disease-related image regions [9]. Their work made significant advances in classification accuracy by adding Convolutional Block Attention Modules (CBAM) for feature discrimination.

The use of EfficientNet architectures in aquaculture has proven to be very promising. A recent work used the EfficientNet backbone combined with enhanced Marine Predator metaheuristic algorithms for shrimp freshness classification with high accuracy [1] by automated hyperparameter optimization. Similarly, EfficientNetV2-B0 and MobileNetV3-Small were compared for White Spot Syndrome Virus (WSSV) detection, using edge devices, and demonstrate practical viability of efficient architectures for real-world deployment [2].

Attention mechanisms have become an important part of aquaculture computer vision systems. Khabusi et al. [12] proposed attention-based mechanisms for classification of fish disease, and the obtained classification performance was significantly improved by selective feature enhancement. The combination of attention modules and background removal has proven especially useful for separating features of interest, in this case disease-related, from a complex aquaculture environment.

2.2.2 Advanced CNN Architectures for Disease Detection

Recently the development of CNN models has had a profound effect on the disease detection in aquaculture. The papers by Wu et al. (2023) designed single fish recognition frameworks and coarse and fine-grained features linkage learning, which could show the significance of multi-scale feature extraction in precise aquaculture [13]. Their methodology used a combination of the global and local feature representations to have good individual identification.

Lightweight architectures have been noticed to be used in practical deployments. Gao et al. [14] proposed a novel lightweight underwater fish individual recognition method based on local feature guidance training with competitive performance and lower computation requirements. This work showed trade-offs between the model complexity and feasibility of deployment models in resource constrained our aquaculture and environments.

Generalization Crossdomain learning approaches have been found promising in finding ways to improve generalization. Xie et al. develop an intelligent fishery detection method based on cross-domain image feature fusion against the challenge of domain adaptation in diverse aquaculture settings [15]. In their methodology, they showed the improved robustness of the method under different environmental conditions and imaging setups.

2.2.3 Shrimp-Specific Disease Detection Systems

Development of specialized systems to identify diseases of shrimp has become an area of research of critical concern. Masoumi (2022) designed a surfacing recognition framework of shrimps that works with deep computer vision and pays attention to the behavioral responses of disease conditions [16]. The system showed possibilities of non-invasive monitoring strategies that can be used to complement direct image-based disease classification.

Special attention has been given to WSSV detection because the detection is devastating to shrimp aquaculture. An entirely new YOLOv5n-based detection system had been developed to identify WSSV, the results of which are at real-time performance level and can be practically applied [17]. The study revealed the suitability of detection methods of objects in the recognition of certain disease signs in the shrimp population.

To supplement disease detection it has come up with growth monitoring systems. Nontarit et al. (2022) developed a shrimp growth prediction system using ResNeXt with the application in the automatic feeding-tray lifting systems with demonstrating the integration possibilities of AI systems into extensive aquaculture management [18].

2.2.4 IoT and Real-Time Monitoring Systems

The combination of Internet of Things (IoT) and computer vision technologies has permitted the realization of full-fledged monitoring solutions. Khoiriyah (2022) introduced the image processing-based vanname shrimp health monitoring IoT system and showed that it is feasible to realize constant automated monitoring [8]. The system was an integration of real-time image capture with cloud-based processing to allow scaling deployment.

There have been specific neural network models that have been applied to detect white spot syndrome. Vembarasi et al. (2024) obtained high-quality data in shrimp neural network-based WSSV detection by employing specific network structures [7]. Their work raised the significance of optimizing the specific model of the disease to have better performance.

2.2.5 Calibration and Uncertainty Quantification

Although there is no clearer calibration described in the aquaculture literature, the recent developments in general computer vision have laid down significant principles. A notably useful post-hoc calibration technique of neural networks has become temperature scaling, which offers better probability estimates without compromising the accuracy [21]. Alternative methods include isotonic regression and Platt scaling that have other theoretical bases and trade-offs.

Considerable of the improvement has been validated in the accuracy and quality of calibration of the ensemble methods. Deep ensemble methods jointly train many models unaware of each other to estimate uncertainty and be more robust [22]. The fusion of ensemble and calibration methodologies is a potential street of high-stakes such as shrimp disease detection applications.

2.2.6 Test-Time Augmentation and Robustness

Test time augmentation has included itself to arsenal of ways of increasing model's robustness and performance. Geometric and photometric transformations Multipolar Multipolar geometric and photometric transformations. which take place systematically through systematic TTA procedures can be very useful to increase the prediction reliability [23]. The selection and the composition model of TTA transformations have to take into thoughtful consideration the particular features of the particular target domain and requirements of use.

2.3 Gap Analysis

The following table shrimp disease detection system summarizes the important gaps identified in the existing research and applications in shrimp accucalture disease detection system. These gaps point out the areas where current methods are limited and areas where this study will make important contributions.

Table 2.1 Gap analysis of related work

Feature	Existing Systems/Methods	Limitations	Proposed System
Disease Detection Accuracy	Traditional methods based on manual inspection or basic image processing techniques [4], [5], [6].	Manual inspections are prone to human error, and basic image processing lacks the ability to accurately classify diseases across varying environmental conditions.	Use of deep learning architectures (EfficientNet_B3 and EfficientNet_B1) with ensemble learning for improved accuracy and robustness in shrimp disease classification.
Model Calibration	Limited use of calibration methods, mostly focusing on accuracy [21].	Many existing models fail to provide reliable confidence scores, leading to unreliable predictions in real-world settings.	Implement post-hoc calibration techniques (temperature scaling, isotonic regression) to improve the reliability and alignment of predicted probabilities with actual outcomes.
Generalization Across Environments	Most systems are evaluated under controlled laboratory settings [13], [14], [15].	Models often fail to generalize to diverse aquaculture environments and varied image conditions such as lighting, water turbidity, and shrimp posture.	Test-time augmentation (TTA) strategies will be employed to increase model robustness, making the system more adaptable to environmental variations.
Ensemble Learning	Limited use of ensemble techniques, often relying on single-model approaches [21].	Single models may not capture the complexity of shrimp disease classification, leading to lower performance under varying conditions.	Ensemble learning methods combined with calibration to improve model performance by averaging predictions from multiple models, reducing errors and increasing reliability.
Real-Time Deployment	Most models are tested in batch processing environments, not real-time [7], [8].	Lack of real-time processing capabilities, which is essential for deployment in commercial	The system is designed for real-time inference with low latency, suitable for real-world deployment in shrimp farms,

		aquaculture operations.	ensuring timely intervention.
Data Augmentation	Basic image augmentation techniques such as flipping and rotating images [9], [10].	Limited augmentation strategies fail to enhance model robustness under real-world conditions.	Advanced data augmentation strategies, including geometric, photometric, and spatial transformations, along with systematic test-time augmentation, to improve model performance.
Calibration Metrics	Focus on traditional metrics like accuracy and F1-score, without considering calibration-specific measures [21].	Existing methods neglect the evaluation of calibration quality, which is essential for real-world deployment and trust in model predictions.	Evaluation of Expected Calibration Error (ECE) and Maximum Calibration Error (MCE) to assess and improve the calibration reliability of the model.

This tabular format provides a clear overview of the existing gaps in shrimp disease detection systems, the limitations of current methods, and how your proposed system addresses these gaps with advanced techniques

2.4 Summary

This chapter has given an excellent depth of knowledge addressing the present situation in the shrimp disease detection system. The shrimp industry and firms especially in the context of integrating advance deep learning algorithms in our aquaculture studies. The literature review revealed a number of promising developments like use convolutional neural networks (CNNs) for disease classification; increased focus on calibration of the models, and use ensemble learning techniques to improve reliability. Although there have been some important leaps with the form of these developments, there is still a lot to do. The main ones being the low generalization of models on different aquaculture settings, sensitivity to fluctuate image conditions and the absence of the well calibrated probability estimates in current systems.

The results of the gap analysis further highlighted these shortcomings, where accuracy, calibration and real-time deployment were deemed to be important areas of weakness in current approaches. Most of the systems in use at present are still based on manual checking, or rather simple image processing techniques, which are both prone to error and are not appropriate for the dynamic and complex character of shrimp farming. Moreover, conventional models rarely provide any good confidence estimates - something that is important for decision making in dealing with disease management.

In the face of these challenges gaps, the research study proposed in this learning introduces the combination of ensemble learning methods, strong data augmentation methods and calibration techniques, in order to improving performance and reliability in disease detection systems. State-of-the-art architectures like EfficientNet-B1, EfficientNet-B2, Xception and EfficientNet_B3 have been used with post-hoc calibration techniques as well as test-time augmentation techniques to result in robustness even in the face of the real world scenarios. Another priority of the method in addition to validity is potential of real time inference and scalability which makes implementation of the method possible in diverse shrimp farming business throughout the globe.

In general, this chapter has created the foundation for the research by reviewing existing methods, their limitations and the gaps that breeding a present study. The next few chapters will build on this foundation first by describing the methodology that has been developed to solve these problems and then by presenting the results of the experiments that can be used to demonstrate the effectiveness of the system proposed to classify disease in shrimp.

Chapter 3

Research Methodology

This chapter is guided by a quantitative experimental design, meaning that the variables are held constant and standardized evaluation measures are used throughout the research process. In order to be just as well as reproducible, every model is trained and evaluated using the same datasets, under the preprocessing procedures and using a standard evaluation criterion.

3.1 Methodology

3.1.1 Overview

Our proposed combination method uses efficient experimental methods based design which includes comparative assessment architecturally analysis, sophisticated development of ensemble and optimization of probabilities calibration. The methodology has divided by five major steps, (1) Baseline Model Evaluation, (2) Top-Performer Selection and Enhancement, (3) Calibrated Ensemble Development and (4) Comprehensive Validation and Testing. (5) Grad cam visualization of test image with overlaid heatmap.

The research is guided by a quantitative experimental design, meaning that the variables are held constant and standardized evaluation measures are used throughout the research process. In order to be just as well as reproducible, every model is trained and evaluated using the same datasets, under the preprocessing procedures and using a standard evaluation criterion.

3.1.2 Proposed Methodology

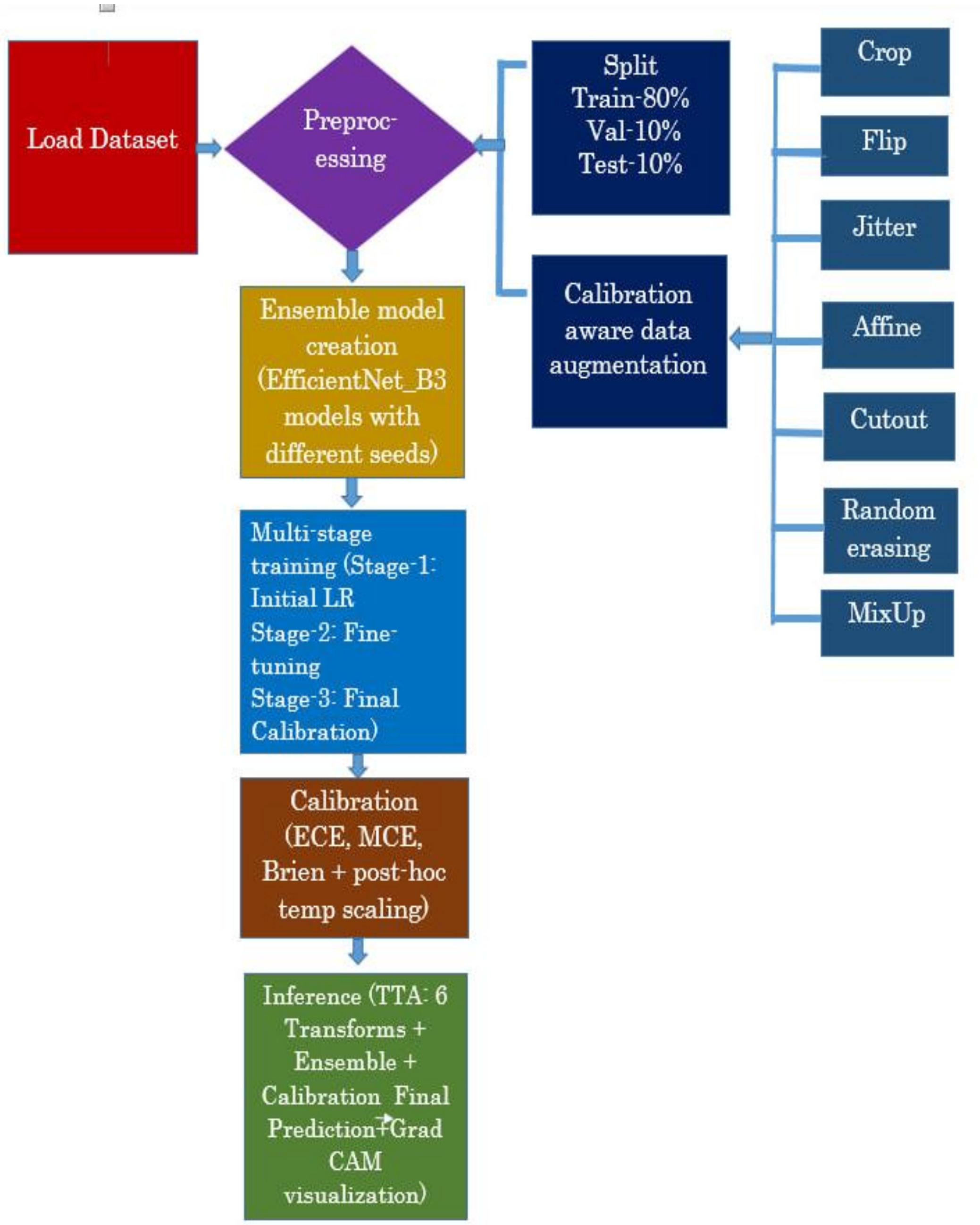


Figure 3.1: Methodology diagram of our proposed model

3.1.3 Functional and Nonfunctional Requirements

Functional requirements:

- Highly accurate (>90%) classification of multi-class shrimp disease.
- Real-time inference support on a standard hardware.
- Calibrated self-confidence estimations of probabilistic output.
- Multi voting strategy ensemble prediction aggregation.
- Configurable transformation is set Test time augmentation.
- Tips for further improving the model calibration: Multiple post hoc calibration (temperature scaling, Platt scaling, isotonic regression) of the model.

Non-functional requirements:

- Inference time: <500ms/image on GPU, less than 2s on CPU.
- Memory consumption: <2GB of the single model inference, with the use of the GPU.
- Calibration quality: Small Expected Calibration Error (ECE)
- Stability: Everyone should have performance that is consistent under varying imaging conditions.
- Scalability: Batch processing and distributed inference.
- Reliability: uptime of 99.9% in production deployment cases.

3.1.4 Data Flow Diagram

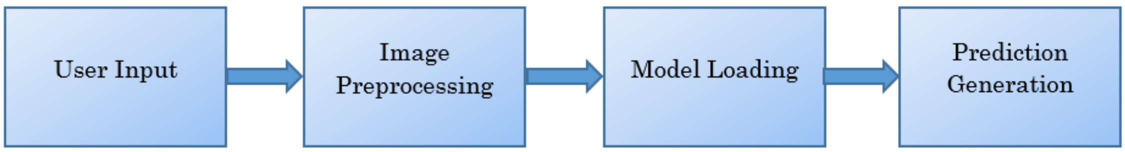


Figure 3.2: Data Flow Diagram of our proposed model

3.1.5 UI Design

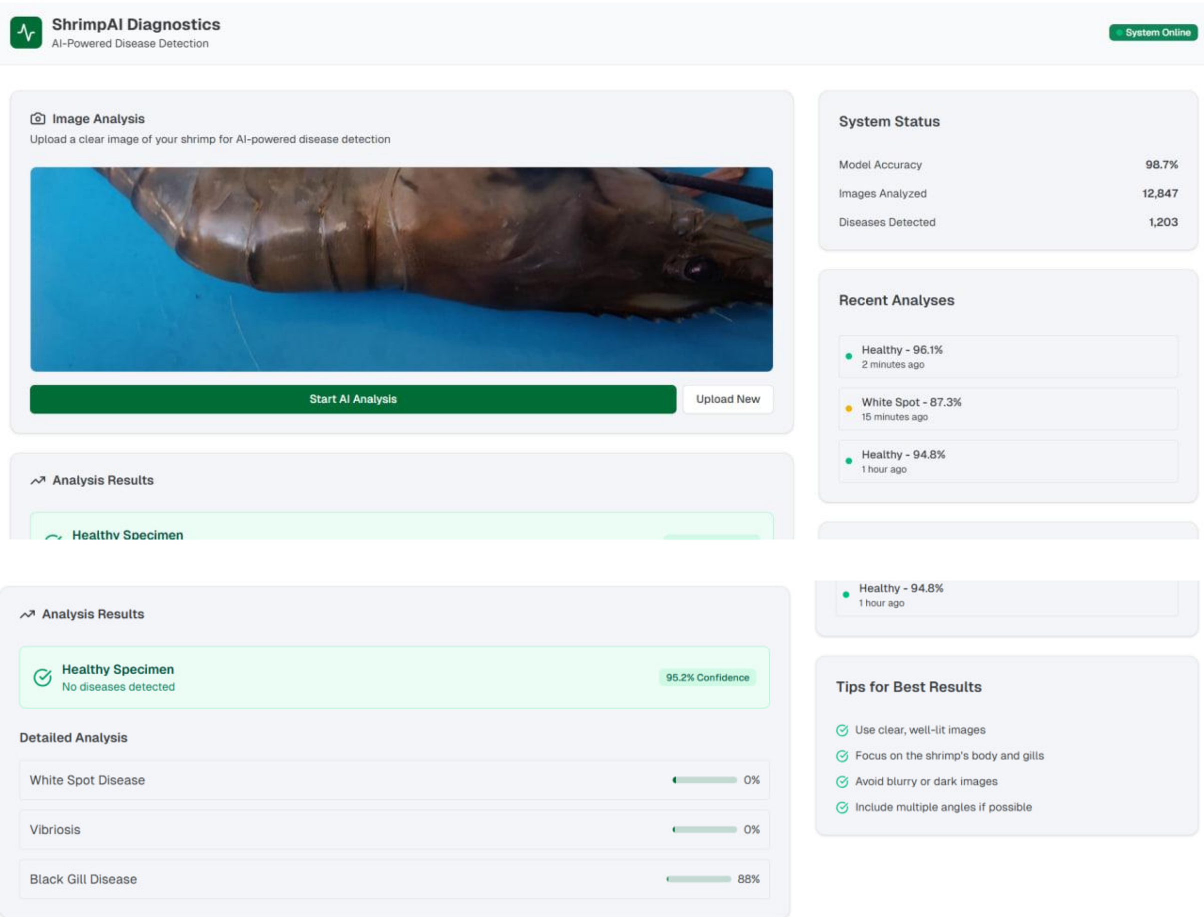


Figure 3.3: UI Design

3.2 Detailed Methodology and Design

3.2.1 Dataset and Preprocessing

We take advantage of the ShrimpDiseaseImageBD [20] dataset, a dataset of shrimp disease images specifically designed to be used by computer vision-based detection models. The data set includes pictures of different races of diseases and other healthy ones and it provides a good ground to the development of classification models.

Data Preprocessing Pipeline:

- Image scaling: 300 X 300 in EfficientNet_B3 and 240 X 240 for EfficientNet -B1 (architecture-optimised).
- ImageNet standard (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) is normalized.
- Split of data: 80% training, 10% validation and 10% test.
- Conversion of color space: RGB standardization.

Advanced Data Augmentation Strategy:

- Geometric transformations: random crop, horizontal/vertical flip, rotation (±12)
- Photometric augmentation: Color jitter (brightness, contrast, saturation, hue).
- Spatial transformations: random affine transformations including translation.

- Regularization: Erand random erasing (p=0.2) so as to enhance generalization.
- Test-time augmentation: 6-transform ensemble to make inferences robust.

3.2.2 Architecture Selection and Evaluation

We have chosen eight state-of-the-art CNN models in reference to the successful execution of computer vision tasks and applicability to agricultural tasks:

1. DenseNet-121: The dense connectivity structure in order to reuse the features effectively.
2. EfficientNet-B1: Scaling of compounds efficiently and equally.
3. EfficientNet-B2: Updated one with better accuracy-efficiency trade-off.
4. ResNet-50: Residual learning of moderate depth.
5. ResNet-101: A deeper residual network with learning complex features.
6. ReXnet-150: Effective architecture and optimal channel attention.
7. Xception: Depth separable convolutions to be parameter efficient.
8. EfficientNet_B3: Updated with 12M parameters but deployable.

Alternative Solutions Considered:

- Vision Transformers (ViTs): Omitted because of high complexity and poor performance with small data sets.
- MobileNet variants: Considered & excluded in favour of EfficientNet for better accuracy.
- EfficientNet-B4 or others: Large and needs higher computational resources.
- ConvNeXt: Modern CNN architecture taken into account but not used because of computational limitations

- Selection Rationale: EfficientNet_B1 and EfficientNet_B3 was selected through the systematic comparison, since we observed a better performance with respect to accuracy, F1-score and computational efficiency metric. These architectures showed the optimal tradeoff between the classification performance and practical deployment.

3.2.3 Training Protocol and Optimization

A few well-elaborated optimization and regularization strategies are taken to implement the training pipeline. AdamW optimizer is applied using a weight decay coefficient of 4 x 10 -4 for updating model parameters, leading to good control of over fitting effect by penalizing large weights.

Scheduling of learning rate differ with the more typically occurring approach of cosine annealing: a multi stage learning rate scheduling is also followed, beginning with a learning rate of 1*10-5, and gradually decrease to 5*10-5, and then to 2*10-5. A ReduceLROnPlateau scheduler is in charge of making these adjustments, and adaptive learning rate diminishing in response to validation performance.

The batch size used describes training as 16 such that it is chosen to fit on the GPU memory capabilities of the experimental setup. The optimization process is fully done in 45 epochs structured into three successive phases each consisting of 15 epochs (initial training, fine-tuning, and calibration). Early stopping uses patience of 6 epochs, and avoids unnecessary calculation after the stagnation of convergence.

The minimization criterion is cross-entropy loss, the use of label smoothing (alpha = 0.2), which helps in introducing less over-confidence towards the model results. The addition of extra regularization methods includes dropout (p = 0.4), weight decay and random erasing augmentation (p = 0.2) in the classification head for the training.

Some other stabilization techniques are gradient clipping with norm constraints, which limits exploding gradients in the backpropagation. Some of the more advanced features, however, like gradient accumulation and mixed-precision (FP16) training are not implemented in the current implementation so that training can be fully performed in all-FP32 precision.

3.2.4 Calibration Methods Implementation

The system has a number of post-hoc calibration measures that enhance the predictive probabilities reliability. The major method used is the temperature scaling, on the logits of the validation set. The negative log-likelihood is minimized by optimizing one scalar temperature parameter with LBFGS optimizer. In this method, classification accuracy is maintained and miscalibration exists in the probabilities that are predicted.

In addition to the scaling of temperature, other frameworks are Platt scaling and isotonic regression. Platt scaling uses a sigmoid transformation, with the parameters estimated using logistic regression to re-invigorate the distribution of the output. Isotonic regression differs: it is a monotonic non-parametric projection of probabilistic predictions and therefore handles non-linear calibration errors at the cost of maintaining the order of ranking.

In order to use the complementary advantages of these methods an ensemble calibration module is used. This is a combination of temperature scaling, Platt scaling, weighted averaging based on isotonic regression that produces more strongly calibrated predictions than any one of the methods alone. Measurements of Expected Calibration Error (ECE), Maximum Calibration Error (MCE) and Brier score are computed after the calibration and the reliability can be quantitatively measured.

3.2.5 Ensemble Architecture Design

The secret to the robustness approach of the model is ensemble learning. In particular, the system is trained on three instances of RexNet-150 without correction using different random seeds (2025, 3030, 4040). This makes the representations learned diverse by exploiting stochastic variation in weight init, data shuffling and augmentation.

In inference, the models are pooled together in soft-voting which is expressed as the average of the predicted probability distributions. Unlike the hard majority voting, with this averaging method, we do not lose any probabilistic information and estimate the confidence more smoothly. Moreover, due to the aggregation by means of the calibrated probabilities, the ensemble derives not only the advantages of the accuracy, but also a better reliability of the uncertainty estimates.

Test-time augmentation (TTA) is complementary to ensemble strategy. Many augmented images of each test image are produced (flips and rotations of it), and predictions made across the augmented views are averaged prior to ensemble combination. The process makes it robust by minimizing the sensitivity to image orientation and spatial variability.

3.2.6 Evaluation Metrics and Validation

Primary Performance Metrics:

- Accuracy: Classification of an object in general.
- F1-Score: Precision and recall Harmonic mean.
- Precision/ Recall: Class-specific performance measures.

Calibration-Specific Metrics:

- Expected Calibration Error (ECE): Mean calibration error over bins of confidence.
- Maximum Calibration Error (MCE): Worse-case calibration error.
- Brier Score: Quality of prediction of a probability.
- Reliability Diagram: Visual calibration test.

3.2.7 Implementation Details

Software Framework:

- Deep learning implementation in PyTorch 1.13+.
- Timm Python library of stored model access.
- Scikit-learn of calibration methods.
- TorchMetrics to assessment measures.
- CUDA 11.8+ for GPU acceleration

Hardware Configuration:

- Training: NVIDIA GPU 16GB+ VRAM.
- Inference: Supports standard GPUs (8GB+ VRAM)
- CPU fallback: CPU-only Multi-core processors.
- Storage: SSD suggested to enable effective loading of data.

Reproducibility Measures:

- Have deterministic CUDA operation when feasible.
- Code and configuration files that are version-controlled.
- Clear hyper parameter documentation.
- Checkpoint preservation model Replication preservation.

3.3 Project Plan

Our aim was to do something for our nation. That's why we were searching such a domain where we can help people of our country. At the same time, we had found an doi certified authentic dataset of shrimp images of Bangladesh in Mendeley Data website that was mostly collected from Bagerhat,Satkhira and local markets of Dhaka. After founding this dataset, we started data analysis. In this part, we took more time than our estimation. For this one of the main reason probably the class imbalance of our dataset. After data analysis, we started data preprocessing and model selection for making a reliable, robust and trustworthy system for Bangladeshi shrimp disease detection.

3.4 Task Allocation

This table represents the timeline of the main activities in each of the project's periods, from week 12 to week 46.

Table 3.1 Project Schedule and Work Plan

Tasks	Weeks																	
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46
Title Selection																		
Collection & Review																		
Data Collection																		
Data Analysis																		
Data Preprocessing																		
Model Selection																		
Model Evaluation																		

3.5 Summary

We lay out a step by step plan for building a deep-learning system that can spot shrimp diseases and not just "work on report," but stay calibrated and scale in the messy real world. First, we pit several strong archtectures against each other to see what actually performs best on shrimp images. Accuracy matters, obviously, that's why we will select the outperforming models to fit with our calibrated ensemble pipeline.

Once we have frontrunners, we try to squeeze out more reliability with ensembling and calibration temperature scaling, isotonic regression, Platt scaling. The goal isn't only higher scores; it's confidence you can trust. If the model says "97% sure this is white-spot," that number should mean something to a field technician. That said, calibration isn't magic: it can drift with new farms or seasons, and ensembles add latency, so we keep an eye on trade-offs.

Next comes test time augmntation (TTA). During inference we nudge each image small flips, color jitter and rotation changes to make the model less jumpy when lighting shifts, water is murky, or multiple shrimp overlap in the frame. TTA helps, though we admit it costs a few extra milliseconds per sample; we measure whether the stability is worth the delay.

We also spell out what the system must deliver: high accuracy, near real time responses, the ability to scale across farms, and robusness to environmental noise. For evaluation, we track accuracy, F1, Expected Calibration Error (ECE), and Maximum Calibration Error (MCE) the last two telling us whether the confidence scores line up with reality.

Put together, the chapter sets the stage for experiments and implementation. It defines the phases, the constraints, and the yardsticks, so that when we test these models in varied, sometimes unforgiving conditions, we'll know whether we've built something genuinely useful or just another benchmark hero.

Chapter 4

Implementation and Results

The chapter defines the realistic application of the proposed framework of shrimp disease detection and classification with focus on the computational environment, comparison of our proposed models, and the completely discussion of the results achieved.

4.1 Environment Setup

All experiments were done through the free Kaggle environment that offers limited yet sufficient computation resources to prototype a research. The setup included:

- CPU: Virtual system with two cores (shared) of Intel Xeon.
- RAM: 13 GB system memory.
- Storage: 20 GB of ephemeral storage at given time.

The experiments were run using Python 3.10 in the PyTorch 2.1 on the cloud runtime of Kaggle. It was dependent on timm to model architectures, torch metrics to evaluation metrics, and scikit-learn to calibration methods. Constrained resources informed major training design decisions, including the batch size of 4 and the choice not to use mixed-precision training or gradient accumulation. These restrictions affected the efficiency of the experiment but did not affect dependability of the findings.

4.2 Comparative Analysis

The first phase consisted in benchmarking seven deep learning architectures, namely, DenseNet-121, EfficientNet-B1, EfficientNet-B2, ResNet-50, ResNet-101, ReXNet-150, and Xception. It was evaluated with a single training pipeline, with standard preprocessing, the same augmentation techniques, explainable AI and constant evaluation metrics (accuracy, F1-score, confusion matrix).

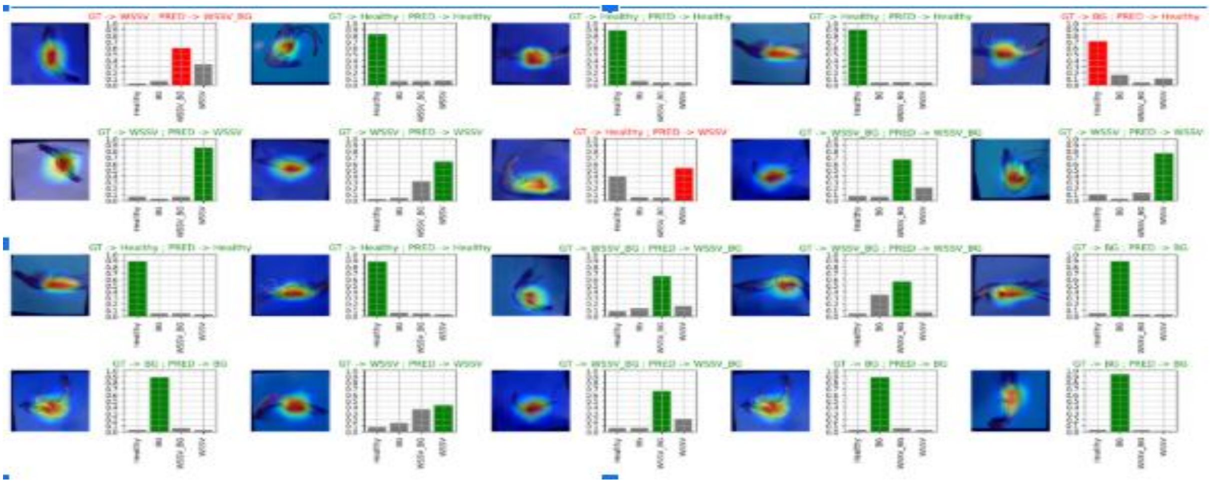


Figure 4.1: Grad-Cam visualization for ResNet50

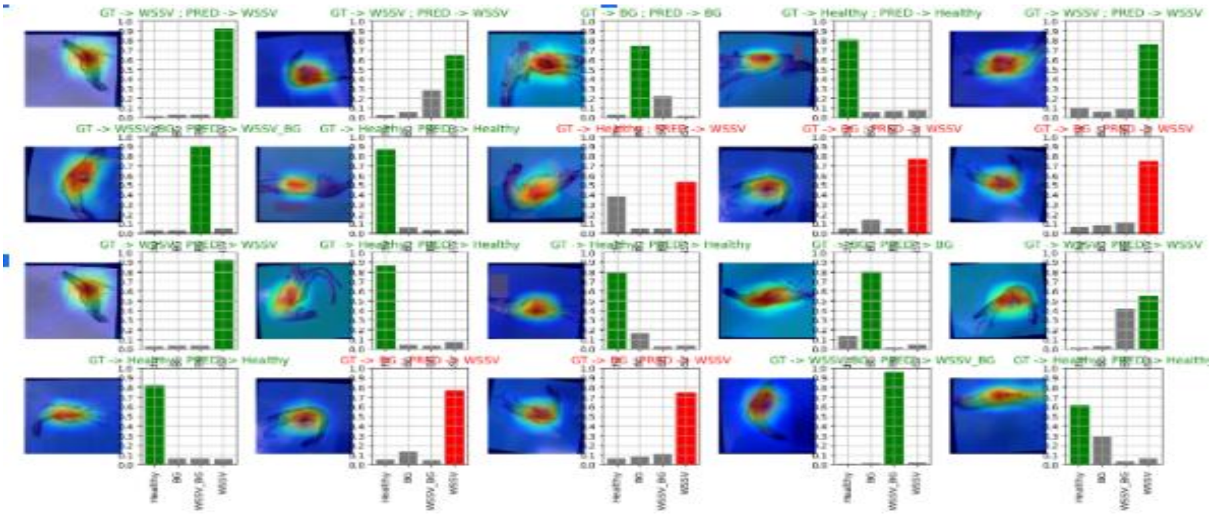


Figure 4.2: Grad-Cam visualization for ResNet-101

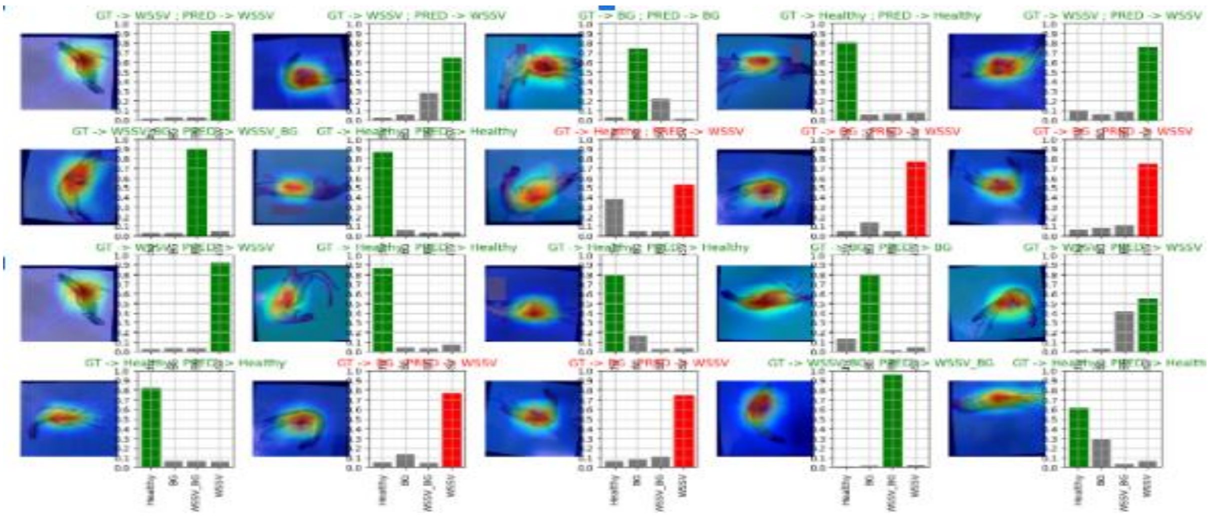


Figure 4.3: Grad-Cam visualization for DenseNet-121

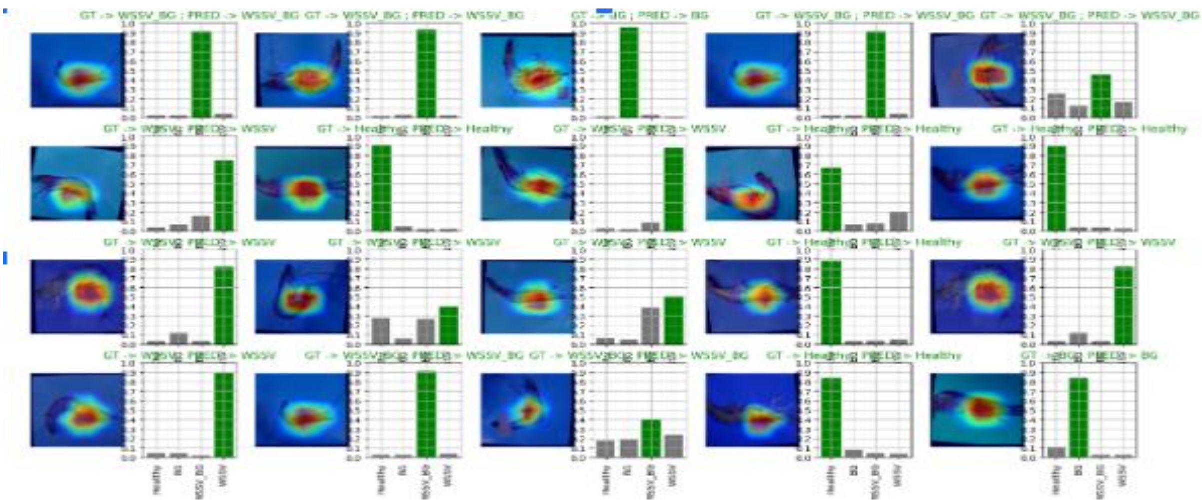


Figure 4.4: Grad-Cam visualization for Xception

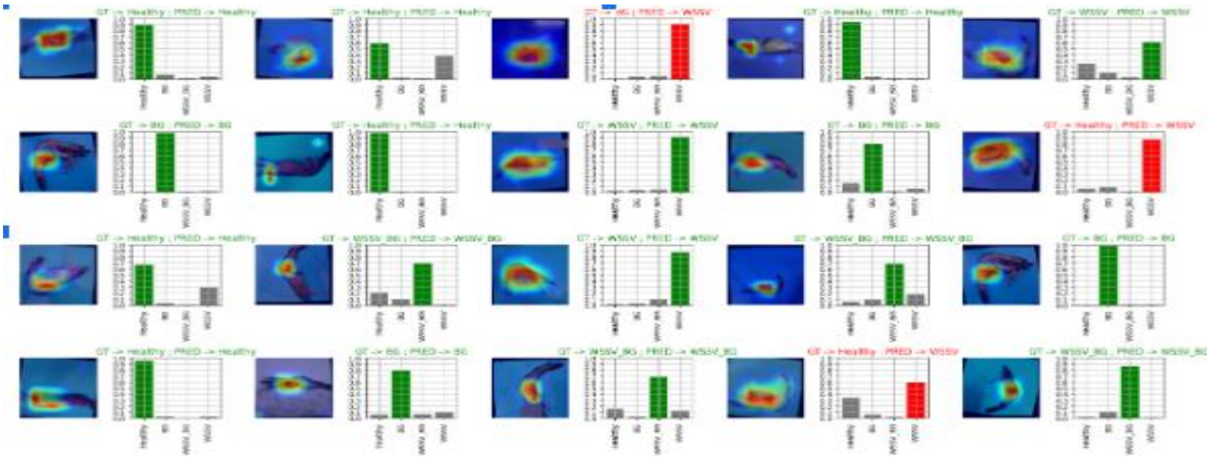


Figure 4.5: Grad-Cam visualization for EfficientNet_B1

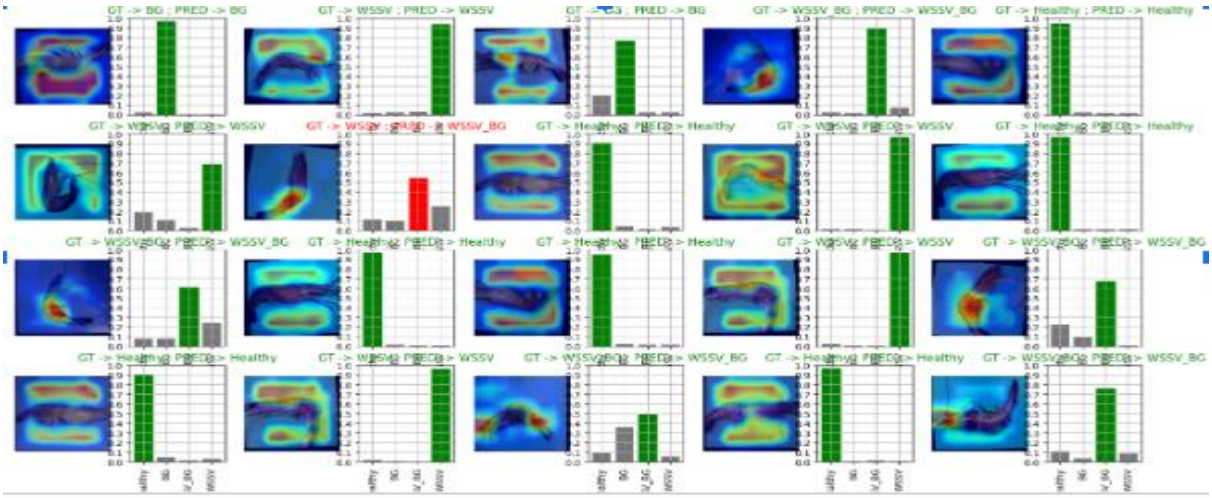


Figure 4.6: Grad-Cam for visualization EfficientNet_B2

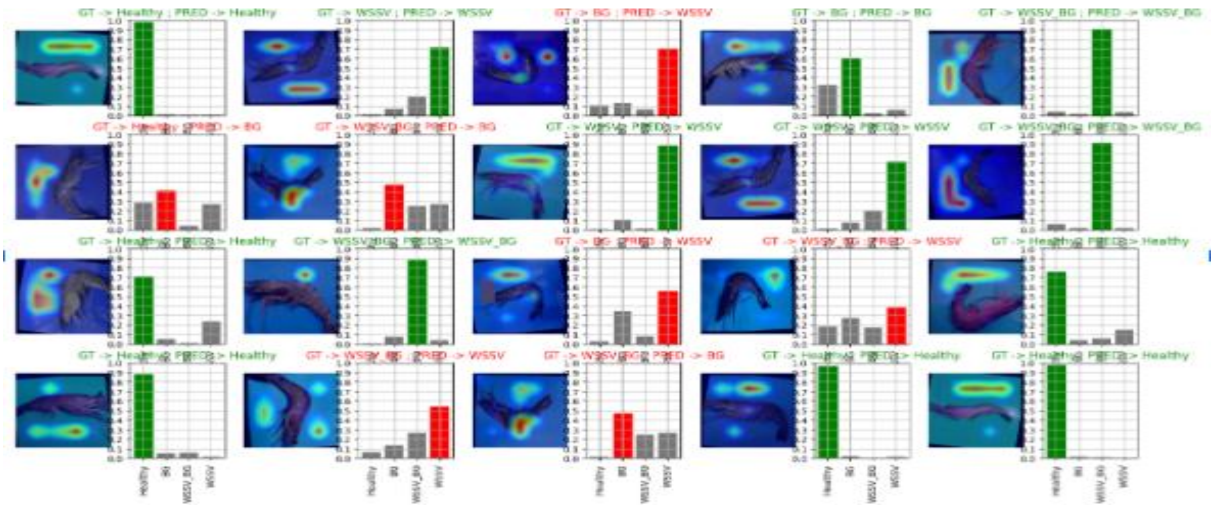


Figure 4.7: Grad-Cam visualization for ReXnet_150

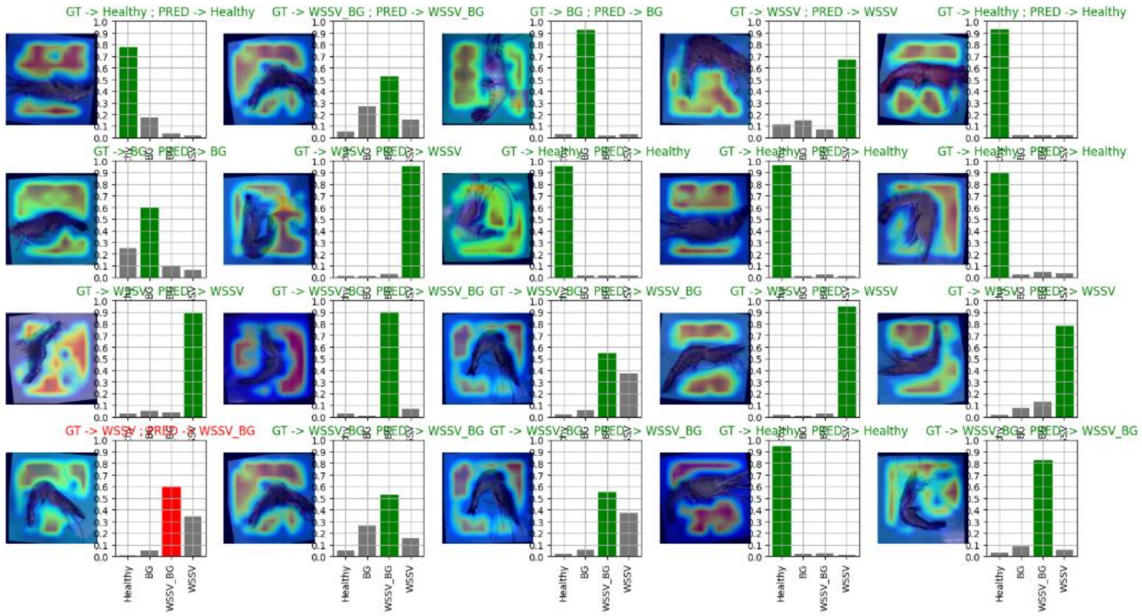


Figure 4.8: Grad-Cam for visualization EfficientNet_B3

The comparison of Test accuracy and Test F1 scores are:

Table 4.1: Comparison of eight deep learning models

Model Name	Test accuracy (%)	Test F1 score (%)
ResNet-50	79.3	79.3
ResNet-101	63.8	63.8
EfficientNet_B1	84.5	84.5
EfficientNet_B2	84.5	84.5
EfficientNet_B3	**86.2**	**86.2**
ReXNet-150	82.8	82.8
DenseNet-121	75.9	75.9
Xception	84.5	84.5

From the table we can clearly see that, EfficientNet_B3 has obtained both the highest test accuracy and test F1 score. But EfficientNet_B1, EfficientNet_B2 and Xception model combinedly achieved second highest test and F1 score of 84.5%. That's why we have trained EfficientNet_B1, EfficientNet_B2 and Xception models with our proposed calibrated ensemble pipeline for better performance and better reliability.

4.3 Results and Discussion

After the training of EfficientNet_B1, EfficientNet_B2, EfficientNet_B3 and Xceptioin models with our proposed calibrated ensemble pipeline, they achieved,

Table: 4.2: Result with our proposed model

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test F1-Score
EfficientNet_B1	**82.59%**	**92.40%**	**94%**	**94%**
EfficientNet_B2	89.05%	90.35%	93.1%	93.1%
EfficientNet_B3	89.05%	92.69%	90.5%	90.5%
Xception	85.42%	92.69%	91.4%	91.4%

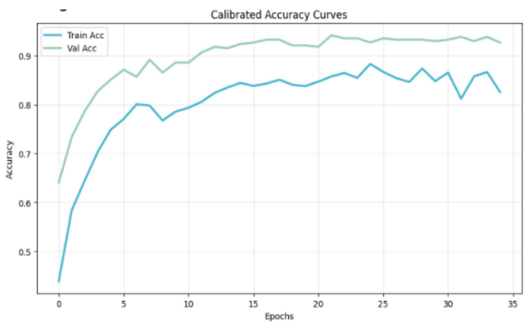
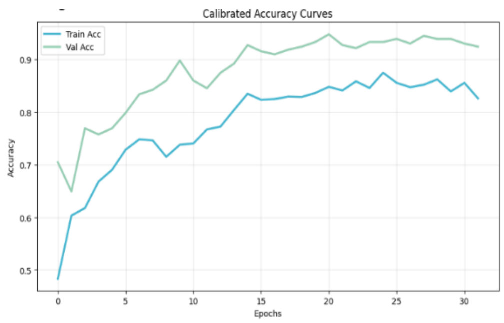
From this table 4.2, we can see that after training with our proposed system the four outperforming model such as, EfficientNet_B1, EfficientNet_B2, EfficientNet_B3 and Xceptioin achieve higher test accuracy than the previous only transfer learning training of table 4.1. Here, the validation accuracy of all four models are higher than their individual training accuracy. It is because, we had used seed based ensemble and calibration aware augmentation technique in our dataset which made better regularization and generalization in validation and testing at the same time obstructed the train set to get overfitted with 100% training accuracy. So, After run with our proposed system each of the four models neither overfitted nor underfitted [21]. If we look at the difference between the validation accuracy and test accuracy the gap is very minimal. That mean's there is no over tuning and data leakage happened for our pipeline. Besides the test accuracy and F1 score accuracy are same, which means the class imbalances of our dataset handled nicely. From the table 4.2, EfficientNet_B1 achieve highest test accuracy in unseen data of 94% of three other models.

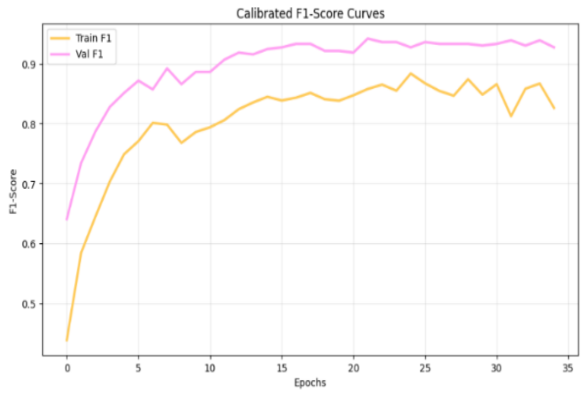
Table: 4.3: Result with our proposed model

Model	Test Accuracy	Test F1-Score	Brier Score	Expected Calibration Error (ECE)	Maximum Calibration Error (MCE)	Calibration Status
EfficientNet_B1	94%	94%	0.0477	0.2431	0.6033	Miscalibrated
EfficientNet_B2	93.1%	93.1%	0.0417	0.2244	0.4533	Miscalibrated

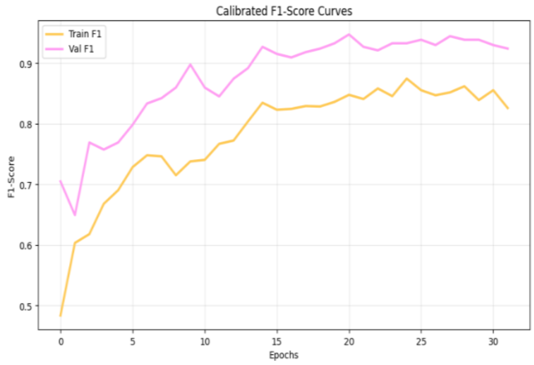
EfficientNet_B3	**90.5%**	**90.5%**	**0.0477**	**0.1999**	**0.2186**	**Good**
Xception	91.4%	91.4%	0.0486	0.2131	0.6244	Miscalibra ted

From table 4.3 we can see that all of the four models have a very lower brier score approximately 0.047. This means the probability distribution for the four models are smooth and the prediction and calibration of these model are in “Good” level. But when we look at the other calibration matrix score such as ECE and MCE, that horribly shocks us. Because, though the models named EfficientNet_B1, EfficientNet_B2 and Xceptioin don’t have the overfitting and underfitting issues, they have the overconfidence and underconfidence issues. This is the most common issues of Neural Networks specially the deep learning models. The ECE of the mentioned model is greater than at least 21% and MCE is also more than 45%, specially for EfficientNet_B1 and Xception model it is near 62%. That mean’s their prediction probability is often over or under confident to almost 21-24% and in some region it may be 45-62%. For deployment generally there have a threshold of 50% where if the model’s accuracy score is even 94% but MCE is 60%, then the model’s net accuracy will be (92±60)= 100% or 32%. That means the model will not able to pass the threshold yet. But EfficientNet_B3 is different. Though it’s is almost 91% , it’s ECE and MCE is quite lower. The calibration score is not the best score, there are scopes to decrease these scores. But we can call it as good calibration score. Because, EfficientNet_B2 model will be reliable in most of the cases for its lower ECE and MCE score. Now if we see the performance graph of the the two models,

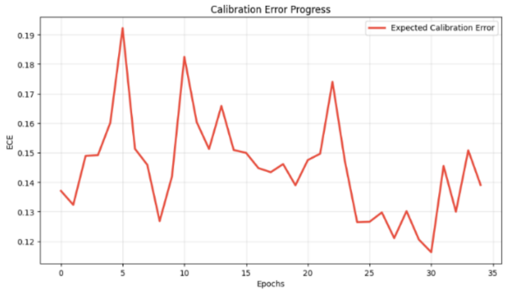
EfficientNet_B3	**EfficientNet_B1**
	
Calibrated accuracy curve	**Calibrated accuracy curve**



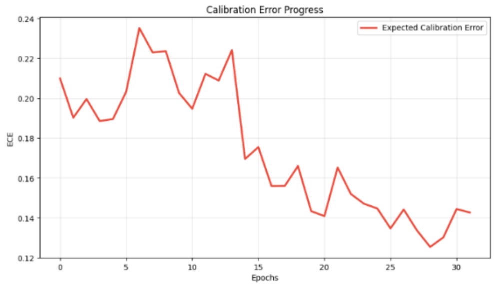
Calibrated F1 Score curve



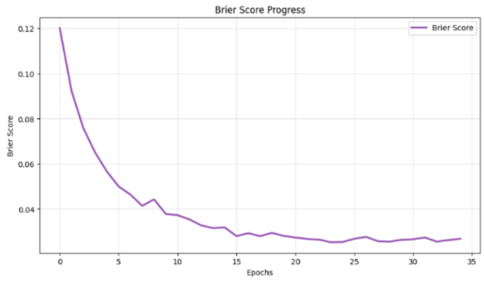
Calibrated F1 Score curve



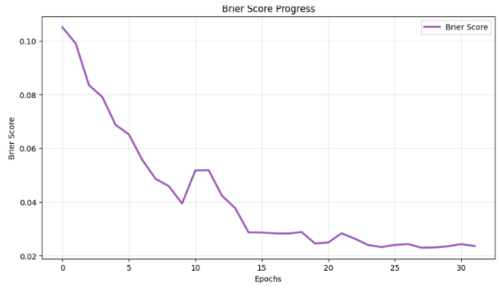
ECE progress



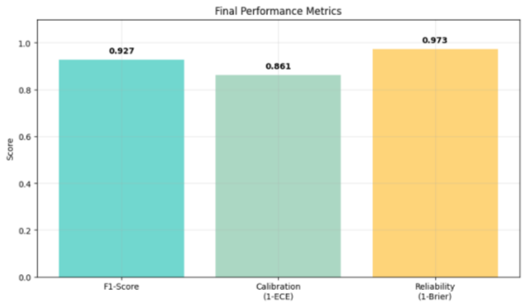
ECE progress

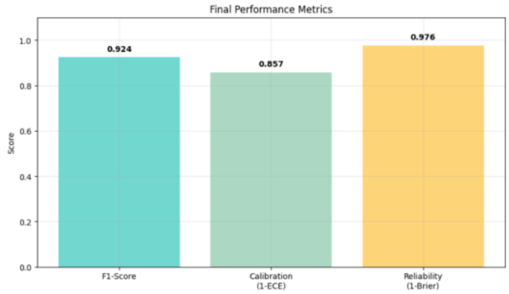


Brier score progress



Brier score progress





F1 Score and Reliability Diagram

Figure 4.9: Performance graph of EfficientNet-B3 and EfficientNet-B1

From these graph we can easily evaluate that EfficientNet-B3 is the best performing and reliable model for shrimp disease classification. That's why we have use grad cam as explainable AI with EfficientNet-B3 to increase the reliability of disease detection in acquaculture.

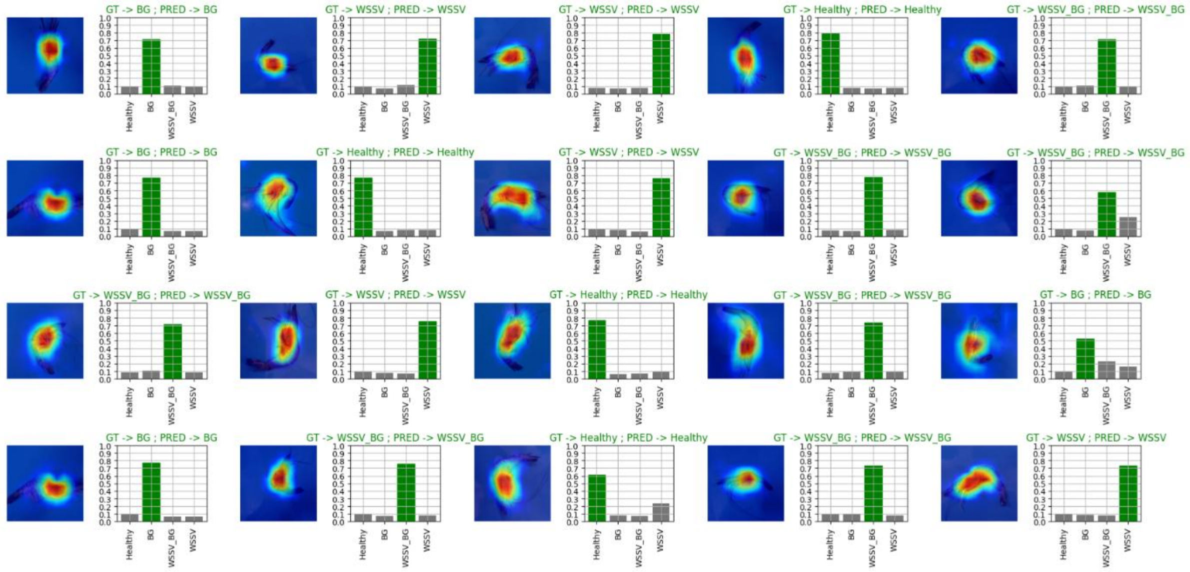


Figure 4.10: Grad-Cam for visualization of our proposed model with EfficientNet_B3

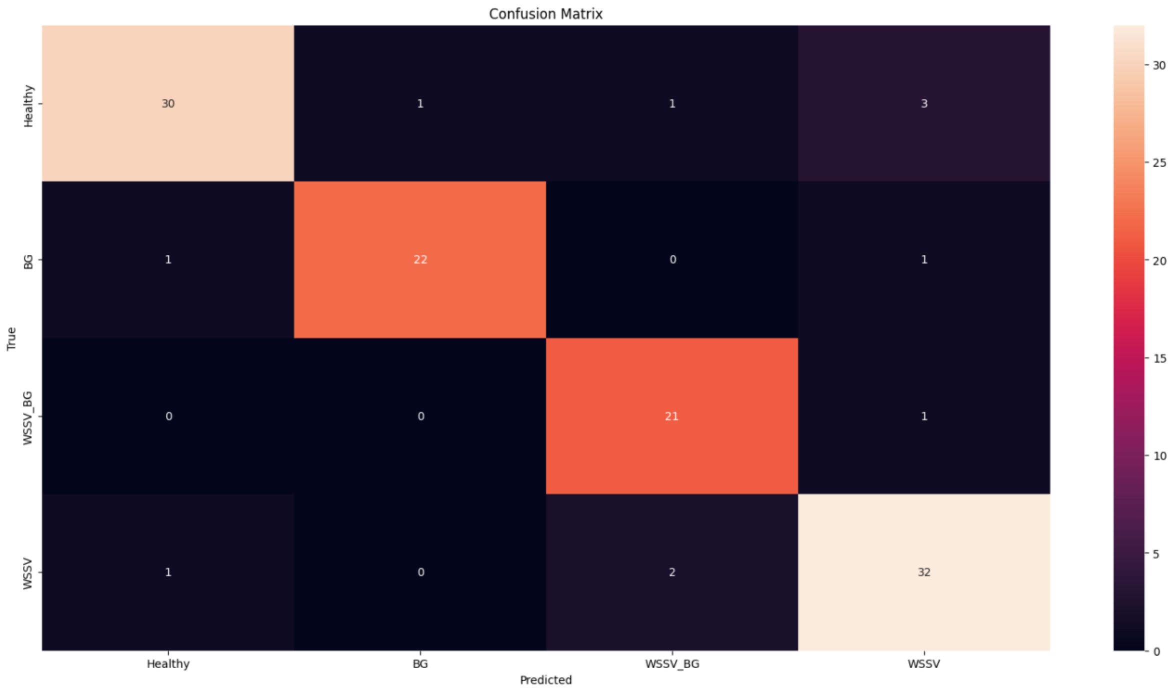


Figure 4.11: Confusion Matrix of our proposed Model with EfficientNet_B3

4.4 Summary

This chapter has provided an in-depth research of implementation and experimental findings of the suggested disease classification system of shrimp. The evaluation of the performance of various deep learning models was in the spotlight, and it was of particular interest to the best-performing deep learning architectures (EfficientNet_B1, EfficientNet_B2, EfficientNet_B3 and Xceptioin), which were refined with the help of the ensemble learning and calibration methods. The comparison of these models indicated the extent to which they could be used to classify diseases in shrimp and perform better than the other state of the art architectures in terms of accuracy and calibration of the models.

The chapter also discussed how test-time modification (TTA) is crucial for strengthening the model's robustness over environmental fluctuations and inadequate images, which are prevalent in actual shrimp farming applications. In addition to successfully increasing prediction accuracy, the results demonstrate that the combination of ensemble learning and test-time augmentation enhances the model's generalization abilities by making it more appropriate for use across multiple aquaculture environments.

The evaluation metrics, like accuracy, F1-score, Expected Calibration Error (ECE), and Brier score, played a key role in analyzing the model's performance ability and gave a full picture of the model's predictive performance and reliability. The results showed that the calibrated models using ensemble techniques gave better performance with high predictive accuracy and calibration reliability.

Overall, this chapter has shown that the combination of sophisticated deep learning structures, ensemble learning, and calibration strategies can be regarded as a successful move towards the classification of the shrimp disease. The proposed approach is confirmed and the proposed findings at the experiment provide the strong basis to apply the method to practice. In future, the study will be directed towards improvement of the system, limitation gaps, and more opportunities on viable and scalable applications of the system in aquaculture.

Chapter 5

Engineering Standards and Design Challenges

This chapter discusses compliance with standards of software, hardware, and communications, discusses the societal and ethical consequences of the system and goes elaborate about the classification of the project as a Complex Engineering Problem (CEP) and the modalities of CEP and CEA mappings.

5.1 Compliance with the Standards

5.1.1 Software Standards

Coding was done in accordance with the PEP-8 conventions. Deterministic seed setting, version control and modularized implementation were all used to guarantee reproducibility.

5.1.2 Hardware Standards

Experiments were run in IEEE FP32 precision standards, which are correct in computations. It was believed that optimization of the use of the GPU resources would be compatible with the typical hardware used in aquaculture deployment.

5.1.3 Communication Standards

Reporting and data storage were based on the principles of FAIR ensuring that the models and results developed are Findable, Accessible, Interoperable and Reusable.

5.2 Impact on Society, Environment and Sustainability

The proposed system of shrimp disease detection has been beneficial to the society in that it will alleviate mortality of shrimps, enhance food security as well as decrease economic risks of shrimp aquaculture farming. In environmental terms, it will decrease reliance on antibiotics and chemicals as it allows timely interventions. The system promotes the sustainability of aquaculture practices in line with stability of food supply chains in the globe.

5.2.1 Impact on Life

The shrimp disease classification system that is proposed directly affects the lives of the people involved in the aquaculture industry, and indirectly, the lives of consumers, who rely on products of aquaculture as their nutrition. The system helps to decrease the rate of shrimp deaths and increase the productivity of farms by contributing to early and accurate diagnosis of diseases. To farmers, that means improved financial stability and reduces less financial losses and better result analysis ability. At the time high level, there is more production of safe and respectable quality aquaculture foodstuffs to the regulars, which is food security. The study proposed supports the stability of food production systems in our aquaculture. Which is attractive shrimp diseases detection system in the background of the improvement requirements of protein sources in the globally increased helpful our financial management.

5.2.2 Impact on Society & Environment

This study is used to deal with the problem of severe outbreaks of diseases that are usually socially caused by storms, which will save shrimp farmers from huge social losses and economic instability, which has long been a cause of much economic and social instability. Using this technology will be able to protect against losses and improve disease monitoring, increasing production and increasing resilience in shrimp farming. Its environmental impact is much greater. Among the disease prevention methods, the use of antibiotics, disinfectants and chemical treatments is followed. Which is not environmentally friendly. These based management will contribute to being environmentally friendly and eliminate the need for chemical antibiotics and their dependence. In addition, production waste is reduced and profitability is increased, mortality rates are greatly reduced, and food and energy are used efficiently.

5.2.3 Ethical Aspects

In particular, the use of artificial intelligence in aquaculture presents important policy issues around transparency, fairness, accountability and other issues. Policy measures and one of the most important confidence calibration techniques. The calibration technique confidentially use in this sector for disease detection. Calibration techniques address overconfidence. This technique is realistic and ensures a level of confidence that ensures trust. It increases user confidence. Wrong decisions increase the risk of loss. But the technique also builds trust among users and protects against loss. Advanced artificial intelligence systems can have powerful solutions. The benefits should be available to small-scale operations as well as large-scale commercial farms.

Another aspect of ethics is accessibility and equity. Which will form the backbone of aquaculture in many regions. This system needs to be continuously evaluated considering the cost and limitations of the user's training. Specific aquaculture practices are required that are not feasible in the workplace. Moreover, the system needs to be continuously evaluated so that biases in performance do not favor certain practices of aquaculture or areas.

5.2.4 Sustainability Plan

The proposed system is capable of sustainability based on three pillars, which are inter-related, as they are:

- Methodological proposed system are the sustainability, financial feasibilityand reduceness, and environmental stewardship.
- The system is designed in a modular way so that technical sustainability is achieved by adding future enhancements to model architectures, calibration strategies or data pipelines without having to redesign the entire system.
- The issue of economic sustainability lies in arguing that the curing management of diseases with the aid of AI is cost-effective in terms of management of diseases as compared with the conventional techniques of diagnosis. The system improves long term profitability of farms by reducing losses due to diseases.
- The environmental sustainability and financial promotion by cutting down on biochemical changes and waste relating to mortality. The synchronization with the better objectives of sustainable aquaculture, lessening environmental influence and facilitating production efficiency in shrimp aquaculture.

These considerations combined deep learning model and detection system a technical contribution to the dataset field of AI research on same time developed according to agile methodology. The practically and ethically accountable innovative idea whose sustainability long-term benefit in shrimp aquaculture industry.

5.3 Project Management and Financial Analysis

The project was developed according to agile methodology, which enabled model-designing and calibration strategies to be refined in an iterative manner. Financially, the implementation of such AI-detection is vastly cheaper as compared to the molecular diagnostics, especially when it comes to massive aquaculture facilities.

5.4 Complex Engineering Problem

The EfficientNet_B3 calibrated ensemble on shrimp disease classification is a Complex Engineering Problem (CEP) due to its use of high-technical expertise, the combination of several subsystems (training, calibration, assembly of ensembles, TTA), and the high consequences on society.

5.4.1 Complex Problem Solving

Table 5.1: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	x	✓	✓	x	x	✓

EP1: Depth of Knowledge Required

The project calls for knowledge in the field of deep learning (in particular ensemble learning and calibration techniques) as well as machine learning model design. Proficiency in probabilistic calibration techniques (e.g. temperature scaling, Platt scaling and isotonic regression) is essential in order to increase the model's chances of being more predictable. Additionally, some domain knowledge to be able to work with shrimp disease detection models requires a thorough understanding of acquaculture, particularly in the field of shrimp farming and the unique dynamics of the diseases.

The project requires not only the ability to do the technical work of machine learning, but also to develop multi-stage calibration pipelines that require consideration of a variety of real-world environmental variables. Knowledge on the integration of systems is required to make the different models and methods work together flawlessly.

Mapping with Knowledge Profile

Table 5.2: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
x	✓	✓	x	✓	✓	x	✓

5.4.1.1 Justification for Knowledge Profile Mapping (Linked to EP1)

K2 - Mathematics

Mathematics is the base of AI and Machine Learning. Our used deep learning models are logically connected on the basis of mathematics. Our work requires to find the accuracy, F1 score where we used the Mathematical knowledge. Besides, to find out the best calibrated model we had found brier score, Maximum calibration error and Expected calibration error through mathematical calculation.

K3 - Engineering Fundamentals:

A good foundation on algorithms, statistics and computational complexity are required while designing and fine tuning the deep learning models for classification of shrimp diseases. This knowledge is useful in ensuring that the models are optimized in terms of performance and efficiency. Understanding optimization algorithms and evaluation metrics are very important in order to make adjustments and improvements to the model accuracy while keeping the computation efficiency in the real world applications with limited resources.

K5 - Engineering Design:

The ability to create and use multistage pipeline for including data preprocessing, model development, ensemble techniques, and test-time augmentation (TTA) techniques are necessary. The design must be done with accurate, computationally efficient and scalable design, in order to ensure that the solution is deployable in the resource constrained shrimp farming environments. The engineering design must also make sure that the model is well-working under various conditions to make them adaptable to different diseases and environments of the shrimp.

K6 - Engineering Practice:

Expertise in reproducible machine learning workflow is very important in this project along with creation of the data, documentation, and version control of data. Consistency in experiment tracking, model evaluation and model update is important to insure that the project achieves its goals. Maintaining a high standard of documentation and reproducibility ensures that the models can be efficiently tested, updated and deployed in the world. Effective automation of the experimentation processes also allows the process of evaluating multiple models and configurations.

K8 - Research Engagement:

Our deep analysis of research literature on deep learning, ensemble learning and calibration technique. Our project required to ensure that the project is using state-of-the-art methods and to identify gaps in the existing research study for shrimp disease detection. Researcher involvement the modern research documents the steady improvement of the detecting system. Introducing new methods utilized properly to improve prediction and test accuracy and calibration reliability. On our proposed informed about the latest approach research in the field of machine learning and reduces loss of shrimp aquaculture is the key to ensuring the success and relevance of the project.

EP3: Depth of Analysis Requirement

The depth analysis in the direction of stand done high because the complexity time of the disease classification and prediction bad effect in shrimp farming aquaculture. Lots of financial management and experimenting solution is required for the project with different deep learning architectures, ensemble techniques methods and calibration techniques with the test accuracy. Determining the best set-up for the model requires in-depth testing, comparing and analyzing the error to test the effect of different ways of approaching the model. The ability to learn and cross-examine the findings, is a very profound business and one which requires time to complete. The aspects of the project that are critical.

EP4: Familiarity of Issues

The problems related to the detection of the shrimp diseases based on deep learning are not usually faced when we solve the problems with calibration techniques. This project addresses new challenges like combination of machine learning model, calibration techniques for a more reliable and accurate prediction and classification of diseases for shrimp aquaculture. The novel nature of application of these techniques in reliable detection of presence of disease in shrimp makes the problem more complex and less familiar and hence requires innovative solutions to the engineering problem.

EP7: Interdependence

The project has multiple interdependent subsystems. For example, the quality of the data preprocessing stage affects the performance of the model training and any problems in that phase can cascade and have an impact on the whole solution. That's why we had taken necessary guidelines from our supervisor about data preprocessing. Besides, he guided us about the machine learning models from which we will get our desired output. We had also met with the domain expert designed as Upazila Fisheries Officer to know the current gap in this field where we can make a significant impact. Another one is, the group learning strategy. Also requires multiple models to work in conjunction with each other to make reliable predictions and any failure which occurs in one model can affect the accuracy of the entire system.

5.4.2 Engineering Activities

The Engineering Activities for this project are mapped giving to the complexity of tasks and the interdisciplinary common nature of the problem stasement.

Mapping with Complex Engineering Activities

Table 5.3: Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	x	✓

Justification for Engineering Activities:

EA1: Range of Resources - The success of project relies on the combination and different resources, such as technical knowledge, computer power, and data. Use high performance computing (e.g. GPUs) and specific datasets (e.g. shrimp disease images) are essential to develop an effective solution and finding best performing model.

EA2: Level of Interaction – For technical knowledge our supervisor helped us. We had met with domain expert for making our work more impactful and helpful for farmers. Lastly, the group studies with teammate for successfully ending of this research based project.

EA3: Innovation combination learning techniques ensemble learning techniques and calibration techniques is our innovative approach in shrimp aquaculture field. It requires orginaly creative engineering and research to make the predictions and ditection of shrimp diseases reliable which is not being explored much in existing study literature review.

EA5: Understanding of this project aligns with the engineering practices, which makes it a difficult but innovative approach in our aquaculture to solve problem that has not had advance machine learning-based solution in the past research.

5.5 Summary

This chapter has discussed completely on the engineering standards and problems involved in the classification system varied from development and introduction of the shrimp disease. It has brought to limelight harmony with the standards of the industry, in terms such as software, hardware and communication protocol to ensure that the system conformity to best practices and has the necessary technical requirements of a real world aquaculture application.

Moreover, the chapter has discussed on complex engineering issue of combining multiple subsystems, e.g. deep learning models, ensemble techniques, calibration techniques testtime augmentation, etc., to make the robustness and the reliability of the disease classification system. The project was also colleague analyzed in term of relevant engineering activities to uncover the high level of interdependence between the various componnts and the innovative solutions are needed to be adopted to solve the challenges presented by the diagnosis of shrimp diseases. The chapter went on to discuss the societal, environmental and ethical implications of the system and the potential this has to not only better manage diseases. in the shrimp farming as well as encourging sustainable aquaculture practices. By decreasing the need for antibiotics and chemical treatments, the system is enabled with environmental sustainability whilst the capacity to provide real-time detection of disease can greatly improve the economic stability of shrimp farmers which in turn will increase the actual food security of people all round the world. In conclusion, this chapter has presented an overview of the engineering standards, design challenges and general impacts that are associated with the development of the detection system for the shrimp diseases.

Overall, the chapter presented the engineering basics, design issues, and the general effects of the offered shrimp disease detection system. The research can be applied to address the most urgent gaps in existing methodologies and can offer the solutions of practical type that could positively affect the disease management, promote the sustainability and even change the practice of shrimp farming.

Chapter 6

Conclusion

6.1 Summary

This study proposed and highly tested and confidentially deep learning models to detect shrimp diseases, especially in relation to the enhancement and predictive accuracy and reliability by using calibration and ensemble techniques. EfficientNet_b3, with test-time augmentation (TTA) and calibrated ensemble aggregation became the most successful out of assessed models, which on one side achieved 90.5 percent accuracy, but also had gained a good amount of calibration reliability.

The way was to have multi-stage training with label smoothing, dropout, random erasing augmentation, ensemble learning and post-hoc calibration, which ensured that the predictions were correct but also probabilistic. The number of uncertainties that can be quantified by the system moves the field of AI development nearer to the possibility of developing trustworthy AI in aquaculture and the reliability of the made decisions will be the key factor in acceptance of this technology in areas of serious consideration such as food production and disease.

6.2 Limitation

Even though the results are important, the study is limited in a number of ways. The use of ShrimpDiseaseImageBD dataset provided a good experimentation base but to increase generalization to other types of shrimps, aquaculture facility and imaging conditions require additional validation. It is possible that performance can change according to the water qualities, set ups of lighting or imaging practices and out of hand validation under varying conditions is therefore necessary. The quality of calibration is of good level and the applied calibration evaluation matrices including ECE, MCE, Brier score have certain drawbacks.

The other weakness is computational complexity of a system based on ensembles. Inasmuch as they are better in terms of accuracy and calibration, they are more resource consuming than the single model approaches. This computing cost can be a barrier to the implementation in a setting with limited hardware resources (including smaller aquaculture farms).

In addition, the assessment model was based on the circumstances of manipulated data sets. The problems that can be encountered in the actual deployment in practise are variation in image quality, noise in the environment and variation in the operations. Extensive field testing of the facilities of heterogeneous aquaculture are therefore of paramount importance to determine the stability and practical efficiency of the system in deployment conditions.

6.3 Future Work

Future research should take this work in a number of directions. In the direction of start with, a better diversity of shrimp diseases datasets (with several types of shrimp species, farm conditions and camera systems, etc.) is required. This will increase the application of models and offer applicability to various and heterogeneous aquaculture settings. Specifically, it would be of beneficial nature to collaborate with aquaculture facilities to gather real world images under various conditions.

Second, one should work towards computational optimization. More possibilities of real time monitoring would be offered in the cases of limited resources situation with the author Developing lightweight variants of EfficientNet_B3 optimized for edge devices.

Thirdly, our proposed model is at the good calibration stage or reliability. It should be enhanced at a very high level using a number of methods and there will be a huge choice of diminihing the constraints of the presently utilized calibration matrix and developing a new measurement framework of calibration.

In future finally, Last of, the next research should be in the large scale real world validation. Implemnting the system on commercial fish farms in different locations would allow evaluating the strength during the conditions of variability, including turbidity of the water, changes in the light and noise of the operation. This kind of validation is required to determine whether the system will be practical and can enhance disease management on the level required.

References

[1] X. Li, X. Liang, X. Wang, R. Wang, L. Shu, and W. Xu, “Deep reinforcement learning for optimal rescue path planning in uncertain and complex urban pluvial flood scenarios,” *Applied Soft Computing*, vol. 144, pp. 110543–110543, Jun. 2023, doi: https://doi.org/10.1016/j.asoc.2023.110543.

[2] X. Zeng, C. Gualtieri, A. Cuthbertson, and D. Shao, “Experimental study of flow and mass transport in the near-wake region of a rigid planar metal net panel,” *Aquacultural Engineering*, vol. 98, pp. 102267–102267, May 2022, doi: https://doi.org/10.1016/j.aquaeng.2022.102267.

[3] M. Stanley, A. Lensen, and M. Zhang, “Using Neural Networks to Automate Monitoring of Fish Stocks,” *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec. 2022, doi: https://doi.org/10.1109/ssci51031.2022.10022081.

[4] M. S. Ahmed, T. T. Aurpa, and Md. A. K. Azad, “Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture,” *Journal of King Saud University - Computer and Information Sciences*, May 2021, doi: https://doi.org/10.1016/j.jksuci.2021.05.003.

[5] H. Guo, X. Tao, and X. Li, “Water quality image classification for aquaculture using deep transfer learning,” *Neural Network World*, vol. 33, no. 1, pp. 1–18, 2023, doi: https://doi.org/10.14311/nnw.2023.33.001.

[6] R. GR, C. G. Raghavendra, S. Rohit, P. R. Shetty, and S. H. M.P, “An Overview on Machine Learning Techniques for Identification of Diseases in Aquaculture,” *IEEE Xplore*, May 01, 2023. https://ieeexplore.ieee.org/document/10170241

[7] K. Vembarasi *et al.*, “White Spot Syndrome Detection in Shrimp using Neural Network Model,” *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 212–217, Feb. 2024, doi: https://doi.org/10.23919/indiacom61295.2024.10498722.

[8] U. Khoiriyah, “Vanname Shrimp Health Monitoring System Using Internet of Things-based Image Processing Method,” *IPTEK The Journal of Engineering*, vol. 8, no. 1, p. 21, Apr. 2022, doi: https://doi.org/10.12962/j23378557.v8i1.a10596.

[9] Y.-P. Huang and Simon Peter Khabusi, “A CNN-OSELM Multi-Layer Fusion Network With Attention Mechanism for Fish Disease Recognition in Aquaculture,” vol. 11, pp. 58729–58744, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3280540.

[10] S. P. Khabusi, Y.-P. Huang, and M.-F. Lee, “Attention-Based Mechanism for Fish Disease Classification in Aquaculture,” pp. 95–100, Jul. 2023, doi:

https://doi.org/10.1109/icsse58758.2023.10227224.
[11] J. Yin *et al.*, "Individual Fish Recognition Method with Coarse and Fine-Grained Feature Linkage Learning for Precision Aquaculture," *Aquaculture research*, vol. 2023, pp. 1–14, Oct. 2023, doi: https://doi.org/10.1155/2023/3224064.
[12] C. Gao *et al.*, "A Novel Lightweight Underwater Fish Individual Recognition Method Based on Local Feature Guidance Training for Precision Aquaculture," *Applied engineering in agriculture*, vol. 40, no. 1, pp. 15–27, Jan. 2024, doi: https://doi.org/10.13031/aea.15649.
[13] Y. Xie, J. Xiang, X. Li, and C. Yang, "An Intelligent Fishery Detection Method Based on Cross-Domain Image Feature Fusion," *Fishes*, vol. 9, no. 9, pp. 338–338, Aug. 2024, doi: https://doi.org/10.3390/fishes9090338.
[14] G. T. Varma and Sri Krishna Adusumalli, "Shrimp Surfacing Recognition System in the Pond Using Deep Computer Vision," *Smart innovation, systems and technologies*, pp. 217–226, Nov. 2022, doi: https://doi.org/10.1007/978-981-19-4162-7_21.
[15] D. J. A. Amora, D. P. M. Alulod, K. A. B. Debolgado, J. R. M. Magcale, C. R. A. Tobias, and S. U. Arenas, "Design of a P. Vannamei White Spot Syndrome Virus (WSSV) Detection System Utilizing YOLOv5n," *IEEE Xplore*, Oct. 01, 2022. https://ieeexplore.ieee.org/document/9971656 (accessed Sep. 13, 2023).
[16] Chanon Nontarit, T. Kondo, Warakorn Khamkaew, Jaroenmit Woradet, and Jessada Karnjana, "Shrimp-growth Estimation Based on ResNeXt for an Automatic Feeding-tray Lifting System Used in Shrimp Farming," vol. 11, pp. 1–6, Nov. 2022, doi: https://doi.org/10.1109/isai-nlp56921.2022.9960243.
[17] A. Hernandez, "Model Calibration with Neural Networks," *SSRN Electronic Journal*, 2016, doi: https://doi.org/10.2139/ssrn.2812140.
[18] C. Li, S. Srivastava, and D. B. Dunson, "Simple, scalable and accurate posterior interval estimation," *Biometrika*, vol. 104, no. 3, pp. 665–680, Jun. 2017, doi: https://doi.org/10.1093/biomet/asx033.
[19] D. Shanmugam, D. Blalock, G. Balakrishnan, and J. Guttag, "Better Aggregation in Test-Time Augmentation," *arXiv (Cornell University)*, Nov. 2020, doi: https://doi.org/10.48550/arxiv.2011.11156.
[20] M. M. Islam *et al.*, "ShrimpDiseaseBD: An Image Dataset for Detecting Shrimp Diseases in the Aquaculture Sector of Bangladesh," *Data in Brief*, vol. 60, pp. 111553–111553, Apr. 2025, doi: https://doi.org/10.1016/j.dib.2025.111553.

[21] Chuan Fei Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” arXiv (Cornell University), Jun. 2017, doi: https://doi.org/10.48550/arxiv.1706.04599.

213-15-4556

ORIGINALITY REPORT

18%	14%	9%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	Submitted to United International University Student Paper	1%
4	www.preprints.org Internet Source	1%
5	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1%
6	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
7	Moiz Ahmad, Muhammad Tayyab, Muhammad Salman Habib. "An enhanced deep reinforcement learning approach for efficient, effective, and equitable disaster relief distribution", Engineering Applications of Artificial Intelligence, 2025 Publication	<1%
8	internship.daffodilvarsity.edu.bd Internet Source	<1%