

A Multi-Stage Deep Learning Framework for Automated Detection and Localization of Shrimp Disease

by

Pritam Sarkar

22101373

Mohammod Tasneem Hasan

21301441

Arjun Saha

24341058

Md. Manzar Siraj

21301273

Mohammed Montasir Abdullah

21201600

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering

BRAC University

January 2026.

© 2026. BRAC University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Pritam Sarkar

22101373

Mohammod Tasneem Hasan

21301441

Arjun Saha

24341058

Md. Manzar Siraj

21301273

Mohammed Montasir Abdullah

21201600

Approval

The thesis titled “A Multi-Stage Deep Learning Framework for Automated Detection and Localization of Shrimp Disease” submitted by

1. Pritam Sarkar (22101373)
2. Mohammad Tasneem Hasan (21301441)
3. Md. Manzar Siraj (21301273)
4. Arjun Saha (24341058)
5. Mohammed Montasir Abdullah (21201600)

of Fall 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in 31st January 2026.

Examining Committee:

Supervisor:
(Member)

Md. Tanzim Reza

Senior Lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Md. Golam Robiul Alam (PhD)

Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi

Chairperson
Department of Computer Science and Engineering
BRAC University

Abstract

This work presents an extensive Multi-Stage Deep Learning Framework to reduce the significant financial effects of White Spot Syndrome Virus (WSSV) and Black Gill (BG) disease on the \$64 billion global shrimp aquaculture sector by fixing the interpretive problems and restricted ability to diagnose ability of current automated systems. The study develops an ordered detection pipeline starting with disease detection using a dataset of 5,498 high-resolution images. Following a comparison of five state-of-the-art architectures, ConvNeXt-Tiny proved to be the best classifier. It accomplished a test accuracy of 96.8% and a Macro F1-score of 0.97. To solve baseline weaknesses, **Projected Gradient Descent (PGD)** adversarial training was executed to restore robust accuracy to 55.5% against environmental disturbances. After classification of disease, a conditional disease specific U-Net++ segmentation module had been developed for precise lesion identification, giving a Dice Score of 0.8217 for Black Gill and 0.71 for WSSV, which enabled the development of a novel pixel-level Severity Staging system (Mild, Moderate, Severe) for actionable intervention. This research successfully bridges the gap between theoretical computation and dependable, practical application by combining Explainable AI (XAI) with Grad-CAM visualisations for assessing histological features and optimizing for edge-deployment efficiency. This has been achieved by both of these techniques.

Keywords: Shrimp Disease Detection, Artificial Intelligence, Deep Learning, Multi-class Classification, White Spot Syndrome Virus (WSSV).

Table of Contents

Declaration	i
Approval	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
List of Tables	vii
Nomenclature	vii
1 Introduction	1
1.1 Background	1
1.2 Motivation:	2
1.3 Problem Statement:	3
1.4 Objectives:	3
1.5 Scopes and Challenges	4
2 Literature Review	5
2.1 Preliminaries:	5
2.2 Review of Current Literature:	5
2.2.1 The Gold Standard: Molecular and Biological Diagnostic Tech- niques	5
2.2.2 The Transition to Automation: Traditional Image Processing and Machine Learning	7
2.2.3 The Deep Learning Revolution in Shrimp Disease Diagnostics	10
2.3 Summary of Key Findings	13
3 Requirements, Impacts and Constraints	15
3.1 Final Specifications and Requirements	15
3.1.1 Final Specifications and Requirements	15
3.1.2 Societal Impact	15
3.1.3 Environmental Impact	16
3.1.4 Ethical Issues	16
3.1.5 Standards	16
3.1.6 Project Management Plan	17

4	Proposed Methodology	19
4.1	Methodology Overview	19
4.1.1	Data Acquisition and Preparation	19
4.1.2	Stage-I: Disease Classification	21
4.1.3	Stage-II: Conditional Segmentation	21
4.1.4	Advanced Optimization and Robustness Techniques	22
4.1.5	Explainable AI (XAI) and Analysis	22
4.2	Preliminary Design and Model Specification	22
4.2.1	Model-Specific Architectures (Classification)	22
4.2.2	Model-Specific Architectures (Segmentation)	25
4.3	Data Collection	27
4.3.1	Data Cleaning	28
4.3.2	Data Transformation	29
4.3.3	Data Integration	33
4.3.4	Final Datasets Overview	35
4.4	Implementation of Selected Design	36
5	Result Analysis	43
5.1	Performance Evaluation	43
5.1.1	Performance Evaluation Metrics and Techniques for Classifi- cation	43
5.1.2	Performance Evaluation Metrics and Techniques for Segmen- tation	46
5.2	Analysis of Design Solutions	49
5.2.1	Classification Solutions	49
5.2.2	Segmentation Solutions	50
5.3	Final Design Adjustments	50
5.3.1	Classification	50
5.3.2	Segmentation	55
5.4	Statistical Analysis	60
5.4.1	Classification	60
5.4.2	Segmentation	68
5.5	Comparisons and Relationships	72
5.5.1	Classification	72
5.5.2	Segmentation	77
5.6	Discussions	79
6	Conclusion	82
6.1	Summary of Findings	82
6.2	Contributions to the Field	82
6.3	Recommendations for Future Work	83
6.3.1	Adversarial Hardening and Certified Robustness	83
6.3.2	Transition to Adaptive Environmental Preprocessing	83
6.3.3	Clinical Scope Expansion and Co-Infection Dynamics	84
6.3.4	Edge Optimization and Federated Learning	84
6.3.5	Advanced Few-Shot and Meta-Learning	84
	Bibliography	88

List of Figures

4.1	Overview of the Proposed Multi-Stage Deep Learning Framework. . .	20
4.2	Multi-angle image acquisition for a single shrimp sample	28
4.3	Background removed image.	30
4.4	Processed Image Samples	32
4.5	Final Dataset Structure.	36
4.6	Overview of the unified model training pipeline.	38
4.7	Visual representation of segmentation pipeline	41
5.1	Mixup image for different values	44
5.2	CutMix image for different values	45
5.3	Examples of adversarial perturbations - 1	45
5.4	Examples of adversarial perturbations - 2	46
5.5	Training curve for DenseNet201.	50
5.6	Training curve for ConvNeXt-Tiny.	51
5.7	Training curve for EfficientNet.	51
5.8	Training curve for ResNet50.	51
5.9	Training curve for Xception.	51
5.10	Training Loss Over Epochs for DenseNet201.	53
5.11	Training Loss Over Epochs for Xception.	54
5.12	Training Loss Over Epochs for ResNet50 V2.	54
5.17	Performance Visualization for ConvNeXt-Tiny Model.	61
5.18	Performance Visualization for Xception Model.	62
5.19	Performance Visualization for ResNet50V2 Model.	64
5.20	Performance Visualization for EfficientNet Model.	65
5.21	Performance Visualization for DenseNet201 Model.	67
5.22	U-net++ (Whitespot)	68
5.23	Segmentation Performance Metrics Over Epochs for WSSV (U-net +).	68
5.24	U-net (Whitespot)	69
5.25	Segmentation Performance Metrics Over Epochs for WSSV.	69
5.26	U-net++	71
5.27	Segmentation Performance Metrics Of U-net + Black Gill.	71
5.28	U-net	72
5.29	Segmentation Performance Metrics Over Epochs for U-net Black Gill	72
5.30	Grad-Cam Analysis Through ConvNeXt-Tiny	76
5.31	Intensity Visualization of Mild, Moderate, and Severe Shrimp Disease Stages(Black gill)	78
5.32	Intensity Level Progression for White Spot Disease (WSSV).	79

List of Tables

4.1	Hyperparameter Configuration for the Unified Training Pipeline . . .	38
4.2	Transfer Learning Pipeline for CNN Models (Simplified)	39
5.1	Evaluation Metrics for Segmentation	47
5.2	Classification of WSSV Intensity Levels	48
5.3	Classification of Black Gill Intensity Levels	49
5.4	Hyper-parameters Tuning for 3 Classification Models	53
5.5	Classification Performance Metrics for ConvNeXt-Tiny Model	61
5.6	Classification Performance Metrics for Xception Model	63
5.7	Classification Performance Metrics for ResNet50_v2 Model	64
5.8	Classification Performance Metrics for EfficientNet Model	66
5.9	Classification Performance Metrics for DenseNet201 Model	67
5.10	Final Segmentation Performance Metrics (U-Net/U-Net++ Results) .	68
5.11	Final Segmentation Performance Metrics (Model specific results) . . .	69
5.12	Final Segmentation Performance Metrics	71
5.13	Final Segmentation Performance Metrics	72
5.14	Comparative Classification Performance Metrics Across All Models .	73
5.15	Performance for Augmented Regularization on ConvNeXt-Tiny	74
5.16	Adversarial robustness for ConvNeXt-Tiny on the shrimp disease dataset under FGSM attacks (baseline) and after PGD training. . . .	76
5.17	Comparison of Segmentation Model Performance on WSSV and Black Gill	77
5.18	Comparison of Research Novelty with Existing Literature	80

Chapter 1

Introduction

1.1 Background

Shrimp farming is a significant component of the worldwide seafood business, which is very broad and of great significance to the economy, and has expanded a lot. When it comes to meeting the growing demand for seafood all over the world as well as helping the economies of nations that are developing, particularly in Asia—which is responsible for producing about 80% of the world’s farmed shrimp [2], [48]—the industry, estimated at more than 64 billion dollars, is crucial.

A huge difficulty is caused by the ongoing threat of epidemics of diseases, which bring immense financial losses every year that can be estimated in the billions of dollars. Virus infections that are extremely dangerous and have a tendency to propagate quickly across crowded aquaculture systems are responsible for the majority of the losses [2], [32].

The **White Spot Syndrome Virus (WSSV)** causes the most severe of these diseases, known as *White Spot Disease (WSD)*. It is widely acknowledged that WSD has the capacity to result in rapid and high-fatality instances, typically reaching a mortality rate of 100% within the first few days of an outbreak. Furthermore, it has a highly broad host range, as it infects almost all decapod crustaceans [3].

The virus can be hard to manage as it may spread vertically (from parents to offspring), laterally (through water supplies), and by means of eating [3], [32]. WSSV has a huge effect on the economy, which explains the reason it is of such importance for present medical research development and enhancement. This involves making use of genetic methods like polymerase chain reaction (PCR), as well as the creation of technologies for diagnosis driven by artificial intelligence. Several additional economically important variables cause loss of production, but usually go overlooked because of the intense emphasis on WSSV. According to certain people, Black Gill (BG) disease represents one of the “most common challenges in shrimp farming” [42]. It is one of the primary diseases affecting shrimp.

Black Gill, in contrast with White Spot Syndrome Virus (WSSV), is a heterogeneous medical condition which can be attributed to an assortment of triggers and is distinguished by gill melanization. The aetiology’s complicated character has been validated via pathological tests. The environment, secondary bacterial infec-

tions (e.g. *Vibrio* species), and initial fungal infections (e.g. *Aspergillus flavus* and *Aspergillus sydowii*) are all related to the disease [42].

Considering the prevalence and impact of an assortment of components, Black Gill is basically absent from the vast majority of AI-driven diagnostics research. Also, there's a notable difference among the reality of farm management and advances in science, demonstrated by this crucial oversight. Considering its widespread incidence and complex impacts, Black Gill is largely ignored in the vast majority of AI-driven assessment research. This important absence additionally serves to highlight the gap which divides advances in science and reality of managing farms. Existing detection methods have serious limitations; therefore, advanced technology is required. Polymerase chain reaction (PCR) along with other standard tests are extremely specific, yet they take an extended period and require advanced laboratory supplies which might not be simple to gain access to. Furthermore, the existence of white spots alongside additional clinical symptoms can often be lacking for obtaining an accurate diagnosis. Artificial Intelligence (AI) is projected to be an important manufacturing development by 2026 due to its capacity to reduce complexity, accelerate, and enhance diagnosis. A great deal of individuals desire AI to assist them, yet it can only do that when all of its capability is implemented. A method which is able to differentiate among more complicated circumstances such as Black Gill and serious conditions like WSSV could be extremely helpful for contemporary agricultural techniques. Shrimp growers are faced with a lot of various diseases every day, leaving intelligent and effective tools for diagnosis increasingly essential.

1.2 Motivation:

The global shrimp aquaculture field is highly vulnerable commercially, experiencing annual losses up to billions of dollars because of viral infections, therefore demanding this research. **Polymerase chain reaction (PCR)** and other genetic diagnostic techniques are not practicable for proactive farm management because of their expensive nature, long process time, and dependency on special laboratory facilities for reliable outcomes. Due to this, they can only be helpful when detecting infections that occur and are not effective for daily control or mitigation. As an outcome, there is a significant demand for methods that farmers can simply apply despite remains clear.

Due to this, AI has become the centre of a lot of research, and a number of them demonstrated a lot of potential by accurately identifying diseases. These models often suffer from serious weaknesses that leave them unusable for use in reality, in spite of their positive results. The basis of this work is that shown accuracy is an insufficient measure of practical practicality. The principle of the “black box” provides the most significant problem. Throughout the steps of selecting a financially important choice, such as initiating a costly treatment or draining a whole pond, it is ridiculous to think that a farmer, whose earnings relies on the health of their livestock, to simply follow the guidance of an AI. The absence of faith grows when things remain unclear, which is maybe the most important barrier to the adoption process.

In addition, the current setting for research is methodologically not enough. An absence of thorough solutions to the complicated disease issues encountered by farmers is an obvious result of the literature’s insisting on WSSV despite the loss of various other significant pathogens, which include Black Gill disease. A widespread concern for computing effectiveness is another cause which contributes to this issue. As a consequence, multiple modern models have become unfeasible for implementation on low-power devices, affordable devices that are beyond an average farmer’s access. Furthermore, the current “big data” model is acute by nature and isn’t willing to deal with the unpredictable nature of aquaculture, where new disease forms can come up.

The need for quickly creating complete, reliable, operationally appropriate, and versatile diagnostic instruments inspired to initiate this research endeavour. This study hopes to offer a broad framework that addresses the biggest obstacles to AI’s development into a trustworthy and crucial companion in sustainable aquaculture on an international level. In order to accomplish this objective, it will do research on few-shot learning with a focus towards future modification, expand its scope by covering neglected diseases and multi-class issues, evaluate simplicity of the methods effectively for the first time, and thoroughly assess machine effectiveness.

1.3 Problem Statement:

Effectively bridging the divide between research findings and actual agricultural use is the primary barrier keeping these techniques from becoming prevalent, yet models often obtain higher accuracy in classification for certain diseases like White Spot Syndrome Virus (WSSV). The vast majority of present-day diagnostic models are “black boxes” as they do not offer end-users with the open and accurate information that deserves to be there. These three important methodical shortcomings make the basic issue of confidence even harder to address:

1. Studies concentrate on WSSV along with other single-disease methods, ignoring the multiple-pathogen characteristic of aquaculture and diseases like Black Gill.
2. Real Limits Ignoring the computing efficiency of models can lead to results that are too costly to be executed on accessible, low-power hardware.
3. Was unable to modify it because it lacks the necessary data. Present strategies aren’t suitable to detect new or prevalent diseases for which large, labeled datasets have not yet been ready.

1.4 Objectives:

This research focuses on creating and assessing a methodology that thoroughly evaluates the differences between performance, model clarity, accuracy in prediction, and processing effectiveness. In an effort to prepare the path for the creation of fresh possibilities, the research was conducted with accurate and effective tools for diagnosis for agricultural management. For achieving this, the work will:

1. Assess deep convolutional neural networks’ performance on a multi-class classification test.
2. Execute a statistical analysis of methods for providing reliable visual testimony, emphasising model clarity via Grad-CAM localization.
3. Figure out deployment-ready models through assessing lightweight architectures to determine the Pareto front of accuracy vs processing efficiency.
4. Find out if few-shot algorithms for learning have the ability of accurately recognizing specific diseases utilizing just a handful of data.

1.5 Scopes and Challenges

A complete detection technique that does more than just describe White Spot Syndrome Virus (WSSV) will be evaluated as a component of the study. The study examines an extensive variety of diseases, among them known as Black Gill disease, in addition to healthy samples. The scope involves an in-depth examination of current deep learning models, like Convolutional Neural Networks (CNNs), and various benchmarks for performance. We study the development of “black box” models into successful tools for diagnosis through using particular ways for producing visible evidence. Towards practical applications that include real-time monitoring and mobile applications, we have been searching for models that are not just lightweight but also suitable for being widely utilized.

While there have been some changes, this field confronts major challenges. Most models are tested on small, private, and non-standardized datasets, which presents serious issues regarding generalizability. A shortage of trustworthiness from users is a major issue making it challenging to use such approaches. Finally, the requirement for effective, low-power device solutions that can be operated on easily accessible technology in rural settings is in conflict with the rising computational challenge of modern facilities modeling.

Chapter 2

Literature Review

2.1 Preliminaries:

The field of science has created multiple new techniques for diagnosis of shrimp diseases in response to the problems. This has ended up in a major shift in the area of detection, shifting from highly trained molecular techniques to harmless, computerised systems that utilise AI and MACHINE LEARNING. The following literature review draws jointly the findings of 27 significant investigations to give a thorough and analytical analysis of the method in question. Molecular and biological testing, which are the gold standard for accuracy in science; classical image processing and MACHINE LEARNING, which are the first measures with automated, non-invasive assessment; and the new and cutting-edge rise of deep learning and computer vision, that delivers accuracy and immediate use that has never been shown earlier.

The goal of this review is to offer an exhaustive understanding of current state of the art through analysing the modelling methodologies, performance measures, and datasets utilised by these research studies. It will evaluate the claims of accuracy with sceptical thinking, highlight ongoing issues (such as an absence of thorough, standardised datasets), and throw insight into the essential research gaps that must be fulfilled.

2.2 Review of Current Literature:

2.2.1 The Gold Standard: Molecular and Biological Diagnostic Techniques

Molecular biology has been the foundation to identify viruses and bacteria responsible for illness in aquaculture for a long time. These approaches give higher sensitivity and specificity by identifying the pathogen's genetic material or certain proteins, leading to an accurate diagnosis that is unaffected by unclear signs and symptoms. The literature review shows an advanced set of molecular tools developed for WSSV, setting an exceptionally high standard for reliability that upcoming technology needs to meet or improve.

PCR-Based Assays: Sensitivity and Specificity

Polymerase Chain Reaction (PCR) and its advanced forms, which include quantitative real-time PCR (qPCR) and nested PCR, are often cited as the “gold standard” for *WSSV* detection [38]. Nested PCR, for instance, is recognized for its capacity to identify significant and chronic illnesses. Due to this, it is essential to spot asymptomatic carriers that may spread the virus between others without anybody acquiring knowledge. The value of this method was proven by a groundbreaking research investigation that evaluated the frequency of *WSSV* in wild *Penaeus monodon* in the Philippines. When examined with one-step PCR, the findings were negative, showing that nested PCR was effective in detecting *WSSV* in shrimp that did not display any clinical symptoms [1]. In the dry season, 9.6% of shrimp confirmed positive, yet in the time of rainfall, this percentage reduced to 0.3%, corresponding to the study. Furthermore, it indicates what significance environmental variables are and how essential reliable diagnoses are in disease monitoring.

By introducing quantification into evaluations, qPCR helped scientists to both recognize and evaluate pathogen burden. [5] developed and evaluated a one-step SYBR Green-based qPCR test that appears for the VP28 gene of *WSSV*. To acknowledge that their approach detects up to 12 copies per reaction and is 100% specific, revealing no cross-reactivity with six typical shrimp viruses. The worldwide value of this study can be seen by its confirmation throughout an extensive variety of 10 crustacean species and various geographic isolates (Mexico, Brazil, China, and Saudi Arabia). The SYBR Green methodology, which is cheaper and simpler than probe-based methods, represents a significant benefit to nations with poor infrastructure.

A bunch of researchers improved the effectiveness of qPCR through the development of a duplex real-time PCR assay for the simultaneous identification of *WSSV* and *Penaeus stylirostris penstyldensovirus 1* (PstDV1).[7] It is essential to make this action forward, as co-infections have become common and may increase the difficulties of diagnosis and treatment. They showed outstanding analytical sensitivity in their duplex test, which allowed them to identify as few as 2 *WSSV* copies and as much as 20 PstDV1 copies in comparable results. After being enhanced with TaqMan Universal Master Mix (TUMM) and a corresponding relative threshold (CRT) investigation, the clinical sensitivities for both viruses reached 100% on their sample set. This enabled the effective identification of three infectious diseases that may have been ignored by sequential singleplex testing.

Field-Applicable and Novel Biological Methods

PCR-based techniques, while highly precise, often require advanced equipment for testing, proficient personnel, and an extensive amount of time to produce satisfactory results. In light of this exception, approaches that are successful for on-site or “pond-side” implementations have been established. Many examples of these advancements are pointed out in the thorough evaluation provided by [38]. LAMP, which refers to loop-mediated isothermal amplification, is a process that may be used inside of a thermal cycler. This strategy is ten times more effective than nested PCR and can be applied at a constant temperature. It has a limit for detection of 1 fg

of viral DNA. The lateral flowing immunoassay (LFIA) is a farmer-friendly immunized technology that can find high virus loads in animals that exhibit symptoms. However, it is less sensitive and is most appropriate for this particular reason. The findings are provided in approximately 20 minutes for roughly USD 3 each sample.

Researchers have been searching for novel biochemical indicators of shrimp health, as well as to direct pathogen detection[16] analyzed shrimp gut microbiota in an innovative manner. They recognized microbial phyla and looked at their abundance for recognizing healthy and ill shrimp. Through the application of this data, they built a qPCR model that could correctly identify the presence of diseases 91% of all the times. What is even more incredible is that the algorithm used was 70% accurate in identifying disease before any obvious symptoms emerged. This has the possibility of helping transform health monitoring by switching the spotlight from spot-on diagnosis to preventative measures that may be adopted in order to stop an epidemic right in its path. Nanotechnology applications are another probable future subject.[23] The researchers in the laboratories have discovered that WSSV-infected shrimp treated with nanoparticles might survive up to 80% longer than shrimp that were never treated. This is not an AI paper by itself, but it provides information about the progress currently being made toward nanosensors (like electrochemical immunosensors and SPR devices) and nanoparticle-based remedies that could effectively observe and deal with diseases.

Analysis and Limitations of Biological Methods

The primary benefit of molecular and biological methodologies is their analytical validity. These techniques offer an obvious diagnostic lead that acts as a benchmark against which all other methods are evaluated. This is because they can detect copies of viruses as low as a single digit [7] and have nearly full accuracy [5]. Image-based approaches have shortcomings for recognizing carriers without symptoms [1] or estimating the initial stages of disease [16].

But there are a lot of legitimate issues with them as well. Since they are so costly, must have particular infrastructure and expertise, and take an extended period to get results, strategies like nested PCR and qPCR are not helpful for farmers who use them for regular surveillance. Field-ready alternatives, like LFIA, self-awareness for ease of use that can result in false negatives in early-stage infections. Those procedures are also harmful by nature, since they need tissue samples (such as gills and pleopods), which stress the shrimp. The academic community has recently come up with non-invasive, image-based procedures that attempt to computerize the visual evaluation procedures in order to bridge an imbalance between mathematical accuracy and application in real life.

2.2.2 The Transition to Automation: Traditional Image Processing and Machine Learning

In an attempt to discover a more affordable, quicker, less distracting solution to molecular diagnostics, scientists explored computer vision. In the initial stages, research in this field used standard image processing techniques and machine learning

models. First, these methods are used to identify unique details from an image, which include texture, color, and shape. These traits are subsequently put into a classifier to generate a prediction. These initial investigations were frequently overlooked by current research on deep learning, yet they were significant in showing that computerized shrimp disease recognition was viable.

Segmentation and Thresholding for Spot Detection

White spots on the exoskeleton of affected shrimp are the most obvious visual symptom of white spot shrimp viremia. Thus, the very first research attempted the segregation of these areas through the use of image segmentation techniques. For this reason, Researchers suggested an innovative application for *K-Means* clustering.[6] Crustaceans’ broad geographical and spectral information has been collected through hyperspectral imaging. Applying *K-Means*, the image was clustered after thoughtfully choosing the region of interest (ROI). Their subjective assessment indicated that WSSV-infected shrimp were capable of being physically identified. The cost-effective technique and the use of *K-Means* for WSSV in the research were unique in regard to biosensors. Still, the investigation has multiple drawbacks which render its outcomes ungeneralizable. These restrictions consist of employing only a single shrimp image in the dataset instead of a total of two, not including any numerical measures of performance (such as sensitivity or accuracy), and not verifying the outcomes against an existing technique like PCR.

A group of researchers performed a more comprehensive equivalent study, additionally enhancing this approach.[26] Their “Optimized Threshold-Based EM Fusion Technique” has been evaluated towards four up-to-date thresholding technique algorithms, namely Niblack’s, Sauvola’s, Bernsen’s, and OTSU. The methodology’s effectiveness was notably improved, as it contains a dynamic median filter for preprocessing and an enhanced probabilistic and Gaussian thresholding approach. In contrast to other models whose SSIMs fluctuated from 0.64 (Niblack’s) to 0.73 (OTSU), the one that was recommended obtained a far higher value of 0.86. While the SSIM score is higher, it shows that the segmented image is more closely comparable to the original, which indicates the White Spot’s important characteristics have been maintained. Furthermore, it obtained a Peak Signal-to-Noise Ratio (PSNR) of 59 and an increased duration of 3014 ms. This research provides clear, statistical evidence of its model’s superiority, which is a significant advance above past efforts. However, it is marked by an identical limitations, which includes the fact that the dataset that was being used wasn’t known with regard to of its size or authenticity, and the overall classification accuracy was not shared. As a result, its real-world relevance could not be fully analyzed or made practical.

Classical Machine Learning on Diverse Data Types

Different data types have been applied to standard machine learning models with the goal to forecast disease threat, along with for controlling image processing. [28] and [27] both carried out a research investigation that used decision tree-based models—Random Tree (RT), Extra Tree (ET), and J48—to predict the vulnerability

of WSSV to a mixture of physio-chemical water characteristics and spatial factors (e.g., proximity to roads). As the very first time that geographic data for WSSV extension had been merged with machine learning, the technique was very fresh and distinctive.

Area Under the ROC Curve (AUC) for the Extra Tree model was 0.713 on the test set, exceeding RT and J48, in a dataset that includes 228 Vietnamese agricultural areas. The significance of the research has been identified in the establishment of susceptibility visualizations that serve as an essential instrument for regional aquaculture management, regardless of the moderate reliability of the outcomes. However, it fails to take into consideration significant factors that have been shown to influence disease, such as application of chemicals and water temperature.

[10] performed an in-depth assessment of machine learning techniques, covering Back-Propagation Neural Networks (BPN), Support Vector Machines (SVM), and K-means clustering, in the fields of medicine, aquaculture, and agriculture. Although it declared 99% accuracy in diabetes prediction and 98% to 100% in cotton plant illness, it pointed out that K-means clustering had been utilized to detect WSSV in shrimp, but provided no details on rate of identification or accuracy. It also suggests that the utilization of machine learning to the identification of diseases in crustacean species was in its early stages and required complete evaluation in 2016, as opposed to other fields.

Analysis and Limitations of Traditional Approaches

By establishing that computerized systems could be invented for disease identification in lieu of being solely human or based in laboratory techniques, these initial approaches made the most major impact. The scientific changes and the growing capacities of machine learning in aquaculture were extensively documented in recent research papers [10], [29], showing a unique shift in the research setting. Work by [26] has demonstrated that thorough quantitative assessments might improve image processing approaches. By combining non-image data, such as spatial and ecological variables [27], an important step has been made toward a more thorough, epidemiological method for prediction.

However, both of these traditional methods to accomplish tasks aren't working anymore. One of the most significant elements which influences the way well they operate is the standard of workmanship of the hand-crafted materials. In addition to its complex nature, demands on time, and dependence on specific knowledge, feature development is not always straightforward to understand. Often models that have been created are not extremely powerful and fail to operate very well when there are variations in the lighting, the background, or shrimp position that were neglected to be taken seriously all through the feature development procedure. Performances from contemporary methods using deep learning significantly exceed the advertised accuracy, such as the decision tree model's 0.713 AUC. An additional problem that continues to come up in these studies is that they utilize confidential, insufficiently documented or small datasets. This causes it extremely hard to duplicate what they find and makes it unclear to what extent their results can be applicable. The existence of these defects became obvious that we require more reliable models that could on their own obtain important features from data that wasn't processed. As

a result, the deep learning approach eventually came into existence.

2.2.3 The Deep Learning Revolution in Shrimp Disease Diagnostics

Deep learning, specifically Convolutional Neural Networks (CNNs), has entirely transformed the field of computer vision. Convolutional Neural Networks (CNNs) are beneficial for image classification, object detection, and segmentation uses since they can separately generate structures of features from pixel input. In traditional methods, feature engineering has to be done manually. These methods are being used extensively and swiftly to identify shrimp diseases, and the research suggests that these methods consistently generate highest-quality outcomes.

Foundational and Comparative CNN Studies

Many initial efforts to use deep learning for disease detection focused on analysing the effectiveness of various existing CNN architectures. Transfer learning, in which a model trained on an immense, universal dataset (such as *ImageNet*) is modified for specific tasks, was often used. This approach provides an important benefit when dealing with limited special data, as it decreases training time while also improving effectiveness.

A fair standard was set by the contrast research study conducted by [27]. Five Transfer Learning models were tested with a traditional CNN base in a two-class (diseased vs. healthy) and a three-class (WSSV, Black Gill, Healthy) classification challenge. In the three-class test, MobileNetV1 reached the maximum accuracy of 95%, outperforming comparable models such as VGG16 (80%), ResNet50 (85%), and the more complex InceptionV3 (90%). The conventional CNN got 70%. The compact shape and efficacy of the MobileNetV1 proved that it might work on machines with limited capabilities or that are flexible. This research has several problems, the most serious of which is that the dataset is very small. The three-class experiment utilised just 91 photos, while a binary trial took 200 images. As a consequence of the small number of samples as well as the risk of overfitting, the incredible 95% accuracy may not be applicable to reality.

The problem of small quantities often occurs. To determine how to classify seven shrimp conditions—two of that were unhealthy and one of which was healthy—[22] performed an identical evaluation of MobileNets-v1 and Inception-v3. Inception-v3 obtained an accuracy rate of 90.02% with a compilation of 348 images, and this was 89.46% more accurate than MobileNets-v1. The most significant insight was the value of taking out image background; lacking this step, the results would be not as reliable. While sample size continues to be an important problem, transfer learning is nevertheless extremely helpful. For example, an outcome of roughly 90 percent accuracy on a seven-class problem via only 348 images is truly extraordinary. Comparable to that, [44] performed a research investigation on fish diseases that is not applicable to shrimp but gives an excellent illustration of how to compare methodologies. They observed that a Vision Transformer (ViT) model was the most precise, with a score of 95.05%, slightly defeating the traditional CNN (94.44%) on a tiny sample of only 266 images.

A group of researchers developed customised approaches that are both light and efficient. To classify six different shrimp creatures, [25] created “ShrimpNet,” a two-layer convolutional neural network (CNN). In general, CNN was correct 95.48 percent of the time. With a result of 98.14%, it did better still at binary classification when it succeeded in determining what was different between white prawns. These outcomes are useful due to the show that a basic, cost-effective system can work quite effectively. Therefore, cultivation can be automated. Researchers did, however, determine that the dataset had a serious imbalance, and it found that the most common causes of errors were reflections and background noise.

Advanced Architectures, Hybrid Models, and Optimization

Recent studies have examined complex architectures, mixed approaches that combine deep learning with different approaches, and advanced optimisation methods to get better results. They use the Iterative Random Forest Algorithm (IRFA) to determine which features to use. K-Means clustering is implemented for unsupervised segmentation. Also, they apply DLCNN for the final classification. Varma and Krishna initially suggested the integrated workflow, that they later called “SD-Net” [35]. On a seven-class disease dataset, our multiple stages strategy acquired an amazing 98.51% accuracy, beating industry standards which include Inception-Net (83.9%) and MobileNet (75.35%). Deep learning is one of its most outstanding characteristics, but it additionally applies unsupervised and machine learning for improving attributes. Deep learning is one of its key advantages for classification, but it also includes machine learning. The most serious weakness of the research is once more the lack of any outside corroboration and reliance on a small collection that is roughly 650 photographs.

A number of research projects have been directed to advancing the creation of new network frameworks and methods to improve their efficiency. One new model has been developed by [43] via the integration of an Enhanced Gated Recurrent Unit (EGRU) with a Wild Geese Migration Optimisation (GMO) approach. By applying LBP for texture preparation and PLDA for obtaining features, the above complicated architecture obtained a success rate of 98.93% on an eight-class dataset, surpassing those of its rivals. Conventional deep learning methods, such as DICNN (97.81%) and ANN (94.67%), lag behind the EGRU-GMO model. An equivalent network, the Enhanced Recurrent Capsule Network (ERCN), was created by [47] by combining the Marine Predator Algorithm along with Harris Hawks Optimisation. It has obtained an accuracy rate of 95.2%. Through most capsules, it is able to understand surrounding spatial frameworks, and by applying an LSTM layer, it recognises temporally the components. The outcomes of this study show the possibilities of metaheuristic optimisation and sophisticated structures that improve the effectiveness of models, which allows the classification of data near-perfectly. The findings are, however, limited by the dataset sizes, which facilitate performance trade-offs (348 photographs for the EGRU-GMO study, and 1599 for the ERCN).

Deep learning can right now examine a broad range of data sources. [15] used CNNs to histological images of shrimp with the goal to detect the various parts that have been identified in the shrimp. The significant initial phase in scientific study was performed. They were able to obtain an accuracy of 84% with their feature engineering method (BOVW); on the other hand, 95% accuracy had been

reached with their fine-tuned VGG16 model. At this point convolutional neural networks (CNNs) have been shown for their capacity of handling images of different dimensions.

Towards Real-World Application: Mobile Systems and Real-Time Monitoring

Producing tools that manufacturers can use on every day is the main objective of this research. Multiple investigations have made significant progress in this field. The “SHRIMPAI” Android app built by [39] uses a convolutional neural network (CNN) model, which can be exported as a TensorFlow Lite model, to effectively recognise WSSV at early stages of development. Supervised testing achieved a perfect accuracy rate of 100% with WSSV-negative samples and an average accuracy rate of 99.9% with WSSV-positive samples, according to results in the research. On the framework of an improved set of 2,400 images, the model went through training. An excellent 4.4775 out of 5 rating provided to the app by 20 stakeholders, including both farmers and experts, is one of the most essential requirements. This attempt is an excellent instance of how to make a high-accuracy model both easily accessible and working.

A further research effort aimed at real-time video applications via the cutting-edge YOLOv8 object recognition model created by [40]. By the technique of segmenting and monitoring every single shrimp in a pond, their technology has the potential to locate diseased shrimp in real time, along with recognising them. In contrast with YOLOv5, their object identification metrics are superior (mAP50 of 0.879). A complete administration system can be created by the implementation of an alert system, detection, and tracking (via the SORT algorithm). The systematic technique applied to the study is the study’s most effective point; nonetheless, it didn’t seem able to provide proper classification precision for its very own dataset of three thousand more images all throughout every stage of its study.

Analysis and Limitations of Deep Learning Approaches

Deep learning models have clearly established new standards for achievement in the area of image-based shrimp disease recognition. Levels of accuracy reach 90% on an ongoing basis as well as approach 98-99% [35], [39], [43]. Previous studies by [27] and [22] found that transfer learning is an extremely efficient approach that may produce greater outcomes even when access to data is low. Mobile apps like SHRIMPAI [39] and real-time monitoring systems [40] provide it with clear how one may put all of this to use on farms.

With this progress, the dataset issue remains an essential obstacle that impacts every aspect of the field. Small, private datasets that were frequently web-scraped or acquired in a monitored setting were the foundation for both the development and assessment of a large number of these high-performing models. [22], [27], [44] have showed that the model is strongly at risk of excess fitting when it recalls the training data rather than collecting universal features in experiments that implement datasets containing only a few hundred images. The successful application of such models in real agricultural environments, which is affected by a number of variables, such as different light, water turbidity, shrimp species, life stages, and backgrounds,

is put into question as a consequence of this. They have major challenges. Data dietary supplements, as demonstrated by the studies of [39], [41], and [40], is a crucial means to address this issue, but it is unable to substitute real data diversity.

In addition, the vast majority of models are very specific, concentrating just on WSSV detection. [22] and [35] claim that a good method of screening should be able identify between a number of common diseases. Finally, the increasing complexity of methods like ERCN-HHO-MPA [47] and EGRU-GMO [43] on low-power edge computing devices that are frequently used in agricultural environments raises serious problems. This is due to the processing cost as well as operational potential.

2.3 Summary of Key Findings

The summary of the reviewed literature shows significant and swift growth in the domain of shrimp disease detection. The field has gone through an evolution from deeply laborious and very accurate molecular techniques [5], [7] to non-invasive, automated visual systems. The thought’s legitimacy was first demonstrated through traditional image processing and machine learning [26]. Progress has been slowed into tough feature engineering and low performance. Deep learning has transformed the field by regularly attaining classification accuracies above 95% [15] and even reaching up to 98–99% [35], [43]. The difference in performance that occurred between humans and machines in visual diagnosis has been properly balanced for by this. Present research patterns suggest the creation of practical, end-to-end systems. Smartphone apps and real-time water surveillance platforms are two examples [39], [40].

In spite of the immense advancement that has been established, an in-depth review of the industry shows that there are crucial problems that now stop the widespread and reliable application of these technologies nowadays. One of the most important problems is the variance in the existing datasets used as benchmarks. Researchers typically utilize confidential, limited (<1000 photos), and poorly reported datasets [22], [25], [27]. This results in it being extremely challenging to compare models effectively or do the same tests yet again. Therefore, individuals in the industry, government organisations such as BFAR, and academic institutions might work together on a complete, publicly available, and carefully reviewed collection of images (similar to ImageNet or PlantVillage) in the near future [39]. This kind of dataset is needed to address the second key problem: insufficient testing for durability and generalisation. Testing models via data that is not coming from the same range doesn’t happen frequently due to the data deficit. Due to this, it is necessary that further research focus rigorous testing on unknown imagery from different kinds of farms, seasons, and challenging climates. Moreover, it is essential to perform longitudinal research investigations that provide a model through every stage of an agricultural calendar.

Also, the studies showed a limited spectrum of disease detection that is mainly affected by WSSV. This limited focus can be observed in the reality that, despite their prevalence, other significant diseases, such as Black Gill disease, obtain little coverage in machine learning research. To offer farmers all the tools they need, future research needs to go beyond binary WSSV detection and create multi-class models

that can identify the difference between a lot of frequent diseases and, perhaps, rank the seriousness of an infection [22], [47]. But we need further definitions of “early detection” as many visual models demonstrate diseases when indications have been noticeable. Research that uses non-visual signs, such as gut microbiota sampling [16], starts to demonstrate potential, in spite of the fact that research on authentic non-symptomatic detection is currently in its beginnings or in its early stages. In order to perform this, future AI systems must be able to incorporate imagery with data from other sensors of the pond, such as sound waves, e-DNA sequences, and water purity reviews.

The lack of explainability (XAI) and farmer confidence might be the most significant factor affecting real-world application. Prior to performing expensive measures that correspond with an AI’s proposals, farmers must immediately solve the “black box” quality that is present in the vast majority of models. Combining methods from right next sectors, like Grad-CAM visualisations utilised in ichthyopathology , or incorporating comprehension within models via attention mechanisms [37], should be a critical component of research. A development of AI that can be both understandable and focused on users should be between the most significant elements of research.

Lastly, diagnostics are most successful when they can be utilised for supervision instead of to projection. The primary objective should be to develop complete decision support systems that combine AI-generated reports with environmental data [31] and recognised disease risk variables [36] to make sure farms are provided with guidance from administrators according to accurate data. So as to assure that these advancements are effortless to apply and trustworthy, it is essential that farmers take part in the creation process from beginning to end [36]. Such a strategy combines sophisticated research with implementation.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

3.1.1 Final Specifications and Requirements

The proposed multi-stage deep learning system will function effectively only if an individual set of technical, data, and methodological requirements are satisfied for high-performance diagnosis in aquatic settings. The approach uses an amalgamation of both data acquisition techniques, resulting in 5,498 high-resolution photos that fall into three primary groups: healthy, White Spot Syndrome Virus (WSSV), and Black Gill (BG). In a bid to preserve methodical uniformity, every image is originally resized to 224×224 pixels and subsequently presented to the Segment Anything Model (SAM) for precise background removal. This distinguishes biological characteristics from background noise. In the context of hardware, the training of complicated techniques which include DenseNet201 and ConvNeXt-Tiny needs strong graphics processing units (GPUs). EfficientNetV2-B0 is used in the final installation in order to save resources by using as few settings as possible. This works well on light device edges and mobile computing platforms. Dependent on the PyTorch framework, the software infrastructure includes particular packages such as `segmentation-models-pytorch` for the U-Net and U-Net++ designs, and additionally Roboflow for manually adding annotations.

3.1.2 Societal Impact

The project's social effects are highly dependent on the prosperity of nations that are developing, particularly those in Asia, whose farmed shrimp makes up nearly eighty percent of the global total. The study improves the security of livelihood for small-scale farmers, who frequently experience fatal 100% mortality rates during White Spot Syndrome Virus (WSSV) attacks, through providing an immediate and accessible detection approach. In addition, maintaining a stable seafood supply is an important factor to international food security, as it protects a global industry estimated at more than \$64 billion. The system's goal of building acceptability and trust through the use of Explainable AI (XAI) is an important social aspect.

By the application of Grad-CAM localization heatmaps, the system changes from a mystery “black box” into an easily understood instrument that technologists and farmers may depend on when making significant choices regarding management. These are done by providing “visual proof” of its diagnostic outcomes.“

3.1.3 Environmental Impact

This approach actively promotes long-term development through active disease management, which prevents the disease from propagating in the first place. This reduces the need for proactive remedies that devastate the environment. Farmers may prevent dangerous diseases and contaminants from going into the water supply of the region around them if they can quickly and accurately spot pathogens. These objectives may be attained through avoiding extreme measures like clearing entire water. Additionally, the research shows how essential it is to preserve energy through placing a focus on tiny layouts that are suitable with devices which require little power use. Employing less power means less carbon pollution throughout agricultural areas. In response to the study, a mix of visual information and non-visual technology, such as vibration analysis, water purity surveillance, and gut microbiome specimens, might be utilized in the future. This could make it simpler to identify clusters faster, that could prevent them from showing up clinically.

3.1.4 Ethical Issues

The concept is founded on a strong adherence to professional and ethical norms regarding information confidentiality and animal welfare. During the imaging technique, every shrimp sample was treated with utmost care, and any specimens when the scan was completed, the results that showed that they were healthy were promptly conveyed. That had to happen to make sure that all of the samples had been treated correctly. For the purpose of ensuring that the data were accounted for and providing a clear record of the evidence, we meticulously documented each stage of the collecting procedure. By integrating supplementary data sets such as ShrimpDiseaseBD [45], the research team implemented substantial measures to mitigate bias resulting from algorithms. They developed a standardized reannotation procedure to eliminate “annotator bias.” Therefore, the same experts who worked on the first piece of data also looked at each of the other images. By implementing the subsequent steps, it was ensured that the diagnostic standard will always stay the same.

3.1.5 Standards

The study maintained technical and professional rigor by contrasting performance to the “gold standard” of molecular and biological diagnostic techniques like PCR and qPCR. All manually polygonal and pixel-level masks were evaluated to an unanimous approval among research experts to ensure absolute biological accuracy, and the annotation procedure kept to rigorous quality assurance procedures. Furthermore, the research made the most of a technical standardization technique in an effort to integrate photos emerging from a variety of resources into a collection that was statistically consistent. It keeps the deep learning models from becoming impacted

by things like moving camera sensors or lighting in the environment and solely looks at significant pathological features, such as calcified spots or melanization on the gills.

3.1.6 Project Management Plan

The project had been carried out in a systematic four-stage procedure, and the goal had been to integrate academic development with real-world application. These were the primary objectives of the first phase: collecting data and making preparations. This consisted of systematic gathering of images from various perspectives of marketplaces and fields in Dhaka and Satkhira. Following the classification step, five new Convolutional Neural Network (CNN) models were evaluated to determine the best balance between speed and accuracy. Throughout the segmentation step, conditional U-Net and UNet++ models were applied to detect specific lesions and estimate the severity of symptoms. The last stage was to look at what came up and determine just how simple they were for one to comprehend. The entire group use Gradient-weighted Class Activation Mapping (Grad-CAM) to examine the images and strengthen the system towards disease threats; that ensures that the system will operate in difficult situations.

3.1.6.1 Risk Management

The research identified and resolved a variety of significant issues in order to be certain that the model would function properly in the real world. Creative methods of regularization, which includes Mixup and Label Smoothing, and high dropout rates (up to 0.6%) the work addressed the significant danger of overfitting that comes with complex models such as DenseNet201. We checked the Fast Gradient Sign Method (FGSM) findings and noticed that the model’s accuracy might decrease considerably when it had been faced with shifts. To prevent this from going once more, we provided adversarial training. A single application of the FGSM was to evaluate possibilities for adversarial fragility. Furthermore, strategic transfer learning has been implemented to resolve the challenge of lack of enough information on new diseases. Under the structure of this technique, weights that had been previously trained on the huge ImageNet dataset have been used for the purpose to offer the models with an elementary knowledge of its physical features.

3.1.6.2 Economic Analysis

The method provides an advantageous expense to benefit ratio compared to conventional laboratory evaluations, indicating a significant economic advantage. The system is an essential financial safeguard for the fish farming industry because it attempts to avert a large amount of the many billions of dollars in revenue losses suffered annually because of viral outbreaks. Conventional Polymerase Chain Reaction (PCR) methods are expensive, time-consuming, and require specialized laboratory equipment and personnel. This innovation is powered by machine learning, making it possible to perform low-cost, "pond-side" testing which provides fast results, in contrast to conventional approaches. The undertaking further safeguards hardware accessibility by organizing itself around affordable models like EfficientNetV2-B0.

Producers can save cost on things they purchase while maintaining accurate diagnostics.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

This research offers a comprehensive, multistage deep learning framework designed for automatically identifying and localizing shrimp diseases, which include Black Gill and White Spot syndromes, in addition to Healthy samples. The methodology includes an ordered pipeline that first evaluates shrimp health and subsequently employs an individual segmentation network to identify infected regions.

The proposed workflow is divided into four primary phases:

1. Data Acquisition and Annotation,
2. Stage-I: Disease Classification,
3. Stage-II: Disease Segmentation,
4. Post-Analysis and Interpretability.

4.1.1 Data Acquisition and Preparation

To enhance model performance, a hybrid dataset was developed by integrating main field data with secondary external data.

Data integration: The primary dataset was augmented using secondary sources to reduce disparities in class and encourage diversity. For the purpose of establishing one repository, every single image was standardized.

Annotation Strategy: Manually annotation was performed to produce ground truth.

Classification Labels: The images have been divided into three distinct categories: Black Gill, White Spot, and Healthy.

Segmentation Masks: A dual-annotation technique was used for diseases that were apparent in the environment. Polygon masks were utilized to label the Black Gill regions with the goal of retaining unusual boundary shapes during the process. As a replacement, pixel-level masks were utilized for identifying the White Spot

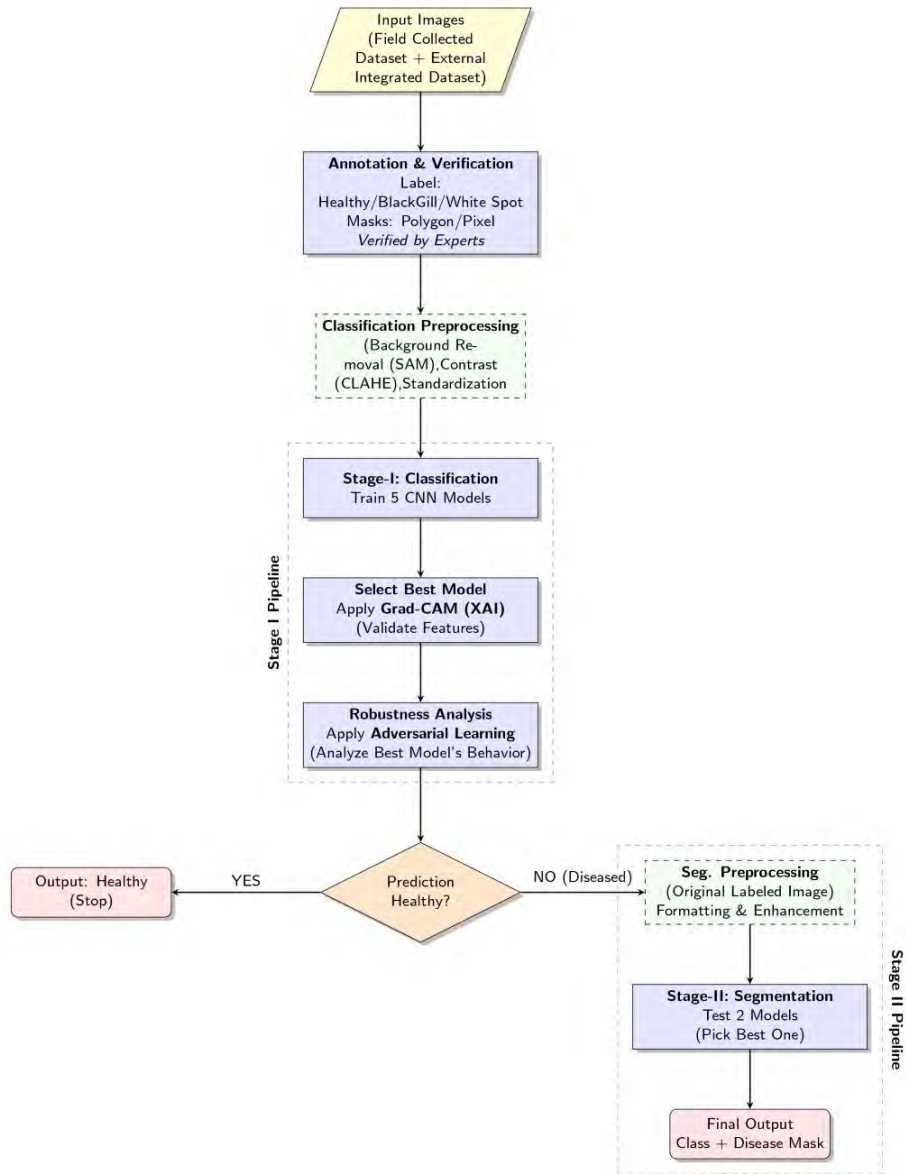


Figure 4.1: Overview of the Proposed Multi-Stage Deep Learning Framework.

lesions in order to account for the reality that the disease has spread out in numerous locations.

Pipelines for Preprocessing: We generated two distinct preprocessing processes to satisfy the requirements of classification and segmentation tasks.

- **Classification Preprocessing:** Images were resized, normalized, and noise was reduced so that Convolutional Neural Network (CNN) models were able to extract features from them.
- **Segmentation Preprocessing:** Original high-resolution images that have been identified as diseased were forwarded through a second pipeline. As a part of this procedure, contrast enhancement and specific structuring have been used so as to ensure the creation of the best feasible mask during the segmentation phase.

Preprocessing Pipelines

Two distinct preprocessing pipelines were developed to cater to the specific requirements of classification and segmentation tasks:

- **Classification Preprocessing:** Images underwent resizing, normalization, and noise reduction to prepare them for feature extraction by Convolutional Neural Network (CNN) models.
- **Segmentation Preprocessing:** For images identified as diseased, the original high-resolution images were routed through a secondary pipeline. This involved contrast enhancement and specific formatting to ensure optimal mask generation during the segmentation phase.

4.1.2 Stage-I: Disease Classification

The initial stage of the framework acts as a diagnostic system. To identify the most effective feature extractor, five distinct Convolutional Neural Network (CNN) structures have been developed and examined.

Custom Transfer Learning: It solves data scarcity issues associated with aquaculture datasets. Pre-trained weights were used, and the classification headings were modified to take into account the three-class problem (Healthy, Black Gill, White Spot).

Model Selection: The five potential models were trained and verified to ensure they performed. On the basis of the concept that produced the best accuracy and F1-score, the classification stage was created.

4.1.3 Stage-II: Conditional Segmentation

The stimulation of this step is dependent upon the occurrence of specific requirements. The next step is contingent upon the fact that the Stage-I classifier says that the image is healthy. Yet, if the image is recognized as a Black Gill or White Spot, it advances to the segmentation stage module that uses an additional preprocessing pipeline.

Architecture Evaluation: Three complex segmentation techniques have been assessed for their value in creating disease masks.

Localization: To produce the binary mask overlay, the best segmentation model was applied for localization. This model indicated specific pixels that related to the symptoms of the disease, including the darkening gills or white patches.

4.1.4 Advanced Optimization and Robustness Techniques

Advanced methods of learning have been employed in order to ensure the model works properly with data it hadn't seen before and doesn't become too optimal.

Adversarial Learning: This was carried out in order to render the model stronger by letting the network learn in a more secure and distinctive manner by finding out how fake and real samples are distinct or by training against modifications made by a rival.

Augmented Regularization: With the aim to avoid overfitting, robust methods for regularizing have been combined with extensive data augmentation (rotations, flips, mild modifications) over the course of training of the best classification model.

4.1.5 Explainable AI (XAI) and Analysis

Gradient weighted Class Activation Mapping, additionally referred to as Grad-CAM, was used to try to validate the reliability of the "Black Box" deep learning models. Visual validation was carried out using Grad-CAM heatmaps. CNN concentrated on particular parts of the image when making its decision about classification. For instance, spots or gill color are significant indications of disease, so this helps make sure the model gets the disease instead of simply background noise.

4.2 Preliminary Design and Model Specification

This study examines the efficacy of six distinct Convolutional Neural Network (CNN) designs for shrimp disease classification. With architectural performance as our primary goal, we developed an integrated training variable. The process makes certain that all of the tests employ the same methods for data augmentation, optimization, and regularization. In this manner, performance indicators show the models' abilities rather than their training technique, which could make it inaccurate.

4.2.1 Model-Specific Architectures (Classification)

The studies employed many neural network (CNN) models, notably EfficientNetV2-B0, ResNet50V2, Xception, DenseNet121, and ConvNeXt-Tiny. Because of their established equilibrium among superior prediction accuracy, computational effectiveness, and applicability for transfer learning in image classification tasks, these models were chosen. The training procedure starts with each model operating on the huge ImageNet dataset. Through doing so, they can effectively gather characteristics for use at a subsequent stage. The models have been changed so that they

are compatible with the data that was utilized in the present research. Further details regarding each model’s basic ideas, mathematical formulations, layer-by-layer arrangements, and how it’s constructed can be obtained in the parts that follow. These subsections concentrate on the elements that make each model special, the novel discoveries it provides, and the manner in which it can be applied with high accuracy or with scarce means.

4.2.1.1 EfficientNetV2-B0

EfficientNetV2-B0 is the latest version of the EfficientNet family. It is intended to enhance both training speed and parameter efficiency through the use of advanced neural architecture search and progressive learning techniques [30]. Using fused-MBConv blocks that combine the expansion and depthwise convolution layers in the early stages, this model improves on the combined scaling method of the original EfficientNetV1, which uniformly scales network depth, width, and input resolution, while decreasing computational latency without sacrificing accuracy. As a result of reducing memory access overheads, this fusion can be particularly useful for edge and mobile devices. In addition, EfficientNetV2 employs regularised training techniques, such as stronger data augmentation for larger variants, to enhance generalisation without producing proportional rises in training time.

The Mobile Inverted Bottleneck Convolution (MBConv) block, which is available in fused or standard forms, acts as the basis of the architecture. To make the channels appear bigger, the standard MBConv first broadens the input features. Following that, it does depthwise convolution to obtain the spatial features, squeeze-and-excite for channel-wise attention, followed by a projection back to the original dimensions, while also maintaining the gradient. Mathematically, this is expressed as:

$$\mathbf{y} = \mathbf{x} + \mathcal{P}_{1 \times 1}(\sigma(\mathcal{D}_{k \times k}(\sigma(\mathcal{P}_{1 \times 1}(\mathbf{x})))) \cdot \text{SE})$$

In this instance, $\mathbf{x}$ indicates the input feature map, $\mathcal{P}_{1 \times 1}$ symbolises pointwise convolution for channel modification, $\mathcal{D}_{k \times k}$ indicates depthwise convolution utilising a unique filter for every channel that is input with a kernel size of k for efficiently capturing local patterns, and σ indicates the Swish activation function, defined as $\sigma(z) = z \cdot \text{sigmoid}(z)$, which introduces non-linearity with a smooth gradient to improve optimisation. The Squeeze-and-Excitation (SE) module enhances pertinent channels by calculating:

$$\text{SE} = \text{sigmoid}(\mathcal{P}_{1 \times 1}(\text{GAP}(\mathbf{z}))) \cdot \mathbf{z}$$

where GAP is global average pooling, which changes spatial data into a channel identifier. The initial growth and depthwise operations are merged into a single standard 3x3 convolution in fused-MBConv variants. This helps our inference by avoiding the requirement for separate kernel launches.

4.2.1.2 ResNet50V2

ResNet50V2 is an enhanced version of the Residual Network architecture, characterised by its pre-activation structure that places Batch Normalisation and ReLU activations before convolutional operations, hence enhancing gradient flow and training stability in very deep networks [9]. This method of pre-activation, known as

identity mapping, solves the vanishing gradient problem better than the original ResNetV1 through making certain that signals flow more smoothly via remaining paths. This makes for deeper patterns without impacting performance. The computational complexity of the model is reduced by the bottleneck structure, which uses 1×1 convolutions for dimension reduction and expansion. This method proves successful for large-scale image allocation and has a strong generalisation.

The pre-activation bottleneck residual unit is the most significant part. It passes inputs via a series of normalisation, activation, and convolutions, concluding with a residual addition. This is mathematically formulated as:

$$\mathbf{y} = \mathbf{x} + \mathcal{P}_{1 \times 1}(\sigma(\text{BN}(\mathcal{D}_{3 \times 3}(\sigma(\text{BN}(\mathcal{P}_{1 \times 1}(\sigma(\text{BN}(\mathbf{x})))))))$$

In this equation, x represents the input tensor, and $\mathcal{P}_{1 \times 1}$ utilizes pointwise convolutions to alter channel dimensions—typically decreasing them by a factor of 4 in the bottleneck to improve computational speed before subsequent expansion. The operator $\mathcal{D}_{3 \times 3}$ denotes a 3×3 convolution. The activation function $\sigma(z) = \max(0, z)$, known as the Rectified Linear Unit (ReLU), introduces non-linearity, while BN represents Batch Normalization, which stabilizes training by normalizing activations across minibatches to minimize internal covariate shift. A 1×1 convolution with a stride of 2 is employed in the shortcut path to match dimensions for blocks that downsample spatial resolutions, ensuring that the residual addition proceeds smoothly.

4.2.1.3 XceptionNet

Extreme Inception (Xception) is a revolutionary framework that substitutes everything of the multi-branch Inception modules with depthwise separable convolutions. It provides this by believing that channel and spatial correlations can be separate for better efficiency and performance, outperforming InceptionV3 with comparable settings [11]. By handling each channel separately before mixing them, this separation scales down the operations. While paired with residual connections for deep training, it performs effectively for tasks requiring strong feature disassociation.

The depthwise separable convolution core decomposes into:

$$\mathbf{y} = \mathcal{P}_{1 \times 1}(\mathcal{D}_{3 \times 3}(\mathbf{x}))$$

with residuals: $\mathbf{y} = \mathbf{x} + \text{sep_conv}(\mathbf{x})$, where $\mathcal{D}_{3 \times 3}$ applies per-channel 3×3 filters for spatial mapping, and $\mathcal{P}_{1 \times 1}$ mixes channels via pointwise operations, using ReLU for non-linearity.

4.2.1.4 DenseNet201

DenseNet201 shows improved feature reuse and parameter efficiency by the dense connection of layers within blocks, enabling each layer to get concatenated inputs from all previous layers, hence reducing vanishing gradients while encouraging compact representations [12]. This close connection cuts off the requirement for wide channels. The amount of new features brought to each layer is regulated by a growth rate, and transition layers get smaller to save space. The dense operation is:

$$\mathbf{x}_l = H_l([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{l-1}])$$

with concatenation $[\cdot]$, and H_i as BN-ReLU-1x1 Conv (bottleneck)-3x3 Conv, with a growth rate $k = 32$.

The stem uses a 7x7 convolution to 64 channels, no growth, followed by max pooling. Dense block 1 has six layers, growth 32, 256 channels, extracting low-level features, then transition with 1x1 conv and 2x2 pooling for compression. Dense block 2: twelve layers, growth 32, 512 channels, mid-level. Transition. Dense block 3: forty-eight layers, growth 32, 1664 channels, high-level. Transition. Dense block 4: thirty-two layers, growth 32, 1920 channels. Head: 7x7 pooling and FC to classes. This promotes efficiency, ideal for thesis comparisons. Key parameters: 20 million weights, 4.3 GFLOPs, input 224x224x3.

4.2.1.5 ConvNeXt-Tiny

ConvNeXt-Tiny refreshes pure convolutional networks through incorporating Transformer-like elements—such as enormous kernels, inverted bottlenecks, and sophisticated normalizations—into a modern ConvNet architecture, achieving Transformer-level performance with the simplest of convolutional methods [34]. The 7x7 depthwise convolutions in this hybridization create large receptive fields. The GELU function gives smooth nonlinearity, and the LayerNorm function guarantees stability. Separate downsampling avoids the data from getting erased.

The block is an inverted bottleneck:

$$\mathbf{y} = \mathbf{x} + \mathcal{P}_{1 \times 1}(\text{GELU}(\mathcal{P}_{1 \times 1}(\text{LN}(\mathcal{D}_{7 \times 7}(\mathbf{x}))))$$

with LN, expansion 4, GELU $\text{GELU}(z) \approx 0.5z \left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}(z + 0.044715z^3)\right)\right)$, and 2x2 strided conv for downsampling.

It begins with the stem: 4x4 convolution to 96 channels, stride: 4. Three blocks, 96 channels, a 7x7 kernel, stride 1, and low-level are composed of the first stage. Reduce the size to 192. Stage 2: three blocks, 192 channels, 7x7, and a stride of 1. Perform a downsampling to 384. Third stage: 9 blocks, 384 channels, 7x7. Resize to 768 pixels. Fourth Stage: 3 blocks, 768 channels, 7x7. Focus: introducing FC and pooling to classes. This works perfectly with sophisticated research investigations since it bridges paradigms. About 28.6 million weights, 4.5 GFLOPs, and 224x224x3 input have important features.

4.2.2 Model-Specific Architectures (Segmentation)

The architectures of the three deep learning models utilized for black gill segmentation in shrimp images are explained in this section: U-Net, U-Net++. Due to problems like changing lighting, occlusions, and irregular disease patterns, each model was selected to assist with semantic and instance segmentation in underwater aquaculture imagery. As the reason for enhancing feature extraction from limited domain-specific data, the models take advantage of encoder-decoder paradigms or region-based convolutional networks that have been trained on ImageNet for transfer learning. We offer a description of the fundamental components, which consist of encoders, decoders, and loss functions, alongside providing mathematical formulations for significant metrics and objectives.

4.2.2.1 U-Net Architecture

Ronneberger et al. (2015) first suggested the U-Net model, which refers to a convolutional neural network (CNN) made for biomedical image segmentation. For preserving spatial information across resolutions, it has an encoder-decoder structure with jump links. For the intent of our implementation, we utilized the capabilities of the segmentation program for Python in order to instantiate U-Net with a ResNet-34 backbone as the encoder. This decoder was pretrained on ImageNet. The decision you made strikes a good mix between the speed at which it can be calculated and how deeply it can represent features. The 34 layers contain residual blocks to stop vanishing gradients from occurring.

The encoder captures hierarchical features F_k at stage k by progressively downsampling the input image $I \in \mathbb{R}^{H \times W \times 3}$ (resized to 512×512) through convolutional layers and max-pooling:

$$F_k = \text{Conv}_{3 \times 3}(\text{ReLU}(\text{BN}(F_{k-1}))) + F_{k-1} \quad (\text{residual block}), \quad (4.1)$$

where BN denotes batch normalization and the downsampling halves spatial dimensions while doubling channels (e.g., from 64 to 512 at the bottleneck).

The decoder upscales these traits by transposed convolutions, combining them with equivalent encoder features via bypass connections to improve boundaries:

$$D_k = \text{UpConv}_{2 \times 2}(D_{k+1}) \oplus F_{m-k}, \quad (4.2)$$

followed by convolutions, where $\oplus$ denotes concatenation. The final output is a sigmoid-activated single-channel mask $M \in [0, 1]^{H \times W}$.

Training improves a custom Masked Tversky Loss, what modifications the Tversky index (Salehi et al., 2017) [13] to focus on regions of interest (ROI) by masking pixels in the background based on the intensity of the image. The Tversky index T has been defined as:

$$T = \frac{TP + \epsilon}{TP + \alpha \cdot FP + \beta \cdot FN + \epsilon}, \quad (4.3)$$

where TP, FP, FN are true positives, false positives, and false negatives; $\alpha = 0.3$, $\beta = 0.7$ (prioritizing recall to detect subtle gill lesions); and $\epsilon = 1$ for stability. The loss is $L = 1 - T$, masked by an ROI filter $R = \max(I_{\max} > -2.0, GT)$, where $I_{\max}$ is the maximum channel intensity and GT is the ground truth mask.

Performance is evaluated using Intersection over Union (IoU):

$$\text{IoU} = \frac{TP}{TP + FP + FN}, \quad (4.4)$$

along with precision $P = \frac{TP}{TP + FP}$ and recall $R = \frac{TP}{TP + FN}$.

4.2.2.2 U-Net++ Architecture

U-Net++ [18] improves U-Net through the use of nested and dense skip paths, enabling multi-scale feature aggregation for better segmentation of complex structures like black gill spots and white spots. We implemented U-Net++ with an

EfficientNet-B0 encoder [19], pretrained on ImageNet. By employing mobile inverted bottleneck convolutions (MBConv) together with squeeze-and-excitation (SE) blocks, EfficientNet-B0 is able to accomplish higher efficiency by means of compound scaling.

The encoder extracts features similarly to U-Net but with MBConv blocks:

$$F_k = \text{SE}(\text{DepthwiseConv}_{k \times k}(\text{PointwiseConv}(F_{k-1}))), \quad (4.5)$$

where SE recalibrates channel-wise dependencies: $\text{SE}(x) = x \cdot \sigma(\text{FC}(\text{AvgPool}(x)))$, and σ is sigmoid.

The decoder enables gradient flow through fewer channels by integrating dense connections across multiple upsampling paths:

$$D_{i,j} = \begin{cases} \text{Down}(D_{i-1,j}) & j = 0, \\ \text{Conv}(\text{Up}(D_{i,j-1}) \oplus [D_{k,j-1}]_{k < i}) & j > 0, \end{cases} \quad (4.6)$$

where i, j index decoder levels and paths, and outputs are deep-supervised by averaging multi-scale predictions.

The loss is a balanced Masked Tversky Loss ($\alpha = \beta = 0.5$), which penalizes false positives and negatives equally, boosting precision in cluttered shrimp images. Calculations of metrics (IoU, precision, dice loss, and recall) are carried out in the same manner as U-Net. A ReduceLROnPlateau scheduler modifies the learning rate based on the validation IoU, which increases convergence.

4.3 Data Collection

Data collection was done systematically to ensure accurate representation and applicability to the Bangladeshi aquaculture environment, focusing on fresh shrimp samples collected from local markets in Dhaka and also from shrimp farms in Satkhira. The 1017 shrimp samples taken were carefully selected from trustworthy suppliers at the Karwan Bazar and also from the shrimp farmers of Satkhira. They have been selected for their wide range of varieties. We have included multiple species of shrimps that are available in Bangladesh; these species are the primary commercial species. The shrimp samples in this dataset range in size from 5 to 15 cm. For each shrimp collected from Dhaka, we have taken images from distinct angles (dorsal, ventral, side to side, left/right, front, and backside, etc.) against a uniform white background under monitored circumstances. This was performed to reduce background variance and improve background segmentation. The requirement of identifying asymmetric disease markers, such as gill infections that aren't apparent from just one aspect, made this multi-angle approach essential. It led to 3475 high-resolution photos (average 6000×4000 pixels) saved in JPEG format of 580 unique shrimps.

On the other hand, for each shrimp collected from Satkhira, we have taken images from two angles for each shrimp as they are small in size and didn't have enough space to capture in multiple angles against different backgrounds under monitored circumstances. It led to 874 high-resolution photos (average 6000×4000 pixels) saved in JPEG format of 437 unique shrimps.

The requirements for ethical behavior were observed, such as the handling of gentle samples and the immediate distribution of healthy samples after photographing. Moreover, each action was documented for the purpose of providing accountability. Each shrimp's all-angle images are stored with the shrimp ID so that one shrimp's images are not mixed with those of other shrimp.

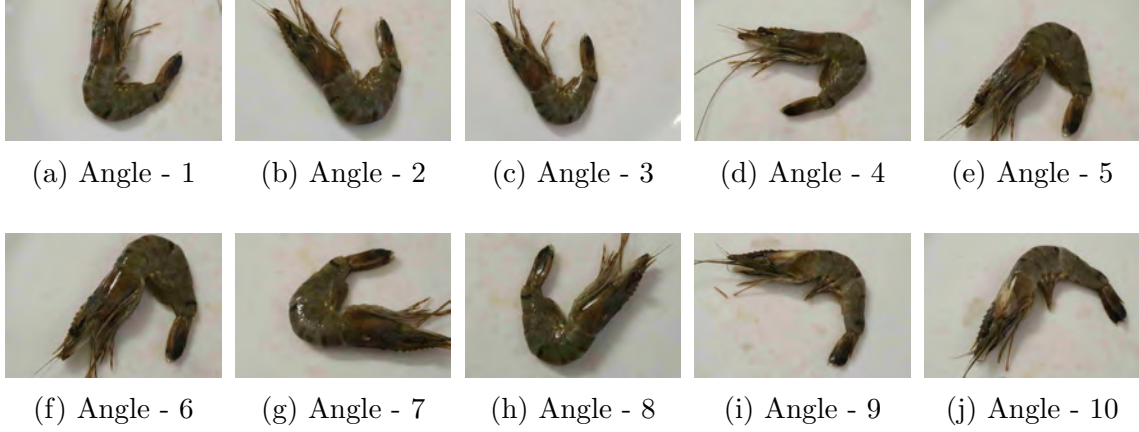


Figure 4.2: Multi-angle image acquisition for a single shrimp sample

Example: Shrimp_305(5) Here, 305 is the shrimp ID, and 5 is the image ID of the 305th shrimp.

Image acquisition was carried out with a Canon M50 Mark II mirrorless camera mounted with a 50mm prime lens. The data was taken indoors with normal lighting (like fluorescent or LED lab bulbs hanging from the ceiling), and no on-camera flash or external flashes were employed. With the reality that the high aperture of the 50mm lens allowed it to capture clear images under these conditions, the placement of the overhead light sources with respect to the subject organically led to the formation of regular specular reflections on the shrimp's wet outer shells. These characteristics appear frequently in testing performed in field aquaculture. Two domain experts (fishery researchers) jointly collaborated in the labeling of images into the Healthy, White Spot, or Black Gill classes, solving annotation differences. The labeling software used was Roboflow, and inconsistencies were solved through a majority vote with other researchers.

4.3.1 Data Cleaning

During the data exploration and data cleaning procedure, images that had issues with quality, like blurry or partly shown shrimp bodies, awful lighting, etc., were eliminated from the list in order to make sure only those with high-quality images were added to the final edition. After an extremely extensive cleaning process, 4349 high-quality pictures were selected, which show 1017 various shrimp samples. To ensure the classification model's robustness and the continued existence of independent class definitions, samples showing signs of co-infection (i.e., the existence of both WSSV and Black Gill disease on the same sample) have been removed from the dataset. This elimination criterion was used to eliminate feature aggregation during the training phase, enabling the model to gain distinctive descriptions for

each disease without the ambiguity resulting from multi-label cases. Moreover, the samples shrimp with both Black Gill and WSSV (BG_WSSV) samples, which are very rare and have very small samples, some samples have major signs of a single disease, which is why we kept them in that disease class. As our dataset images are from different, diverse angles, we skipped this multi-label class to reduce the complexity of the data.

These shrimps were classified as follows:

1. **Healthy Shrimp:** no visible symptoms of disease.
2. **Shrimp with Black Gill (BG):** shrimp diagnosed with Black Gill disease, which have black or blackish spots in the gill area.
3. **Shrimp with White Spot Syndrome Virus (WSSV):** shrimp infected with White Spot Disease, which have tiny white dots within 0.5–2.0 mm in mainly the head area and can be in any part of the body.

4.3.2 Data Transformation

4.3.2.1 Preprocessing for Classification

To ensure the model focuses exclusively on relevant morphological features—specifically white spot lesions—and to mitigate the influence of irrelevant environmental context, a robust data transformation pipeline was implemented. This pipeline focuses on isolating the region of interest (ROI) and standardizing input quality prior to network ingestion.

Background Removal and Image Enhancement The raw dataset underwent a preprocessing routine designed to clean the visual data and enhance disease markers.

- **Background Removal via SAM:** To remove background noise that could end up in incorrect relationships, we used the Segment Anything Model (SAM). This fundamental model was used to produce high-accuracy segmentation masks for the shrimp subjects. By using these masks, background pixels have been removed (turned black), ensuring that the classification model examines only the biological characteristics of the individual, eliminating environmental distortions.



(a) a



(b) b



(c) c

Figure 4.3: Background removed image.

- **Adaptive Contrast Enhancement (CLAHE):** After the background had been taken out, image measures were looked at to observe how good the picture was. Particularly, images with low contrast (standard deviation ≤ 25) were subjected to a Conditional Contrast Limited Adaptive Histogram Equalisation (CLAHE). For a sensitive enhancement that makes likely spots easier to see without adding extra noise or over-amplifying the low background, we chose a grid size of 16×16 and a clip limit of 1.2.
- **Standardization:** Ultimately, both preprocessed images were scaled to a

consistent intermediate dimension of 224×224 pixels and limited to the valid pixel range of $[0, 255]$ to preserve data integrity before dataset splitting.

4.3.2.2 Preprocessing for Segmentation

We will explain the preprocessing methods for the black gill and white spot datasets in the next section. These datasets have been designed to address the particular difficulties that accompany shrimp imagery regarding the detection of diseases. This process may enhance models' generalization through rendering characteristics clearer, maintaining consistency, and adding data. For the purpose of carrying out transformations, both pipelines take advantage of the Albumentations library. The modifications comprise scaling, normalization, and augmentations that simulate actual aquaculture variations such light inconsistency and shrimp direction.

Preprocessing for Black Gill Dataset

To prevent background noise impact and highlight boundary specifics, the black gill dataset contains shrimp images with broad, diffuse gill lesions against dark backgrounds. Augments are required. Following the masks are binarized (> 0 as focus), images are uploaded from the separated folders (train, val, and test). The pipeline has:

- **Resizing and Padding:** To preserve aspect ratios with no distortion, images are reduced to 512×512 utilizing maximum scaling, then converted to zero-padding.
- **Normalization:** ImageNet data (mean: $[0.485, 0.456, 0.406]$; standard deviation: $[0.229, 0.224, 0.225]$) are employed to standardize the values of pixels, thereby rendering them consistent with trained encoders.
- **Training Augmentations** (utilized probabilistically to render the dataset broader):
 - Horizontal/Vertical Flips ($p = 0.5$)
 - Rotation ($\pm 20-30^\circ$, $p = 0.5$)
 - Elastic Transform or Grid Distortion ($p = 0.3-0.5$)
 - Random Brightness/Contrast (± 0.2 , $p = 0.5$)
 - Hue/Saturation/Value Shifts ($\pm 10/20/10$, $p = 0.3-0.5$)
 - CLAHE (Contrast Limited Adaptive Histogram Equalization, clip_limit=2.0, $p = 0.3$)
- **Validation/Test Transforms:** resizing, padding, and normalization, without stochastic augmentations.

By employing strength masking for focusing loss calculation on regions of significance, this pipeline fixes class imbalance, assuring that models value particular regions throughout the background.



(a) Sample A



(b) Sample B



(c) Sample C

Figure 4.4: Processed Image Samples

Preprocessing for White Spot Dataset

Spot lesions in shrimp shells are small, colorful dots that might be mistaken for other diseases like bubble. Improved edge and precise improvements are the main goals of the training. Similar to black gill, the information is separated into train, val, and test sets, and an effective mask is created.

- **Resizing and Padding:** Zero-padding via uniform scaling to 1024×1024 .
- **Normalization:** The remaining ImageNet amounts from before.
- **Training Augmentations:**
 - Horizontal/Vertical Flips ($p = 0.5$)
 - Rotation ($\pm 15^\circ$, $p = 0.5$)
 - CLAHE (clip_limit=3.0, tile_grid_size= 8×8 , $p = 0.5$)
 - Grid Distortion ($p = 0.3$)
 - Random Brightness/Contrast (± 0.2 , $p = 0.5$)
- **Validation/Test Transforms:** Resizing, padding, and normalization only.

The augmentations are tested with great care so as to avoid the development of incorrect outcomes. CLAHE plays a significant role in enhancing spot accessibility, this is just one of the explanations that is performed.

4.3.3 Data Integration

Combining data from different sources, if applicable. Here is the revised Data Integration section. I have removed the "Balancing" section and focused entirely on the details of the dataset, the justification for adding it, and the critical methodology regarding the re-annotation and segmentation by your specific expert.

4.3.3.1 External Dataset Integration

We mixed the ShrimpDiseaseBD dataset [45] with the field data used in this research to make the prediction models more applicable and to ensure the shrimp were exposed to a greater variety of environmental variables.

4.3.3.2 Dataset Overview and Composition

The ShrimpDiseaseBD dataset [45] is a recently released, open-access image repository for disease detection in *Litopenaeus vannamei* (Pacific White Shrimp). The experimental data came from shrimp farms in the Satkhira and Bagerhat districts of Bangladesh, which have become recognized for their extensive aquaculture operations.

Bangladeshi districts that have become famous for their large aquaculture activities. High-resolution RGB images from the assortment demonstrate three major physiological states:

- **Healthy:** Specimens that are free of any noticeable visible lesions along with clear bodies and gill fibers.
- **White Spot Syndrome Virus (WSSV):** Specimens showing the distinctive clinical feature of rounded white calcium deposits (0.5–2.0 mm) situated inside the carapace membrane.
- **Black Gill Disease:** Samples that show different levels of pigmentation of the gill lamellae, a sign that the connective tissue has been harmed by fungi or pathogens, and develop a black or blackish colour.

The images show considerable variations in background complexity (dirt, water, containers), lighting conditions (daylight versus artificial flash), and figure position. This variation is essential for training deep learning models that are robust to situations in the real world.

4.3.3.3 Justification for Dataset Integration

The addition of this external dataset has been driven by the need to solve the "small data" problem that is frequently encountered in aquaculture research. The following tasks took place once ShrimpDiseaseBD [45] had been incorporated into our main dataset:

- **Feature Robustness:** The model cannot learn features that are specific to a specific farm or camera setup, as a consequence of the differences in the camera sensors and water quality are introduced by the external data.
- **Class Density:** Because of the brief duration of disease epidemics, smaller personal collections usually do not have enough specimens for the WSSV and Black Gill classes. Yet, this data set fixes this.

4.3.3.4 Data Harmonization and Preprocessing for Classification

To guarantee experimental accuracy and avoid neural networks from learning particular source imperfections, the external ShrimpDiseaseBD dataset [45] has been processed in the same way as the main field data.

Raw images from the external repository were not ingested directly; instead, they were passed through the unified image enhancement workflow defined in Section ??:

- **Background Suppression:** To remove various backgrounds that were unique to the external gathering sites, segmentation masks were created for every one of the 1,149 external images, utilising the Segment Anything Model (SAM).
- **Contrast Equalization:** To normalise the histogram distribution while making guarantees that the differences in lighting among the two data sources were as slight as feasible, the CLAHE technique (Clip Limit 1.2) was employed.
- **Resolution Standardization:** Using the same interpolated algorithms, every single picture was expanded to satisfy the various input specifications of the specific architectures (for example, so as to fit the intermediate storage that was 224×224 pixels).

The two data streams were successfully "harmonised" by the use of this identical transformation protocol. This assures that the combined training corpus is statistically uniform, which forces the deep learning models to focus solely on biological characteristics instead of environmental context or acquisition source.

4.3.3.5 Uniform Annotation and Segmentation Protocol

The inaccuracy that occurs when various experts label data based on somewhat variable diagnostic criteria is a key difficulty in integrating external datasets, known as "annotator bias". Furthermore, the raw ShrimpDiseaseBD [45] dataset contained image-level classification labels but lacked the pixel-level segmentation masks necessary for our methodology architectures.

To fix this, we established a consistent methodology for re-annotation:

- **Raw Data Extraction:** Only the raw, unlabelled images have been obtained from the source repository. The first labels provided by the dataset producers have been removed to reduce label noise.
- **In-House Segmentation:** The Regions of Interest (ROI) consisted of manual annotations that were performed by our study team.
 - For WSSV, every visible white spot was enclosed by pixel-precise binary masks.
 - Polygonal protections have been created for Black Gill that cover the infected gill regions.
- **Unified Expert Validation:** The same domain expert who verified our main field data also handled the annotation of this external dataset to ensure that the diagnosis was accurate. For our research, we depended on the expert's confirmation of each image to make sure that the segmentation boundaries were correct from the standpoint of biology and that the labelling had been determined on the clinical criteria used.

This comprehensive procedure ensured each of the training samples had the same ground truth standard, despite the fact that they came from different producers.

4.3.4 Final Datasets Overview

A dataset including 5,498 pictures representing three classes—Healthy ($n = 2,490$), White Spot Syndrome Virus (WSSV) ($n = 1,609$), and Black Gill (BG) ($n = 1,399$)—was derived from the experimental dataset through the combination of initially acquired data with information from external repositories. Throughout the phase of classification, the data was separated into three sets: the training set, the testing set, and the validation set. The ratio involved was 75-15-10%. Better the data led to a balanced training set with 2,000 pictures in each class, for a total of 6,000 images. The following fixed the problem of uneven classes. Evaluation was carried out on 823 independent test samples and 552 validation samples in order to ensure the robustness of the model. A segmentation-specific dataset was additionally generated for the sick classes at the same time. This dataset included 3,008 image-mask pairs that were labeled using specialist algorithms. We

divided this group into three parts: 80% for training, 10% for testing, and 10% for validation. Each component got 2,406 pairs for training and 301 pairs for testing and validation. This double-track data preparation approach makes sure that the classification models receive balanced class distributions along with the segmentation models get excellent, disease-specific spatial data for training purposes.

Augmentation Strategy to Balance the Classification Train Set Classes

- **Flipping:** Mirroring the image horizontally or vertically. Horizontal is standard; vertical is only used for data where "up" doesn't matter (e.g., satellite or microscopic images).
- **Brightness/Contrast:** Adjusting how light or "punchy" the image is.
- **Cropping & Scaling:** Taking a random sub-section of the image and resizing it back to original dimensions.

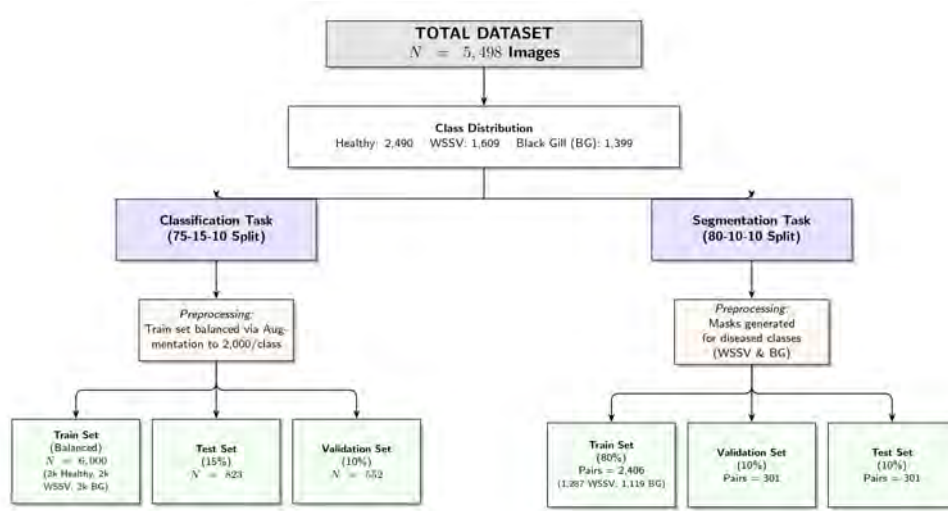


Figure 4.5: Final Dataset Structure.

4.4 Implementation of Selected Design

The present paper enables a good foundation in multi-stage deep learning on aquaculture; but, to construct an environment of diagnosis which will be both robust and independent, there are several significant directions to which we need to discover future studies. In addition to these recommendations, the model should be enhanced to be more resistive to the changes in the external world, it should be utilized on larger terms, and it should mix different types of data to detect the problems before they start manifesting.

Classification

The Unified Training Pipeline

We have constructed a very strict experimental framework so as to guarantee reproducibility and fair comparison. The pipeline manages the whole of the model lifecycle from data ingestion to inference.

A. Optimization and Loss Function We utilized the Adam optimizer with a conservative learning rate of $\alpha = 1 \times 10^{-4}$. This low learning rate is critical during fine-tuning to prevent large gradient updates from destroying the pre-learned weights of the backbone model (a phenomenon known as "catastrophic forgetting").

For the loss function, we employed Categorical Cross-Entropy with Label Smoothing ($\epsilon = 0.1$). Standard cross-entropy forces the model to predict the correct class with 100% probability (1.0) and others with 0%. This often leads to overconfidence and poor generalization. Label smoothing relaxes this target, asking the model to predict 0.9 for the correct class and spread the remaining 0.1 probability among the incorrect classes.

$$L_{\text{smooth}} = (1 - \epsilon) \cdot y_{\text{true}} + \frac{\epsilon}{K}$$

Where K is the number of classes (3). This encourages the model to learn clustered representations rather than just maximizing the margin for training examples.

B. Callbacks and Training Control To prevent computational waste and overfitting, three callbacks were utilized:

- **ModelCheckpoint:** Saves only the model weights that achieved the highest Validation Accuracy.
- **ReduceLROnPlateau:** Dynamic learning rate scheduling. If the validation loss stagnates for 3 epochs, the learning rate is halved. This allows the model to settle into finer local minima.
- **EarlyStopping:** Terminates training if validation loss does not improve for 10 consecutive epochs.

Table 4.1: Hyperparameter Configuration for the Unified Training Pipeline

Hyperparameter	Value / Setting	Description
Optimizer	Adam	Selected for efficient convergence on non-convex surfaces.
Learning Rate (α)	1×10^{-4} (0.0001)	Initial learning rate for fine-tuning.
Batch Size	32	Balanced to fit GPU memory while maintaining gradient stability.
Epochs	60	Maximum iterations with Early Stopping enabled.
Scheduler	ReduceLROnPlateau	Factor: 0.5, Patience: 3 epochs, Min LR: 1×10^{-6} .
Stopping Condition	EarlyStopping	Monitor: val_loss, Patience: 7
Loss Function	Categorical Crossentropy	Used with Label Smoothing ($\epsilon = 0.1$) for DenseNet, ResNet, ConvNeXt.
Input Size	224×224	Standard for DenseNet, ResNet, MobileNet, ConvNeXt, EfficientNet, Xception

The Unified Training Pipeline

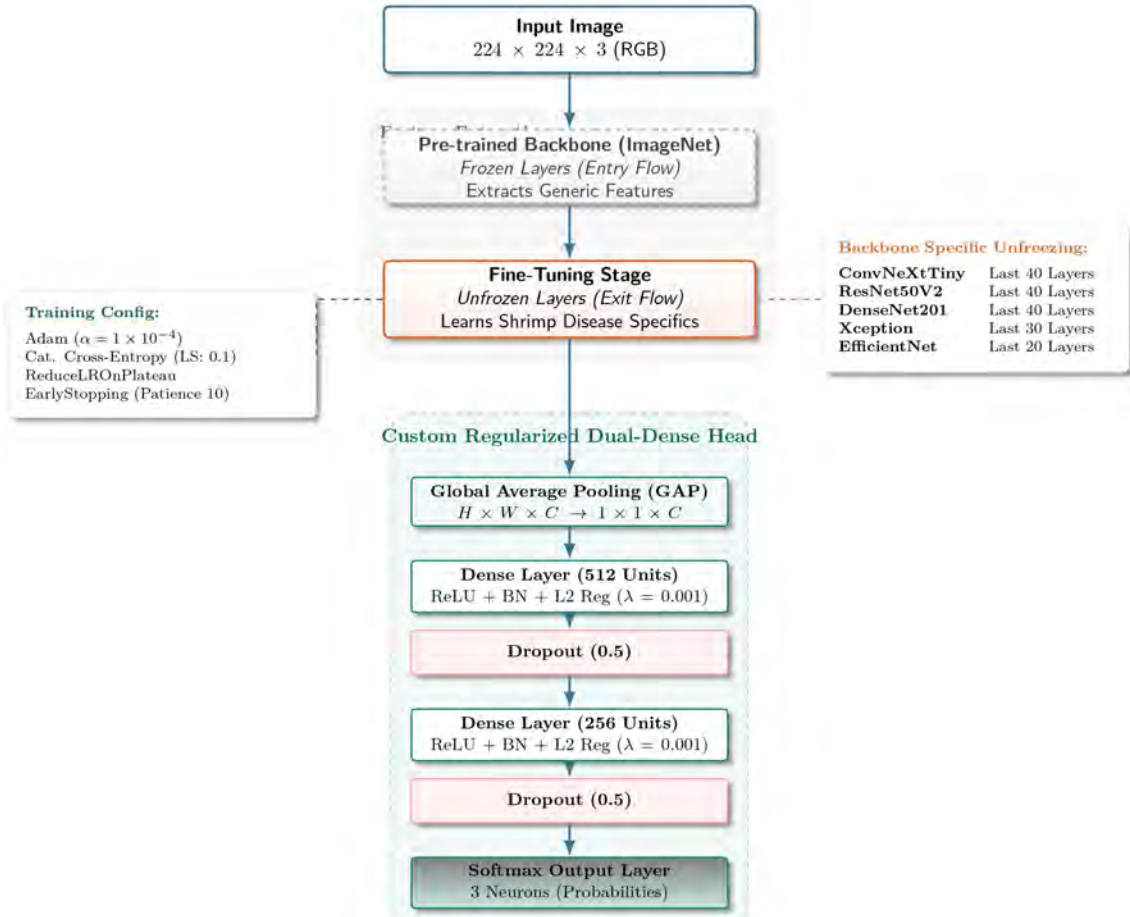


Figure 4.6: Overview of the unified model training pipeline.

Table 4.2: Transfer Learning Pipeline for CNN Models (Simplified)

Model	Fine-Tuning Strategy	Loss Function	Custom Classification Head Structure
Xception	Last 30 Layers Unfrozen	Cat. Crossentropy (LS: 0.1)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.5) Dense (256, L2=0.001, ReLU, BN) Dropout (0.5) Softmax (3 classes)
ConvNeXtTiny	Last 40 Layers Unfrozen	Cat. Crossentropy (LS: 0.1)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.5) Dense (256, L2=0.001, ReLU, BN) Dropout (0.5) Softmax (3 classes)
EfficientNetB0V2	Last 20 Layers Unfrozen	Cat. Crossentropy (LS: 0.1)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.5) Dense (256, L2=0.001, ReLU, BN) Dropout (0.5) Softmax (3 classes)
ResNet50V2	Last 40 Layers Unfrozen	Cat. Crossentropy (LS: 0.1)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.5) Dense (256, L2=0.001, ReLU, BN) Dropout (0.5) Softmax (3 classes)
DenseNet201	Last 40 Layers Unfrozen	Cat. Crossentropy (LS: 0.1)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.5) Dense (256, L2=0.001, ReLU, BN) Dropout (0.5) Softmax (3 classes)

1. Regularization and Classification Head Strategy

We made a strict, multi-stage regularization pipeline that reduces the risk of overfitting, which comes frequently when fine-tuning large, pre-trained models on specific datasets. This particular arrangement was selected after a lot of evaluation since it consistently provided the highest validation accuracy while maintaining the disparity between training and validation performance to an absolute minimum.

The classification head attached to all backbones consists of a standardized topology designed to enforce sparsity and robustness:

- **L2 Weight Decay** ($\lambda = 0.001$): This is applied to the kernels of all dense layers, penalizes significant weight adjustments, and prevents the model from

excessively depending on any singular feature.

- **Dropout (0.5):** A 50% loss rate was established. This leads the network to learn different representations of features, thereby making sure that predictions stay correct even if certain neurons are switched off.
- **Batch Normalization:** Added before functions of activation to maintain learning dynamics steady and speeding up convergence.
- **Label Smoothing (0.1):** To mitigate the model’s tendency for overconfidence in its predictions, known as ”calibration error,” we replaced conventional hard objectives (0 or 1) with smoothed soft targets. As a consequence, the model gets inspired to cluster classes closer together in the feature space, thereby improving generalization to unknown information.

2. Architecture-Specific Fine-Tuning Strategy

While the regularization pipeline was kept constant, the fine-tuning strategy (specifically the depth of unfrozen layers) was calibrated to the unique topological characteristics of each backbone architecture. A uniform unfreezing approach was deemed unsuitable given the significant variance in model depth and block structure.

- **EfficientNetV2B0 (20 Layers):** EfficientNet is sensitive to weight disturbance because of its extremely compact architecture that has been optimized via neural architecture search. In order to prevent the calamitous loss of its parameter-efficient representations, the fine-tuning was carefully limited to the final 20 layers, focusing exclusively on the top-most MBConv blocks.
- **Xception (30 Layers):** 30 layers were selected as the unfreezing level that encompasses the architecture’s ”Exit Flow.” The feature-extraction ”Entry Flow” was frozen to preserve low-level visual filters, while the depthwise separable convolutions responsible for high-level class discrimination were adaptable as a consequence of this approach.
- **ResNet50V2 (40 Layers):** ResNet used its residual skip connections, which enable gradients to move gradually over complex networks, to make way for a more aggressive approach. By unfreezing 40 layers, the model was capable of retraining the final two convolutional stages (Conv4 and Conv5), which allowed it to adapt both texture and shape-based features to the specific dataset.
- **ConvNeXtTiny (40 Layers):** The ”Tiny” name reflects the fact that ConvNeXt utilizes a gigantic (7×7) kernel-based macro-architecture similar to Vision Transformers. In order for the model to be able to alter its global receptive fields, it was necessary to set up a forty-layer unfreezing technique at the final processing stage.
- **DenseNet201 (40 Layers):** Because of its significant depth (201 layers) and feature-reuse mechanism, unfreezing 40 layers concentrated solely on the last ”Dense Block” and transition layer. Without disrupting the long-range connections which are important to the network’s feature hierarchy, this approach facilitated semantic adaptation at the decision boundary.

Segmentation

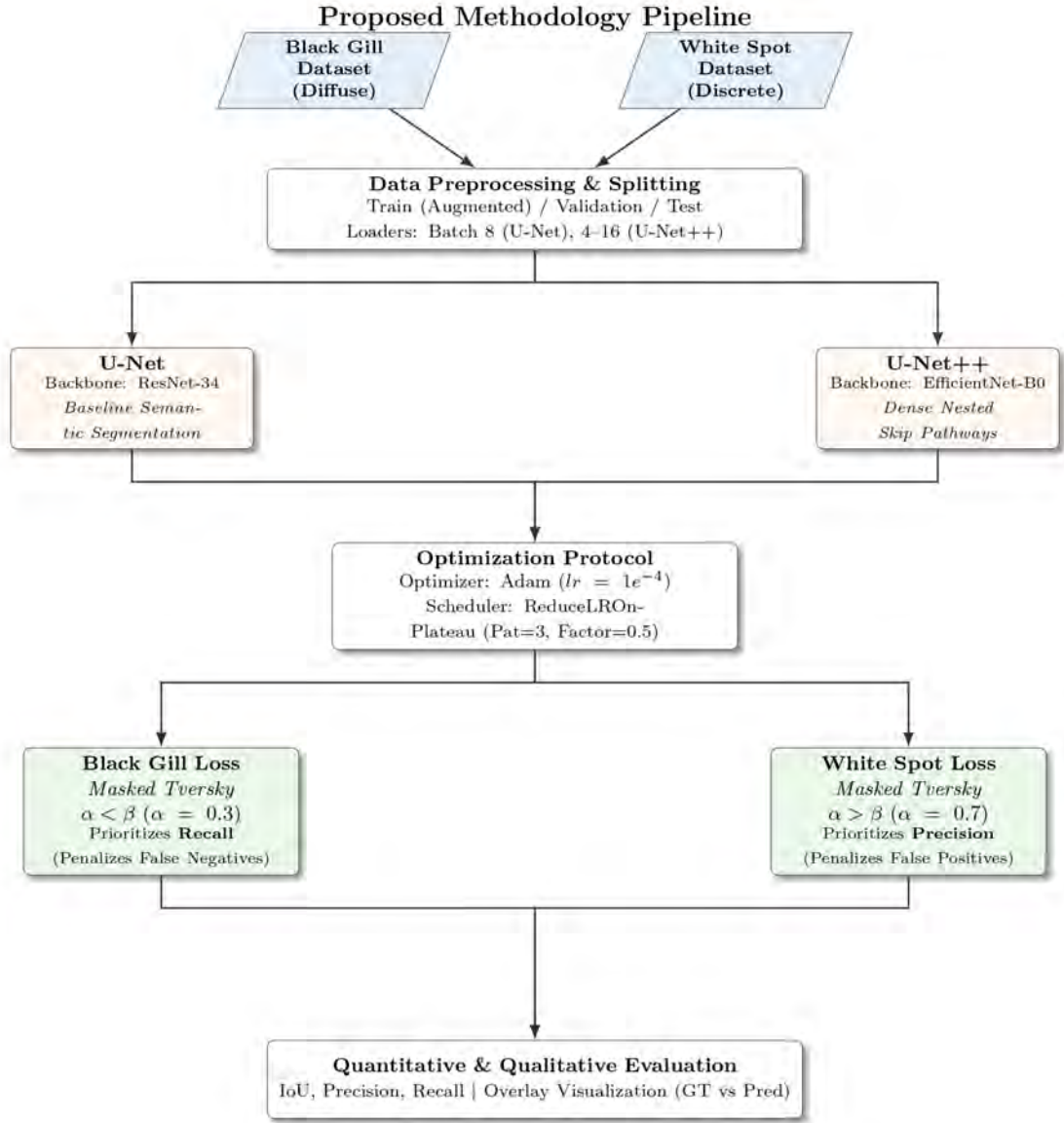


Figure 4.7: Visual representation of segmentation pipeline

Methodology for Black Gill Segmentation

Black gill detection involves three progressive models (U-Net, U-Net++) to compare semantic and instance segmentation efficacy on diffuse lesions.

- **Data Handling:** Dataset split into train (augmented), val, and test loaders (batch sizes: 8 for U-Net, 4 for U-Net++).
- **Model Training:**
 - Models: U-Net (ResNet-34), U-Net++ (EfficientNet-B0)
 - Optimizer: Adam ($lr=1e-4$) for U-Net/U-Net++;
 - Loss: Masked Tversky ($\alpha=0.3-0.5$, $\beta=0.5-0.7$) for U-Net/U-Net++.

- Scheduler: ReduceLROnPlateau (patience=3, factor=0.5) on val IoU for U-Net++.
- Epochs: 30–40; Early saving of best IoU model.
- **Evaluation:** Pixel-level IoU, precision, recall on val/test; Visual overlays (green GT, red pred) for 15 test images.

This methodology evolves from baseline semantic to advanced instance segmentation, focusing on recall for early disease detection.

Methodology for White Spot Segmentation

White spot detection, which assumes a similar experimental setup to black gill for comprehensive comparison, incorporates three models (U-Net, U-Net++) to evaluate semantic and instance segmentation on small, discrete lesions. The emphasis is on precision in order to avoid false positives from abnormalities.

- **Data Handling:** Dataset split into train (augmented), val, and test loaders (batch sizes: 8 for U-Net, 4 for U-Net++).
- **Model Training:**
 - Models: U-Net (ResNet-34), U-Net++ (EfficientNet-B0).
 - Optimizer: Adam (lr=1e−4) for U-Net/U-Net++.
 - Loss: Masked Tversky ($\alpha=0.7$, $\beta=0.3$) for U-Net/U-Net++ to penalize false positives heavily.
 - Scheduler: ReduceLROnPlateau (patience=3, factor=0.5) on val IoU for all models.
 - Epochs: 30–40; Best model saved by IoU.
- **Evaluation:** IoU, precision, recall; Visuals with 0.5 threshold for binary masks and overlays (green GT, red pred) for test images.
- **Special Adaptations:** Logging to CSV is employed for epoch-wise metrics, and NaN loss checks are carried out when needed.

It uses instance segmentation to adjust for potential overlap in aquaculture monitoring and adjusts loss weights for precision, comparable to the black gill approach.

Chapter 5

Result Analysis

5.1 Performance Evaluation

5.1.1 Performance Evaluation Metrics and Techniques for Classification

The efficacy of the proposed classification model is evaluated using conventional metrics along with techniques to improve comprehension and durability. The following few parts clarify the main evaluation criteria, what they are for, and how to use metrics to figure them out.

Standard Classification Metrics

Accuracy Determines the percentage of right predictions, both positive and negative, out of every potential case. Despite its straightforwardness and simplicity, it may be deceiving in indicating the existence of class imbalance [4], [24].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5.1)$$

Precision signifies the percentage of projected positive cases that are actually positive. In situations where false positives have a large cost, it is of greatest significance [4], [24].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5.2)$$

Recall (also known as Sensitivity or True Positive Rate) evaluates the ratio of accurately detected favorable instances by the model. It is essential when missing positive situations produces serious consequences [4], [24].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.3)$$

F1 Score gives a single balanced score by harmonically averaging recall and precision; this can be especially useful when utilizing datasets that are not equally distributed [4], [24].

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.4)$$

Confusion Matrix A contingency table that encapsulates the prediction outcomes by presenting the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It acts as a basis for the computation of the majority of performance metrics [4], [24].

Advanced Performance and Interpretability Measures

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) tests how well the model identifies the difference between classes at all classification tiers. The AUC value ranges from 0 to 1; higher values reflect higher discriminative ability [4], [24].

As an indicator of both True Positive Rate (Recall) and False Positive Rate, the ROC curve generates a graph. AUC is the probability that a randomly selected positive example is ranked higher than a randomly picked negative example.

Grad-CAM (Gradient-weighted Class Activation Mapping) interpretability method makes class-discriminative localization heatmaps via the gradients of the target class inserted into the final convolutional layer. This helps understand CNN decisions without modifying the core neural network framework [14].

Regularization and Robustness Techniques

Mixup is a simple yet effective data augmentation and regularization approach that generates more training instances by linearly interpolating pairings of samples and labels. This facilitates the development of more fluid decision boundaries and enhances generalization [17].

Given input samples x_i, x_j and corresponding labels y_i, y_j , with mixing coefficient $\lambda \sim \text{Beta}(\alpha, \alpha)$:

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda) x_j, \\ \tilde{y} &= \lambda y_i + (1 - \lambda) y_j.\end{aligned}\tag{5.5}$$

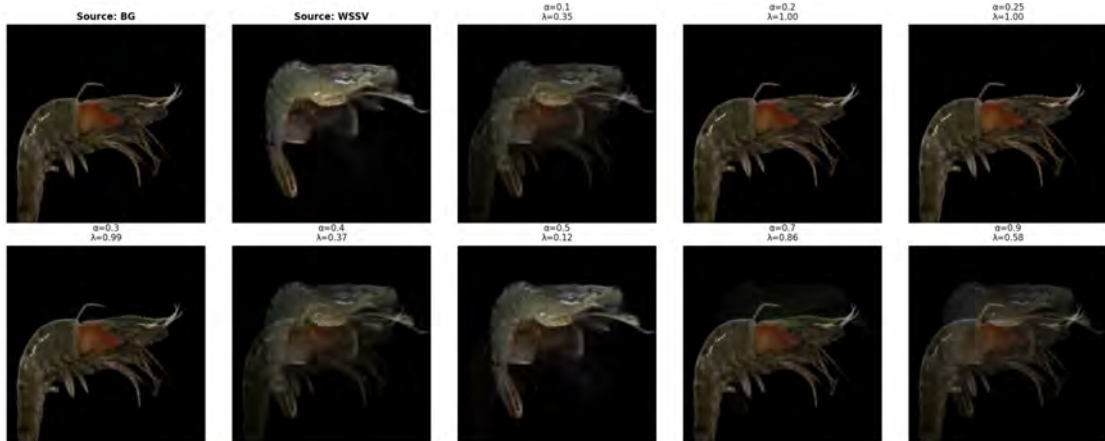


Figure 5.1: Mixup image for different values

CutMix is an advanced regularization technique that proportionally mixes the ground-truth labels according to the patch area, cutting a patch from one image

and putting it onto a different one. This improves classification performance and localization capacities [20].

Given a binary mask $\mathbf{M}$ with area ratio λ :

$$\begin{aligned}\tilde{x} &= \mathbf{M} \odot x_i + (\mathbf{1} - \mathbf{M}) \odot x_j, \\ \tilde{y} &= \lambda y_i + (1 - \lambda) y_j.\end{aligned}\tag{5.6}$$



Figure 5.2: CutMix image for different values

Fast Gradient Sign Method (FGSM) For Adversarial Learning

Adversarial training utilizes a one-step adversarial attack approach. Adversarial instances are generated by changing the input in the direction of the loss gradient's sign, which are afterward utilized to improve model stability [8].

$$x' = x + \epsilon \cdot \text{sign}(\nabla_x J(\theta, x, y))\tag{5.7}$$

where $J(\theta, x, y)$ is the loss function, θ are the model parameters, and ϵ controls the perturbation magnitude.



Figure 5.3: Examples of adversarial perturbations - 1

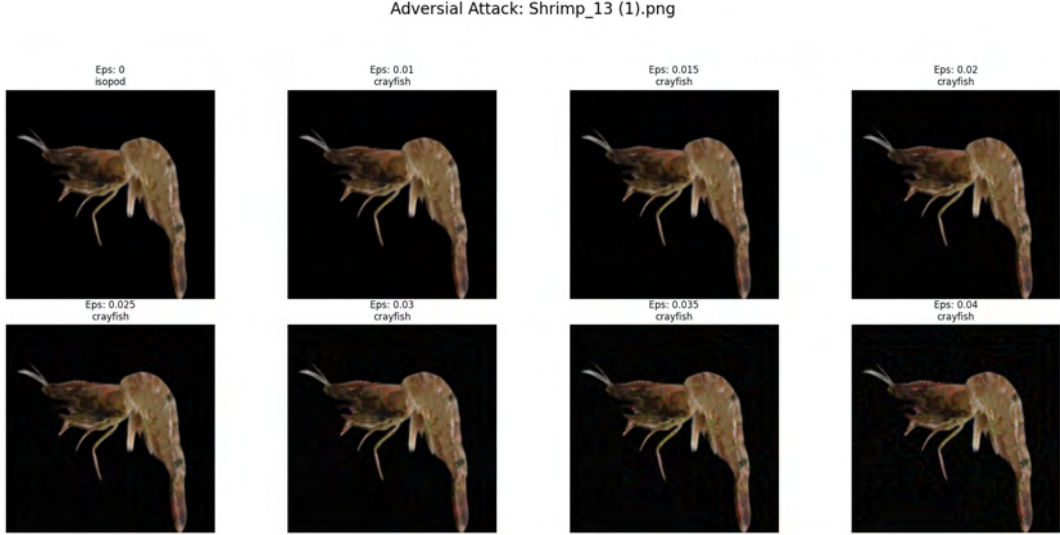


Figure 5.4: Examples of adversarial perturbations - 2

Projected Gradient Descent (PGD) Adversarial Training utilizes an iterative multi-step adversarial assault approach, which provides more accurate estimates than one-step methods like FGSM. Continually taking small steps toward the direction of the loss gradient’s sign, with projections to stay within the disturbance bound, creates adversarial cases. These are used to make the model less vulnerable to attacks.

The update rule for generating the adversarial example is:

$$x^{t+1} = \Pi_{x+\mathcal{B}_\epsilon} (x^t + \alpha \cdot \text{sign}(\nabla_x J(\theta, x^t, y))) \quad (5.8)$$

where $J(\theta, x, y)$ is the loss function, θ are the model parameters, ϵ defines the maximum perturbation (e.g., ℓ_∞ -norm ball $\mathcal{B}_\epsilon$), α is the step size, t is the iteration, and Π is the projection operator to clip values back into the allowed range (often including normalization to $[0,1]$ for images). This process begins from $x^0 = x + \eta$ (random perturbation η within the ball) and runs for a fixed number of steps (e.g., 10–40 iterations) to find a strong adversary.

5.1.2 Performance Evaluation Metrics and Techniques for Segmentation

There is an established procedure to evaluate segmentation models for both black gill and white spot diseases in shrimp images. Models are evaluated on the way they line up, overlap, and balance anticipated and ground truth masks. Dependable performance in aquaculture applications along with tackling issues like class imbalance—where diseased areas tend to be small compared to backgrounds—are both made feasible by these measurements. By combining Intersection over Union (IoU), accuracy, recall, and Dice loss (or coefficient), we succeeded to complete image segmentation tasks that are frequently employed in the medical and semantic domains. Each of these indicators provide multiple types of data. IoU and Dice focus on how much information they share, while accuracy and recall focus on the types of errors

they make as well. Computations were performed at the pixel level employing utilities from the `segmentation_models_pytorch` library, and predictions were binarized at reliability standards like 0.5. Masks were semantically aggregated in instance-based models (for example, Mask R-CNN). Every statistic will be explained here, including with its formulas and relevance to both the circumstances.

IoU, that is also known as the Jaccard index, quantifies the overlap between predicted and actual segmented regions, rendering it an accurate approach to evaluate boundary accuracy in asymmetrical lesions such as black gill or discrete patches in white spot disease. It's defined as the ratio of the overlap to the union of the sets, and both excessive and insufficient segmentation is punished in a comparable way. Precision is the proportion of correctly identified positive pixels among all predicted positives. It is crucial to decrease the amount of false alarms in white spot identification, which is difficult because bubbles can be mistaken for lesions. Recall, or sensitiveness, quantifies the fraction of actual diseased pixels identified. This is crucial for black gill since missing diffuse areas can delay management. Commonly used in benchmarks, dice loss, which is derived from the Sørensen-Dice coefficient, strikes a balance between recall and precision, rendering it sensitive to small regions that include white spot lesions .

The table below provides a summary of the metrics which can be employed for both black gill and white spot segmentation tests. Let TP indicate true positives, FP indicate false positives, FN signify false negatives, and TN represent true negatives.

IoU was the primary criterion for model selection because of its holistic nature, and these metrics were collected over validation and test sets . To guarantee numerical stability in lesion-free instances, a modest epsilon ($1e-7$) was added to the denominators. Metric evolution graphs along with additional visualizations enabled us recognize the compromises which made certain that the models performed well with real-life shrimp images.

Table 5.1: Evaluation Metrics for Segmentation

Metric	Formula	Relevance to Black Gill and White Spot
IoU	$IoU = \frac{TP}{TP+FP+FN}$	Assesses overall overlap; robust for diffuse (black gill) and small (white spot) regions
Precision	$P = \frac{TP}{TP+FP}$	Reduces false positives; key for white spot to avoid artifact confusion.
Recall	$R = \frac{TP}{TP+FN}$	Ensures detection completeness; critical for early black gill identification.
Dice Loss	$D = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \text{ Loss} = 1 - D$	Balances precision/recall; sensitive to small lesions in both diseases.

Intensity Level Classification

Intensity classification extends segmentation outputs to quantify disease severity, providing practical metrics for aquaculture decision-making. For both black gill and white spot, intensity is computed as the percentage of affected shrimp tissue, normalized against an estimated body area to account for image variations. This post-processing uses predicted masks and simple thresholding for body approximation, enabling categorical staging. While sharing a similar framework, thresholds and interpretations differ by disease: black gill emphasizes progressive necrosis stages, white spot on rapid lesion spread. Formulas and tables below detail the processes, with limitations noted for thesis rigor.

Intensity Level Classification for White Spot Disease

For white spot disease (WSSV), intensity represents infection severity as the ratio of infected to total shrimp tissue area, highlighting early-stage dots that can proliferate quickly. This aids in monitoring outbreaks, where even low percentages signal urgency. Severity is calculated post-segmentation, integrating with models for end-to-end analysis. The approach assumes dark backgrounds for tissue isolation, yielding actionable categories from healthy to severe.

The severity percentage is:

$$\text{Severity (\%)} = \frac{\text{Infected Area (predicted mask pixels)}}{\text{Total Shrimp Tissue Area}} \times 100, \quad (5.9)$$

with tissue area estimated via grayscale thresholding (> 20 intensity as tissue). Infected pixels are those with model confidence > 0.5 . Categories use empirical thresholds for WSSV progression.

Severity % (Infected Area)	Intensity Level	Meaning
$\leq 0.1\%$	Healthy	No visible WSSV
$0.1 - 2\%$	Mild	Early-stage infection
$2 - 10\%$	Moderate	Visible progression
$> 10\%$	Severe	Critical infection

Table 5.2: Classification of WSSV Intensity Levels

Limitations include threshold sensitivity to lighting, potentially over/underestimating tissue; future work could incorporate adaptive methods.

Intensity Level Detection for Black Gill Disease

The intensity of black gill assesses the geographical spread of gill necrosis as a proportion of the body, allowing progressive evaluations for diseases connected to environmental stress. This normalization additionally takes into account the different sizes of shrimp while concentrating on how diffuse injury gets worse. Whenever it is implemented after prediction, it utilizes a lower threshold for body detection, which makes it suitable for images which are focused on the gills.

The intensity percentage is:

$$\text{Disease Intensity (\%)} = \frac{\text{Predicted Diseased Pixels}}{\text{Estimated Shrimp Body Pixels}} \times 100, \quad (5.10)$$

with body pixels thresholded at grayscale > 15 . Diseased pixels exceed 0.5 confidence. Stages reflect necrosis advancement.

Infected Body Area (%)	Intensity / Stage	Clinical Meaning
$\leq 0.05\%$	Healthy	No visible disease
$0.05 - 5\%$	Stage 1 (Mild)	Localized lesions
$5 - 10\%$	Stage 2 (Moderate)	Progressive gill damage
$> 10\%$	Stage 3 (Severe)	Extensive necrosis

Table 5.3: Classification of Black Gill Intensity Levels

Fixed thresholding may vary with brightness; changes that involve dedicated body segmentation must be made for greater accuracy.

5.2 Analysis of Design Solutions

A dual-stage approach is necessary for the detection of shrimp diseases. At first, a robust classification of the sample’s overall health state is necessary, followed by a high-precision segmentation to quantify the severity of the infection. We selected the following designs as they each have particular computational advantages to deal with the unique textural and spatial characteristics of shrimp disease.

5.2.1 Classification Solutions

We evaluated five architectures to determine the optimal balance between diagnostic accuracy and computational efficiency.

- **ResNet50-v2 and DenseNet201 (Gradient & Feature Stability):** ResNet50-v2 employs skip connections to address the problem of the vanishing gradient. With "residual" learning, the network may concentrate on the small, high-frequency variations of white lesions on the shrimp’s shell, which are crucial to identify WSSV.

DenseNet201 improves feature reuse by connecting each layer to all subsequent layers. This "dense connectivity" is surprisingly effective in identifying the presence of Blackgill and White Spot, as it permits the model to protect subtle color gradients and necrotic patterns that could be hidden in more sparse designs.

- **Xception and EfficientNetB0-V2 (Efficiency & Pattern Separation):** Xception utilizes convolutions that are depth-separable rather than conventional convolutions. It is proficient at the isolation of the distinct "dot" geometry of Whitespot lesions from the detailed textures of crustacean pleopods by distinguishing spatial and channel-wise correlations.

EfficientNetB0-V2 employs compound scaling to improve depth, width, and resolution. This model is small and light without losing accuracy, thereby rendering it an excellent choice for real-time tracking of vehicles on farms.

- **ConvNeXt.tiny (Modern CNN Evolution):** Using design concepts from Vision Transformers (ViTs), such as larger kernel sizes (7×7) and LayerNorm, ConvNeXt.tiny re-engineers the traditional CNN. This creates a "global" receptive field, allowing the network to understand the spatial distribution of spots across the entire shrimp body, instead of merely localized regions.

5.2.2 Segmentation Solutions

Segmentation allows us to isolate and measure the specific area of infection.

- **U-Net:** U-Net has the position of medical segmentation benchmark due to its symmetric encoder-decoder design. Having a high degree of pixel accuracy, it shows the expansive and irregular limits of the Blackgill and Whitespot discoloration.
- **U-Net++:** U-Net++ (Nested U-Net) has more advantages than the initial version because it adds nested and dense skip paths. These routes fill in the "semantic gap" between the encoder and decoder's feature maps. That renders it simpler for the model to tell the difference between light-colored natural shell reflections and true early-stage Whitespot lesions, which drops the number of false positives.

5.3 Final Design Adjustments

5.3.1 Classification

Comparative Analysis of Deep Learning Architectures and Generalization Performance

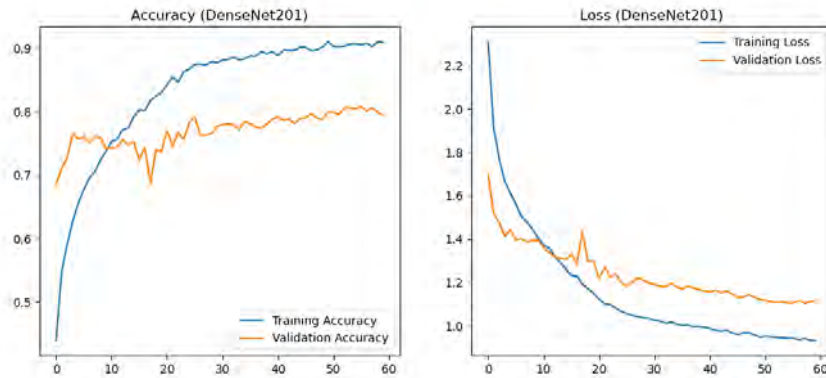


Figure 5.5: Training curve for DenseNet201.

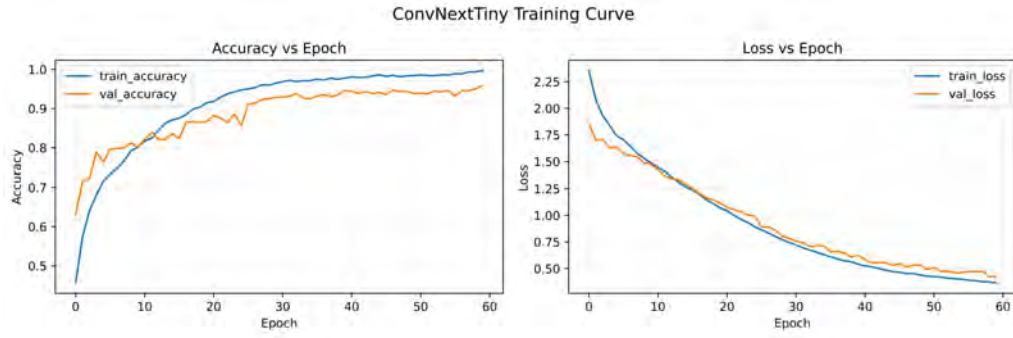


Figure 5.6: Training curve for ConvNeXt-Tiny.

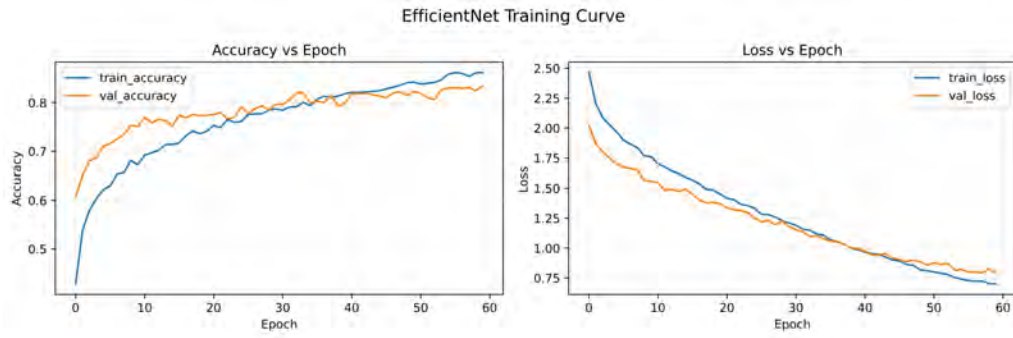


Figure 5.7: Training curve for EfficientNet.

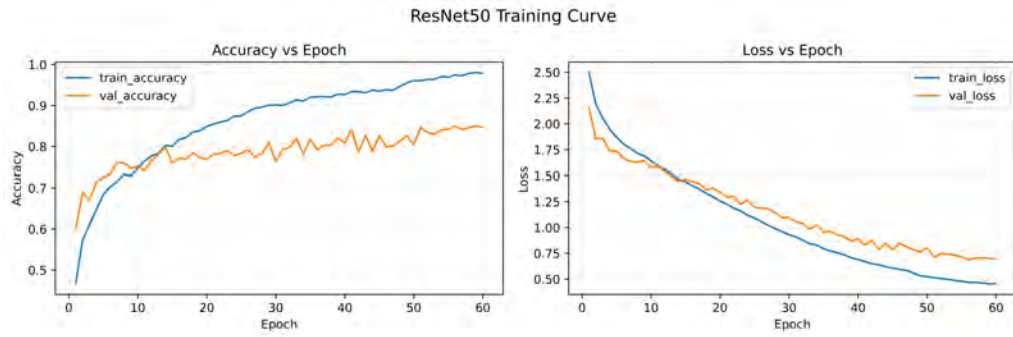


Figure 5.8: Training curve for ResNet50.

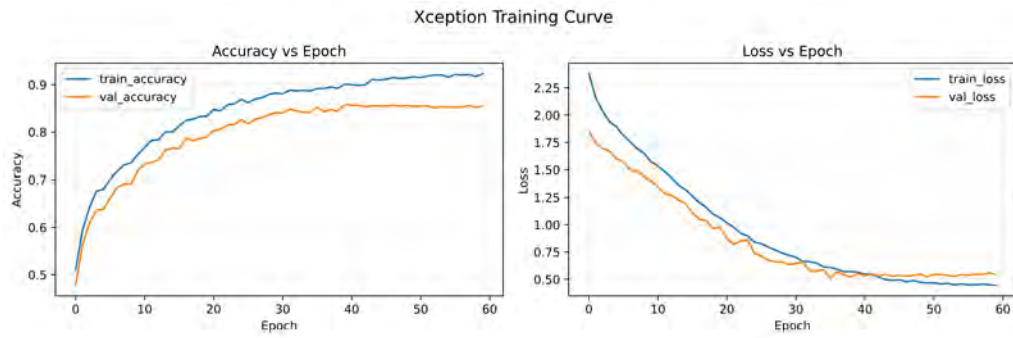


Figure 5.9: Training curve for Xception.

Empirical Review of Classification Architectures

The empirical review of the five deep learning architectures—ConvNeXt-Tiny, EfficientNet, Xception, ResNet50, and DenseNet201—provides important insights into the correlation between model complexity and generalization. An ideal fit and robust feature extraction was demonstrated by the tightly coupled training and validation curves that ConvNeXt-Tiny preserved throughout the 60-epoch duration, and it was found that it reached a peak validation accuracy close to 95% while maintaining a negligible generalization gap. In a comparable way, EfficientNet showed great stability and excellent convergence characteristics. Yet, its ultimate accuracy was significantly more moderate, indicating that it valued a steady decrease in loss higher than rapid fitting characteristics.

A significant pattern of overfitting was noted in the Xception, ResNet50, and DenseNet201 models. These architectures performed a good job minimizing training loss; however, their validation metrics frequently diverged, indicating that they were unable to generalize to new data. ResNet50 showed high-frequency oscillations (jitter) in its validation loss, while Xception and ResNet50 both showed a plateau in validation accuracy despite ongoing improvements in training accuracy. The greatest drop in performance occurred in DenseNet201, which had an ordinary overfitting pattern: as training accuracy climbed over 90%, validation accuracy declined to near 80%, and after epoch 20, validation loss started to rise significantly. This variation demonstrates how these three models' strict connectivity and high range of parameters induced them to memorize random noise instead of learning transferable latent representations for the particular dataset. The "Overfitting Trio" requires more regularization, such as greater dropout, weight decay, or advanced information augmentation, with the goal to address the substantial gap that exists in their training performance and validation effectiveness. Accordingly, ConvNeXt-Tiny is the most achievable option for deployment; however, overfitting is a challenge that requires to be resolved.

Hypermeter Tuning for 3 Models

The Xception, ResNet50V2, and DenseNet201 designs have been shown to be highly vulnerable to overfitting in the comparative analysis of the initial training results. Consequently, a more rigorous regularization framework is required to guarantee the generalizability of the model. Utilizing a dynamic, augmentation-heavy method for the prior static transfer learning method enhanced the experimental pipeline and decreased the apparent difference between training and validation measurements. At the core of this enhancement was an on-the-fly stochastic augmentation suite that flipped the image horizontally and vertically, randomly rotated and zoomed it by 30%, and adjusted the brightness and contrast along with various geometric and photonic modifications. These changes were introduced at the starting point of each epoch to stop the model from settling on stable spatial patterns; this extended the training set's "effective" size and encouraged the deeper layers to extract features that were more robust to transformations.

Table 5.4: Hyper-parameters Tuning for 3 Classification Models

Model	Loss Function	Augmentation Strategy	Custom Classification Head Structure
Xception	Categorical Crossentropy (LS: 0.1)	Flip H/V, Rot/Zoom (30%), Contr/Bright (20%/10%)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.6) Dense (256, L2=0.001, ReLU, BN) Dropout (0.6) Softmax (3 classes)
ResNet50V2	Categorical Crossentropy (LS: 0.1)	Flip H/V, Rot (30%), Zoom (30%), Contrast (20%), Brightness (10%)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.6) Dense (256, L2=0.001, ReLU, BN) Dropout (0.6) Softmax (3 classes)
DenseNet201	Categorical Crossentropy (LS: 0.1)	Flip H/V, Rot (30%), Zoom (30%), Contrast (20%), Brightness (10%)	Dense (512, L2=0.001, ReLU, BN) Dropout (0.6) Dense (256, L2=0.001, ReLU, BN) Dropout (0.6) Softmax (3 classes)

Post-Hyperparameter Tuning: Comparative Analysis of Generalization Performance:

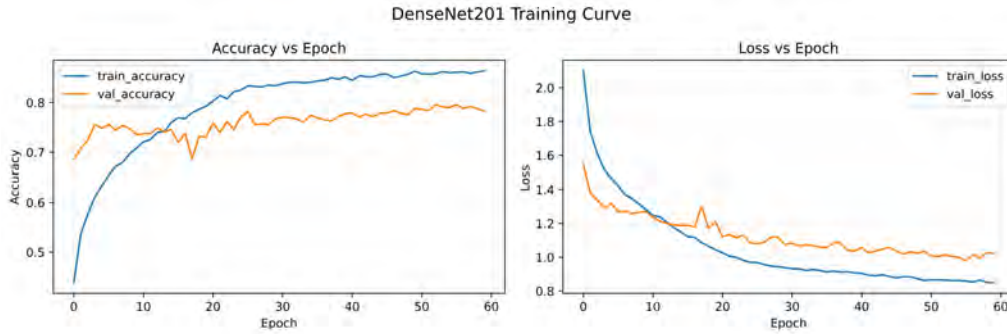


Figure 5.10: Training Loss Over Epochs for DenseNet201.

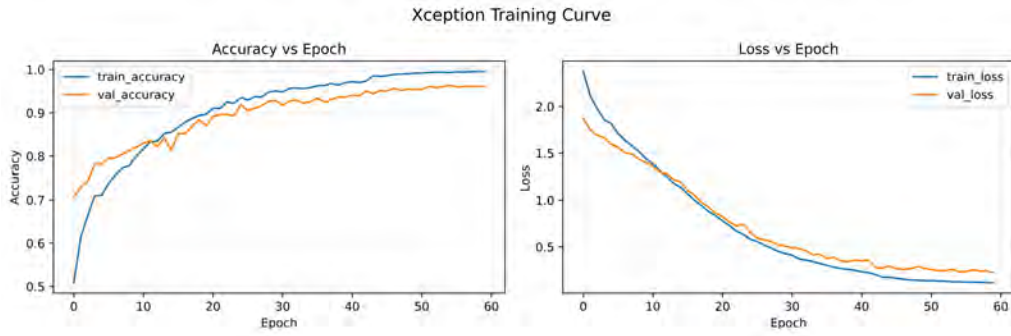


Figure 5.11: Training Loss Over Epochs for Xception.

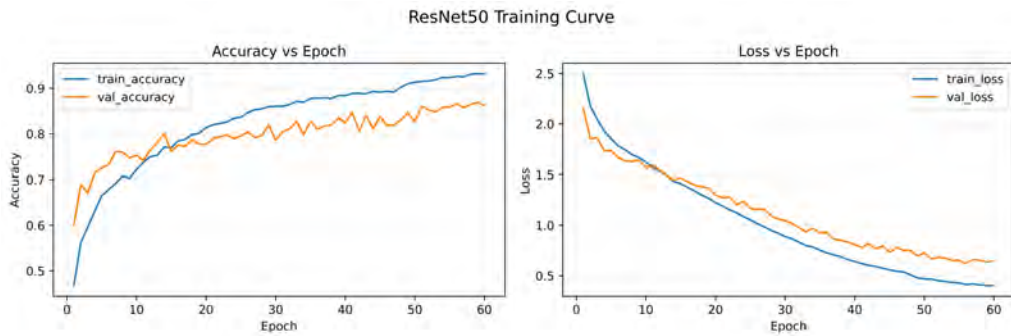


Figure 5.12: Training Loss Over Epochs for ResNet50 V2.

The stability and convergence profiles of the Xception, ResNet50V2, and DenseNet201 models got better when a new regularization strategy was implemented. This strategy featured aggressive on-the-fly data augmentation and more dropouts. In particular, the Xception architecture got its initial and final trajectories to line up nearly perfectly, which fixed its previous susceptibility to overfit and led it to a high-accuracy equilibrium at the 60-epoch point in time. The validation loss trajectory of ResNet50V2 has stabilized significantly when compared to the initial trials, notwithstanding the fact that a gap between its training and validation accuracy still remains. The geometric and photometric augmentations have largely erased out the previously seen erratic "jitter."

The results indicate that DenseNet201 is the most challenging model to regularize within this set of models, as it continues in showing minor overfitting. Compared to the initial training phase, the enhancement is notable because the validation accuracy curve now shows a more steady upward trend, as compared to the volatile stagnation that was noticed previously, and the substantial divergence in validation loss has been decreased. While a generalization gap is still present, as evidenced by the training accuracy exceeding the validation accuracy by approximately 5-8%, the improved strategy efficiently mitigated the calamitous overfitting that characterized the baseline model. It also suggests that the model's predictive utility has been maintained by the integration of 0.6 Dropout and diversified augmentation, ignoring the case that the high-density connectivity of DenseNet201 inherently encourages over-parameterization for this particular dataset.

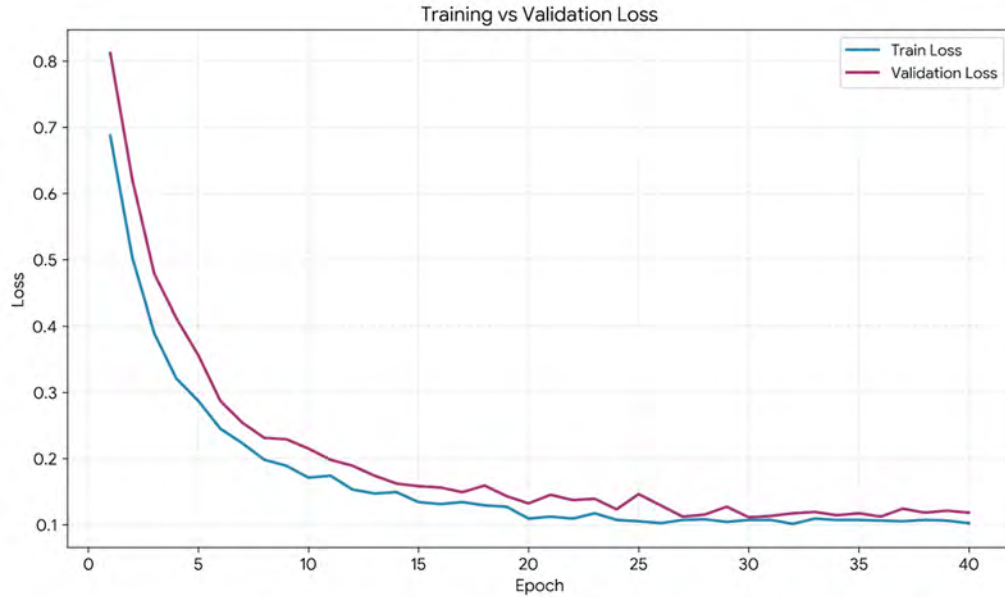
5.3.2 Segmentation

5.3.2.1 White Spot

5.3.2.1.1 U net++



(a) Performance Metrics Over Epochs

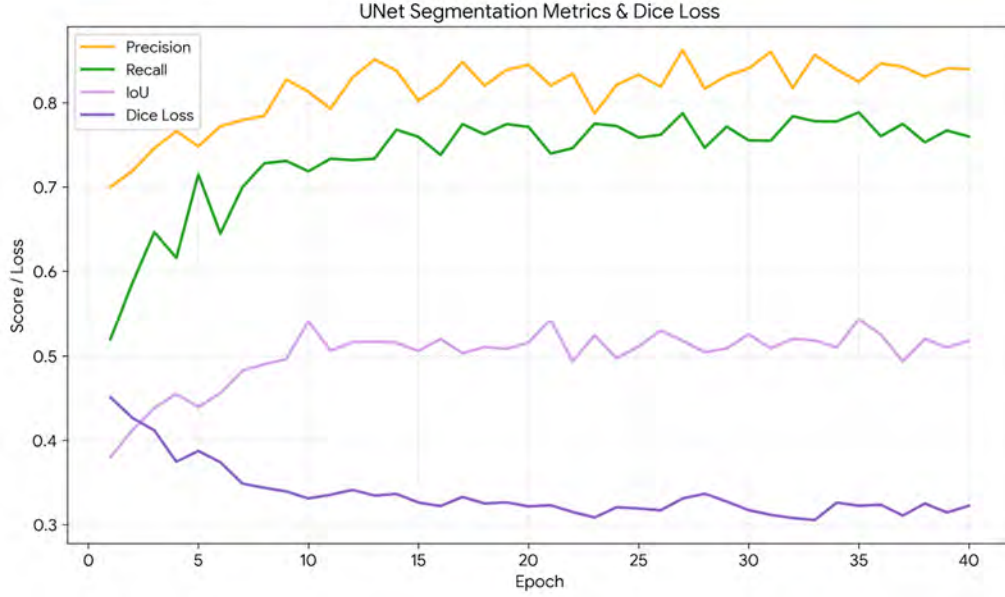


(b) Performance Metrics

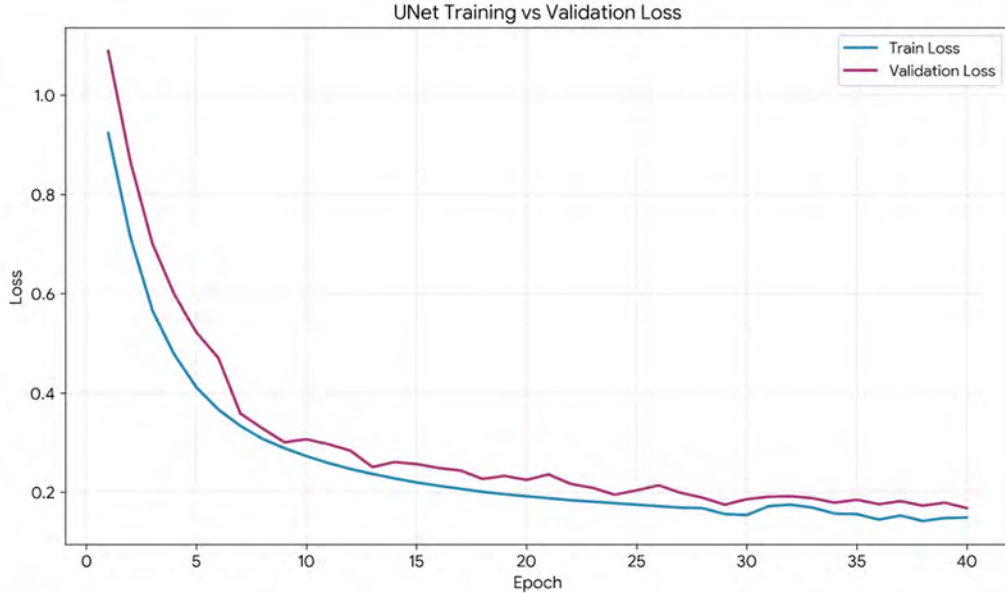
U-net++'s segmentation model is extremely precise as well as capable of generalization, as shown by its results. In the beginning, both the training and validation loss curves decrease rapidly. After a brief but normal change, they return to normal, which indicates the model is learning well despite getting too perfect. The segmentation indicators of performance back this claim: accuracy stays high (about 0.72–0.75), which indicates that the model could effectively choose target regions with few false positives. Recall carefully fluctuates between 0.70 and 0.73, which indicates it is exceptionally effective at identifying relevant patterns. Following the

inherent challenge of pixel-level segmentation tasks, the IoU values of 0.52 to 0.56 demonstrate both the predicted and ground-truth masks have a significant overlap. Over the years, dice loss drops to a low level (about 0.29-0.31), which results in enhanced accuracy and stability. An accurate reliable model match for precise semantic segmentation that meets the prerequisites for scientific accuracy as well as evaluation has been demonstrated by the continual closing pattern, balanced precision-recall dispute and equivalent dice loss.

5.3.2.1.1 U net



(a) Performance Metrics Over Epochs



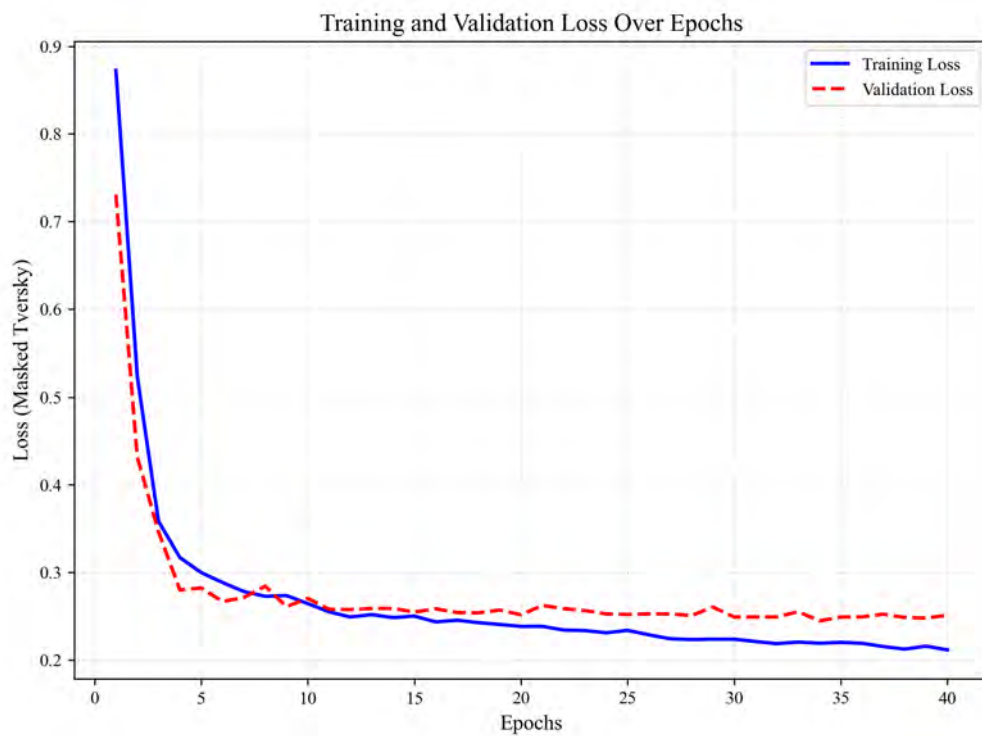
(b) Performance Metrics

After an ongoing decrease in validation and training losses, the U-Net architecture that was suggested showed good convergence and generalisation. It is eventually rounded out, which shows it. In a lack of any major overfitting to the training

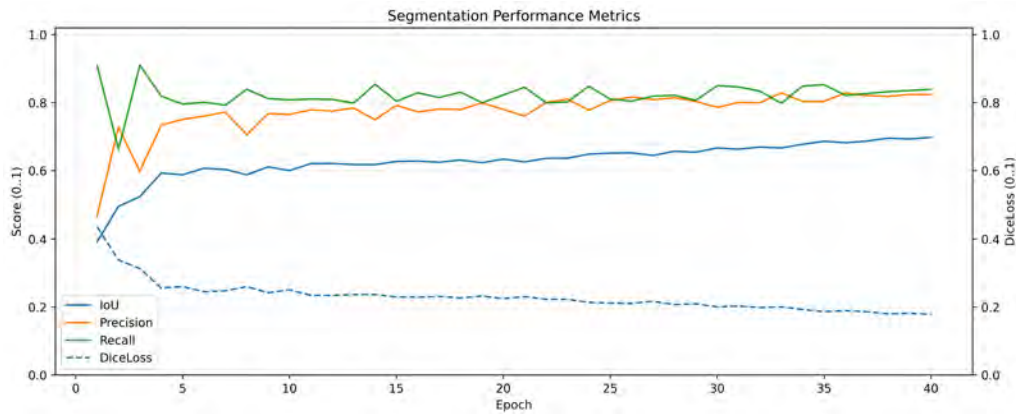
data, the validation loss maintained at roughly 0.18. According to performance evaluations, positive predictions are pretty accurate, with Precision achieving its highest point at approximately 0.85 and Recall achieving its maximum point at 0.77. The intersection over Union (IoU) score of 0.52 of the framework indicates it is fairly precise in the setting of limits. Yet, the Dice Loss' progressive decrease to 0.32 demonstrates that the model works well in improving the overall overlapping of areas. These numerical findings indicate the model's outstanding ability to preserve constant learning dynamics while decreasing false positives, which underlines its entire segmentation performance.

5.3.2.2 Black gill

5.3.2.2.1 U net ++



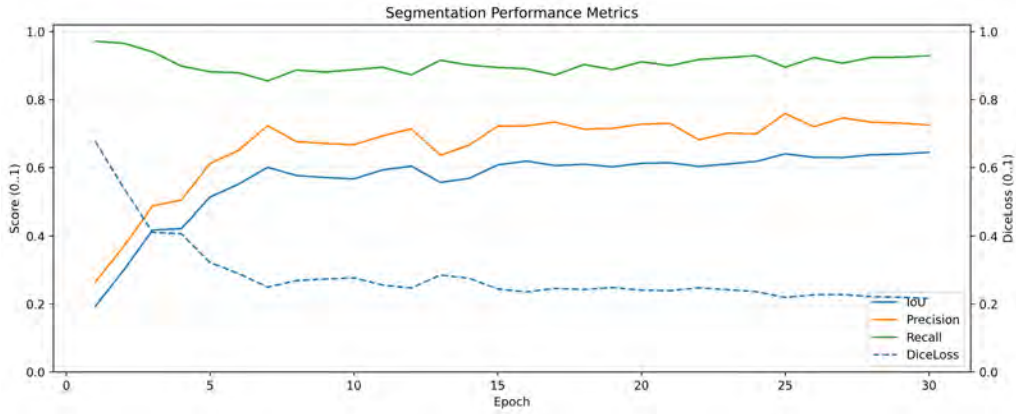
(a) Training and Validation Loss Over Epochs



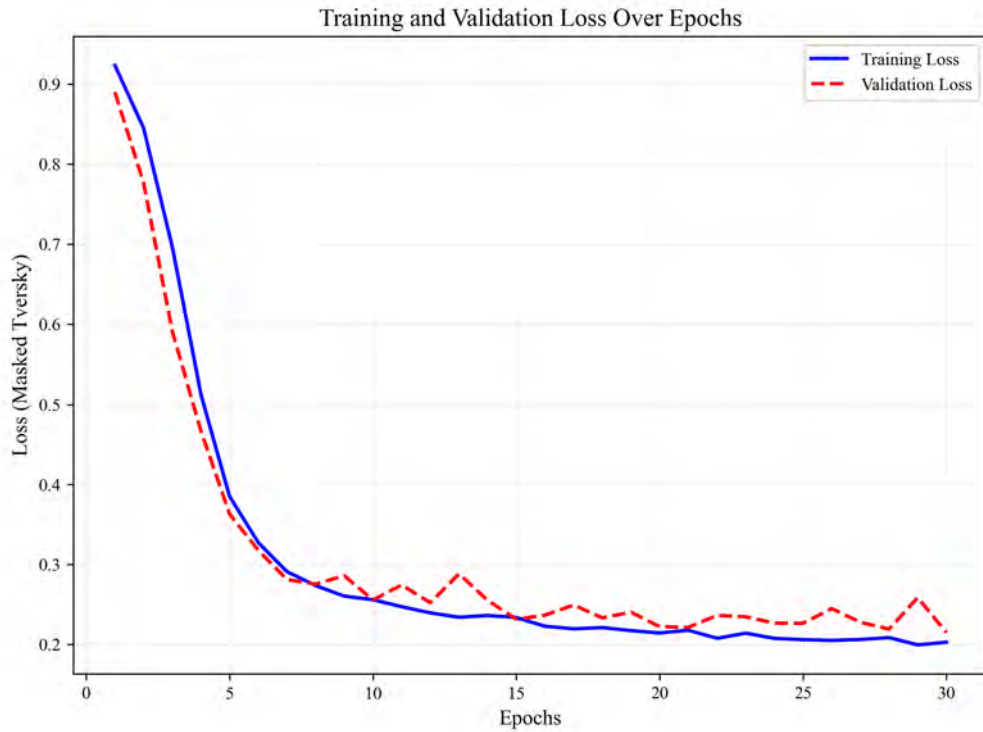
(b) Performance Metrics

The training segmentation model works quite well, showing high accuracy and consistent convergence. Due to this, it can accurately identify features in the environment while decreasing the amount of false positives. Overall the training and validation datasets of the Masked Tversky Loss showed an ongoing declining slope throughout the training period. Due to this, the problems that have been brought about by the disparity of categories were successfully solved and it was guaranteed that the model wasn't biased towards the previous class that was the majority in any significant way. The Idea was successful in capturing the inner features and physical boundaries of the shrimps, as indicated by the high levels of the Mean Intersection over Union (mIoU) and Dice Coefficient. Additionally, recognizing that the training and validation metrics are so closely related confirms the model's strong generalization abilities; it indicates that the model acquired fundamental trends rather than simply recalling the training distribution, resulting in it perfect for making reliable projections with unseen data.

5.3.2.2.2 U net



(a) Training and Validation Loss Over Epochs



(b) Performance Metrics

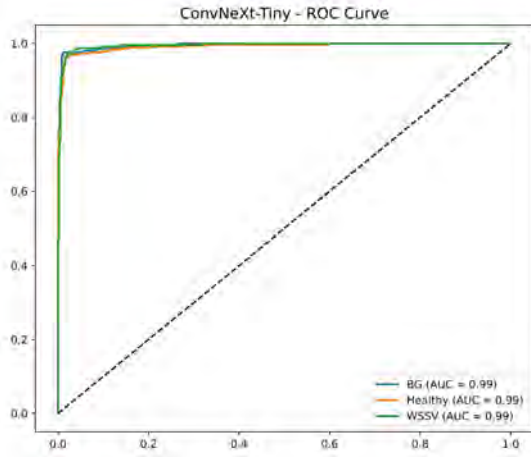
The model obtains successful convergence with an ending Masked Tversky Loss of around 0.20, showing an established and predictable learning pathway. The training and validation arcs completely line with each other, providing indicative both of the lack of overfitting and the outstanding generalization of the data. With a Recall that keeps near 0.93, the model certainly places an excessive amount of emphasis on high sensitivity. Precision and IoU, on the contrary, achieve plateaus of 0.75 and 0.65 in regards to how accurately they segment, correspondingly. The model's metric distribution shows high efficacy in choosing target places, but it could generate too wide boundary estimations. Efficiency comes near to a reliable constant state by epoch 20, leading to a high recall benchmark that is effective for tasks in which every aspect of a characteristic is of greater significance.

5.4 Statistical Analysis

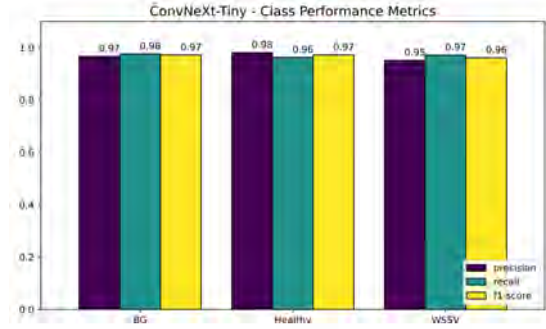
5.4.1 Classification

5.4.1.1 ConvNextTiny

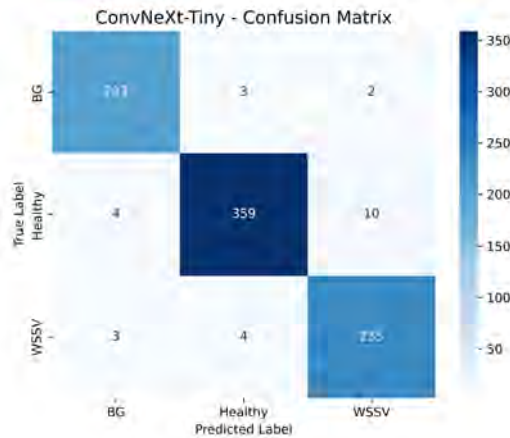
ConvNeXt-Tiny architecture works very well for detecting shrimp health problems. The model achieved an impressive overall accuracy of 96.8% after testing just 823 samples. The average score ranged from 86.10% to a maximum of 91.86%. This means that most of the classifications were completed with high accuracy. Healthy, White Spot Syndrome Virus (WSSV) and Blackgill (BG) are the three groups of shrimp that we tested. And the accuracy of the results is relatively divided. With a score of 0.98, the “Healthy” group, which had 373 samples, was the most accurate. Among them, the “BG” class was able to detect 0.97 samples, which shows that they were very accurate in their detection. The “WSSV” disease group (242 samples) achieved a very good F1-score of 0.96 and a recall of 0.97, which increased its accuracy. This allows us to easily find sick shrimps. The macro and weighted means both equate 0.97, indicating that the model’s progress is relatively stable. Even though class support is shifting, this indicates the mechanism is still fair. The confusion matrix and ROC curve lines show that ConvNeXt-Tiny rarely gets it wrong, which makes it a nearly perfect diagnostic tool. It has the ability to correctly recognize objects and apply its understanding to an assortment of environmental and disease conditions. Additionally, it is a diagnostic solution that can identify issues with high precision and function in a range of environments to reduce false positives while maintaining the process of learning ongoing. As a result, we can easily find sick shrimp. Moreover, both the macro and weighted mean are equal to 0.97, indicating that the model’s performance is relatively stable. Although the class support is changing, which indicates that the process is still fair. The confusion matrix and ROC curve line show that ConvNeXt-Tiny rarely makes mistakes, making convNeXt-Tiny an almost perfect diagnostic tool. It has the ability to accurately recognize objects and apply its understanding to a wide variety of environmental and disease conditions. Ultimately, it is a diagnostic solution that can detect problems with high accuracy and reduce false positives while keeping the learning process running by working in different environments.



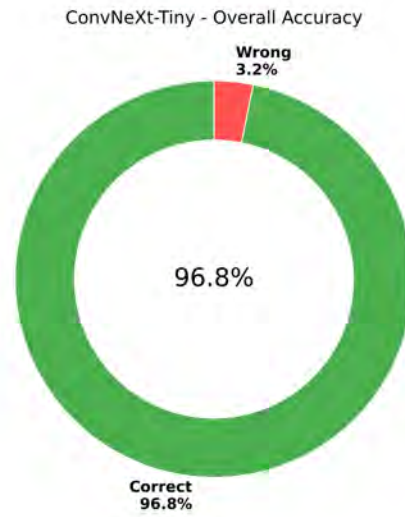
(a) ROC Curve



(b) Metrics Bar Chart



(c) Confusion Matrix



(d) Accuracy Donut

Figure 5.17: Performance Visualization for ConvNeXt-Tiny Model.

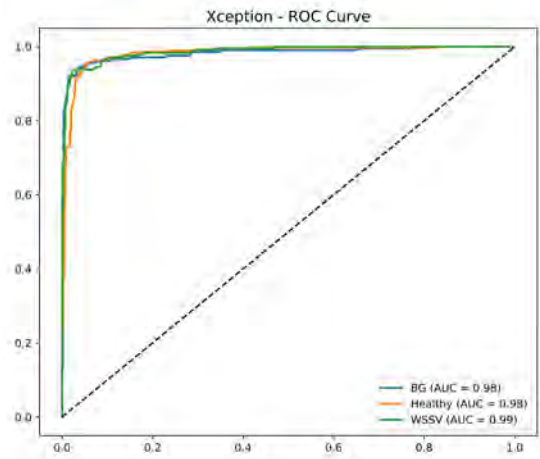
Table 5.5: Classification Performance Metrics for ConvNeXt-Tiny Model

Class	Precision	Recall	F1-Score	Support
BG	0.97	0.98	0.97	208
Healthy	0.98	0.96	0.97	373
WSSV	0.95	0.97	0.96	242
Accuracy		0.97		823
Macro Avg	0.97	0.97	0.97	823
Weighted Avg	0.97	0.97	0.97	823

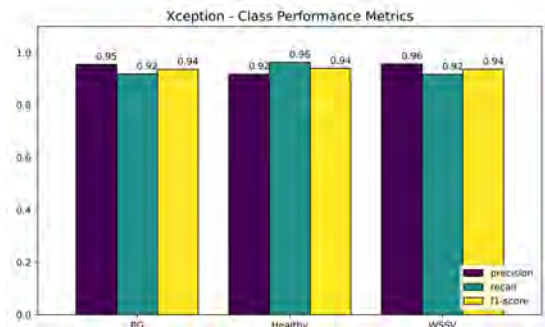
5.4.1.2 Xception

The xception model showed an overall accuracy score of 87.7 percent on just 823 different samples, suggesting that it is relatively good at making accurate predic-

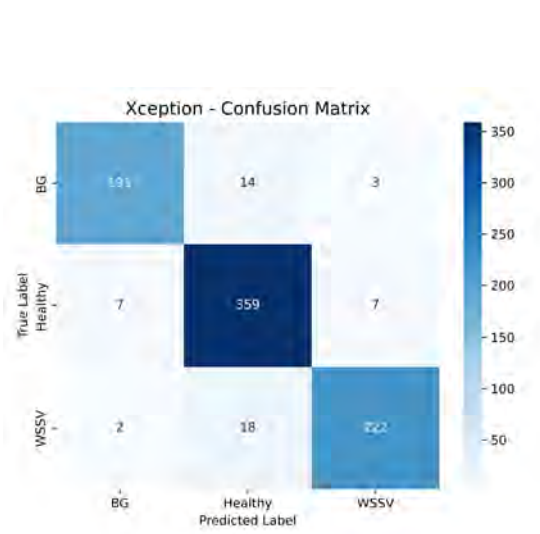
tions. Also, a closer look shows that the results vary significantly among different factors. And the recall percentages for the "healthy" class and the "blackgill" class are 91% and 88%, respectively. On the other hand, there is much concern about the "WSSV" disease class, as it showed an accuracy score of only 83%. And the model incorrectly classified 36 diseased "WSSV" samples as "healthy", which is a very dangerous form of false negative. It was discovered in the course of a massive investigation of the false outcomes, which indicated that there was a safety risk. The statistics indicate a potential remedy to the situation and indicate that the system is aware of its own uncertainty based on a statistical method. Though the model makes these misguided classifications, its overall confidence rating goes down to an optimal level of 95.7 percent on the correct classification, to approximately 80 percent. Certain groups of errors are not accidental. A total of 34 distinct images of shrimps contribute to the majority of the changes. The gap in performance is mostly explained by an issue with specific photos, rather than the fact that the model in its entirety is bad. It is not the factor of the mistakes of the model, but the misunderstanding of its essential issues.



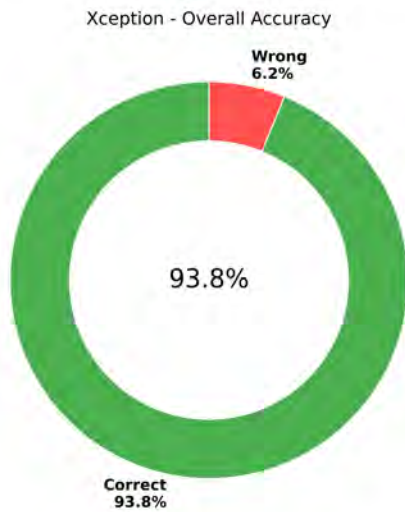
(a) ROC Curve



(b) Metrics Bar Chart



(c) Confusion Matrix



(d) Accuracy Donut

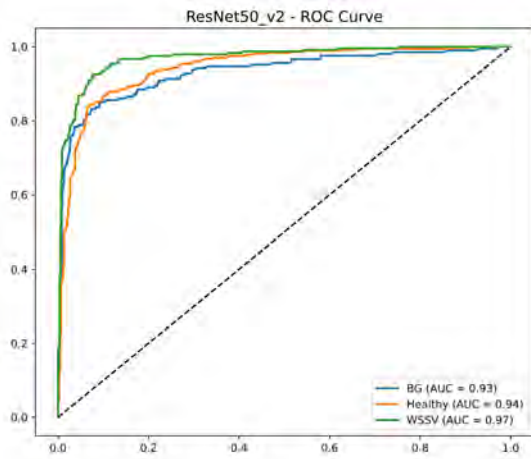
Figure 5.18: Performance Visualization for Xception Model.

Table 5.6: Classification Performance Metrics for Xception Model

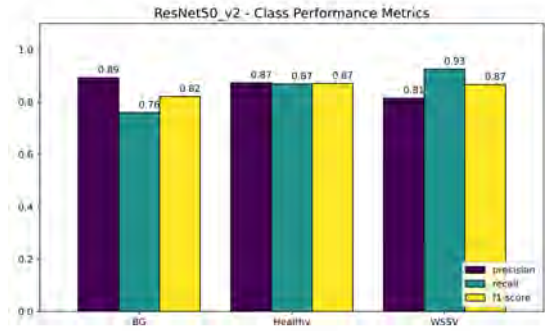
Class	Precision	Recall	F1-Score	Support
BG	0.95	0.92	0.94	208
Healthy	0.92	0.96	0.94	373
WSSV	0.96	0.92	0.94	242
Accuracy		0.94		823
Macro Avg	0.94	0.93	0.94	823
Weighted Avg	0.94	0.94	0.94	823

5.4.1.3 ResNet50 V2

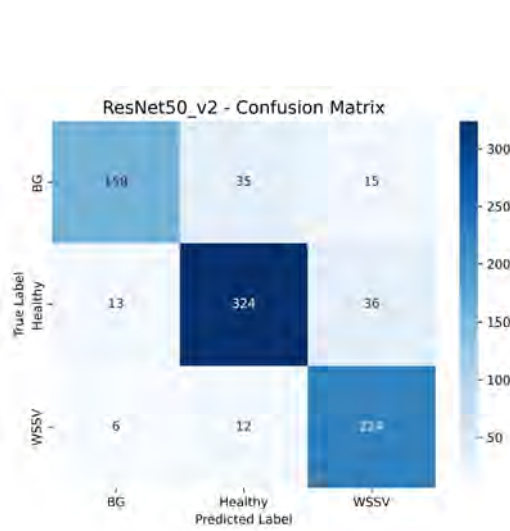
The ResNet50 v2 architecture was analyzed with the help of the dataset which included 823 photos. The total accuracy rate was 86 percent, 683 of the total were correct on identifying them and 140 were incorrect. It was demonstrated by conducting a thorough analysis of the incorrect allocation of errors that each of the three groups possesses a characteristic collection of qualities of performance. This method identified 215 of the 242 cases of infection in the right way with the best performance in the hazardous classification of the WSSV. Nevertheless, it did not detect 27 instances, and falsely classified 20 and 7 Healthy and BlackGill respectively. It predicted a 7% error rate. On the other hand, the model accurately categorized only 149 of the 208 samples in the BlackGill (BG) group. It improperly designated 41 BlackGill pics as Healthy and 18 as WSSV, suggesting that only 76% of the samples seemed preserved. A balanced profile has been maintained by the Healthy category, comprising a large amount of data (373), having 319 accurate predictions, with 39 healthy samples being incorrectly classified as WSSV, resulting in an increase in false alarm rate. In the end, while the system is effective at identifying illness, the fact that the BlackGill and Healthy classes keep getting mixed up shows that more work needs to be done to separate biological problems from nonbiological ones in order to make the system accurate.



(a) ROC Curve



(b) Metrics Bar Chart



(c) Confusion Matrix



(d) Accuracy Donut

Figure 5.19: Performance Visualization for ResNet50V2 Model.

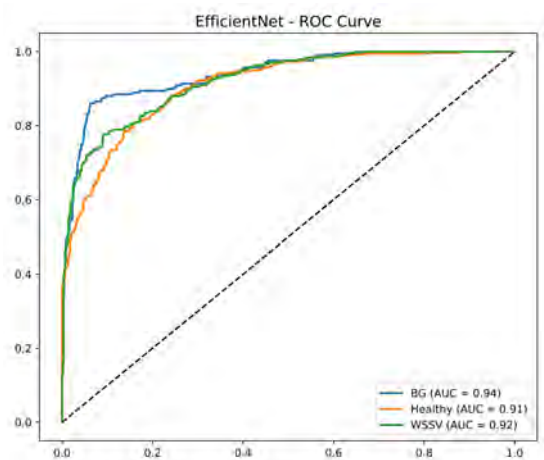
Table 5.7: Classification Performance Metrics for ResNet50_v2 Model

Class	Precision	Recall	F1-Score	Support
BG	0.89	0.76	0.82	208
Healthy	0.87	0.87	0.87	373
WSSV	0.81	0.93	0.87	242
Accuracy		0.86		823
Macro Avg	0.86	0.85	0.85	823
Weighted Avg	0.86	0.86	0.86	823

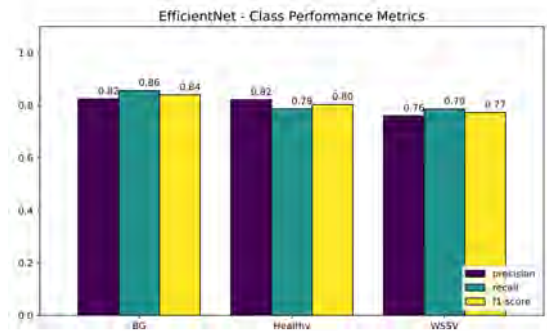
5.4.1.4 EfficientNet

The EfficientNet model obtained an accuracy of 80%, suggesting a high level of trust in anticipating shrimp health conditions. Following an extensive review of the

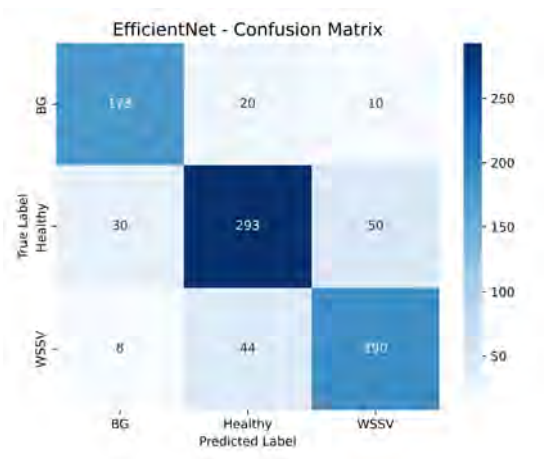
model, it turned out that it functioned exceptionally well in detecting the BlackGill category. The model got the best F1-score of 0.84 and a high recall rate of 0.86. This indicates that the model is useful in reducing false negatives on the part of the BlackGill class. There was a precision of 0.82 in the healthy group. But the recall rate decreased a little to 0.79, meaning that the healthy model was not as biased in its classification. Its findings did not correspond to the healthy group. The categorization of the White Spot Syndrome Virus (WSSV) was the most complicated part of the procedure. The lowest accuracy was of 0.79 in the model depicted a recall. With an F1 of 0.77, meaning that it identified more false positives than any of the other models, it identified 0.76 as more positive. The average correlation of 0.80 demonstrates the overall that the result is justifiable. The statistics however, clearly demonstrate that improvements are required in the detection of WSSV signals of positive or background variations in order to make the model more accurate.



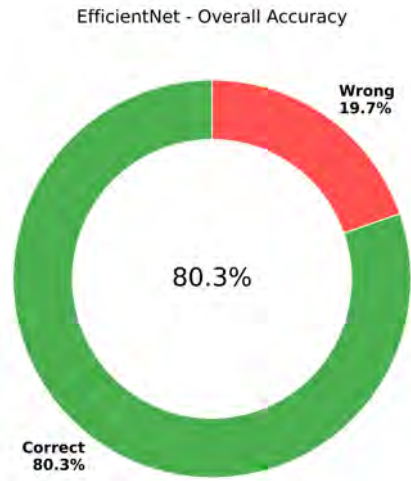
(a) ROC Curve



(b) Metrics Bar Chart



(c) Confusion Matrix



(d) Accuracy Donut

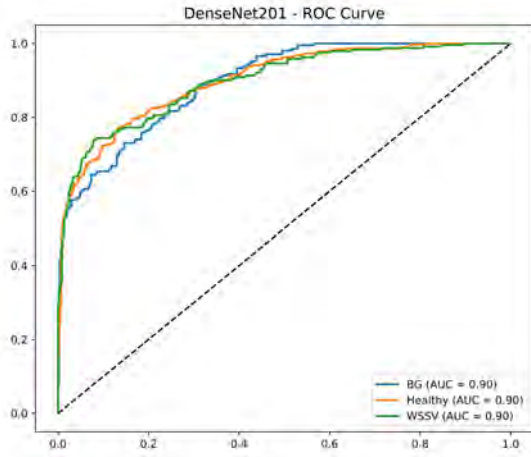
Figure 5.20: Performance Visualization for EfficientNet Model.

Table 5.8: Classification Performance Metrics for EfficientNet Model

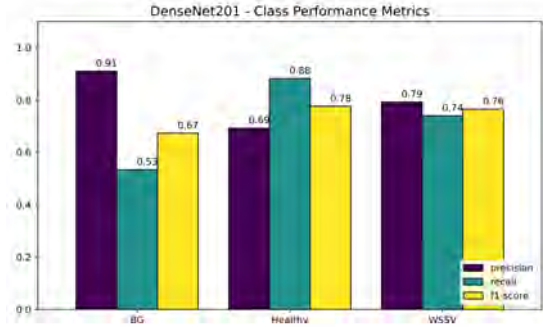
Class	Precision	Recall	F1-Score	Support
BG	0.82	0.86	0.84	208
Healthy	0.82	0.79	0.80	373
WSSV	0.76	0.79	0.77	242
Accuracy		0.80		823
Macro Avg	0.80	0.81	0.80	823
Weighted Avg	0.80	0.80	0.80	823

5.4.1.5 DenseNet201

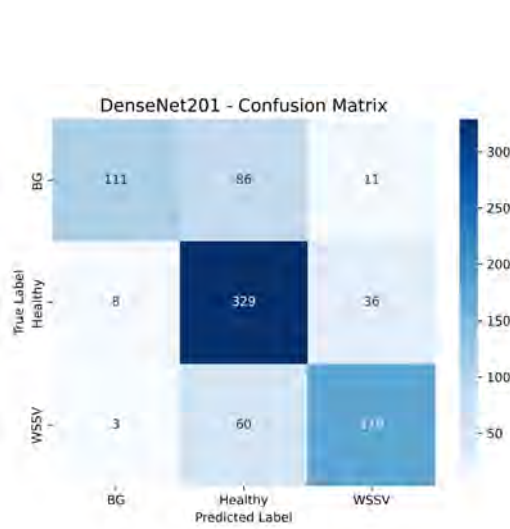
The DenseNet201 model showed great accuracy and sensitivity in shrimp health classification. The model achieved 75% accuracy after testing with 823 samples. The model showed high accuracy for detecting black gill (BG) disease, with 0.91, which shows a very low precision, but a high recall. In contrast, the healthy group had a mean of 88%, but the mean was recorded as 0.69, indicating that a large proportion of clinical samples were incorrectly classified as healthy. The WSSV format also performed well, with an accuracy of 0.79 and a recall rate of 0.74, which resulted in an average F1 score of 0.76. While the high prediction accuracy rate - often above 90% - indicates that the model is reliable to correctly identify the class, the assignment of a good BG classifier often highlights the need for major improvements in the field to prevent the spread of serious diseases in farms.



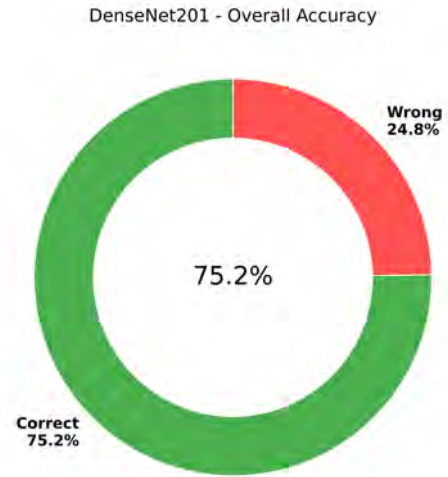
(a) ROC Curve



(b) Metrics Bar Chart



(c) Confusion Matrix



(d) Accuracy Donut

Figure 5.21: Performance Visualization for DenseNet201 Model.

Table 5.9: Classification Performance Metrics for DenseNet201 Model

Class	Precision	Recall	F1-Score	Support
BG	0.91	0.53	0.67	208
Healthy	0.69	0.88	0.78	373
WSSV	0.79	0.74	0.76	242
Accuracy		0.75		823
Macro Avg	0.80	0.72	0.74	823
Weighted Avg	0.78	0.75	0.75	823

5.4.2 Segmentation

5.4.2.1 Whitespot

5.4.2.1.1 U-Net ++

Table 5.10: Final Segmentation Performance Metrics (U-Net/U-Net++ Results)

Metric	Value
IoU (Intersection over Union)	0.54
Precision	0.72
Recall	0.70
Dice (F1)	0.71
Dice Loss	0.29
Train Loss	0.10
Validation Loss	0.11

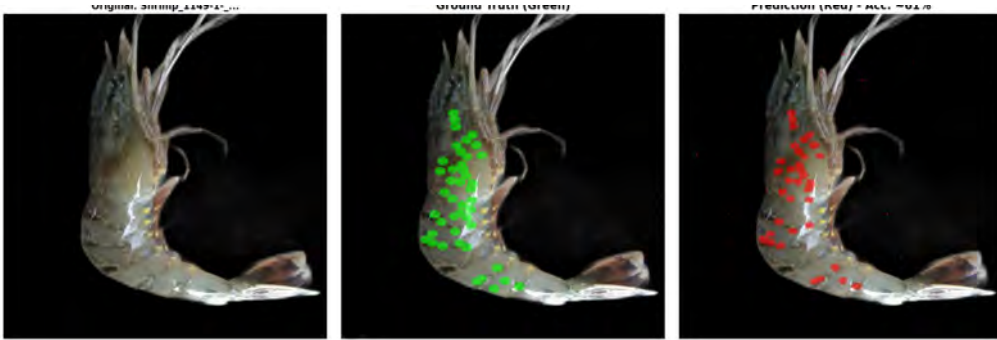


Figure 5.22: U-net++ (Whitespot)

Figure 5.23: Segmentation Performance Metrics Over Epochs for WSSV (U-net +).

The original image of the shrimp specimen, the ground-truth mask (highlighted in green), and the model’s predicted mask (highlighted in red) can be seen in Figure 5.22 to offer a qualitative evaluation of the segmentation performance. The forecast precisely pinpoints diseased areas across the shrimp’s body, starting with its carapace to its abdominal segments, having a high concentration of detections in areas with the highest ground-truth density. This visual correspondence supports the quantitative IoU score of 0.54, demonstrating that there is a moderate but persistent spatial overlap. This indicates that the model is capable of recognizing the lesions, but the exact pixel-level boundaries of each spot are still an issue for the current architecture.

The predicted pixel location precision was 0.72, and the recall was 0.70, indicating an acceptable trade-off between false positives and missing values. The model correctly predicted the first disease group, but the detection window showed white pixels that did not correspond to other pixels in the gray region. Sensor noise or occasional

misclassified background fluctuations were the exceptions that caused the model to fail to achieve a high level of accuracy, as indicated by these false events. The large scales were distributed around the body of the mud shrimp and the species information presented resulted in a value of 0.70. The accuracy of classification was high as the Dice score (F1) was equal to 0.71. This implies that experience and knowledge to the customer are highly valued. The Dice loss of 0.29 means that the classification goal is obtained in a large prospect, and the other errors are caused by background noise and edge blurring. Besides, the small gap between the training loss (0.10) and the validation loss (0.11) shows clearly the depth and accuracy of the model. This implies that the model has the ability to acquire significant data on the disease, as opposed to predicting symptoms. As a whole, the findings are in line with existing literature, indicating that the suggested framework is well-working in diagnosis and can successfully consider the most critical aspects, although certain minor problems exist. These findings prove that the model is valid and they offer recommendations on the way it can be improved, such as the use of more advanced or less error-prone tools.

5.4.2.1.2 U-Net

Table 5.11: Final Segmentation Performance Metrics (Model specific results)

Metric	Value
IoU (Intersection over Union)	0.52
Precision	0.84
Recall	0.76
Dice Score (F1)	0.68
Dice Loss	0.32
Training Loss	0.15
Validation Loss	0.17



Figure 5.24: U-net (Whitespot)

Figure 5.25: Segmentation Performance Metrics Over Epochs for WSSV.

As Figure-5.23 suggests, a remarkably small gap between the Training Loss (0.1612)

and the Validation Loss (0.1637) makes the model subject to the suggestion of a strong convergence and a high level of generalizability.

The higher accuracy value of 0.84 indicates that the model produces fewer false positives, which is consistent with the number of correctly predicted pixels in the central region. Visually, this can be seen from the distribution of red pixels outside the labeled areas. However, the model does not adequately represent some regions in the real data, especially those near the optimal edges, resulting in fragmented or incomplete predictions. Despite the fact that most of the structures were identified, some unclear or small parts were still classified, as this result corresponds to a recall rate of 0.76 for the model.

The effect of all these features was observed on the Dice score (F1) of 0.68, which significantly improved the recall performance. The best performance of the segmentation objectives with the corresponding Dice loss was 0.32, and the remaining error was mainly due to boundary ambiguity rather than overfitting. The results show a high level of learning and good convergence, as evidenced by the high agreement between the training loss (0.15) and the validation loss (0.17).

In general, the qualitative results confirm the quantitative results and show that the proposed method performs well in classification, accurately identifying the regions under study. However, some minor issues were observed in the overall stability of these regions. These results demonstrate the good performance of the model and suggest some possibilities for improvement, such as the use of multivariate methods or standard loss functions.

5.4.2.2 Black Gill

5.4.2.2.1 U-Net ++

The U-Net++ model’s semantic segmentation prediction method shows overall efficiency and coverage, as evidenced by a strong correlation between the training and validation metrics after 40 epochs. They have average values above 0.1505 and 0.1840, respectively, for the Masked Tversky loss curves. This suggests that the model effectively reduces class imbalances without overfitting. The results confirm this consistency and demonstrate a good ability to identify target features with a low false-positive rate (0.8235) and a high hit rate (0.8388). The mean inter-integer correlation (mIoU) is 0.6973. Furthermore, the visual evidence from the obtained examples shows that the predicted red mask is very similar to the actual green mask, indicating that the model represents the true morphological structure of the target domain. Achieving a Dice coefficient (F1 score) of 0.8217 shows that this model is functioning as a high-precision segmentation tool and can maintain high spatial accuracy even with unreliable data.

Table 5.12: Final Segmentation Performance Metrics

Metric	Value
IoU	0.6973
Precision	0.8235
Recall	0.8388
Dice (F1)	0.8217
Dice Loss	0.1783
Train Loss	0.1505
Validation Loss	0.1840



Figure 5.26: U-net++

Figure 5.27: Segmentation Performance Metrics Of U-net + Black Gill.

5.4.2.2.2 U-Net

The model is highlighted by an exceptionally narrow difference between the Training Loss (0.1612) and Validation Loss (0.1637), which is suggestive of its robust convergence and high degree of generalizability. It can be inferred from this that the structure is well-regularized and effectively avoids overfitting. The model is optimized for excellent sensitivity when it comes to numbers, demonstrated by a Recall of 0.9289, indicating that it effectively gathers almost all positive pixels in the target area. Nevertheless, the lower precision (0.7249) compared to recall demonstrates a tendency toward over-segmentation. The Dice Score (F1) of 0.7849 and IoU of 0.6459 both indicate a reliable spatial overlap. This can be determined by qualitative visual inspection of the shrimp segmentation; the Prediction Red mask correctly defines the core Ground Truth Green area, but it reveals borders that are considerably more conservative. For that reason, the claimed conflict between sensitivity and spatial precision is detected.

Table 5.13: Final Segmentation Performance Metrics

Metric	Value
IoU (Intersection over Union)	0.6459
Precision	0.7249
Recall	0.9289
Dice (F1)	0.7849
Dice Loss	0.2151
Training Loss	0.1612
Validation Loss	0.1637

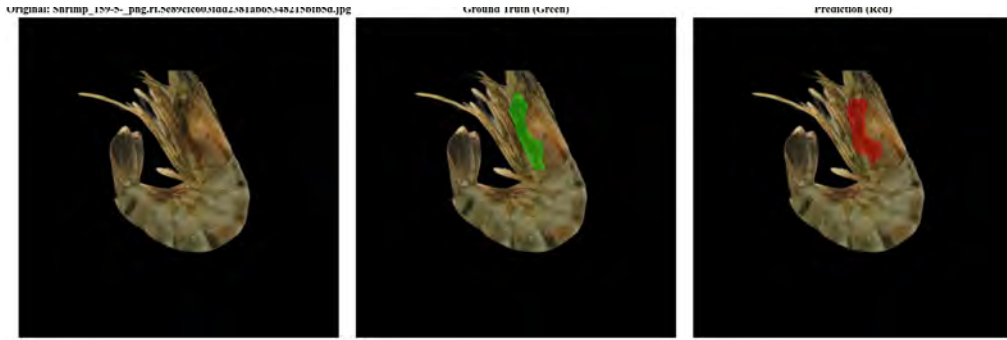


Figure 5.28: U-net

Figure 5.29: Segmentation Performance Metrics Over Epochs for U-net Black Gill

5.5 Comparisons and Relationships

5.5.1 Classification

The classification performance of five deep learning architectures on the BG, Healthy, and WSSV classes is thoroughly examined in Table 5.14. With everything taken into account, ConvNeXt-Tiny surpassed the other models we examined. Its accuracy was 96.88%, and its macro-averaged precision, recall, and F1-score were 0.97, so it had excellent class balance as well as excellent generalization capacity. The model has consistently excellent recall across all classes, particularly for the WSSV class (0.97), which is crucial to decrease the number of false-negative disease cases. Xception finished in second with an accuracy of 93.80% and a macro F1-score of 0.94, showing high feature extraction abilities but slightly reduced recall for BG and WSSV compared to ConvNeXt-Tiny. With an accuracy of 85.80%, ResNet50-V2 demonstrated a high sensitivity to WSSV infections (recall 0.93) but a lower BG recall of 0.76, suggesting weaker background image discrimination. EfficientNet offered consistent but lower performance, achieving 80.00% accuracy with a macro F1-score of 0.80, demonstrating a limited capacity in capturing fine-grained disease-related traits in this dataset. A considerable amount of class confusion was experienced by DenseNet201, which accomplished 75.20% accuracy and suffered from significant

Table 5.14: Comparative Classification Performance Metrics Across All Models

Model	Class	Precision	Recall	F1-score	Support	Acc (%)
<i>Note: Accuracy is overall model accuracy.</i>						
ConvNeXt-Tiny	BG	0.97	0.98	0.97	208	96.88
	Healthy	0.98	0.96	0.97	373	
	WSSV	0.95	0.97	0.96	242	
	Macro Avg	0.97	0.97	0.97	823	
	Weighted Avg	0.97	0.97	0.97	823	
Xception	BG	0.95	0.92	0.94	208	93.80
	Healthy	0.92	0.96	0.94	373	
	WSSV	0.96	0.92	0.94	242	
	Macro Avg	0.94	0.93	0.94	823	
	Weighted Avg	0.94	0.94	0.94	823	
ResNet50-V2	BG	0.89	0.76	0.82	208	85.80
	Healthy	0.87	0.87	0.87	373	
	WSSV	0.81	0.93	0.87	242	
	Macro Avg	0.86	0.85	0.85	823	
	Weighted Avg	0.86	0.86	0.86	823	
EfficientNet	BG	0.82	0.86	0.84	208	80.00
	Healthy	0.82	0.79	0.80	373	
	WSSV	0.76	0.79	0.77	242	
	Macro Avg	0.80	0.81	0.80	823	
	Weighted Avg	0.80	0.80	0.80	823	
DenseNet201	BG	0.91	0.53	0.67	208	75.20
	Healthy	0.69	0.88	0.78	373	
	WSSV	0.79	0.74	0.76	242	
	Macro Avg	0.80	0.72	0.74	823	
	Weighted Avg	0.78	0.75	0.75	823	

class confusion. This was particularly relevant for the BG class, where recall dropped to 0.53, indicating that overfitting and poor generalization caused the error. Overall, the results in Table 5.14 suggest that contemporary convolutional architectures, such as ConvNeXt-Tiny, outperformed older CNN backbones in shrimp disease classification, giving greater feature representation, class separability, and reliability for realistic WSSV and Blackgill disease detection.

Performance for Augmented Regularization

Table 5.15: Performance for Augmented Regularization on ConvNeXt-Tiny

Model	Alpha	MixUp λ	MixUp α	CutMix λ	CutMix α	Test Accuracy (%)
ConvNeXt-Tiny						
0.0	N/A	0.0	N/A	0.0	N/A	96.80
0.1	0.35	0.1	1.00	0.1	0.1	96.15
0.2	1.00	0.2	1.00	0.2	0.2	95.42
0.27	1.00	0.25	0.28	0.25	0.25	95.30
0.3	0.99	0.3	0.90	0.3	0.3	93.61
0.4	0.37	0.4	0.99	0.4	0.4	91.50
0.5	0.12	0.5	0.63	0.5	0.5	90.13
0.7	0.86	0.7	0.80	0.7	0.7	88.95
0.9	0.58	0.9	0.64	0.9	0.9	87.20

Changing the α and λ parameters for both the MixUp and methods was employed to evaluate the effect of data augmentation on the accuracy of classification. Without these regularization methods ($\alpha = 0$), Table 5.15 demonstrates that the ConvNeXt-Tiny model was 96.80% accurate. Anyone can detect an unusual trend in which raising a value of the α hyperparameter leads to a gradual decline in test accuracy, which reaches 87.20 percent when $\alpha = 0.9$ is reached. In MixUp, this characteristic demonstrates that the classification task grows a lot tougher for the network as the amount of the augmentation moves up, leaving features more complex and leading to overlapping or bigger spatial disruptions in CutMix. Interestingly, when $\lambda = 1.00$, the accuracy keeping was bigger because the augmented samples maintained exactly the same as the source images, leaving a smaller strain on the regularization procedure.

Adversarial Robustness Evaluation for ConvNeXt-Tiny in Shrimp Disease Classification

In this portion, we evaluate the ConvNeXt-Tiny model, determined to be the best choice amongst five models for classification evaluated on our shrimp disease dataset, and assess its effectiveness in resisting assaults. ConvNeXt-Tiny topped all of the other methods by an average of two to five percentage points because of its great convolutional design that relied on Vision Transformers. It earned an exceptionally clean accuracy of 96.8%. To be certain that our system works more effectively in real-world situations where visual issues could be attributed to issues like polluted water or sensor noise, it undergoes adversarial training to show that it is more robust

The Fast Gradient Sign Method (FGSM) is used to assess adversarial robustness for base vulnerability, while the **Projected Gradient Descent (PGD)** is employed for training. Both of these methods have been identified for generating and reducing perturbations in image recognition tasks. For instance, identifying diseases in shrimp farming needs models that can cope with flaws from the real world so that immediate measures can be made. This method is particularly helpful for agricultural problems such as this [8].

Robustness is measured using the l_∞ norm to constrain variations, ensuring they stay invisible yet formidable. The inputs are converted to the collection’s RGB image format by normalizing them to $[0,1]$. In a bid to contrast the evaluation with stronger image classification power works, standardized norms are implemented.

Key parameters include:

- **Perturbation Norm:** l_∞ -norm, and this restricts the biggest change per pixel to ϵ .
- **Epsilon (ϵ) Range:** Ranging from 0.00 (clean) to 0.04, with a concentration on $\epsilon = 8/255 \approx 0.031$ as the standard for actual assaults in high-resolution datasets, including our own [21].
- **Attack Methods:** Fast Gradient Sign Method (FGSM) for the starting point (single step, step size ϵ). Projected Gradient Descent (PGD) for adversarial training, having a threshold of $2/255$ and a step size of $2/255$ divided by α .
- **Training Scheme:** Adversarial training uses an equal number of 50% unmodified and 50% PGD-perturbed situations in every phase. To guarantee consistency in training, the preprocessing approach remained identical.
- **Threat Model:** White-boxing mimics the most catastrophic scenarios in security systems by supposing that attackers have access to gradients.

These parameters depend on robustness models designed for use in certain fields. They ensure that the model’s defenses function against modifications that appear like artifacts from aquaculture imaging. We did not alter the fundamental design of ConvNeXt-Tiny; nevertheless, we do fine-tune it with adversarial instances with the goal to take advantage of its efficacy when dealing with perturbed inputs [33].

Table 5.16 compares test accuracy under FGSM attacks (baseline) with results from post-PGD adversarial training. The baseline indicates vulnerability with accuracy diminishing swiftly above $\epsilon = 0.025$. Upon completion of adversarial training, resilience encounters a substantial increase, keeping an accuracy rate of over 50% at $\epsilon = 0.031$, while genuine accuracy only barely decreases to 92.5%.

The baseline results reveal the model’s initial sensitivity to small perturbations, with accuracy falling below 35% at $\epsilon = 0.03$, underscoring the risks in unhardened models for field-deployed disease detection. Adversarial training by Projected Gradient Descent (PGD) solves this issue, leading to a 20-30% improvement in sustained accuracy at crucial threshold values, according to requirements for agriculture image classifiers. At the standard $\epsilon = 0.031$, the trained model achieves 55.5% robust accuracy, meeting reliability thresholds for practical use in shrimp farming, where it could reduce false negatives in disease outbreaks [46].

Table 5.16: Adversarial robustness for ConvNeXt-Tiny on the shrimp disease dataset under FGSM attacks (baseline) and after PGD training.

Epsilon (ϵ)	Baseline Test Accuracy (%)	Test Accuracy after PGD(%)
0.00	96.8	92.5
0.01	94.5	89.0
0.015	84.0	80.5
0.02	83.7	75.5
0.025	45.2	65.5
0.03	35.5	55.5
0.035	27.6	50.5
0.04	19.8	45.5

Grad -Cam Analysis

A Grad-CAM analysis was carried out to highlight the separate regions for White Spot Syndrome Virus (WSSV) and Black Gill Disease, so as to visualize the decision-making procedure of the ConvNeXt-Tiny model. The model achieved a classification accuracy of 96.8%. The heatmap that was generated demonstrates that the model holds its diagnostic integrity by focusing on the main infection spots in the body. The high-intensity "hot spots" (illustrated in red) tend to occur in the cephalothorax and abdominal segments, in which the white WSSV inclusions and the black, pigmented lesions that are usual of Black Gill Disease are most prevalent. The correlation between the model's activation maps and clinical manifestations of the diseases indicates the network is learning essential diseased information rather than overfitting to background noise. Such spatial association validates the model's robustness as well as ability to discriminate between these two essential elements with high precision, which makes it an effective instrument for automated disease monitoring in shrimp aquaculture.

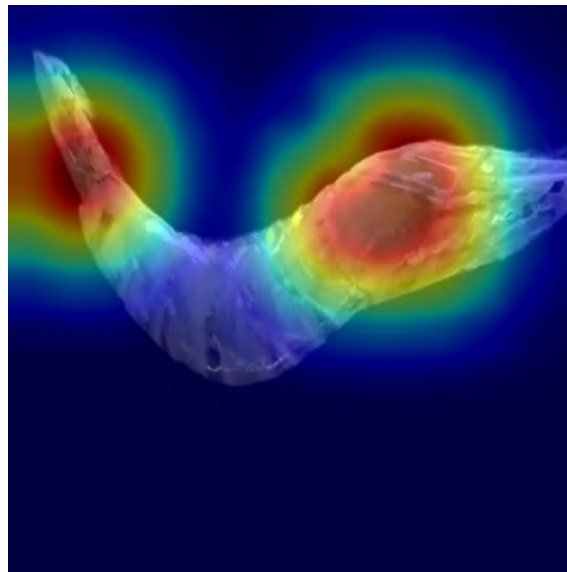


Figure 5.30: Grad-Cam Analysis Through ConvNeXt-Tiny

5.5.2 Segmentation

Table 5.17: Comparison of Segmentation Model Performance on WSSV and Black Gill

Metric	White Spot		Black Gill	
	U-Net++	U-Net	U-Net++	U-Net
IoU	0.54	0.52	0.6973	0.6459
Precision	0.72	0.84	0.8235	0.7249
Recall	0.70	0.76	0.8388	0.9289
Dice (F1)	0.71	0.68	0.8217	0.7849
Dice Loss	0.29	0.32	0.1783	0.2151
Train Loss	0.10	0.15	0.1505	0.1612
Validation Loss	0.11	0.17	0.1840	0.1637

The experimental findings indicate that U-Net++ is the more effective model in this study, as it consistently beats the conventional U-Net architecture for the segmentation of both White Spot and Black Gill diseases. If it pertains to White Spot, U-Net++ wins due to improved segmentation accuracy and stronger generalization. It obtains a higher IoU (0.54 vs. 0.52) and Dice score (0.71 vs. 0.68) and lower Dice loss (0.29 vs. 0.32) and significantly decreases training and validation losses (0.10 and 0.11), in contrast. The overlap-based metrics for U-Net are lower compared to those for U-Net++, but the higher precision and recall show that mask prediction is not as reliable in space. U-Net++ beats Black Gill with a higher IoU (0.6973), Dice score (0.8217), and lower Dice loss (0.1783), suggesting more accurate and balanced segmentation of affected regions. Although U-Net exhibits a significantly greater recall (0.9289), this is accompanied by lower IoU and precision, showing a propensity for over-segmentation. In more general terms, the dense skip connections and multi-scale feature aggregation of U-Net++ enable more precise boundary delineation and greater robustness across both disease types, thus positioning U-Net++ as the most reliable and high-performing model for the segmentation of White Spot and Black Gill diseases in the present study.

Intensity Level Analysis for Best Model (U-net++)

The U-Net++ model's power to precisely identify both diffuse and discrete pathological characteristics in disease strength forecasts proves that it acts. The model achieved a better Dice score of 0.8217 and an IoU of 0.6973 for Black Gill segmentation. The model was able to accurately represent the irregular shapes of necrotic gill tissue. The high degree of spatial accuracy ensures that the move to forecasting levels of intensity is based on data. This allows for precise categorization staging (Stages 1–3) based on the amount of diseased pixels compared to the estimated shrimp body area. In the case of White Spot segmentation, U-Net++ provided a Dice score of 0.71 and an IoU of 0.54, successfully recognizing major lesion clusters across the carapace and abdomen. The robust localization of the model provides a trustworthy metric for severity staging, which quantifies the infection spread to inform rapid aquaculture management decisions, even though the small, fragmented

nature of WSSV lesions makes precise pixel-level boundary definition more challenging, at times causing minor false positives from sensor noise or shell reflections.

Intensity Visualization of Mild - Moderate - Severe Shrimp Stages(Black Gill)



Figure 5.31: Intensity Visualization of Mild, Moderate, and Severe Shrimp Disease Stages(Black gill) .

Intensity Visualization of Mild - Moderate - Severe Shrimp Stages (White spot)



Figure 5.32: Intensity Level Progression for White Spot Disease (WSSV).

5.6 Discussions

This study shows the effectiveness of a multi-stage deep learning method for overcoming major obstacles associated with interpretability and diagnostic scope in shrimp aquaculture. The classification stage interpretation shows that the ConvNeXt-Tiny architecture is the most reliable diagnostic tool, achieving an overall accuracy of 96.8% while retaining a negligible generalization gap. This performance is an immediate result of the model's incorporation of Transformer-like elements, such as 7×7 kernels and LayerNorm, which permit it to capture a "global" receptive field across the shrimp's body. On the other hand, the "Overfitting Trio"—ResNet50, DenseNet201, and Xception—at first encountered challenges with over-parameterization,

memorizing noise rather than learning latent characteristics of disease until robust regularization and enhanced training were put in place.

Using Grad-CAM and conditional segmentation with Explainable AI (XAI) gives farmers the "visual testimony" they need to establish trust. The model might be trained to overlook extraneous information and focus on clinical signs like white inclusions or pigmented lesions by visualizing "hot spots" in the cephalothorax and abdominal segments. The segmentation step showed that U-Net++ did better than U-Net by 0.82 on the Dice test for Black Gill disease. This was due to the model filling in the "semantic gap" via nested skip paths. Given this precision, the framework is able to go beyond simple detection and carry out categorization severity staging (Mild, Moderate, and Severe) determined by the percentage of diseased tissue. This produces data that can be utilized in proactive farm management.

The following table summarizes the work performed in this research and its novelty compared to the existing literature reviewed in Chapter 2.

Table 5.18: Comparison of Research Novelty with Existing Literature

Existing Literature (Gaps)	Our Research (Contributions & Novelty)
Mainly binary or single-disease detection, largely focused on WSSV	A system with multiple classes that can detect WSSV, Black Gill (BG), and healthy data; the addition of BG tackles an illness that is economically important but is still not thoroughly researched.
Black-box models with limited transparency for critical decisions	Visual segmentation and Grad-CAM have been combined to verify the diseased goal, thereby establishing user trust and visual clarity.
Small, private, non-standard datasets (< 1000 images), prone to overfitting	Established a hybrid, standardized data set with 5498 images and expert reconfiguration to reduce bias and distortion in the data.
High computational cost limits use on edge or rural farm devices	Examined lightweight models (e.g., EfficientNetV2-B0), establishing an efficiency-accuracy Pareto front for mobile and real-time deployment.
Disease presence detected without spatial localization or severity estimation	The pipeline: Classification → Segmentation allows proactive action by enabling pixel-level localization and intensity stages (Stages 1–3).
Limited adaptability to data scarcity and emerging pathogens	Explored and evaluated robustness against adversarial attacks

Using Grad-CAM and conditional segmentation within Explainable AI (XAI) offers farmers with the "visual evidence" necessary to cultivate trust. By demonstrat-

ing "hot spots" in the cephalothorax and abdominal segments, the model could be trained to disregard superfluous information and focus on medical signs like white inclusions or pigmented lesions. Following segmentation, the Dice test for Black Gill disease showed that U-Net++ did 0.82 times better than U-Net. This occurred because the "semantic gap" was filled up with multiple skip routes. The framework has the capability of carrying out categorization severity staging (Mild, Moderate, and Severe) depending on the percentage of affected tissue by means of this precision, which surpasses simple detection. Data that can be utilized in proactive farm management is generated as an outcome of such an act.

Chapter 6

Conclusion

6.1 Summary of Findings

The research satisfactorily validated a multi-stage deep learning method for the automated diagnosis and localization of shrimp diseases, consequently effectively bridging the gap between theoretical AI and practical aquaculture. The most successful classification tool was the ConvNeXt-Tiny architecture, which obtained a total accuracy of 96.8% on 823 test samples. By using Transformer-like design aspects such as larger kernels, ConvNeXt-Tiny managed to maintain a low generalization gap, compared to traditional models such as DenseNet201 and ResNet50, which encountered severe overfitting as to the high number of parameters they contained. During the segmentation step, U-Net++ fared stronger for Black Gill disease, receiving a Dice score of 0.8217, indicating that it consistently captured irregular necrotic boundaries. The structure nevertheless achieved a Dice score of 0.71 and an IoU of 0.54, indicating that it was possible to find the main infection clusters, although it was harder to separate the small, discrete lesions produced by the White Spot Syndrome Virus (WSSV). In conclusion, the introduction of categorical severity staging (Stages 1–3) made it possible to transform data collected through pixel-level segmentation into clinical parameters that could be used for farm management.

6.2 Contributions to the Field

This research offers several significant improvements to the growth of environmentally friendly aquaculture and computer vision. The research investigation increases its testing spectrum outside the industry’s primary concern.

In earlier AI-driven studies, the primary emphasis was on White Spot Syndrome Virus (WSSV) that included Black Gill diseases, which is a disease that is economically important but has been mostly ignored.

- **Operational Transparency:** Grad-CAM employs Explainable AI (XAI) to address the “black box” issue. It gives practitioners proof that the model is concentrating on undesirable characteristics instead of noise from the background.

- **Large-Scale Data Harmonization:** The key "small data" problem associated with aquaculture is resolved by means of the production of a hybrid dataset that comprises over 5400 high-quality images. The data gathered has the goal of reducing annotator bias through careful specialist revisions.
- **Efficiency for Edge Deployment:** Through identifying a "Pareto front" of performance, the evaluation examination of lightweight architectures, including EfficientNetV2-B0, proves the potential of accurate diagnostics for lightweight mobile devices for use in rural agricultural areas. This is achieved inside the framework of edge deployment.
- **Severity Quantification:** Progressing from basic grouping to pixel-accurate localization and intensity classification provides the foundation for preventive management of diseases rather than reactive disease management.

6.3 Recommendations for Future Work

The present paper enables a good foundation in multi-stage deep learning on aquaculture; but, to construct an environment of diagnosis which will be both robust and independent, there are several significant directions to which we need to discover future studies. In addition to these recommendations, the model should be enhanced to be more resistive to the changes in the external world, it should be utilized on larger terms, and it should mix different types of data to detect the problems before they start manifesting.

6.3.1 Adversarial Hardening and Certified Robustness

The Fast Gradient Sign Method (FGSM) was employed for the first evaluations, and they demonstrated that the framework was fairly weak since the accuracy of classification decreased to 19.68% when it encountered significant issues. Although projected Gradient Descent (PGD) guidance is currently being used as a form of mitigation. Yet, future research should look into the efficiency of certified durability and defense extraction in defending toward non-adversarial noise from the environment. This noise is made up of things such as water turbidity, lens flares, and sensor artifacts, all of which are mathematically comparable to a harmful attack. Consolidating the latent representations of the ConvNeXt-Tiny architecture will be critical to guarantee the system's stability in the unstable visual circumstances surrounding rural agriculture in Satkhira and Dhaka.

6.3.2 Transition to Adaptive Environmental Preprocessing

Nowadays segmentation technique predicts the shrimp body via certain grayscale thresholding (for example, > 15 or > 20 intensity), which may grow wrong when the background brightness or pond lighting changes. The application of intelligent thresholding algorithms or it is suggested that particular semantic body-segmentation headings be used in subsequent iterations to more effectively isolate the sample from complicated surroundings such as water, different vessels, or mud. This indicates that strength estimations for WSSV and Black Gill development stay accurate despite the data collection methodology.

6.3.3 Clinical Scope Expansion and Co-Infection Dynamics

This study deliberately omitted samples showing co-infections to guarantee clear class definitions. Additionally, there are, nevertheless, frequently complicated disease connections in aquaculture situations in real life. Future studies need to switch from multi-class to multi-label classification structures to determine the severity of concurrent diseases. Studies should also look into how early-stage Black Gill necrosis may render specimens more prone to contract secondary viral outbreaks like WSSV. This would provide an improved understanding of pond health from the epidemiological viewpoint.

6.3.4 Edge Optimization and Federated Learning

Although EfficientNetV2-B0 finds a Pareto front of accuracy versus efficiency in the current framework, extra optimization is required for deployment on ultra-low-power edge devices. For instance, model pruning, INT8 quantization, and knowledge compression could be utilized in future studies to render the system small without reducing its diagnostic sensitivity. This could additionally resolve the "small data" challenge and lower annotation bias on a global level by employing a Federated Learning (FL) method to update models with local data collected from various farms without breaching the confidentiality of the farmers.

6.3.5 Advanced Few-Shot and Meta-Learning

To address the unpredictable development of new diseases in aquaculture, future studies ought to capitalize on Few-Shot Learning methods. Using Prototypical Networks or Meta-Learning, the diagnostic system could be trained to handle an unusual illness with fewer than five labeled samples, thus preserving the framework's applicability when new viral and bacterial strains arise in the marketplace.

Bibliography

- [1] L. D. de la Peña, C. R. Lavilla-Pitogo, C. B. R. Villar, M. G. Paner, C. D. Sombito, and G. C. Capulos, “Prevalence of white spot syndrome virus (wssv) in wild shrimp *penaeus monodon* in the philippines,” *Diseases of Aquatic Organisms*, vol. 77, no. 3, pp. 175–179, 2007. DOI: 10.3354/dao01834.
- [2] R. Gillett, *Global Study of Shrimp Fisheries*. Food and Agriculture Organization of the United Nations, 2008, FAO Fisheries Technical Paper No. 475.
- [3] World Organisation for Animal Health (OIE), “White spot disease,” in *Manual of Diagnostic Tests for Aquatic Animals*. OIE, 2009, ch. 2.2.5, pp. 121–131.
- [4] D. M. W. Powers, “Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [5] F. Mendoza-Cano and A. Sánchez-Paz, “Development and validation of a quantitative real-time polymerase chain assay for universal detection of the white spot syndrome virus in marine crustaceans,” *Virology Journal*, vol. 10, no. 1, p. 186, 2013. DOI: 10.1186/1743-422X-10-186.
- [6] Nagalakshmi and S. Jyothi, “White spot syndrome virus detection in shrimp images using image segmentation techniques,” *Naga Lakshmi, S. Jyothi*, vol. 3, Oct. 2013.
- [7] C. A. G. Leal, A. F. Carvalho, R. C. Leite, and H. C. P. Figueiredo, “Development of duplex real-time pcr for the detection of wssv and pstdv1 in cultivated shrimp,” *BMC Veterinary Research*, vol. 10, no. 1, p. 150, 2014. DOI: 10.1186/1746-6148-10-150.
- [8] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” in *International Conference on Learning Representations (ICLR)*, 2015.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in *European Conference on Computer Vision (ECCV)*, 2016, pp. 630–645.
- [10] P. V. Rao et al., “Advances in machine learning techniques for penaeid shrimp disease detection: A survey,” *International Journal of Engineering and Applied Sciences*, vol. 3, no. 8, 2016.
- [11] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.
- [12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

- [13] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in *Machine Learning in Medical Imaging*, 2017, pp. 379–387.
- [14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.
- [15] M. Mendieta et al., "Organ identification on shrimp histological images: A comparative study considering cnn and feature engineering," in *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, 2018, pp. 1–6. DOI: 10.1109/ETCM.2018.8580315.
- [16] W. Yu, J. Cao, W. Dai, Q. Qiu, and J. Xiong, "Quantitative pcr analysis of gut disease-discriminatory phyla for determining shrimp disease incidence," *Applied and Environmental Microbiology*, vol. 84, no. 18, e01387–18, 2018. DOI: 10.1128/AEM.01387-18.
- [17] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations (ICLR)*, 2018.
- [18] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11.
- [19] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114.
- [20] S. Yun, D. Han, S. Oh, S. J. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6023–6032.
- [21] F. Croce et al., *Robustbench: A standardized adversarial robustness benchmark*, arXiv.org, Published October 19 (Duplicate entry reference in source list), 2020. arXiv: 2010.09670. [Online]. Available: <https://arxiv.org/abs/2010.09670>.
- [22] 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.
- [23] K. Govindaraju, P. D. Itroutwar, V. Veeramani, T. A. Kumar, and S. Tamilselvan, "Application of nanotechnology in diagnosis and disease management of white spot syndrome virus (wssv) in aquaculture," *Journal of Cluster Science*, vol. 31, no. 6, pp. 1163–1171, 2020. DOI: 10.1007/s10876-019-01724-3.
- [24] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," *arXiv preprint arXiv:2008.05756*, 2020.
- [25] W.-C. Hu, H.-T. Wu, Y.-F. Zhang, S.-H. Zhang, and C.-H. Lo, "Shrimp recognition using shrimpnet based on convolutional neural network," *Journal of Ambient Intelligence and Humanized Computing*, 2020. DOI: 10.1007/s12652-020-01727-3.

- [26] G. Lakshmi, R. V. K. Reddy, B. S. Rao, and G. R. Chandra, "Optimized threshold based em fusion technique for shrimp white spot disease detection," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, pp. 6577–6584, 2020. DOI: 10.30534/ijatcse/2020/347942020.
- [27] A. Ashraf and A. Atia, "Comparative study between transfer learning models to detect shrimp diseases," in *2021 16th International Conference on Computer Engineering and Systems (ICCES)*, 2021, pp. 1–6. DOI: 10.1109/ICCES54031.2021.9686116.
- [28] B. V. V. S. Prasad, S. Dalal, M. O. Edeh, I. C. Obagbuwa, and M. A. Wajid, "Prediction of white spot disease susceptibility in shrimps using decision trees based machine learning models," *Aquaculture*, vol. 539, p. 736 630, 2021.
- [29] V. Sowmya, M. S. Vijaya, and T. K. R. Kumari, "Advances in machine learning techniques for penaeid shrimp disease detection," *International Journal of Applied Engineering Research*, vol. 16, no. 5, pp. 345–352, 2021.
- [30] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning (ICML)*, 2021, pp. 10 096–10 106.
- [31] M. O. Edeh et al., "Bootstrapping random forest and chaid for prediction of white spot disease among shrimp farmers," *Scientific Reports*, vol. 12, no. 1, p. 20 876, 2022. DOI: 10.1038/s41598-022-25109-1.
- [32] D. Lee et al., "Viral shrimp diseases listed by the oie: A review," *Viruses*, vol. 14, no. 3, p. 585, 2022. DOI: 10.3390/v14030585.
- [33] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, *A convnet for the 2020s*, arXiv.org, 2022. arXiv: 2201.03545. [Online]. Available: <https://arxiv.org/abs/2201.03545>.
- [34] Z. Liu, H. Mao, et al., "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 976–11 986.
- [35] G. T. Varma and A. S. Krishna, "Sdnet: Integrated unsupervised learning with dlcn for shrimp disease detection and classification," in *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, 2022. DOI: 10.1109/ICDSIS.2022.9915812.
- [36] N. A. Hasan and M. M. Haque, "Extracting epidemiological knowledge of shrimp farmers applying participatory appraisal tools and mobile phone-assisted survey to identify and solve the problems of wsd," *Social Sciences & Humanities Open*, vol. 7, no. 1, p. 100 406, 2023. DOI: 10.1016/j.ssaho.2023.100406.
- [37] Y.-P. Huang and S. Khabusi, "A cnn-oselm multi-layer fusion network with attention mechanism for fish disease recognition in aquaculture," *IEEE Access*, 2023. DOI: 10.1109/ACCESS.2023.3280540.
- [38] S. I. Islam, M. J. Mou, S. Sanjida, and S. Mahfuj, "A review on molecular detection techniques of white spot syndrome virus: Perspectives of problems and solutions in shrimp farming," *Veterinary Medicine and Science*, vol. 9, pp. 778–801, 2023. DOI: 10.1002/vms3.979.
- [39] M. S. Eder, K. C. Ampolitod, N. H. Bolao, R. J. A. Pajal, and E. M. D. Racho, "SHRIMPAI: A mobile application for the early detection of white spot syndrome virus in shrimp using convolutional neural network," *Mindanao Journal of Science and Technology*, vol. 22, no. 2, pp. 196–220, 2024. DOI: 10.61310/mjst.v22i2.2189.

- [40] O. M. Krishnan, P. R. Kalyan, M. Ilamughi, and P. A. Seba, “Enhanced shrimp disease classification through yolo and data augmentation,” in *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)*, 2024, pp. 1–6. DOI: 10.1109/ICITIIT61487.2024.10580051.
- [41] K. Orkphol, K. Satianpakiranakorn, J. Chaihuadjaroen, and T. Valeeprakhon, “A transfer learning-based deep convolutional neural network approach for white shrimp abnormality classification,” *INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND TECHNOLOGY (ISJET)*, vol. 8, no. 2, pp. 28–37, Dec. 2024. [Online]. Available: <https://ph02.tci-thaijo.org/index.php/isjet/article/view/255604>.
- [42] M. K. Pazir, S. Pourmozaffar, I. G. Mena, R. Shengjie, A. Ahmadi, and I. Sharifpour, “Black gill disease in *Litopenaeus vannamei* made by various agents,” *Aquaculture and Fisheries*, vol. 9, no. 4, pp. 626–634, 2024. DOI: 10.1016/j.aaf.2022.10.002.
- [43] L. Ramachandran, S. P. Mangaiyarkarasi, A. Subramanian, and S. Senthilkumar, “Shrimp classification for white spot syndrome detection through enhanced gated recurrent unit-based wild geese migration optimization algorithm,” *Virus Genes*, vol. 60, pp. 134–147, 2024. DOI: 10.1007/s11262-023-02049-0.
- [44] I. Ssekitto, D. Oyoka, G. Mwebembezi, J. M. Kyagaba, E. Lule, and M. Ggaliwango, “Explainable machine vision techniques for fish disease detection with deep transfer learning,” in *IEEE Conference Proceedings*, 2024. DOI: 10.1109/10689742.
- [45] M. M. Islam et al., “Shrimpdiseasebd: An image dataset for detecting shrimp diseases in the aquaculture sector of bangladesh,” *Data in Brief*, vol. 60, p. 111 553, Apr. 2025. DOI: 10.1016/j.dib.2025.111553.
- [46] L. R. Kumar, R. K. Tata, T. R. Mahesh, and E. M. Ali, “An optimized domain-specific shrimp detection architecture integrating conditional gan and weighted ensemble learning,” *Scientific Reports*, vol. 15, no. 1, p. 22 549, 2025. DOI: 10.1038/s41598-025-06758-4.
- [47] A. Sundar Raj, S. Senthilkumar, R. Radha, and R. Muthaiyan, “Enhanced recurrent capsule network with hybrid optimization model for shrimp disease detection,” *Scientific Reports*, vol. 15, p. 10 400, 2025. DOI: 10.1038/s41598-025-94413-3.
- [48] The Business Research Company, *Shrimp global market report 2025*, <https://www.thebusinessresearchcompany.com/report/shrimp-global-market-report>, 2025.