Application of Machine Learning to Determine Fish Freshness Based on Eye Images

John P. Q. Tomas [1,*], Maria R. J. C. Caranay [1], Eonn G. M. G. Domingo [1], Mia B. Enciso [1], and Bonifacio T. Doma [2]

[1] School of Information Technology, Mapua University, Makati City, Philippines
[2] School of Health Sciences, Mapúa University, Manila, Philippines
Email: jptomas@mapua.edu.ph (J.P.Q.T.); mrjccaranay@mymail.mapua.edu.ph (M.R.J.C.C.); egmgdomingo@mymail.mapua.edu.ph (E.G.M.G.D.); mbenciso@mymail.mapua.edu.ph (M.B.E.); btdoma@mapua.edu.ph (B.T.D.)
*Corresponding author

***Abstract*—The assessment of fish freshness is crucial for ensuring food safety and quality within the seafood industry. Traditional methods of freshness evaluation rely on sensory and instrumental assessments, which can be subjective and require expertise. This study explores the application of machine learning techniques to classify fish freshness based on eye images. A total of 880 images were collected from two fish species—milkfish (*Chanos Chanos)* and tilapia (*Oreochromis Niloticus*)—with distinct eye characteristics, spanning four freshness categories: excellent, good, average, and not fit for consumption. Features extracted from the eye regions, including RGB, CIE Lab*, and GLCM descriptors, were used to train three classification models: Naïve Bayes (NB), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN). Among the models, KNN achieved the highest accuracy of 77%. The study demonstrates the potential of automated, non-destructive, and objective machine learning-based approaches for evaluating fish freshness, contributing to improved quality control in the seafood industry.**

***Keywords*—fish freshness, image processing, k-Nearest Neighbors, machine learning, Naïve Bayes (NB), seafood quality, Support Vector Machine (SVM)**

I. Introduction

The freshness of fish is an essential consideration for both customers and the seafood business. It is of the utmost importance to ensure that fish is safe for eating and of excellent quality, as damaged seafood poses health hazards and economic losses for fishermen, suppliers, and customers. Traditional methods of assessing fish freshness rely on sensory evaluation, such as evaluating the fish's odor, texture, and color [1]. According to Nguyen *et al.* [2], applying machine learning to determine fish freshness based on eye images is an innovative and promising solution. Fisheyes can give vital information regarding their freshness since particular changes occur in the eyes as fish decay. Clouding of the cornea, color changes, and other apparent abnormalities are among these changes. Medeiros *et al*. [3] mentioned that machine learning algorithms could be trained to recognize these subtle variations in fisheye images, allowing for quick and objective freshness assessment.

Several things influenced the choice of this theme. The fish sector is an essential component of the global food supply chain, and guaranteeing the safety and quality of seafood products is critical for public health and economic sustainability. Second, traditional techniques for judging seafood freshness often require expertise, whether the evaluation is sensory or instrumental, which makes outsourcing the assessment to machine learning-based automation, especially for consumers who might not have the expertise to distinguish between different levels of fish freshness. Finally, advances in computer vision and machine learning have made it more possible to create accurate and trustworthy models for judging fish freshness based on eye images. This research on using machine learning to detect fish freshness can benefit consumers as it can provide foundations for the automation of image-based fish freshness assessment with machine learning. By utilizing an automated assessment process, consumers may be better equipped to make informed decisions regarding their purchase of fish products. As a result, this issue has the potential to make a significant contribution to both the seafood sector and the larger field of machine learning applications in food quality evaluation.

The Food and Agriculture Organization (FAO) of the United Nations equates fish freshness, or the degree of spoilage that the fish has undergone, with the quality of fish [4]. There are two main ways to determine fish freshness according to FAO, namely, sensory and instrumental, where the former entails using the five senses to measure and interpret the characteristics of food. Table I outlines a set of criteria utilized within the College of Fisheries in Central Luzon State University (CLSU) for the sensory evaluation of fish, which is based on the work of Larsen *et al*. [5].

The scoring in Table I is the basis of the definition of

Manuscript received March 28, 2025; revised April 29, 2025; accepted June 30, 2025; published January 16, 2026.

freshness levels provided by the Department of Post-Harvest in CLSU College of Fisheries [6]. Fish of excellent quality are those that have a score in the 0–5 range. They are fish with very few or no noticeable defects in appearance, odor, texture, or taste. These fish are considered fresh and highly desirable for consumption or commercial purposes. Fish of good quality are those that have a score in the 6–10 range. Minor defects may be present, but overall, the fish is still of acceptable quality. These fish may have slight blemishes or a slightly less than ideal odor, texture, or taste, but they are still suitable for consumption. Fish of average quality have a score in the 11–15 range. Fish in this stage may have noticeable defects in one or more aspects of appearance, odor, texture, or taste. They may not be as fresh or appealing as fish with lower scores, but they are still edible. Fish that are not fit for consumption have a score in the 16–20 range. These are fish with significant defects or signs of spoilage. These fish may have strong off odors, mushy texture, or unpleasant flavors, indicating advanced deterioration. They are generally not recommended for consumption and may need to be discarded.

TABLE I. SENSORY EVALUATION TABLE FOR FISH FRESHNESS [4, 5]

Quality Parameter	Character	Score
General Appearance	Skin	0—Bright, shining 1—Bright 2—Dull
	Bloodspot on gill cover	0—None 1—Small, 10–30% 2—Big, 30–50% 3—Very big, 50–100%
	Stiffness	0—Stiff, in rigor mortis 1—Elastic 2—Firm 3—Soft
	Belly	0—Firm 1—Soft 2—Belly burst
	Smell	0—Fresh, seaweed/metallic 1—Neutral 2—Musty/sour 3—Stale meat/rancid
Eyes	Clarity	0—Clear 1—Cloudy
	Shape	0—Normal 1—Plain 2—Sunken
Gills	Color	0—Characteristics, red 1—Faded, discolored
	Smell	0—Fresh, seaweed/metallic 1–Neutral 2—Musty/sour 3—Stale meat/rancid
Sum of Scores	Min. 0, max. 20	

The identified gaps in fish freshness assessment research include the need for classifiers with multiple detailed levels of freshness, the development of classifiers that can assess the freshness of fish species with differing eye characteristics, and the creation of a dataset that comprehensively captures fish deterioration across different freshness levels. Addressing these gaps will contribute to advancing fish freshness assessment techniques and their practical applications in the food industry.

II. LITERATURE REVIEW

In recent times, the utilization of advanced technologies such as Deep Learning (DL) and Machine Learning (ML) has exhibited great promise in improving the detection of fish freshness. DL and ML enable computers to efficiently process and analyze extensive data, enabling the creation of accurate and automated models that can interpret intricate patterns related to diverse fish quality attributes. These technologies represent a significant shift, offering efficient, cost-effective, and consistent methods for evaluating fish freshness. Contemporary machine learning algorithms like Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have displayed encouraging outcomes in assessing fish freshness [7]. Furthermore, image-based segmentation approaches utilizing methods such as K-means clustering, and wavelet transformations have demonstrated effectiveness in identifying pertinent features for precise quality assessment [8]. It is important to note that these technologies have the potential to transform the seafood industry significantly, ensuring safer consumption and optimizing resource utilization.

Different machine learning approaches have been developed and refined for fish classification and freshness assessment, drawing from various preprocessing techniques and model architecture. Delineated preprocessing steps encompass image segmentation through rotation, cropping, masking, K-means clustering, and morphology, particularly for gill segmentation [9]. These preprocessing steps were crucial for enhancing the input data quality and facilitating subsequent feature extraction. Feature extraction primarily leveraged Red, Green, and Blue (RGB) feature extraction and RGB values, Lab* values, and delta E, c* values [9–11] as the features provided a rich representation of the visual characteristics of the fish images, enabling effective discrimination between different fish species and assessing freshness levels.

Various ML models were employed, including feed-forward neural networks (ANNs), regression models, and classifiers such as SVM, random forest, and Naïve Bayes [11–16]. These models exhibited high accuracy rates, ranging from 84% to 100%, depending on the task, such as fish classification or freshness assessment.

Model performance was augmented through resolution downscaling, median filtering, thresholding, and open-close filtering [2, 15–17]. These preprocessing and enhancement techniques contributed to refining the input data and reducing noise, ultimately enhancing the robustness and accuracy of the ML models.

Several image processing techniques, including histogram equalization, thresholding, and blob extraction, were integrated into the workflow [2, 12, 14]. These techniques played a vital role in enhancing image quality, isolating relevant features, and improving model interpretability. The machine learning framework exhibited impressive performance metrics across multiple evaluation criteria, including accuracy, precision, sensitivity, specificity, and Area Under the Curve (AUC), as outlined in references [2, 13]. These metrics underscored the efficacy of the proposed approach in accurately classifying fish species and assessing their freshness levels.

The machine learning presented a systematic and robust fish classification and freshness assessment methodology, integrating various preprocessing techniques, feature extraction methods, and model architectures. The comprehensive evaluation across various performance metrics demonstrated the effectiveness and reliability of the proposed approach in addressing the targeted objectives.

A. Image Acquisition

In training deep learning and machine learning models, a good quality dataset is necessary for accurate results. Two main options for building this dataset include (1) gathering fisheye images online and (2) building a dataset by taking images of fish at different time intervals to get different levels of freshness. In the context of fish freshness, however, it is important for objectivity in the labeling of freshness levels, and the researchers have no objective way to ensure the appropriate labelling of freshness. Therefore, building a dataset appears to be the best approach to acquire images of fisheyes. This is often done by either gathering fish samples from the market [18] or farms [13, 19], then taking their photos in a proper setting to ensure uniformity in lighting. Additionally, as deep convolutional networks normally use RGB images as their input [13, 18], the images taken must also be in RGB. Another viable option for image acquisition is to use available datasets like [20].

B. Preprocessing

Regardless of the model that will take the image data as input, it is important to ensure the quality of the dataset by performing image preprocessing. This includes the basics, such as the removal of unusable images, cropping, rotation, resizing, and normalization. Some studies, like that of Taheri-Garavand *et al*. [13] and Lalabadi *et al*. [11] utilize other image preprocessing techniques like noise removal filtering and grayscale conversion. It is important to note that they were done in preparation for manual feature extraction, and thus, studies that will rely on manual feature extraction for feature engineering can greatly benefit from performing such preprocessing methods.

C. Data Augmentation

Volume is another factor in image preparation. Deep convolutional images consider many factors, so it is important to have a large number of images in order to avoid overfitting [13, 18, 21]. Overfitting is seen in a model performing perfectly with training data but poorly on testing data, and it can happen when the dataset is too small. Therefore, the dataset must be relatively large achieved through processes like photographing a lot of samples or augmenting the images. Augmentation is the process of making slight modifications on the dataset to artificially increase its size.

D. Image Segmentation

While the images in the dataset may already be cropped, it is possible that the background of the Region of Interest (ROI), which in this context is the fisheye, is inconsistent and potentially affecting the details of the ROI. Therefore, image segmentation, or the division of an image into segments for the purposes of only processing the ROI, is often performed during image processing. Some techniques include clustering, thresholding, and watershed transformations.

E. Feature Extraction

Feature extraction takes the characteristics of the images (such as color, texture, a shape), and turns them into features that will be the input of the models in the classification phase. In the context of fisheye images, color and texture are important as they both change as the eye of a fish changes in appearance as it reduces in freshness. Color values, such as RGB, HSI, and CIE Lab can be used as features or as ingredients of features [18]. The same treatment can be used on texture values like the Gray Level Co-occurrence Matrix (GLCM). Feature extraction can also be performed by deep learning models as it learns from the dataset.

F. Classification

Classifiers lead to the proper labeling of individual records from the dataset based on their respective features. To achieve this, classification has two major processes: modeling, which is the process of training a model to learn about a specific dataset whose labels are provided, and validation, which is the process of testing the model if it has correctly learned the data by giving it a subset of data that it has yet to see, and letting it label that data [18]. Classifiers can include machine learning and data learning models, which are both possible models for classifying fish freshness.

Building on these observations, our study contributes in four key ways: (1) we construct a labeled dataset of fish eye images by capturing progressive stages of spoilage across two commonly consumed species; (2) we apply tailored preprocessing and segmentation to isolate the region of interest (fish eyes) for feature extraction; (3) we extract interpretable features—RGB, CIE Lab, and GLCM texture descriptors—chosen for their relevance to visual freshness indicators; and (4) we evaluate three machine learning classifiers (Naïve Bayes, KNN, and SVM), selected for their interpretability and efficiency with limited data. These decisions are guided by both empirical insights from prior studies and the practical constraints of low-cost, scalable freshness assessment systems. The methodological details are discussed in the following section.

III. Methodology

This study aims to evaluate the performances of Naïve Bayes, Support Vector Machine (SVM), and K-Nearest

Neighbors (KNN) classifiers in classifying two different species based on features from fisheye images. The structured and systematic framework for acquiring, categorizing, and analyzing fisheye images is shown in Fig. 1 and is expounded on in this section.

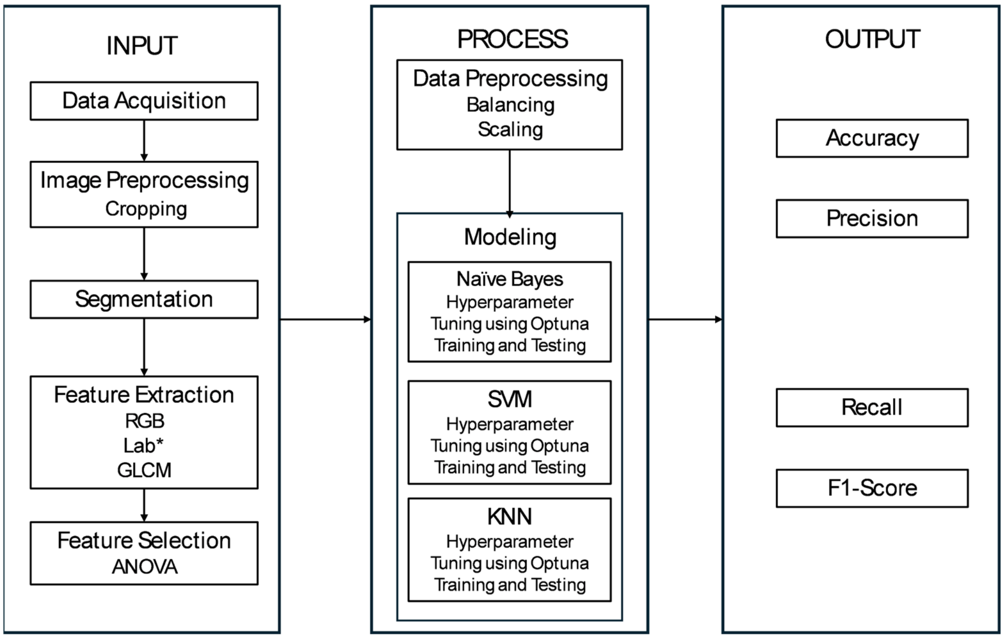


Fig. 1. Conceptual framework.

A. Sample Acquisition

The species chosen comprises two of the most common fish seen in the Philippines' crowded markets. These species, namely milkfish (*Chanos chanos*) and tilapia (*Oreochromis niloticus*), were chosen both because they have differing eye characteristics and because they are widely available in Philippine markets. The study employed a total of twenty fish samples, ten per species. These fishes were acquired at a local market near the image acquisition site and transported in Styrofoam boxes with ice. Each sample was laid out on their respective locations on trays, with labels to identify each fish and associate scores with each fish during all of the image acquisition times. Table II lists the labels that were assigned to the images based on the sensory evaluation score of the fish. The fish's eyes discolored throughout this time, and other changes to their appearance also occurred due to natural decomposition processes that had transpired. These variations in eye condition are the basis of the predictor variables this study uses to determine fish freshness.

TABLE II. SENSORY EVALUATION SCORES AND THEIR CORRESPONDING LABELS

Sensory evaluation score range	Label
0–5	Excellent
6–10	Good
11–15	Average
16–20	Not fit for consumption

B. Setup for Image Acquisition

A controlled environment ensures the consistency and quality of acquired high-quality fisheye images. The choice of this controlled setting includes a custom-designed box equipped with an internal light to maintain standardized and consistent illumination throughout the image acquisition process. The selected image capture tool is the smartphone camera of a Samsung Galaxy A54, enabling the capture of sharp and clear images of the fisheyes. To eliminate potential sources of distortion and enhance image clarity, the fish were carefully placed on a clean, plain, non-reflective background to minimize distractions and reflections that could compromise the image quality. To maintain stability and ensure consistent framing during image capture, the phone was secured onto the lid of the box, aligned with the hole created for the phone's cameras. When not in the box, the samples were on their respective trays.

C. Image Acquisition

In image acquisition, we captured images from both sides of each fish. In the image acquisition phase, we captured photographs from both sides of each fish to train the machine learning model for accurate freshness assessment, enabling the tracking of changes in fisheye appearance over time. The primary objective was to document different stages of deterioration by capturing one photo of each sample's side per hour, resulting in two images per sample per hour. This procedure was repeated hourly over a 21-hour period, yielding 44 photographs per sample starting from the 0th hour. While the dataset was limited to 10 samples per species due to logistical and resource constraints, this setup allowed for highly controlled image acquisition and consistent monitoring across time points. Therefore, 880 base images resulted from this meticulous image acquisition procedure. Fig. 2

shows the steps each sample underwent for every hour of the image acquisition.

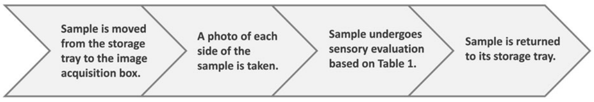


Fig. 2. A sample's hourly journey during image acquisition.

The dynamic changes in fisheye appearance throughout the 21-hour period were thoroughly documented in these photographs, which serve as essential visual records of spoilage progression. Such systematic documentation is vital for analyzing freshness fluctuations. In addition to image capture, it was crucial to maintain detailed records of relevant metadata. After imaging, each sample was transferred to the sensory evaluation station, where researchers assessed freshness. Individual scores were recorded, grouped by sample ID, and compiled into a spreadsheet to calculate the total score and assign a freshness label based on the scale in Table III: "excellent", "good", "fair to average", and "not fit for consumption". These labels, along with the images and raw scores, were reviewed and validated by experts from the Department of Aquatic Post-Harvest under the Central Luzon State University–College of Fisheries to establish ground truth. This integrated metadata collection process facilitates systematic categorization and robust analysis of the dataset.

TABLE III. DEFINITIONS OF DIFFERENT FRESHNESS LEVELS [6]

Sensory evaluation score range	Freshness level	Description
0–5	Excellent	Fish with very few or no noticeable defects in appearance, odor, texture, or taste. These fish are considered fresh and highly desirable for consumption or commercial purposes.
6–10	Good	Minor defects may be present, but overall, the fish is still of acceptable quality. These fish may have slight blemishes or a slightly less-than-ideal odor, texture, or taste, but they are still suitable for consumption.
11–15	Fair to average	Fish in this stage may have noticeable defects in one or more aspects of appearance, odor, texture, or taste. They may not be as fresh or appealing as fish with lower scores, but they are still edible.
16–20	Not fit for consumption	Fish with significant defects or signs of spoilage. These fish may have strong off odors, mushy texture, or unpleasant flavors, indicating advanced deterioration. They are generally not recommended for consumption and may need to be discarded.

TABLE IV. SAMPLE COMPUTATIONS FOR THE MINIMUM AND MAXIMUM SCORES PER LABEL FOR TILAPIA

Freshness	Total Score	Image	General Appearance—Skin	General Appearance—Bloodspot on gill cover	General Appearance—Stiffness	General Appearance—Belly	General Appearance—Smell	Eyes—Clarity	Eyes—Shape	Gills—Color	Gills—Smell
Excellent	0		0	0	0	0	0	0	0	0	0
Excellent	5		1	0	1	1	1	0	0	0	1
Good	6		2	0	1	1	1	0	0	0	1
Good	10		2	1	2	1	1	1	0	1	1
Fair to average	11		2	1	2	1	1	1	1	1	1
Fair to average	15		2	1	2	1	1	1	1	1	1
Not fit for consumption	16		2	2	2	1	2	1	1	1	3
Not fit for consumption	19		2	2	3	1	2	1	1	1	3

Table IV shows the computed freshness scores for Tilapia range from 0 to 19, with each category capturing a spectrum of quality degradation. A fish with a total score of 0 exhibits no signs of deterioration, characterized by bright, shining skin, absence of bloodspots, stiff flesh in rigor mortis, a firm belly, fresh seaweed-like smell, clear and normally shaped eyes, and red, odor-free gills. However, as the score reaches 5, subtle changes begin to appear—the skin may lose some brightness, stiffness may transition to slight elasticity, the belly may soften, and a faint neutral smell might develop. In the "good" category (6–10), the fish's condition continues to decline, with duller skin, the emergence of small bloodspots (10–30% coverage), increasing elasticity, and occasional cloudiness in the eyes. By 10, the stiffness may become firm rather than elastic, and the gills may start to fade in color. Moving into the "fair to average" category (11–15), the degradation becomes more pronounced; fish at 11 may show larger bloodspots (30–50%), sunken eyes, and a musty or sour odor, while those closer to 15 may have widespread bloodspots, increased flesh softness, and a stronger stale smell. In the "not fit for consumption" category (16–19), fish at 16 exhibit very large bloodspots (50–100%), fully softened flesh, and discolored, foul-smelling gills, while a score of 19 represents near-complete spoilage, with heavily sunken eyes, a burst belly, rancid odor, and severe loss of characteristic red gill color. This detailed scoring system allows for a nuanced, objective assessment of fish freshness, ensuring that even small degradations are systematically captured.

TABLE V. SAMPLE COMPUTATIONS FOR THE MINIMUM AND MAXIMUM SCORES PER LABEL FOR MILKFISH

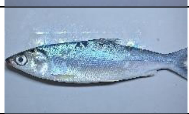
Freshness	Total Score	Image	General Appearance —Skin	General Appearance —Bloodspot on gill cover	General Appearance —Stiffness	General Appearance —Belly	General Appearance —Smell	Eyes—Clarity	Eyes—Shape	Gills—Color	Gills—Smell
excellent	0		0	0	0	0	0	0	0	0	0
excellent	5		1	0	1	1	0	0	2	0	0
good	6		1	0	2	1	0	0	2	0	0
good	10		2	1	2	1	0	1	2	1	0
fair to average	11		2	1	3	1	0	1	2	1	0
fair to average	15		2	2	3	1	1	1	2	1	2
not fit for consumption	16		2	2	3	1	2	1	2	1	2
not fit for consumption	19		2	3	3	2	3	1	1	1	3

Table V shows computations for the minimum and maximum scores that a selection of our milkfish samples received, starting from 0 to 19. For the lowest score for the "excellent" label, the sample was evaluated with all 0's, meaning that there were no defects in the sample's appearance, odor, texture, or taste. The maximum total score for excellent is 5, which showed slight degradations in the different parts of the fish—it skin looked dull, the flesh was firm, the belly felt stiff, and the eye shape was plain. The good category (6–10), the degradation furthers in more aspects. For the score of 6, the difference between 5 was that its flesh felt soft in addition to the same degradations. For 10, the skin was dull, a small bloodspot (10–30%) appeared on the gill cover, its flesh firm and belly soft, its eyes were cloudy and sunken, and its gills faded and discolored. For "fair to average" (11–15), the starting score of 11's difference from that of 10's was attributed to the flesh being soft, represented by the worst score a sample could get for the fish's stiffness. The max score of 15 had dull skin, big bloodspot on the gill cover (30–50%), soft flesh and belly, a neutral smell, cloudy and sunken eyes, faded and discolored gills, and a neutral smell. The "not fit for consumption" category, represented here by scores 16–19, starts with the addition of a musty/sour smell and ends with a burst belly and a rancid smell in addition to previous degradations.

TABLE VI. SCORE BREAKDOWN OF MILKFISH

Freshness Level	Total Score	Description
excellent	0	No degradation evaluated.
	1	Samples have elastic stiffness.
	2	Samples have either elastic stiffness or soft belly.
	3	Samples have elastic stiffness and either soft belly or faded/discolored gills.
	4	Samples have elastic stiffness, soft belly, and sunken eyes.
	5	Samples have bright but not shining skin, elastic stiffness, soft belly, and sunken eyes.
good	6	Samples have bright but not shining skin, firm stiffness, soft belly, and sunken eyes.
	7	Samples have dull skin, firm stiffness, soft belly, and sunken eyes.
	8	Samples have dull skin, firm stiffness, soft belly, sunken eyes, and either small (10–30%) bloodspot on gill cover or faded/discolored gills.
	9	Samples have dull skin, small (10–30%) bloodspot on gill cover, firm stiffness, sunken eyes, and either cloudy eyes or faded/discolored gills.
	10	Samples have dull skin, small (10–30%) bloodspot on gill cover, firm stiffness, soft belly, cloudy and sunken eyes, and faded/discolored gills.
fair to average	11	Samples have dull skin, small (10–30%) bloodspot on gill cover, firm or soft stiffness, soft belly, cloudy and sunken eyes, and either neutral smell overall and in gills or faded/discolored gills.
	12	Samples have dull skin, small (10–30%) bloodspot on gill cover, firm stiffness, soft belly, cloudy and sunken eyes, neutral smell overall and in gills and faded/discolored gills.
	13	Samples have dull skin, small (10–30%) bloodspot on gill cover, soft stiffness, soft belly, cloudy and sunken eyes, neutral smell overall and in gills and faded/discolored gills.
	14	Samples have dull skin, small (10–30%) or big (30–50%) bloodspot on gill cover, soft stiffness, soft belly, cloudy and sunken eyes, neutral smell overall, either neutral or musty/sour smell in gills and faded/discolored gills.
	15	Samples have dull skin, big (30–50%), or very big (50–100%) bloodspot on gill cover, soft stiffness, soft belly, cloudy and sunken eyes, either neutral or musty/sour smell overall, musty/sour smell in gills and faded/discolored gills.
not fit for consumption	16	Samples have dull skin, very big (50–100%) bloodspot on gill cover, soft stiffness, soft belly, cloudy and sunken eyes, musty/sour smell overall and in gills, faded/discolored gills, and cloudy eyes that are either sunken or plain.
	17	Samples have dull skin, very big (50–100%) bloodspot on gill cover, soft stiffness, either soft or burst belly, cloudy eyes that were either sunken or plain, musty/sour smell overall and in gills, and faded/discolored gills.
	18	Samples have dull skin, very big (50–100%) bloodspot on gill cover, soft stiffness, soft or burst belly, cloudy and plain eyes, stale/rancid smell overall and in gills, faded/discolored gills.
	19	Samples have dull skin, very big (50–100%) bloodspot on gill cover, soft stiffness, soft or burst belly, cloudy and either plain or sunken eyes, stale/rancid smell overall and in gills, faded/discolored gills.

By the 21st hour of image acquisition, the largest score that the milkfish samples received was 19. For each score computed, Table VI outlines the most common attributes that the samples have. At the "excellent" level (scores 0–5), samples showed minimal to no degradation, maintaining elastic stiffness, soft bellies, and clear eyes, with minor signs such as faded gills appearing at higher scores. The "good" level (scores 6–10) showed more deterioration, through the samples' dull skin, sunken eyes, and small bloodspots on the gill cover, with some samples also having cloudy eyes or faded gills. In the "fair to average" category (scores 11–15), the milkfish samples displayed more greater degradation, such as larger bloodspots, cloudy and sunken eyes, and neutral to musty/sour odors overall and specifically in the gills. Finally, the milkfish samples classified as not fit for consumption (scores 16–19) exhibit severe degradation, including very large bloodspots, burst bellies, plain or sunken eyes, and strong stale or rancid odors overall and in the gills.

Table VII shows the descriptions of the tilapia samples at the different scores they received based on the attributes that most of the samples exhibited at each hour. The maximum score they received by the 21st hour is 19. The "excellent" scores (0–5) were given to tilapia samples that showed minimal signs of degradation, with attributes such as elastic stiffness, soft belly, and neutral smell. Some samples may have slightly diminished freshness in one or more aspects, like no longer having shining skin or a firm belly. The "good" scores (6–10) were given to tilapia samples with dull skin, firm or elastic stiffness, soft bellies, and neutral smells, with increasing signs of deterioration like small bloodspots on the gill cover, faded gills, or cloudy eyes at higher scores. The "fair to average" scores (11–15) indicated more noticeable decline, with defects becoming widespread, including larger bloodspots, musty or sour odors overall and in the gills, and cloudy or plain eyes. Finally, tilapia samples scored with "not fit for consumption" scores (16–19) exhibit severe spoilage, such as very large bloodspots, soft or burst bellies, stale or rancid smells overall and in the gills, cloudy and sunken eyes, and heavily faded or discolored gills with strong odors.

D. Preparation of Images

To alleviate the additional computational burden associated with potentially processing images of differing dimensions, the images were programmatically standardized. The images were standardized to a size of 224×224 pixels. This size was chosen as it strikes a balance between preserving essential features and reducing computational overhead. By cropping images to this uniform size, extraneous details were removed, focusing the analysis on the relevant fish features.

TABLE VII. SCORE BREAKDOWN OF TILAPIA

Freshness Level	Total Score	Description
Excellent	0	No degradation evaluated.
	1	These samples possess one attribute that has diminished freshness, usually one of the following: their general stiffness being described as elastic and no longer in rigor mortis, their belly being soft and not firm, or their skin being bright but no longer shining.
	2	Samples have mostly elastic stiffness and a soft belly.
	3	Samples have a combination of three of the following: elastic stiffness, soft belly, bright but not shining skin, and plain eye shape.
	4	Samples have a combination of bright or dull skin, elastic or firm stiffness, soft belly, neutral overall smell.
	5	Samples have bright but not shining skin, soft belly, neutral overall and gill smell, and elastic stiffness.
Good	6	Samples mostly have dull skin, either elastic or firm stiffness, soft belly, and neutral smell both overall and specifically in the gills.
	7	Samples have dull skin, firm stiffness, soft belly, and neutral smell both overall and specifically in the gills.
	8	Samples have dull skin, firm stiffness, soft belly, neutral overall and gill smell, and either plain eyes or small (10–30%) bloodspot on gill cover.
	9	Samples have dull skin, firm stiffness, soft belly, neutral overall and gills smell, faded/discolored gills, and small (10–30%) bloodspot on gill cover.
	10	Samples have dull skin, firm stiffness, soft belly, cloudy eyes, faded/discolored gills, small (10–30%) bloodspot on gill cover, and neutral smell overall and specifically in the gills.
Fair to average	11	Samples now have defects across the board. They have dull skin, small (10–30%) bloodspot on gill cover, firm stiffness, soft belly, neutral overall and gills smell, cloudy and plain eyes, and faded/discolored gills.
	12	Samples have dull skin, small (10–30%) bloodspot on gill cover, firm stiffness, soft belly, neutral overall smell, cloudy and plain eyes, faded/discolored gills, and musty/sour gills smell.
	13	Samples have dull skin, firm stiffness, soft belly, cloudy and plain eyes, faded/discolored gills, musty/sour smelling gills, and either musty/sour overall smell with small (10–30%) bloodspot on gill cover, or big (30–50%) bloodspot on gill cover with neutral overall smell.
	14	Samples have dull skin, big (30–50%) bloodspot on gill cover, firm stiffness, soft belly, musty/sour overall smell and gills smell, cloudy and plain eyes, and faded/discolored gills.
	15	Samples have dull skin, big (30–50%) bloodspot on gill cover, firm stiffness, soft belly, musty/sour overall smell, cloudy and plain eyes, faded/discolored gills that smell stale/rancid.
Not fit for consumption	16	Samples have dull skin, big (30–50%) bloodspot on gill cover, firm or soft stiffness, soft belly, musty/sour smell, cloudy and plain eyes, faded/discolored gills that smell stale/rancid.
	17	Samples have dull skin, big (30–50%) bloodspot on gill cover, soft stiffness, soft belly, either musty/sour or stale/rancid overall smell, cloudy and plain eyes, and faded/discolored gills that smell stale/rancid.
	18	Samples have dull skin, big (30–50%) or very big (50–100%) bloodspot on gill cover, soft stiffness and belly, stale/rancid smell overall and specifically in the gills, cloudy and sunken eyes, and stale/rancid smelling gills that are faded/discolored.
	19	Samples have dull skin, very big (50–100%) bloodspot on gill cover, soft stiffness, burst belly, stale/rancid overall and gills smell, cloudy and sunken eyes, and faded/discolored gills.

E. *Segmentation*

Roboflow was utilized for the segmentation process. We manually annotated the fisheye images by segmenting the eye regions to create accurate training data for our model, which also resulted in bounding boxes being generated shown at the Fig. 3. These annotations generated detailed JSON data containing both bounding box coordinates and segmentation masks, which serve as essential inputs for training our machine learning models to assess fish freshness effectively. This careful segmentation ensures the model focuses on the most relevant eye features, enabling it to learn and distinguish between different freshness levels with precision.

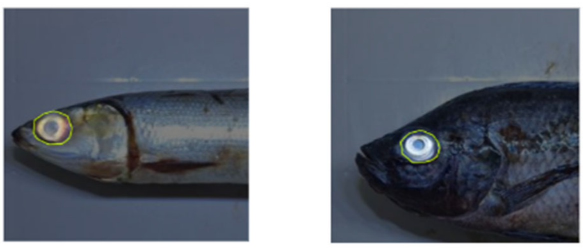


Fig. 3. Segmented images of Milfish and Tilapia.

F. *Feature Extraction*

The features employed in this study include color features, namely RGB and Lab* features, and texture features from the Gray Level Cooccurrence Matrix (GLCM). Prior research has shown that these features are useful to assess fish freshness and quality due to their specific advantages in capturing relevant visual and textural information. RGB is utilized by studies such as [r / 20] because its computation and interpretation are simple. This color space directly represents raw color information captured by cameras, making it straightforward to analyze changes in the fish eyes that indicate freshness. The Lab* space separates luminance (L*) from chromaticity (a* and b*), which ensures that color descriptors are not confounded by illumination variations. Features such as a* and b* components proved useful for capturing color shifts associated with spoilage, such as increased redness, as shown in Taheri-Garavand *et al*. [13] and Lalabadi *et al*. [11]. Taheri-Garavand *et al*. [13] and Nguyen *et al*. [22] also emphasize that textural changes are induced by spoilage, which GLCM features can reliably interpret. These features are fast to compute and do not require complex processing or data processing pipelines. The combination of these three provides a holistic assessment covering visual appearance and the structural integrity of tissue. The JSON data from segmentation included a part called "segmented pixels", which are the pixels in the segment shown as RGB values. To prepare these values for the model, the average was taken and added to a different column. The raw RGB values were preserved to be converted to the Lab* and GLCM values.

In our Lab conversion process, we transformed the segmented RGB pixels into the CIE Lab* color space using the rgb2lab function from the skimage color library. This conversion is crucial because the Lab* color space is designed to closely match human visual perception, making it particularly useful for detecting subtle color differences that are less discernible in the standard RGB space. After conversion, we computed the mean values for each channel—mean L*, mean a*, and mean b*—across the segmented pixels. These mean values serve as key features for our machine learning models, as they reflect the nuanced color changes in fisheyes that occur over time.

Finally, we applied Gray Level Co-occurrence Matrix (GLCM) analysis to capture texture-based features from the segmented fisheye images. First, we converted the RGB eye images into grayscale using the rgb 2 gray function, as GLCM works on intensity values rather than color. The grayscale image was then transformed into an 8-bit unsigned integer format, which is a common requirement for GLCM computation. GLCM was calculated with a distance of 1 pixel and across four standard angles (0°, 45°, 90°, and 135°), which allows the model to analyze texture patterns in multiple directions.

G. Feature Selection

The resulting number of features from the feature extraction processes was twenty-two in total. We tested different feature selection approaches and found the best results with ANOVA, which stands for Analysis of Variance, as it allowed us to check the variance within the features and between them and evaluate the importance of each feature. To reduce the dimensionality of the dataset and improve model performance, we utilized ANOVA to identify the top ten features that had an impact on the target variable, freshness level, based on the feature's F-statistic. As a result, the dataset retained the 'mean_R', 'mean_b', 'mean_L', 'mean_a*', 'mean_G', 'contrast_90', 'contrast_45', 'contrast_135', 'correlation_90', and 'homogeneity_90'** features, listed below in decreasing order of F-statistic, along with their respective purposes.

H. Data Preparation for Modeling

With the selected features, the next step was preparing the data for model training. To ensure a balanced representation of each class in the target variable, we first oversampled the dataset. The imbalance in class distribution was addressed by matching the number of samples in each class to the highest count, as indicated by the maximum value in the label column. The resulting balanced dataset was then scaled using Standard Scaler, which standardized the features to have a mean of 0 and a standard deviation of 1. This was essential, particularly for models like KNN and SVM, where distance metrics are sensitive to the scale of the input features. Finally, we split the dataset into training and test sets, with 80% used for training and 20% reserved for testing.

I. Model Development and Evaluation

1) Naïve Bayes

Naive Bayes is a probabilistic classifier based on Bayes' theorem, which assumes that the features are conditionally independent given the class label. This assumption allows the model to compute the posterior probabilities of each class and choose the class with the highest probability. For continuous features, Naive Bayes assumes that they follow a Gaussian distribution, and it estimates the mean and variance for each class.

In order to improve the model's accuracy, we used Optuna to optimize the hyperparameters. Optuna is an automatic hyperparameter optimization framework that uses an efficient sampling strategy to explore the hyperparameter search space and find the best parameters. For Naive Bayes, we focused on optimizing the var_smoothing parameter, which helps stabilize the variance estimation, especially in the case of small or zero variance values. The var_smoothing parameter was set to a range of values using np.logspace(0, −9, num = 100), which generates 100 logarithmically spaced values between 10^0 (1) and 10^{-9}. Var smoothing adds a small value to the variance of each feature, which helps avoid numerical issues and also reduces the model's tendency to overfit by not placing excessive reliance on small variances.

We ran 500 trials in Optuna to identify the best value for this parameter. After optimizing the hyperparameter, we trained the model and then tested its performance on the test set, calculating its accuracy and generating a classification report to assess the precision, recall, F1-Score, and other metrics.

2) k-Nearest Neighbors (KNN)

K-Nearest Neighbors is a non-parametric algorithm that classifies data based on the majority class among its k-Nearest Neighbors. It computes the distance between the test point and training samples, then predicts the class of the test point by considering the k closest training samples. The performance of KNN heavily depends on the choice of hyperparameters. Therefore, Optuna was used to tune several important parameters.

First, the n_neighbors parameter was varied between 1 and 30 to determine the optimal number of neighbors. Too few neighbors might cause overfitting, while too many could smooth the decision boundary too much. Additionally, we tuned the weights parameter, with options for uniform (equal weight for all neighbors) and distance (closer neighbors have more weight). The algorithm parameter, which controls the method used to find the nearest neighbors, was tested with auto, ball_tree, kd_tree, and brute options. Each algorithm has different computational efficiencies, and the optimal choice depends on the data's structure. Finally, we adjusted the leaf_size (between 1 and 100) to control the size of the leaf nodes in tree-based algorithms, and the p parameter (1 for Manhattan distance, 2 for Euclidean distance) was fine-tuned to assess its effect on the model's accuracy.

After running 500 trials in Optuna to find the best combination of these parameters, we trained the KNN model. We then tested the model by predicting the fish freshness, calculated its accuracy, and generated the classification report, which helped assess the model's effectiveness.

3) Support Vector Machine (SVM)

Support Vector Machine is a powerful classification algorithm that works by finding the hyperplane that best separates the classes in the feature space. SVM can handle both linear and non-linear decision boundaries, depending on the choice of kernel. Similar to the previous models, Optuna was used to fine-tune several SVM hyperparameters, beginning with the *C* parameter, which controls the regularization of the model. The *C* parameter varied between 10^{-3} and 10^{3} allowing us to balance the margin maximization and classification error minimization. A high C value places more emphasis on minimizing errors, while a smaller value allows for a larger margin and potentially better generalization. The kernel parameter, which determines the type of decision boundary, was tested with different values: linear, poly, rbf, and sigmoid. For non-linear data, the rbf kernel is particularly effective in mapping data to higher-dimensional spaces where a linear boundary can separate the classes. Additionally, the gamma parameter, which influences the influence of individual data points on the decision boundary, varied between scale and auto. A higher gamma value allows the model to create more complex decision boundaries, while a lower value leads to a simpler, smoother boundary. Finally, we optimized the degree parameter, which controls the degree of the polynomial used in the poly kernel, adjusting it between 2 and 5.

After running 500 trials with Optuna to identify the optimal hyperparameters, we trained the SVM model. We then tested the model's predictive performance by calculating its accuracy and generating a classification report. This helped us evaluate the SVM model's overall performance in predicting fish freshness.

The classifiers selected for this study, Naïve Bayes (NB), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM), were chosen based on their suitability for small to moderately sized datasets, interpretability, and computational efficiency. These classical machine learning models are well-suited for scenarios where handcrafted features, such as color and texture descriptors, can be extracted from the data.

In contrast, deep learning models, particularly Convolutional Neural Networks (CNNs), generally require large-scale datasets to perform effectively and avoid overfitting. Training such models is resource-intensive, demanding significant computational power, longer training times, and greater data volume to generalize well. Given the controlled but limited dataset used in this study (880 images), deep learning approaches were deemed impractical.

Moreover, traditional machine learning models allow for feature engineering and targeted input selection, enabling the use of domain knowledge, such as the visual characteristics of fisheyes known to correlate with freshness. This controlled approach offers better model transparency and faster training, making NB, KNN, and SVM appropriate choices for evaluating the feasibility of automated fish freshness classification in resource-constrained or small-sample settings.

To address the risk of overfitting posed by the small dataset and the use of oversampling, several strategies were implemented, with a key emphasis on hyperparameter tuning using Optuna. By optimizing parameters such as 'var_smoothing' in Naive Bayes, 'n_neighbors' in KNN, and 'C', 'kernel', and 'gamma' in SVM, the model search prioritized configurations that reduced complexity and improved generalization. For instance, tuning 'var_smoothing' helped regularize low-variance features common in oversampled data, while adjusting 'n_neighbors' balanced KNN's sensitivity to noise versus its ability to generalize.

In SVM, controlling 'C' and 'gamma' allowed for a trade-off between margin width and decision boundary complexity. These hyperparameters directly influence model flexibility and were critical in preventing overfitting to synthetic or redundant patterns. Complementary to this, ANOVA-based feature selection removed irrelevant or noisy predictors, reducing the risk of the model fitting to spurious correlations. Oversampling was followed by scaling to prevent any one feature from dominating the learning process, and an 80/20 stratified split was used to maintain class distribution and enable a fair evaluation of model generalizability. Together, these steps formed a cohesive approach to mitigating overfitting despite the constraints of limited data.

IV. Results and Discussion

The analysis of fish samples using a combination of RGB values (segmented pixels), Lab color values, and texture features from the Gray-Level Co-occurrence Matrix (GLCM) reveals critical insights into the relationship between visual characteristics and fish freshness. The dataset consists of various fish images, each represented by segmented pixel values in RGB format, capturing the raw color information of the fish. These RGB values were further converted into the Lab color space to capture more precise color variations and enhance the differentiation between fresh and aging fish. Texture features such as contrast, correlation, energy, and homogeneity were also extracted using GLCM to quantify surface properties.

In terms of the hourly breakdown of the samples' freshness, Fig. 4 gives us insight into the points in time at which the fish start changing from "excellent" to "good" to "fair to average" to "not fit for consumption". For the 0th to 2nd hour, all the milkfish samples are graded within the "excellent" range. By the 3rd hour, half of the samples moved into the "good" category, with two more changing by the 4th hour. By the 5th hour, there were no more excellent examples. One sample moved into "fresh to average" by the 10th hour, then another moved into "not fit for consumption" by the 13th hour. Starting at the 18th hour, all samples were unfit for consumption.

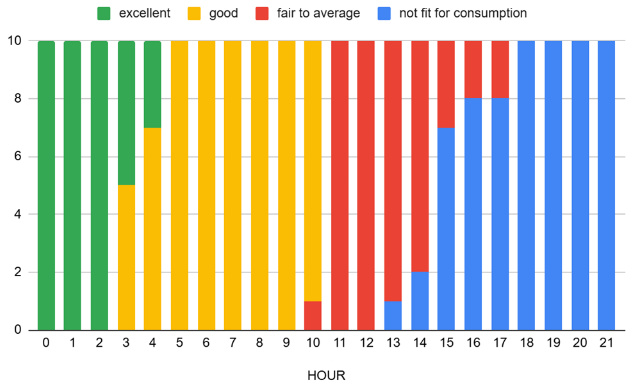


Fig. 4. Hourly freshness breakdown for milkfish.

Unlike the hourly breakdown of the milkfish samples' freshness, Fig. 5 shows that the tilapia samples only started moving into the "good" label at the 5th hour. The hour at which they moved into "fair to average" was the same as the milkfish samples', which is at the 10th hour. The tilapia samples started moving into the "not fit for consumption" category at the 15th hour. The tilapia samples stayed fresher for longer overall, but both species' samples were completely under the "not fit for consumption" category by the 18th hour. These differences between the times at which the milkfish samples started moving into less fresh categories vs the times at which the tilapia samples point to how there could be merit into looking into the inherent differences between the two species that could possibly affect when and how they deteriorate.

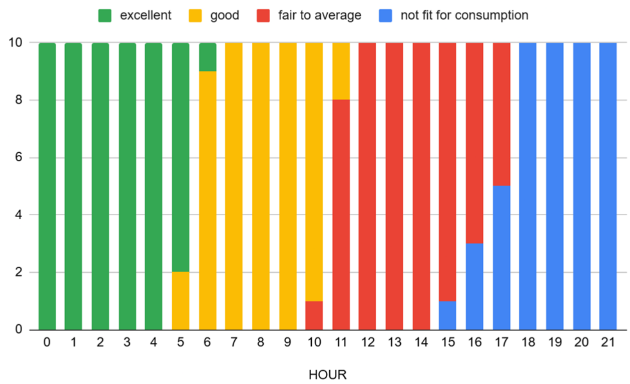


Fig. 5. Hourly freshness breakdown for tilapia.

A. Naive Bayes

In this study, the performance of a Gaussian Naive Bayes model was optimized using Optuna to predict fish freshness based on features derived from RGB values, Lab* values, and GLCM texture properties. The key hyperparameter of the Naive Bayes model, var_smoothing, was fine-tuned by conducting 500 optimization trials across a range of values. The optimal var_smoothing value suggested by Optuna was 0.0035, resulting in the best accuracy of 61.67% on the test dataset shown at Table VIII.

Fig. 6 shows the breakdown of the Naive Bayes model's performance on the dataset, with "excellent" and "good" getting the greatest number of accurate predictions. However, it seems like the model had some difficulty differentiating them from each other, with thirteen "excellent" records misclassified as "good", and 6 "good" records misclassified as "excellent". Five "good" records were also misclassified as "not fit for consumption". The model had the most difficulty with the "fair to average" label, with only 17 accurate predictions. It also had 17 misclassifications with "good" and 13 with "not fit for consumption". With 29 accurate predictions, the model had average performance with the "not fit for consumption" label, which was only misclassified as "fair to average".

TABLE VIII. NAIVE BAYES CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score
Excellent	0.82	0.7	0.75
Fair to Average	0.57	0.35	0.44
Good	0.52	0.72	0.61
Not Fit for Consumption	0.6	0.72	0.66
Accuracy			0.62

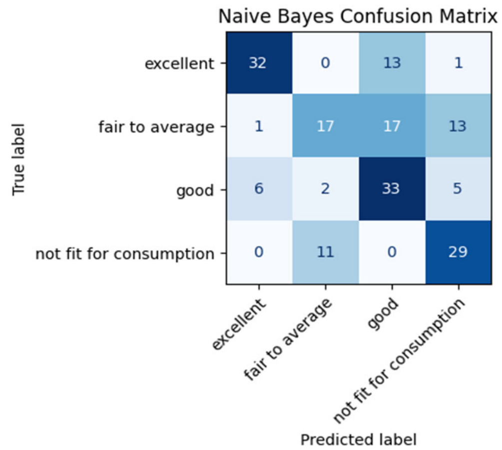


Fig. 6. Naïve Bayes confusion matrix.

B. k-Nearest Neighbors (KNN)

In the KNN classifier, it was optimized using Optuna to improve its ability to classify fish freshness based on image features. The optimization process focused on tuning key hyperparameters, including the number of neighbors (n_neighbors), weight function (weights), algorithm (algorithm), leaf size (leaf_size), and power parameter (p) for the Minkowski distance. After conducting 500 trials, the best set of hyperparameters was identified as 4 neighbors, a 'distance' weight function, the 'kd_tree' algorithm, a leaf size of 76, and a Minkowski distance parameter $p = 3$. Therefore, Table IX showed the accuracy of 0.89.

TABLE IX. KNN CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score
Excellent	0.93	0.87	0.9
Fair to Average	0.88	0.9	0.89
Good	0.81	0.93	0.87
Not Fit for Consumption	1	0.88	0.93
Accuracy			0.89

The breakdown of the KNN's predictions can be seen in Fig. 7, which shows that the model had more accurate predictions with the "good" label, followed by "fair to average", which was misclassified with only "good". The "excellent" label is next with 40 accurate predictions. These three labels had no misclassifications with "not fit for consumption", which is the label with which the model had mediocre performance, with 35 accurate predictions and some misclassifications.

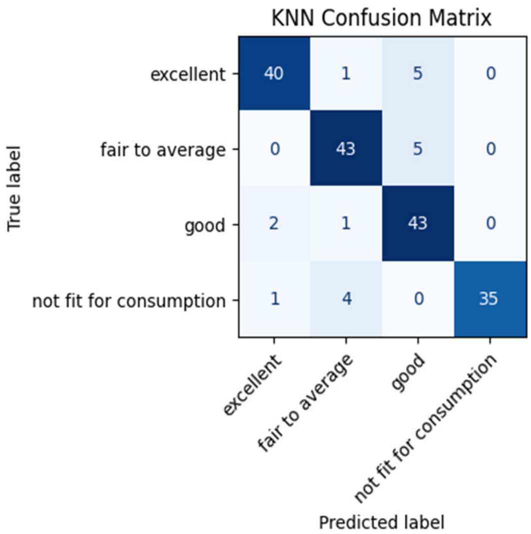


Fig. 7. KNN confusion matrix.

C. Support Vector Machine

The Support Vector Machine (SVM) model was optimized using Optuna to enhance its classification performance for fish freshness labels. The optimization process focused on tuning key hyperparameters, including the regularization parameter (C), kernel type (kernel), kernel coefficient (gamma), and, where applicable, the degree of the polynomial kernel. A total of 500 trials were conducted to determine the optimal combination of hyperparameters for maximizing accuracy. The best configuration identified was an RBF kernel with a regularization parameter C of 131.80 and gamma set to 'scale' shown in Table X.

TABLE X. SVM CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score
Excellent	0.92	0.96	0.94
Fair to Average	0.85	0.83	0.84
Good	0.86	0.93	0.9
Not Fit for Consumption	0.91	0.8	0.85
Accuracy			0.88

In Fig. 8, the SVM model's predictions across labels are shown with "excellent" having 44 accurate predictions and only 2 misclassifications with the "good" label. For the "good" label, it was only misclassified thrice as "excellent", and the rest are accurate predictions. Here, we can see that the model is performing well in differentiating the differences between these two classes. The model also performed decently in the "fair to average" label, with 40 accurate predictions and misclassifications with "good" and "not fit for consumption." There were 32 accurate predictions for the "not fit for consumption" label, with 8 misclassifications. Overall, the SVM model showed good performance in identifying "excellent", "good", and "fair to average" labels, while "not fit for consumption" identification was subpar.

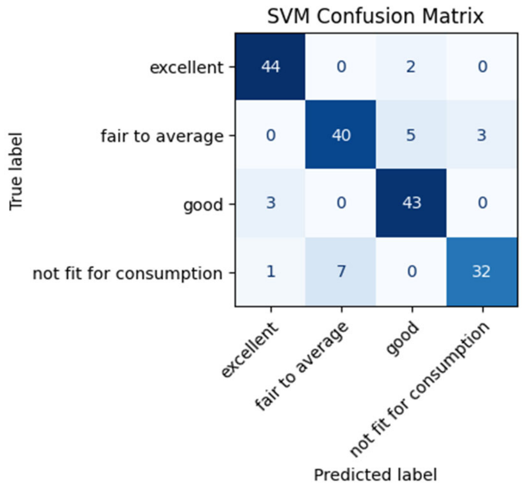


Fig. 8. SVM confusion matrix.

Although the models demonstrated strong performance in classifying the "Excellent" and "Good" freshness levels, their predictive accuracy was notably lower for the "Fair to Average" and "Not Fit for Consumption" categories. This discrepancy may be attributed to two primary factors. First, the visual features of fisheyes within the "Fair to Average" category tend to overlap with those of adjacent freshness levels. The subtle variations in color and texture may not be sufficiently distinguishable using the current set of handcrafted features, thereby contributing to classification ambiguity. Similar challenges in classifying intermediate freshness stages have been noted in previous studies where overlapping color and textural patterns reduced classification accuracy in non-destructive freshness monitoring of fisheyes [16].

Second, the dataset exhibited class imbalance, with fewer samples assigned to the "Fair to Average" category. This imbalance likely introduced bias during the training phase, resulting in reduced model sensitivity to minority classes and a higher misclassification rate for intermediate freshness states. The impact of class imbalance on model accuracy has also been emphasized in food freshness prediction studies, where underrepresented classes tend to be misclassified without appropriate balancing strategies [11, 12, 22].

Interestingly, previous studies using similar machine learning approaches have reported higher classification accuracies. For instance, Tolentino *et al.* [23] employed Support Vector Machine (SVM) to classify the freshness of milkfish, round scad, and short mackerel scad based on eye and gill redness using RGB features, achieving a 98% accuracy rate aligned with manual sensory evaluations. Similarly, Yudhana *et al.* [24] compared k-Nearest Neighbors (KNN) and Naïve Bayes (NB) classifiers for fish freshness detection and reported high accuracies of 97% and 94%, respectively. While our models achieved slightly lower accuracy, 89.44% for KNN, 88.33% for SVM, and 61.67% for NB, this can be attributed to differences in dataset size, species variability, image preprocessing, and the complexity introduced by four freshness levels rather than binary classification.

To address these limitations, several methodological improvements are proposed for future work. Data augmentation techniques, such as color jittering, contrast adjustment, or synthetic simulation, may be employed to enrich the diversity and representation of minority classes. Similar techniques have been successfully applied in prior image-based food quality studies to enhance model robustness and generalization [19]. Furthermore, the extraction of more granular texture and morphological features could improve the discriminative capacity of the models. The integration of ensemble methods or probabilistic classification approaches may also offer improved handling of ambiguous cases by enabling softer decision boundaries and reducing the risk of overfitting to dominant classes [16].

V. Conclusion

This study investigated the efficacy of various machine learning algorithms—namely Gaussian Naïve Bayes (NB), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—in classifying fish freshness based on features extracted from RGB, CIE Lab*, and Gray Level Co-occurrence Matrix (GLCM) values. By applying ANOVA for feature selection and balancing the dataset through oversampling, we effectively reduced dimensionality and addressed class imbalance, resulting in improved model accuracy. Among the models tested, KNN achieved the highest performance with an accuracy of 89.44%, followed by SVM at 88.33%. NB, while computationally efficient, achieved a more modest accuracy of 61.67%. Notably, SVM and KNN excelled in predicting "excellent" and "good" classes but struggled with lower precision in the "fair to average" and "not fit for consumption" categories—likely due to subtle visual differences and fewer samples per class.

To further enhance classification performance, several recommendations are proposed. First, class imbalance remains a key challenge. We suggest the use of targeted resampling strategies such as Synthetic Minority Over-sampling Technique or Adaptive Synthetic Sampling for undersampling. These methods can help increase the representation of underrepresented classes without overfitting. Additionally, class-weighted loss functions could be explored, especially in SVM or ensemble classifiers, to improve sensitivity to minority classes. Second, future models may benefit from a hierarchical classification structure—first distinguishing the fish species and then assessing freshness within each species. This two-step approach could reduce inter-class variability caused by anatomical differences between species and improve overall classification accuracy. Future studies may consider incorporating k-fold cross-validation during hyperparameter tuning to improve the robustness and generalizability of model performance, particularly in cases of limited or imbalanced datasets.

Moreover, the integration of external contextual features, such as ambient storage temperature, humidity, and time since harvest, could provide additional predictive power. These variables often correlate with freshness and may help resolve ambiguities in intermediate classes like "fair to average". While deep learning methods were not explored in this study due to dataset constraints, future research with a larger and more diverse dataset could investigate Convolutional Neural Networks (CNNs) or hybrid deep learning pipelines. Transfer learning from pre-trained models could also be considered to overcome limitations related to small sample sizes.

Lastly, future research should explore the integration of external factors, such as environmental conditions and handling practices, which may influence fish freshness. By incorporating these elements into the classification framework, the models could provide more comprehensive insights into fish quality assessment. This research not only advances the understanding of machine learning applications in food safety but also lays the groundwork for future studies aimed at enhancing the accuracy and reliability of fish freshness classification systems.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contributions

J. P. Q. Tomas supervised the study, administered the project, and contributed to writing, review, and editing. M. R. J. C. Caranay was responsible for conceptualization, methodology, data curation, investigation, formal analysis, and contributed to writing both the original draft and the review and editing of the manuscript. E. G. M. G. Domingo contributed to conceptualization, methodology, data curation, investigation, formal analysis, and writing, including both the original draft and subsequent review and editing. M. B. Enciso also contributed to conceptualization, methodology, data curation, investigation, formal analysis, and writing, including the original draft and review and editing. B. T. Doma supervised the study and acquired funding. All authors have read and approved the final version of the manuscript.

References

[1] P. Howgate, "A history of the development of sensory methods for the evaluation of freshness of fish," *J. Aquat. Food Prod. Technol.*, vol. 24, pp. 516–532, 2015.

[2] A. T. Nguyen *et al.*, "Proposed novel fish freshness classification using effective low-cost threshold-based and neural network models on extracted image features," in *Advances in Intelligent Systems and Computing*, vol. 1084, 2020, pp. 60–71. doi: 10.1007/978-3-030-62324-1_6

[3] E. C. Medeiros, L. Almeida, and J. G. De A. T. Filho, "Computer vision and machine learning for tuna and salmon meat classification," *Informatics*, vol. 8, no. 4, p. 70, 2021. doi: 10.3390/informatics8040070

[4] Assessment of fish quality. in *Quality and Quality Changes in Fresh Fish*. FAO, (n.d.). [Online]. Available: https://www.fao.org/3/V7180E/V7180e09.htm

[5] E. Larsen, J. Heldbo, C. M. Jespersen, and J. Nielsen, "Development of a method for quality assessment of fish for human consumption based on sensory evaluation," *Developments in Food Science*, vol. 29, pp. 351–358, 1992.

[6] J. F. Aguilar and L. M. Pengson, "Physicochemical indicators of spoilage in marinated fish products: A review," *CLSU International*

Journal of Science and Technology, vol. 8, no. 1, pp. 55–77, 2024. https://ijst.clsu.edu.ph/index.php/ijst/article/view/76/44

[7] C. Shi *et al.*, "Developing a machine vision system for simultaneous prediction of freshness indicators based on tilapia (*Oreochromis niloticus*) pupil and gill color during storage at 4 °C," *Food Chem.*, vol. 243, pp. 134–140, 2018. doi: 10.1016/j.foodchem.2017.09.047

[8] X. Guo, S. Gu, and D. Wei, "Fish freshness evaluation based on Mobilenetv1 and attention mechanism," in *Proc. 2022 6th Int. Conf. Electron. Inf. Technol. Comput. Eng.*, 2022, pp. 270–276. doi: 10.1145/3573428.3573475

[9] I. Navotas *et al.*, "Fish identification and freshness classification through image processing using artificial neural network," *J. Eng. Appl. Sci.*, vol. 13, pp. 4912–4922, 2018.

[10] A. Banwari, R. C. Joshi, N. Sengar, and M. Dutta, "Computer vision technique for freshness estimation from segmented eye of fish image," *Ecol. Inform.*, vol. 69, 101602, 2022. doi: 10.1016/j.ecoinf.2022.101602

[11] H. M. Lalabadi, M. Sadeghi, and S. A. Mireei, "Fish freshness categorization from eyes and gills color features using multi-class artificial neural network and support vector machines," *Aquac. Eng.*, vol. 90, 102076, 2020. doi: 10.1016/j.aquaeng.2020.102076

[12] L. K. Tolentino *et al.*, "Fish freshness determination through support vector machine," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, pp. 139–143, 2017.

[13] A. Taheri-Garavand, S. Fatahi, A. Banan, and Y. Makino, "Real-time nondestructive monitoring of common carp fish freshness using robust vision-based intelligent modeling approaches," *Comput. Electron. Agric.*, vol. 159, pp. 16–27, 2019. doi: 10.1016/j.compag.2019.02.023

[14] D. Sutaji and R. Dijaya, "Classification of milk fish quality using fuzzy K-nearest neighbor method based on form descriptor and co-occurrence matrix," *J. Phys.: Conf. Ser.*, vol. 1179, 012021, 2019. doi: 10.1088/1742-6596/1179/1/012021

[15] M. A. Rayan, A. Rahim, M. A. Rahman, M. A. Marjan, and U. A. M. E. Ali, "Fish freshness classification using combined deep learning model," in *Proc. 2021 Int. Conf. Autom., Control Mechatron. Ind. 4.0 (ACMI)*, Rajshahi, Bangladesh, 2021, pp. 1–5. doi: 10.1109/ACMI53878.2021.9528138

[16] M. Dowlati *et al.*, "Freshness assessment of gilthead sea bream (*Sparus aurata*) by machine vision based on gill and eye color changes," *J. Food Eng.*, vol. 119, no. 2, pp. 277–287, 2013. doi: 10.1016/j.jfoodeng.2013.05.023

[17] S. Kunjulakshmi *et al.*, "Development of portable, non-destructive freshness indicative sensor for Indian mackerel (*Rastrelliger kanagurta*) stored under ice," *J. Food Eng.*, vol. 287, 110132, 2020. doi: 10.1016/j.jfoodeng.2020.110132

[18] A. Issac, M. K. Dutta, and B. Sarkar, "Computer vision based method for quality and freshness check for fish from segmented gills," *Comput. Electron. Agric.*, vol. 139, pp. 10–21, 2017. doi: 10.1016/j.compag.2017.05.006

[19] A. Taheri-Garavand, A. Nasiri, A. Banan, and Y. Zhang, "Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish," *J. Food Eng.*, vol. 278, 109930, 2020. doi: 10.1016/j.jfoodeng.2020.109930

[20] N. Sengar, M. K. Dutta, and B. Sarkar, "Computer vision based technique for identification of fish quality after pesticide exposure," *Int. J. Food Prop.*, pp. 1–15, 2017. doi: 10.1080/10942912.2017.1368553

[21] E. Yavuzer and M. Köse, "Prediction of fish quality level with machine learning," *Int. J. Food Sci. Technol.*, vol. 57, no. 8, pp. 5250–5255, 2022. doi: 10.1111/ijfs.15853

[22] A. T. Nguyen *et al.*, "Proposed novel fish freshness classification using effective low-cost threshold-based and neural network models on extracted image features," in *Proc. Int. Conf. Adv. Intell. Syst. Comput.*, vol. 1084, 2020, pp. 60–71.

[23] A. Yudhana, R. Umar, and S. Saputra, "Fish freshness identification using machine learning: Performance comparison of k-NN and naïve Bayes classifier," *J. Comput. Sci. Eng.*, vol. 16, no. 3, pp. 153–164, 2022. doi: 10.5626/JCSE.2022.16.3.153

[24] L. K. Tolentino *et al.*, "Fish freshness determination through support vector machine," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 2–9, pp. 139–143, 2017.