Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)

A Special Problem Proposal
Presented to
the Faculty of the Division of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Visayas
Miag-ao, Iloilo

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science by

ADRICULA, Briana Jade
PAJARILLA, Gliezel Ann
VITO, Ma. Christina Kane

Francis DIMZON
Adviser
Victor Marco Emmanuel FERRIOLS
Co-Adviser

April 28, 2025

Abstract

Tegillarca granosa (Linnaeus, 1758), commonly known as blood cockles, is one of the most well-known marine bivalve for its nutritional benefits and economic significance. Determining their sex is essential for maintaining a balanced male-to-female ratio, which is crucial for preventing overexploitation of this shellfish resource. The sex-determining mechanism in the shell morphology of bivalves is challenging macroscopically due to the limited literature regarding this expertise. In addition, no current technologies are employed to classify the sex based on shell morphology. This study proposes a machine learning approach for classifying the sex of blood cockles using various linear measurements (length, width, height, distance between the hinge line, distance between umbos, and rib count) and angles (dorsal, ventral, anterior, posterior, left lateral, and right lateral) collected from male and female specimens. Available machine learning models in MATLAB were trained to discern sexual dimorphism. Among the models, Linear SVM performed best, achieving an accuracy of 69.80%, precision of 69.82%, recall of 69.80%, and an F1-score of 69.73%. Feature importance analysis indicated that the distance between the umbos and height were the most significant features.

Keywords: deep learning, supervised machine learning , convolutional neural network, blood cockle, sex identification, *Tegillarca granosa*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

The Philippines is a global center of marine biodiversity and has established aquaculture as a significant contributor to total fishery production (Aypa & Baconguis, 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annually and is the 11th largest seafood producer in the world. Aquaculture is deeply integrated into Filipinos' livelihoods, encompassing fish cultivation and the production of various aquatic species, including bivalves. Among these, blood cockles *(Tegillarca granosa)* hold considerable economic and environmental significance, making it essential to ensure sustainable production and population balance.

Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent overharvesting and ensure sustainability. An imbalanced ratio can lead to overexploitation and negatively impact the population's viability. However, there is limited literature on *T. granosa* that provides a thorough understanding of its sex-determining mechanisms, particularly regarding sexual dimorphism based on morphological and morphometric characteristics (Breton, Capt, Guerra, & Stewart, 2017).

Currently, sex determination methods for blood cockles are invasive, including dissection and histological examinations, which often result in the death of the species. While there is growing literature on sex identification in aquaculture commodities using machine learning and deep learning, there is a notable scarcity of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

This study, titled "Morphometric-Based Non-Invasive Sex Identification of

Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)," aims to provide a detailed baseline analysis of blood cockles by leveraging their morphological and morphometric characteristics. Sexual dimorphism in bivalves is often subtle and challenging to establish mascropically (Karapunar, Werner, Fürsich, & Nützel, 2021). However, by integrating machine learning and deep learning, the study seeks to identify distinct features that may indicate sexual dimorphism between male and female blood cockles.

1.2 Problem Statement

Identifying the sex of *T. granosa* is important for promoting sustainable aquaculture and biodiversity by maintaining a balanced male-to-female ratio. A balanced ratio helps prevent overharvesting. Although sex identification is crucial for blood cockle population management and sustainable aquaculture, there is a notable lack of research on creating non-invasive methods for determining the sex of *T. granosa*. Many recent studies and approaches rely on invasive methods like dissection or histological analysis, which are impractical for large-scale aquaculture operations focused on conservation.

Current methods for determining the sex of *T. granosa* are invasive and involve dissection, which requires cutting open the shell to visually inspect the gonads (Erica, 2018). This procedure can cause harm to the specimens and frequently leads to their death. Another method is histological examination, where tissue samples are analyzed under a microscope (May, Maung, Phyu, & Tun, 2021). Both approaches are labor-intensive and time-consuming, and can pose risks to population management, particularly when maintaining a balanced sex ratio for breeding programs is essential. Moreover, these invasive methods require specialized technical skills for accurate execution. Resource-limited aquaculture operations face significant challenges in accessing the necessary laboratory equipment, such as microscopes and staining tools, complicating the process.

A less invasive approach employed by aquaculturists involves monitor spawning behavior, where individuals are separated and stimulated to reproduce in order to determine their sex through the release of gametes (Miranda & Ferriols, 2023). Although this method is indeed less invasive than dissection, it still induces stress in blood cockles and may not be completely effective for fast identification in large populations.

Given the limitations of both invasive and less invasive methods, there is a clear need for a more advanced approach. An alternative, non-invasive method involving machine and deep learning technologies could address these issues by providing a fast, accurate, and effective solution without harming or stressing the blood cockles.

1.3 Research Objectives

1.3.1 General Objective

The general objective of this study is to develop a non-invasive method for identifying the sex of *Tegillarca granosa* using machine and deep learning integrated with computer vision technologies. This method aims to provide accurate and streamlined sex identification without causing harm to the specimens, thus supporting sustainable aquaculture practices.

1.3.2 Specific Objectives

To achieve the overall general objective of developing a non-invasive sex identification of *T. granosa* using machine learning, deep learning, and computer vision technologies, the following specific objectives have been established:

1. To collect and organize a comprehensive dataset of *T. granosa* which will include high-quality images and relevant morphological measurements that will serve as the basis for the machine-learning model.
2. To develop and implement machine learning models that can classify the sex of *T. granosa* based on the collected linear measurements and images of different angles of the sample.
3. To evaluate the performance of the models used using performance metrics such as accuracy, precision, recall, and F1-score.
4. To develop a system that can identify the sex of *T. granosa* based on its morphological characteristics.

1.4 Scope and Limitations of the Research

This study is conducted alongside the ongoing research by the UPV DOST-PCAARRD, titled "Establishment of the Center for Mollusc Research and Development: Development of Spawning and Hatchery Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."The ongoing research primarily involves the rearing of *T. granosa* from spat to larvae, as well as feeding experiments, stocking density evaluations, substrate selection, and settlement rate assessments.

In contrast, this study mainly focuses on developing a non-invasive method for identifying the sex of *Tegillarca granosa* using machine learning, deep learning, and computer vision technologies. The goal is to provide an accurate and efficient means of sex identification without causing harm to the samples, contributing to sustainable aquaculture practices.

The researchers work with 500 already sex-identified blood cockles taken from Panay Island, specifically from Zarraga Iloilo and Ivisan Capiz. These samples, equally divided between 250 males and 250 females, were obtained through induced spawning via temperature shock and dissection. Samples subjected to data collection of *T.granosa* are only limited to the spawned stage among the five gonadal stages - immature, developing, mature, spawning, and spent stages. The other stages are not preferable due to indistinguishable gonads and their inability to perform induced spawning (May et al., 2021). Thus, the researchers only focused on the samples undergoing the spawned stage.

In collecting the data, the researchers will personally gather linear measurements, including length, width, height, rib count, length of the hinge line, and distance between the umbos through the vernier caliper. Images of the specimens, captured from various angles, will also be gathered under the supervision of University Research Associates from the Institute of Aquaculture, College of Fisheries and Ocean Sciences. Collection of the images of the sample is non-invasive due to the blood cockle-built ability to survive in low oxygen areas and having the intertidal mudflats as their natural habitat (Zhan & Bao, 2022).

The method developed in this study is specific to *Tegillarca granosa* and may not be applicable to other bivalve species. The model will be trained exclusively for *Tegillarca granosa* and morphological features including length, width, height, rib count, length of the hinge line, and distance between the umbos may not be consistent across other shellfish species.

1.5 Significance of the Research

This study will give us a significant advancement in non-invasive sex identification methods in *T. granosa* providing innovative solutions that could solve the challenges in identifying sex and reshape sustainable approaches to aquaculture. The significance of this study extends to the following:

Research Institution. The result of this study focusing on the sex-identification mechanism of bivalves, specifically *Tegillarca granosa*, will provide valuable insights into universities and research centers that focus on fisheries and coastal management, such as the UPV Institute of Aquaculture, that aim to develop sustainable development and suitable culture techniques.

Fishermen. By developing a non-invasive method in sex identification, this study can help long-term harvest efficiency and maintain the ratio of the harvest which can help prevent overexploitation of the *T. granosa.*

Coastal Communities. The result of this study would be beneficial for the coastal communities that are reliant on their source of income with aquaculture commodities like blood cockles. Maintaining the diversity and aspect ratio of male and female may increase the market value of blood cockle production since cockle aquaculture faces significant obstacles worldwide due to the fluctuating seed supplies and scarcity of broodstock from the wild.

Future Researchers. The result of this study would serve as the basis for studies that involve sex identification in bivalves such as *T. granosa*. Some technologies are yet to be explored in machine learning, deep learning, and computer vision technologies that can lead to higher accuracy and distinguish the presence of sexual dimorphism in the *T. granosa*.

Chapter 2

Review of Related Literature

Aquaculture is the fastest-growing industry in animal food production and has great potential as a sustainable solution to global food security, nutrition, and development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply integrated into the livelihoods of Filipinos, not only through fish cultivation but also through the production of other aquatic species, including mollusks, oysters, clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood clams *Tegillarca granosa*, have economic and environmental significance. It has been a collective effort to maintain an ideal male-to-female ratio to avoid overharvesting and maintain the optimal ratio to preserve the population and production of the blood cockles.

The members of the Arcidae Family, including *T. granosa* are important sources of food and livelihood. Cockle aquaculture meets rising demands, however, it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols, 2023). To solve the problem, researchers exert a considerable amount of effort, developing a broader understanding of bivalves, including their sex-determining mechanism, due to their notable importance in terms of diversity, environmental benefits, and economic and market importance (Breton et al., 2017). Despite the promising idea of identifying sex, there is limited research reported in terms of sexual dimorphism, making it harder to distinguish through its morphological and morphometric characteristics.

By addressing the challenges in the sex identification of *T. granosa* , it would be able to address one problem at a time. Currently, there are no recent documented publications that integrate machine learning and computer vision in characterizing sexual dimorphism, reducing complexity, variability in sex determination, and differentiation mechanisms in bivalves, including *T. granosa* specifically.

2.1 Background on *Tegillarca granosa* and Their Importance

Tegillarca granosa (Linnaeus, 1758) is also known as blood cockles or blood clam. In the Philippines, it is commonly known as a Litob, a marine bivalve species from the family Arcidae. Litob is widely distributed in the world including Southeast Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat as T.granosa's habitat, they experience severe hypoxia or low oxygen levels in the blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood phenotype where it serves two purposes the hemocyte carries oxygen around the body and strengthens immune defenses. In addition, it possesses a unique ability to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

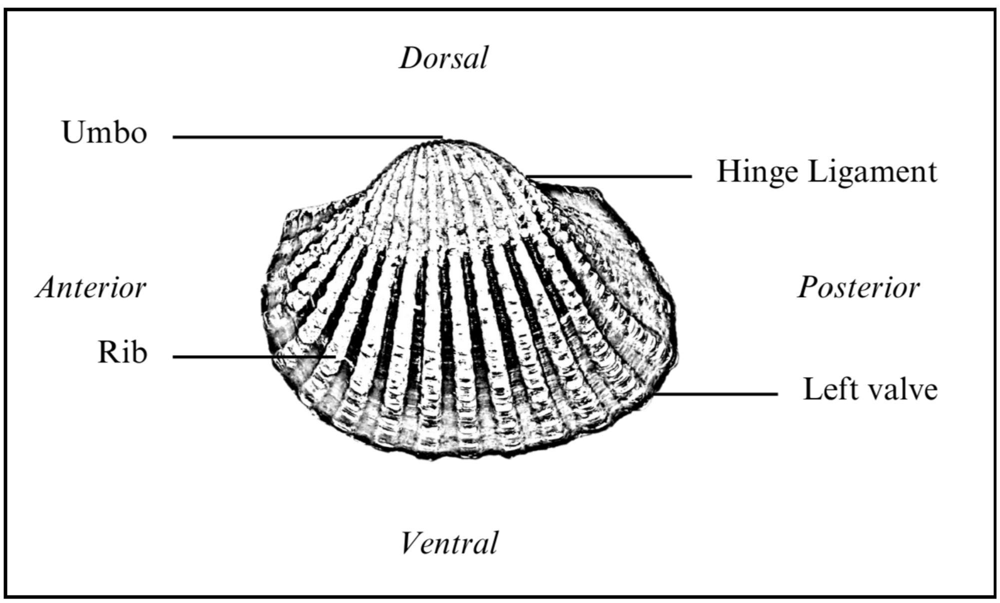

Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

T. granosa shell is medium-sized, fairly thick, ovate, and convex, with both valves being equal in size but asymmetrical from the hinge. The top edge of the dorsal margin is straight, while the front is rounded and slopes downward, with its back being obliquely rounded with a concave bottom edge. It has a narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings, although some may be incomplete. The shell's outer layer, or the periostracum, is smooth and brown with a straight hinge line and 40-68 fine short teeth arranged in a straight line. The beak, or prosogyrate, curves forward, with the shell having 18–21 raised ribs with blunt nodules and spaces between them. The inner shell is white with crenulations along the valves' ventral, anterior, and posterior margins. The posterior adductor scar is elongated and squarish, while the anterior adductor scar is similar but smaller in size. The mantle covering the bulk of *T. granosa's* visceral mass is thin but the edges are thick and muscular. It bears the impression of the crenulated shell edges. Their foot is large with a ventral grove with no byssus or thread-like attachment. The *T. granosa's* soft body is blood red (Narasimham, 1988).

T. granosa is one of the most well-known marine bivalves given that they are a protein-rich food, known for their rich flavor, substantial nutritional benefits, a good source of vitamins, low in fat, and contain a considerable amount of iron, important in combating anemia (Zha et al., 2022). Blood cockles were collected by locals inhabiting the brackish mudflats during the low tides for consumption and sold in the market as a source of livelihood (Miranda & Ferriols, 2023). *T. granosa* is not only valuable for its market and food purposes but also facilitates an important role in marine ecosystems as a food source for various organisms like wading birds, intertidal-feeding fish, and crustaceans such as shore crabs and shrimp (Burdon, Callaway, Elliott, Smith, & Wither, 2014). Blood cockles can act as sentinel species and a bioindicator of marine pollutants such as heavy metals (Ishak, Mohamad, Soo, & Hamid, 2016) and polycyclic aromatic hydrocarbons (PAHs) (Sany et al., 2014). Additionally, cockle shells can be utilized to create a cost-effective catalyst for biodiesel production by providing calcium oxide (Boey, Maniam, Hamid, & Ali, 2011).

Determining the sex of bivalves is important for three reasons: diversity, environmental benefits, and economic significance (Breton et al., 2010). Firstly, with the estimated 25, 000 living species under class Bivalvia, it would be a suitable resource to develop a broader understanding of their evolution of the sex and sex determination mechanism (Breton et al., 2010). Second, studying sex determination is important since bivalves are utilized as bioindicators of environmental health. This would pave the way for understanding bivalves' life cycle and population dynamics in determining different factors that affect them (Campos, Tedesco, Vasconcelos, & Cristobal, 2012). Thirdly, the immediate and practical reason to unveil the sex determination mechanism is the economic and nutritional importance of bivalves as a large population of people relies on fish and shellfish as sources of food and nutrition (Naylor et al., 2000). Additionally, male and female aquaculture commodities have different growth and economic values. Male Nile tilapia, for example, grow faster and have lower feed conversion rates than females, female Kuruma prawns (*Penaeus japonicus*) are generally larger than males at the time of harvest (Budd, Banh, Domingos, & Jerry, 2015).

Clearly, much more work is required to understand the mechanisms underlying sexual dimorphism in bivalves, specifically *T. granosa*. Just like the other aquaculture commodities, sex affects not just reproduction but it can affect market preference and underlying economic value, making the determination of sex important for meeting consumer demands. These are the increasing significance of the *T. granosa* despite the lack of reviewed articles in the Philippines.

2.2 Current Methods of Sex Identification in *Tegillarca granosa*

The current sex identification methods in *Tegillarca granosa* range from invasive histological techniques to less invasive methodologies like temperature-induced spawning. Each approach comes with its pros and cons regarding accuracy, feasibility, and impact on natural populations.

Induced spawning and larval rearing are considered the less invasive techniques used to study *Tegillarca granosa*. In the Philippines, limited research has been done on the *Tegillarca granosa* (Linnaeus, 1758), and this study, titled Initial Attempts on Spawning and Larval Rearing of the Blood Cockle, *Tegillarca granosa* in the Philippines, is conducted by Denise Vergara Miranda and Victor Marco Emmanuel Nuestro Ferriols (2023). The researchers conducted experiments on induced spawning and larval rearing, discovering that the eggs of female *T. granosa* were salmon pink, while the sperm released by males looked milky. After spawning, the researchers successfully generated 6, 531, 000 fertilized eggs.

They highlighted the importance of *T. granosa* and other anadarinids as a food source that was established worldwide, especially in Malaysia and Korea. However, in the Philippines, the bivalve aquaculture of the clam species is still limited. The experiment which focuses on the culture and rearing of *T. granosa* was attempted by subjecting the wild broodstocks to a series of temperature fluctuations to induce the spawning of gametes. This is currently the most natural and least invasive method for bivalves (Aji, 2011). The study of Miranda and Ferriols aimed to pave the way to the sustainable production of *T. granosa* seeds for aquaculture production and stock enhancement despite the scarcity of documented hatchery culture of *T. granosa* from larvae to adults that is available in the Philippines.

In the study entitled "The earliest example of sexual dimorphism in bivalves —

evidence from the astartid Nicaniella (Lower Jurassic, southern Germany)," the researchers utilized Principal Component Analysis and Fourier Analysis as a non-invasive method that investigates sexual expression in the *Nicaniella rakoveci*. In the study, researchers discovered that the bivalves with crenulations were found to have a different shell shape, which made them more inflated than those without crenulations. This suggests that when they became females, they adapted to hold more eggs rather than for protection from predators as previously thought. The formation of crenulations is likely part of the genetic process that controls both the sex change and the changes in shell structure (Karapunar et al., 2021). Overall, the findings demonstrate that the genetic mechanisms for sex change and shell morphology in bivalves existed as early as the Early Jurassic, contributing to our understanding of bivalve diversity and evolution. Thus, the researchers concluded that crenulations serve as a morphological marker for identifying the sex and reproductive stage of these bivalves (Karapunar et al., 2021).

On the other hand, invasive techniques such as histological analysis offer a more thorough but harmful method for determining the sex of *T. granosa*. A study on the Spawning Period of Blood Cockle *Tegillarca granosa* (Linnaeus, 1758) in Myeik Coastal. 240 blood cockle samples were examined for sex and gonad maturity stages using histological examination, with shell lengths ranging from 26-35mm and shell weights from 8.1-33g. For histological analysis, the whole soft tissues were removed from the shell and the flesh containing most parts of the gonads was fixed in formalin, dehydrated in an upgraded series of ethanol, and cleared in xylene. This invasive method allows for precise identification of the gonadal maturation stages based on the cellular and structural changes in the gonads.

The classification of the gonad stages used was by Yurimoto et al. (2014). There are five maturation stages of gonadal development: immature (Stage I), developing (Stage II), mature (Stage III), spawning (Stage IV), and spent (Stage V) stages. The sex of the *T. granosa* was confirmed by the color of the gonad and by conducting a histological examination of the gonads. During the immature stage, sex determination was indistinguishable due to the difficulties of observing the germ cells. In the developing stage, the spermatocytes and a few spermatids can be seen for males, and immature oocytes are attached to the tube wall for the female. In the mature stage, the follicles are full of spermatozoa with their tails pointing towards the center of the tube for the male, and the female is full of mature oocytes that are irregular or polygonal in shape with the oval nucleus. Upon reaching spawning, some spermatozoa are released, causing the empty space in the follicle wall for males and females. There is a decrease in the number of mature oocytes and it exhibits nuclear disappearance due to the breakdown of the germinal vesicle. Lastly, the spent stage is where the genital tube is deformed and devoid of spermatocytes which have completely spawned. In the female, the genital tube is deformed and degenerated, making it empty. The morphology of the cockle gonad shows that the area of the gonad increases according to the increased levels of gonad maturity. The coloration of the gonad tissue layer in the blood cockle varies from orange-red to pale orange in females and from white to grayish-white in males for different maturity stages (May et al., 2021).

Although the histological examination is the most reliable method for obtaining accurate information on the reproductive biology and sex determination of *T. granosa*, it has limitations. Given its invasive nature, this approach requires the dissection and destruction of specimens, making it unsuitable for continuous monitoring and conservation efforts. Moreover, the current understanding of sex determination in bivalves and mollusks is poor, and no chromosomes that can be differentiated based on their morphology have been discovered (Afiati, 2007). There exists a study that can provide insight into the sex-determining factor in bivalves but *N. schoberti* is more difficult to analyze concerning potential sexual dimorphism. Thickening the edges of the shell increases its inflation, which means the shell can hold more space inside. This extra space helps protandrous females accommodate more eggs.

2.3 Machine Learning and Deep Learning in Biological Studies

Machine learning has the potential to improve the quality of life of human beings and has a wide range of applications in terms of research and development. The term machine learning refers to the invention and algorithm evaluation that enables pattern recognition, classification, and prediction based on models generated from available data (Tarca, Carey, Chen, Romero, & Drăghici, 2007). The study of machine learning methods has advanced in the last several years, including biological studies. In biological studies, machine learning has been used for discovery and prediction. This section will explore existing machine learning studies that are applied in biological sciences, highlighting the identification of sex in shells, bivalves, and mollusks.

2.3.1 Deep Learning for Phenotype Classification in Ark Shells

In the study, the researchers utilized three (3) convolutional neural network (CNN) models: the Visual Geometry Group Network (VGGnet), the Inception Residual Network (ResNet), and the SqueezeNet (Kim, Yang, Cha, Jung, & Kim, 2024). These deep learning models are utilized for the ark shells, namely *Anadara kagoshimensis*, *Tegillarca granosa*, and *Anadara broughtonii*, to identify the phenotype classification.

The researchers classified the ark shells based on radial rib count where they investigated the difference in the number of radial ribs between three species and were counted. Their CNN-based model that classifies images of three ark shells can provide a theoretical basis for bivalve classification and enable the tracking of the entire production process of ark shells from catching to selling with the support of big data, which is useful for improving food safety, production efficiency, and economic benefits (Kim et al., 2024).

2.3.2 Geometric Morphometrics and Machine Learning for Species Delimitation

In *Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail Placostylus (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia*, the shell size was quantified using centroid size from the Procrustes analysis, and both the shape and size information were used in training the machine learning model. Their study concluded that the researchers support utilizing both methods: supervised and unsupervised machine learning, rather than choosing either of them individually. In general, their research contributes to the growing number of studies that have combined geometric morphometrics with the aid of machine learning, which is helpful in biological innovation and breakthrough (Quenu, Trewick, Brescia, & Morgan-Richards, 2020).

2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning

Tuset et al. (2020), in their study, *Recognising mollusc shell contours with enlarged spines: Wavelet vs Elliptic Fourier analyses*, mentioned that gastropod shells have large spines and sharp shapes that differ based on environmental, taxonomic, and evolutionary influences. The researchers stated that classic morphometric methods may not accurately depict morphological features of the shell, especially when using the angular decomposition of the contour. The current research examined and compared the robustness of the contour analysis using wavelet transformed and Elliptic Fourier descriptors for gastropod shells with enlarged spines. For that, the researchers analyzed two geographically and ecologically separated populations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed that contour analysis of gastropod shells with enlarged spines can be analyzed using both methodologies, but the wavelet analysis provided better local discrimination. From an ecological perspective, shells with various sizes of spines in both areas indicate the broad adaptability of the species.

2.3.4 Machine Learning for Shape Analysis of Marine Organisms

In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses to Recording Structures of Marine Organisms for Stock Discrimination and Taxonomic Purposes*, they utilized geometric morphometrics (GM) as an approach to the traditional method of collecting linear measurements with the application of multivariate statistical methods and outline analysis in recording the structures of marine organisms. The main taxonomic categories (mollusks, teleost fish, and elasmobranchs) with their hard bodies have been used as an indication of age and a determinable time-scale and structure continue to go through life (Arkhipkin, 2005; Kerr & Campana, 2014). This study has explored variations in the morphometry of recording structures in stock discrimination and systematics. The researchers utilized the principal component analysis rather than the traditional approach, which helps simplify the data without losing important information. They utilized landmark-based geometric morphometrics, which has three different types, namely: discrete juxtaposition of tissue, maxima or curvature, or other morphogenetic processes, and lastly, the extremal points are constructed landmarks.

Generalized Procrustes Analysis (GPA) is a common superimposition technique in landmark-based geometric morphometrics that aligns landmarks via translation, scaling, and rotation to eliminate non-shape deviations (Zelditch, Swiderski, & Sheets, 2004). However, there is a limit to the amount of smooth areas that may be captured, and it is possible to overlook significant shape details. Utilization of the semi-landmarks enhanced the shape description (Adams, Rohlf, & Slice, 2004). The researchers observed that using an outline-based approach would be more effective than using a landmark-based approach.

Another approach is the Fourier analysis which is a curve-fitting approach commonly used due to its well-known mathematical background and how general functions can be decomposed into trigonometric or exponential functions with definite frequencies. It has two main approaches, namely: Polar Transform (PT) in which it expresses the outline using equally spaced radii, and Elliptical Fourier Analysis (EFA) which separately analyzes the x and y coordinates of the shape. The PT works for simple rounded outlines and has the tendency to miss details in more complex shapes, unlike the EFA which can handle complex, convoluted outlines (Zahn & Roskies, 1972; Doering & Ludwig, 1990; Ponton, 2006). Many researchers view EFA as the most effective Fourier method for providing a comprehensive and detailed description of recording structures (Mérigot, Letourneur, & Lecomte-Finiger, 2007; Ferguson, Ward, & Gillanders, 2011; Leguá, Plaza, Pérez, & Arkhipkin, 2013; Mahé et al., 2016).

Landmark-based methods used in the study showed that there are detectable differences between male and female octopuses. However, the accuracy of determining sex based on these differences was low, similar to the results obtained with traditional morphometric techniques. The study involved a relatively small sample size of 160 individuals, and the structure being analyzed (the stylet, or internalized shell) varies significantly between individuals. Although the results aligned with findings from other studies that attempted to identify gender differences in cephalopods, the researchers concluded that the approach might not be accurate enough for reliable sex determination.

2.3.5 Deep Learning for Landmark-Free Morphological Feature Extraction

In another study, *a deep learning approach for morphological feature extraction based on variational auto-encoder: an application to mandible shape*, the Morpho-VAE machine learning approach was used to conduct a landmark-free shape analysis. Morpho-Vae reduces dimensions by concentrating on morphological features that distinguish data with different labels using an image-based deep learning framework that combines unsupervised and supervised machine learning. After utilizing the method in primate mandible images, the morphological features reveal the characteristics to which family they belonged. Based on the result, the method applied provides a versatile and promising tool for evaluating a wide range of image data of biological shapes including those missing segments.

2.3.6 Machine Learning for Sex Differentiation in Abalone

In the study, *Towards Abalone Differentiation Through Machine Learning*, researchers identified a problem in abalone farming which is having to identify the sex of abalone to apply measures for its growth or preservation. The researchers classified abalone sex using machine learning. Researchers trained the machine to classify different types of classes which are male, female, and immature. The results demonstrated the effectiveness of utilizing linear classifiers for this task.

Similarly, in the study, *Data scaling performance on various machine learning algorithms to identify abalone sex*, the researchers of the University of India (2022) focused on the data scaling performance of various machine learning algorithms to identify the abalone sex, specifically using min-max normalization and zero-mean standardization. The different machine learning algorithms are the Supervised Vector Machine (SVM), Random Forest, Naive Bayesian, and Decision Tree. Their study aims to utilize machine learning in terms of identifying the trends and distribution patterns in the abalone dataset. Eight features of the abalone dataset (length, diameter, height, whole weight, shucked weight, viscera weight, shell weight, ring) were used to determine the three sexes of Abalone. Their data has been grouped based on sex which are Female, Male, and Infant. They utilized the Synthetic Minority Oversampling Technique (SMOTE) in data balancing for the preprocessing of the data. Followed by data scaling or normalization where it converts numeric values in a data set to a general scale without distorting differences in the range of values. Then they classified by splitting the data into training and testing sets (Arifin, Ariawan, Rosalia, Lukman, & Tufailah, 2021).

The study found that Naive Bayes consistently performed better than other algorithms. However, when applied to both min-max and zero-mean normalization, the average accuracies of the algorithms were as follows: Random Forest (62.37%), SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear kernel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with normalization, Random Forest achieved the highest overall metrics, including an average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of 83.31%. Liu et al. concluded that Random Forest is highly accurate because it can handle large, complex datasets, run processes in parallel using multiple trees, and select the most relevant features to enhance model performance (Arifin et al., 2021).

2.3.7 Machine Learning for Geographical Traceability in Bivalves

In the study, *BivalveNet: A hybrid deep neural network for common cockle (Cerastoderma edule) geographical traceability based on shell image analysis*, the researchers incorporated computer vision and machine learning technologies for an efficient determination of blood cockle harvesting origin based on the shell geometric and morphometric analysis. It aims to improve the traceability methodologies in these organisms and its potential as a reliable traceability tool. Thirty *Cerastoderma edule* samples were collected along the five locations on the Atlantic West and South Portuguese coast with individual images processed using lazy snapping segmentation, spectro-textural-morphological phenotype extraction, and feature selection through hybrid Principal Component Analysis and Neighborhood Component Analysis (Concepcion, Guillermo, Tanner, Fonseca, & Duarte, 2023).

The researchers developed a non-invasive image-based traceability technique, an alternative to the chemical and biochemical analysis of the bivalves. It was able to incorporate machine learning methods to promote lesser human intervention. The researchers discovered that BivalveNet emerged as the superior model for bivalves with 96.91% accuracy which is comparable to the accuracy of the destructive methods with 97% and 97.2% accuracy rates. The result of the study aided the researchers in concluding that there is a possibility of on-site evaluation of the bivalve through the implementation of a mobile app that would allow the public and official entities to obtain information regarding the provenance of seafood products' traceability because of its non-invasive and image-based aspects (Concepcion et al., 2023).

Tegillarca granosa is known for having no sexual dimorphism. However, through several related studies, the researchers can apply how family shells of *Tegillarca granosa* have been identified based on its morphological and morphometric characteristics and the methods used in machine learning in identifying its sex.

2.4 Limitations on Sex Identification in *Tegillarca granosa*

To date, no distinction has been made between the male and female *T. granosa* in sexing methodology. In cockle aquaculture without clearly apparent sexual dimorphism, sexing can be performed using invasive methods such as chemical stimulation, dissection, and gonad-stripping. Induced spawning, specifically temperature shock, is the most natural and least invasive method for bivalves (Aji, 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water from hot to cold with a specific temperature requires deliberate and careful manipulation of the temperature over a specific period and would require constant management and monitoring.

Recent studies involved non-invasive methods, with a specific emphasis on morphological characteristics as indicators of sex differentiation. However, Tatsuya Yurimoto et al. (2014) stated that the existing methods for determining the sex of bivalves and mollusks in general are somewhat limited (Afiati, 2007). At present, there is no recorded evidence of sexual dimorphism in *Tegillarca granosa*. Gonochoristic is the classification given to *Tegillarca granosa* (Lee, 1997). However, Lee et al. (2012) reported that the sex ratio varied with shell length, suggesting that sex might alter.

Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous (synchronous or functional) characteristics. Sequential hermaphrodites switch genders after being male or female for one or multiple yearly cycles. (Heller, 1993; Gosling, 2004; Collin, 2013). Sex change and consecutive hermaphroditism have been observed in different bivalve species, including Ostreidae, Pectinidae, Veneridae, and Patellidae. However, macroscopically differentiating bivalve sex is challenging. The only way it may be identified is through histological analysis of gonad remains but to do so there is an act of killing the organism (Coe, 1943; Gosling, 2004). Verification of sex change in bivalves to classify whether male or female while they are alive is challenging since they need to be re-confirmed and re-evaluated to be the same individual after a year.

Lee et al. (2012) found out that *T. granosa*, a species in Arcidae, has been discovered to be a sequential hermaphrodite, with the sex ratio changing with an increase in the shell size. In bivalves, sex changes usually happen when the gonad is not differentiated between spawning seasons (Thompson, Newell, Kennedy, & Mann, 1996). But in *T. granosa*, after the spawning season, sex changes during its inactive phase. Results showed a 15.1% sex change ratio, with males having a higher sex change ratio (21.2%) than females (6.2%). The 1+ year class had a higher ratio (17.8%) than the 2+ year class (12.1%). Thus, this study indicates that *T. granosa* is a sequential hermaphrodite. The results of the study demonstrated that the bivalve's age affects the sex ratio and degree of sex change, but additional in-depth investigation is required to determine the role that genetic and environmental factors play in these changes.

No literature in the study of mollusks specifically addresses the machine learning algorithm used to determine the sex of *T. granosa* bivalves in various models. Nevertheless, various techniques such as shape analysis, morphometric analysis, Wavelet, and Fourier analysis, as well as different deep learning models like VGNet, ResNet, and SqueezeNet in CNN networks, are utilized for phenotype classification, while different machine learning algorithms could serve as the foundation for this research project.

2.5 Synthesis of the Study

This section of the paper summarizes the technologies used in the different studies related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca granosa* (Linnaeus, 1758).

Author	Technology / Method Used	Description of Problem	Pros	Cons
D. V. Miranda and V. M. E. N. Ferriols	Temperature shock	No recent studies are available on the production and rearing of *T. granosa* in the Philippines.	Employed less invasive techniques which minimize the stress in *T. granosa* and can lead to better survival rates.	Time-consuming as the entire process from fertilization to the spat stage took 120 days.
Karapunar, Baran and Werner, W. and Fürsich, F. T. and Nützel, A.	Morphometric analysis, microscope imaging, principal component analysis (PCA), and Fourier shape analysis	To address the observed shell dimorphism in the Early Jurassic bivalve *Nicaniella rakoveci*, namely the presence or lack of crenulations on the ventral shell margin, and whether these variations represent sexual dimorphism and sequential hermaphroditism.	The methods used reveal significant morphological differences with regard to sexual dimorphism.	There could be misinterpretation of the shape differences of bivalves due to the constraints and resolution of technologies used.
K. May and C. Maung and E. Phyu and N. Tun	Histological examination	The need to understand the reproductive period of *T. granosa* in Myeik to ensure sustainable aquaculture and to prevent overexploitation.	Method used allows for accurate sex identification based on the histological characteristics and color of the gonads.	Invasive technique used to determine the sex of *T. granosa* through gonad histological analysis.
E. Kim and S.-M. Yang and J.-E. Cha and D.-H. Jung and H.-Y. Kim	Convolutional neural network (CNN) models, VGGNet, Inception-ResNet, SqueezeNet	Traditional methods of recognizing and classifying ark shell species based on shell traits are time-consuming and inaccurate.	Automated classification of the three ark shells using a deep learning model obtained an accuracy of 92.4%.	Challenges may arise with certain ark shells that share similar morphology.
Mathieu Quenu and S. A. Trewick and F. Brescia and M. Morgan-Richards	Neural network analysis (supervised learning) and Gaussian mixture models (unsupervised learning)	To determine whether the shape and size of the snail's shells can distinguish between two *Placostylus* species, particularly in groups that appear to be hybrids.	Combining geometric morphometrics and machine learning effectively answers biological issues, providing insights into species classification and possible hybridization.	Difficulty classifying intermediate phenotypes, with potential for overfitting and misclassification in both learning methods.
V. M. Tuset and E. Galimany and A. Farrés and E. Marco-Herrero and J. L. Otero-Ferrer and A. Lombarte and M. Ramón	Wavelet functions and Elliptic Fourier descriptors	Addresses the difficulty of accurately defining phenotypic diversity in gastropod shells.	Advanced contour analysis methods allow accurate differentiation of gastropod shell forms.	Cannot clarify the causes of phenotypic variation in the two populations studied.
Fedor Lishchenko and Jones, J. B.	Landmark- and outline-based Geometric Morphometric methods	To address difficulties in differentiating between stocks of marine organisms to prevent misidentification that could affect conservation and management.	Shape analysis improves taxonomic classification precision and offers close distinction between related species or organisms.	Landmark-based methods can be sensitive to landmark placement.
M. Tsutsumi and N. Saito and D. Koyabu and C. Furusawa	Morphological regulated variational AutoEncoder (Morpho-VAE)	The need for reliable, landmark-free methods, such as a modified variational autoencoder, to extract and decipher complex shapes from image data.	Employs dimension reduction and feature extraction, making it a user-friendly tool for biology non-experts.	Limited sample size in certain families presented challenges.
Barrera-Hernandez, R. and Barrera-Soto, V. and Martinez-Rodriguez, J. L. and Ríos-Alvarado, A. B. and Ortiz-Rodríguez, F.	Machine learning algorithms	Identifying the sex of abalones is challenging for producers applying specific growth or preservation strategies.	Machine learning algorithms accurately classify abalone sex into three categories: male, female, and immature.	Selected features may not fully capture the complexity of abalone morphology.
Concepcion, R. and Guillermo, M. and Tanner, S. E. and Fonseca, V. and Duarte, B.	EfficientNet-Bo, ResNet101, MobileNetV2, InceptionV3	Addresses the difficulty of accurately tracing bivalve harvesting origins using computer vision and machine learning algorithms to enhance seafood traceability and combat food fraud.	Non-invasive, image-based tools for bivalve traceability provide faster, cheaper, and equally accurate alternatives to traditional chemical analysis methods.	Small sample size (only 30 cockles) limits model reliability.

Table 2.1: Comparison of the Methods Used in Bivalves Studies

Recent developments and breakthroughs in machine learning offer hopeful solutions for biological issues. Research findings indicate that various machine learning techniques such as CNNs, geometric morphometrics, and deep learning models. They are deemed effective for identifying phenotypes and determining the gender of various aquaculture commodities, such as mollusks and abalones. These techniques provide a starting point for creating new, non-invasive ways to differentiate male and female *T. granosa*, potentially addressing the drawbacks of manual and invasive methods. Thus, machine learning to examine morphological and morphometric features may streamline the process of sex identification.

Nevertheless, the use of machine learning to determine the sex of *T. granosa* has not been fully explored. It lacks up-to-date and significant related literature on using machine learning to identify sex in *T. granosa*, particularly given the species' possible sequential hermaphroditism and lack of obvious external sexual distinctions.

Chapter 3

Research Methodology

This chapter discussed the materials and methods employed in the study, focusing on the development requirements, as well as the software and programming languages utilized. It also detailed the overall workflow in conducting the study, Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca granosa* (Linnaeus), 1758) using machine learning and deep learning technologies.

Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture, oversaw the overall workflow and conduct of the experiment. The researchers were also guided by research associates LC Mae Gasit and Allena Esther Artera. Consequently, the entire dataset collection process was conducted at the University of the Philippines Visayas hatchery facility.

The methodology consisted of eight parts: (1) Sample Collection, (2) Ethical Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteristics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Software Configuration,(7) Morphometric Characteristics Evaluation Using Machine Learning, and (8) Morphological Characteristics Evaluation Using Deep Learning

3.1 Sample Collection

The collection of *T. granosa* samples used in this study was part of an ongoing research project by UPV DOST-PCAARRD titled ”Establishment of the Center for Mollusc Research and Development: Development of Spawning and Hatchery Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture.” A total of 271 samples were provided for this study to classify the sex of *T. granosa*.

The samples, ranging in size from 34 to 61 mm, were sourced from the coastal area of Zaraga, Iloilo, and fish markets in Ivisan, Capiz, Philippines (*see Figure 3.1*).

The research and experimentation were conducted at the University of the Philippines Visayas hatchery facility in Miagao, Iloilo, where the samples were maintained in 200 L fiberglass-reinforced plastic (FRP) tanks containing filtered seawater with 35 ppt salinity (Miranda & Ferriols, 2023).

As part of the data collection process, the researchers utilized induced spawning and dissection to classify the sex of the samples. Induced spawning through temperature fluctuations was the most natural and least invasive method for bivalves compared to other approaches (Aji, 2011). However, since not all samples exhibited gamete release, the researchers also performed dissections, assisted by hatchery staff, to expedite data collection. The sex of the dissected samples was identified based on the coloration of gonad tissue, which varies according to sex and maturity stage. Females exhibited orange-red to pale orange gonads, while males displayed white to grayish-white gonads (May et al., 2021).

The methods used for data collection were considered noninvasive, particularly given that *T. granosa* are oxygen regulators well adapted to tidal exposure and hypoxia (Davenport & Wong, 1986).

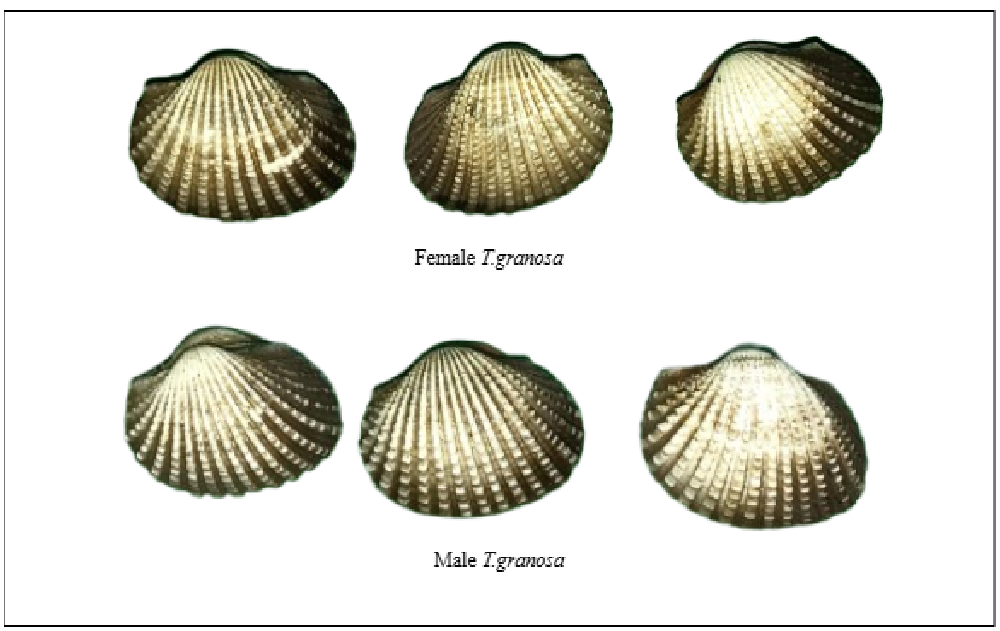

Figure 3.1: Male and Female *Tegillarca granosa* shells

3.2 Ethical Considerations

The ongoing research project titled "Establishment of the Center for Mollusc Research and Development: Development of Spawning and Hatchery Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture"—from which the samples used in this study were obtained—was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of the University of the Philippines Visayas.

3.3 Creating *T. granosa* Dataset

The experiment began with the collection of preliminary observations from 100 *T. granosa* samples. For the actual experimentation, the researchers collected the full dataset in batches until a total sample size of 271 *T. granosa* was reached. Linear measurements—including width, height, length, rib count, hinge line length, and the distance between the umbos—were recorded and organized into a CSV file. This dataset served as the foundation for training and testing machine learning models, as well as for establishing a baseline for the Convolutional Neural Networks.

Images of each sample were captured and saved in JPG format using a standardized file naming convention that included the sample's sex, the shell's orientation or view, and its corresponding number out of the 271 total samples. File names for female *T. granosa* samples began with "0", while those for male samples began with "1". Each file name also included one of the six captured views: (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right lateral (refer to Figure 3.2), followed by a unique sample number. For example, "010001" denoted the first female sample taken from the dorsal view, while "110001" represented the first male sample from the same view. This naming convention was implemented to prevent data leakage and ensure accurate labeling of images according to their respective samples.

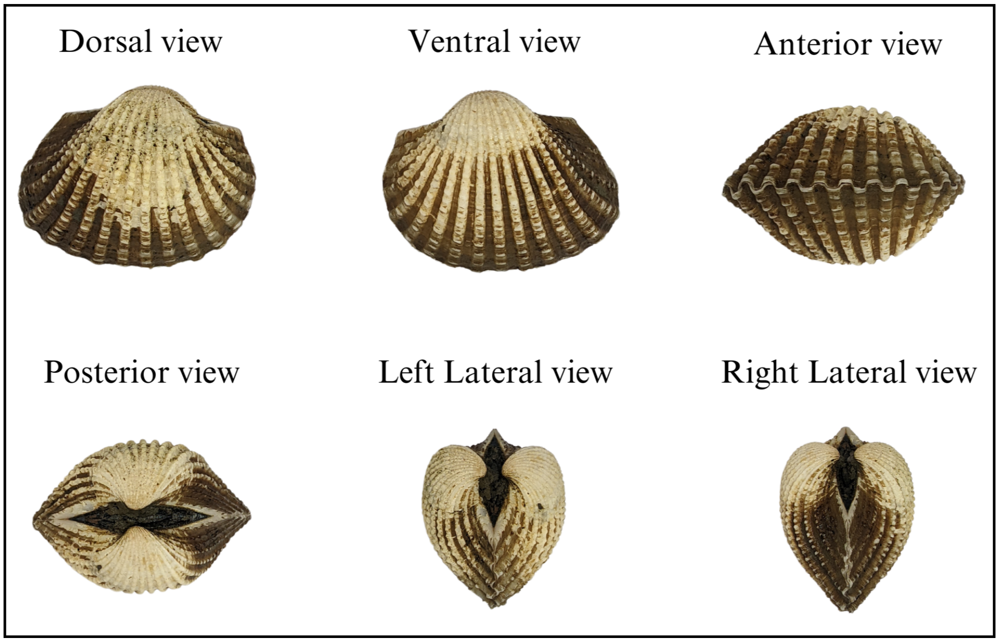

Figure 3.2: Different Views of the *T. granosa* Shell Captured

3.4 Morphological and Morphometric Characteristics Collection

Morphology refers to biological form and is one of the most visually recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023). In this study, morphological characteristics describe the structural features of *T. granosa*, focusing on measurable attributes such as shape, size, and color. Morphometric characteristics, on the other hand, refer to specific quantifiable features of T. granosa, including length, width, height, hinge line length, distance between the umbos, and rib count. As stated by the researchers, quantifying and characterizing these traits is essential for understanding and visualizing variations in *T. granosa* morphology.

The researchers measured the height, width, and length of *T. granosa* using a Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the corresponding measurement diagram. Length (A) refers to the distance from the anterior to the posterior of the shell. Width (B) is defined as the widest span across the shell from the left to the right valve. Height (C) measures the distance from the base to the apex of the shell. In addition, the hinge line length (D) near the hinge and the distance between the umbos (E) were recorded.

Reyment and Kennedy (1998) emphasized that including rib count as supplementary information can enhance identification accuracy. Following this insight, the researchers also recorded the rib count for both male and female *T. granosa*, adjusting the values by calculating ratios to account for natural size variation among specimens.

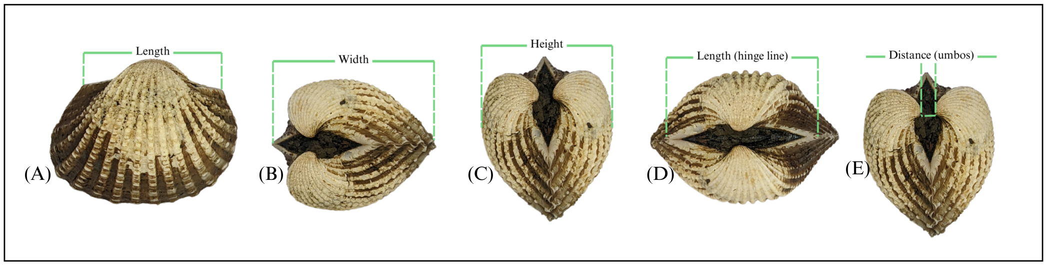

Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

3.5 Image Acquisition and Data Gathering

This study comprised 144 male and 127 female *T. granosa* samples, resulting in a total of 1,626 images captured from various angles. To ensure consistency during image acquisition, the researchers constructed a box-like structure with a white background to control the imaging environment. This setup allowed for uniform image captures by fixing the camera at a consistent angle directly above the *T. granosa*. A ring light was positioned in front of the box to enhance image quality, eliminate shadows, and ensure clarity of the samples throughout the image acquisition process.

The images were captured using a Google Pixel 3 XL smartphone, which features a resolution of 2960 × 1440 pixels and a 12.2 MP camera (4032 × 3024 pixels). Additional camera specifications include an f/1.8 aperture, 28mm wide lens, $\frac{1}{2}$.55” sensor size, 1.4µm pixel size, dual-pixel phase detection autofocus (PDAF), and optical image stabilization (OIS) (Concepcion et al., 2023).

Figure 3.4: Image Acquisition Setup for *T. granosa* Samples

3.6 Hardware and Software Configuration

This section of the paper discusses the software, programming languages, and tools used for sex identification. Data collection, preprocessing, and model training were conducted on a Windows 11 operating system using an ACER Aspire 3 general-purpose unit (GPU) equipped with an AMD Ryzen 3 7320U CPU with Radeon Graphics (8 cores) @ 2.395 GHz and 8 GB of RAM. Google Colaboratory was utilized for collaborative preprocessing, computer vision tasks, and model training. Image preprocessing was performed using computer vision techniques in Python, while machine learning and deep learning models were developed using Python libraries, including Keras. The results of the gathered measurements were stored and managed using spreadsheet software. GitHub was employed for version control, documentation, and activity tracking throughout the study.

3.7 Morphometric Characteristics Evaluation Using Machine Learning

This section of the paper discusses the machine learning operations that served as a baseline prior to implementing more complex deep learning methods for image classification. The study utilized collected variables including linear measurements—length, width, height, hinge line length, distance between the umbos, and rib count—along with derived features used as predictors. These included the length-to-width ratio, length-to-height ratio, width-to-height ratio, umbo distance-to-length ratio, hinge line length-to-length ratio, umbo distance-to-height ratio, and rib density. The samples were classified by sex, with females labeled as 0 and males as 1, which served as the response variable.

3.7.1 Data Preprocessing

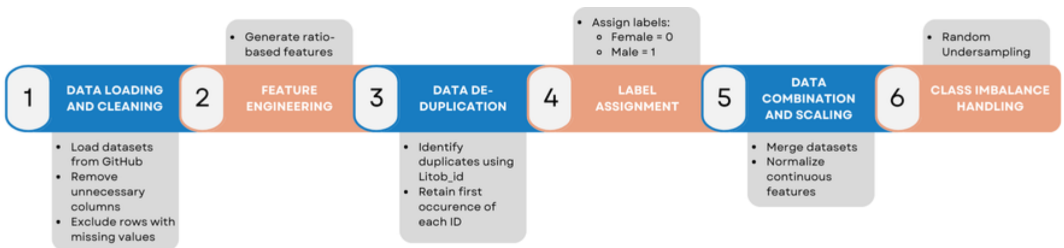

Figure 3.5: Data Preprocessing Pipeline

The preprocessing of the dataset involved several essential steps, carried out using Python in Google Colaboratory, in preparation for machine learning analysis (see Figure 3.5).

Data Loading and Cleaning

The process began by loading two separate datasets for male and female *T. granosa* directly from GitHub using `pd.read_csv()`. Unnecessary columns were removed, and rows containing missing values were excluded using the `dropna()` function to ensure data completeness and reliability.

Feature Engineering

Additional ratio-based features were generated to augment the existing measurements. These included the length-to-width ratio, length-to-height ratio, width-to-height ratio, hinge line length-to-length ratio, umbos distance-to-length ratio, umbos distance-to-height ratio, and rib density. These derived features aimed to emphasize shape characteristics independent of size, improving the models' ability to distinguish morphological differences between sexes.

Data De-duplication

To avoid redundancy and ensure each specimen was uniquely represented, the last three digits of each `Litob_id` were used to identify duplicates. Only the first occurrence of each unique ID was retained, reducing potential bias caused by repeated entries.

Label Assignment

A new column labeled `Label` was added to both datasets. Female specimens were assigned a label of 0, and male specimens a label of 1. This column served as the target variable for classification.

Data Combination and Scaling

After cleaning and feature engineering, the male and female datasets were merged into a single DataFrame. The `Litob_id` column was removed post deduplication. All continuous numeric features were normalized using `MinMaxScaler` to scale values to the range [0, 1].

Rib count was excluded from normalization because it is a discrete feature with biologically meaningful bounds. According to best practices in machine learning, normalizing discrete or categorical features can distort their meaning and is often unnecessary (Jaiswal, 2024). In this study, rib count was treated as a categorical attribute due to its biological significance and finite, non-continuous nature.

Class Imbalance Handling

After normalization, class imbalance was addressed by applying Random Undersampling to the male dataset. This technique randomly reduced the number of male samples to match the number of female samples (127 each), ensuring equal class representation. By using this approach, model bias was minimized, and the classification performance became more reliable across both classes.

3.7.2 Machine Learning Models Training

Model Selection and Hyperparameter Tuning

To establish a baseline for classification, various models were evaluated: Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, AdaBoost, Extra Trees, and Gradient Boosting. Hyperparameter tuning was conducted using `GridSearchCV`, which systematically identified the optimal settings for each model to enhance accuracy and performance.

Cross-Validation

A five-fold cross-validation approach was implemented. The dataset was divided into five subsets, with four used for training and one for testing. This process was repeated five times, with each fold serving as the test set once. This method ensured that model evaluation was robust and generalizable, minimizing the bias that may result from a single train-test split. (GeeksforGeeks, 2024)

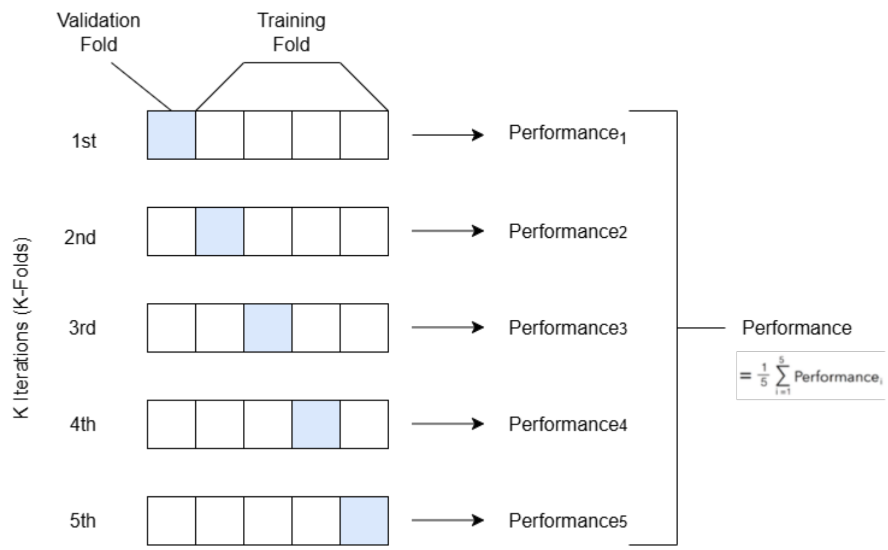

Figure 3.6: Diagram of k-fold cross-validation with k = 5

3.7.3 Evaluation Metrics for Machine Learning

Evaluating the performance of the binary classification model is important as well as selecting the appropriate metrics that is based on the requirements of the user. The performance of the supervised machine learning models will be measured based on four metrics namely: accuracy, precision, recall, and F1 score.

Accuracy (ACC) is the ratio of the overall correctly predicted samples to the total number of examples in the evaluation dataset (Cui, Pan, Chen, & Zou, 2020). The overall correctness of the model in predicting male and female blood cockles. This metric could help in understanding how well the model performs across all classifications. The formula for the accuracy is:

$$\text{ACC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

Precision (PREC) is the ratio between correctly predicted samples in all samples that are assigned to the positive class (Cui et al., 2020). This metric promotes fair representation and prevents the misidentification of blood cockles as it identifies potential inaccuracies or biases. The formula for precision is:

$$\text{PREC} = \frac{\text{True positive samples}}{\text{Samples assigned to class}} = \frac{TP}{TP + FP} \tag{3.2}$$

Recall (REC) is known as the sensitivity or the true positive rate (TPR) which is the ratio of the correctly predicted cases from all the samples assigned to the actual positive cases (Cui et al., 2020). This metric is the ability of the model to correctly identify positive male and female samples. The formula for the recall is:

$$\text{REC} = \frac{\text{True positive samples}}{\text{Samples classified positive}} = \frac{TP}{TP + FN} \tag{3.3}$$

F1 score is defined as the mean of the precision and recall in which it penalizes the extreme values of either of the two (Cui et al., 2020). The formula for the F1 is:

$$\text{F1} = \frac{precision \times recall}{precision \times recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.4}$$

Chapter 4

Results and Discussions

This chapter presents the results of the machine learning and deep learning analyses conducted on the preprocessed dataset. Preprocessing was performed using Python in Google Colaboratory. The chapter includes the evaluation of various machine learning classifiers, analysis of feature importance, and the application of deep learning models for image-based classification. These approaches aim to identify key predictors and assess classification performance for sex identification in *T. granosa*.

4.1 Machine Learning Analysis

4.1.1 Data Exploration

Exploratory data analysis was performed to characterize the dataset using visualizations to understand the patterns and correlations within the data. A correlation heatmap was created to assess the relationship between the predictors and the target variable.

The heatmap (see Figure 4.1) revealed three features most correlated with the sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos-length ratio ($r = 0.12$), and the distance between the umbos ($r = 0.12$). Each of these features demonstrated a weak positive relationship with the target variable.

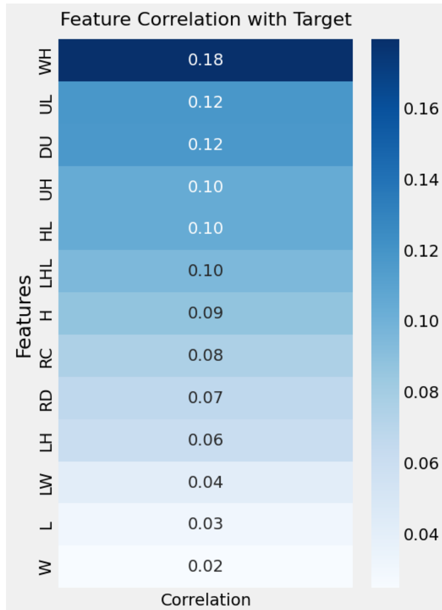

Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

4.1.2 Statistical Analysis

Variable	p-value
Length	0.334
Width	0.753
Height	0.124
Rib count	0.251
Length (Hinge Line)	0.120
Distance Umbos	0.025
LW_ratio	0.011
LH_ratio	0.490
WH_ratio	0.003
UL_ratio	0.019
HL_ratio	0.079
UH_ratio	0.036
Rib Density	0.181

Table 4.1: Mann-Whitney U Test Results for Sex-Based Feature Comparison

As part of the exploratory data analysis, statistical testing confirmed that the dataset did not follow a normal distribution. Consequently, the Mann-Whitney U test was applied with a significance level of $\alpha = 0.05$ to compare male and female samples. Out of thirteen features, five showed statistically significant differences. These included: distance between umbos ($p = 0.025$), length-width ratio ($p = 0.011$), umbos-length ratio ($p = 0.019$), width-height ratio ($p = 0.003$), and umbos-height ratio ($p = 0.036$).

It is important to note that statistical significance does not imply predictive importance. Therefore, further analysis, such as feature importance evaluation, was performed to identify the most informative predictors for classification.

4.1.3 Feature Importance Analysis

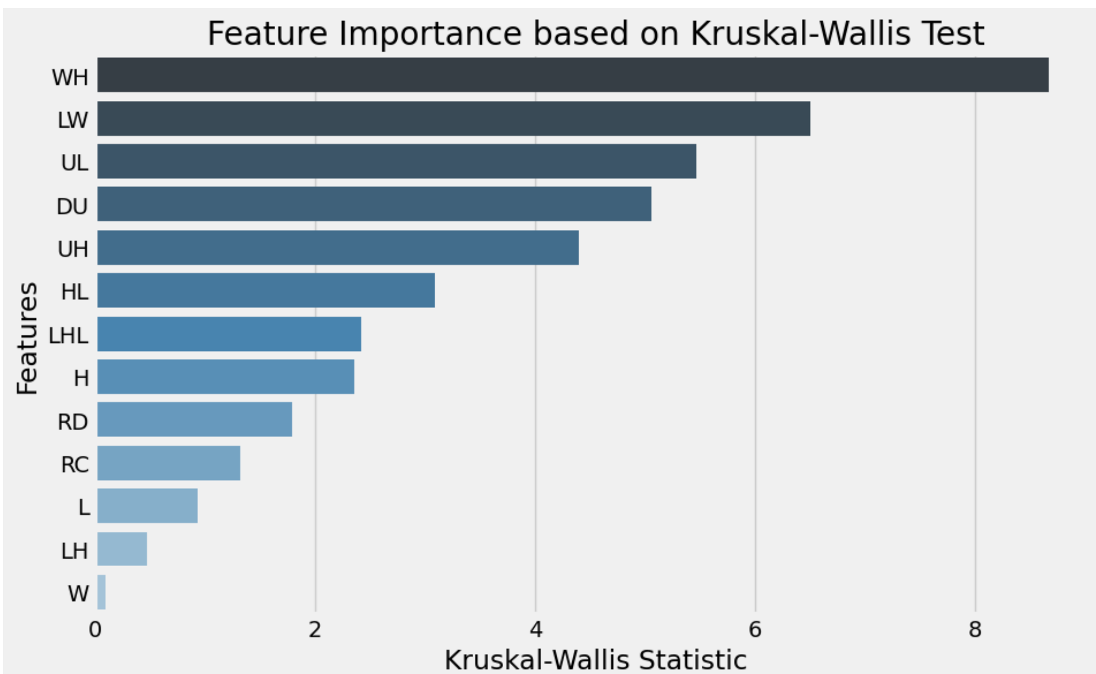

Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

Feature importance was assessed using the Kruskal-Wallis test, a non-parametric method that is suitable for evaluating differences in distributions across groups when the data does not follow a normal distribution. This approach was chosen because of the non-normality of the dataset and its robustness in handling continuous and ordinal data without assuming homogeneity of variances. (Ribeiro, 2024)

The analysis showed that the width-to-height ratio (WH_ratio) had the highest importance score, indicating it is the most statistically significant feature for distinguishing the sex of *T. granosa*. Other notable features included the length-to-width ratio (LW_ratio), umbos-to-length ratio (UL_ratio), and the distance between the umbos, all of which contributed significantly to the classification task.

4.1.4 Performance Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	58.62	58.62	58.62	58.44
Logistic Regression	57.83	57.83	57.83	57.61
K-Nearest Neighbors	51.18	51.31	51.18	50.77
Extra Trees	60.24	56.98	56.69	56.39
Random Forest	59.07	59.46	59.06	58.74
Gradient Boosting	60.27	60.98	60.27	59.96
AdaBoost	60.63	60.98	60.63	60.39

Table 4.2: Performance Metrics for Models with All 13 Features

In table 4.2, the performance of different machine learning models is presented using the full set of 13 features from the dataset. AdaBoost emerges as the highest-performing model, with an accuracy of 60.63%, precision of 60.98%, recall of 60.63%, and an F1-score of 60.39%. These results suggest that AdaBoost is particularly effective when utilizing all available features, likely due to its ability to combine multiple weak learners into a more robust model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	63.77	64.47	63.77	63.42
Logistic Regression	63.75	63.87	63.75	63.70
K-Nearest Neighbors	64.16	64.97	64.16	63.75
Extra Trees	62.20	59.69	59.08	58.69
Random Forest	62.96	60.10	59.85	59.54
Gradient Boosting	63.39	64.24	64.16	64.04
AdaBoost	61.02	61.26	61.02	60.82

Table 4.3: Performance Metrics for Models with 5 Features

Table 4.3 presents the performance of the same models using only the top 5 features identified through Kruskal-Wallis feature importance analysis. The top 5 features selected are distance between the umbos, length-to-width ratio, width-to-height ratio, umbos-to-height ratio and umbos-to-length ratio.

Interestingly, the performance of the models improves with the reduced feature set. K-Nearest Neighbors (KNN) achieves the highest performance in this scenario, with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1-score of 63.75%. These results suggest that KNN benefits from using only the most significant features, showing a notable improvement over its performance when all 13 features are used.

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

5.2 Recommendations

This special problem entitled Morphometric-Based Non-invasive Sex Identification of *T. granosa* focuses on creating a baseline study that will serve as a foundation for further studies involving *Tegillarca granosa*, blood cockles using machine learning, deep learning, and computer vision technologies in determining the sex of the samples is a salient need in aquaculture practices. Thus, the proceeding recommendations are the future applications to improve and have detailed analysis such as focusing on shape analysis, exploring other state-of-the-art CNN such as ResNet, SqueezeNet, and InceptionNet, and comparing the analysis result. Furthermore, the main goal of conducting this is to have the ability to identify the sex of the samples by taking real-time angles by rotating from the dorsal, lateral, and ventral.

Future studies could also invest in a much sturdier and more controlled environment by using a green background and positioning a webcam at a fixed angle. In addition, experiment with other image processing techniques such as scaling, rotating, and augmentation. The dataset can be utilized for further analysis using deep learning and computer vision to make sense of the images gathered and discern sexual dimorphism for *T.granosa* or will serve as the basis for conducting similar studies to other bivalve species.

References

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology*, *71*, 5–16. doi: 10.1080/11250000409356545

Afiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara granosa (l.) and anadara antiquate (l.) (bivalvia: Arcidae) in central java. , *10*.

Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian Sains*, *14*, 14207.

Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021). Data scaling performance on various machine learning algorithms to identify abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi: 10.14710/jtsiskom.2021.14105

Arkhipkin, A. I. (2005). Statoliths as 'black boxes' (life recorders) in squid. *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158

Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquaculture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Surtida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on mangrove-friendly aquaculture organized by the seafdec aquaculture department, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).

BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound, Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic Resources.

Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of waste cockle shell (anadara granosa) in biodiesel production from palm olein: Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358. doi: 10.1016/j.fuel.2011.03.002

Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining mechanisms in bivalves.* Preprints.org. doi: 10.20944/preprints201706.0127.v1

Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman, E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtdna: A new sex determination system in freshwater mussels (bivalvia: Unionoida)? *Molecular Biology and Evolution*, *28*(5), 1645–1659. doi: 10.1093/molbev/msq345

Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Approaches, challenges and opportunities for aquaculture. *Journal of Marine Science and Engineering*, *3*(2), 329–355. doi: 10.3390/jmse3020329

Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04). Mass mortalities in bivalve populations: A review of the edible cockle *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, *150*. doi: 10.1016/j.ecss.2014.04.011

Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic research in bivalves: Towards the identification of molecular markers of aquatic pollution. *Proteomic Research in Bivalves*, *75*(14), 4346–4359. doi: 10.1016/j.jprot.2012.04.027

Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly Review of Biology*, *18*, 154–164. doi: 10.1086/qrb.1943.18.issue-2

Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan sexual systems. *Integrative and Comparative Biology*, *53*, 723–735. doi: 10.1093/icb/ict076

Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023). Bivalvenet: A hybrid deep neural network for common cockle (cerastoderma edule) geographical traceability based on shell image analysis. *Ecological Informatics*, *78*, 102344. doi: 10.1016/j.ecoinf.2023.102344

Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method for chinese mitten crab using deep convolutional neural network. *Multimedia Tools and Applications*, *79*(11-12), 7669–7684. doi: https://doi.org/10.1007/s11042-019-08355-w

Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial exposure. *Aquaculture*, *56*(2), 151–162. Retrieved from `https://doi.org/10.1016/0044-8486(86)90024-4` doi: 10.1016/0044-8486(86)90024-4

Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect ageing of eel, anguilla anguilla (l.)? *International Review of Hydrobiology*, *75*(6), 737–743. doi: 10.1002/iroh.19900750607

Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and anatomy for young learners.* Rosie Research. Retrieved from `https://rosieresearch.com/clam-dissection-anatomy/`

Fao 2024 report: Sustainable aquatic food systems important for global food security – european fishmeal. (2024). `https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/`.

Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and elemental composition: Complementary tools for stock discrimination of mulloway (argyrosomus japonicus) in southern australia. *Fish Research*, *110*, 75–83. doi: 10.1016/j.fishres.2011.03.014

GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn.* Retrieved from `https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/` (Accessed: 2025-04-23)

Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture.* Oxford: Blackwell Science.

Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, *48*, 19–42. doi: 10.1111/bij.1993.48.issue-1

Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and surface water characterization and heavy metal health risk on cockles in kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp. 263–271). doi: 10.1016/j.sbspro.2016.05.156

Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a comprehensive guide to data rescaling.* DataCamp. Retrieved from `https://www.datacamp.com/tutorial/normalization-in-machine-learning` (Accessed 2025-04-23)

Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest example of sexual dimorphism in bivalves—evidence from the astartid *Nicaniella* (lower jurassic, southern germany). *Journal of Paleontology*, *95*(6), 1216–1225. doi: 10.1017/jpa.2021.48

Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4

Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep learning-based phenotype classification of three ark shells: Anadara kagoshimensis, tegillarca granosa, and anadara broughtonii. *Frontiers in Marine Science*, *11*. doi: 10.3389/fmars.2024.1356356

Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the granulated ark, tegillarca granosa (linne). *The Korean Journal of Malacology*, *13*, 55–64.

Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as a tool for stock identification of the southern blue whiting, micromesistius australis. *Latin American Journal of Aquatic Research*, *41*, 479–489.

Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., & et al. (2016). Identifying blue whiting (micromesistius poutassou) stock structure in the northeast atlantic by otolith shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*, 1363–1371. doi: 10.1139/cjfas-2015-0332

May, K., Maung, C., Phyu, E., & Tun, N. (2021). Spawning period of blood cockle tegillarca granosa (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.*

Arts Sci, *4*.

Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and larval rearing of the blood cockle, tegillarca granosa (linnaeus, 1758), in the philippines. *Asian Fisheries Science*, *36*(2). doi: 10.33997/j.afs.2023.36.2.001

Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of local populations of the common sole solea solea (pisces, soleidae) in the nw mediterranean through otolith morphometrics and shape analysis. *Marine Biology*, *151*(3), 997–1008. doi: 10.1007/s00227-006-0549-0

Narasimham, K. A. (1988). Taxonomy of the blood clams anadara (tegillarca) granosa (linnaeus, 1758) and a. (t.) rhombea (born, 1780).

Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge, M. C. M., Clay, J., … Troell, M. (2000). Effect of aquaculture on world fish supplies. *Nature*, *405*(6790), 1017–1024. doi: 10.1038/35016500

Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith shape of different fish species? *Journal of Morphology*, *267*(7), 750–757. doi: 10.1002/jmor.10439

Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail placostylus (pulmonata: Bothriembryontidae) on the isle of pines, new caledonia. *Journal of Molluscan Studies*, *86*(1), 35–41. doi: 10.1093/mollus/eyz031

Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test.* Retrieved from `https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/` (Accessed: 2025-04-23)

Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O., & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydrocarbons from consumption of blood cockle and exposure to contaminated sediments and water along the klang strait, malaysia. *Marine Pollution Bulletin*, *84*(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004

Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence of seasonal variation and anthropogenic stress on blood cockle (tegillarca granosa) production potential. *Influence of Seasonal Variation and Anthropogenic Stress on Blood Cockle (Tegillarca Granosa) Production Potential*, *44*(2), 62–82.

Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Machine learning and its applications to biology. *PLoS Computational Biology*, *3*(6), e116. doi: 10.1371/journal.pcbi.0030116

Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Reproductive process and early development. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370). College Park, MD: Maryland Sea Grant.

Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning approach for morphological feature extraction based on variational auto-encoder: An application to mandible shape. *Npj Systems Biology and Applications*, *9*(1), 1–12. doi: 10.1038/s41540-023-00293-6

Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in the laboratory, malaysia.* (Handle.net) doi: 10.3366/in_3366.pdf

Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, *C-21*, 269–281. doi: 10.1109/tc.1972.5008949

Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.

Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Impacts of four commonly used nanoparticles on the metabolism of a marine bivalve species, tegillarca granosa. *Chemosphere*, *296*, 134079. doi: 10.1016/j.chemosphere.2022.134079

Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the blood clam tegillarca granosa. *Marine Environmental Research*, *177*. Retrieved from `https://doi.org/10.1016/j.marenvres.2022.105632`

Appendix A

Data Gathering Documentation and Supplementary Analysis

Figure A.1: Sex Identification Through Spawning of *Tegillarca granosa*

Figure A.2: Separating Male and Female Samples After Spawning of *Tegillarca granosa*

Figure A.3: Sex Identified Female Through Dissecting of *Tegillarca granosa*

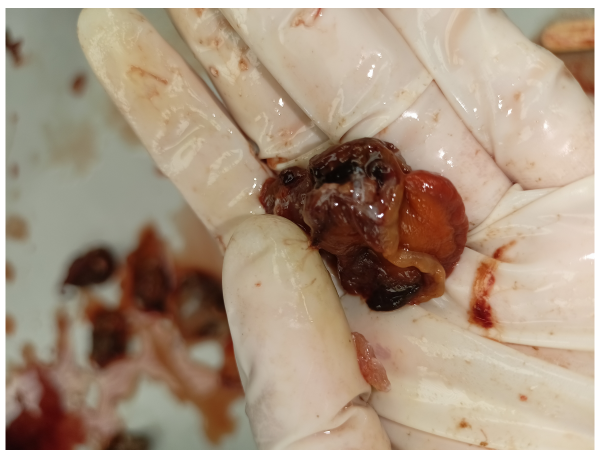

Figure A.4: Sex Identified Male Through Dissecting of *Tegillarca granosa*

Litob_id	Length	Width	Height	Rib count	Length (Hinge Line)	Distance Umbos
10001	48.05	37.6	32.15	20	33.55	4.1
20001	48.05	37.6	32.15	20	33.55	4.1
30001	48.05	37.6	32.15	20	33.55	4.1
40001	48.05	37.6	32.15	20	33.55	4.1
50001	48.05	37.6	32.15	20	33.55	4.1
60001	48.05	37.6	32.15	20	33.55	4.1
10002	47.4	32.5	32.25	20	33.1	3.05
20002	47.4	32.5	32.25	20	33.1	3.05
30002	47.4	32.5	32.25	20	33.1	3.05
40002	47.4	32.5	32.25	20	33.1	3.05
50002	47.4	32.5	32.25	20	33.1	3.05
60002	47.4	32.5	32.25	20	33.1	3.05
10003	43.3	34.1	31.25	21	32.05	4.5
20003	43.3	34.1	31.25	21	32.05	4.5
30003	43.3	34.1	31.25	21	32.05	4.5
40003	43.3	34.1	31.25	21	32.05	4.5
50003	43.3	34.1	31.25	21	32.05	4.5
60003	43.3	34.1	31.25	21	32.05	4.5
10075	50.05	35.05	32.05	21	30.05	4.1
20075	50.05	35.05	32.05	21	30.05	4.1

female

Figure A.5: Linear Measurements of Female *Tegillarca granosa*

Litob_id	Length	Width	Height	Rib count	Length (Hinge Line)	Distance Umbos
110004	43.1	33.05	28.15	21	28.5	3.05
120004	43.1	33.05	28.15	21	28.5	3.05
130004	43.1	33.05	28.15	21	28.5	3.05
140004	43.1	33.05	28.15	21	28.5	3.05
150004	43.1	33.05	28.15	21	28.5	3.05
160004	43.1	33.05	28.15	21	28.5	3.05
110005	41.1	31.05	27.6	20	23.05	3.35
120005	41.1	31.05	27.6	20	23.05	3.35
130005	41.1	31.05	27.6	20	23.05	3.35
140005	41.1	31.05	27.6	20	23.05	3.35
150005	41.1	31.05	27.6	20	23.05	3.35
160005	41.1	31.05	27.6	20	23.05	3.35
110006	43.2	33.45	29.35	20	29.35	3.3
120006	43.2	33.45	29.35	20	29.35	3.3
130006	43.2	33.45	29.35	20	29.35	3.3
140006	43.2	33.45	29.35	20	29.35	3.3
150006	43.2	33.45	29.35	20	29.35	3.3
160006	43.2	33.45	29.35	20	29.35	3.3
110007	41.5	32.55	27.7	20	24.1	3.7
120007	41.5	32.55	27.7	20	24.1	3.7

male

Figure A.6: Linear Measurements of Male *Tegillarca granosa*

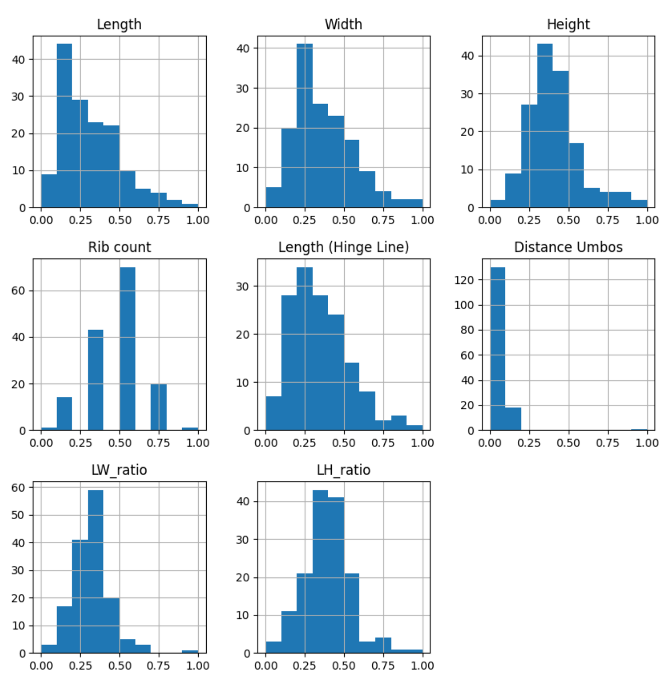

Figure A.7: Distribution of the Features of *Tegillarca granosa*