

A NON-INVASIVE SEX IDENTIFICATION OF BLOOD
COCKLES *TEGILLARCA GRANOSA* (LINNAEUS,
1758) USING MACHINE LEARNING

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

December 1, 2024

Abstract

Tegillarca granosa (Linnaeus, 1758), known as blood cockles, is one of the most well-known marine bivalves due to its nutritional benefits and one of the major sources of livelihood. Due to these, it is crucial to determine their sex to maintain a balanced male-to-female ratio to avoid exploitation of this shellfish commodity. 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 paper proposes a classification approach for the blood cockle's sex analyzed based on the linear measurements (length, width, height, distance hinge line, distance between umbos, rib count) and different angles (dorsal, ventral, anterior, posterior, left, and right lateral). This study aims to discern the sexual dimorphism present between the male and female *T. granosa* utilizing supervised machine learning models for linear measurements (logistic regression, random forest, support vector machine, k-nearest neighbors, and extreme gradient boosting) as well as Convolutional Neural Networks (CNN) for image analysis using the widely used models for bivalve studies (VGGNet, Inception-ResNet, SqueezeNet).

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

Contents

1	Introduction	1
1.1	Overview	1
1.2	Problem Statement	2
1.3	Research Objectives	3
1.3.1	General Objective	3
1.3.2	Specific Objectives	3
1.4	Scope and Limitations of the Research	3
1.5	Significance of the Research	4
2	Review of Related Literature	6
2.1	Background on <i>Tegillarca granosa</i> and Their Importance	7
2.2	Current Methods of Sex Identification in <i>Tegillarca granosa</i>	9
2.3	Machine Learning and Deep Learning in Biological Studies	11
2.3.1	Deep Learning for Phenotype Classification in Ark Shells .	11
2.3.2	Geometric Morphometrics and Machine Learning for Species Delimitation	12
2.3.3	Contour Analysis in Mollusc Shells Using Machine Learning	12
2.3.4	Machine Learning for Shape Analysis of Marine Organisms	13

2.3.5	Deep Learning for Landmark-Free Morphological Feature Extraction	14
2.3.6	Machine Learning for Sex Differentiation in Abalone	14
2.3.7	Machine Learning for Geographical Traceability in Bivalves	15
2.4	Limitations on Sex Identification in <i>Tegillarca granosa</i>	16
2.5	Synthesis of the Study	17
3	Research Methodology	20
3.1	Sample Collection	20
3.2	Ethical Considerations	22
3.3	Creating <i>T. granosa</i> Dataset	22
3.4	Morphological and Morphometric Characteristics Collection	23
3.5	Image Acquisition and Pre-Processing	24
3.6	Hardware and Software Configuration	25
3.7	Data Augmentation	25
3.8	Machine and Deep Learning Technologies	26
3.9	Morphometric Characteristics Evaluation using Machine Learning	26
3.9.1	Extreme Gradient Boosting	26
3.9.2	Logistic Regression	27
3.9.3	Random Forest	27
3.9.4	Support Vector Machine	27
3.9.5	k-Nearest Neighbors	28
3.10	Deep Learning for Image-Based Classification	28
3.10.1	CNN Model: Visual Geometry Group	28

3.10.2 CNN Model: Inception ResNet	28
3.10.3 CNN Model: SqueezeNet	29
3.11 Evaluation Metrics for Machine Learning and CNN Models	29
4 Preliminary Results/System Prototype	31
4.1 Data Summary	31
4.1.1 Dataset Overview	31
4.1.2 Preprocessing Results	31
4.2 Hyperparameter Optimization	32
4.3 Tuned Classifier Comparison	33
4.4 Comparison of Model Performance	36
4.5 Feature Importance Analysis	36
References	38
A Appendix Title	43

List of Figures

2.1	Diagram of <i>Tegillarca granosa</i> Anatomy	7
3.1	Male and Female <i>Tegillarca granosa</i> shells	21
3.2	Different Views of the <i>T. granosa</i> Shell Captured	23
3.3	Linear Measurements of <i>Tegillarca granosa</i> shell.	24
3.4	Image Acquisition Setup for <i>T. granosa</i> Samples	25
4.1	Confusion Matrices of the Machine Learning Models	35

List of Tables

2.1	Comparison of the methods used in bivalves studies	18
4.1	Best parameters and accuracy scores for the machine learning models.	32
4.2	Balanced accuracy and training time for each machine learning model.	33
4.3	Model Performance Comparison	36

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 considered to be 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 are blood cockles (*Tegillarca granosa*) which hold considerable economic and environmental significance.

Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent overharvesting and ensure sustainable production. An imbalanced ratio can lead to overexploitation and can impact the population's sustainability. However, there is limited literature on *T. granosa* that has a thorough understanding of its sex-determining mechanisms, particularly concerning 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 aquaculture commodities sex identification using machine learning and deep learning, there is a notable scarcity of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

This study, titled "A Non-Invasive Sex Identification of *T. granosa* using Ma-

chine Learning,” aims to provide a comprehensive analysis of blood cockles by leveraging their morphological and morphometric characteristics. Sexual dimorphism in bivalves is hardly expressed macroscopically (Karapunar, Werner, Fürsich, & Nützel, 2021). However, by integrating machine learning and computer vision techniques, 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 to promote sustainable aquaculture and biodiversity by maintaining a balanced male-to-female ratio. A balanced ratio helps prevent overharvesting. Although sex identification is important for blood cockle population management and sustainable aquaculture, there is a notable lack of research in creating non-invasive methods to identify the sex of *T. granosa*. Many of the latest studies and approaches are based on invasive methods like dissection or histological analysis, which are impractical for large-scale aquaculture operations focused on conservation.

The existing invasive methods for identifying the sex of *T. granosa* often require dissection, a technique that involves cutting open the shell to visually inspect the gonads (Erica, 2018). This causes harm and death to the specimens. In some cases, histological examination is used to examine tissue samples through a microscope, leading to further destruction of the organism (May, Maung, Phy, & Tun, 2021). These methods are time-consuming, labor-intensive, and can pose a threat to population management, especially when it is essential to maintain a balanced sex ratio for breeding programs. Moreover, invasive methods also require technical skills to execute properly. Aquaculture operations, particularly in resource-limited settings, face challenges in accessing laboratory equipment like microscopes and staining tools which complicate the process.

A less invasive approach employed by aquaculturists is to monitor spawning behavior in which individuals are separated and stimulated to reproduce to determine their sex through the release of gametes (Miranda & Ferriols, 2023). Although it is indeed less invasive than dissection, spawning still involves inducing 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, this highlights the need for a more advanced approach. An alternative, non-invasive method involving machine and deep learning technologies might solve these issues by pro-

viding 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 De-

velopment: 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 will work with 500 spawned blood cockles taken from Panay Island, specifically from Zarraga Iloilo and Ivisan Capiz, equally divided between 250 males and 250 females, obtained through induced spawning through temperature shock. The researchers will personally gather linear measurements, including length, width, height, rib count, length of the hinge line, and distance between the umbos using 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.

Data collection will take place at the hatchery facility of the University of the Philippines Visayas and will be conducted in batches, depending on the availability of spawned samples.

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).

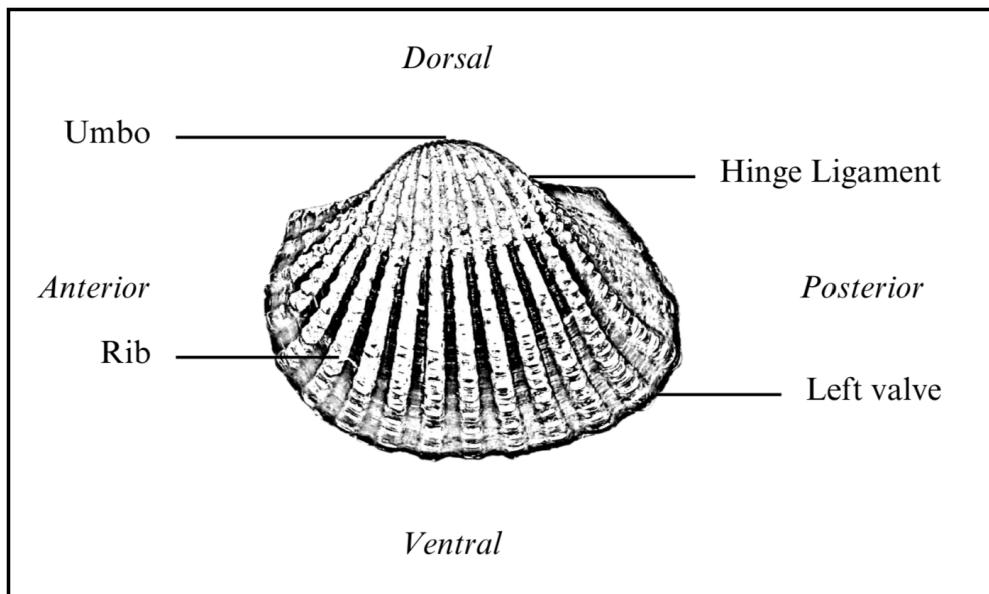


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

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 groove 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 (Tarcă, 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, Letourneau, & 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 interven-

tion. 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, Non-invasive Sex Identification of *T. granosa* using machine learning.

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 <i>T. granosa</i> in the Philippines.	Employed less invasive techniques which minimize the stress in <i>T. granosa</i> 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 <i>Nicanella rakoveci</i> , 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 <i>T. granosa</i> 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 <i>T. granosa</i> 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 Quennec 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 <i>Placostylus</i> 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-Ferret 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 discusses the materials and methods to be employed in the study, focusing on the development requirements and the software, and languages utilized. This will also entail the overall workflow in conducting the study, Non-Invasive Methods in Determining the Sex of *Tegillarca granosa* (blood cockles) using machine learning technologies. The different machine and deep learning and computer vision technologies will be thoroughly discussed to ensure a comprehensive understanding of the entity of the research endeavor and its processes.

Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture, will oversee the overall workflow and conduct of this experiment. The researchers will also be guided by the research associates, LC Mae Gasit and Allena Esther Artera. Consequently, the whole dataset collection process will be done at the University of the Philippines Visayas hatchery facility.

The methodology consists of ten 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) Data Augmentation (8) Machine and Deep Learning Technologies, (9) Machine Learning Models for Pre-evaluation, and (10) Deep Learning for Image-Based Classification.

3.1 Sample Collection

The collection of *T. granosa* samples used in this study is part of an ongoing research project 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.” This research project provides a total of 500 adult *T. granosa* samples which undergo spawning and are classified by sex as either male or female. The samples, ranging in size from 34 to 61 mm, are sourced from the coastal area of Zaraga, Iloilo, Philippines and fish markets in Ivisan, Capiz, Philippines.

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

Spawning is induced through temperature fluctuations which is the most natural and least invasive method for bivalves compared to other methods (Aji, 2011). However, due to a shortage of female spawned samples, additional blood cockles were dissected. This dissection process is carried out personally by the researchers, assisted by hatchery staff. The sex of the dissected samples is identified based on the coloration of the gonad tissue, which varies by sex and maturity stage. Females exhibit orange-red to pale orange gonads while males display white to grayish-white gonads (May et al., 2021).

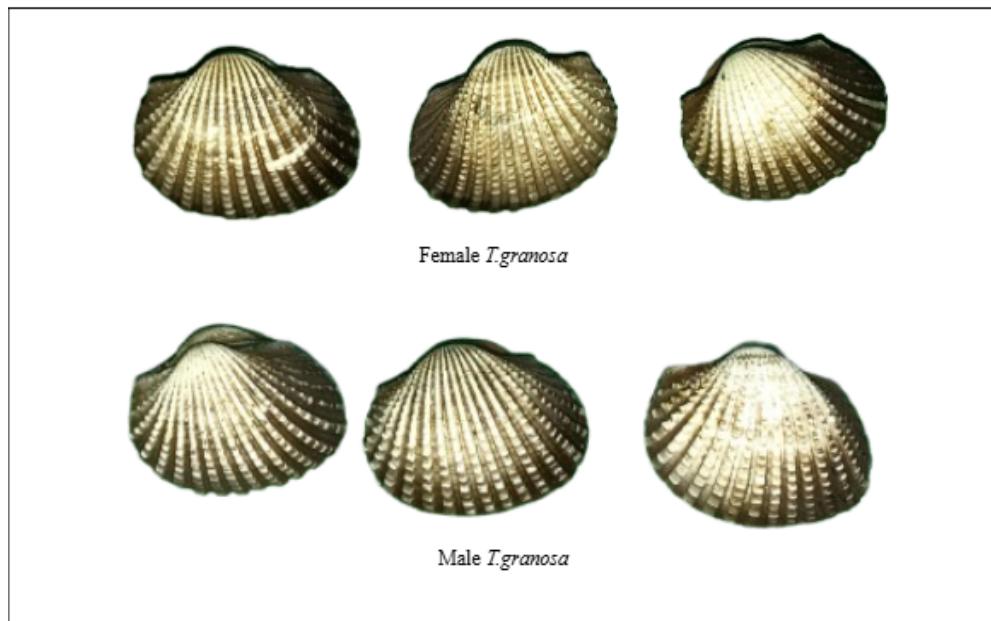


Figure 3.1: Male and Female *Tegillarca granosa* shells

3.2 Ethical Considerations

This study did not require ethical approval for working with animals, as per local legislation and institutional guidelines, since the experiments were performed on species commonly consumed as food by humans.

3.3 Creating *T. granosa* Dataset

For the initial preparation of the experiment, the researchers will collect primary observations for 100 samples of *T. granosa*. For the actual experimentation, the researchers will collect the original dataset by batch eventually comprising 500 samples of *T. granosa*. The linear measurements were measured and gathered manually by the researchers and compiled in the CSV file separating male and female *T. granosa*.

On the other hand, the images captured for the dataset will be saved in jpg format with a file naming convention of the sample's sex, the orientation or view of the shell, and its corresponding number out of the total 500 samples. Female *T. granosa* samples will begin with 0 in their file name, while males will begin with 1, followed by the views captured such as (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right lateral (refer to Figure 3.2), and lastly, a unique sample number. For example, “010001” will be the file name for the first female sample taken from the dorsal view, and “110001” for the first male sample also taken from the dorsal view.

The dataset will be organized in a CSV file that lists each image's file name along with their shell's width, height, length, rib count, length of the hinge line, and distance between their umbos. This dataset will be essential for machine learning model training and testing.

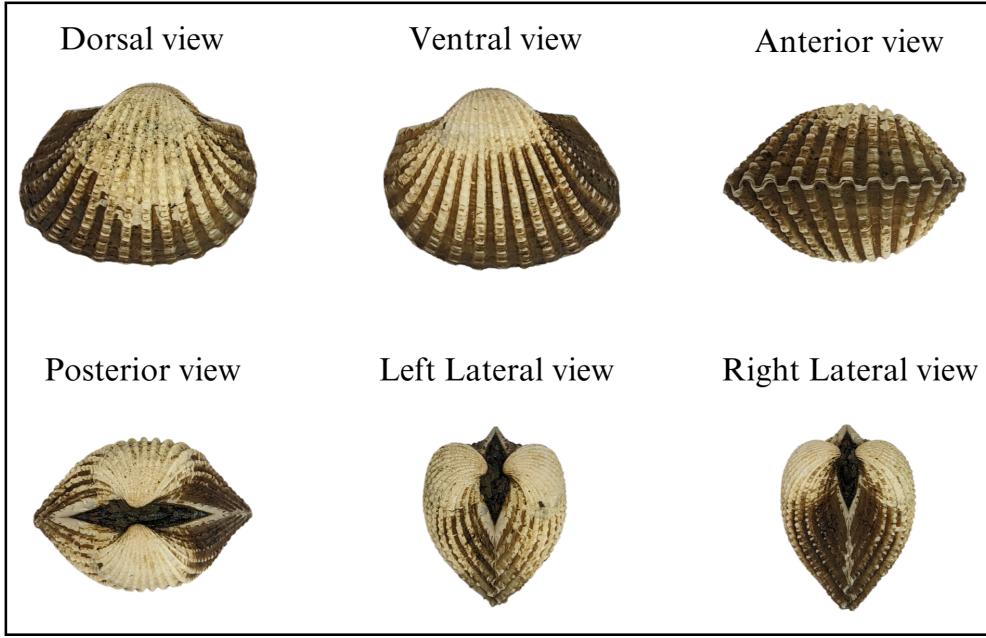


Figure 3.2: Different Views of the *T. granosa* Shell Captured

3.4 Morphological and Morphometric Characteristics Collection

Morphology refers to the biological form and represents one of the most visually recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023). Morphology is a term that describes structural characteristics by measuring specific components, namely, dimensions such as shapes, sizes, and colors. As stated by the researchers, quantifying and characterizing the shape is essential to understanding and visualizing the variations in *T. granosa*'s morphology.

In this study, the researchers are going to measure the height, width, and length of *T. granosa*. The dimensions will be recorded using a Vernier caliper to the nearest 0.01 mm. For the measurements, refer to Figure 3.3. The length (A) of the *T. granosa* refers to the measurement from the anterior to the posterior of the shell. The width (B) is measured through the shell's widest point from the left to the right valve. The height (C) is measured from the base of the shell to the shell's apex. The height of the gap between the valves (C) near the hinge will also be measured for the length of the hinge line and the distance between the umbos. The authors, Reymant and Kennedy (1998), indicated that the use of counts of the shell ribs as supplementary information increases identification

accuracy. Thus, the researchers will also take into account the difference in the rib count of the male and female *T. granosa*, and the ratio will be calculated since the sizes of the blood cockles vary. Sex ratio, size frequency distribution, and relative growth rates will be used to investigate sexual dimorphism.

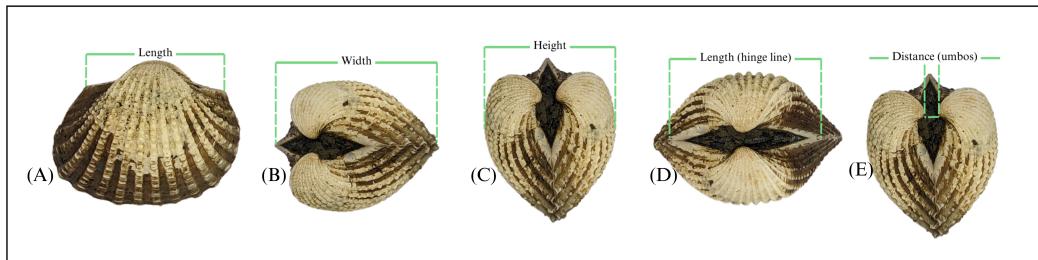


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

3.5 Image Acquisition and Pre-Processing

In this study, an original dataset of *Tegillarca granosa* will be built with 500 images (250 male blood cockles images and 250 female images). Both the male and female *T. granosa* images will contain 1500 images from different angles. During the data gathering process, there seemed to be an unbalanced number of spawned samples in favor with the male blood cockles. Thus, this can cause the data to be unbalanced. Unbalanced samples will lead to train model slowly and affect the gradient update. So, the dataset needs to meet sample equilibrium criteria (Cui & Zou, 2020). The researchers constructed a controlled environment for capturing the images of the samples utilizing a box-like structure with a white background surface. This setup was designed to maintain uniform captures of the images, and a consistent measurement between the sample and the camera, fixing the camera at a consistent angle above the *T. granosa*. To ensure the image quality, eliminate shadows, and clarity of the sample during the image acquisition process, the ring light will be placed in front of the box and use a camera without flash. Google Pixel 3 XL is the smartphone that will be utilized with the following specifications: 2960 x 1440 for the resolution, 4,032 x 3,024 pixels (12.2 MP) for the dimensions, f/1.8 for the fstop, 28mm (wide), $\frac{1}{2.55}$ ", 1.4 μ m, dual pixel PDAF, OIS (Concepcion et al., 2023)



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

3.6 Hardware and Software Configuration

The machine learning and deep learning algorithms would be trained on ACER Aspire 3 general processing unit (GPU) which has a central processing unit (CPU) of AMD Ryzen 3 7320U with Radeon Graphics (8) @ 2.395GHz and with a memory of 8 gigabytes (GB). The model will be performed on Keras which is a deep learning framework integrated with TensorFlow (Cui & Zou, 2020).

3.7 Data Augmentation

In the practice of deep learning, there is a possibility of a small dataset to being filled with bias and overfitting. The most obvious solution is to add more samples however, it is time-consuming and labor-intensive to collect a massive amount of data. In dealing with an unbalanced dataset, various data augmentation techniques can be used, such as rotation, scaling, shearing, and translation. These four data augmentation technique will be applied to generate 10 new images for each image. (Cui & Zou, 2020). Thus, from a total of 3000 images, where 6 angles are captured for each of the 500 samples, it would enlarge to 120,000 images that will be used in training the state-of-the-art CNN models.

3.8 Machine and Deep Learning Technologies

This section of the paper will discuss the technologies to be used in training, and testing the model as well as associated techniques and algorithms. Since obtaining the induced samples will be done per batch, the researchers will conduct an initial run with supervised machine learning models to determine the significance of the collected linear measurements (length, width, height, length hinge line, distance between umbos, and rib count) before delving into more complex methods such as deep learning CNN models.

3.9 Morphometric Characteristics Evaluation using Machine Learning

The shape of recording structures was first analyzed by collecting measurements of linear distances and applying multivariate statistical methods to these data (traditional linear measurement method) (Rohlf & Archie, n.d.). Geometric morphometric (GM) methods are an alternative way of analyzing and quantifying shape, which in theory retains more detail about the geometry of the structure than could be obtained from linear measurements (Adams et al., 2004). Machine learning techniques such as decision tree classification, support vector machines (SVMs), and k-nearest neighbors (KNN) have been applied to the analysis of bivalve shell geometry and morphology to classify shells based on morphological features, including shell shape, size, and texture, among others (Kiel, n.d.). The results of these studies have shown that machine learning algorithms can accurately classify bivalve shells and provide insights into the relationships between shell morphology and various environmental factors. Following this, the researchers are going to conduct a pre-evaluation of the linear measurements for 100 samples of *T. granosa* using a Support Vector Machine in order to determine whether these measurements can serve as a reliable factor in determining the sex of the samples before proceeding to more complex methods.

3.9.1 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a DT-based ensemble machine learning algorithm that combines the gradient boosting framework and decision trees to reach a final decision. It has the ability to achieve improved performance and higher accuracy compared to other supervised machine learning models. This

model would be beneficial in terms of its key capabilities of computation of the importance of the attributes based on the overall contribution to the accuracy of the predictions (Torres, 2023).

3.9.2 Logistic Regression

Logistic Regresion (LR) is the process of modeling the probability of a discrete outcome given the input variable. The outcomes in the logistic regression is measured using a binary variable and is a transformation of the linear regression using the sigmoid function (Cui & Zou, 2020). Although there are no related studies about sex identification in mollusks, logistic regression is a useful analysis method in classification problems such as determining the sexual dimorphism given the linear measurements measured.

3.9.3 Random Forest

Random Forest (RF) is a type of supervised machine learning algorithm that is often used in regression and classification problems (Cui & Zou, 2020). It has also been utilized in mollusks studies such as sex identification of abalone wherein the Random Forest achieved the highest average balance accuracy for all the datasets (Arifin et al., 2021).

3.9.4 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that classifies data by finding the optimal hyperplane which is also used in classification and regression problems (Cui & Zou, 2020). Support vector machine is also widely used in aquaculture related studies such as the sex determining mandible shapes due to its ability to solve classification problems with high validation accuracy especially when the clusters of the data with varying labels. The SVM has the ability to reflect the degree of cluster separation. It has the ability to be used as a tool in the variety of the image data with biological shapes (Tsutsumi et al., 2023).

3.9.5 k-Nearest Neighbors

The k-Nearest Neighbor (KNN) algorithm is a method commonly used for regression and classification tasks. It was applied in the study "Identification of Asian Green Mussel *Perna Viridis'* Sex Using Image Processing, Fuzzy Logic, and K-Nearest Neighbor," where KNN achieved 100% accuracy in classifying the sex of 50 male and 50 female mussels during the initial training phase with a dataset of 100 samples (Magbayao, Arboleda, & Galas, 2020).

3.10 Deep Learning for Image-Based Classification

After collecting a sufficient number of images and identifying initial patterns, convolutional neural networks (CNNs) will be used. CNNs, models like VGGNet, Inception ResNet, and SqueezeNet have been effectively applied in phenotype classification (Kim et al., 2024). In this study, the deep learning model will be specifically adapted for the sex identification of *T. granosa* based on shell images. CNN has achieved breakthroughs in terms of image classification, image segmentation, and speech recognition (He & Zhang, 2018). In this study, CNNs will analyze the images and learn important details about their shapes that can help identify whether they are male or female. Due to the limitation of literature focusing on the sex determining mechanism for bivalves, particularly *T. granosa*, the researchers will implement and compare three of the state-of-the-art CNN models namely VGGNet, Inception ResNet, and SqueezeNet Kim et al. (2024).

3.10.1 CNN Model: Visual Geometry Group

Visual Geometry Group (VGG) is considered to be a standard Convolutional Neural Network (CNN) architecture that is comprised of multiple layers that has been a ground-breaking object detection models and one of the most popular image recognition architecture (Boesch, 2021).

3.10.2 CNN Model: Inception ResNet

Inception ResNet is a convolutional neural network that has been trained on more than a million images and has 164 deep layers that have the capacity to classify

varied images (MathWorks, n.d.). This model is mainly used in bivalve studies such as in identifying different ark shells (Kim et al., 2024).

3.10.3 CNN Model: SqueezeNet

SqueezeNet is particularly advantageous because it reduces the number of parameters and amount of memory required to store the model without sacrificing accuracy (Koonce, n.d.). Its ability to achieve high accuracy in classifying shell images makes it a suitable choice for distinguishing between male and female *T. granosa*.

3.11 Evaluation Metrics for Machine Learning and CNN Models

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 three 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 & 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:

$$\text{PREC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

Precision (PREC) is the ratio between correctly predicted samples in all samples that are assigned to the positive class (Cui & Zou, 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:

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

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 & Zou, 2020). This metric is the ability of the model to correctly identify positive male and female samples. The formula for the recall is:

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

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

$$\text{REC} = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3.4)$$

Chapter 4

Preliminary Results/System Prototype

This chapter presents the preliminary results between the comparison of trained machine learning models: Logistic Regression, Random Forest, Decision Trees, Support Vector Machine (SVM), and k-nearest Neighbors (KNN). This chapter will also present the model evaluation and the comparison of model performance.

4.1 Data Summary

4.1.1 Dataset Overview

The dataset contains the morphological measurements collected from the 50 male and 50 female *T. granosa* samples. The morphological characteristics included in the dataset are the length, width, height, rib count, hinge length, and distance between the umbos. Additional ratio-based characteristics, such as Length-to-Width ratio, and Length-to-Height ratio, were computed to take into consideration the size variations. These ratios were calculated by dividing the shell length by the shell width and height, respectively.

4.1.2 Preprocessing Results

The preprocessing of the data includes data cleaning, feature engineering, and encoding. Any missing data entries were removed throughout the data cleaning

process. The LW and LH ratios were calculated and added by feature engineering to standardize size variations among the samples while maintaining significant morphological patterns. Additionally, a categorical label was encoded using binary values. It was done by indicating male as 1 and female as 0. As for the final preparation, the dataset was split into 70% as the training and 30% as testing to ensure balanced male and female sample representation in the training set as well as in testing sets.

4.2 Hyperparameter Optimization

Hyperparameter optimization was conducted using Grid Search Cross-Validation (GridSearchCV) to improve the performance and reliability of the machine learning models. This was done systematically by fine-tuning key parameters to find the best combination that maximizes model accuracy, precision, recall, and F1-score (Belcic and Stryker, 2024). Table 4.1 shows the summarized resulting parameters and corresponding scores for each model.

Model	Best Parameters	Score
Logistic Regression	{'classifier_C': 5}	0.744654
Decision Tree	{'criterion': 'gini', 'max_depth': 10}	0.986164
Random Forest	{'n_estimators': 50}	0.991195
SVM	{'classifier_C': 20, 'classifier_kernel': 'rbf'}	0.862893
KNN	{'classifier_metric': 'manhattan', 'classifier_n_neighbors': 9, 'classifier_weights': 'distance'}	0.991195

Table 4.1: Best parameters and accuracy scores for the machine learning models.

For Logistic Regression, the best hyperparameter was the regularization strength, C (1, 5, and 10), with the best score being obtained when C=5. The score was 0.7447, which implies a balance between model complexity and overfitting prevention.

The Decision Tree model was optimized by tuning the criterion for splitting (gini and entropy) and the maximum tree depth (5 and 10). The best combination used the Gini impurity criterion and a maximum depth of 10. It resulted in a score of 0.9862. This setup effectively balanced the trade-off between catching data patterns and preventing over-complexity.

For the Random Forest model, n_estimators (10, 15, 20, 50, 100, and 200) were tuned. Using 50 estimators gave the best performance with a score of 0.9912 which indicates the strength of the ensemble model in capturing diverse data patterns and reducing variance.

A support vector machine was tuned using regularization parameters (1, 10,

and 20) and kernel (rbf and linear) where RBF kernel and a regularization parameter of 20 showed superior performance with a score of 0.8629. This configuration well captures how the model can really handle the non-linear relationship in the dataset.

The k-Nearest Neighbors (KNN) algorithm was tuned using the number of neighbors (3, 5, 7, and 9), distance metric (euclidean and manhattan), and weighting scheme (uniform and distance), with n_neighbors of 9, manhattan distance metric, and distance-based weighting yielding a satisfactory score of 99.1%.

4.3 Tuned Classifier Comparison

After hyperparameter tuning, the performance of the classifiers was re-evaluated to compare their efficiency and accuracy. The evaluation of the models considered balanced accuracy and training time to know which model is more appropriate for practical use in predicting sex in *T. granosa*. Table 4.2 shows the balanced accuracy percentages obtained by each tuned classifier along with the corresponding training times.

Model	Balanced Accuracy (%)	Training Time (s)
Logistic Regression	74.35	3.07
Decision Tree	95.44	0.16
Random Forest	98.73	1.29
SVM	74.64	0.12
KNN	83.05	0.15

Table 4.2: Balanced accuracy and training time for each machine learning model.

Among the models, Decision Tree had the highest balanced accuracy of 99.44% and took 0.16 seconds for training. Random Forest showed a balanced accuracy of 98.73% but took 1.29 seconds for training. The model SVM was also very close with a training time of 0.12 seconds but a much lower balanced accuracy at 74.64%. In terms of training time, Logistic Regression was the slowest to train at 3.07 seconds and recorded the lowest accuracy at 74.35%. On the other hand, KNN showed a pretty balanced accuracy of 83.05% and took only 0.15 seconds to train. However, it wasn't as good as Decision Tree and Random Forest. These results indicate that while the Decision Tree proved to be the most accurate and fastest classifier for the dataset, the Random Forest had comparable accuracy with a slightly longer training time.

Figure 4.1 shows the confusion matrices that allow for a detailed breakdown of classifier predictions including true positives (correctly identified females), true negatives (correctly identified males), false positives (males incorrectly classified as females), and false negatives (females incorrectly classified as males).

Logistic Regression had achieved 69 true positives and 104 true negatives. However, there were also 22 false positives and 33 false negatives, so it actually has not been optimal and so does not differ as perfectly between male and female samples of *T. granosa*, agreeing to its low balanced accuracy with 74.35%. The large number of the wrongly classified data points can only indicate that Logistic Regression suffers from the complexity of this dataset.

The Decision Tree classifier gave perfect results, true positives at 102, and true negatives at 126, without any false positives or negatives. This reflects the ability of the model to capture very fine patterns in the morphological traits, thus having an excellent balanced accuracy of 99.44%.

Similar to the Decision Tree, Random Forest achieved perfect classification with 102 true positives and 126 true negatives with no false positives or false negatives. The ensemble approach of this model contributed to its robustness, which gave it an excellent balanced accuracy of 98.73%.

SVM correctly classified 87 samples as true positives and 124 as true negatives but had 2 false positives and 15 false negatives. Although better than Logistic Regression, misclassifications in both indicate difficulties in capturing all the complexities of the dataset completely. It is evident from the fact that its balanced accuracy is also lower at 74.64%.

KNN was perfect in classification results with 126 true positives and 102 true negatives, and no misclassifications. Its balanced accuracy is 83.05%, which shows that it has a good ability to work well with local data patterns.

The confusion matrix results show that Decision Tree, Random Forest, and KNN performed great in classifying male and female *T. granosa* with no mistake. In contrast, Logistic Regression and SVM showed more errors, and their performance underscores the limitations of such models. These results prove that non-parametric methods like KNN and ensemble approaches like Random Forest are suitable for this kind of data set of morphological features.

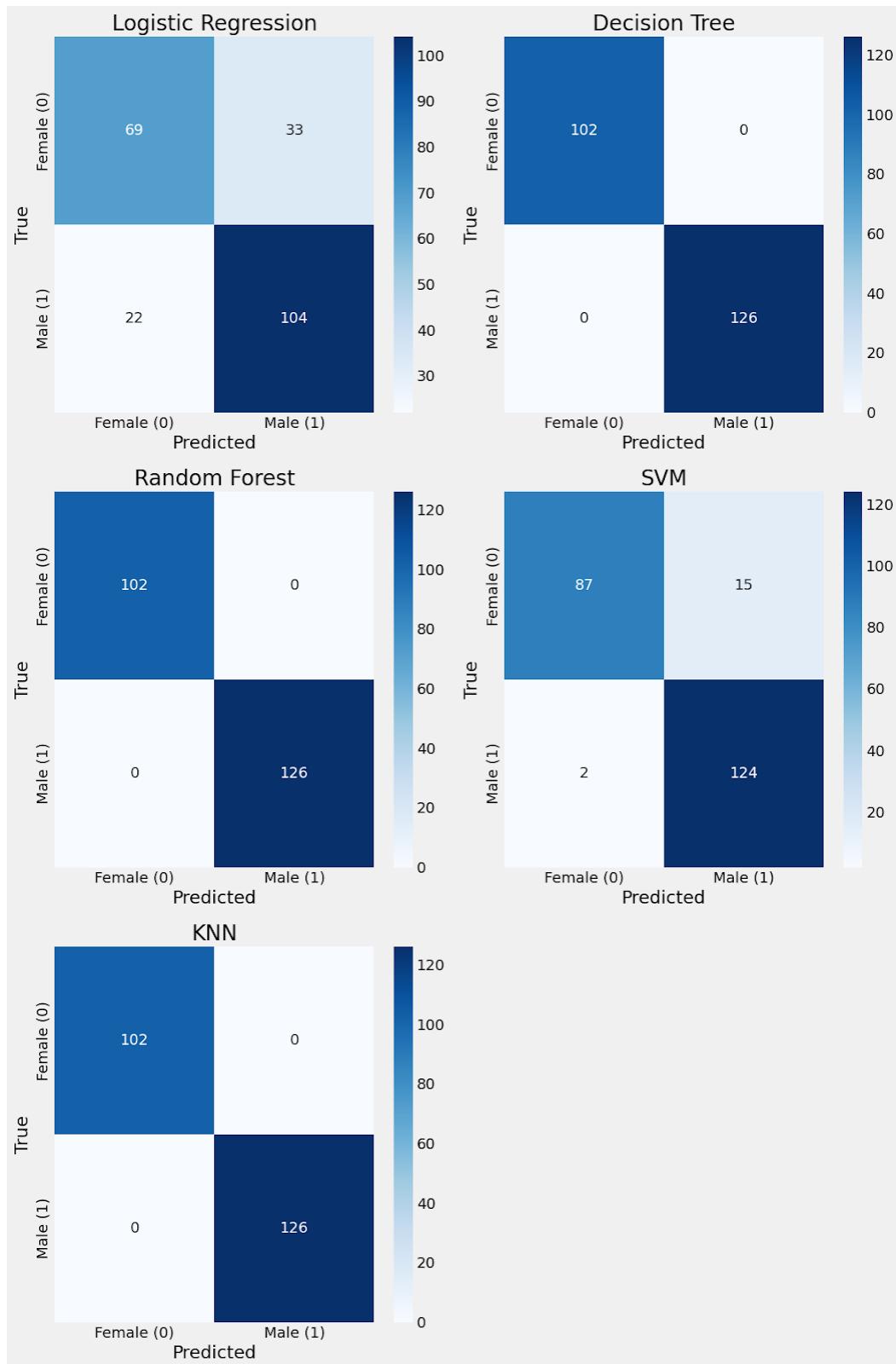


Figure 4.1: Confusion Matrices of the Machine Learning Models

4.4 Comparison of Model Performance

To evaluate the performance of the different models used, the effectiveness of the models were evaluated and compared in predicting the sex of the T.granosa based on morphological traits. The use of performance metrics such as accuracy, precision, recall, and F1-score is used to evaluate the performance of the different models. By analyzing the performance metrics the researchers can identify the most effective and best model for the classification of male and female T.granosa.

Model	Ave Accuracy (%)	Ave Precision (%)	Ave Recall (%)	Ave F1-score (%)
Logistic Regression	78.86	74.96	74.86	74.55
Decision Tree	98.11	98.22	98.11	98.12
Random Forest	98.68	98.74	98.68	98.68
SVM	87.34	87.79	87.34	87.22
KNN	99.43	99.45	99.43	99.43

Table 4.3: Model Performance Comparison

Table 4.3 presents the comparison results of machine learning models on the morphological traits of the combined male and female T.granosa datasets. The results indicate that all models demonstrated high performance in predicting male and female, with F1-score ranging between 74-99%. The KNN model achieved the highest accuracy (99.43%), precision (99.45%), recall (99.43%) and F1-score (99.43%), followed by random forest, decision tree, and SVM respectively. However, the logistic regression was the worst classifier having an accuracy of (74.86%), precision (74.96%), recall (74.86%) and (74.55%) F1-score. Overall, the results seen in this comparison highlights that machine learning models are highly effective in predicting sex identification of T.granosa based on their morphological features. With KNN being the best model it suggests that non-parametric data and pattern recognition (H. Benhar, 2020) is well suited for this dataset.

4.5 Feature Importance Analysis

After processing the dataset and splitting it into training and testing sets, the models are trained and their important features are computed. Feature Analysis helps in identifying which morphological features contribute most in classifying male and female T.granosa. The study uses models such as decision trees, and random forests to generate the scores. The models explore a wide range of features and then produce a score that reflects the feature's importance in predicting the variable (Anitha & Neelakandan, 2024).

To determine the important features that contribute to the classification of sex in *T.granosa*, two machine learning models, decision trees, and random forest were compared based on the importance of the features. The results in figures ? indicates variations in feature importance between two models, however certain features such as distance of the umbos, LW_ratio, length, and height consistently emerge as influential predictors. Hence, this analysis allowed the researchers to identify which features were most predictive that can improve the model's performance.

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 antiquata (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.
- Boesch, G. (2021, October 6). *Vgg very deep convolutional networks (vggnet) - what you need to know*. Retrieved from <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/> (Viso.ai)
- 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, P. T. C. S., Y., & 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>
- 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
- Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Blackwell Science.
- He, X., & Zhang, W. (2018). Emotion recognition by assisted learning with convolutional neural networks. *Neurocomputing*, *291*, 187–194.
- 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
- Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest example of sexual dimorphism in bivalves—evidence from the astartiid *Nicanella* (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
- Kiel, S. (n.d.). Assessing bivalve phylogeny using deep learning and computer vision approaches.
doi: <https://doi.org/10.1101/2021.04.08.438943>
- 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
- Koonce, B. (n.d.). SqueezeNet. *Convolutional Neural Networks with Swift for Tensorflow*, 73–85. *International Journal of Scientific & Technology Research*, *9*, 1199-1202.
- 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.
- Magbayao, R., Arboleda, E., & Galas, E. (2020, 01). Identification of asian green mussel *perna viridis'* sex using image processing, fuzzy logic and k -nearest neighbor. *International Journal of Scientific & Technology Research*, *9*, 1199-1202.
- 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
- MathWorks. (n.d.). *Pretrained inception-resnet-v2 convolutional neural network - matlab inceptionresnetv2*. Retrieved from <https://www.mathworks.com/help/deeplearning/ref/inceptionresnetv2.html> (MathWorks)
- May, K., Maung, C., Phy, 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
- Rohlf, F. J., & Archie, J. W. (n.d.). A comparison of fourier methods for the description of wing shape in mosquitoes (diptera: Culicidae). *Systematic Zoology*, 33(3), 302–302. doi: <https://doi.org/10.2307/2413076>
- 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.
- Torres, L. F. (2023, April). *Xgboost: The king of machine learning algorithms- latinixinai - medium. medium*. <https://medium.com/latinixinai/xgboost-the-king-of-machine-learning-algorithms-6b5c0d4acd87>.
- 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

Appendix A

Appendix Title