

¹ MORPHOMETRIC-BASED NON-INVASIVE SEX
² IDENTIFICATION OF BLOOD COCKLES *TEGILLARCA*
³ *GRANOSA* (LINNAEUS, 1758)

⁴ A Special Problem Proposal
⁵ Presented to
⁶ the Faculty of the Division of Physical Sciences and Mathematics
⁷ College of Arts and Sciences
⁸ University of the Philippines Visayas
⁹ Miag-ao, Iloilo

¹⁰ In Partial Fulfillment
¹¹ of the Requirements for the Degree of
¹² Bachelor of Science in Computer Science by

¹³ ADRICULA, Briana Jade
¹⁴ PAJARILLA, Gliezel Ann
¹⁵ VITO, Ma. Christina Kane

¹⁶ Francis DIMZON
¹⁷ Adviser
¹⁸ Victor Marco Emmanuel FERRIOLS
¹⁹ Co-Adviser

²⁰ January 24, 2025

Abstract

22 *Tegillarca granosa* (Linnaeus, 1758), commonly known as blood cockles, is one
23 of the most well-known marine bivalve for its nutritional benefits and economic
24 significance. Determining their sex is essential for maintaining a balanced male-
25 to-female ratio, which is crucial for preventing overexploitation of this shellfish
26 resource. The sex-determining mechanism in the shell morphology of bivalves is
27 challenging macroscopically due to the limited literature regarding this expertise.
28 In addition, no current technologies are employed to classify the sex based on shell
29 morphology. This study proposes a machine learning approach for classifying the
30 sex of blood cockles using various linear measurements (length, width, height,
31 distance between the hinge line, distance between umbos, and rib count) and
32 angles (dorsal, ventral, anterior, posterior, left lateral, and right lateral) collected
33 from male and female specimens. Available machine learning models in MATLAB
34 were trained to discern sexual dimorphism. Among the models, Linear SVM
35 performed best, achieving an accuracy of 69.80%, precision of 69.82%, recall of
36 69.80%, and an F1-score of 69.73%. Feature importance analysis indicated that
37 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

³⁹ Contents

⁴⁰ 1 Introduction	1
⁴¹ 1.1 Overview	1
⁴² 1.2 Problem Statement	2
⁴³ 1.3 Research Objectives	3
⁴⁴ 1.3.1 General Objective	3
⁴⁵ 1.3.2 Specific Objectives	3
⁴⁶ 1.4 Scope and Limitations of the Research	4
⁴⁷ 1.5 Significance of the Research	5
⁴⁸ 2 Review of Related Literature	6
⁴⁹ 2.1 Background on <i>Tegillarca granosa</i> and Their Importance	7
⁵⁰ 2.2 Current Methods of Sex Identification in <i>Tegillarca granosa</i>	9
⁵¹ 2.3 Machine Learning and Deep Learning in Biological Studies	11
⁵² 2.3.1 Deep Learning for Phenotype Classification in Ark Shells .	12
⁵³ 2.3.2 Geometric Morphometrics and Machine Learning for Species ⁵⁴ Delimitation	12
⁵⁵ 2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning	12
⁵⁶ 2.3.4 Machine Learning for Shape Analysis of Marine Organisms	13

57	2.3.5	Deep Learning for Landmark-Free Morphological Feature Extraction	14
58			
59	2.3.6	Machine Learning for Sex Differentiation in Abalone	15
60			
61	2.3.7	Machine Learning for Geographical Traceability in Bivalves	16
62	2.4	Limitations on Sex Identification in <i>Tegillarca granosa</i>	16
63	2.5	Synthesis of the Study	18
63	3	Research Methodology	21
64	3.1	Sample Collection	21
65	3.2	Ethical Considerations	23
66	3.3	Creating <i>T. granosa</i> Dataset	23
67	3.4	Morphological and Morphometric Characteristics Collection	24
68	3.5	Image Acquisition and Data Gathering	25
69	3.6	Hardware and Software Configuration	25
70	3.7	Morphometric Characteristics Evaluation Using Machine Learning	26
71	3.7.1	Preprocessing and Model Training	26
72	3.7.2	Evaluation Metrics for Machine Learning	28
73	4	Preliminary Results	29
74	4.1	Data Summary	29
75	4.1.1	Dataset Overview and Exploration	29
76	4.1.2	Statistical Analysis of <i>T. granosa</i> Features by Sex	31
77	4.2	Comparison of Model Performance	33
78	4.2.1	Performance Evaluation	33
79	4.2.2	Confusion Matrix Analysis	34

80	4.2.3 Feature Importance Analysis	35
81	5 Conclusion and Recommendations	38
82	5.1 Conclusion	38
83	5.2 Recommendations	38
84	References	39
85	A Data Gathering Documentation and Supplementary Analysis	44

⁸⁶ List of Figures

87	2.1 Diagram of <i>Tegillarca granosa</i> Anatomy	7
88	3.1 Male and Female <i>Tegillarca granosa</i> shells	22
89	3.2 Different Views of the <i>T. granosa</i> Shell Captured	24
90	3.3 Linear Measurements of <i>Tegillarca granosa</i> shell.	25
91	3.4 Image Acquisition Setup for <i>T. granosa</i> Samples	26
92	3.5 Preprocessing and Model Training Pipeline	27
93	4.1 Proportion of Male and Female <i>T. granosa</i>	30
94	4.2 Correlation Matrix of Predictors and Target Variables	30
95	4.3 Sex-based Differences in Morphological Features of <i>T. granosa</i>	32
96	4.4 Comparison of Model Performance	33
97	4.5 Confusion Matrix of Linear SVM	35
98	4.6 Feature Importance Scores Sorted Using the MRMR Algorithm.	35
99	4.7 Feature Importance Scores Sorted Using the Chi2 Algorithm.	36
100	4.8 Feature Importance Scores Sorted Using the ReliefF Algorithm.	36
101	4.9 Feature Importance Scores Sorted Using the Kruskal Wallis Algo- 102 rithm.	36
103	A.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i>	44

104	A.2 Separating Male and Female Samples After Spawning of <i>Tegillarca</i>	44
105	<i>granosa</i>	
106	A.3 Sex Identified Female Through Dissecting of <i>Tegillarca granosa</i> .	45
107	A.4 Sex Identified Male Through Dissecting of <i>Tegillarca granosa</i> . .	45
108	A.5 Linear Measurements of Female <i>Tegillarca granosa</i>	45
109	A.6 Linear Measurements of Male <i>Tegillarca granosa</i>	46
110	A.7 Distribution of the Features of <i>Tegillarca granosa</i>	46

¹¹¹ List of Tables

¹¹²	2.1 Comparison of the Methods Used in Bivalves Studies	19
¹¹³	4.1 Performance Metrics of Machine Learning Models for Sex Identifi-	
¹¹⁴	cation	33

¹¹⁵ **Chapter 1**

¹¹⁶ **Introduction**

¹¹⁷ **1.1 Overview**

¹¹⁸ The Philippines is a global center of marine biodiversity and has established aqua-
¹¹⁹ culture as a significant contributor to total fishery production (Aypa & Baconguis,
¹²⁰ 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-
¹²¹ ally and is the 11th largest seafood producer in the world. Aquaculture is deeply
¹²² integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-
¹²³ duction of various aquatic species, including bivalves. Among these, blood cockles
¹²⁴ (*Tegillarca granosa*) hold considerable economic and environmental significance,
¹²⁵ making it essential to ensure sustainable production and population balance.

¹²⁶ Maintaining a balanced male-to-female ratio of blood cockles is crucial to pre-
¹²⁷ vent overharvesting and ensure sustainability. An imbalanced ratio can lead to
¹²⁸ overexploitation and negatively impact the population's viability. However, there
¹²⁹ is limited literature on *T. granosa* that provides a thorough understanding of its
¹³⁰ sex-determining mechanisms, particularly regarding sexual dimorphism based on
¹³¹ morphological and morphometric characteristics (Breton, Capt, Guerra, & Stew-
¹³² art, 2017).

¹³³ Currently, sex determination methods for blood cockles are invasive, including
¹³⁴ dissection and histological examinations, which often result in the death of the
¹³⁵ species. While there is growing literature on sex identification in aquaculture
¹³⁶ commodities using machine learning and deep learning, there is a notable scarcity
¹³⁷ of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

¹³⁸ This study, titled "Morphometric-Based Non-Invasive Sex Identification of

¹³⁹ Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)," aims to provide a detailed
¹⁴⁰ baseline analysis of blood cockles by leveraging their morphological and morpho-
¹⁴¹ metric characteristics. Sexual dimorphism in bivalves is often subtle and chal-
¹⁴² lenging to establish mascropically (Karapunar, Werner, Fürsich, & Nützel, 2021).
¹⁴³ However, by integrating machine learning and deep learning, the study seeks to
¹⁴⁴ identify distinct features that may indicate sexual dimorphism between male and
¹⁴⁵ female blood cockles.

¹⁴⁶ 1.2 Problem Statement

¹⁴⁷ Identifying the sex of *T. granosa* is important for promoting sustainable aquacul-
¹⁴⁸ ture and biodiversity by maintaining a balanced male-to-female ratio. A balanced
¹⁴⁹ ratio helps prevent overharvesting. Although sex identification is crucial for blood
¹⁵⁰ cockle population management and sustainable aquaculture, there is a notable
¹⁵¹ lack of research on creating non-invasive methods for determining the sex of *T.*
¹⁵² *granosa*. Many recent studies and approaches rely on invasive methods like dis-
¹⁵³ section or histological analysis, which are impractical for large-scale aquaculture
¹⁵⁴ operations focused on conservation.

¹⁵⁵ Current methods for determining the sex of *T. granosa* are invasive and in-
¹⁵⁶ volve dissection, which requires cutting open the shell to visually inspect the
¹⁵⁷ gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-
¹⁵⁸ quently leads to their death. Another method is histological examination, where
¹⁵⁹ tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,
¹⁶⁰ 2021). Both approaches are labor-intensive and time-consuming, and can pose
¹⁶¹ risks to population management, particularly when maintaining a balanced sex
¹⁶² ratio for breeding programs is essential. Moreover, these invasive methods require
¹⁶³ specialized technical skills for accurate execution. Resource-limited aquaculture
¹⁶⁴ operations face significant challenges in accessing the necessary laboratory equip-
¹⁶⁵ ment, such as microscopes and staining tools, complicating the process.

¹⁶⁶ A less invasive approach employed by aquaculturists involves monitor spawning
¹⁶⁷ behavior, where individuals are separated and stimulated to reproduce in order
¹⁶⁸ to determine their sex through the release of gametes (Miranda & Ferriols, 2023).
¹⁶⁹ Although this method is indeed less invasive than dissection, it still induces stress
¹⁷⁰ in blood cockles and may not be completely effective for fast identification in large
¹⁷¹ populations.

¹⁷² Given the limitations of both invasive and less invasive methods, there is a
¹⁷³ clear need for a more advanced approach. An alternative, non-invasive method

¹⁷⁴ involving machine and deep learning technologies could address these issues by
¹⁷⁵ providing a fast, accurate, and effective solution without harming or stressing the
¹⁷⁶ blood cockles.

¹⁷⁷ 1.3 Research Objectives

¹⁷⁸ 1.3.1 General Objective

¹⁷⁹ The general objective of this study is to develop a non-invasive method for iden-
¹⁸⁰ tifying the sex of *Tegillarca granosa* using machine and deep learning integrated
¹⁸¹ with computer vision technologies. This method aims to provide accurate and
¹⁸² streamlined sex identification without causing harm to the specimens, thus sup-
¹⁸³ porting sustainable aquaculture practices.

¹⁸⁴ 1.3.2 Specific Objectives

¹⁸⁵ To achieve the overall general objective of developing a non-invasive sex identifi-
¹⁸⁶ cation of *T. granosa* using machine learning, deep learning, and computer vision
¹⁸⁷ technologies, the following specific objectives have been established:

- ¹⁸⁸ 1. To collect and organize a comprehensive dataset of *T. granosa* which will
¹⁸⁹ include high-quality images and relevant morphological measurements that
¹⁹⁰ will serve as the basis for the machine-learning model.
- ¹⁹¹ 2. To develop and implement machine learning models that can classify the
¹⁹² sex of *T. granosa* based on the collected linear measurements and images of
¹⁹³ different angles of the sample.
- ¹⁹⁴ 3. To evaluate the performance of the models used using performance metrics
¹⁹⁵ such as accuracy, precision, recall, and F1-score.
- ¹⁹⁶ 4. To develop a system that can identify the sex of *T. granosa* based on its
¹⁹⁷ morphological characteristics.

198 1.4 Scope and Limitations of the Research

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

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

211 The researchers work with 500 already sex-identified blood cockles taken from
212 Panay Island, specifically from Zarraga Iloilo and Ivisan Capiz. These samples,
213 equally divided between 250 males and 250 females, were obtained through in-
214 duced spawning via temperature shock and dissection. Samples subjected to data
215 collection of *T. granosa* are only limited to the spawned stage among the five go-
216 nadal stages - immature, developing, mature, spawning, and spent stages. The
217 other stages are not preferable due to indistinguishable gonads and their inability
218 to perform induced spawning. Thus, the researchers only focused on the samples
219 undergoing the spawned stage.

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

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

233 1.5 Significance of the Research

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

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

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

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

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

²⁵⁷ **Chapter 2**

²⁵⁸ **Review of Related Literature**

²⁵⁹ Aquaculture is the fastest-growing industry in animal food production and has
²⁶⁰ great potential as a sustainable solution to global food security, nutrition, and
²⁶¹ development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
²⁶² *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
²⁶³ tegrated into the livelihoods of Filipinos, not only through fish cultivation but
²⁶⁴ also through the production of other aquatic species, including mollusks, oysters,
²⁶⁵ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
²⁶⁶ clams *Tegillarca granosa*, have economic and environmental significance. It has
²⁶⁷ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
²⁶⁸ vesting and maintain the optimal ratio to preserve the population and production
²⁶⁹ of the blood cockles.

²⁷⁰ The members of the Arcidae Family, including *T. granosa* are important
²⁷¹ sources of food and livelihood. Cockle aquaculture meets rising demands, however,
²⁷² it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,
²⁷³ 2023). To solve the problem, researchers exert a considerable amount of effort,
²⁷⁴ developing a broader understanding of bivalves, including their sex-determining
²⁷⁵ mechanism, due to their notable importance in terms of diversity, environmental
²⁷⁶ benefits, and economic and market importance (Breton et al., 2017). Despite the
²⁷⁷ promising idea of identifying sex, there is limited research reported in terms of
²⁷⁸ sexual dimorphism, making it harder to distinguish through its morphological and
²⁷⁹ morphometric characteristics.

²⁸⁰ By addressing the challenges in the sex identification of *T. granosa*, it would be
²⁸¹ able to address one problem at a time. Currently, there are no recent documented
²⁸² publications that integrate machine learning and computer vision in characterizing
²⁸³ sexual dimorphism, reducing complexity, variability in sex determination, and

²⁸⁴ differentiation mechanisms in bivalves, including *T. granosa* specifically.

²⁸⁵ **2.1 Background on *Tegillarca granosa* and Their 286 Importance**

²⁸⁷ *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.
²⁸⁸ In the Philippines, it is commonly known as a Litob, a marine bivalve species from
²⁸⁹ the family Arcidae. Litob is widely distributed in the world including Southeast
²⁹⁰ Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest
²⁹¹ (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat
²⁹² as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the
²⁹³ blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood
²⁹⁴ phenotype where it serves two purposes the hemocyte carries oxygen around the
²⁹⁵ body and strengthens immune defenses. In addition, it possesses a unique ability
²⁹⁶ to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

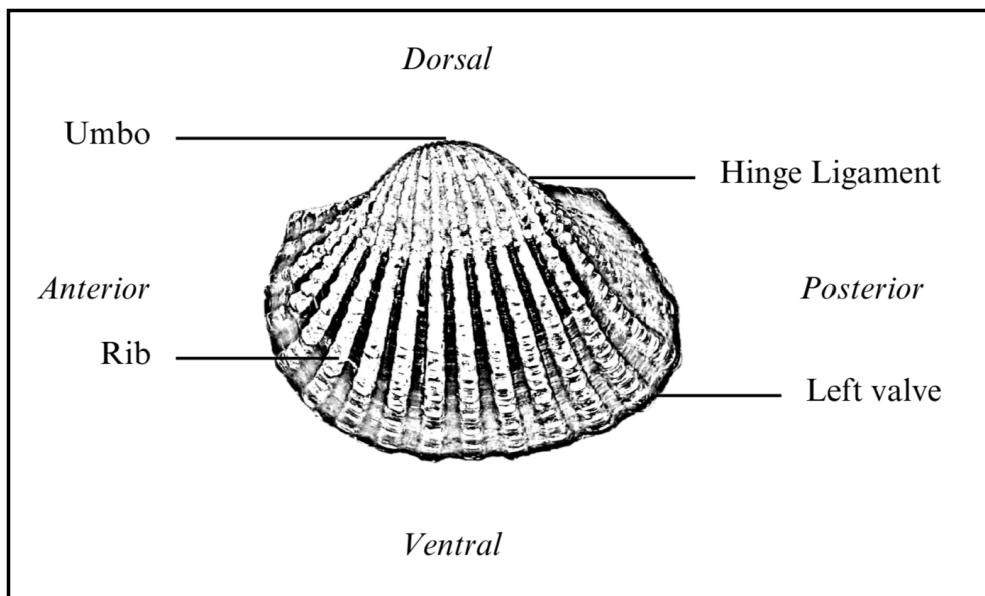


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

²⁹⁷ *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both
²⁹⁸ valves being equal in size but asymmetrical from the hinge. The top edge of
²⁹⁹ the dorsal margin is straight, while the front is rounded and slopes downward,
³⁰⁰ with its back being obliquely rounded with a concave bottom edge. It has a
³⁰¹ narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,
³⁰² 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

³⁴³ males at the time of harvest (Budd, Banh, Domingos, & Jerry, 2015).

³⁴⁴ Clearly, much more work is required to understand the mechanisms under-
³⁴⁵ lying sexual dimorphism in bivalves, specifically *T. granosa*. Just like the other
³⁴⁶ aquaculture commodities, sex affects not just reproduction but it can affect mar-
³⁴⁷ ket preference and underlying economic value, making the determination of sex
³⁴⁸ important for meeting consumer demands. These are the increasing significance
³⁴⁹ of the *T. granosa* despite the lack of reviewed articles in the Philippines.

³⁵⁰ **2.2 Current Methods of Sex Identification in *Tegillarca granosa***

³⁵¹

³⁵² The current sex identification methods in *Tegillarca granosa* range from invasive
³⁵³ histological techniques to less invasive methodologies like temperature-induced
³⁵⁴ spawning. Each approach comes with its pros and cons regarding accuracy, feasi-
³⁵⁵ bility, and impact on natural populations.

³⁵⁶ Induced spawning and larval rearing are considered the less invasive techniques
³⁵⁷ used to study *Tegillarca granosa*. In the Philippines, limited research has been
³⁵⁸ done on the *Tegillarca granosa* (Linnaeus, 1758), and this study, titled Initial At-
³⁵⁹ tempts on Spawning and Larval Rearing of the Blood Cockle, *Tegillarca granosa*
³⁶⁰ in the Philippines, is conducted by Denise Vergara Miranda and Victor Marco
³⁶¹ Emmanuel Nuestro Ferriols (2023). The researchers conducted experiments on
³⁶² induced spawning and larval rearing, discovering that the eggs of female *T. gra-*
³⁶³ *nosa* were salmon pink, while the sperm released by males looked milky. After
³⁶⁴ spawning, the researchers successfully generated 6, 531, 000 fertilized eggs.

³⁶⁵ They highlighted the importance of *T. granosa* and other anadarinids as a
³⁶⁶ food source that was established worldwide, especially in Malaysia and Korea.
³⁶⁷ However, in the Philippines, the bivalve aquaculture of the clam species is still
³⁶⁸ limited. The experiment which focuses on the culture and rearing of *T. granosa*
³⁶⁹ was attempted by subjecting the wild broodstocks to a series of temperature fluc-
³⁷⁰ tuations to induce the spawning of gametes. This is currently the most natural
³⁷¹ and least invasive method for bivalves (Aji, 2011). The study of Miranda and
³⁷² Ferriols aimed to pave the way to the sustainable production of *T. granosa* seeds
³⁷³ for aquaculture production and stock enhancement despite the scarcity of docu-
³⁷⁴ mented hatchery culture of *T. granosa* from larvae to adults that is available in
³⁷⁵ the Philippines.

³⁷⁶ In the study entitled "The earliest example of sexual dimorphism in bivalves —

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

391 On the other hand, invasive techniques such as histological analysis offer a
392 more thorough but harmful method for determining the sex of *T. granosa*. A
393 study on the Spawning Period of Blood Cockle *Tegillarca granosa* (Linnaeus,
394 1758) in Myeik Coastal. 240 blood cockle samples were examined for sex and
395 gonad maturity stages using histological examination, with shell lengths ranging
396 from 26-35mm and shell weights from 8.1-33g. For histological analysis, the whole
397 soft tissues were removed from the shell and the flesh containing most parts of
398 the gonads was fixed in formalin, dehydrated in an upgraded series of ethanol,
399 and cleared in xylene. This invasive method allows for precise identification of
400 the gonadal maturation stages based on the cellular and structural changes in the
401 gonads.

402 The classification of the gonad stages used was by Yurimoto et al. (2014).
403 There are five maturation stages of gonadal development: immature (Stage I),
404 developing (Stage II), mature (Stage III), spawning (Stage IV), and spent (Stage
405 V) stages. The sex of the *T. granosa* was confirmed by the color of the gonad and
406 by conducting a histological examination of the gonads. During the immature
407 stage, sex determination was indistinguishable due to the difficulties of observing
408 the germ cells. In the developing stage, the spermatocytes and a few spermatids
409 can be seen for males, and immature oocytes are attached to the tube wall for
410 the female. In the mature stage, the follicles are full of spermatozoa with their
411 tails pointing towards the center of the tube for the male, and the female is full
412 of mature oocytes that are irregular or polygonal in shape with the oval nucleus.
413 Upon reaching spawning, some spermatozoa are released, causing the empty space
414 in the follicle wall for males and females. There is a decrease in the number of
415 mature oocytes and it exhibits nuclear disappearance due to the breakdown of
416 the germinal vesicle. Lastly, the spent stage is where the genital tube is deformed

417 and devoid of spermatocytes which have completely spawned. In the female, the
418 genital tube is deformed and degenerated, making it empty. The morphology
419 of the cockle gonad shows that the area of the gonad increases according to the
420 increased levels of gonad maturity. The coloration of the gonad tissue layer in the
421 blood cockle varies from orange-red to pale orange in females and from white to
422 grayish-white in males for different maturity stages (May et al., 2021).

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

435 **2.3 Machine Learning and Deep Learning in Bi- 436 ological Studies**

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

447 **2.3.1 Deep Learning for Phenotype Classification in Ark
448 Shells**

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

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

462 **2.3.2 Geometric Morphometrics and Machine Learning for
463 Species Delimitation**

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

474 **2.3.3 Contour Analysis in Mollusc Shells Using Machine
475 Learning**

476 Tuset et al. (2020), in their study, *Recognising mollusc shell contours with enlarged*
477 *spines: Wavelet vs Elliptic Fourier analyses*, mentioned that gastropod shells have
478 large spines and sharp shapes that differ based on environmental, taxonomic, and

479 evolutionary influences. The researchers stated that classic morphometric meth-
480 ods may not accurately depict morphological features of the shell, especially when
481 using the angular decomposition of the contour. The current research examined
482 and compared the robustness of the contour analysis using wavelet transformed
483 and Elliptic Fourier descriptors for gastropod shells with enlarged spines. For
484 that, the researchers analyzed two geographically and ecologically separated pop-
485 ulations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed
486 that contour analysis of gastropod shells with enlarged spines can be analyzed
487 using both methodologies, but the wavelet analysis provided better local discrim-
488 ination. From an ecological perspective, shells with various sizes of spines in both
489 areas indicate the broad adaptability of the species.

490 2.3.4 Machine Learning for Shape Analysis of Marine Or- 491 ganisms

492 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*
493 to *Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*
494 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to
495 the traditional method of collecting linear measurements with the application of
496 multivariate statistical methods and outline analysis in recording the structures
497 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and
498 elasmobranchs) with their hard bodies have been used as an indication of age and
499 a determinable time-scale and structure continue to go through life (Arkhipkin,
500 2005; Kerr & Campana, 2014). This study has explored variations in the mor-
501 phometry of recording structures in stock discrimination and systematics. The
502 researchers utilized the principal component analysis rather than the traditional
503 approach, which helps simplify the data without losing important information.
504 They utilized landmark-based geometric morphometrics, which has three differ-
505 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other
506 morphogenetic processes, and lastly, the extremal points are constructed land-
507 marks.

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

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

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

538 2.3.5 Deep Learning for Landmark-Free Morphological Fea- 539 ture Extraction

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

550 2.3.6 Machine Learning for Sex Differentiation in Abalone

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

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

573 The study found that Naive Bayes consistently performed better than other al-
574 gorithms. However, when applied to both min-max and zero-mean normalization,
575 the average accuracies of the algorithms were as follows: Random Forest (62.37%),
576 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-
577 nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with
578 normalization, Random Forest achieved the highest overall metrics, including an
579 average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of
580 83.31%. Liu et al. concluded that Random Forest is highly accurate because it
581 can handle large, complex datasets, run processes in parallel using multiple trees,
582 and select the most relevant features to enhance model performance (Arifin et al.,
583 2021).

584 **2.3.7 Machine Learning for Geographical Traceability in**
585 **Bivalves**

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

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

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

612 **2.4 Limitations on Sex Identification in *Tegillarca***
613 ***granosa***

614 To date, no distinction has been made between the male and female *T. granosa*
615 in sexing methodology. In cockle aquaculture without clearly apparent sexual
616 dimorphism, sexing can be performed using invasive methods such as chemical
617 stimulation, dissection, and gonad-stripping. Induced spawning, specifically tem-

618 perature shock, is the most natural and least invasive method for bivalves (Aji,
619 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water
620 from hot to cold with a specific temperature requires deliberate and careful ma-
621 nipulation of the temperature over a specific period and would require constant
622 management and monitoring.

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

631 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous
632 (synchronous or functional) characteristics. Sequential hermaphrodites switch
633 genders after being male or female for one or multiple yearly cycles. (Heller,
634 1993; Gosling, 2004; Collin, 2013). Sex change and consecutive hermaphroditism
635 have been observed in different bivalve species, including Ostreidae, Pectinidae,
636 Veneridae, and Patellidae. However, macroscopically differentiating bivalve sex is
637 challenging. The only way it may be identified is through histological analysis of
638 gonad remains but to do so there is an act of killing the organism (Coe, 1943;
639 Gosling, 2004). Verification of sex change in bivalves to classify whether male or
640 female while they are alive is challenging since they need to be re-confirmed and
641 re-evaluated to be the same individual after a year.

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

654 No literature in the study of mollusks specifically addresses the machine learn-
655 ing algorithm used to determine the sex of *T. granosa* bivalves in various mod-
656 els. Nevertheless, various techniques such as shape analysis, morphometric ana-

657 lysis, Wavelet, and Fourier analysis, as well as different deep learning models like
658 VGNet, ResNet, and SqueezeNet in CNN networks, are utilized for phenotype
659 classification, while different machine learning algorithms could serve as the foun-
660 dation for this research project.

661 **2.5 Synthesis of the Study**

662 This section of the paper summarizes the technologies used in the different studies
663 related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex
664 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 <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

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

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

⁶⁷⁹ **Chapter 3**

⁶⁸⁰ **Research Methodology**

⁶⁸¹ This chapter discusses the materials and methods to be employed in the study, fo-
⁶⁸² cusing on the development requirements and the software, and languages utilized.
⁶⁸³ This will also entail the overall workflow in conducting the study, Morphometric-
⁶⁸⁴ Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca granosa* (Lin-
⁶⁸⁵ naeus, 1758) using machine learning technologies.

⁶⁸⁶ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,
⁶⁸⁷ will oversee the overall workflow and conduct of this experiment. The researchers
⁶⁸⁸ will also be guided by the research associates, LC Mae Gasit and Allena Esther
⁶⁸⁹ Artera. Consequently, the whole dataset collection process will be done at the
⁶⁹⁰ University of the Philippines Visayas hatchery facility.

⁶⁹¹ The methodology consists of seven parts: (1) Sample Collection, (2) Ethical
⁶⁹² Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-
⁶⁹³ tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-
⁶⁹⁴ ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine
⁶⁹⁵ Learning

⁶⁹⁶ **3.1 Sample Collection**

⁶⁹⁷ The collection of *T. granosa* samples used in this study is part of an ongoing re-
⁶⁹⁸ search project by the UPV DOST-PCAARRD titled "Establishment of the Center
⁶⁹⁹ for Mollusc Research and Development: Development of Spawning and Hatchery
⁷⁰⁰ Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."
⁷⁰¹ Furthermore, a total of 500 samples were provided in this study to classify the sex

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

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

709 As part of the data collection process, the researchers utilized induced spawn-
710 ing and dissection to classify the sex of the samples. Induced spawning through
711 temperature fluctuations is the most natural and least invasive method for bi-
712 valves compared to other methods (Aji, 2011). However, since not all samples
713 exhibited gamete release, the researchers carried out a dissection process assisted
714 by hatchery staff, to expedite data collection. The sex of the dissected samples is
715 identified based on the coloration of gonad tissue, which varies by sex and matu-
716 rity stage. Females exhibit orange-red to pale orange gonads while males display
717 white to grayish-white gonads (May et al., 2021).

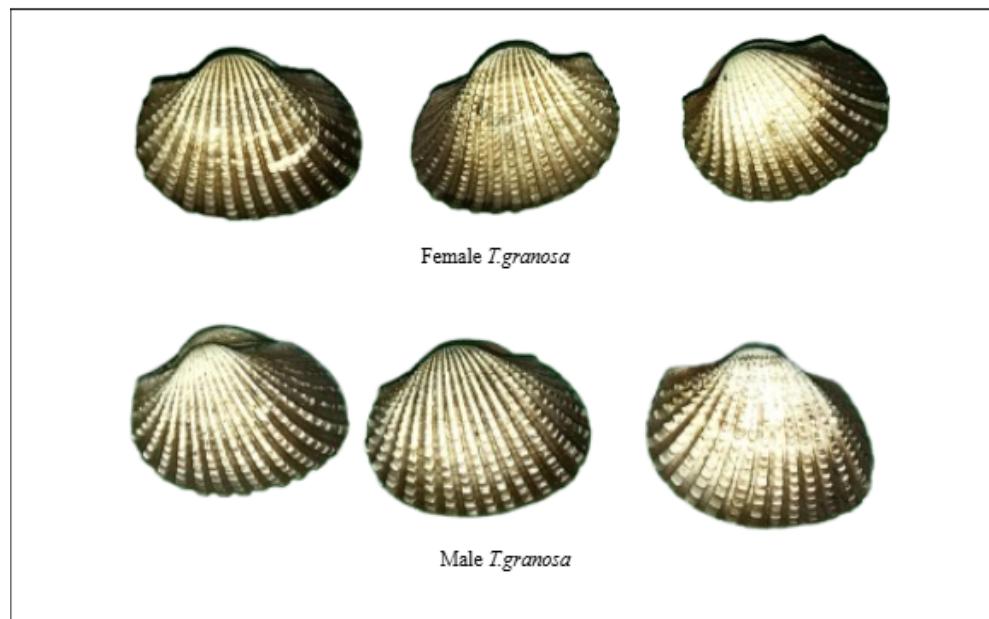


Figure 3.1: Male and Female *Tegillarca granosa* shells

718 3.2 Ethical Considerations

719 The experiment involving blood cockles will be conducted according to the Animal
720 Research: Reporting of In Vivo Experiments (ARRIVE) guidelines and will be
721 reviewed by the Institutional Animal Care and Use Committee (IACUC) of the
722 University of the Philippines Visayas.

723 3.3 Creating *T. granosa* Dataset

724 The experiment began by collecting primary observations for 100 samples of *T.*
725 *granosa*. For the actual experimentation, the researchers will collect the original
726 dataset by batch until a sample size of 500 *T. granosa* is reached. Linear mea-
727 surements were gathered by measuring the width, height, length, rib count, length
728 of the hinge line, and distance between the umbos, and these measurements were
729 organized in a CSV file. This dataset is essential for training and testing ma-
730 chine learning models and establishing the baseline for the Convolutional Neural
731 Networks.

732 The images captured for each sample were saved in JPG format using a file
733 naming convention that includes the sample sex, the orientation or view of the
734 shell, and its corresponding number out of the total 500 samples. Female *T.*
735 *granosa* samples will have file names starting with 0, while males begin with 1.
736 Each file name will include the views captured, such as (1) dorsal, (2) ventral, (3)
737 anterior, (4) posterior, (5) left lateral, and (6) right lateral (refer to Figure 3.2),
738 followed by a unique sample number. For example, “010001” will be the file name
739 for the first female sample taken from the dorsal view, and “110001” will be the
740 file name for the first male sample also taken from the dorsal view. This naming
741 convention is designed to prevent data leakage and ensure that images are correctly
742 labeled according to their respective samples. Assigning the naming convention
743 ensures smooth flow in training and testing of the gathered data points.

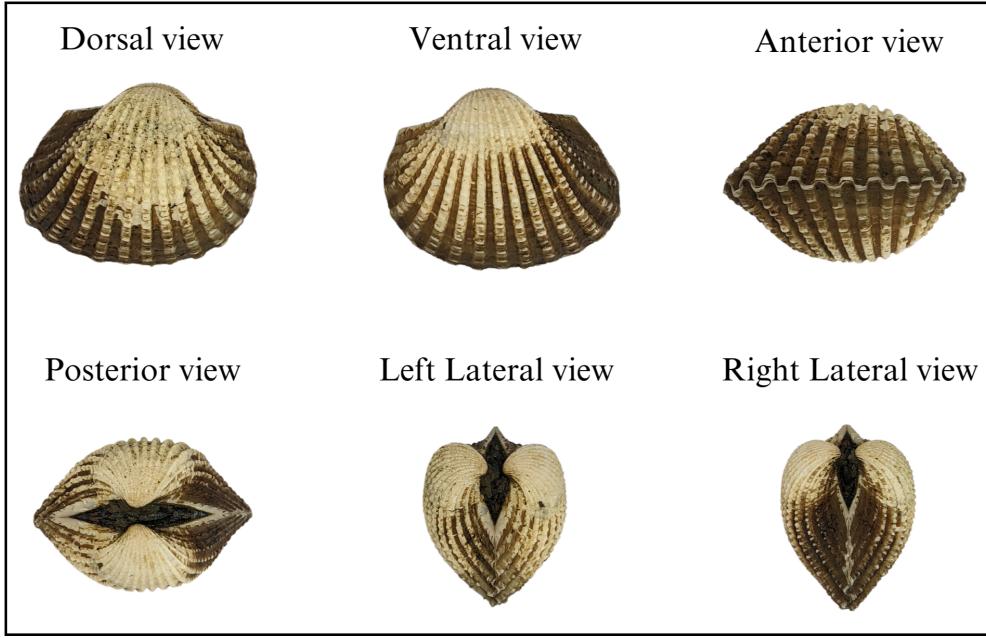


Figure 3.2: Different Views of the *T. granosa* Shell Captured

⁷⁴⁴ 3.4 Morphological and Morphometric Characteristics Collection

⁷⁴⁵

⁷⁴⁶ Morphology refers to the biological form and represents one of the most visually
⁷⁴⁷ recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Fu-
⁷⁴⁸ rusawa, 2023). In this study, morphological characteristics describe *T. granosa*
⁷⁴⁹ structural characteristics by measuring specific components, and dimensions such
⁷⁵⁰ as shapes, sizes, and colors. In terms of morphometric characteristics, this refers
⁷⁵¹ to the measurable features of *T. granosa* which are the length, width, height,
⁷⁵² length of the hinge line, the distance between the umbos, and the rib count. As
⁷⁵³ stated by the researchers, quantifying and characterizing the shape is essential to
⁷⁵⁴ understanding and visualizing the variations in *T. granosa*'s morphology.

⁷⁵⁵ In this study, the researchers measured the height, width, and length of *T.*
⁷⁵⁶ *granosa*. using a Vernier caliper to the nearest 0.01 mm. For the measurements,
⁷⁵⁷ refer to Figure 3.3. The length (A) of the *T. granosa* refers to the measurement
⁷⁵⁸ from the anterior to the posterior of the shell. The width (B) is the distance across
⁷⁵⁹ the shell's widest point from the left to the right valve. The height (C) refers to
⁷⁶⁰ the measurement from the base of the shell to the shell's apex. The length of
⁷⁶¹ the hinge line (D) near the hinge was measured, along with the distance between
⁷⁶² the umbos (E). Reymant and Kennedy (1998) indicated that incorporating rib

763 count as supplementary information increases identification accuracy. Following
764 this, the researchers recorded the rib count of the male and female *T. granosa*,
765 calculating the ratio since the sizes of the blood cockles vary.

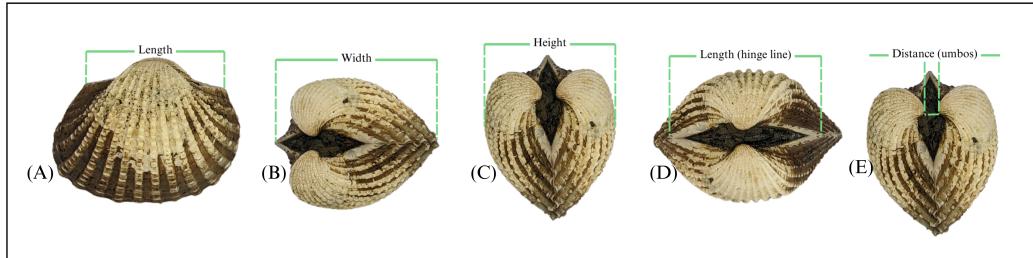


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

766 3.5 Image Acquisition and Data Gathering

767 This study will comprise 250 male and 250 female *T. granosa* images, resulting in
768 a total of 3000 images taken from different angles. The researchers constructed
769 a box-like structure with a white background to control the environment while
770 capturing images of the samples. This setup aimed to maintain uniform captures
771 of the images, fixing the camera at a consistent angle above the *T. granosa*. A
772 ring light was positioned in front of the box to ensure the image quality, eliminate
773 shadows, and ensure the clarity of the sample during the image acquisition process.
774 Google Pixel 3 XL is the smartphone used with the following specifications: 2960
775 x 1440 for the resolution, 4,032 x 3,024 pixels (12.2 MP) for the dimensions, f/1.8
776 for the fstop, 28mm (wide), $\frac{1}{2.55}$ ", 1.4 μ m, dual pixel PDAF, OIS (Concepcion et
777 al., 2023)

778 3.6 Hardware and Software Configuration

779 This section of the paper discusses the software, programming language, and nec-
780 essary tools for sex identification. Data collection, preprocessing, and model train-
781 ing were conducted on the Windows 11 operating system using an ACER Aspire
782 3 general-purpose unit (GPU) with an AMD Ryzen 3 7320U CPU with Radeon
783 Graphics (8) @ 2.395 GHz and 8 gigabytes (GB) of memory. Google Collaboratory
784 was utilized for collaborative preprocessing and data visualization. The results of
785 the gathered measurements were stored and managed in a spreadsheet. GitHub
786 was used for version control, documentation, and activity tracking throughout the



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

787 study. Python served as the primary programming language, while MATLAB was
788 used for machine learning operations and training machine learning algorithms.

789

3.7 Morphometric Characteristics Evaluation Us- 790 ing Machine Learning

791 This section of the paper discusses the machine learning operations that serve
792 as a baseline before delving into more complex deep learning methods for image
793 classification. The study variables collected included linear measurements (length,
794 width, height, length of the hinge line, distance between umbos, and rib count),
795 along with additional features such as the length-width ratio and the length-height
796 ratio as predictors. Samples were then classified by sex (female = 0, male = 1),
797 which serves as the response variable.

798

3.7.1 Preprocessing and Model Training

799 The preprocessing of the dataset involved several steps as preparation for the
800 machine learning analysis (see *Figure 3.5*). These steps included handling miss-
801 ing values, assigning labels, feature engineering, data merging and cleaning, and
802 scaling the data. Missing values present in both male and female datasets were
803 addressed by removing entries with NaN values. This approach ensured that sub-
804 sequent analyses were performed on complete data, minimizing the risk of bias or
805 errors and enhancing the reliability of the results.

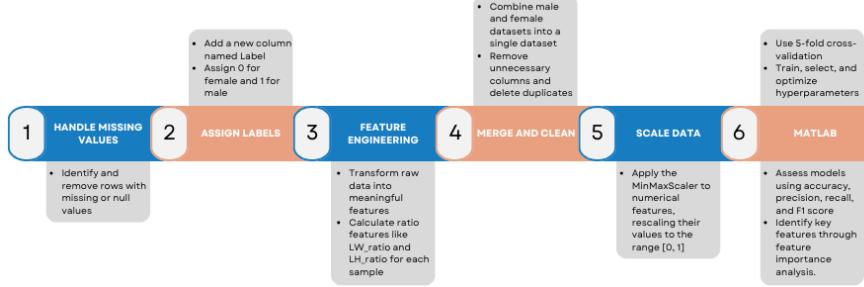


Figure 3.5: Preprocessing and Model Training Pipeline

806 Label assignment was conducted by creating a new column named "Label" in
 807 both datasets. A label of 0 was assigned to female samples, and a label of 1 was
 808 assigned to male samples, thereby establishing the target variable for the machine
 809 learning models.

810 In this study, the researchers performed feature engineering, which involves
 811 selecting, extracting, and transforming raw data into meaningful features suit-
 812 able for accurate classification. Additional ratio-based features were introduced:
 813 Length-to-Width (LW_ratio) and Length-to-Height (LH_ratio), calculated sepa-
 814 rately for both male and female datasets. These features aimed to capture size-
 815 normalized morphometric traits to improve model performance and serve as a
 816 foundation for developing a more complex model for morphological classification
 817 using deep learning technologies.

818 Subsequently, the male and female datasets were merged to create a combined
 819 dataset for further machine learning analysis and exploration. Several operations
 820 were carried out to eliminate redundancy and potential bias by removing unnec-
 821 essary columns and excluding duplicate rows to avoid bias in the analysis.

822 Lastly, numerical features were scaled using MinMaxScaler, which shifts and
 823 rescales the values to fit within a specific range, usually [0,1]. Scaling features uni-
 824 formly helps machine learning models perform well, particularly for those sensitive
 825 to the varying scales of different features.

826 Moreover, machine learning operations, including model training and selection,
 827 were conducted using MATLAB Software version 24.2.0.2712019 (R2024b). A 5-
 828 fold cross-validation technique was applied to partition the dataset and assess
 829 accuracy in each fold. In MATLAB, researchers performed data standardization
 830 and model evaluation using metrics such as accuracy, precision, recall, and F1
 831 score. Additionally, they determined the optimal hyperparameters and feature
 832 importance to identify the most relevant features for determining the sex of *T.*
granosa.

3.7.2 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:

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

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

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

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 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:

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

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:

$$F1 = \frac{precision \times recall}{precision + recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3.4)$$

⁸⁵⁵ **Chapter 4**

⁸⁵⁶ **Preliminary Results**

⁸⁵⁷ This chapter outlines the results of preprocessing, training of machine learning
⁸⁵⁸ models, and feature importance analysis. The dataset was preprocessed using
⁸⁵⁹ Python in Google Colab. After preprocessing, the dataset was imported to MAT-
⁸⁶⁰ LAB to train and evaluate the performance of various classifiers. It was followed
⁸⁶¹ by assessing the performance of different classifiers and conducting feature impor-
⁸⁶² tance analysis to identify the most significant predictors for sex identification in
⁸⁶³ *T. granosa*.

⁸⁶⁴ **4.1 Data Summary**

⁸⁶⁵ **4.1.1 Dataset Overview and Exploration**

⁸⁶⁶ The dataset contains the morphometric measurements collected from the 77 male
⁸⁶⁷ and 72 female *T. granosa* samples. Figure no. shows the proportion of male and
⁸⁶⁸ female samples, a total of 149 samples collected by the researchers and classified
⁸⁶⁹ through spawning and dissection.

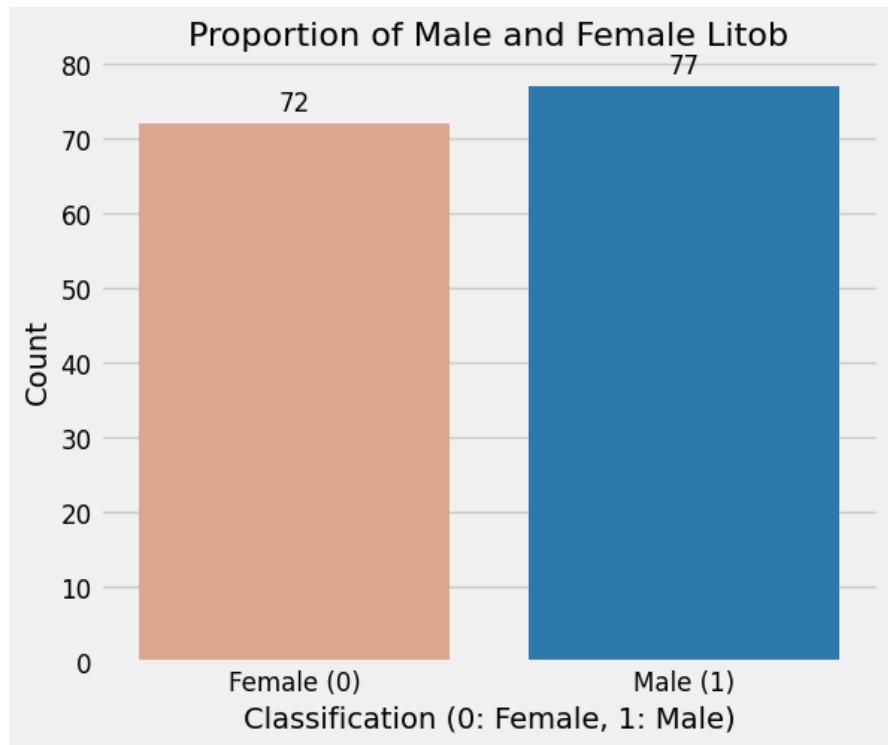


Figure 4.1: Proportion of Male and Female *T. granosa*

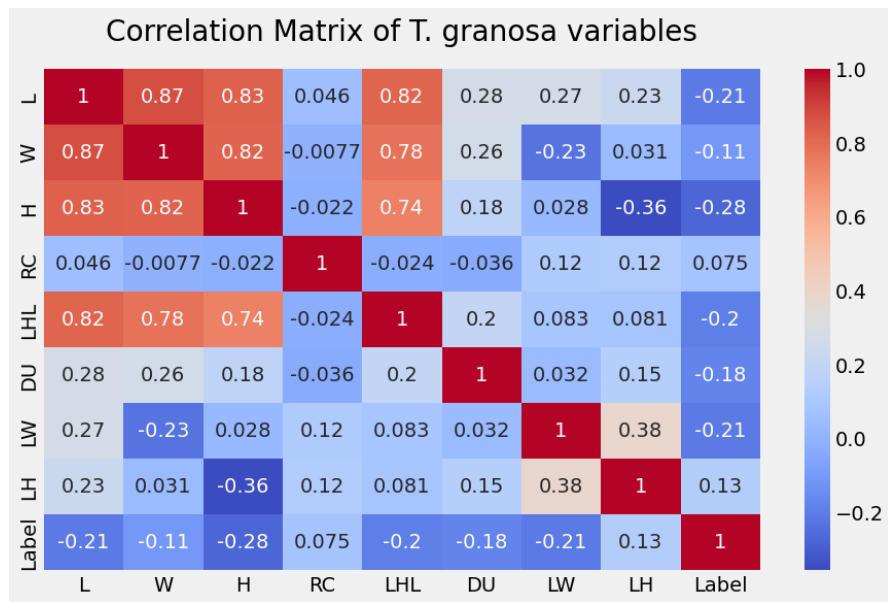


Figure 4.2: Correlation Matrix of Predictors and Target Variables

870 Figure 4.2 shows the correlation matrix between the variables and the corre-
871 lation with the target was identified and displayed in the heatmap. The positive
872 correlations observed in the matrix are the length (L) and Width (W) having (r
873 = 0.87) and height (H) (r = 0.83), width (W) and Height (H) with (r = 0.82),
874 and length (L) and hinge line length (LHL) with (r = 0.82). These features show
875 high multicollinearity since the correlation is greater than 0.8 (J. H. Kim, 2019).
876 This feature indicates that as the length of the shell increases, its width, height,
877 and hinge line length increases as well. In contrast, the rib count (RC) and dis-
878 tance of the umbos (DU) have a weak correlation from other features, with (r =
879 0.0046) and (r = 0.28) being the highest, respectively. This indicates that fea-
880 tures such as the rib count and distance of the umbos do not strongly depend on
881 the length, width, and height of the shell. The correlation analysis between the
882 predictors and the target (label, male or female) showed that most features had
883 a weak negative correlation with the label. Specifically, the highest negative cor-
884 relations were observed for the length (L) (r = -0.21), and height (H) (r = -0.28)
885 being the highest, indicating that these linear measurements only slightly differ
886 between males and females, making it challenging to distinguish morphometric
887 differences. Conversely, a weak positive correlation was found between the label
888 and the rib count (RC) and the length-to-height ratio (LH ratio), implying that
889 as these variables increase, the likelihood of classifying the sex improves.

890 Overall, the results show that while linear measurements such as length (L),
891 width (W), height (H), and hinge line length (LHL) are interdependent, the rib
892 count (RC) and distance between umbos (DU) are mostly independent features.
893 Additionally, the weak correlations between the linear measurements and the label
894 suggest that distinguishing between male and female *T. granosa* based on these
895 traits alone is difficult. To enhance predictive power, a combination of features
896 should be considered. Feature selection could be employed to identify meaningful
897 combinations of features and evaluate their performance using machine learning
898 metrics. Identifying these patterns is crucial for understanding complex biological
899 processes, as traditional correlation coefficients that capture only linear relation-
900 ships may overlook nonlinear interactions (Pividori, 2024).

901 **4.1.2 Statistical Analysis of *T. granosa* Features by Sex**

902 Figure 4.3 illustrates the mean and standard deviation (SD) of features by sex
903 in *T. granosa*. The differences in the average values of several features between
904 males and females are apparent, though the degree of variability within each
905 group, as indicated by the error bars, suggests varying levels of confidence in
906 these differences.

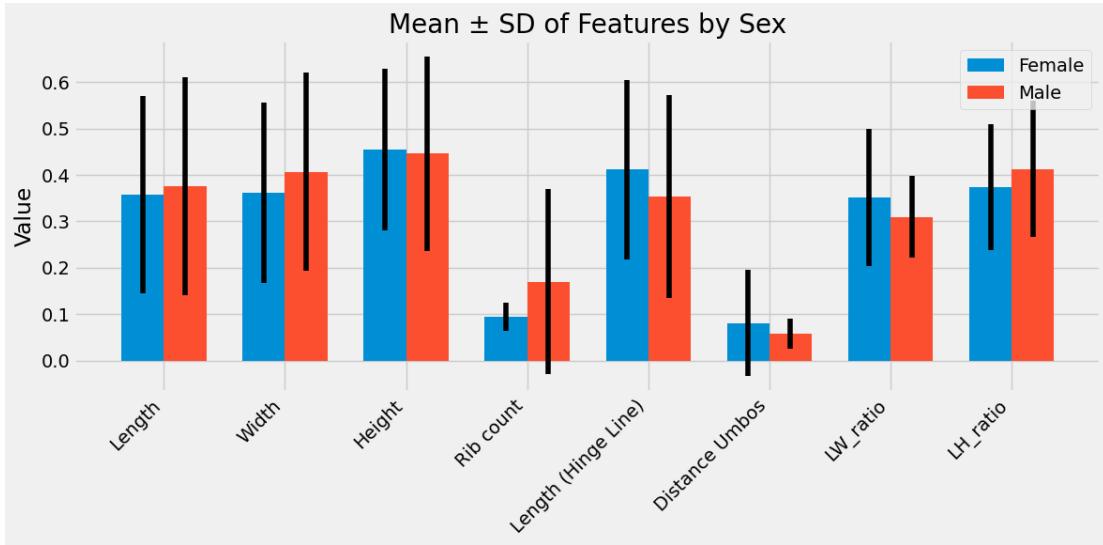


Figure 4.3: Sex-based Differences in Morphological Features of *T. granosa*

907 Males generally have larger average values for size-related features, including
 908 length, width, and height, compared to females. These differences are supported
 909 by relatively small error bars, suggesting greater confidence in the observed dis-
 910 tinctions.

911 While males show slightly higher average rib counts than females, the overlap
 912 in error bars indicates greater variability and less confidence in the difference for
 913 this feature.

914 The distance between umbos presents the most pronounced difference, with
 915 males having a significantly larger average value and minimal overlap in error
 916 bars. This strongly suggests a reliable distinction between sexes based on this
 917 feature. Similarly, hinge line length also exhibits a clear distinction, with males
 918 possessing longer average hinge lines and relatively small error bars.

919 The ratios of length to width (LW ratio) and length to height (LH ratio)
 920 show smaller differences between males and females, with overlapping error bars
 921 indicating that these features might be less reliable for sex differentiation, despite
 922 males tending to have slightly higher average values.

923 4.2 Comparison of Model Performance

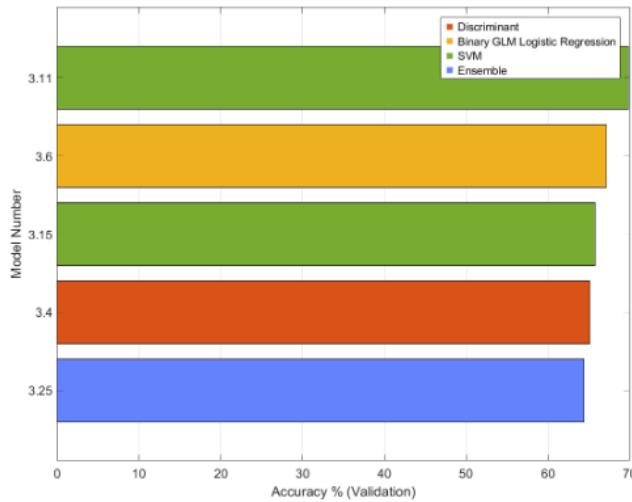


Figure 4.4: Comparison of Model Performance

924 Figure 4.4 shows the comparison of the accuracy in classifying the sex of
 925 T.granosa across different models including Discriminant, Binary GLM Logistic
 926 Regression, SVM, and Ensemble. Based on the figure above, the SVM achieved the
 927 highest accuracy percentage of 69.80%. This indicates that the SVM performed
 928 best among the models in the validation set, followed by the Binary GLM Logistic
 929 Regression and then the Discriminant. On the other hand, the Ensemble had the
 930 lowest accuracy making it the least effective model in the validation set.

931 4.2.1 Performance Evaluation

932 To evaluate the performance of the different models used, the effectiveness of each
 933 model in predicting the sex of *T. granosa* based on morphometric characteristics
 934 was assessed and compared. Performance metrics such as accuracy, precision,
 935 recall, and F1-score were utilized to evaluate the models. By analyzing these
 936 metrics, the researchers can identify the most effective model for classifying male
 937 and female *T. granosa*.

Model	Accuracy (Validation)	Weighted Precision	Weighted Recall	Weighted F1-score	Training Time (sec)
Linear SVM	69.80(%)	69.82(%)	69.80(%)	69.73(%)	2.354
Binary GLM Logistic Regression	67.11 (%)	67.16(%)	67.11(%)	66.99(%)	1.9415
Medium Gaussian SVM	65.77(%)	65.77(%)	65.77(%)	65.69 (%)	1.0323
Linear Discriminant	65.10(%)	65.22(%)	65.10(%)	64.86(%)	2.333
Subspace Discriminant	64.43(%)	64.50(%)	64.43(%)	64.23(%)	7.708

Table 4.1: Performance Metrics of Machine Learning Models for Sex Identification

938 Table 4.1 presents the comparison results of machine learning models on
939 the morphometric characteristics of the combined- male and female *T.granosa*
940 datasets. The results indicate that all models demonstrated moderate to high
941 performance in predicting males and females, with accuracies ranging between
942 64.43% to 69.80%.

943 The Linear SVM performs as the best model achieving the highest accuracy
944 (69.80%), precision (69.82%), recall (69.80%), and F1-score (69.73%), with a train-
945 ing time of 2.354s. This indicates that SVM is well-suited in identifying sex of
946 *T.granosa* based on its morphological features.

947 The Binary GLM Logistic Regression also performed well having an accuracy
948 of 67.11%, precision of 67.16%, recall of 67.11%, and F1-score of 66.99%, with a
949 training time of 1.9415s which is faster than Linear SVM.

950 The Medium Gaussian SVM, and Linear Discriminant closely followed each
951 other with accuracies of 65.77% and 65.10%, precisions of 65.77% and 65.22%,
952 recalls of 65.77% and 65.10%, and F1-scores of 65.69% and 64.86%, with a training
953 time of 1.0323s and 2.333s, respectively.

954 The Subspace discriminant, however, performed as the worst classifier with an
955 accuracy of 64.43%, precision (64.50%), recall(64.43%), and F1-score (64.23%),
956 having the longest training time of 7.708s.

957 Overall, the results seen in this comparison highlight that machine learning
958 models are effective in predicting sex identification of *T.granosa* based on their
959 morphometric characteristic with Linear SVM performing as the best model for
960 this dataset.

961 **4.2.2 Confusion Matrix Analysis**

962 Figure 4.5 displays the confusion matrix that provides a detailed breakdown of
963 classifier predictions, including true positives (correctly identified females), true
964 negatives (correctly identified males), false positives (males incorrectly classified
965 as females), and false negatives (females incorrectly classified as males).

966 The Linear SVM, being the best performing model, achieved 57 true positives
967 and 47 true negatives. However, it also had 25 false positives and 20 false nega-
968 tives. This indicates that the model did not accurately differentiate between male
969 and female, aligning with its accuracy of 69.80%. The large number of incorrectly
970 classified data points suggests while the Linear SVM is the best model compared
971 to others, it still struggles with the complexity of this dataset.

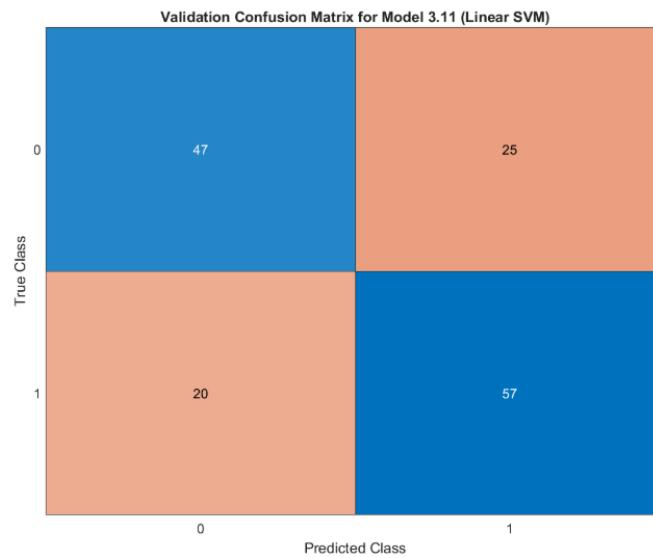


Figure 4.5: Confusion Matrix of Linear SVM

972 4.2.3 Feature Importance Analysis

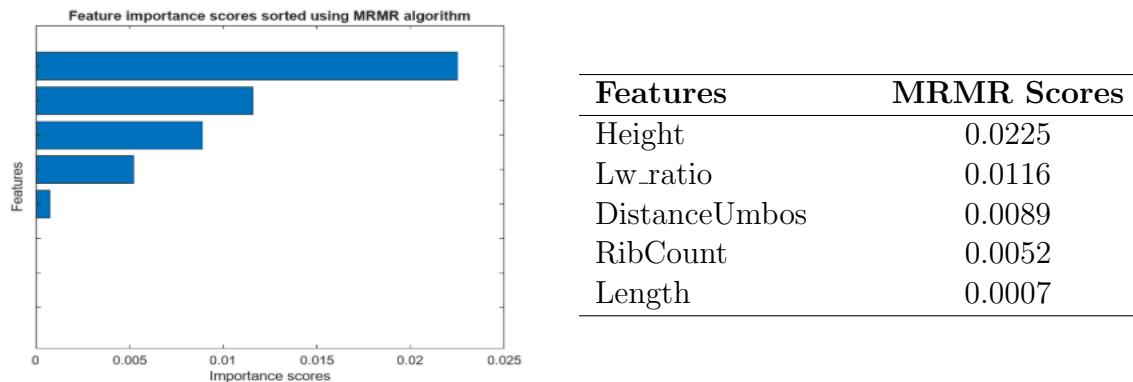


Figure 4.6: Feature Importance Scores Sorted Using the MRMR Algorithm.

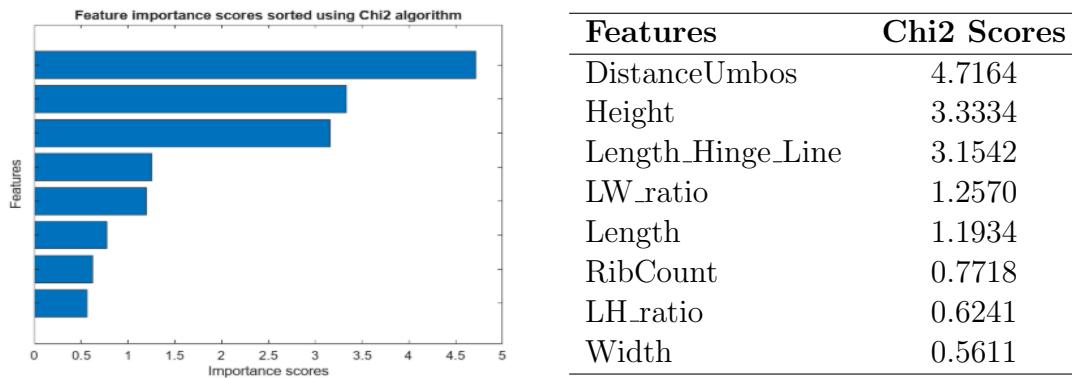


Figure 4.7: Feature Importance Scores Sorted Using the Chi2 Algorithm.

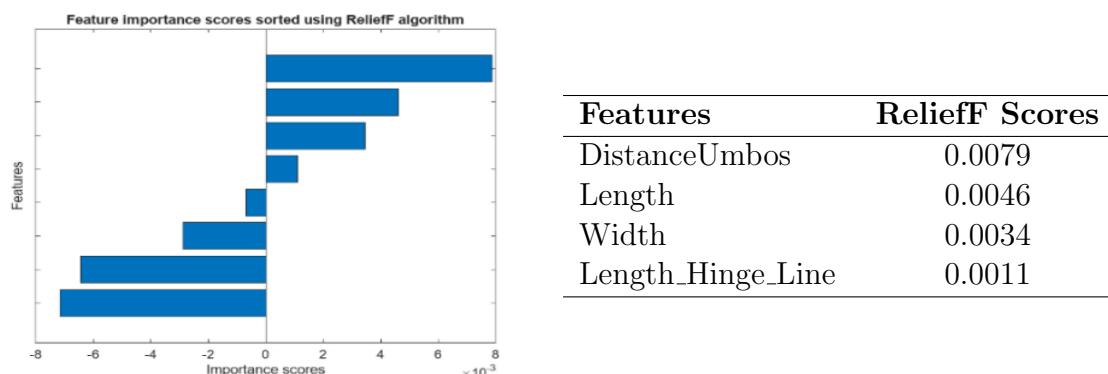


Figure 4.8: Feature Importance Scores Sorted Using the ReliefF Algorithm.

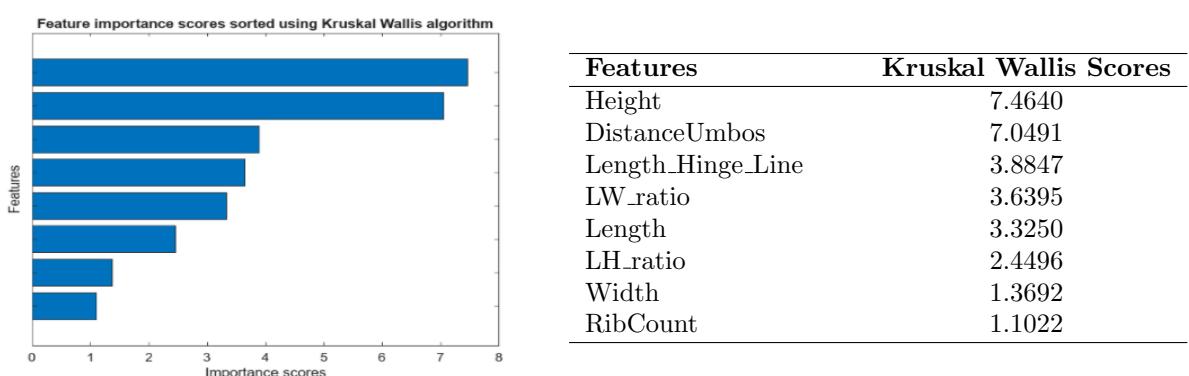


Figure 4.9: Feature Importance Scores Sorted Using the Kruskal Wallis Algorithm.

973 After processing the dataset and splitting it into training and testing sets,
974 the models are trained and their important features are computed. The features
975 are further reduced in the table removing zeros and negative results that do not
976 contribute to the scores. Feature Analysis helps in identifying which morphological
977 features contribute most in classifying male and female *T.granosa*. The study
978 employed models such as Minimum Redundancy Maximum Relevance (mRMR),
979 Chi-square (Chi2), ReliefF, Analysis of Variance (ANOVA), and Kruskal Wallis
980 feature selection algorithms.

981 The Minimum Redundancy Maximum Relevance (mRMR) identified the best
982 features as height, LW ratio, distance of the umbos, rib count, and length respec-
983 tively, which contribute most to sex classification. The Chi-square (Chi2) analysis
984 includes all eight features: however, the distance of the umbos is the most signifi-
985 cant, followed by height, length of the hinge line, LW ratio, length, rib count, LH
986 ratio, and width. In the ReliefF scores, the key features include the distance of
987 the umbos, length, width, and length of the hinge line, while height, LW ratio,
988 LH ratio, and rib count did not contribute to sex classification. Furthermore, in
989 the Kruskal-Wallis analysis, height is the most significant feature, followed by the
990 distance of the umbos, length of the hinge line, LW ratio, length, LH ratio, width,
991 and rib count as the least significant.

992 The results in figures 4.6, 4.7, 4.8, and 4.9 indicate variations in feature im-
993 portance among the four algorithm models. However, certain features, such as
994 the distance between the umbos, are present in the best features of all algorithms,
995 followed closely by height, which is identified as a key feature in three of the four
996 models, except for ReliefF. Therefore, features such as the distance between the
997 umbos and height consistently emerge as influential predictors. This analysis en-
998 abled the researchers to identify the most predictive features, which can serve as a
999 baseline for sex identification of *T. granosa* based on morphological characteristics.

1000 **Chapter 5**

1001 **Conclusion and**
1002 **Recommendations**

1003 **5.1 Conclusion**

1004 **5.2 Recommendations**

1005 This special problem entitled Morphometric-Based Non-invasive Sex Identification
1006 of *T. granosa* focuses on creating a baseline study that will serve as a foundation
1007 for further studies involving *Tegillarca granosa*, blood cockles using machine learn-
1008 ing, deep learning, and computer vision technologies in determining the sex of the
1009 samples is a salient need in aquaculture practices. Thus, the proceeding rec-
1010 ommendations are the future applications to improve and have detailed analysis
1011 such as focusing on shape analysis, exploring other state-of-the-art CNN such as
1012 ResNet, SqueezeNet, and InceptionNet, and comparing the analysis result. Fur-
1013 thermore, the main goal of conducting this is to have the ability to identify the
1014 sex of the samples by taking real-time angles by rotating from the dorsal, lateral,
1015 and ventral.

1016 Future studies could also invest in a much sturdier and more controlled envi-
1017 ronment by using a green background and positioning a webcam at a fixed angle.
1018 In addition, experiment with other image processing techniques such as scaling,
1019 rotating, and augmentation. The dataset can be utilized for further analysis us-
1020 ing deep learning and computer vision to make sense of the images gathered and
1021 discern sexual dimorphism for *T.granosa* or will serve as the basis for conducting
1022 similar studies to other bivalve species.

¹⁰²³ References

- ¹⁰²⁴ Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten
¹⁰²⁵ years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,
¹⁰²⁶ 5–16. doi: 10.1080/11250000409356545
- ¹⁰²⁷ Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara
¹⁰²⁸ granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,
¹⁰²⁹ *10*.
- ¹⁰³⁰ Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian
1031 Sains*, *14*, 14207.
- ¹⁰³² Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).
¹⁰³³ Data scaling performance on various machine learning algorithms to identify
¹⁰³⁴ abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:
¹⁰³⁵ 10.14710/jtsiskom.2021.14105
- ¹⁰³⁶ Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.
¹⁰³⁷ *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- ¹⁰³⁸ Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquacul-
¹⁰³⁹ ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-
¹⁰⁴⁰ tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on
1041 mangrove-friendly aquaculture organized by the seafdec aquaculture depart-
1042 ment, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- ¹⁰⁴³ BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
¹⁰⁴⁴ Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
¹⁰⁴⁵ Resources.
- ¹⁰⁴⁶ Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of
¹⁰⁴⁷ waste cockle shell (anadara granosa) in biodiesel production from palm olein:
¹⁰⁴⁸ Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.
¹⁰⁴⁹ doi: 10.1016/j.fuel.2011.03.002
- ¹⁰⁵⁰ Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining
1051 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127
.v1
- ¹⁰⁵² Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
¹⁰⁵³ E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A

- 1055 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1056 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/
1057 msq345
- 1058 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1059 proaches, challenges and opportunities for aquaculture. *Journal of Marine
1060 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1061 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).
1062 Mass mortalities in bivalve populations: A review of the edible cockle
1063 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1064 10.1016/j.ecss.2014.04.011
- 1065 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1066 research in bivalves: Towards the identification of molecular markers of
1067 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1068 10.1016/j.jprot.2012.04.027
- 1069 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly
1070 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1071 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1072 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1073 10.1093/icb/ict076
- 1074 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1075 Bivalvenet: A hybrid deep neural network for common cockle (cerastoderma
1076 edule) geographical traceability based on shell image analysis. *Ecological
1077 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1078 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1079 for chinese mitten crab using deep convolutional neural network. *Multi-
1080 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1081 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1082 ageing of eel, anguilla anguilla (l.)? *International Review of Hydrobiology*,
1083 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1084 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and
1085 anatomy for young learners*. Rosie Research. Retrieved from <https://rosieresearch.com/clam-dissection-anatomy/>
- 1086 Fao 2024 report: Sustainable aquatic food systems important for global food
1087 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1088 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1089 elemental composition: Complementary tools for stock discrimination of
1090 mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,
1091 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1092 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Black-

- 1097 well Science.
- 1098 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean*
1099 *Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1100 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1101 surface water characterization and heavy metal health risk on cockles in
1102 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1103 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1104 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The ear-
1105 liest example of sexual dimorphism in bivalves—evidence from the astar-
1106 tid *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,
1107 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1108 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1109 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1110 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1111 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1112 learning-based phenotype classification of three ark shells: Anadara kagoshi-
1113 mensis, tegillarca granosa, and anadara broughtonii. *Frontiers in Marine*
1114 *Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1115 Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean*
1116 *Journal of Anesthesiology*, 72(6), 558–569. Retrieved from <https://doi.org/10.4097/kja.19087> doi: 10.4097/kja.19087
- 1117 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the
1118 granulated ark, *tegillarca granosa* (linne). *The Korean Journal of Malacol-
1119 ogy*, 13, 55–64.
- 1120 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as
1121 a tool for stock identification of the southern blue whiting, *micromesistius*
1122 *australis*. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1123 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1124 et al. (2016). Identifying blue whiting (*micromesistius poutassou*) stock
1125 structure in the northeast atlantic by otolith shape analysis. *Canadian*
1126 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1127 cjfas-2015-0332
- 1128 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1129 *tegillarca granosa* (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.
1130 Arts Sci*, 4.
- 1131 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and
1132 larval rearing of the blood cockle, *tegillarca granosa* (linnaeus, 1758), in the
1133 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2
1134 .001
- 1135 Mérigot, B., Letourneau, Y., & Lecomte-Finiger, R. (2007). Characterization of
1136 local populations of the common sole *solea solea* (pisces, soleidae) in the nw
1137 mediterranean through otolith morphometrics and shape analysis. *Marine*
1138

- 1139 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1140 Narasimham, K. A. (1988). Taxonomy of the blood clams *anadara* (*tegillarca*)
1141 *granosa* (linnaeus, 1758) and *a.* (*t.*) *rhombea* (born, 1780).
- 1142 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1143 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1144 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1145 Pividori, R. M. D. M. D. H. . G. C. S., M. (2024). An efficient, not-only-linear
1146 correlation coefficient based on clustering. *Cell Systems*, 15, 854-868. doi:
1147 <https://doi.org/10.1016/j.cels.2024.08.005>
- 1148 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1149 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1150 doi: 10.1002/jmor.10439
- 1151 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1152 morphometrics and machine learning challenge currently accepted species
1153 limits of the land snail *placostylus* (pulmonata: Bothriembryontidae) on the
1154 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1155 doi: 10.1093/mollus/eyz031
- 1156 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,
1157 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1158 carbons from consumption of blood cockle and exposure to contaminated
1159 sediments and water along the klang strait, malaysia. *Marine Pollution
1160 Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1161 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1162 of seasonal variation and anthropogenic stress on blood cockle (*tegillarca*
1163 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-
1164 pogenic Stress on Blood Cockle (Tegillarca Granosa) Production Potential*,
1165 44(2), 62–82.
- 1166 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1167 chine learning and its applications to biology. *PLoS Computational Biology*,
1168 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1169 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1170 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1171 & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).
1172 College Park, MD: Maryland Sea Grant.
- 1173 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1174 approach for morphological feature extraction based on variational auto-
1175 encoder: An application to mandible shape. *Npj Systems Biology and Ap-
1176 plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1177 Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in
1178 the laboratory, malaysia.* (Handle.net) doi: 10.3366/in.3366.pdf
- 1179 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1180 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972

- 1181 .5008949
- 1182 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1183 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1184 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Impacts
1185 of four commonly used nanoparticles on the metabolism of a marine bivalve species,
1186 *tegillarca granosa*. *Chemosphere*, 296, 134079. doi:
1187 10.1016/j.chemosphere.2022.134079
- 1188 Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1189 blood clam *tegillarca granosa*. *Marine Environmental Research*, 177. Retrieved from
1190 <https://doi.org/10.1016/j.marenvres.2022.105632>

₁₁₉₁ **Appendix A**

₁₁₉₂ **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

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

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

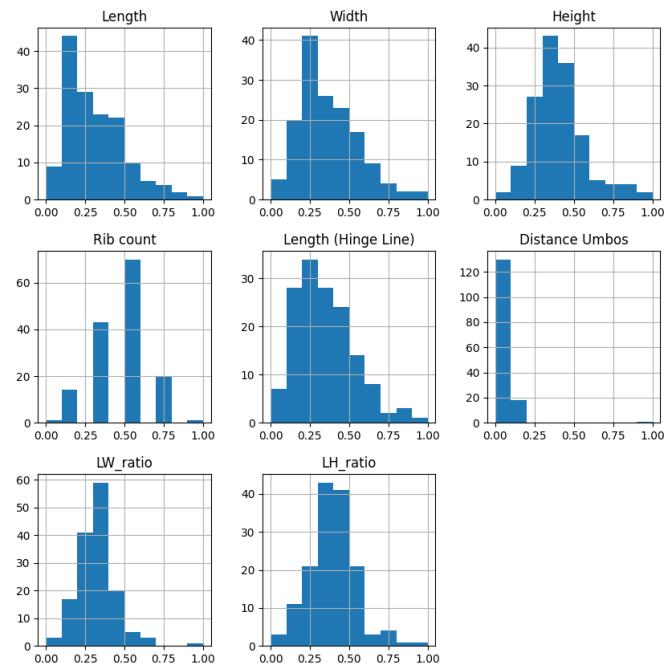


Figure A.7: Distribution of the Features of *Tegillarca granosa*