

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 April 28, 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	2.3.7	Machine Learning for Geographical Traceability in Bivalves	16
61	2.4	Limitations on Sex Identification in <i>Tegillarca granosa</i>	16
62	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	26
70	3.7	Morphometric Characteristics Evaluation Using Machine Learning	26
71	3.7.1	Data Preprocessing	27
72	3.7.2	Machine Learning Models Training	28
73	3.8	Morphological Characteristics Evaluation Using Deep Learning	29
74	3.8.1	Convolutional Neural Network	30
75	3.8.2	CNN Training	32
76	3.9	Evaluation Metrics	34
77	4	Results and Discussions	36
78	4.1	Machine Learning Analysis	36
79	4.1.1	Data Exploration	36

80	4.1.2 Statistical Analysis	37
81	4.1.3 Feature Importance Analysis	38
82	4.1.4 Performance Evaluation	39
83	5 Conclusion and Recommendations	41
84	5.1 Conclusion	41
85	5.2 Recommendations	41
86	References	42
87	A Data Gathering Documentation and Supplementary Analysis	47

⁸⁸ List of Figures

89	2.1 Diagram of <i>Tegillarca granosa</i> Anatomy	7
90	3.1 Male and Female <i>Tegillarca granosa</i> shells	22
91	3.2 Different Views of the <i>T. granosa</i> Shell Captured	24
92	3.3 Linear Measurements of <i>Tegillarca granosa</i> shell.	25
93	3.4 Image Acquisition Setup for <i>T. granosa</i> Samples	26
94	3.5 Data Preprocessing Pipeline	27
95	3.6 Diagram of k-fold cross-validation with k = 5	29
96	3.7 Shadows removed from male samples at different angles	30
97	3.8 Shadows removed from female samples at different angles	30
98	3.9 Architecture of Convolutional Neural Network (CNN)	31
99	3.10 Diagram of stratified k-fold cross-validation with k=5	32
100	3.11 Data Augmentation Techniques	33
101	4.1 Correlation heatmap of morphometric features with the sex of <i>T.</i>	
102	<i>granosa</i>	37
103	4.2 Feature Importance Scores Using the Kruskal-Wallis Test	38
104	A.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i>	47

105	A.2 Separating Male and Female Samples After Spawning of <i>Tegillarca</i>	47
106	<i>granosa</i>	
107	A.3 Sex Identified Female Through Dissecting of <i>Tegillarca granosa</i> .	48
108	A.4 Sex Identified Male Through Dissecting of <i>Tegillarca granosa</i> . .	48
109	A.5 Linear Measurements of Female <i>Tegillarca granosa</i>	48
110	A.6 Linear Measurements of Male <i>Tegillarca granosa</i>	49
111	A.7 Distribution of the Features of <i>Tegillarca granosa</i>	49

¹¹² List of Tables

¹¹³	2.1 Comparison of the Methods Used in Bivalves Studies	19
¹¹⁴	4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison	37
¹¹⁵	4.2 Performance Metrics for Models with All 13 Features	39
¹¹⁶	4.3 Performance Metrics for Models with 5 Features	39

¹¹⁷ **Chapter 1**

¹¹⁸ **Introduction**

¹¹⁹ **1.1 Overview**

¹²⁰ The Philippines is a global center of marine biodiversity and has established aqua-
¹²¹ culture as a significant contributor to total fishery production (Aypa & Baconguis,
¹²² 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-
¹²³ ally and is the 11th largest seafood producer in the world. Aquaculture is deeply
¹²⁴ integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-
¹²⁵ duction of various aquatic species, including bivalves. Among these, blood cockles
¹²⁶ (*Tegillarca granosa*) hold considerable economic and environmental significance,
¹²⁷ making it essential to ensure sustainable production and population balance.

¹²⁸ Maintaining a balanced male-to-female ratio of blood cockles is crucial to pre-
¹²⁹ vent overharvesting and ensure sustainability. An imbalanced ratio can lead to
¹³⁰ overexploitation and negatively impact the population's viability. However, there
¹³¹ is limited literature on *T. granosa* that provides a thorough understanding of its
¹³² sex-determining mechanisms, particularly regarding sexual dimorphism based on
¹³³ morphological and morphometric characteristics (Breton, Capt, Guerra, & Stew-
¹³⁴ art, 2017).

¹³⁵ Currently, sex determination methods for blood cockles are invasive, including
¹³⁶ dissection and histological examinations, which often result in the death of the
¹³⁷ species. While there is growing literature on sex identification in aquaculture
¹³⁸ commodities using machine learning and deep learning, there is a notable scarcity
¹³⁹ of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

¹⁴⁰ This study, titled "Morphometric-Based Non-Invasive Sex Identification of

¹⁴¹ Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)," aims to provide a detailed
¹⁴² baseline analysis of blood cockles by leveraging their morphological and morpho-
¹⁴³ metric characteristics. Sexual dimorphism in bivalves is often subtle and chal-
¹⁴⁴ lenging to establish mascropically (Karapunar, Werner, Fürsich, & Nützel, 2021).
¹⁴⁵ However, by integrating machine learning and deep learning, the study seeks to
¹⁴⁶ identify distinct features that may indicate sexual dimorphism between male and
¹⁴⁷ female blood cockles.

¹⁴⁸ 1.2 Problem Statement

¹⁴⁹ Identifying the sex of *T. granosa* is important for promoting sustainable aquacul-
¹⁵⁰ ture and biodiversity by maintaining a balanced male-to-female ratio. A balanced
¹⁵¹ ratio helps prevent overharvesting. Although sex identification is crucial for blood
¹⁵² cockle population management and sustainable aquaculture, there is a notable
¹⁵³ lack of research on creating non-invasive methods for determining the sex of *T.*
¹⁵⁴ *granosa*. Many recent studies and approaches rely on invasive methods like dis-
¹⁵⁵ section or histological analysis, which are impractical for large-scale aquaculture
¹⁵⁶ operations focused on conservation.

¹⁵⁷ Current methods for determining the sex of *T. granosa* are invasive and in-
¹⁵⁸ volve dissection, which requires cutting open the shell to visually inspect the
¹⁵⁹ gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-
¹⁶⁰ quently leads to their death. Another method is histological examination, where
¹⁶¹ tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,
¹⁶² 2021). Both approaches are labor-intensive and time-consuming, and can pose
¹⁶³ risks to population management, particularly when maintaining a balanced sex
¹⁶⁴ ratio for breeding programs is essential. Moreover, these invasive methods require
¹⁶⁵ specialized technical skills for accurate execution. Resource-limited aquaculture
¹⁶⁶ operations face significant challenges in accessing the necessary laboratory equip-
¹⁶⁷ ment, such as microscopes and staining tools, complicating the process.

¹⁶⁸ A less invasive approach employed by aquaculturists involves monitor spawning
¹⁶⁹ behavior, where individuals are separated and stimulated to reproduce in order
¹⁷⁰ to determine their sex through the release of gametes (Miranda & Ferriols, 2023).
¹⁷¹ Although this method is indeed less invasive than dissection, it still induces stress
¹⁷² in blood cockles and may not be completely effective for fast identification in large
¹⁷³ populations.

¹⁷⁴ Given the limitations of both invasive and less invasive methods, there is a
¹⁷⁵ clear need for a more advanced approach. An alternative, non-invasive method

¹⁷⁶ involving machine and deep learning technologies could address these issues by
¹⁷⁷ providing a fast, accurate, and effective solution without harming or stressing the
¹⁷⁸ blood cockles.

¹⁷⁹ 1.3 Research Objectives

¹⁸⁰ 1.3.1 General Objective

¹⁸¹ The general objective of this study is to develop a non-invasive method for iden-
¹⁸² tifying the sex of *Tegillarca granosa* using machine and deep learning integrated
¹⁸³ with computer vision technologies. This method aims to provide accurate and
¹⁸⁴ streamlined sex identification without causing harm to the specimens, thus sup-
¹⁸⁵ porting sustainable aquaculture practices.

¹⁸⁶ 1.3.2 Specific Objectives

¹⁸⁷ To achieve the overall general objective of developing a non-invasive sex identifi-
¹⁸⁸ cation of *T. granosa* using machine learning, deep learning, and computer vision
¹⁸⁹ technologies, the following specific objectives have been established:

- ¹⁹⁰ 1. To collect and organize a comprehensive dataset of *T. granosa* which will
¹⁹¹ include high-quality images and relevant morphological measurements that
¹⁹² will serve as the basis for the machine-learning model.
- ¹⁹³ 2. To develop and implement machine learning models that can classify the
¹⁹⁴ sex of *T. granosa* based on the collected linear measurements and images of
¹⁹⁵ different angles of the sample.
- ¹⁹⁶ 3. To evaluate the performance of the models used using performance metrics
¹⁹⁷ such as accuracy, precision, recall, and F1-score.
- ¹⁹⁸ 4. To develop a system that can identify the sex of *T. granosa* based on its
¹⁹⁹ morphological characteristics.

200 1.4 Scope and Limitations of the Research

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

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

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

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

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

235 1.5 Significance of the Research

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

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

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

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

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

²⁵⁹ **Chapter 2**

²⁶⁰ **Review of Related Literature**

²⁶¹ Aquaculture is the fastest-growing industry in animal food production and has
²⁶² great potential as a sustainable solution to global food security, nutrition, and
²⁶³ development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
²⁶⁴ *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
²⁶⁵ tegrated into the livelihoods of Filipinos, not only through fish cultivation but
²⁶⁶ also through the production of other aquatic species, including mollusks, oysters,
²⁶⁷ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
²⁶⁸ clams *Tegillarca granosa*, have economic and environmental significance. It has
²⁶⁹ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
²⁷⁰ vesting and maintain the optimal ratio to preserve the population and production
²⁷¹ of the blood cockles.

²⁷² The members of the Arcidae Family, including *T. granosa* are important
²⁷³ sources of food and livelihood. Cockle aquaculture meets rising demands, however,
²⁷⁴ it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,
²⁷⁵ 2023). To solve the problem, researchers exert a considerable amount of effort,
²⁷⁶ developing a broader understanding of bivalves, including their sex-determining
²⁷⁷ mechanism, due to their notable importance in terms of diversity, environmental
²⁷⁸ benefits, and economic and market importance (Breton et al., 2017). Despite the
²⁷⁹ promising idea of identifying sex, there is limited research reported in terms of
²⁸⁰ sexual dimorphism, making it harder to distinguish through its morphological and
²⁸¹ morphometric characteristics.

²⁸² By addressing the challenges in the sex identification of *T. granosa*, it would be
²⁸³ able to address one problem at a time. Currently, there are no recent documented
²⁸⁴ publications that integrate machine learning and computer vision in characterizing
²⁸⁵ sexual dimorphism, reducing complexity, variability in sex determination, and

²⁸⁶ differentiation mechanisms in bivalves, including *T. granosa* specifically.

²⁸⁷ **2.1 Background on *Tegillarca granosa* and Their ²⁸⁸ Importance**

²⁸⁹ *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.
²⁹⁰ In the Philippines, it is commonly known as a Litob, a marine bivalve species from
²⁹¹ the family Arcidae. Litob is widely distributed in the world including Southeast
²⁹² Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest
²⁹³ (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat
²⁹⁴ as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the
²⁹⁵ blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood
²⁹⁶ phenotype where it serves two purposes the hemocyte carries oxygen around the
²⁹⁷ body and strengthens immune defenses. In addition, it possesses a unique ability
²⁹⁸ to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

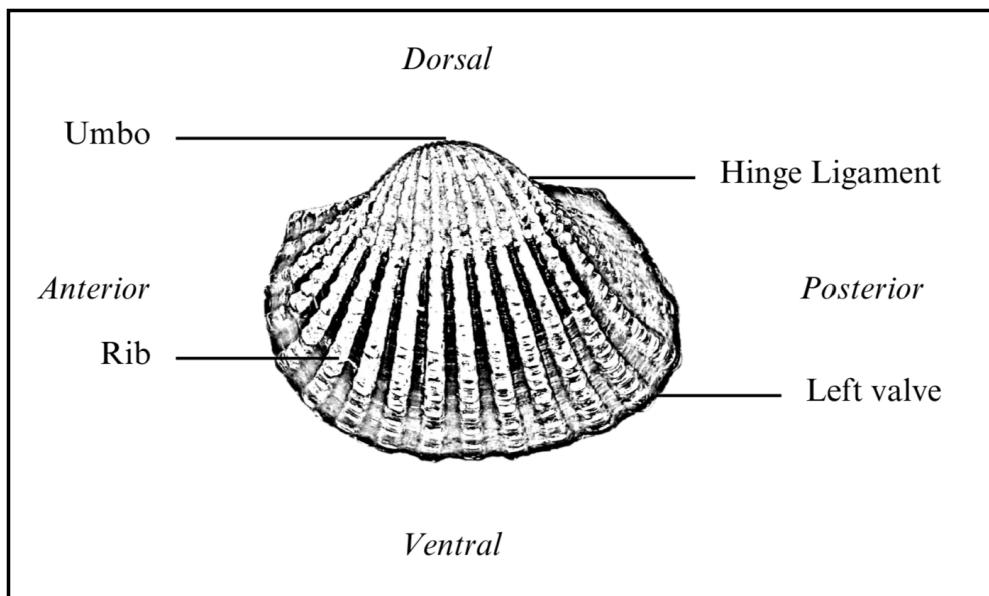


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

²⁹⁹ *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both
³⁰⁰ valves being equal in size but asymmetrical from the hinge. The top edge of
³⁰¹ the dorsal margin is straight, while the front is rounded and slopes downward,
³⁰² with its back being obliquely rounded with a concave bottom edge. It has a
³⁰³ narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,
³⁰⁴ although some may be incomplete. The shell's outer layer, or the periostracum, is

smooth and brown with a straight hinge line and 40-68 fine short teeth arranged in a straight line. The beak, or prosogyrate, curves forward, with the shell having 18–21 raised ribs with blunt nodules and spaces between them. The inner shell is white with crenulations along the valves' ventral, anterior, and posterior margins. The posterior adductor scar is elongated and squarish, while the anterior adductor scar is similar but smaller in size. The mantle covering the bulk of *T. granosa*'s visceral mass is thin but the edges are thick and muscular. It bears the impression of the crenulated shell edges. Their foot is large with a ventral grove with no byssus or thread-like attachment. The *T. granosa*'s soft body is blood red (Narasimham, 1988).

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

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

³⁴⁵ males at the time of harvest (Budd, Banh, Domingos, & Jerry, 2015).

³⁴⁶ Clearly, much more work is required to understand the mechanisms under-
³⁴⁷ lying sexual dimorphism in bivalves, specifically *T. granosa*. Just like the other
³⁴⁸ aquaculture commodities, sex affects not just reproduction but it can affect mar-
³⁴⁹ ket preference and underlying economic value, making the determination of sex
³⁵⁰ important for meeting consumer demands. These are the increasing significance
³⁵¹ of the *T. granosa* despite the lack of reviewed articles in the Philippines.

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

³⁵³

³⁵⁴ The current sex identification methods in *Tegillarca granosa* range from invasive
³⁵⁵ histological techniques to less invasive methodologies like temperature-induced
³⁵⁶ spawning. Each approach comes with its pros and cons regarding accuracy, feasi-
³⁵⁷ bility, and impact on natural populations.

³⁵⁸ Induced spawning and larval rearing are considered the less invasive techniques
³⁵⁹ used to study *Tegillarca granosa*. In the Philippines, limited research has been
³⁶⁰ done on the *Tegillarca granosa* (Linnaeus, 1758), and this study, titled Initial At-
³⁶¹ tempts on Spawning and Larval Rearing of the Blood Cockle, *Tegillarca granosa*
³⁶² in the Philippines, is conducted by Denise Vergara Miranda and Victor Marco
³⁶³ Emmanuel Nuestro Ferriols (2023). The researchers conducted experiments on
³⁶⁴ induced spawning and larval rearing, discovering that the eggs of female *T. gra-*
³⁶⁵ *nosa* were salmon pink, while the sperm released by males looked milky. After
³⁶⁶ spawning, the researchers successfully generated 6, 531, 000 fertilized eggs.

³⁶⁷ They highlighted the importance of *T. granosa* and other anadarinids as a
³⁶⁸ food source that was established worldwide, especially in Malaysia and Korea.
³⁶⁹ However, in the Philippines, the bivalve aquaculture of the clam species is still
³⁷⁰ limited. The experiment which focuses on the culture and rearing of *T. granosa*
³⁷¹ was attempted by subjecting the wild broodstocks to a series of temperature fluc-
³⁷² tuations to induce the spawning of gametes. This is currently the most natural
³⁷³ and least invasive method for bivalves (Aji, 2011). The study of Miranda and
³⁷⁴ Ferriols aimed to pave the way to the sustainable production of *T. granosa* seeds
³⁷⁵ for aquaculture production and stock enhancement despite the scarcity of docu-
³⁷⁶ mented hatchery culture of *T. granosa* from larvae to adults that is available in
³⁷⁷ the Philippines.

³⁷⁸ In the study entitled "The earliest example of sexual dimorphism in bivalves —

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

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

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

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

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

437 **2.3 Machine Learning and Deep Learning in Bi- 438 ological Studies**

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

449 **2.3.1 Deep Learning for Phenotype Classification in Ark
450 Shells**

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

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

464 **2.3.2 Geometric Morphometrics and Machine Learning for
465 Species Delimitation**

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

476 **2.3.3 Contour Analysis in Mollusc Shells Using Machine
477 Learning**

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

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

492 2.3.4 Machine Learning for Shape Analysis of Marine Or- 493 ganisms

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

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

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

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

540 **2.3.5 Deep Learning for Landmark-Free Morphological Fea- 541 ture Extraction**

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

552 2.3.6 Machine Learning for Sex Differentiation in Abalone

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

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

575 The study found that Naive Bayes consistently performed better than other al-
576 gorithms. However, when applied to both min-max and zero-mean normalization,
577 the average accuracies of the algorithms were as follows: Random Forest (62.37%),
578 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-
579 nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with
580 normalization, Random Forest achieved the highest overall metrics, including an
581 average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of
582 83.31%. Liu et al. concluded that Random Forest is highly accurate because it
583 can handle large, complex datasets, run processes in parallel using multiple trees,
584 and select the most relevant features to enhance model performance (Arifin et al.,
585 2021).

586 **2.3.7 Machine Learning for Geographical Traceability in**
587 **Bivalves**

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

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

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

614 **2.4 Limitations on Sex Identification in *Tegillarca***
615 ***granosa***

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

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

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

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

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

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

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

663 2.5 Synthesis of the Study

664 This section of the paper summarizes the technologies used in the different studies
665 related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex
666 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 Quemu 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. Tusset and E. Galimany and A. Farrés and E. Marco-Herrero and J. L. Otero-Ferrer and A. Lombarte and M. Ramón	Wavelet functions and Elliptic Fourier descriptors	Addresses the difficulty of accurately defining phenotypic diversity in gastropod shells.	Advanced contour analysis methods allow accurate differentiation of gastropod shell forms.	Cannot clarify the causes of phenotypic variation in the two populations studied.
Fedor Lishchenko and Jones, J. B.	Landmark- and outline-based Geometric Morphometric methods	To address difficulties in differentiating between stocks of marine organisms to prevent misidentification that could affect conservation and management.	Shape analysis improves taxonomic classification precision and offers close distinction between related species or organisms.	Landmark-based methods can be sensitive to landmark placement.
M. Tsutsumi and N. Saito and D. Koyabu and C. Furusawa	Morphological regulated variational AutoEncoder (Morpho-VAE)	The need for reliable, landmark-free methods, such as a modified variational autoencoder, to extract and decipher complex shapes from image data.	Employs dimension reduction and feature extraction, making it a user-friendly tool for biology non-experts.	Limited sample size in certain families presented challenges.
Barrera-Hernandez, R. and Barrera-Soto, V. and Martinez-Rodriguez, J. L. and Ríos-Alvarado, A. B. and Ortiz-Rodriguez, 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

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

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

⁶⁸¹ Chapter 3

⁶⁸² Research Methodology

⁶⁸³ This chapter discussed the materials and methods employed in the study, focusing
⁶⁸⁴ on the development requirements, as well as the software and programming
⁶⁸⁵ languages utilized. It also detailed the overall workflow in conducting the study,
⁶⁸⁶ Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca*
⁶⁸⁷ *granosa* (Linnaeus), 1758) using machine learning and deep learning technologies.

⁶⁸⁸ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,
⁶⁸⁹ oversaw the overall workflow and conduct of the experiment. The researchers were
⁶⁹⁰ also guided by research associates LC Mae Gasit and Allena Esther Artera. Con-
⁶⁹¹ sequently, the entire dataset collection process was conducted at the University of
⁶⁹² the Philippines Visayas hatchery facility.

⁶⁹³ The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical
⁶⁹⁴ Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-
⁶⁹⁵ tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-
⁶⁹⁶ ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine
⁶⁹⁷ Learning, (8) Morphological Characteristics Evaluation Using Deep Learning, and
⁶⁹⁸ (9) Evaluation Metrics

⁶⁹⁹ 3.1 Sample Collection

⁷⁰⁰ The collection of *T. granosa* samples used in this study was part of an ongoing
⁷⁰¹ research project by UPV DOST-PCAARRD titled "Establishment of the Center
⁷⁰² for Mollusc Research and Development: Development of Spawning and Hatchery
⁷⁰³ Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."

⁷⁰⁴ A total of 271 samples were provided for this study to classify the sex of *T. granosa*.
⁷⁰⁵ The samples, ranging in size from 34 to 61 mm, were sourced from the coastal area
⁷⁰⁶ of Zaraga, Iloilo, and fish markets in Ivisan, Capiz, Philippines (see Figure 3.1).

⁷⁰⁷ The research and experimentation were conducted at the University of the
⁷⁰⁸ Philippines Visayas hatchery facility in Miagao, Iloilo, where the samples were
⁷⁰⁹ maintained in 200 L fiberglass-reinforced plastic (FRP) tanks containing filtered
⁷¹⁰ seawater with 35 ppt salinity (Miranda & Ferriols, 2023).

⁷¹¹ As part of the data collection process, the researchers utilized induced spawning
⁷¹² and dissection to classify the sex of the samples. Induced spawning through
⁷¹³ temperature fluctuations was the most natural and least invasive method for bi-
⁷¹⁴ valves compared to other approaches (Aji, 2011). However, since not all samples
⁷¹⁵ exhibited gamete release, the researchers also performed dissections, assisted by
⁷¹⁶ hatchery staff, to expedite data collection. The sex of the dissected samples was
⁷¹⁷ identified based on the coloration of gonad tissue, which varies according to sex
⁷¹⁸ and maturity stage. Females exhibited orange-red to pale orange gonads, while
⁷¹⁹ males displayed white to grayish-white gonads (May et al., 2021).

⁷²⁰ The methods used for data collection were considered noninvasive, particularly
⁷²¹ given that *T. granosa* are oxygen regulators well adapted to tidal exposure and
⁷²² hypoxia (Davenport & Wong, 1986).

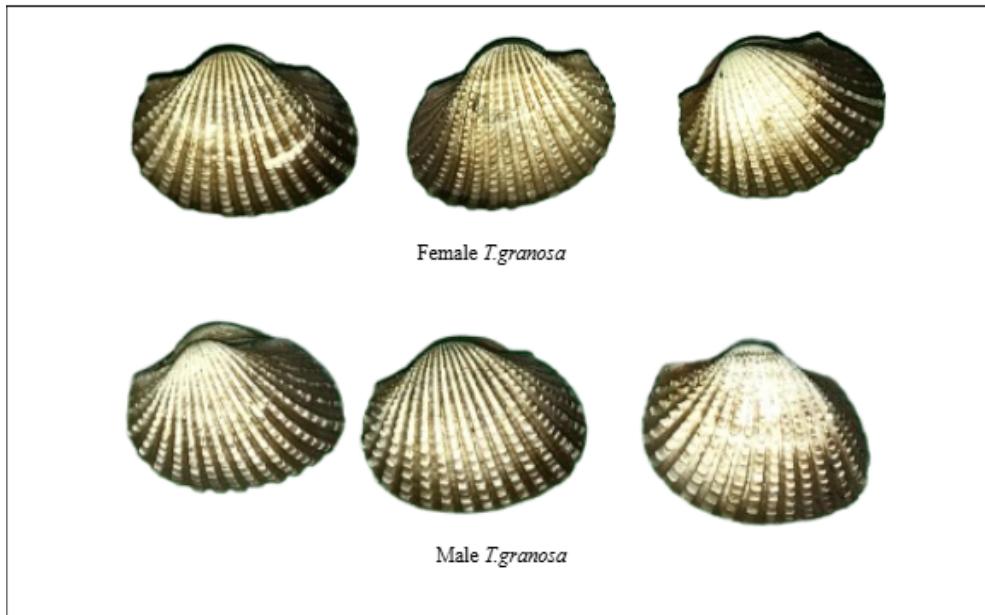


Figure 3.1: Male and Female *Tegillarca granosa* shells

723 3.2 Ethical Considerations

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

730 3.3 Creating *T. granosa* Dataset

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

739 Images of each sample were captured and saved in JPG format using a stan-
740 dardized file naming convention that included the sample's sex, the shell's ori-
741 entation or view, and its corresponding number out of the 271 total samples. File
742 names for female *T. granosa* samples began with "0", while those for male sam-
743 ples began with "1". Each file name also included one of the six captured views:
744 (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right
745 lateral (refer to Figure 3.2), followed by a unique sample number. For exam-
746 ple, "010001" denoted the first female sample taken from the dorsal view, while
747 "110001" represented the first male sample from the same view. This naming
748 convention was implemented to prevent data leakage and ensure accurate labeling
749 of images according to their respective samples.

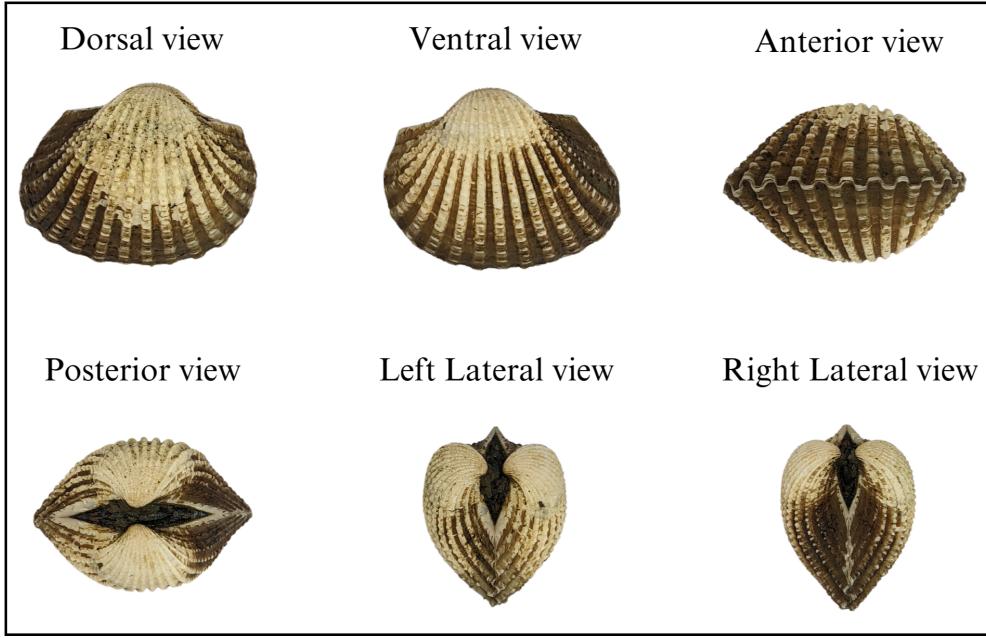


Figure 3.2: Different Views of the *T. granosa* Shell Captured

⁷⁵⁰ 3.4 Morphological and Morphometric Characteristics Collection

⁷⁵¹ Morphology refers to biological form and is one of the most visually recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).
⁷⁵² In this study, morphological characteristics describe the structural features of *T. granosa*, focusing on measurable attributes such as shape, size, and color.
⁷⁵³ Morphometric characteristics, on the other hand, refer to specific quantifiable features of *T. granosa*, including length, width, height, hinge line length, distance between the umbos, and rib count. As stated by the researchers, quantifying and characterizing these traits is essential for understanding and visualizing variations in *T. granosa* morphology.

⁷⁶¹ The researchers measured the height, width, and length of *T. granosa* using
⁷⁶² a Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the
⁷⁶³ corresponding measurement diagram. Length (A) refers to the distance from the
⁷⁶⁴ anterior to the posterior of the shell. Width (B) is defined as the widest span
⁷⁶⁵ across the shell from the left to the right valve. Height (C) measures the distance
⁷⁶⁶ from the base to the apex of the shell. In addition, the hinge line length (D) near
⁷⁶⁷ the hinge and the distance between the umbos (E) were recorded.

⁷⁶⁸ Reament and Kennedy (1998) emphasized that including rib count as supple-

mentary information can enhance identification accuracy. Following this insight, the researchers also recorded the rib count for both male and female *T. granosa*, adjusting the values by calculating ratios to account for natural size variation among specimens.

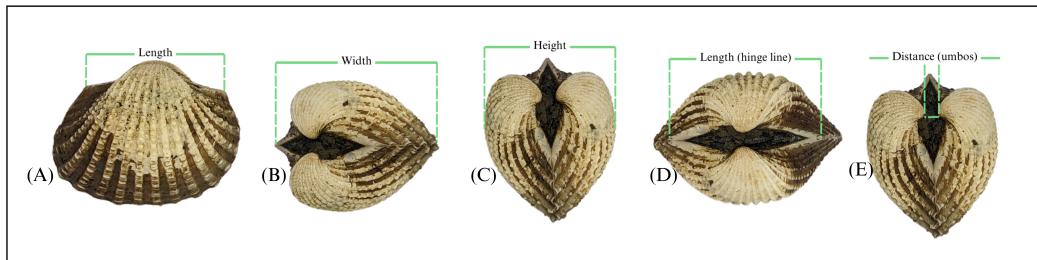


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

3.5 Image Acquisition and Data Gathering

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

The images were captured using a Google Pixel 3 XL smartphone, which features a resolution of 2960×1440 pixels and a 12.2 MP camera (4032×3024 pixels). Additional camera specifications include an f/1.8 aperture, 28mm wide lens, $\frac{1}{2.55}$ " sensor size, 1.4 μm pixel size, dual-pixel phase detection autofocus (PDAF), and optical image stabilization (OIS) (Concepcion et al., 2023).



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

787 3.6 Hardware and Software Configuration

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

799 3.7 Morphometric Characteristics Evaluation Us- 800 ing Machine Learning

801 This section of the paper discusses the machine learning operations that served
802 as a baseline prior to implementing more complex deep learning methods for
803 image classification. The study utilized collected variables including linear mea-
804 surements—length, width, height, hinge line length, distance between the um-
805 bos, and rib count—along with derived features used as predictors. These in-
806 cluded the length-to-width ratio, length-to-height ratio, width-to-height ratio,
807 umbo distance-to-length ratio, hinge line length-to-length ratio, umbo distance-

808 to-height ratio, and rib density. The samples were classified by sex, with females
809 labeled as 0 and males as 1, which served as the response variable.

810 **3.7.1 Data Preprocessing**

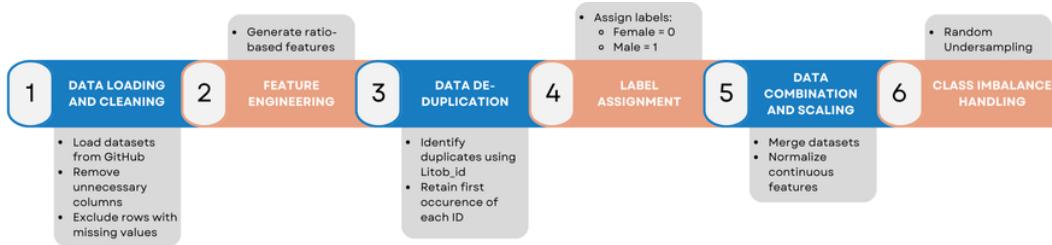


Figure 3.5: Data Preprocessing Pipeline

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

814 ***Data Loading and Cleaning***

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

819 ***Feature Engineering***

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

826 ***Data De-duplication***

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

831 ***Label Assignment***

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

835 ***Data Combination and Scaling***

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

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

845 ***Class Imbalance Handling***

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

851 **3.7.2 Machine Learning Models Training**

852 ***Model Selection and Hyperparameter Tuning***

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

858 ***Cross-Validation***

859 A five-fold cross-validation approach was implemented. The dataset was di-
860 vided into five subsets, with four used for training and one for testing. This
861 process was repeated five times, with each fold serving as the test set once. This

862 method ensured that model evaluation was robust and generalizable, minimizing
863 the bias that may result from a single train-test split. (GeeksforGeeks, 2024)

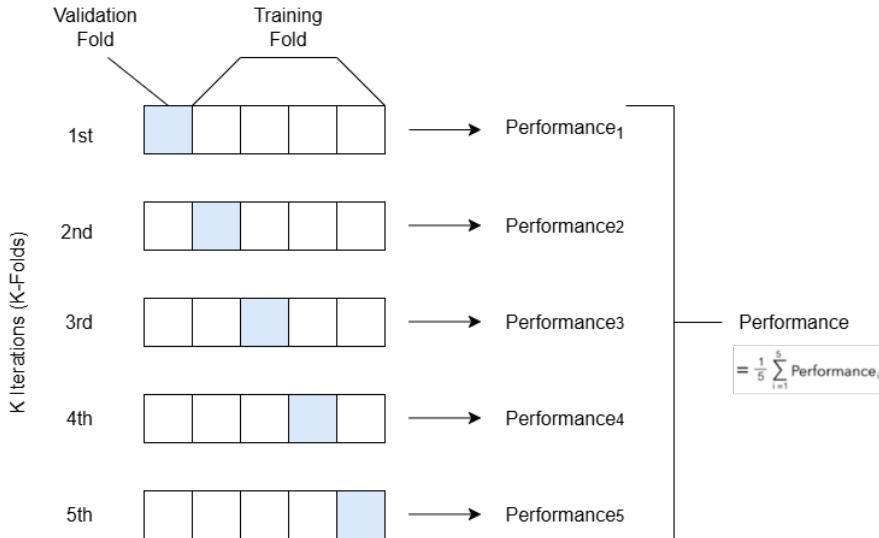


Figure 3.6: Diagram of k-fold cross-validation with $k = 5$

864 3.8 Morphological Characteristics Evaluation Us- 865 ing Deep Learning

866 This section outlines the application of deep learning techniques in analyzing the
867 morphological characteristics of *Tegillarca granosa* to identify their sex based on
868 shell images. A Convolutional Neural Network (CNN) architecture was imple-
869 mented and trained on preprocessed images using cross-validation.

870 *Image Preprocessing*

871 This subsection details the image processing techniques applied to raw shell
872 images of *T. granosa* using computer vision methods before training the deep
873 learning model. The image preprocessing techniques include standardizing input
874 dimensions and removing shadows, background, and noise. Each image under-
875 went data augmentation to enhance feature visibility for effective learning. Image
876 preprocessing ensures consistent and high-quality input data for model training.

877 *Adjusting Dimensions*

878 All images were resized to a consistent dimension of 256x256 pixels to ensure
879 uniformity throughout the dataset. This standardization is essential for Convo-

880 lutional Neural Networks (CNNs), as a consistent input dimension is required.
881 While resizing, the aspect ratio was maintained to prevent distortion of the mor-
882 phological features, and padding was added to retain the original format.

883 ***Background Removal***

884 Background removal was performed to maintain a consistent white background
885 throughout the dataset. The tool `rembg` was used to efficiently remove the original
886 background, retaining the foreground from the raw images. This method resulted
887 in clear images with a white background, enhancing focus on the morphological
888 features and defining the shell boundaries.

889 ***Shadow Removal***

890 To minimize noise caused by shadows around the shell, HSV thresholding,
891 contours, and morphological thresholds were applied to isolate and remove shad-
892 owed regions. This approach preserved the natural color of the blood cockles and
893 eliminated shadows and noise from the surrounding area.

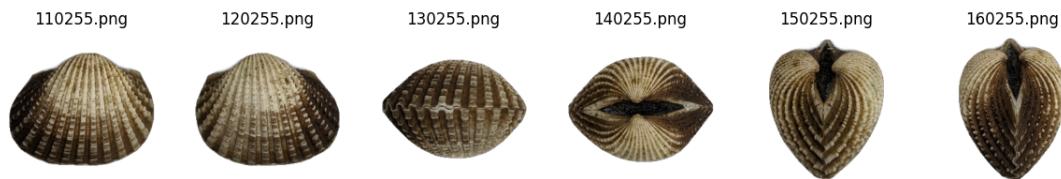


Figure 3.7: Shadows removed from male samples at different angles

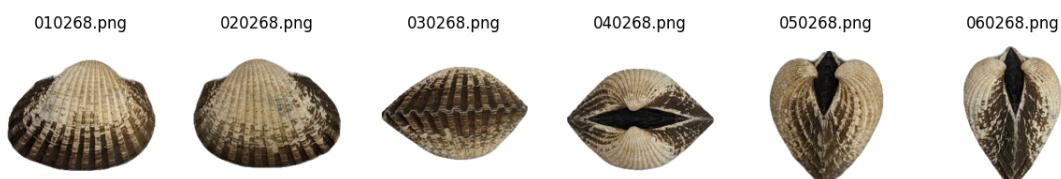


Figure 3.8: Shadows removed from female samples at different angles

894 **3.8.1 Convolutional Neural Network**

895 Convolutional Neural Networks are the main deep learning tool used in image
896 classification, specifically binary classification. CNNs leverage their ability to
897 share weights and use pooling techniques, reducing the number of parameters (Cui,
898 Pan, Chen, & Zou, 2020). The proposed CNN architecture for sex identification of
899 blood cockles employs 12 layers designed to extract features from the input image

900 with dimensions of (256, 256, 3). The layers consist of four convolution layers,
901 a flatten layer, and two dense layers. The CNN framework used in this study is
902 shown in Figure 3.9.

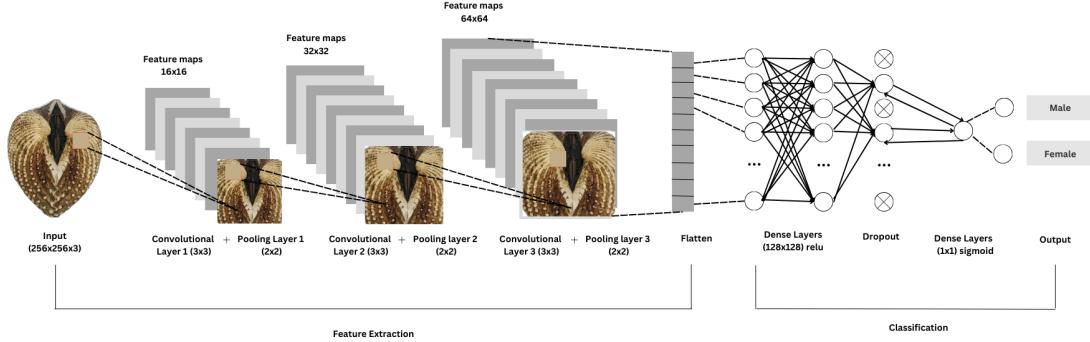


Figure 3.9: Architecture of Convolutional Neural Network (CNN)

903 ***Convolution Layer***

904 The convolution layers of CNN extract the features from the input image
905 through the convolution operation. This study uses four convolution layers with
906 a 3x3 kernel size and filter sizes of 16, 32, 64, and 128. The first layer extracts
907 the low-level features, such as edges, lines, and corners, while the deeper layers
908 iteratively extract more complex information from these low-level features. The
909 ReLU activation function was used, allowing the model to learn the complex
910 patterns within the data.

911 ***Pooling Layer***

912 A pooling layer was added after the convolution layer to enhance calculation
913 speed and prevent overfitting (Cui et al., 2020). In this study, max pooling was
914 applied with a (3,3) kernel size.

915 ***Fully Connected and Dropout***

916 Fully connected layers follow after the convolution and pooling layers. Each
917 neuron connects to all neurons of the previous layer. The output values from the
918 fully connected layers are sent to an output layer. It was classified using different
919 sigmoid functions appropriate for binary classification.

920 A large number of parameters in the training process can lead to overfitting.
921 It occurs when the model learns the training data too well, including its noise and
922 irrelevant details. This results in poor performance on unseen data. To mitigate
923 the overfitting, the dropout layer was employed. Dropout works by temporarily
924 discarding a portion of the neurons in the network with probability p ($0 < p < 1$).

925 During this process, these neurons do not participate in the forward propagation
926 process of CNN and the backward propagation process (Cui et al., 2020).

927 3.8.2 CNN Training

928 The dataset consists of 1626 samples, with 127 samples from females and 144 sam-
929 ples from males, individually for each angle. Given the minimal class imbalance,
930 random undersampling was carried out to create a balanced dataset. All images
931 were resized to 256x256 pixels and normalized using a Rescaling layer, ensuring
932 pixel values were within the range [0, 1].

933 *Data Splitting*

934 Due to the limited dataset size, a traditional train-test split was not adopted.
935 Instead, a 5-fold stratified cross-validation approach was used to maximize the
936 use of available data while preserving the class distribution within each fold.
937 **StratifiedKFold** was applied to ensure that the distribution of male and female
938 samples remained consistent across all folds, thereby enabling fair and robust
939 model evaluation (GeeksforGeeks, 2020).

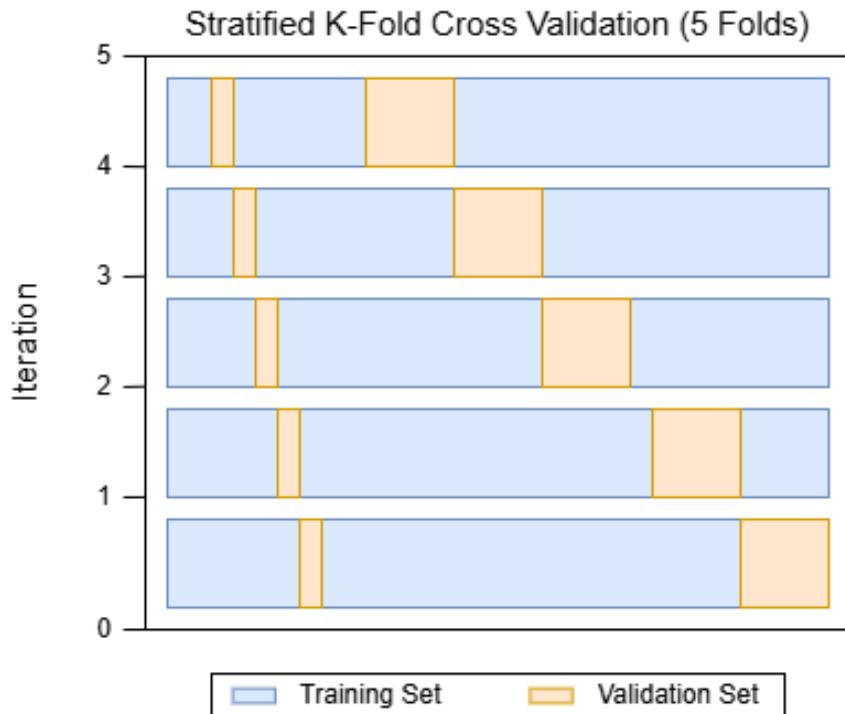


Figure 3.10: Diagram of stratified k-fold cross-validation with k=5

940 ***Data Augmentation***

941 Before model training, online data augmentation was applied exclusively to
942 the training data within each fold, creating new data variations on the fly. The
943 augmentations included random horizontal flips, slight rotations, and zoom trans-
944 formations to enhance data diversity and improve model generalization (Awan,
945 2022). All augmentation was strictly applied only to the training subset of each
946 fold to prevent data leakage and maintain the validity of the results. On-the-fly
947 data augmentation (OnDAT) generates augmented data during each iteration,
948 exposing the model to constantly changing data variations. Augmenting the origi-
949 nal data allows better exploration of the underlying data generation process and
950 has the potential to prevent the model from overfitting spurious patterns, thereby
951 improving performance (Cerqueira, Santos, Baghoussi, & Soares, 2024).

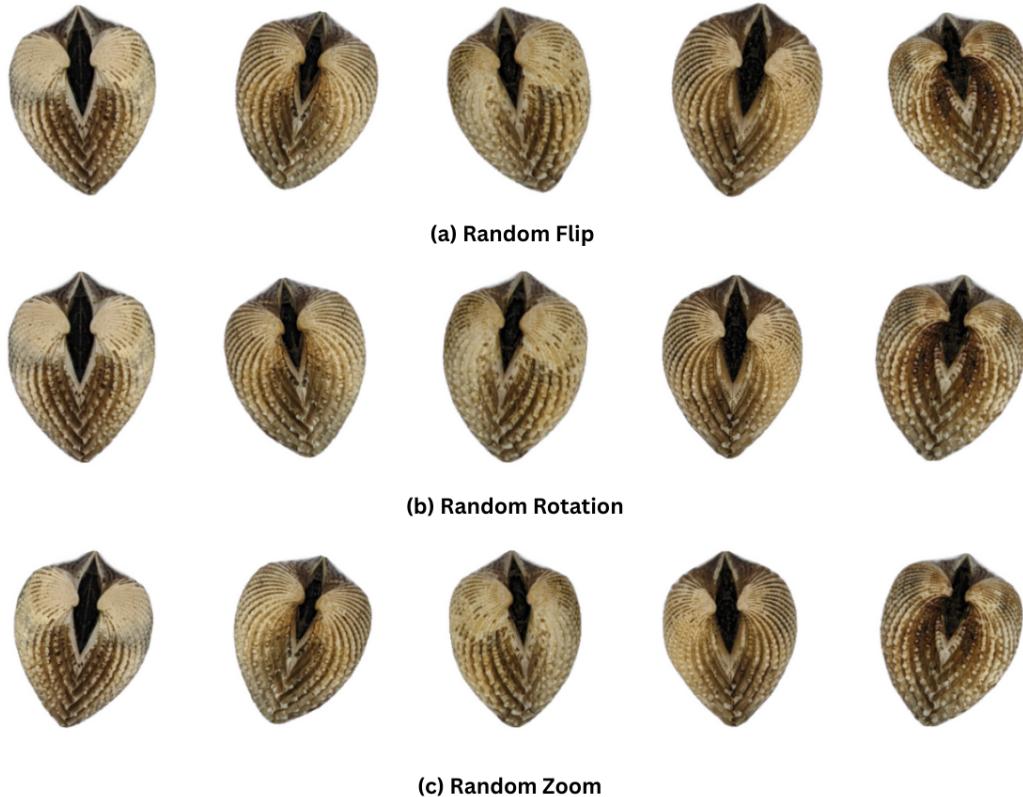


Figure 3.11: Data Augmentation Techniques

952 ***Training Procedure***

953 During the training process, model performance per fold was carefully mon-
954 itored. One important thing to observe is the consistency in the performance,
955 whether the model is still learning or is at high risk of overfitting. Early stopping

956 was applied to ensure the stable performance of the model across folds. This
957 technique allows for monitoring the training of the neural network, stopping when
958 the performance metrics, in this case, validation loss, cease to improve. Further-
959 more, to enhance the learning process, `ReduceLROnPlateau` was applied, which
960 decreased the learning rate if there was no improvement in the model for a speci-
961 fied number of epochs (Team, n.d.).

962 The model was trained using the Adam optimization algorithm, with an ini-
963 tial learning rate of 0.001. Binary cross-entropy, commonly known as the log loss,
964 was employed as the loss function due to its effectiveness in binary classifica-
965 tion tasks. To reduce the risk of overfitting, a dropout rate of 0.5 was applied, ran-
966 domly deactivating half of the neurons during the training process to improve
967 generalization.

968 3.9 Evaluation Metrics

969 Evaluating the performance of a binary classification model is essential, and se-
970 lecting appropriate metrics depends on the specific requirements of the user. The
971 performance of both supervised machine learning and deep learning models will
972 be measured using several key metrics, including accuracy, precision, recall, F1
973 score, and the AUC-ROC score.

974 Accuracy (ACC) is the ratio of the overall correctly predicted samples to the
975 total number of examples in the evaluation dataset (Cui et al., 2020). It measures
976 the overall correctness of the model in predicting both male and female blood
977 cockles. This metric provides insight into how well the model performs across all
978 classifications. The formula for accuracy is:

$$974 \text{ACC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

979 Precision (PREC) is the ratio of correctly predicted positive samples to all
980 samples assigned to the positive class (Cui et al., 2020). This metric helps in
981 evaluating the fairness of the model and prevents the misclassification of blood
982 cockles as it identifies potential inaccuracies or biases. The formula for precision
983 is:

$$983 \text{PREC} = \frac{\text{True positive samples}}{\text{Samples assigned to positive class}} = \frac{TP}{TP + FP} \quad (3.2)$$

984 Recall (REC), also known as sensitivity or the true positive rate (TPR), is the
985 ratio of correctly predicted positive cases to all the actual positive samples (Cui
986 et al., 2020). It represents the ability of the model to correctly identify positive
987 male and female samples. The formula for recall is:

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

988 The F1 score is the harmonic mean of precision and recall, which penalizes
989 extreme values of either of the two metrics (Cui et al., 2020). It is particularly
990 useful when the class distribution is imbalanced. The formula for the F1 score is:

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

991 The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is
992 a performance measurement for classification problems, particularly used in deep
993 learning in this study. The ROC curve is a plot of the true positive rate (recall)
994 against the false positive rate (1 - specificity), and the AUC score quantifies the
995 overall ability of the model to discriminate between positive and negative classes.
996 A higher AUC indicates better model performance. (Nahm, 2022)

997 **Chapter 4**

998 **Results and Discussions**

999 This chapter presents the results of the machine learning and deep learning anal-
1000 yses conducted on the preprocessed dataset. Preprocessing was performed using
1001 Python in Google Colaboratory. The chapter includes the evaluation of various
1002 machine learning classifiers, analysis of feature importance, and the application
1003 of deep learning models for image-based classification. These approaches aim to
1004 identify key predictors and assess classification performance for sex identification
1005 in *T. granosa*.

1006 **4.1 Machine Learning Analysis**

1007 **4.1.1 Data Exploration**

1008 Exploratory data analysis was performed to characterize the dataset using visu-
1009 alizations to understand the patterns and correlations within the data. A corre-
1010 lation heatmap was created to assess the relationship between the predictors and
1011 the target variable.

1012 The heatmap (see Figure 4.1) revealed three features most correlated with the
1013 sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos-length ratio (r
1014 = 0.12), and the distance between the umbos ($r = 0.12$). Each of these features
1015 demonstrated a weak positive relationship with the target variable.

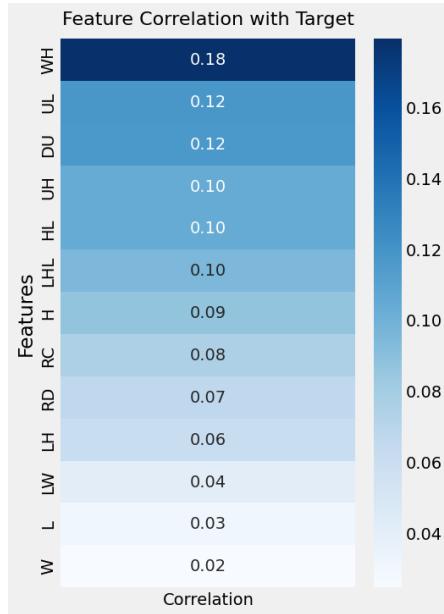


Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

¹⁰¹⁶ 4.1.2 Statistical Analysis

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

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

¹⁰¹⁷ As part of the exploratory data analysis, statistical testing confirmed that the
¹⁰¹⁸ dataset did not follow a normal distribution. Consequently, the Mann-Whitney
¹⁰¹⁹ U test was applied with a significance level of $\alpha = 0.05$ to compare male and

1020 female samples. Out of thirteen features, five showed statistically significant dif-
1021 ferences. These included: distance between umbos ($p = 0.025$), length-width ratio
1022 ($p = 0.011$), umbos-length ratio ($p = 0.019$), width-height ratio ($p = 0.003$), and
1023 umbos-height ratio ($p = 0.036$).

1024 It is important to note that statistical significance does not imply predictive
1025 importance. Therefore, further analysis, such as feature importance evaluation,
1026 was performed to identify the most informative predictors for classification.

1027 4.1.3 Feature Importance Analysis

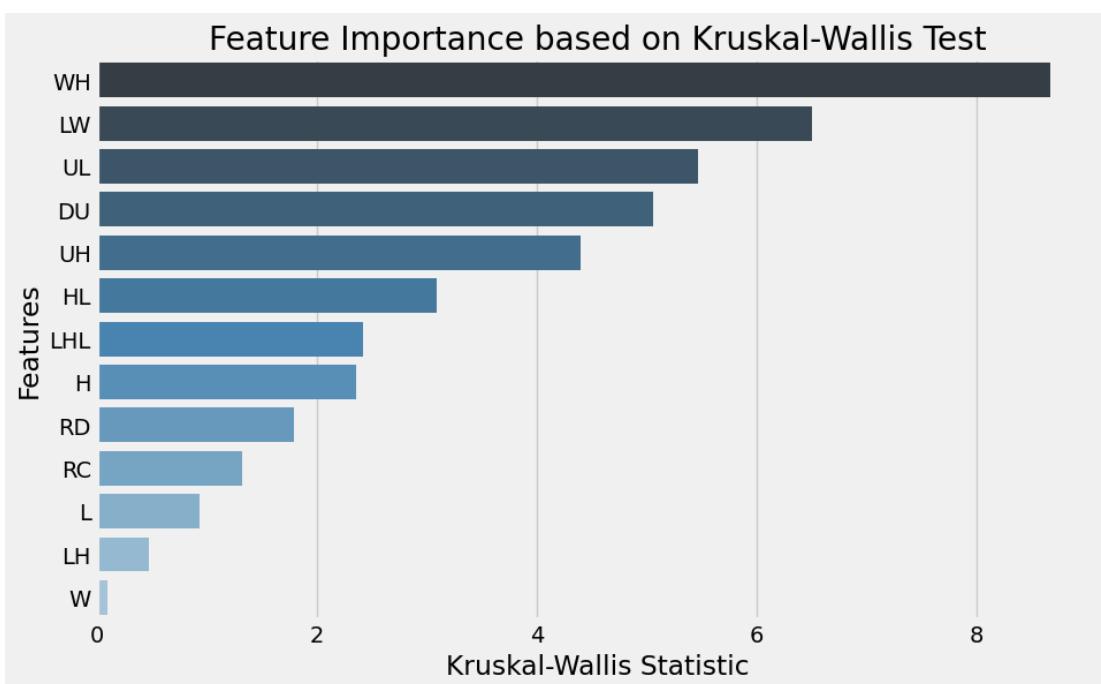


Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

1028 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric
1029 method that is suitable for evaluating differences in distributions across groups
1030 when the data does not follow a normal distribution. This approach was chosen
1031 because of the non-normality of the dataset and its robustness in handling con-
1032 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,
1033 2024)

1034 The analysis showed that the width-to-height ratio (WH_ratio) had the high-
1035 est importance score, indicating it is the most statistically significant feature for

1036 distinguishing the sex of *T. granosa*. Other notable features included the length-to-width ratio (LW_ratio), umbos-to-length ratio (UL_ratio), and the distance between the umbos, all of which contributed significantly to the classification task.
1037
1038
1039

1040 4.1.4 Performance Evaluation

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

Table 4.2: Performance Metrics for Models with All 13 Features

1041 In table 4.2, the performance of different machine learning models is presented
1042 using the full set of 13 features from the dataset. AdaBoost emerges as the
1043 highest-performing model, with an accuracy of 60.63%, precision of 60.98%, recall
1044 of 60.63%, and an F1-score of 60.39%. These results suggest that AdaBoost is
1045 particularly effective when utilizing all available features, likely due to its ability
1046 to combine multiple weak learners into a more robust model.

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

Table 4.3: Performance Metrics for Models with 5 Features

1047 Table 4.3 presents the performance of the same models using only the top 5
1048 features identified through Kruskal-Wallis feature importance analysis. The top
1049 5 features selected are distance between the umbos, length-to-width ratio, width-
1050 to-height ratio, umbos-to-height ratio and umbos-to-length ratio.

1051 Interestingly, the performance of the models improves with the reduced fea-
1052 ture set. K-Nearest Neighbors (KNN) achieves the highest performance in this

1053 scenario, with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and
1054 an F1-score of 63.75%. These results suggest that KNN benefits from using only
1055 the most significant features, showing a notable improvement over its performance
1056 when all 13 features are used.

1057 **Chapter 5**

1058 **Conclusion and
1059 Recommendations**

1060 **5.1 Conclusion**

1061 **5.2 Recommendations**

1062 This special problem entitled Morphometric-Based Non-invasive Sex Identification
1063 of *T. granosa* focuses on creating a baseline study that will serve as a foundation
1064 for further studies involving *Tegillarca granosa*, blood cockles using machine learn-
1065 ing, deep learning, and computer vision technologies in determining the sex of the
1066 samples is a salient need in aquaculture practices. Thus, the proceeding rec-
1067 ommendations are the future applications to improve and have detailed analysis
1068 such as focusing on shape analysis, exploring other state-of-the-art CNN such as
1069 ResNet, SqueezeNet, and InceptionNet, and comparing the analysis result. Fur-
1070 thermore, the main goal of conducting this is to have the ability to identify the
1071 sex of the samples by taking real-time angles by rotating from the dorsal, lateral,
1072 and ventral.

1073 Future studies could also invest in a much sturdier and more controlled envi-
1074 ronment by using a green background and positioning a webcam at a fixed angle.
1075 In addition, experiment with other image processing techniques such as scaling,
1076 rotating, and augmentation. The dataset can be utilized for further analysis us-
1077 ing deep learning and computer vision to make sense of the images gathered and
1078 discern sexual dimorphism for *T.granosa* or will serve as the basis for conducting
1079 similar studies to other bivalve species.

1080 References

- 1081 Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten
1082 years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,
1083 5–16. doi: 10.1080/11250000409356545
- 1084 Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara
1085 granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,
1086 *10*.
- 1087 Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian
1088 Sains*, *14*, 14207.
- 1089 Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).
1090 Data scaling performance on various machine learning algorithms to identify
1091 abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:
1092 10.14710/jtsiskom.2021.14105
- 1093 Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.
1094 *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- 1095 Awan, A. A. (2022, November). *A complete guide to data augmentation*. Retrieved from <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>
- 1096 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquacul-
1097 ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-
1098 tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on
1099 mangrove-friendly aquaculture organized by the seafdec aquaculture depart-
1100 ment, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- 1101 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
1102 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
1103 Resources.
- 1104 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of
1105 waste cockle shell (anadara granosa) in biodiesel production from palm olein:
1106 Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.
1107 doi: 10.1016/j.fuel.2011.03.002
- 1108 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining
1109 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127

- 1112 .v1
- 1113 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
1114 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A
1115 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1116 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/
1117 msq345
- 1118 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1119 proaches, challenges and opportunities for aquaculture. *Journal of Marine
1120 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1121 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).
1122 Mass mortalities in bivalve populations: A review of the edible cockle
1123 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1124 10.1016/j.ecss.2014.04.011
- 1125 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1126 research in bivalves: Towards the identification of molecular markers of
1127 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1128 10.1016/j.jprot.2012.04.027
- 1129 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data
1130 augmentation for forecasting with deep learning. *ArXiv (Cornell Univer-
1131 sity)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1132 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly
1133 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1134 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1135 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1136 10.1093/icb/ict076
- 1137 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1138 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma
1139 edule*) geographical traceability based on shell image analysis. *Ecological
1140 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1141 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1142 for chinese mitten crab using deep convolutional neural network. *Multi-
1143 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1144 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle
1145 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-
1146 posure. *Aquaculture*, 56(2), 151–162. Retrieved from [https://doi.org/10.1016/0044-8486\(86\)90024-4](https://doi.org/10.1016/0044-8486(86)90024-4) doi: 10.1016/0044-8486(86)90024-4
- 1147 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1148 ageing of eel, *anguilla anguilla* (l.)? *International Review of Hydrobiology*,
1149 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1150 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and
1151 anatomy for young learners*. Rosie Research. Retrieved from <https://>

- 1154 rosiereresearch.com/clam-dissection-anatomy/
- 1155 Fao 2024 report: Sustainable aquatic food systems important for global food
1156 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1159 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1160 elemental composition: Complementary tools for stock discrimination of
1161 mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,
1162 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1163 GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*. Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1166 GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn*. Retrieved from <https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/> (Accessed: 2025-04-23)
- 1169 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Blackwell Science.
- 1171 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1173 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1174 surface water characterization and heavy metal health risk on cockles in
1175 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1176 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1177 Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a comprehensive guide to data rescaling*. DataCamp. Retrieved from <https://www.datacamp.com/tutorial/normalization-in-machine-learning> (Accessed 2025-04-23)
- 1181 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest
1182 example of sexual dimorphism in bivalves—evidence from the astartid
1183 *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,
1184 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1185 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1186 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1187 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1188 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1189 learning-based phenotype classification of three ark shells: *Anadara kagoshimensis*,
1190 *tegillarca granosa*, and *anadara broughtonii*. *Frontiers in Marine Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1192 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the
1193 granulated ark, *tegillarca granosa* (linne). *The Korean Journal of Malacology*, 13, 55–64.
- 1195 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as

- 1196 a tool for stock identification of the southern blue whiting, micromesistius
1197 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1198 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1199 et al. (2016). Identifying blue whiting (micromesistius poutassou) stock
1200 structure in the northeast atlantic by otolith shape analysis. *Canadian*
1201 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1202 cjfas-2015-0332
- 1203 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1204 tegillarca granosa (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.*
1205 *Arts Sci*, 4.
- 1206 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and
1207 larval rearing of the blood cockle, tegillarca granosa (linnaeus, 1758), in the
1208 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2
1209 .001
- 1210 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of
1211 local populations of the common sole solea solea (pisces, soleidae) in the nw
1212 mediterranean through otolith morphometrics and shape analysis. *Marine*
1213 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1214 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and
1215 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–
1216 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/
1217 kja.21209
- 1218 Narasimham, K. A. (1988). Taxonomy of the blood clams anadara (tegillarca)
1219 granosa (linnaeus, 1758) and a. (t.) rhombea (born, 1780).
- 1220 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1221 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1222 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1223 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1224 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1225 doi: 10.1002/jmor.10439
- 1226 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1227 morphometrics and machine learning challenge currently accepted species
1228 limits of the land snail placostylus (pulmonata: Bothriembryontidae) on the
1229 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1230 doi: 10.1093/mollus/eyz031
- 1231 Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test*.
1232 Retrieved from https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/ (Accessed: 2025-04-23)
- 1233 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,
1234 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1235 carbons from consumption of blood cockle and exposure to contaminated
1236 sediments and water along the klang strait, malaysia. *Marine Pollution*

- 1238 *Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1239 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1240 of seasonal variation and anthropogenic stress on blood cockle (*tegillarca*
1241 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-*
1242 *pogenic Stress on Blood Cockle (*Tegillarca Granosa*) Production Potential*,
1243 44(2), 62–82.
- 1244 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1245 chine learning and its applications to biology. *PLoS Computational Biology*,
1246 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1247 Team, K. (n.d.). *Keras documentation: Reducelronplateau*. Retrieved from
1248 https://keras.io/api/callbacks/reduce_lr_on_plateau/
- 1249 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1250 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1251 & A. F. Eble (Eds.), *The eastern oyster *crassostrea virginica** (pp. 335–370).
1252 College Park, MD: Maryland Sea Grant.
- 1253 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1254 approach for morphological feature extraction based on variational auto-
1255 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*
1256 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1257 Wong, T. M., & Lim, T. G. (2018). *Cockle (*anadara granosa*) seed produced in*
1258 *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in.3366.pdf
- 1259 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1260 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972
1261 .5008949
- 1262 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1263 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1264 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-
1265 pacts of four commonly used nanoparticles on the metabolism of a ma-
1266 rine bivalve species, *tegillarca granosa*. *Chemosphere*, 296, 134079. doi:
1267 10.1016/j.chemosphere.2022.134079
- 1268 Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1269 blood clam *tegillarca granosa*. *Marine Environmental Research*, 177. Re-
1270 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

¹²⁷¹ **Appendix A**

¹²⁷² **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*

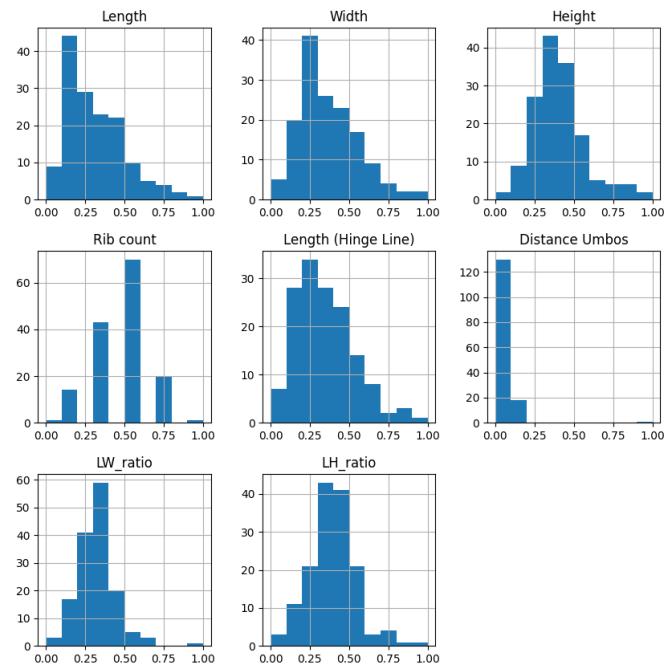


Figure A.7: Distribution of the Features of *Tegillarca granosa*