

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 | 38 |
| 81 | 4.1.3 Feature Importance Analysis | 38 |
| 82 | 4.1.4 Performance Evaluation | 39 |
| 83 | 4.1.5 Confusion Matrix Analysis | 40 |
| 84 | 5 Conclusion and Recommendations | 42 |
| 85 | 5.1 Conclusion | 42 |
| 86 | 5.2 Recommendations | 42 |
| 87 | References | 43 |
| 88 | A Data Gathering Documentation and Supplementary Analysis | 48 |

⁸⁹ **List of Figures**

| | | |
|-----|--|----|
| 90 | 2.1 Diagram of <i>Tegillarca granosa</i> Anatomy | 7 |
| 91 | 3.1 Male and Female <i>Tegillarca granosa</i> shells | 22 |
| 92 | 3.2 Different Views of the <i>T. granosa</i> Shell Captured | 24 |
| 93 | 3.3 Linear Measurements of <i>Tegillarca granosa</i> shell. | 25 |
| 94 | 3.4 Image Acquisition Setup for <i>T. granosa</i> Samples | 26 |
| 95 | 3.5 Data Preprocessing Pipeline | 27 |
| 96 | 3.6 Diagram of k-fold cross-validation with k = 5 | 29 |
| 97 | 3.7 Shadows removed from male samples at different angles | 30 |
| 98 | 3.8 Shadows removed from female samples at different angles | 30 |
| 99 | 3.9 Architecture of Convolutional Neural Network (CNN) | 31 |
| 100 | 3.10 Diagram of stratified k-fold cross-validation with k=5 | 32 |
| 101 | 3.11 Data Augmentation Techniques | 33 |
| 102 | 4.1 Correlation heatmap of morphometric features with the sex of <i>T.</i> | |
| 103 | <i>granosa</i> | 37 |
| 104 | 4.2 Feature Importance Scores Using the Kruskal-Wallis Test | 39 |
| 105 | 4.3 Feature Importance Scores Using the Kruskal-Wallis Test | 41 |
| 106 | A.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i> | 48 |

| | | |
|-----|---|----|
| 107 | A.2 Separating Male and Female Samples After Spawning of <i>Tegillarca</i> | 48 |
| 108 | <i>granosa</i> | |
| 109 | A.3 Sex Identified Female Through Dissecting of <i>Tegillarca granosa</i> . | 49 |
| 110 | A.4 Sex Identified Male Through Dissecting of <i>Tegillarca granosa</i> . . | 49 |
| 111 | A.5 Linear Measurements of Female <i>Tegillarca granosa</i> | 49 |
| 112 | A.6 Linear Measurements of Male <i>Tegillarca granosa</i> | 50 |
| 113 | A.7 Distribution of the Features of <i>Tegillarca granosa</i> | 50 |

¹¹⁴ List of Tables

| | | |
|----------------|---|----|
| ¹¹⁵ | 2.1 Comparison of the Methods Used in Bivalves Studies | 19 |
| ¹¹⁶ | 4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison | 38 |
| ¹¹⁷ | 4.2 Performance Metrics for Models with All 13 Features | 39 |
| ¹¹⁸ | 4.3 Performance Metrics for Models with 5 Features | 40 |

¹¹⁹ **Chapter 1**

¹²⁰ **Introduction**

¹²¹ **1.1 Overview**

¹²² The Philippines is a global center of marine biodiversity and has established aqua-
¹²³ culture as a significant contributor to total fishery production (Aypa & Baconguis,
¹²⁴ 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-
¹²⁵ ally and is the 11th largest seafood producer in the world. Aquaculture is deeply
¹²⁶ integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-
¹²⁷ duction of various aquatic species, including bivalves. Among these, blood cockles
¹²⁸ (*Tegillarca granosa*) hold considerable economic and environmental significance,
¹²⁹ making it essential to ensure sustainable production and population balance.

¹³⁰ Maintaining a balanced male-to-female ratio of blood cockles is crucial to pre-
¹³¹ vent overharvesting and ensure sustainability. An imbalanced ratio can lead to
¹³² overexploitation and negatively impact the population's viability. However, there
¹³³ is limited literature on *T. granosa* that provides a thorough understanding of its
¹³⁴ sex-determining mechanisms, particularly regarding sexual dimorphism based on
¹³⁵ morphological and morphometric characteristics (Breton, Capt, Guerra, & Stew-
¹³⁶ art, 2017).

¹³⁷ Currently, sex determination methods for blood cockles are invasive, including
¹³⁸ dissection and histological examinations, which often result in the death of the
¹³⁹ species. While there is growing literature on sex identification in aquaculture
¹⁴⁰ commodities using machine learning and deep learning, there is a notable scarcity
¹⁴¹ of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

¹⁴² This study, titled "Morphometric-Based Non-Invasive Sex Identification of

¹⁴³ Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)," aims to provide a detailed
¹⁴⁴ baseline analysis of blood cockles by leveraging their morphological and morpho-
¹⁴⁵ metric characteristics. Sexual dimorphism in bivalves is often subtle and chal-
¹⁴⁶ lenging to establish mascropically (Karapunar, Werner, Fürsich, & Nützel, 2021).
¹⁴⁷ However, by integrating machine learning and deep learning, the study seeks to
¹⁴⁸ identify distinct features that may indicate sexual dimorphism between male and
¹⁴⁹ female blood cockles.

¹⁵⁰ 1.2 Problem Statement

¹⁵¹ Identifying the sex of *T. granosa* is important for promoting sustainable aquacul-
¹⁵² ture and biodiversity by maintaining a balanced male-to-female ratio. A balanced
¹⁵³ ratio helps prevent overharvesting. Although sex identification is crucial for blood
¹⁵⁴ cockle population management and sustainable aquaculture, there is a notable
¹⁵⁵ lack of research on creating non-invasive methods for determining the sex of *T.*
¹⁵⁶ *granosa*. Many recent studies and approaches rely on invasive methods like dis-
¹⁵⁷ section or histological analysis, which are impractical for large-scale aquaculture
¹⁵⁸ operations focused on conservation.

¹⁵⁹ Current methods for determining the sex of *T. granosa* are invasive and in-
¹⁶⁰ volve dissection, which requires cutting open the shell to visually inspect the
¹⁶¹ gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-
¹⁶² quently leads to their death. Another method is histological examination, where
¹⁶³ tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,
¹⁶⁴ 2021). Both approaches are labor-intensive and time-consuming, and can pose
¹⁶⁵ risks to population management, particularly when maintaining a balanced sex
¹⁶⁶ ratio for breeding programs is essential. Moreover, these invasive methods require
¹⁶⁷ specialized technical skills for accurate execution. Resource-limited aquaculture
¹⁶⁸ operations face significant challenges in accessing the necessary laboratory equip-
¹⁶⁹ ment, such as microscopes and staining tools, complicating the process.

¹⁷⁰ A less invasive approach employed by aquaculturists involves monitor spawning
¹⁷¹ behavior, where individuals are separated and stimulated to reproduce in order
¹⁷² to determine their sex through the release of gametes (Miranda & Ferriols, 2023).
¹⁷³ Although this method is indeed less invasive than dissection, it still induces stress
¹⁷⁴ in blood cockles and may not be completely effective for fast identification in large
¹⁷⁵ populations.

¹⁷⁶ Given the limitations of both invasive and less invasive methods, there is a
¹⁷⁷ clear need for a more advanced approach. An alternative, non-invasive method

¹⁷⁸ involving machine and deep learning technologies could address these issues by
¹⁷⁹ providing a fast, accurate, and effective solution without harming or stressing the
¹⁸⁰ blood cockles.

¹⁸¹ 1.3 Research Objectives

¹⁸² 1.3.1 General Objective

¹⁸³ The general objective of this study is to develop a non-invasive method for iden-
¹⁸⁴ tifying the sex of *Tegillarca granosa* using machine and deep learning integrated
¹⁸⁵ with computer vision technologies. This method aims to provide accurate and
¹⁸⁶ streamlined sex identification without causing harm to the specimens, thus sup-
¹⁸⁷ porting sustainable aquaculture practices.

¹⁸⁸ 1.3.2 Specific Objectives

¹⁸⁹ To achieve the overall general objective of developing a non-invasive sex identifi-
¹⁹⁰ cation of *T. granosa* using machine learning, deep learning, and computer vision
¹⁹¹ technologies, the following specific objectives have been established:

- ¹⁹² 1. To collect and organize a comprehensive dataset of *T. granosa* which will
¹⁹³ include high-quality images and relevant morphological measurements that
¹⁹⁴ will serve as the basis for the machine-learning model.
- ¹⁹⁵ 2. To develop and implement machine learning models that can classify the
¹⁹⁶ sex of *T. granosa* based on the collected linear measurements and images of
¹⁹⁷ different angles of the sample.
- ¹⁹⁸ 3. To evaluate the performance of the models used using performance metrics
¹⁹⁹ such as accuracy, precision, recall, and F1-score.
- ²⁰⁰ 4. To develop a system that can identify the sex of *T. granosa* based on its
²⁰¹ morphological characteristics.

202 1.4 Scope and Limitations of the Research

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

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

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

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

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

237 1.5 Significance of the Research

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

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

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

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

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

²⁶¹ **Chapter 2**

²⁶² **Review of Related Literature**

²⁶³ Aquaculture is the fastest-growing industry in animal food production and has
²⁶⁴ great potential as a sustainable solution to global food security, nutrition, and
²⁶⁵ development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
²⁶⁶ *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
²⁶⁷ tegrated into the livelihoods of Filipinos, not only through fish cultivation but
²⁶⁸ also through the production of other aquatic species, including mollusks, oysters,
²⁶⁹ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
²⁷⁰ clams *Tegillarca granosa*, have economic and environmental significance. It has
²⁷¹ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
²⁷² vesting and maintain the optimal ratio to preserve the population and production
²⁷³ of the blood cockles.

²⁷⁴ The members of the Arcidae Family, including *T. granosa* are important
²⁷⁵ sources of food and livelihood. Cockle aquaculture meets rising demands, however,
²⁷⁶ it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,
²⁷⁷ 2023). To solve the problem, researchers exert a considerable amount of effort,
²⁷⁸ developing a broader understanding of bivalves, including their sex-determining
²⁷⁹ mechanism, due to their notable importance in terms of diversity, environmental
²⁸⁰ benefits, and economic and market importance (Breton et al., 2017). Despite the
²⁸¹ promising idea of identifying sex, there is limited research reported in terms of
²⁸² sexual dimorphism, making it harder to distinguish through its morphological and
²⁸³ morphometric characteristics.

²⁸⁴ By addressing the challenges in the sex identification of *T. granosa*, it would be
²⁸⁵ able to address one problem at a time. Currently, there are no recent documented
²⁸⁶ publications that integrate machine learning and computer vision in characterizing
²⁸⁷ sexual dimorphism, reducing complexity, variability in sex determination, and

²⁸⁸ differentiation mechanisms in bivalves, including *T. granosa* specifically.

²⁸⁹ 2.1 Background on *Tegillarca granosa* and Their ²⁹⁰ Importance

²⁹¹ *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.
²⁹² In the Philippines, it is commonly known as a Litob, a marine bivalve species from
²⁹³ the family Arcidae. Litob is widely distributed in the world including Southeast
²⁹⁴ Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest
²⁹⁵ (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat
²⁹⁶ as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the
²⁹⁷ blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood
²⁹⁸ phenotype where it serves two purposes the hemocyte carries oxygen around the
²⁹⁹ body and strengthens immune defenses. In addition, it possesses a unique ability
³⁰⁰ to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

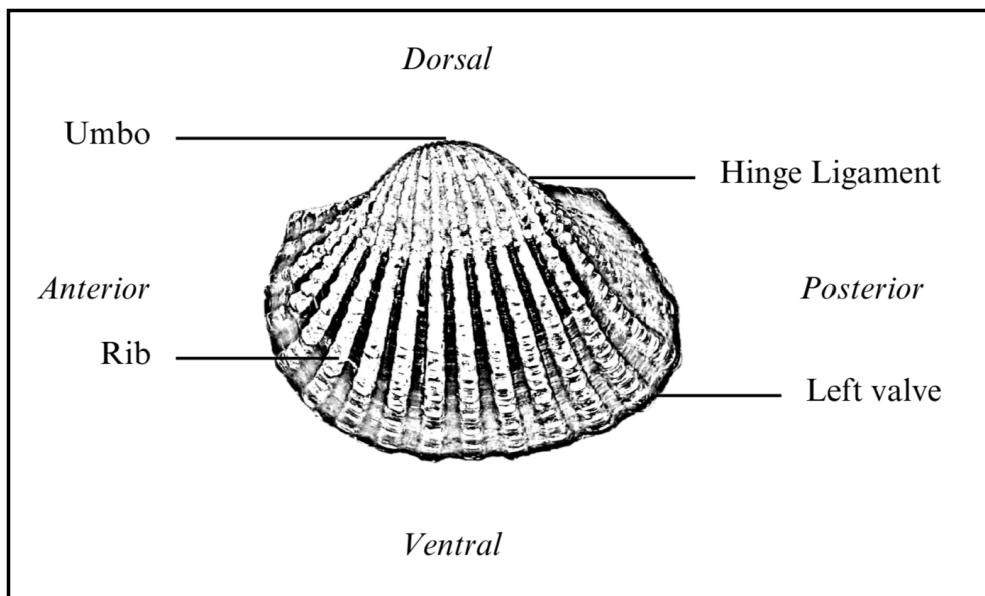


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

³⁰¹ *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both
³⁰² valves being equal in size but asymmetrical from the hinge. The top edge of
³⁰³ the dorsal margin is straight, while the front is rounded and slopes downward,
³⁰⁴ with its back being obliquely rounded with a concave bottom edge. It has a
³⁰⁵ narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,
³⁰⁶ 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

³⁴⁷ males at the time of harvest (Budd, Banh, Domingos, & Jerry, 2015).

³⁴⁸ Clearly, much more work is required to understand the mechanisms under-
³⁴⁹ lying sexual dimorphism in bivalves, specifically *T. granosa*. Just like the other
³⁵⁰ aquaculture commodities, sex affects not just reproduction but it can affect mar-
³⁵¹ ket preference and underlying economic value, making the determination of sex
³⁵² important for meeting consumer demands. These are the increasing significance
³⁵³ of the *T. granosa* despite the lack of reviewed articles in the Philippines.

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

³⁵⁵

³⁵⁶ The current sex identification methods in *Tegillarca granosa* range from invasive
³⁵⁷ histological techniques to less invasive methodologies like temperature-induced
³⁵⁸ spawning. Each approach comes with its pros and cons regarding accuracy, feasi-
³⁵⁹ bility, and impact on natural populations.

³⁶⁰ Induced spawning and larval rearing are considered the less invasive techniques
³⁶¹ used to study *Tegillarca granosa*. In the Philippines, limited research has been
³⁶² done on the *Tegillarca granosa* (Linnaeus, 1758), and this study, titled Initial At-
³⁶³ tempts on Spawning and Larval Rearing of the Blood Cockle, *Tegillarca granosa*
³⁶⁴ in the Philippines, is conducted by Denise Vergara Miranda and Victor Marco
³⁶⁵ Emmanuel Nuestro Ferriols (2023). The researchers conducted experiments on
³⁶⁶ induced spawning and larval rearing, discovering that the eggs of female *T. gra-*
³⁶⁷ *nosa* were salmon pink, while the sperm released by males looked milky. After
³⁶⁸ spawning, the researchers successfully generated 6, 531, 000 fertilized eggs.

³⁶⁹ They highlighted the importance of *T. granosa* and other anadarinids as a
³⁷⁰ food source that was established worldwide, especially in Malaysia and Korea.
³⁷¹ However, in the Philippines, the bivalve aquaculture of the clam species is still
³⁷² limited. The experiment which focuses on the culture and rearing of *T. granosa*
³⁷³ was attempted by subjecting the wild broodstocks to a series of temperature fluc-
³⁷⁴ tuations to induce the spawning of gametes. This is currently the most natural
³⁷⁵ and least invasive method for bivalves (Aji, 2011). The study of Miranda and
³⁷⁶ Ferriols aimed to pave the way to the sustainable production of *T. granosa* seeds
³⁷⁷ for aquaculture production and stock enhancement despite the scarcity of docu-
³⁷⁸ mented hatchery culture of *T. granosa* from larvae to adults that is available in
³⁷⁹ the Philippines.

³⁸⁰ In the study entitled "The earliest example of sexual dimorphism in bivalves —

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

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

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

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

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

439 **2.3 Machine Learning and Deep Learning in Bi- 440 ological Studies**

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

451 **2.3.1 Deep Learning for Phenotype Classification in Ark**
452 **Shells**

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

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

466 **2.3.2 Geometric Morphometrics and Machine Learning for**
467 **Species Delimitation**

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

478 **2.3.3 Contour Analysis in Mollusc Shells Using Machine**
479 **Learning**

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

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

494 2.3.4 Machine Learning for Shape Analysis of Marine Or- 495 ganisms

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

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

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

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

542 2.3.5 Deep Learning for Landmark-Free Morphological Fea- 543 ture Extraction

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

554 2.3.6 Machine Learning for Sex Differentiation in Abalone

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

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

577 The study found that Naive Bayes consistently performed better than other al-
578 gorithms. However, when applied to both min-max and zero-mean normalization,
579 the average accuracies of the algorithms were as follows: Random Forest (62.37%),
580 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-
581 nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with
582 normalization, Random Forest achieved the highest overall metrics, including an
583 average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of
584 83.31%. Liu et al. concluded that Random Forest is highly accurate because it
585 can handle large, complex datasets, run processes in parallel using multiple trees,
586 and select the most relevant features to enhance model performance (Arifin et al.,
587 2021).

588 **2.3.7 Machine Learning for Geographical Traceability in**
589 **Bivalves**

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

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

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

616 **2.4 Limitations on Sex Identification in *Tegillarca***
617 ***granosa***

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

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

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

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

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

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

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

665 **2.5 Synthesis of the Study**

666 This section of the paper summarizes the technologies used in the different studies
667 related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex
668 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

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

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

⁶⁸³ Chapter 3

⁶⁸⁴ Research Methodology

⁶⁸⁵ This chapter discussed the materials and methods employed in the study, focusing
⁶⁸⁶ on the development requirements, as well as the software and programming
⁶⁸⁷ languages utilized. It also detailed the overall workflow in conducting the study,
⁶⁸⁸ Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca*
⁶⁸⁹ *granosa* (Linnaeus), 1758) using machine learning and deep learning technologies.

⁶⁹⁰ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,
⁶⁹¹ oversaw the overall workflow and conduct of the experiment. The researchers were
⁶⁹² also guided by research associates LC Mae Gasit and Allena Esther Artera. Con-
⁶⁹³ sequently, the entire dataset collection process was conducted at the University of
⁶⁹⁴ the Philippines Visayas hatchery facility.

⁶⁹⁵ The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical
⁶⁹⁶ Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-
⁶⁹⁷ tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-
⁶⁹⁸ ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine
⁶⁹⁹ Learning, (8) Morphological Characteristics Evaluation Using Deep Learning, and
⁷⁰⁰ (9) Evaluation Metrics

⁷⁰¹ 3.1 Sample Collection

⁷⁰² The collection of *T. granosa* samples used in this study was part of an ongoing
⁷⁰³ research project by UPV DOST-PCAARRD titled "Establishment of the Center
⁷⁰⁴ for Mollusc Research and Development: Development of Spawning and Hatchery
⁷⁰⁵ Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."

⁷⁰⁶ A total of 271 samples were provided for this study to classify the sex of *T. granosa*.
⁷⁰⁷ The samples, ranging in size from 34 to 61 mm, were sourced from the coastal area
⁷⁰⁸ of Zaraga, Iloilo, and fish markets in Ivisan, Capiz, Philippines (see Figure 3.1).

⁷⁰⁹ The research and experimentation were conducted at the University of the
⁷¹⁰ Philippines Visayas hatchery facility in Miagao, Iloilo, where the samples were
⁷¹¹ maintained in 200 L fiberglass-reinforced plastic (FRP) tanks containing filtered
⁷¹² seawater with 35 ppt salinity (Miranda & Ferriols, 2023).

⁷¹³ As part of the data collection process, the researchers utilized induced spawning
⁷¹⁴ and dissection to classify the sex of the samples. Induced spawning through
⁷¹⁵ temperature fluctuations was the most natural and least invasive method for bi-
⁷¹⁶ valves compared to other approaches (Aji, 2011). However, since not all samples
⁷¹⁷ exhibited gamete release, the researchers also performed dissections, assisted by
⁷¹⁸ hatchery staff, to expedite data collection. The sex of the dissected samples was
⁷¹⁹ identified based on the coloration of gonad tissue, which varies according to sex
⁷²⁰ and maturity stage. Females exhibited orange-red to pale orange gonads, while
⁷²¹ males displayed white to grayish-white gonads (May et al., 2021).

⁷²² The methods used for data collection were considered noninvasive, particularly
⁷²³ given that *T. granosa* are oxygen regulators well adapted to tidal exposure and
⁷²⁴ hypoxia (Davenport & Wong, 1986).

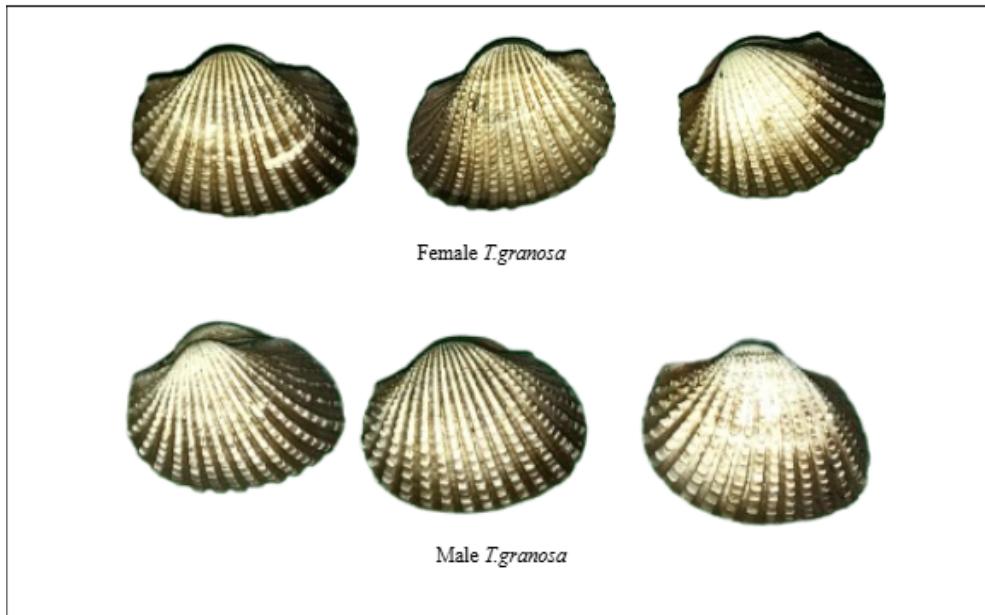


Figure 3.1: Male and Female *Tegillarca granosa* shells

725 3.2 Ethical Considerations

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

732 3.3 Creating *T. granosa* Dataset

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

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

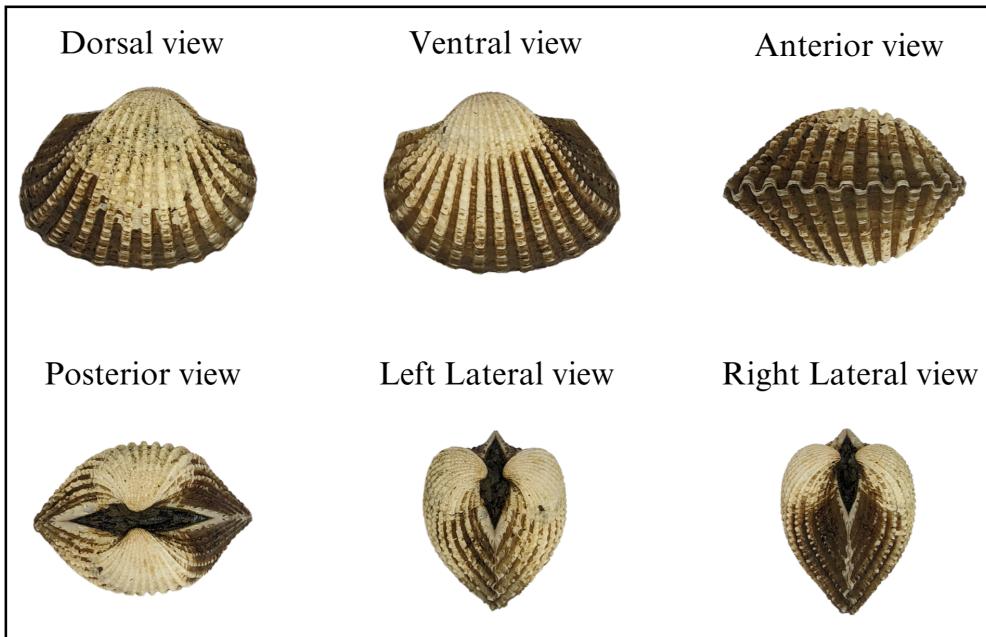


Figure 3.2: Different Views of the *T. granosa* Shell Captured

⁷⁵² 3.4 Morphological and Morphometric Characteristics Collection

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

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

⁷⁷⁰ 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.

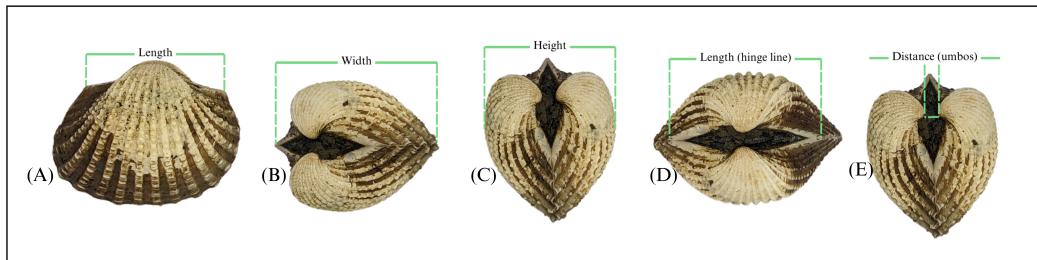


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

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

789 3.6 Hardware and Software Configuration

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

801 3.7 Morphometric Characteristics Evaluation Us- 802 ing Machine Learning

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

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

812 3.7.1 Data Preprocessing

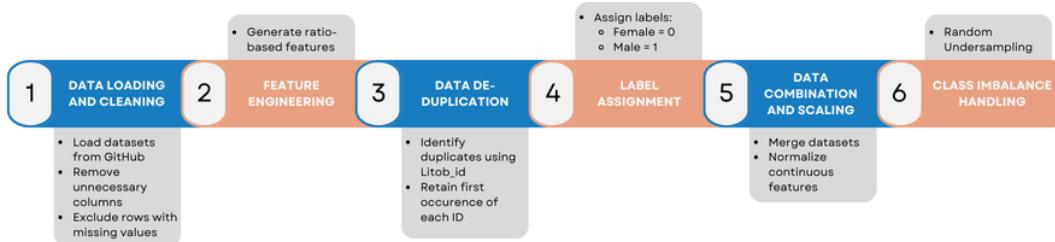


Figure 3.5: Data Preprocessing Pipeline

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

816 *Data Loading and Cleaning*

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

821 *Feature Engineering*

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

828 *Data De-duplication*

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

833 ***Label Assignment***

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

837 ***Data Combination and Scaling***

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

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

847 ***Class Imbalance Handling***

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

853 **3.7.2 Machine Learning Models Training**

854 ***Model Selection and Hyperparameter Tuning***

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

860 ***Cross-Validation***

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

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

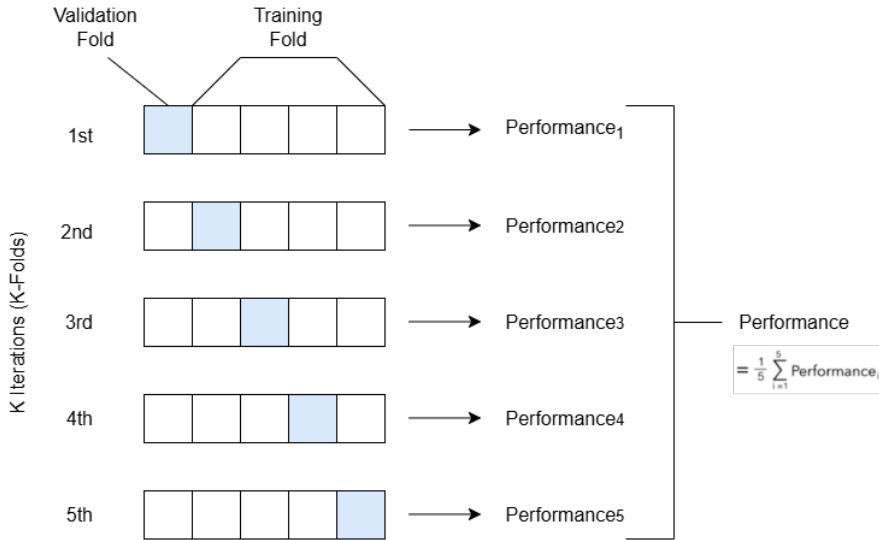


Figure 3.6: Diagram of k-fold cross-validation with $k = 5$

866 3.8 Morphological Characteristics Evaluation Us- 867 ing Deep Learning

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

872 *Image Preprocessing*

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

879 *Adjusting Dimensions*

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

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

885 ***Background Removal***

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

891 ***Shadow Removal***

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

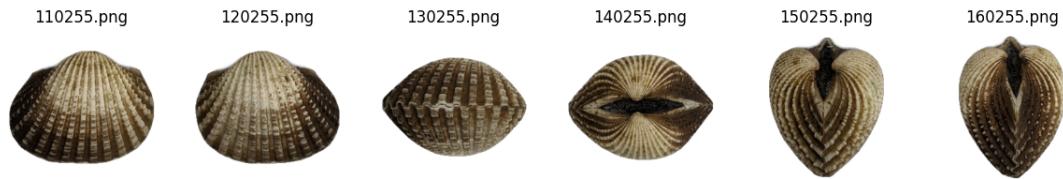


Figure 3.7: Shadows removed from male samples at different angles

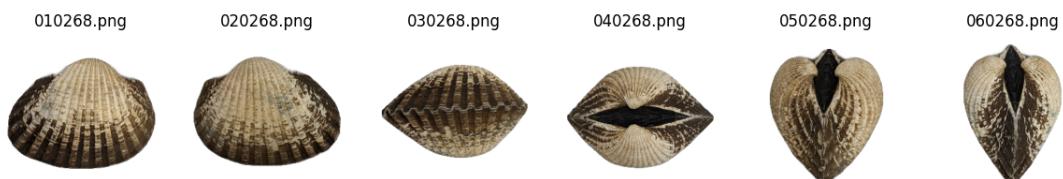


Figure 3.8: Shadows removed from female samples at different angles

896 **3.8.1 Convolutional Neural Network**

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

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

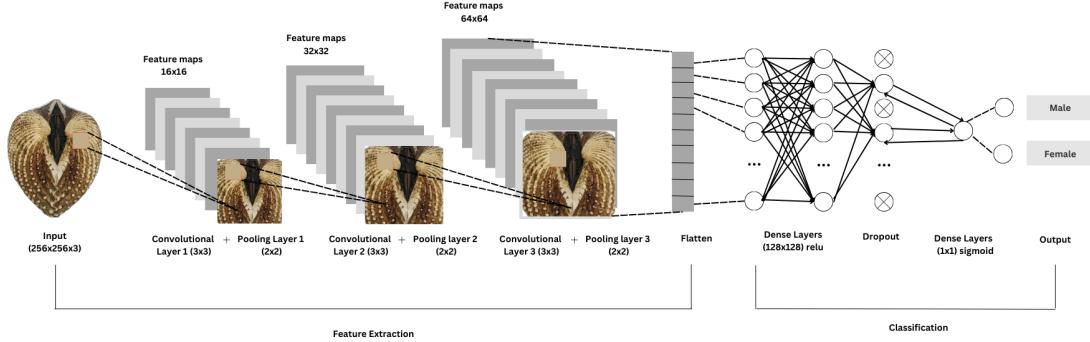


Figure 3.9: Architecture of Convolutional Neural Network (CNN)

905 ***Convolution Layer***

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

913 ***Pooling Layer***

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

917 ***Fully Connected and Dropout***

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

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

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

929 3.8.2 CNN Training

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

935 *Data Splitting*

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

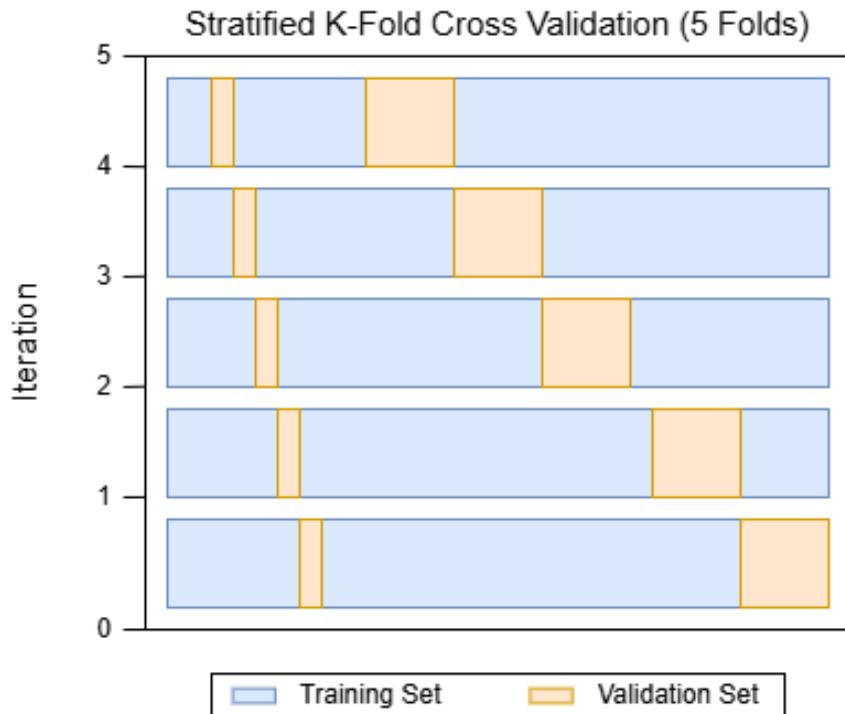


Figure 3.10: Diagram of stratified k-fold cross-validation with k=5

942 ***Data Augmentation***

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

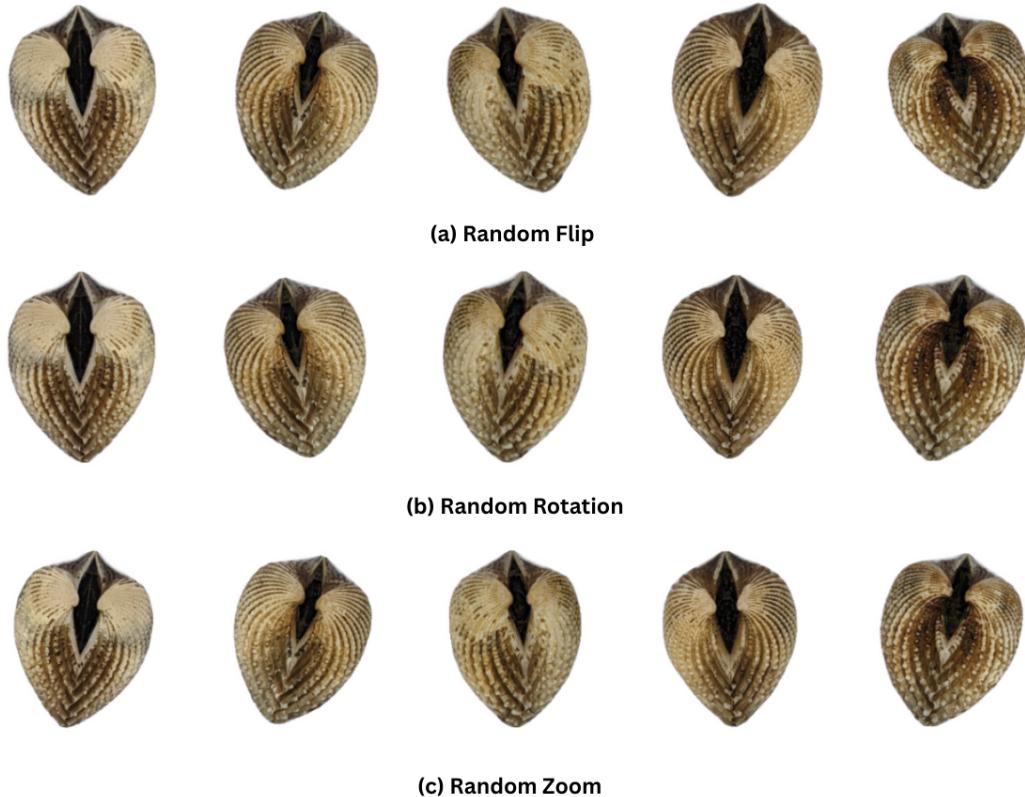


Figure 3.11: Data Augmentation Techniques

954 ***Training Procedure***

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

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

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

970 3.9 Evaluation Metrics

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

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

$$976 \text{ACC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

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

$$986 \text{PREC} = \frac{\text{True positive samples}}{\text{Samples assigned to positive class}} = \frac{TP}{TP + FP} \quad (3.2)$$

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

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

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

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

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

999 **Chapter 4**

1000 **Results and Discussions**

1001 This chapter presents the results from the machine learning and deep learning
1002 analyses conducted on the preprocessed dataset. It includes an evaluation of
1003 various machine learning classifiers and the application of deep learning models
1004 for image-based classification. The primary focus is on identifying key predictors
1005 and assessing classification performance for sex identification in *T. granosa*.

1006 **4.1 Machine Learning Analysis**

1007 This chapter outlines the results of preprocessing, training of machine learning
1008 models, and feature importance analysis, all conducted in Google Colab using
1009 Python. The dataset was preprocessed in Colab, and the training and evaluation
1010 of various classifiers were performed entirely within this environment. This part of
1011 the paper includes five subsections: data exploration, statistical analysis, feature
1012 importance analysis, performance evaluation, and confusion matrix analysis.

1013 **4.1.1 Data Exploration**

1014 Exploratory data analysis was performed to characterize the dataset using visu-
1015 alizations to understand the patterns and correlations within the data. A corre-
1016 lation heatmap was created to assess the relationship between the predictors and
1017 the target variable.

1018 The heatmap (see Figure 4.1) revealed three features most correlated with the

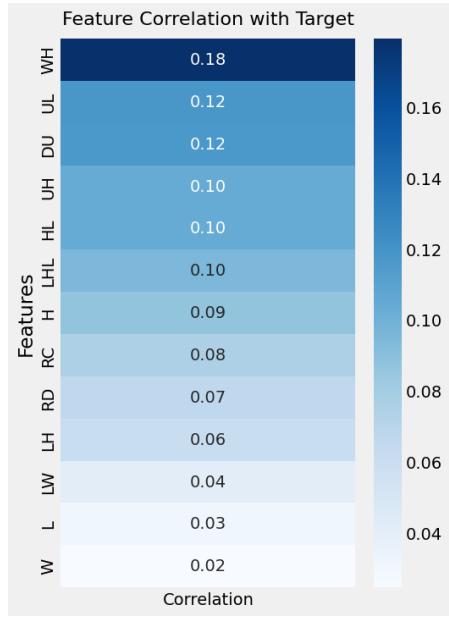


Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

1019 sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos-length ratio (r
 1020 $= 0.12$), and the distance between the umbos ($r = 0.12$). Each of these features
 1021 demonstrated a weak positive relationship with the target variable.

1022 **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

1023 As part of the exploratory data analysis, statistical testing confirmed that the
1024 dataset did not follow a normal distribution. Consequently, the Mann-Whitney
1025 U test was applied with a significance level of $\alpha = 0.05$ to compare male and
1026 female samples. Out of thirteen features, five showed statistically significant dif-
1027 ferences. These included: distance between umbos ($p = 0.025$), length-width ratio
1028 ($p = 0.011$), umbos-length ratio ($p = 0.019$), width-height ratio ($p = 0.003$), and
1029 umbos-height ratio ($p = 0.036$).

1030 It is important to note that statistical significance does not imply predictive
1031 importance. Therefore, further analysis, such as feature importance evaluation,
1032 was performed to identify the most informative predictors for classification.

1033 **4.1.3 Feature Importance Analysis**

1034 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric
1035 method that is suitable for evaluating differences in distributions across groups
1036 when the data does not follow a normal distribution. This approach was chosen
1037 because of the non-normality of the dataset and its robustness in handling con-
1038 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,
1039 2024)

1040 The analysis showed that the width-to-height ratio (WH ratio) had the high-

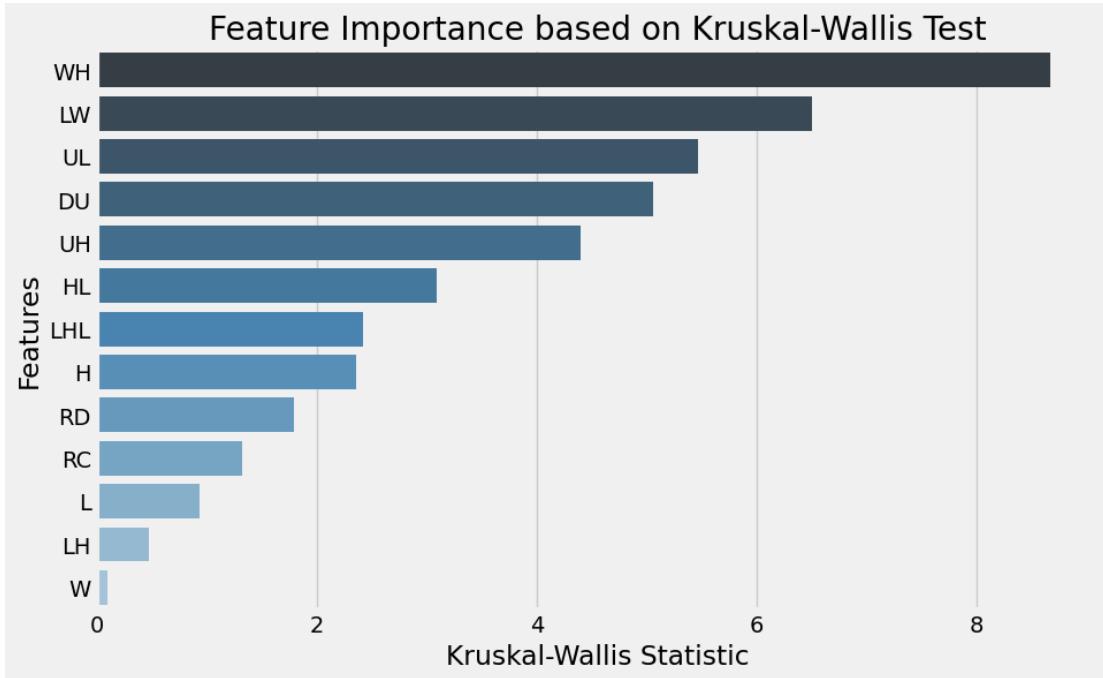


Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

1041 est importance score, indicating it is the most statistically significant feature for
 1042 distinguishing the sex of *T. granosa*. Other notable features included the length-
 1043 to-width ratio (LW ratio), umbo distance-to-length ratio (UL ratio), distance
 1044 between the umbos, and umbo distance-to-height ratio (UH ratio), all of which
 1045 contributed significantly to the classification task.

1046 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 | 59.07 | 59.54 | 59.07 | 58.45 |
| Random Forest | 59.85 | 59.99 | 59.85 | 59.80 |
| Gradient Boosting | 61.03 | 61.32 | 61.03 | 60.81 |
| AdaBoost | 60.63 | 60.98 | 60.63 | 60.39 |

Table 4.2: Performance Metrics for Models with All 13 Features

1047 Table 4.2 shows the performance metrics of different machine learning models
 1048 trained using all 13 features from the dataset. Among the models, Gradient

1049 Boosting achieved the highest accuracy of 61.03%, along with strong precision,
 1050 recall, and F1-score values. AdaBoost also performed competitively, with an ac-
 1051 curacy of 60.63%. These results highlight the effectiveness of ensemble methods
 1052 such as Gradient Boosting and AdaBoost when utilizing the full feature set, likely
 1053 because of their capability to combine multiple weak learners into a more robust
 1054 predictive model (Hussain & Zaidi, 2024).

| 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 | 61.04 | 61.68 | 61.04 | 60.67 |
| Random Forest | 61.01 | 61.12 | 61.01 | 60.91 |
| Gradient Boosting | 64.15 | 64.24 | 64.15 | 64.01 |
| AdaBoost | 61.02 | 61.26 | 61.02 | 60.82 |

Table 4.3: Performance Metrics for Models with 5 Features

1055 Table 4.3 presents the performance of the same models using only the top
 1056 5 features identified through Kruskal-Wallis feature importance analysis. The
 1057 selected features are the distance between the umbos, length-to-width ratio, width-
 1058 to-height ratio, umbo distance-to-height ratio, and umbo distance-to-length ratio.

1059 Interestingly, the overall performance of the models improved when using only
 1060 the top 5 features compared to using all 13. K-Nearest Neighbors (KNN) achieved
 1061 the best results with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%,
 1062 and an F1-score of 63.75%. Gradient Boosting followed closely behind. These find-
 1063 ings suggest that reducing the feature set to the most relevant variables helped
 1064 simplify the models, improved generalization, and enhanced predictive per-
 1065 formance—particularly for KNN, which showed a notable improvement over its ear-
 1066 lier results with the full feature set.

1067 4.1.5 Confusion Matrix Analysis

1068 Figure 4.3 summarizes the performance of the K-Nearest Neighbors model in
 1069 classifying *T. granosa* based on their sex, where 0 represents female samples and
 1070 1 represents male samples. From the matrix, we observe that out of all the actual
 1071 female samples (true label 0), 91 were correctly predicted as female (true positive
 1072 for class 0), while 36 were incorrectly classified as male (false negative for class
 1073 0). On the other hand, out of all the actual male samples (true label 1), 72 were
 1074 correctly predicted as male (true positive for class 1), while 55 were incorrectly
 1075 classified as female (false negative for class 1).

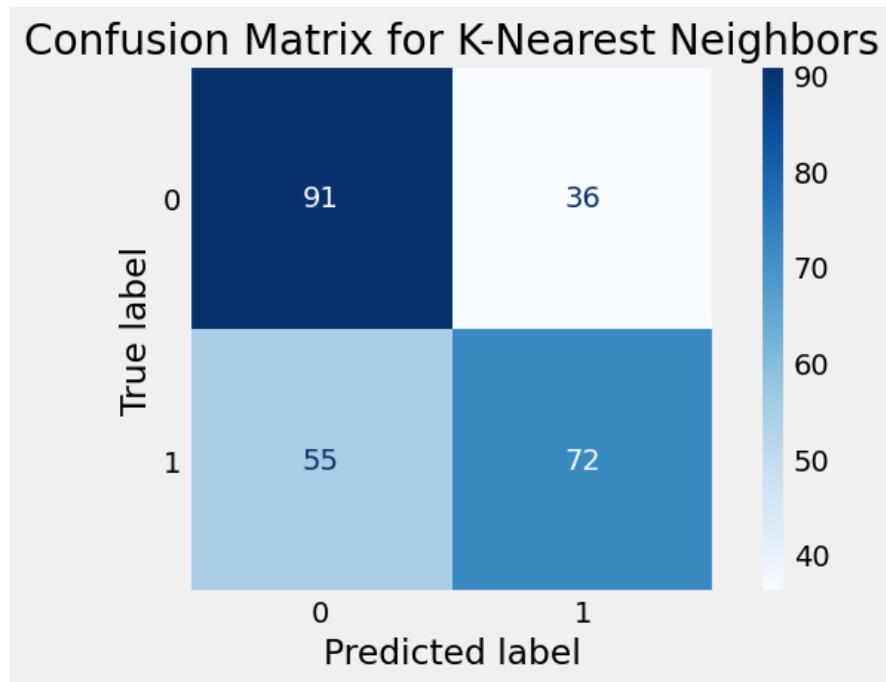


Figure 4.3: Feature Importance Scores Using the Kruskal-Wallis Test

1076 The distribution of correct and incorrect predictions suggests that the model
1077 performs slightly better at identifying female samples compared to male samples.
1078 Nevertheless, there is a noticeable amount of misclassification in both classes,
1079 which indicates room for improvement in the model's predictive performance.

1080 **Chapter 5**

1081 **Conclusion and**
1082 **Recommendations**

1083 **5.1 Conclusion**

1084 **5.2 Recommendations**

1085 This special problem entitled Morphometric-Based Non-invasive Sex Identification
1086 of *T. granosa* focuses on creating a baseline study that will serve as a foundation
1087 for further studies involving *Tegillarca granosa*, blood cockles using machine learn-
1088 ing, deep learning, and computer vision technologies in determining the sex of the
1089 samples is a salient need in aquaculture practices. Thus, the proceeding rec-
1090 ommendations are the future applications to improve and have detailed analysis
1091 such as focusing on shape analysis, exploring other state-of-the-art CNN such as
1092 ResNet, SqueezeNet, and InceptionNet, and comparing the analysis result. Fur-
1093 thermore, the main goal of conducting this is to have the ability to identify the
1094 sex of the samples by taking real-time angles by rotating from the dorsal, lateral,
1095 and ventral.

1096 Future studies could also invest in a much sturdier and more controlled envi-
1097 ronment by using a green background and positioning a webcam at a fixed angle.
1098 In addition, experiment with other image processing techniques such as scaling,
1099 rotating, and augmentation. The dataset can be utilized for further analysis us-
1100 ing deep learning and computer vision to make sense of the images gathered and
1101 discern sexual dimorphism for *T.granosa* or will serve as the basis for conducting
1102 similar studies to other bivalve species.

¹¹⁰³ References

- ¹¹⁰⁴ Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten
¹¹⁰⁵ years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,
¹¹⁰⁶ 5–16. doi: 10.1080/11250000409356545
- ¹¹⁰⁷ Afiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara
¹¹⁰⁸ granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,
¹¹⁰⁹ *10*.
- ¹¹¹⁰ Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian
1111 Sains*, *14*, 14207.
- ¹¹¹² Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).
¹¹¹³ Data scaling performance on various machine learning algorithms to identify
¹¹¹⁴ abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:
¹¹¹⁵ 10.14710/jtsiskom.2021.14105
- ¹¹¹⁶ Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.
¹¹¹⁷ *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- ¹¹¹⁸ Awan, A. A. (2022, November). *A complete guide to data augmentation*. Retrieved from <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>
- ¹¹¹⁹ Aypa, S. M., & Bacongus, S. R. (2000). Philippines: mangrove-friendly aquaculture.
¹¹²⁰ In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Surtida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on mangrove-friendly aquaculture organized by the seafdec aquaculture department, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- ¹¹²¹ BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
¹¹²² Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
¹¹²³ Resources.
- ¹¹²⁴ Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of
¹¹²⁵ waste cockle shell (anadara granosa) in biodiesel production from palm olein:
¹¹²⁶ Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.
¹¹²⁷ doi: 10.1016/j.fuel.2011.03.002
- ¹¹²⁸ Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining
1129 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127

- 1135 .v1
- 1136 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
1137 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A
1138 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1139 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/
1140 msq345
- 1141 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1142 proaches, challenges and opportunities for aquaculture. *Journal of Marine
1143 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1144 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).
1145 Mass mortalities in bivalve populations: A review of the edible cockle
1146 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1147 10.1016/j.ecss.2014.04.011
- 1148 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1149 research in bivalves: Towards the identification of molecular markers of
1150 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1151 10.1016/j.jprot.2012.04.027
- 1152 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data
1153 augmentation for forecasting with deep learning. *ArXiv (Cornell Univer-
1154 sity)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1155 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly
1156 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1157 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1158 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1159 10.1093/icb/ict076
- 1160 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1161 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma
1162 edule*) geographical traceability based on shell image analysis. *Ecological
1163 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1164 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1165 for chinese mitten crab using deep convolutional neural network. *Multi-
1166 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1168 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle
1169 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-
1170 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
- 1172 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1173 ageing of eel, *anguilla anguilla* (l.)? *International Review of Hydrobiology*,
1174 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1175 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and
1176 anatomy for young learners*. Rosie Research. Retrieved from <https://>

- 1177 rosiereresearch.com/clam-dissection-anatomy/
- 1178 Fao 2024 report: Sustainable aquatic food systems important for global food
1179 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1180 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1181 elemental composition: Complementary tools for stock discrimination of
1182 mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,
1183 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1184 GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*. Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1185 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)
- 1186 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Blackwell Science.
- 1187 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1188 Hussain, S. S., & Zaidi, S. S. H. (2024). Adaboost ensemble approach with
1189 weak classifiers for gear fault diagnosis and prognosis in dc motors. *Applied Sciences*, 14(7), 3105. doi: 10.3390/app14073105
- 1190 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1191 surface water characterization and heavy metal health risk on cockles in
1192 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1193 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1194 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)
- 1195 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest
1196 example of sexual dimorphism in bivalves—evidence from the astartid
1197 *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,
1198 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1199 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1200 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1201 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1202 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1203 learning-based phenotype classification of three ark shells: Anadara kagoshimensis,
1204 tegillarca granosa, and anadara broughtonii. *Frontiers in Marine Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1205 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the

- 1219 granulated ark, tegillarca granosa (linne). *The Korean Journal of Malacology*, 13, 55–64.
- 1220
- 1221 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as
1222 a tool for stock identification of the southern blue whiting, micromesistius
1223 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1224 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1225 et al. (2016). Identifying blue whiting (micromesistius poutassou) stock
1226 structure in the northeast atlantic by otolith shape analysis. *Canadian
1227 Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1228 cjfas-2015-0332
- 1229 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1230 tegillarca granosa (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.
1231 Arts Sci*, 4.
- 1232 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and
1233 larval rearing of the blood cockle, tegillarca granosa (linnaeus, 1758), in the
1234 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2
1235 .001
- 1236 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of
1237 local populations of the common sole solea solea (pisces, soleidae) in the nw
1238 mediterranean through otolith morphometrics and shape analysis. *Marine
1239 Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1240 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and
1241 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–
1242 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/
1243 kja.21209
- 1244 Narasimham, K. A. (1988). Taxonomy of the blood clams anadara (tegillarca)
1245 granosa (linnaeus, 1758) and a. (t.) rhombea (born, 1780).
- 1246 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1247 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1248 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1249 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1250 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1251 doi: 10.1002/jmor.10439
- 1252 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1253 morphometrics and machine learning challenge currently accepted species
1254 limits of the land snail placostylus (pulmonata: Bothriembryontidae) on the
1255 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1256 doi: 10.1093/mollus/eyz031
- 1257 Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test*.
1258 Retrieved from https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/ (Accessed: 2025-04-23)
- 1259
- 1260 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,

- 1261 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1262 carbons from consumption of blood cockle and exposure to contaminated
1263 sediments and water along the klang strait, malaysia. *Marine Pollution*
1264 *Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1265 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1266 of seasonal variation and anthropogenic stress on blood cockle (*Tegillarca*
1267 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-*
1268 *pogenic Stress on Blood Cockle (*Tegillarca Granosa*) Production Potential*,
1269 44(2), 62–82.
- 1270 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1271 chine learning and its applications to biology. *PLoS Computational Biology*,
1272 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1273 Team, K. (n.d.). *Keras documentation: ReduceLronplateau*. Retrieved from
1274 https://keras.io/api/callbacks/reduce_lr_on_plateau/
- 1275 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1276 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1277 & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).
1278 College Park, MD: Maryland Sea Grant.
- 1279 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1280 approach for morphological feature extraction based on variational auto-
1281 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*
1282 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1283 Wong, T. M., & Lim, T. G. (2018). *Cockle (*Anadara granosa*) seed produced in*
1284 *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in_3366.pdf
- 1285 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1286 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972
1287 .5008949
- 1288 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1289 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1290 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-
1291 pacts of four commonly used nanoparticles on the metabolism of a ma-
1292 rine bivalve species, *Tegillarca granosa*. *Chemosphere*, 296, 134079. doi:
1293 10.1016/j.chemosphere.2022.134079
- 1294 Zhan, P. L., Zha, S., & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1295 blood clam *Tegillarca granosa*. *Marine Environmental Research*, 177. Re-
1296 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

¹²⁹⁷ **Appendix A**

¹²⁹⁸ **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*

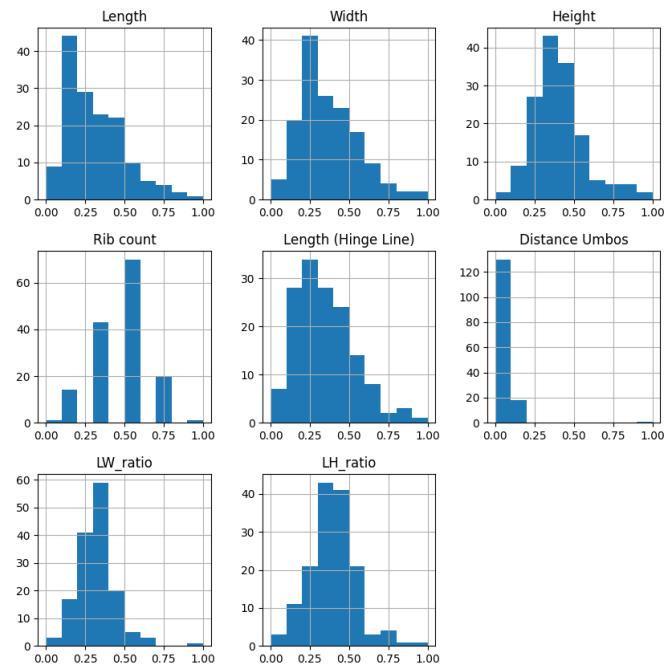


Figure A.7: Distribution of the Features of *Tegillarca granosa*