

1 MORPHOMETRIC AND MORPHOLOGICAL-BASED
2 NON-INVASIVE SEX IDENTIFICATION OF BLOOD
3 COCKLES *Tegillarca granosa* (LINNAEUS, 1758)

4 A Special Problem

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, Ph.D.

17 Adviser

18 Victor Marco Emmanuel FERRIOLS, Ph.D.

19 Co-Adviser

20 May 2025

21

Approval Sheet

22

The Division of Physical Sciences and Mathematics, College of Arts and
Sciences, University of the Philippines Visayas

24

certifies that this is the approved version of the following special problem:

25

**MORPHOMETRIC AND MORPHOLOGICAL-BASED
NON-INVASIVE SEX IDENTIFICATION OF BLOOD
COCKLES *Tegillarca granosa* (LINNAEUS, 1758)**

26

27

28

Approved by:

29

Name	Signature	Date
Francis D. Dimzon, Ph.D. (Adviser)	_____	_____
Victor Marco Emmanuel N. Ferriols, Ph.D. (Co-Adviser)	_____	_____
Ara Abigail E. Ambita (Panel Member)	_____	_____
Kent Christian A. Castor (Division Chair)	_____	_____

30 Division of Physical Sciences and Mathematics
31 College of Arts and Sciences
32 University of the Philippines Visayas

33 **Declaration**

34 We, Briana Jade Adricula, Gliezel Ann Pajarilla, and Ma. Christina Kane
35 Vito, hereby certify that this Special Problem has been written by us and is the
36 record of work carried out by us. Any significant borrowings have been properly
37 acknowledged and referred.

Name	Signature	Date
Briana Jade D. Adricula (Student)	_____	_____
Gliezel Ann T. Pajarilla (Student)	_____	_____
Ma. Christina Kane B. Vito (Student)	_____	_____

Dedication

“Hello, world.”

41

Acknowledgment

42

“Hello, world.”

Abstract

44 *Tegillarca granosa*, commonly known as blood cockles, is a significant marine bi-
45 valve species due to its nutritional benefits and economic importance. Accurate
46 sex determination is essential for maintaining a balanced male-to-female ratio, sus-
47 tainable harvesting, and resource management. However, sex-determining mecha-
48 nisms based on shell morphology are challenging macroscopically, and no existing
49 technologies are available for non-invasive sex classification. This study proposes
50 the use of machine learning and deep learning techniques to classify the sex of
51 blood cockles based on various shell measurements (length, width, height, hinge
52 line distance, umbo distance, and rib count) and images from multiple camera
53 angles (dorsal, ventral, anterior, posterior, and lateral views). The initial ma-
54 chine learning analysis using K-Nearest Neighbor (KNN) achieved an accuracy
55 of 64.16%, a precision of 64.97%, a recall of 64.16%, and an F1-score of 63.75%.
56 In contrast, deep learning with Convolutional Neural Networks (CNN) achieved
57 an accuracy of 71.68%, a precision of 72.52%, a recall of 69.29%, an F1-score of
58 69.12%, and an AUC score of 77.34% using images captured from the left lateral
59 angle. These results offer a non-invasive method for sex identification, which could
60 help in sustainable resource management and serve as a baseline for future studies
61 on blood cockles classification.

62 **Keywords:** deep learning, supervised machine learning, computer vision,
convolutional neural network, blood cockle, sex identifica-
tion, *Tegillarca granosa*

Contents

⁶³	1	Introduction	1
⁶⁴	1.1	Overview	1
⁶⁵	1.2	Problem Statement	2
⁶⁶	1.3	Research Objectives	4
⁶⁷	1.3.1	General Objective	4
⁶⁸	1.3.2	Specific Objectives	4
⁶⁹	1.4	Scope and Limitations of the Research	5
⁷⁰	1.5	Significance of the Research	6

⁷¹	2	Review of Related Literature	9
---------------	----------	-------------------------------------	----------

⁷²	2.1	Background on <i>T. granosa</i> and Their Importance	10
⁷³	2.2	Sex Identification Methods in <i>T. granosa</i>	14

75	2.3 Machine Learning and Deep Learning in Biology	16
76	2.3.1 Deep Learning for Phenotype Classification in Ark Shells .	17
77	2.3.2 Geometric Morphometrics and Machine Learning for Species	
78	Delimitation	18
79	2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning	18
80	2.3.4 Machine Learning for Shape Analysis of Marine Organisms	19
81	2.3.5 Deep Learning for Landmark-Free Morphological Feature	
82	Extraction	21
83	2.3.6 Machine Learning for Sex Differentiation in Abalone . . .	22
84	2.3.7 Machine Learning for Geographical Traceability in Bivalves	23
85	2.4 Limitations on Sex Identification in <i>Tegillarca granosa</i>	24
86	2.5 Chapter Summary	26
87	3 Research Methodology	31
88	3.1 Sample Collection	32
89	3.2 Ethical Considerations	33
90	3.3 Creating <i>T. granosa</i> Dataset	34
91	3.4 Morphometric and Morphological Characteristics Collection . . .	35
92	3.5 Image Acquisition and Data Gathering	36

CONTENTS ix

93	3.6 Hardware and Software Configuration	38
94	3.7 Morphometric Characteristics Evaluation Using Machine Learning	38
95	3.7.1 Data Preprocessing	39
96	3.7.2 Machine Learning Models Training	42
97	3.8 Morphological Characteristics Evaluation Using Deep Learning	42
98	3.8.1 Convolutional Neural Network	45
99	3.8.2 CNN Training	46
100	3.9 Evaluation Metrics	49
101	4 Results and Discussions	53
102	4.1 Machine Learning Analysis	53
103	4.1.1 Data Exploration	54
104	4.1.2 Statistical Analysis	55
105	4.1.3 Feature Importance Analysis	56
106	4.1.4 Performance Evaluation	56
107	4.1.5 Confusion Matrix Analysis	58
108	4.2 Deep Learning Analysis	59
109	4.2.1 Baseline Model	60

110	4.2.2 Comparison of Individual and Combined Angles	61
111	4.2.3 Training Result and Hyperparameter Tuning	62
112	4.2.4 Proposed Model	63
113	4.2.5 Learning Rates and Training Behavior per Fold	64
114	4.2.6 Visualizations	65
115	4.3 Discussions	69
116	5 Conclusion and Recommendations	73
117	5.1 Conclusion	73
118	5.2 Recommendations	75
119	6 References	77
120	A Code Snippets	85
121	B Resource Persons	87
122	C Data Gathering Documentation	89

¹²³ List of Figures

¹²⁴	2.1 Diagram of <i>T. granosa</i> 's external anatomy.	12
¹²⁵	3.1 Male and female <i>T. granosa</i> shells.	33
¹²⁶	3.2 Different views of the <i>T. granosa</i> shell captured	35
¹²⁷	3.3 Linear measurements of the <i>T. granosa</i> shell.	36
¹²⁸	3.4 Image acquisition setup for <i>T. granosa</i> samples.	37
¹²⁹	3.5 Data preprocessing in machine learning pipeline.	39
¹³⁰	3.6 Diagram of k-fold cross-validation with k = 5.	43
¹³¹	3.7 Shadows removed from male samples at different angles.	44
¹³²	3.8 Shadows removed from female samples at different angles.	44
¹³³	3.9 Architecture of convolutional neural network (CNN).	45
¹³⁴	3.10 Diagram of stratified k-fold cross-validation with k=5.	47
¹³⁵	3.11 Data augmentation techniques.	48

136	4.1 Heatmap of morphometric correlations with <i>T. granosa</i> sex.	54
137	4.2 Feature importance scores using the Kruskal-Wallis test.	57
138	4.3 KNN confusion matrix for <i>T. granosa</i> sex classification.	59
139	4.4 Training and validation accuracy per fold.	65
140	4.5 Average training and validation accuracy across folds.	66
141	4.6 Training and validation loss per fold.	66
142	4.7 Average training and validation loss across folds.	67
143	4.8 Confusion matrix for final model predictions.	68
144	4.9 ROC curve with AUC score for the proposed model.	69
145	C.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i>	89
146	C.2 Sex-Based Separation of <i>Tegillarca granosa</i> Samples Post-Spawning	89
147	C.3 Sex Identified Female Through Dissection of <i>Tegillarca granosa</i> .	90
148	C.4 Sex Identified Male Through Dissection of <i>Tegillarca granosa</i> .	90
149	C.5 Linear Measurements of Female <i>Tegillarca granosa</i>	90
150	C.6 Linear Measurements of Male <i>Tegillarca granosa</i>	91
151	C.7 Captured Images of Female <i>Tegillarca granosa</i>	91
152	C.8 Captured Images of Male <i>Tegillarca granosa</i>	91

List of Tables

153	2.1 Comparison of the Methods Used in Bivalves Studies	28
154	4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison	55
155	4.2 Performance Metrics for Models with All 13 Features	56
156	4.3 Performance Metrics for Models with 5 Features	57
157	4.4 Performance Metrics for Unbalanced vs. Balanced Datasets (Batch Size: 16, Epochs: 20)	60
158	4.5 Performance Metrics for Individual and Combined Angles (Batch Size: 16, Epochs: 20)	61
159	4.6 Effect of Batch Size and Epoch Values on CNN Model Performance	62
160	4.7 Performance Metrics for Different Activation Functions (Batch Size: 32, Epochs: 50)	63
161	4.8 Per-Fold Performance Metrics (Batch Size: 32, Epochs: 50, Activation Function: ReLU)	64

167	4.9 Learning Rate Reductions, Early Stopping, and Best Epochs per	
168	Fold During 5-Fold Cross-Validation	65

¹⁶⁹ Chapter 1

¹⁷⁰ Introduction

¹⁷¹ 1.1 Overview

¹⁷² The Philippines is a global center of marine biodiversity and has established aqua-
¹⁷³ culture as a significant contributor to total fishery production (Aypa & Baconguis,
¹⁷⁴ 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-
¹⁷⁵ ally and is the 11th largest seafood producer in the world. Aquaculture is deeply
¹⁷⁶ integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-
¹⁷⁷ duction of various aquatic species, including bivalves. Among these, blood cockles
¹⁷⁸ (*Tegillarca granosa*) hold considerable economic and environmental significance,
¹⁷⁹ making it essential to ensure sustainable production and population balance.

¹⁸⁰ Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent
¹⁸¹ overharvesting and ensure sustainability. An imbalanced ratio can lead to over-
¹⁸² exploitation and negatively impact the population's viability. However, there is

¹⁸³ limited literature on *T. granosa* that provides a thorough understanding of its
¹⁸⁴ sex-determining mechanisms, particularly regarding sexual dimorphism based on
¹⁸⁵ morphometric and morphological characteristics (Breton, Capt, Guerra, & Stew-
¹⁸⁶ art, 2017).

¹⁸⁷ Currently, sex determination methods for blood cockles are invasive, including
¹⁸⁸ dissection and histological examinations, which often result in the death of the
¹⁸⁹ species. While there is growing literature on sex identification in aquaculture
¹⁹⁰ commodities using machine learning and deep learning, there is a notable scarcity
¹⁹¹ of research specific to *T. granosa* (Miranda & Ferriols, 2023).

¹⁹² This study aims to provide a detailed baseline analysis of blood cockles by lever-
¹⁹³ aging their morphometric and morphological characteristics. Sexual dimorphism
¹⁹⁴ in bivalves is often subtle and challenging to establish macroscopically (Karapunar,
¹⁹⁵ Werner, Fürsich, & Nützel, 2021). However, by integrating machine learning and
¹⁹⁶ deep learning, the study seeks to identify distinct features that may indicate sexual
¹⁹⁷ dimorphism between male and female blood cockles.

¹⁹⁸ 1.2 Problem Statement

¹⁹⁹ Identifying the sex of *T. granosa* is important for promoting sustainable aquacul-
²⁰⁰ ture and biodiversity by maintaining a balanced male-to-female ratio. A balanced
²⁰¹ ratio helps prevent overharvesting. Although sex identification is crucial for blood
²⁰² cockle population management and sustainable aquaculture, there is a notable
²⁰³ lack of research on creating non-invasive methods for determining the sex of *T.*
²⁰⁴ *granosa*. Many recent studies and approaches rely on invasive methods like dis-

205 section or histological analysis, which are impractical for large-scale aquaculture
206 operations focused on conservation.

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

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

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

²²⁹ **1.3 Research Objectives**

²³⁰ **1.3.1 General Objective**

²³¹ The general objective of this study is to develop a non-invasive method for iden-
²³² tifying the sex of *Tegillarca granosa* using machine and deep learning integrated
²³³ with computer vision technologies. This method aims to provide accurate and
²³⁴ streamlined sex identification without causing harm to the specimens, thus sup-
²³⁵ porting sustainable aquaculture practices.

²³⁶ **1.3.2 Specific Objectives**

²³⁷ To achieve the overall general objective of developing a non-invasive sex identifi-
²³⁸ cation of *T. granosa* using machine learning, deep learning, and computer vision
²³⁹ technologies, the following specific objectives have been established:

- ²⁴⁰ 1. to collect and organize a comprehensive dataset of *T. granosa*, which will
²⁴¹ include linear measurements and images captured from different camera an-
²⁴² gles that will serve as the basis for training and evaluating the machine
²⁴³ learning and deep learning models,
- ²⁴⁴ 2. to develop and implement machine learning and deep learning models that
²⁴⁵ can classify the sex of *T. granosa* based on the collected linear measurements
²⁴⁶ and images of different camera angles of the sample, and determine the best
²⁴⁷ performing models, and
- ²⁴⁸ 3. to evaluate the model performance using performance metrics such as accu-

249 racy, precision, recall, and F1-score, AUC-ROC score for deep learning, and
250 optimize the performance by performing hyperparameter optimization.

251 1.4 Scope and Limitations of the Research

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

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

263 The researchers worked with 271 blood cockles that had been sex-identified and
264 taken from Panay Island, specifically sourced from Zarraga Iloilo and Ivisan Capiz.
265 These samples, divided between 144 males and 127 females, were obtained through
266 induced spawning via temperature shock and dissection. Data collection was lim-
267 ited to the spawned stage among the five gonadal stages - immature, developing,
268 mature, spawning, and spent stages. The other stages were not preferable due to
269 indistinguishable gonads and their inability to undergo induced spawning (May
270 et al., 2021). Thus, the researchers only focused on the samples undergoing the

²⁷¹ spawned stage.

²⁷² During the data collection, the researchers personally gathered linear measure-
²⁷³ ments, including length, width, height, rib count, length of the hinge line, and
²⁷⁴ distance between the umbos through the vernier caliper. The data-gathering pro-
²⁷⁵ cess was supervised by the University Research Associates from the Institute of
²⁷⁶ Aquaculture, College of Fisheries and Ocean Sciences. Aside from linear measure-
²⁷⁷ ments, images were taken from six different angles. The process of linear measure-
²⁷⁸ ments and image collection were non-invasive, considering the blood cockle-built
²⁷⁹ ability to survive in low oxygen environments and naturally inhabit intertidal
²⁸⁰ mudflats (Zhan & Bao, 2022).

²⁸¹ The method developed in this study is specific to *Tegillarca granosa* and may
²⁸² not apply to other bivalve species. The model was trained exclusively for *Te-*
²⁸³ *gillarca granosa* and morphometric and morphological features, which may not be
²⁸⁴ consistent and applicable across other shellfish species.

²⁸⁵ 1.5 Significance of the Research

²⁸⁶ This study will give us a significant advancement in non-invasive sex identifica-
²⁸⁷ tion methods in *T. granosa* providing innovative solutions that could solve the
²⁸⁸ challenges in identifying sex and reshape sustainable approaches to aquaculture.
²⁸⁹ The significance of this study extends to the following:

²⁹⁰ *Research Institution.* The result of this study focusing on the sex-identification
²⁹¹ mechanism of bivalves, specifically *Tegillarca granosa*, will provide valuable in-

²⁹² sights into universities and research centers that focus on fisheries and coastal
²⁹³ management, such as the UPV Institute of Aquaculture, that aim to develop
²⁹⁴ sustainable development and suitable culture techniques.

²⁹⁵ *Fishermen.* By developing a non-invasive method in sex identification, this study
²⁹⁶ can help long-term harvest efficiency and maintain the ratio of the harvest which
²⁹⁷ can help prevent exploitation of the *T. granosa*.

²⁹⁸ *Coastal Communities.* The result of this study would be beneficial for the coastal
²⁹⁹ communities that are reliant on their source of income with aquaculture com-
³⁰⁰ modities like blood cockles. Maintaining the diversity and aspect ratio of male
³⁰¹ and female may increase the market value of blood cockle production since cockle
³⁰² aquaculture faces significant obstacles worldwide due to the fluctuating seed sup-
³⁰³ plies and scarcity of broodstock from the wild.

³⁰⁴ *Future Researchers.* The result of this study would serve as the basis for studies
³⁰⁵ that involve sex identification in bivalves such as *T. granosa*. Some technologies
³⁰⁶ are yet to be explored in machine learning, deep learning, and computer vision
³⁰⁷ technologies that can lead to higher accuracy and distinguish the presence of
³⁰⁸ sexual dimorphism in the *T. granosa*.

³⁰⁹ Chapter 2

³¹⁰ Review of Related Literature

³¹¹ Aquaculture is the fastest-growing industry in animal food production and has
³¹² great potential as a sustainable solution to global food security, nutrition, and
³¹³ development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
³¹⁴ *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
³¹⁵ tegrated into the livelihoods of Filipinos, not only through fish cultivation but
³¹⁶ also through the production of other aquatic species, including mollusks, oysters,
³¹⁷ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
³¹⁸ clams *Tegillarca granosa*, have economic and environmental significance. It has
³¹⁹ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
³²⁰ vesting and maintain the optimal ratio to preserve the population and production
³²¹ of the blood cockles.

³²² The members of the Arcidae Family, including *T. granosa* are important sources
³²³ of food and livelihood. Cockle aquaculture meets rising demands, however, it
³²⁴ faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,

325 2023). To solve the problem, researchers exert a considerable amount of effort,
326 developing a broader understanding of bivalves, including their sex-determining
327 mechanism, due to their notable importance in terms of diversity, environmental
328 benefits, and economic and market importance (Breton et al., 2017). Despite the
329 promising idea of identifying sex, there is limited research reported in terms of
330 sexual dimorphism, making it harder to distinguish through its morphological and
331 morphometric characteristics.

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

337 **2.1 Background on *T. granosa* and Their Im- 338 portance**

339 *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.
340 In the Philippines, it is known locally as Litob and Bakalan, a marine bivalve
341 species from the family Arcidae. Litob is widely distributed in the world including
342 Southeast Asia. They can be found in the intertidal mudflats adjacent to the
343 mangrove forest (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With
344 the intertidal mudflat as *T. granosa*'s habitat, they experience severe hypoxia or
345 low oxygen levels in the blood tissues during the tidal cycle. The blood clams
346 exhibit a unique red-blood phenotype where it serves two purposes the hemocyte

³⁴⁷ carries oxygen around the body and strengthens immune defenses. In addition,
³⁴⁸ it possesses a unique ability to absorb oxygen at similar rates in water and air
³⁴⁹ (Zhan & Bao, 2022).

³⁵⁰ *T. granosa* shell (refer to Figure 2.1) is medium-sized, fairly thick, ovate, and
³⁵¹ convex, with both valves being equal in size but asymmetrical from the hinge. The
³⁵² top edge of the dorsal margin is straight, while the front is rounded and slopes
³⁵³ downward, with its back being obliquely rounded with a concave bottom edge.
³⁵⁴ It has a narrow diamond-shaped ligament near the hinge with 3-4 dark chevron
³⁵⁵ markings, although some may be incomplete. The shell's outer layer, or the
³⁵⁶ periostracum, is smooth and brown with a straight hinge line and 40-68 fine short
³⁵⁷ teeth arranged in a straight line. The beak, or prosogyrate, curves forward, with
³⁵⁸ the shell having 18–21 raised ribs with blunt nodules and spaces between them.
³⁵⁹ The inner shell is white with crenulations along the valves' ventral, anterior, and
³⁶⁰ posterior margins. The posterior adductor scar is elongated and squarish, while
³⁶¹ the anterior adductor scar is similar but smaller in size. The mantle covering the
³⁶² bulk of *T. granosa*'s visceral mass is thin but the edges are thick and muscular.
³⁶³ It bears the impression of the crenulated shell edges. Their foot is large with a
³⁶⁴ ventral grove with no byssus or thread-like attachment. The *T. granosa*'s soft
³⁶⁵ body is blood red (Narasimham, 1988).

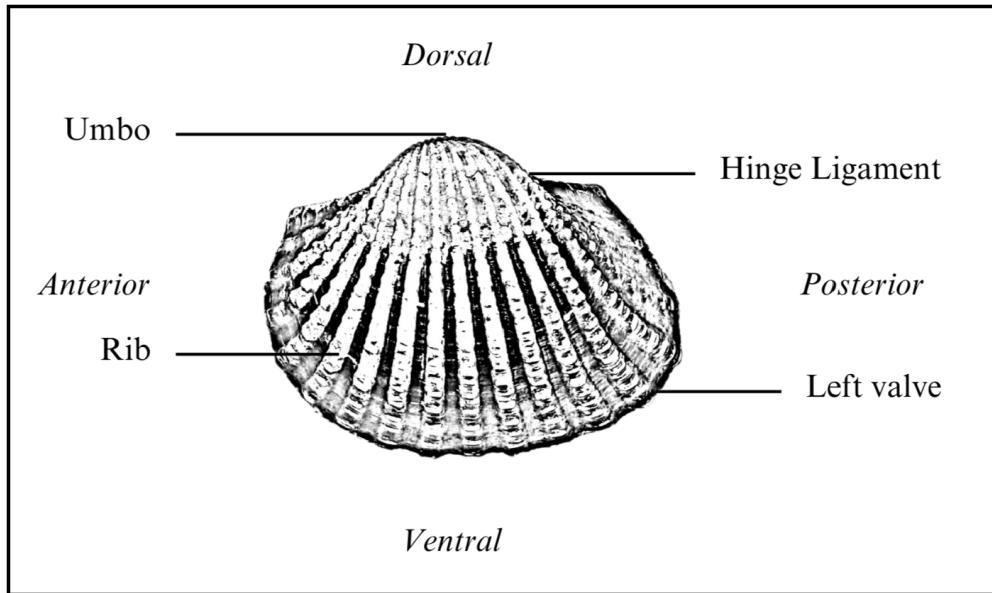


Figure 2.1: Diagram of *T. granosa*'s external anatomy.

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

380 Maniam, Hamid, & Ali, 2011).

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

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

403 2.2 Sex Identification Methods in *T. granosa*

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

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

417 The researchers highlighted the importance of *T. granosa* and other anadarinids
418 as a food source established worldwide, especially in Malaysia and Korea. How-
419 ever, in the Philippines, the bivalve aquaculture of the clam species is still limited.
420 The experiment, which focused on the culture and rearing of *T. granosa*, was at-
421 tempted by subjecting the wild broodstocks to a series of temperature fluctuations
422 to induce the spawning of gametes. This is currently the most natural and least
423 invasive method for bivalves (Aji, 2011). The study of Miranda and Ferriols aimed
424 to pave the way for the sustainable production of *T. granosa* seeds for aquaculture
425 and stock enhancement, despite the scarcity of documented hatchery culture of
426 *T. granosa* from larvae to adults in the Philippines.

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

437 The classification of the gonad stages used was by Yurimoto et al. (2014). There
438 are five maturation stages of gonadal development: immature (Stage I), devel-
439 oping (Stage II), mature (Stage III), spawning (Stage IV), and spent (Stage V)
440 stages. The sex of the *T. granosa* was confirmed by the color of the gonad and
441 by conducting a histological examination of the gonads. During the immature
442 stage, sex determination was indistinguishable due to the difficulties of observing
443 the germ cells. In the developing stage, the spermatocytes and a few spermatids
444 can be seen for males, and immature oocytes are attached to the tube wall for
445 the female. In the mature stage, the follicles are full of spermatozoa with their
446 tails pointing towards the center of the tube for the male, and the female is full
447 of mature oocytes that are irregular or polygonal in shape with the oval nucleus.
448 Upon reaching spawning, some spermatozoa are released, causing the empty space
449 in the follicle wall for males and females. There is a decrease in the number of
450 mature oocytes and it exhibits nuclear disappearance due to the breakdown of
451 the germinal vesicle. Lastly, the spent stage is where the genital tube is deformed

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

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

470 2.3 Machine Learning and Deep Learning in Bi- 471 ology

472 Machine learning has the potential to improve the quality of life of human beings
473 and has a wide range of applications in terms of research and development. The

474 term machine learning refers to the invention and algorithm evaluation that en-
475 ables pattern recognition, classification, and prediction based on models generated
476 from available data (Tarcă, Carey, Chen, Romero, & Drăghici, 2007). The study
477 of machine learning methods has advanced in the last several years, including bio-
478 logical studies. In biological studies, machine learning has been used for discovery
479 and prediction. This section will explore existing machine learning studies that
480 are applied in biological sciences, highlighting the identification of sex in shells,
481 bivalves, and mollusks.

482 2.3.1 Deep Learning for Phenotype Classification in Ark 483 Shells

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

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

497 **2.3.2 Geometric Morphometrics and Machine Learning for**
498 **Species Delimitation**

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

509 **2.3.3 Contour Analysis in Mollusc Shells Using Machine**
510 **Learning**

511 Tuset et al. (2020), in their study, *Recognising mollusc shell contours with enlarged*
512 *spines: Wavelet vs Elliptic Fourier analyses*, mentioned that gastropod shells have
513 large spines and sharp shapes that differ based on environmental, taxonomic, and
514 evolutionary influences. The researchers stated that classic morphometric meth-
515 ods may not accurately depict morphological features of the shell, especially when
516 using the angular decomposition of the contour. The current research examined
517 and compared the robustness of the contour analysis using wavelet transformed
518 and Elliptic Fourier descriptors for gastropod shells with enlarged spines. For

519 that, the researchers analyzed two geographically and ecologically separated pop-
520ulations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed
521 that contour analysis of gastropod shells with enlarged spines can be analyzed
522 using both methodologies, but the wavelet analysis provided better local discrim-
523ination. From an ecological perspective, shells with various sizes of spines in both
524 areas indicate the broad adaptability of the species.

525 **2.3.4 Machine Learning for Shape Analysis of Marine Or-
526 ganisms**

527 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*
528 *to Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*
529 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to
530 the traditional method of collecting linear measurements with the application of
531 multivariate statistical methods and outline analysis in recording the structures
532 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and
533 elasmobranchs) with their hard bodies have been used as an indication of age and
534 a determinable time-scale and structure continue to go through life (Arkhipkin,
535 2005; Kerr & Campana, 2014). This study has explored variations in the mor-
536 phometry of recording structures in stock discrimination and systematics. The
537 researchers utilized the principal component analysis rather than the traditional
538 approach, which helps simplify the data without losing important information.
539 They utilized landmark-based geometric morphometrics, which has three differ-
540 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other
541 morphogenetic processes, and lastly, the extremal points are constructed land-

542 marks.

543 Generalized Procrustes Analysis (GPA) is a common superimposition technique in
544 landmark-based geometric morphometrics that aligns landmarks via translation,
545 scaling, and rotation to eliminate non-shape deviations (Zelditch, Swiderski, &
546 Sheets, 2004). However, there is a limit to the amount of smooth areas that may
547 be captured, and it is possible to overlook significant shape details. Utilization
548 of the semi-landmarks enhanced the shape description (Adams, Rohlf, & Slice,
549 2004). The researchers observed that using an outline-based approach would be
550 more effective than using a landmark-based approach.

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

564 Landmark-based methods used in the study showed that there are detectable
565 differences between male and female octopuses. However, the accuracy of deter-

566 mining sex based on these differences was low, similar to the results obtained
567 with traditional morphometric techniques. The study involved a relatively small
568 sample size of 160 individuals, and the structure being analyzed (the stylet, or
569 internalized shell) varies significantly between individuals. Although the results
570 aligned with findings from other studies that attempted to identify gender differ-
571 ences in cephalopods, the researchers concluded that the approach might not be
572 accurate enough for reliable sex determination.

573 **2.3.5 Deep Learning for Landmark-Free Morphological Fea-** 574 **ture Extraction**

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

585 2.3.6 Machine Learning for Sex Differentiation in Abalone

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

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

608 The study found that Naive Bayes consistently performed better than other algo-

609 rithms. However, when applied to both min-max and zero-mean normalization,
610 the average accuracies of the algorithms were as follows: Random Forest (62.37%),
611 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-
612 nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with
613 normalization, Random Forest achieved the highest overall metrics, including an
614 average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of
615 83.31%. Liu et al. concluded that Random Forest is highly accurate because it
616 can handle large, complex datasets, run processes in parallel using multiple trees,
617 and select the most relevant features to enhance model performance (Arifin et al.,
618 2021).

619 2.3.7 Machine Learning for Geographical Traceability in 620 Bivalves

621 In the study, *BivalveNet: A hybrid deep neural network for common cockle (*Ceras-**

622 *toderma edule*) geographical traceability based on shell image analysis, the re-
623 searchers incorporated computer vision and machine learning technologies for an
624 efficient determination of blood cockle harvesting origin based on the shell geomet-
625 ric and morphometric analysis. It aims to improve the traceability methodologies
626 in these organisms and its potential as a reliable traceability tool. Thirty *Cerasto-*
627 *derma edule* samples were collected along the five locations on the Atlantic West
628 and South Portuguese coast with individual images processed using lazy snapping
629 segmentation, spectro-textural-morphological phenotype extraction, and feature
630 selection through hybrid Principal Component Analysis and Neighborhood Com-
631 ponent Analysis (Concepcion, Guillermo, Tanner, Fonseca, & Duarte, 2023).

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

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

647 **2.4 Limitations on Sex Identification in *T. gra-***

648 ***nosa***

649 To date, no distinction has been made between the male and female *T. granosa*
650 in sexing methodology. In cockle aquaculture without clearly apparent sexual
651 dimorphism, sexing can be performed using invasive methods such as chemical
652 stimulation, dissection, and gonad-stripping. Induced spawning, specifically tem-
653 perature shock, is the most natural and least invasive method for bivalves (Aji,

654 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water
655 from hot to cold with a specific temperature requires deliberate and careful ma-
656 nipulation of the temperature over a specific period and would require constant
657 management and monitoring.

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

666 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous (syn-
667 chronous or functional) characteristics. Sequential hermaphrodites switch genders
668 after being male or female for one or multiple yearly cycles. (Heller, 1993; Gosling,
669 2004; Collin, 2013). Sex change and consecutive hermaphroditism have been ob-
670 served in different bivalve species, including Ostreidae, Pectinidae, Veneridae,
671 and Patellidae. However, macroscopically differentiating bivalve sex is challeng-
672 ing. The only way it may be identified is through histological analysis of gonad
673 remains but to do so there is an act of killing the organism (Coe, 1943; Gosling,
674 2004). Verification of sex change in bivalves to classify whether male or female
675 while they are alive is challenging since they need to be re-confirmed and re-
676 evaluated to be the same individual after a year.

677 Lee et al. (2012) found out that *T. granosa*, a species in Arcidae, has been dis-

678 covered to be a sequential hermaphrodite, with the sex ratio changing with an
679 increase in the shell size. In bivalves, sex changes usually happen when the gonad
680 is not differentiated between spawning seasons (Thompson, Newell, Kennedy, &
681 Mann, 1996). But in *T. granosa*, after the spawning season, sex changes during
682 its inactive phase. Results showed a 15.1% sex change ratio, with males having
683 a higher sex change ratio (21.2%) than females (6.2%). The 1+ year class had a
684 higher ratio (17.8%) than the 2+ year class (12.1%). Thus, this study indicates
685 that *T. granosa* is a sequential hermaphrodite. The results of the study demon-
686 strated that the bivalve's age affects the sex ratio and degree of sex change, but
687 additional in-depth investigation is required to determine the role that genetic
688 and environmental factors play in these changes.

689 No literature in the study of mollusks specifically addresses the machine learn-
690 ing algorithm used to determine the sex of *T. granosa* bivalves in various mod-
691 els. Nevertheless, various techniques such as shape analysis, morphometric ana-
692 lysis, Wavelet, and Fourier analysis, as well as different deep learning models like
693 VGNet, ResNet, and SqueezeNet in CNN networks, are utilized for phenotype
694 classification, while different machine learning algorithms could serve as the foun-
695 dation for this research project.

696 2.5 Chapter Summary

697 This section of the paper summarizes the technologies used in the different studies
698 related to the pursuit of the study entitled, Morphometric and Morphological-
699 Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca granosa* (Lin-

700 naeus, 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. Phyus 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. Tuset and E. Galimany and A. Farrés and E. Marco-Herrero and J. L. Otero-Ferrer and A. Lombarte and M. Ramón	Wavelet functions and Elliptic Fourier descriptors	Addresses the difficulty of accurately defining phenotypic diversity in gastropod shells.	Advanced contour analysis methods allow accurate differentiation of gastropod shell forms.	Cannot clarify the causes of phenotypic variation in the two populations studied.
Fedor Lishchenko and Jones, J. B.	Landmark- and outline-based Geometric Morphometric methods	To address difficulties in differentiating between stocks of marine organisms to prevent misidentification that could affect conservation and management.	Shape analysis improves taxonomic classification precision and offers close distinction between related species or organisms.	Landmark-based methods can be sensitive to landmark placement.
M. Tsutsumi and N. Saito and D. Koyabu and C. Furusawa	Morphological regulated variational AutoEncoder (Morpho-VAE)	The need for reliable, landmark-free methods, such as a modified variational autoencoder, to extract and decipher complex shapes from image data.	Employs dimension reduction and feature extraction, making it a user-friendly tool for biology non-experts.	Limited sample size in certain families presented challenges.
Barrera-Hernandez, R. and Barrera-Soto, V. and Martinez-Rodriguez, J. L. and Ríos-Alvarado, A. B. and Ortiz-Rodríguez, F.	Machine learning algorithms	Identifying the sex of abalones is challenging for producers applying specific growth or preservation strategies.	Machine learning algorithms accurately classify abalone sex into three categories: male, female, and immature.	Selected features may not fully capture the complexity of abalone morphology.
Concepcion, R. and Guillermo, M. and Tanner, S. E. and Fonseca, V. and Duarte, B.	EfficientNet-Bo, ResNet101, MobileNetV2, InceptionV3	Addresses the difficulty of accurately tracing bivalve harvesting origins using computer vision and machine learning algorithms to enhance seafood traceability and combat food fraud.	Non-invasive, image-based tools for bivalve traceability provide faster, cheaper, and equally accurate alternatives to traditional chemical analysis methods.	Small sample size (only 30 cockles) limits model reliability.

Table 2.1: Comparison of the Methods Used in Bivalves Studies

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

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

⁷¹⁵ Chapter 3

⁷¹⁶ Research Methodology

⁷¹⁷ This chapter discusses the materials and methods employed in the study, focusing
⁷¹⁸ on the development requirements, as well as the software and programming
⁷¹⁹ languages utilized. It also detailed the overall workflow in conducting the study,
⁷²⁰ Morphometric and Morphological-Based Non-Invasive Sex Identification of Blood
⁷²¹ Cockles *Tegillarca granosa* (Linnaeus), 1758) using machine learning and deep
⁷²² learning technologies.

⁷²³ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture, oversaw
⁷²⁴ the overall workflow by providing baseline characteristics of the samples that
⁷²⁵ the researchers could focus on. Additionally, guidance was offered by the research
⁷²⁶ associates LC Mae Gasit and Allena Esther Artera. Consequently, the entire
⁷²⁷ dataset collection process was conducted at the University of the Philippines Visayas
⁷²⁸ hatchery facility.

⁷²⁹ The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical Con-

730 siderations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteristics
731 Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Software
732 Configuration,(7) Morphometric Characteristics Evaluation Using Machine Learn-
733 ing, (8) Morphological Characteristics Evaluation Using Deep Learning, and (9)
734 Evaluation Metrics

735 3.1 Sample Collection

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

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

743 The research and experimentation were conducted at the University of the Philip-
744 pines Visayas hatchery facility in Miagao, Iloilo, where the samples were main-
745 tained in 200 L fiberglass-reinforced plastic (FRP) tanks containing filtered sea-
746 water with 35 ppt salinity (Miranda & Ferriols, 2023).

747 As part of the data collection process, the researchers utilized induced spawn-
748 ing and dissection to classify the sex of the samples. Induced spawning through
749 temperature fluctuations was the most natural and least invasive method for bi-
750 valves compared to other approaches (Aji, 2011). However, since not all samples
751 exhibited gamete release, the researchers also performed dissections, assisted by

752 hatchery staff, to expedite data collection. The sex of the dissected samples was
753 identified based on the coloration of gonad tissue, which varies according to sex
754 and maturity stage. Females exhibited orange-red to pale orange gonads, while
755 males displayed white to grayish-white gonads (May et al., 2021).

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

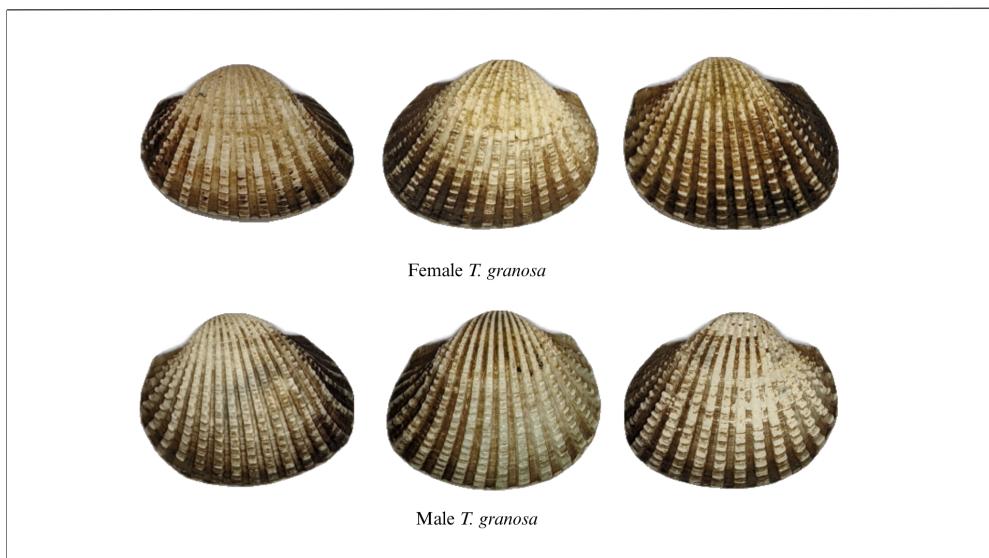


Figure 3.1: Male and female *T. granosa* shells.

759 3.2 Ethical Considerations

760 The ongoing research project titled "Establishment of the Center for Mollusc Re-
761 search and Development: Development of Spawning and Hatchery Techniques for
762 the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture"—from which
763 the samples used in this study were obtained—was reviewed and approved by the

764 Institutional Animal Care and Use Committee (IACUC) of the University of the
765 Philippines Visayas.

766 3.3 Creating *T. granosa* Dataset

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

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

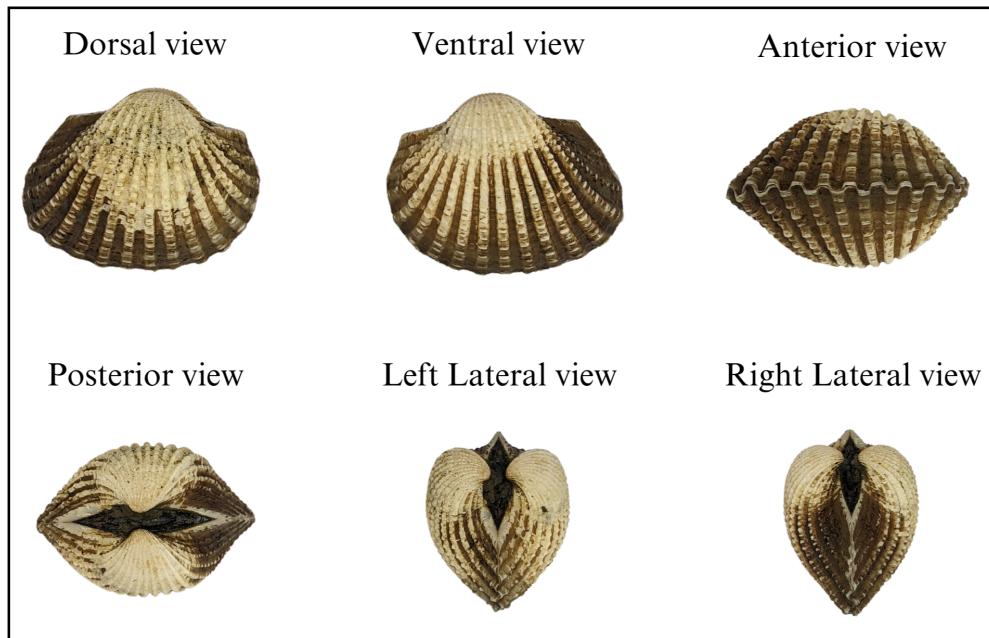


Figure 3.2: Different views of the *T. granosa* shell captured

786 3.4 Morphometric and Morphological Charac- 787 teristics Collection

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

797 The researchers measured the height, width, and length of *T. granosa* using a

798 Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the
 799 corresponding measurement diagram. Length (A) refers to the distance from the
 800 anterior to the posterior of the shell. Width (B) is defined as the widest span
 801 across the shell from the left to the right valve. Height (C) measures the distance
 802 from the base to the apex of the shell. In addition, the hinge line length (D) near
 803 the hinge and the distance between the umbos (E) were recorded.
 804 Reymant and Kennedy (1998) emphasized that including rib count as supplemen-
 805 tary information can enhance identification accuracy. Following this insight, the
 806 researchers also recorded the rib count for both male and female *T. granosa*, ad-
 807 justing the values by calculating ratios to account for natural size variation among
 808 specimens.

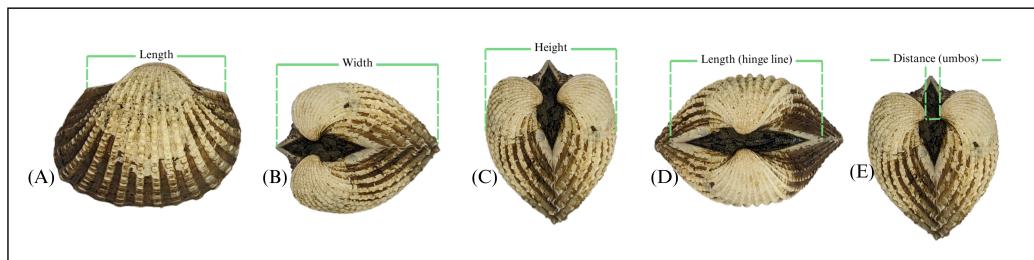


Figure 3.3: Linear measurements of the *T. granosa* shell.

809 3.5 Image Acquisition and Data Gathering

810 This study comprised 144 male and 127 female *T. granosa* samples, resulting
 811 in a total of 1,626 images captured from various angles. To ensure consistency
 812 during image acquisition, the researchers constructed a box-like structure with
 813 a white background to control the imaging environment (see Figure 3.4). This
 814 setup allowed for uniform image captures by fixing the camera at a consistent

815 angle directly above the *T. granosa*. A ring light was positioned in front of the
816 box to enhance image quality, eliminate shadows, and ensure clarity of the samples
817 throughout the image acquisition process.

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



Figure 3.4: Image acquisition setup for *T. granosa* samples.

823 **3.6 Hardware and Software Configuration**

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

835 **3.7 Morphometric Characteristics Evaluation Us- 836 ing Machine Learning**

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

3.7. MORPHOMETRIC CHARACTERISTICS EVALUATION USING MACHINE LEARNING 3

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

846 3.7.1 Data Preprocessing

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

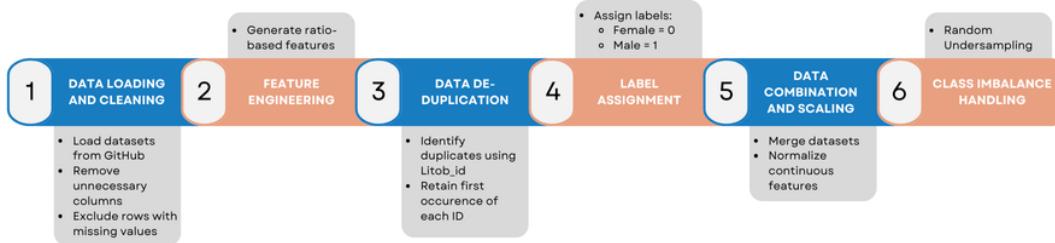


Figure 3.5: Data preprocessing in machine learning pipeline.

850 *Data Loading and Cleaning*

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

855 *Feature Engineering*

856 Additional ratio-based features were generated to augment the existing measure-
857 ments. These included the length-to-width ratio, length-to-height ratio, width-
858 to-height ratio, hinge line length-to-length ratio, umbos distance-to-length ratio,
859 umbos distance-to-height ratio, and rib density. These derived features aimed to

860 emphasize shape characteristics independent of size, improving the models' ability
861 to distinguish morphological differences between sexes.

862 ***Data De-duplication***

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

3.7. MORPHOMETRIC CHARACTERISTICS EVALUATION USING MACHINE LEARNING 4

867 ***Label Assignment***

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

871 ***Data Combination and Scaling***

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

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

881 ***Class Imbalance Handling***

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

887 3.7.2 Machine Learning Models Training**888 *Model Selection and Hyperparameter Tuning***

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

894 *Cross- Validation*

895 A five-fold cross-validation approach was implemented (refer to Figure 3.6). The
896 dataset was divided into five subsets, with four used for training and one for
897 testing. This process was repeated five times, with each fold serving as the test set
898 once. This method ensured that model evaluation was robust and generalizable,
899 minimizing the bias that may result from a single train-test split. (GeeksforGeeks,
900 2024)

**901 3.8 Morphological Characteristics Evaluation Us-
902 ing Deep Learning**

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

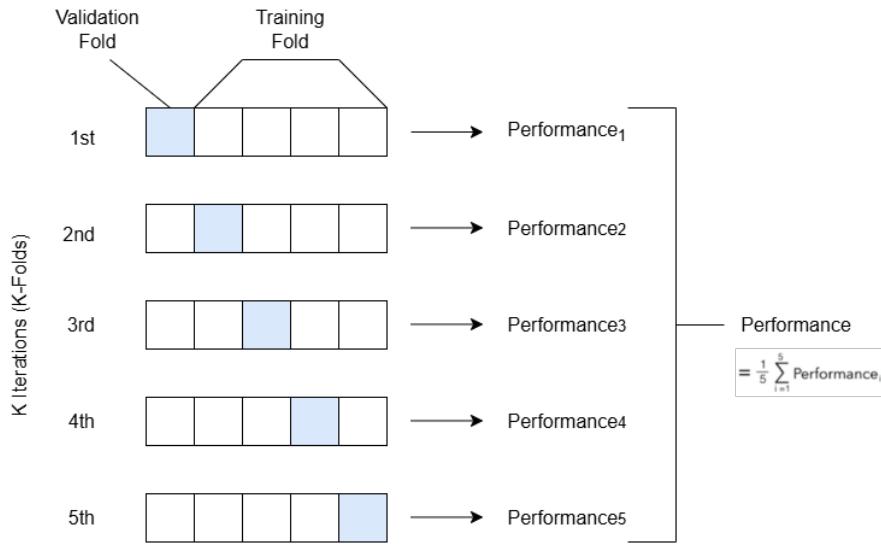


Figure 3.6: Diagram of k-fold cross-validation with $k = 5$.

907 *Image Preprocessing*

908 This subsection details the image processing techniques applied to raw shell images
 909 of *T. granosa* using computer vision methods before training the deep learning
 910 model. The image preprocessing techniques include standardizing input dimen-
 911 sions and removing shadows, background, and noise. Each image underwent data
 912 augmentation to enhance feature visibility for effective learning. Image prepro-
 913 cessing ensures consistent and high-quality input data for model training.

914 *Adjusting Dimensions*

915 All images were resized to a consistent dimension of 256x256 pixels to ensure
 916 uniformity throughout the dataset. This standardization is essential for Convo-
 917 lutional Neural Networks (CNNs), as a consistent input dimension is required.
 918 While resizing, the aspect ratio was maintained to prevent distortion of the mor-
 919 phological features, and padding was added to retain the original format.

920 *Background Removal*

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

926 *Shadow Removal*

927 To minimize noise caused by shadows around the shell, HSV thresholding, con-
928 tours, and morphological thresholds were applied to isolate and remove shadowed
929 regions. This approach preserved the natural color of the blood cockles and elim-
930 inated shadows and noise from the surrounding area (see Figures 3.7 and 3.8).

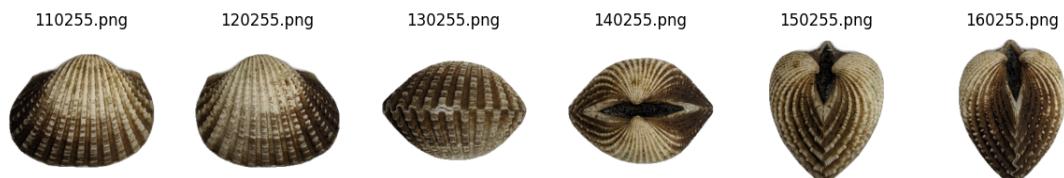


Figure 3.7: Shadows removed from male samples at different angles.

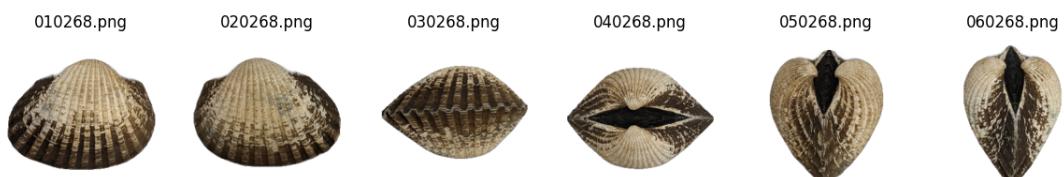


Figure 3.8: Shadows removed from female samples at different angles.

931 3.8.1 Convolutional Neural Network

932 Convolutional Neural Networks are the main deep learning tool used in image
 933 classification, specifically binary classification. CNNs leverage their ability to
 934 share weights and use pooling techniques, reducing the number of parameters (Cui,
 935 Pan, Chen, & Zou, 2020). The proposed CNN architecture for sex identification
 936 of blood cockles employs 12 layers designed to extract features from the input
 937 image with dimensions. The layers consist of four convolution layers, a flatten
 938 layer, dropout and two dense layers. The CNN framework used in this study is
 939 shown in Figure 3.9.

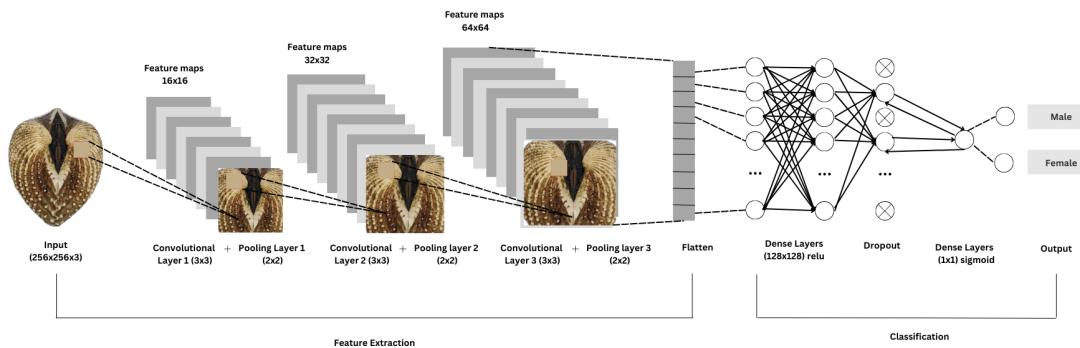


Figure 3.9: Architecture of convolutional neural network (CNN).

940 *Convolution Layer*

941 The convolution layers of CNN extract the features from the input image through
 942 the convolution operation. This study uses four convolution layers with a 3x3
 943 kernel size and filter sizes of 16, 32, 64, and 128. The first layer extracts the
 944 low-level features, such as edges, lines, and corners, while the deeper layers it-
 945 eratively extract more complex information from these low-level features. The
 946 ReLU activation function is used as the baseline for this model, and experiments
 947 are conducted with different activation functions, such as ELU and PReLU, to

948 evaluate their impact on learning complex patterns within the data.

949 ***Pooling Layer***

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

953 ***Fully Connected and Dropout***

954 Fully connected layers follow after the convolution and pooling layers. Each neu-
955 ron connects to all neurons of the previous layer. The output values from the
956 fully connected layers are sent to an output layer. It was classified using different
957 sigmoid functions appropriate for binary classification.

958 A large number of parameters in the training process can lead to overfitting. It
959 occurs when the model learns the training data too well, including its noise and
960 irrelevant details. This results in poor performance on unseen data. To mitigate
961 the overfitting, the dropout layer was employed. Dropout works by temporarily
962 discarding a portion of the neurons in the network with probability p ($0 < p < 1$).
963 During this process, these neurons do not participate in the forward propagation
964 process of CNN and the backward propagation process (Cui et al., 2020).

965 **3.8.2 CNN Training**

966 The dataset consists of 1626 samples, with 127 samples from females and 144 sam-
967 ples from males, individually for each angle. Given the minimal class imbalance,
968 random undersampling was carried out to create a balanced dataset. All images

969 were resized to 256x256 pixels and normalized using a Rescaling layer, ensuring
 970 pixel values were within the range [0, 1].

971 ***Data Splitting***

972 Due to the limited dataset size, a traditional train-test split was not adopted.
 973 Instead, a 5-fold stratified cross-validation approach was used to maximize the
 974 use of available data while preserving the class distribution within each fold (refer
 975 to Figure 3.10). `StratifiedKFold` was applied to ensure that the distribution of
 976 male and female samples remained consistent across all folds, thereby enabling
 977 fair and robust model evaluation (GeeksforGeeks, 2020).

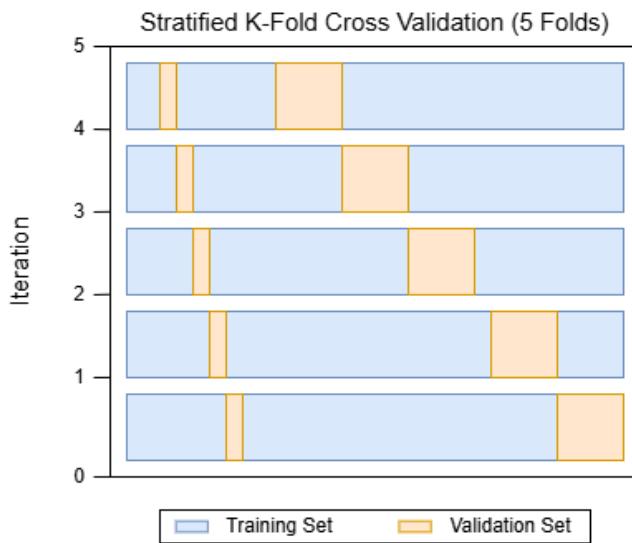


Figure 3.10: Diagram of stratified k-fold cross-validation with k=5.

978 ***Data Augmentation***

979 Before model training, online data augmentation was applied exclusively to the
 980 training data within each fold, creating new data variations on the fly. The aug-
 981 mentations included random horizontal flips, slight rotations, and zoom trans-
 982 formations to enhance data diversity and improve model generalization (Awan,

983 All augmentation was strictly applied only to the training subset of each
984 fold to prevent data leakage and maintain the validity of the results (*Figure 3.11*).

985 On-the-fly data augmentation (OnDAT) generates augmented data during each
986 iteration, exposing the model to constantly changing data variations. Augmenting
987 the original data allows better exploration of the underlying data generation pro-
988 cess and has the potential to prevent the model from overfitting spurious patterns,
989 thereby improving performance (Cerqueira, Santos, Baghoussi, & Soares, 2024).

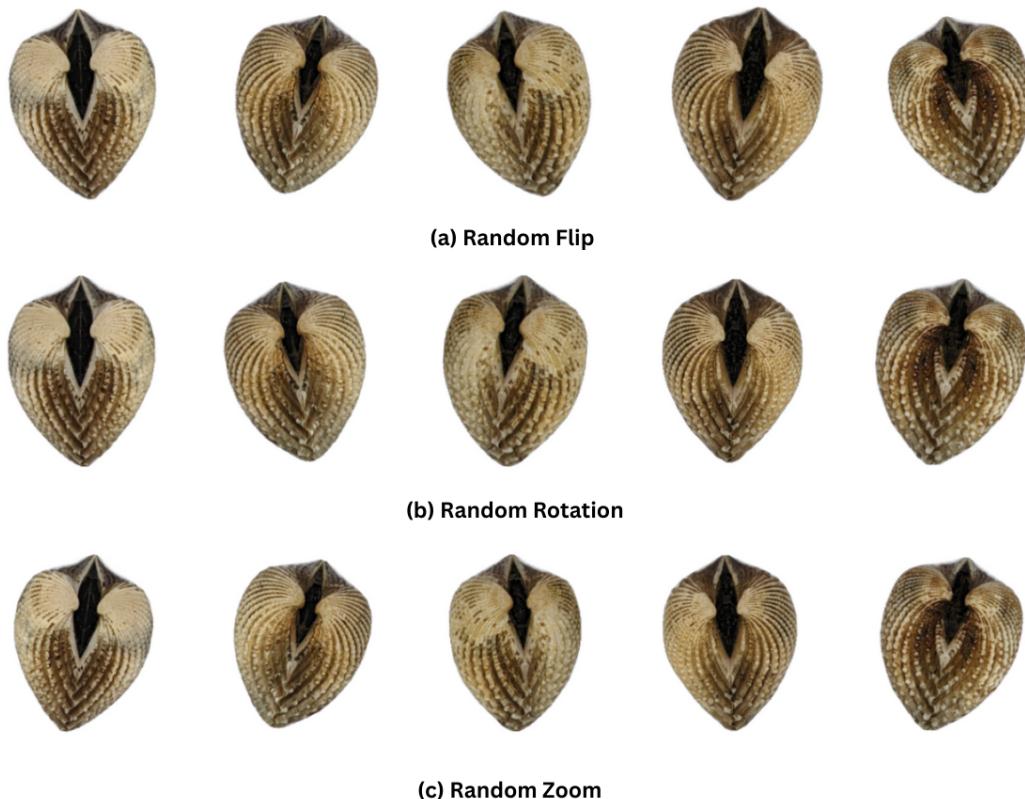


Figure 3.11: Data augmentation techniques.

990 ***Training Procedure***

991 During the training process, model performance per fold was carefully monitored.
992 One important thing to observe is the consistency in the performance, whether

the model is still learning or is at high risk of overfitting. Early stopping was applied to ensure the stable performance of the model across folds. This technique allows for monitoring the training of the neural network, stopping when the performance metrics, in this case, validation loss, cease to improve. Furthermore, to enhance the learning process, `ReduceLROnPlateau` was applied, which decreased the learning rate if there was no improvement in the model for a specified number of epochs (Team, n.d.).

The model was trained using the Adam optimization algorithm, with an initial learning rate of 0.001. Binary cross-entropy, commonly known as the log loss, was employed as the loss function due to its effectiveness in binary classification tasks. To reduce the risk of overfitting, a dropout rate of 0.5 was applied, randomly deactivating half of the neurons during the training process to improve generalization.

3.9 Evaluation Metrics

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

Accuracy (ACC) is the ratio of the overall correctly predicted samples to the total number of examples in the evaluation dataset (Cui et al., 2020). It measures the overall correctness of the model in predicting both male and female blood

₁₀₁₅ cockles. This metric provides insight into how well the model performs across all
₁₀₁₆ classifications. The formula for accuracy is:

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

₁₀₁₇ Precision (PREC) is the ratio of correctly predicted positive samples to all samples
₁₀₁₈ assigned to the positive class (Cui et al., 2020). This metric helps in evaluating
₁₀₁₉ the fairness of the model and prevents the misclassification of blood cockles as it
₁₀₂₀ identifies potential inaccuracies or biases. The formula for precision is:

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

₁₀₂₁ Recall (REC), also known as sensitivity or the true positive rate (TPR), is the
₁₀₂₂ ratio of correctly predicted positive cases to all the actual positive samples (Cui
₁₀₂₃ et al., 2020). It represents the ability of the model to correctly identify positive
₁₀₂₄ male and female samples. The formula for recall is:

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

₁₀₂₅ The F1 score is the harmonic mean of precision and recall, which penalizes extreme
₁₀₂₆ values of either of the two metrics (Cui et al., 2020). It is particularly useful when
₁₀₂₇ the class distribution is imbalanced. The formula for the F1 score is:

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

1028 The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a
1029 performance measurement for classification problems, particularly used in deep
1030 learning in this study. The ROC curve is a plot of the true positive rate (recall)
1031 against the false positive rate (1 - specificity), and the AUC score quantifies the
1032 overall ability of the model to discriminate between positive and negative classes.
1033 A higher AUC indicates better model performance. (Nahm, 2022)

¹⁰³⁴ **Chapter 4**

¹⁰³⁵ **Results and Discussions**

¹⁰³⁶ This chapter presents the results from the machine learning and deep learning
¹⁰³⁷ analyses conducted on the preprocessed dataset. It includes an evaluation of
¹⁰³⁸ various machine learning classifiers and the application of deep learning models
¹⁰³⁹ for image-based classification. The primary focus is on identifying key predictors
¹⁰⁴⁰ and assessing classification performance for sex identification in *T. granosa*.

¹⁰⁴¹ **4.1 Machine Learning Analysis**

¹⁰⁴² This chapter outlines the results of preprocessing, training of machine learning
¹⁰⁴³ models, and feature importance analysis, all conducted in Google Colab using
¹⁰⁴⁴ Python. The dataset was preprocessed in Colab, and the training and evaluation
¹⁰⁴⁵ of various classifiers were performed entirely within this environment. This part of
¹⁰⁴⁶ the paper includes five subsections: data exploration, statistical analysis, feature
¹⁰⁴⁷ importance analysis, performance evaluation, and confusion matrix analysis.

1048 **4.1.1 Data Exploration**

1049 Exploratory data analysis was performed to characterize the dataset using visu-
1050 alizations to understand the patterns and correlations within the data. A corre-
1051 lation heatmap was created to assess the relationship between the predictors and
1052 the target variable.

1053 The heatmap (see Figure 4.1) revealed three features most correlated with the
1054 sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos-length ratio (r
1055 $= 0.12$), and the distance between the umbos ($r = 0.12$). Each of these features
1056 demonstrated a weak positive relationship with the target variable.

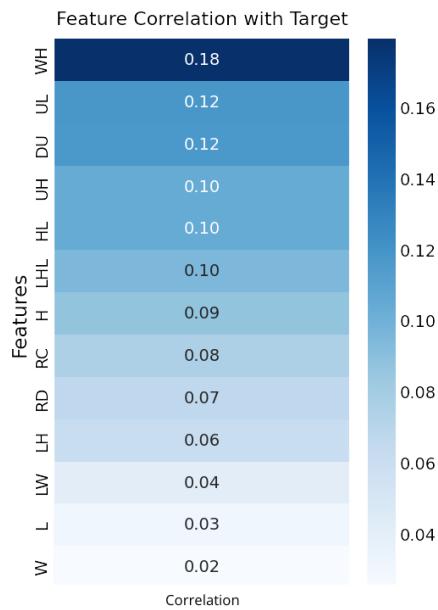


Figure 4.1: Heatmap of morphometric correlations with *T. granosa* sex.

¹⁰⁵⁷ **4.1.2 Statistical Analysis**

¹⁰⁵⁸ As part of the exploratory data analysis, statistical testing confirmed that the
¹⁰⁵⁹ dataset did not follow a normal distribution (*see Table 4.1*). Consequently, the
¹⁰⁶⁰ Mann-Whitney U test was applied with a significance level of $\alpha = 0.05$ to com-
¹⁰⁶¹ pare male and female samples. Out of thirteen features, five showed statistically
¹⁰⁶² significant differences. These included: distance between umbos ($p = 0.025$),
¹⁰⁶³ length-width ratio ($p = 0.011$), umbos-length ratio ($p = 0.019$), width-height
¹⁰⁶⁴ ratio ($p = 0.003$), and umbos-height ratio ($p = 0.036$).

¹⁰⁶⁵ It is important to note that statistical significance does not imply predictive im-
¹⁰⁶⁶ portance. Therefore, further analysis, such as feature importance evaluation, was
¹⁰⁶⁷ performed to identify the most informative predictors for classification.

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

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

1068 4.1.3 Feature Importance Analysis

1069 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric
 1070 method that is suitable for evaluating differences in distributions across groups
 1071 when the data does not follow a normal distribution. This approach was chosen
 1072 because of the non-normality of the dataset and its robustness in handling con-
 1073 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,
 1074 2024)

1075 The analysis showed that the width-to-height ratio (WH ratio) had the high-
 1076 est importance score, indicating it is the most statistically significant feature for
 1077 distinguishing the sex of *T. granosa*. Other notable features included the length-
 1078 to-width ratio (LW ratio), umbo distance-to-length ratio (UL ratio), distance
 1079 between the umbos, and umbo distance-to-height ratio (UH ratio), all of which
 1080 contributed significantly to the classification task (refer to Figure 4.2).

1081 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

1082 Table 4.2 shows the performance metrics of different machine learning models
 1083 trained using all 13 features from the dataset. Among the models, Gradient

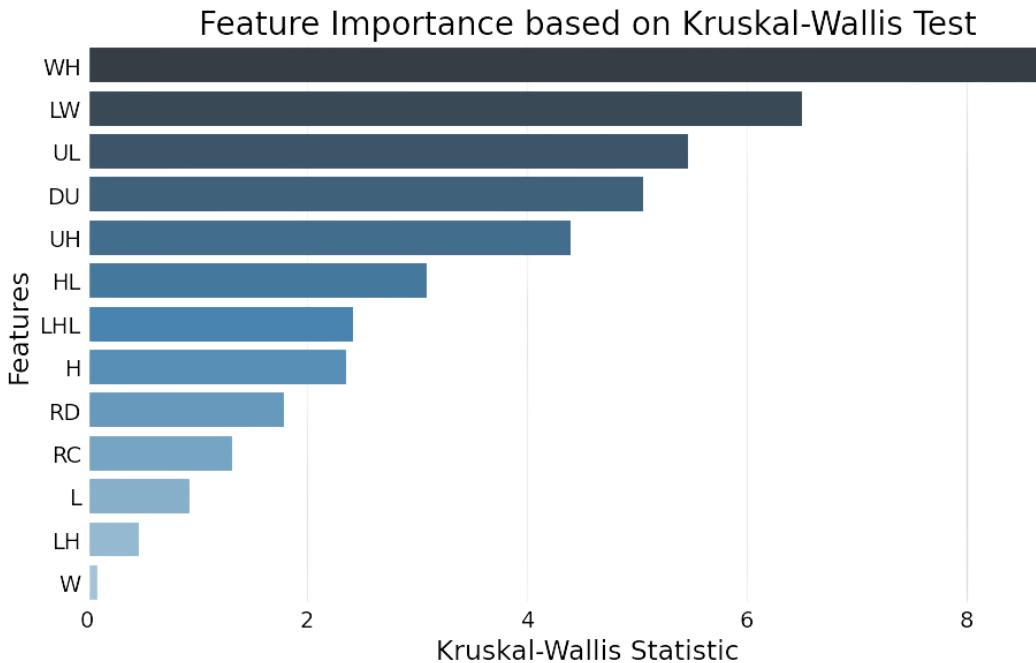


Figure 4.2: Feature importance scores using the Kruskal-Wallis test.

1084 Boosting achieved the highest accuracy of 61.03%, along with strong precision,
 1085 recall, and F1-score values. AdaBoost also performed competitively, with an ac-
 1086 curacy of 60.63%. These results highlight the effectiveness of ensemble methods
 1087 such as Gradient Boosting and AdaBoost when utilizing the full feature set, likely
 1088 because of their capability to combine multiple weak learners into a more robust
 1089 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

1090 Table 4.3 presents the performance of the same models using only the top five fea-

tures identified through Kruskal-Wallis feature importance analysis. The selected features are the distance between the umbos, length-to-width ratio, width-to-height ratio, umbo distance-to-height ratio, and umbo distance-to-length ratio.

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

4.1.5 Confusion Matrix Analysis

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

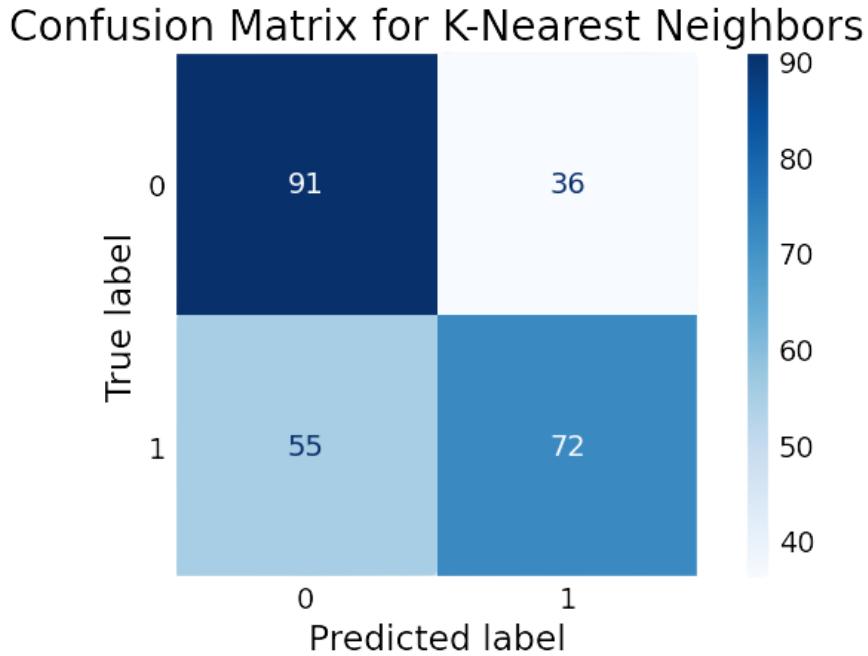


Figure 4.3: KNN confusion matrix for *T. granosa* sex classification.

4.2 Deep Learning Analysis

This section presents the performance of the Convolutional Neural Network (CNN) model in classifying the sex of *T. granosa* based on shell morphology. The analysis evaluates the model's ability to distinguish between male and female shell images using various evaluation metrics. This part of the paper includes six subsections: baseline model, comparison of individual and combined angles, training result and hyperparameter tuning, proposed model, learning rates and training behavior per fold, and visualizations.

The machine learning analysis (see Figure 4.3) revealed that five of the original features produced significant results. The K-Nearest Neighbor (KNN) model achieved an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1 score of 63.75%. This section compares the model's performance across differ-

₁₁₂₃ ent angles based on the results of the machine learning and feature importance
₁₁₂₄ analysis.

₁₁₂₅ 4.2.1 Baseline Model

₁₁₂₆ This section presents the baseline model with a batch size of 16 and 20 epochs,
₁₁₂₇ which will serve as the starting point for comparison and provide a guideline for
₁₁₂₈ hyperparameter tuning. The focus will be on one of the angles, specifically the
₁₁₂₉ Left Lateral view, since the feature importance analysis using the Kruskal-Wallis
₁₁₃₀ Test indicated that the width-to-height ratio had the highest importance score,
₁₁₃₁ which is most visible from the Left Lateral view.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC score (%)	Loss (%)
Unbalanced	65.27	71.82	58.99	63.99	73.08	0.6122
Balanced	67.34	69.43	64.06	65.60	74.31	0.5981

Table 4.4: Performance Metrics for Unbalanced vs. Balanced Datasets (Batch Size: 16, Epochs: 20)

₁₁₃₂ The unbalanced dataset, which consisted of 144 male samples and 127 female
₁₁₃₃ samples, achieved an accuracy of 65.27%, precision of 71.82%, recall of 58.99%,
₁₁₃₄ an F1-score of 63.99%, an AUC score of 73.08%, and a loss of 0.6122. However, to
₁₁₃₅ address the class imbalance and enhance model performance, random undersam-
₁₁₃₆ pling was performed. This approach resulted in improved performance metrics for
₁₁₃₇ the balanced dataset, with an accuracy of 67.34%, precision of 69.43%, a recall
₁₁₃₈ of 64.06%, an F1-score of 65.60%, an AUC score of 74.31%, and a lower loss of
₁₁₃₉ 0.5981.

1140 **4.2.2 Comparison of Individual and Combined Angles**

1141 Using the same batch size and number of epochs, performance was compared
1142 across all individual angles and the combination of the two highest-performing
1143 angles based on accuracy, using a balanced dataset. For the combined analysis,
1144 samples from the two selected angles were placed side by side, and a new dataset
1145 folder was created for male and female samples.

Angle	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC score (%)	Loss (%)
Dorsal	66.54	63.76	77.88	69.96	73.09	0.6152
Ventral	67.30	69.33	66.18	66.53	74.87	0.6159
Anterior	51.57	31.11	6.31	10.02	65.87	0.6825
Posterior	61.43	63.48	51.17	54.25	70.12	0.6257
Left Lateral	67.34	69.43	64.06	65.60	74.31	0.5981
Right Lateral	65.37	67.18	59.82	62.99	71.02	0.6115
Ventral + Left Lateral	62.60	67.02	57.85	58.57	70.37	0.6433

Table 4.5: Performance Metrics for Individual and Combined Angles (Batch Size: 16, Epochs: 20)

1146 Table 4.5 presents the performance metrics for each individual angle and the com-
1147 bination of the two highest-performing angles in terms of accuracy. The Left Lat-
1148 eral view achieved the highest accuracy (67.34%) and precision (69.43%), while the
1149 Dorsal view obtained the highest recall (77.88%) and F1-score (69.96%). Mean-
1150 while, the Ventral view recorded the highest AUC score (74.87%), indicating its
1151 strong ability to distinguish between classes. Combining the Ventral and Left
1152 Lateral views resulted in an overall accuracy of 62.60%, suggesting that while
1153 combined images may provide complementary information, individual angle views
1154 still outperformed the combined views under the current experimental setup.

1155 4.2.3 Training Result and Hyperparameter Tuning

1156 The Left Lateral angle was selected for further optimization. Several experiments
 1157 were conducted by tuning hyperparameters such as batch size, number of epochs,
 1158 and activation functions. Each adjustment was compared against the baseline
 1159 model to enhance performance and develop a robust CNN for sex classification of
 1160 *T. granosa*.

1161 The Left Lateral angle was chosen because it achieved the highest accuracy and
 1162 precision among all individual views, and because the Kruskal-Wallis feature im-
 1163 portance analysis indicated that the width-to-height ratio, a feature most visible
 1164 from the lateral perspective, was the most significant morphological trait for clas-
 1165 sification. Therefore, focusing on this view was expected to maximize the model's
 1166 learning capacity and improve classification performance.

1167 A. Batch Size and Number of Epochs

Batch Size	No. of Epoch	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC score (%)	Loss (%)
16	20	67.34	69.43	64.06	65.60	74.31	0.5981
16	30	67.73	70.17	64.06	65.72	75.76	0.5900
16	50	67.73	70.17	64.06	65.72	75.76	0.5900
32	20	68.13	72.25	58.95	62.34	74.76	0.6041
32	30	71.28	73.17	66.89	68.27	76.76	0.5832
32	50	71.68	72.52	69.29	69.12	77.34	0.5824
64	20	56.71	65.96	36.83	41.46	71.28	0.6692
64	30	57.95	61.94	48.12	52.66	71.22	0.6241
64	50	61.10	62.68	56.12	56.83	73.46	0.6086

Table 4.6: Effect of Batch Size and Epoch Values on CNN Model Performance

1168 Table 4.6 shows the results indicating that a batch size of 32 with 50 epochs
 1169 achieved the best overall performance, with an accuracy of 71.68%, a precision of
 1170 72.52%, a recall of 69.29%, an F1-score of 69.12%, and AUC score of 77.34%.

1171 In contrast, increasing the batch size to 64 resulted in lower recall and F1-scores,

1172 suggesting that smaller batch Sizes (16 or 32) are more effective for this dataset.
1173 A moderate batch size of 32 allowed the model to generalize better and maintain
1174 stable learning, while too large batch sizes may have led to underfitting.

1175 **B. Activation Functions**

Activation Functions	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC score (%)	Loss (%)
ReLU	71.68	72.52	69.29	69.12	77.34	0.5824
ELU	53.14	32.91	53.08	39.95	58.23	0.6796
PreLU	62.64	66.59	50.43	56.96	72.33	0.6162

Table 4.7: Performance Metrics for Different Activation Functions (Batch Size: 32, Epochs: 50)

1176 Table 4.7 the performance of different activation functions applied to the CNN
1177 model trained with a batch size of 32 and 50 epochs. Based on the results, the
1178 ReLU activation function achieved the best overall performance, with an accu-
1179 racy of 71.68%, precision of 72.52%, recall of 69.29%, F1-score of 69.12%, and
1180 AUC score of 77.34%, along with the lowest loss at 0.5824. This suggests that
1181 ReLU remains an effective activation function for the classification of *T. granosa*,
1182 outperforming both ELU and PReLU in this setup.

1183 **4.2.4 Proposed Model**

1184 This section presents the performance evaluation of the proposed Convolutional
1185 Neural Network (CNN) model, trained with a batch size of 32, 50 epochs, and us-
1186 ing the ReLU activation function. The model's effectiveness was assessed through
1187 5-fold cross-validation to ensure robustness and generalizability across different
1188 data partitions.

Fold no.	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC score (%)	Loss (%)
Fold 1	76.47	70.59	92.31	80.00	73.08	0.5975
Fold 2	62.75	70.59	46.15	55.81	71.85	0.6202
Fold 3	78.43	75.00	84.00	79.25	84.92	0.5392
Fold 4	62.75	71.43	40.00	51.28	71.08	0.6331
Fold 5	78.00	75.00	84.00	79.25	85.76	0.5219

Table 4.8: Per-Fold Performance Metrics (Batch Size: 32, Epochs: 50, Activation Function: ReLU)

1189 The proposed model consistently achieved high performance in Folds 1, 3, and
 1190 5, with accuracies above 76% and strong recall and AUC scores, demonstrating
 1191 its potential for reliable sex identification of *T. granosa*. The slight variation
 1192 in performance across folds may be attributed to differences in data distribution,
 1193 emphasizing the importance of further data augmentation and balancing for future
 1194 work.

1195 4.2.5 Learning Rates and Training Behavior per Fold

1196 This section presents the learning rate adjustments, early stopping events, and
 1197 best epoch selections for each fold during the 5-fold cross-validation of the pro-
 1198 posed model. During training, the ReduceLROnPlateau callback was employed
 1199 to monitor the validation loss and automatically reduce the learning rate when
 1200 performance plateaued. Additionally, EarlyStopping was utilized to halt training
 1201 once no further improvement was observed after a set patience, and the model
 1202 weights were restored from the end of the best-performing epoch to ensure optimal
 1203 performance.

1204 The following table summarizes the epochs where learning rate reductions oc-
 1205 curred, the adjusted learning rates, the epochs at which early stopping took place,

1206 and the best epochs from which model weights were restored for each fold.

Fold no.	Epoch (LR Reduced)	Learning Rate After Reduction	Early Stopping Epoch	Best Epoch (Restored)
Fold 1	20	0.0005000	25	17
	23	0.0002500		
Fold 2	9	0.0005000	19	11
	14	0.0002500		
	17	0.0001250		
Fold 3	15	0.0005000	20	12
	18	0.0002500		
Fold 4	12	0.0005000	32	24
	15	0.0002500		
	27	0.0001250		
	30	0.0000625		
Fold 5	20	0.0005000	25	17
	23	0.0002500		

Table 4.9: Learning Rate Reductions, Early Stopping, and Best Epochs per Fold During 5-Fold Cross-Validation

1207 4.2.6 Visualizations

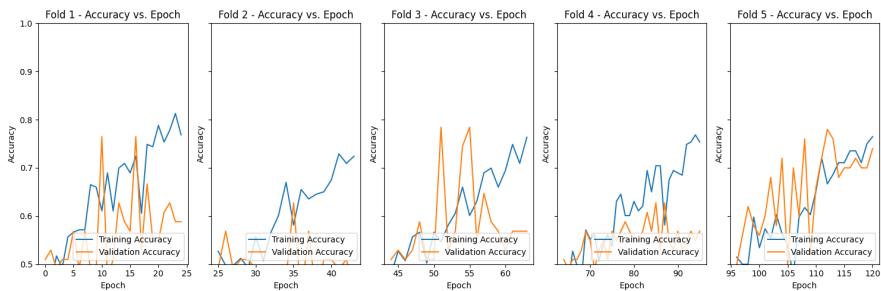


Figure 4.4: Training and validation accuracy per fold.

1208 Figure 4.4 shows the performance of the model in the training and validation in terms of accuracy across five folds. The graph across folds displays a consistent upward trend for the training accuracy. However, there is an observable change in 1210 the performance, particularly in Folds 1 and 2, where it shows a slight downward 1211 trend in the validation accuracy. 1212

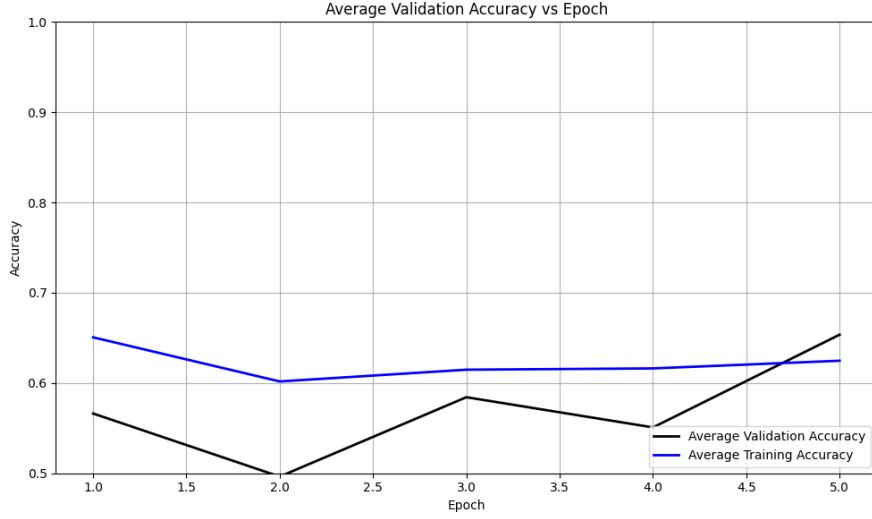


Figure 4.5: Average training and validation accuracy across folds.

1213 Figure 4.5 shows the average performance of the model in both training and accu-
 1214 racy in terms of accuracy across five folds. Similar to the individual performances,
 1215 there is an observable upward trend, which shows that the accuracy score improves
 1216 with the number of folds. The validation accuracy shows a downward and upward
 1217 trend that shows that it gradually improves on later epochs. The accuracy in
 1218 the training is slightly higher than the accuracy when validating the model, it
 1219 indicates that the model learns during training.

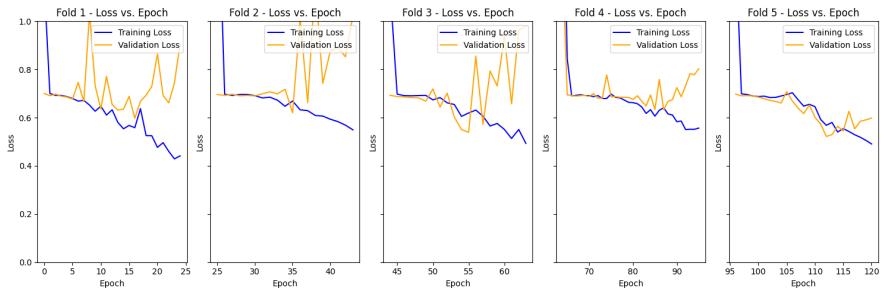


Figure 4.6: Training and validation loss per fold.

1220 Figure 4.6 shows the performance of the model in the training and validation in
 1221 terms of the training and validation loss across five folds. The graph across folds

1222 displays a consistent downward trend for the training loss. On the other hand,
 1223 there is an observable change in the performance, especially in Folds 1,2,3, and 4,
 1224 where it shows an upward trend in the validation loss. This is an implication for
 1225 the learning performance of the model, as it may not be learning effectively.

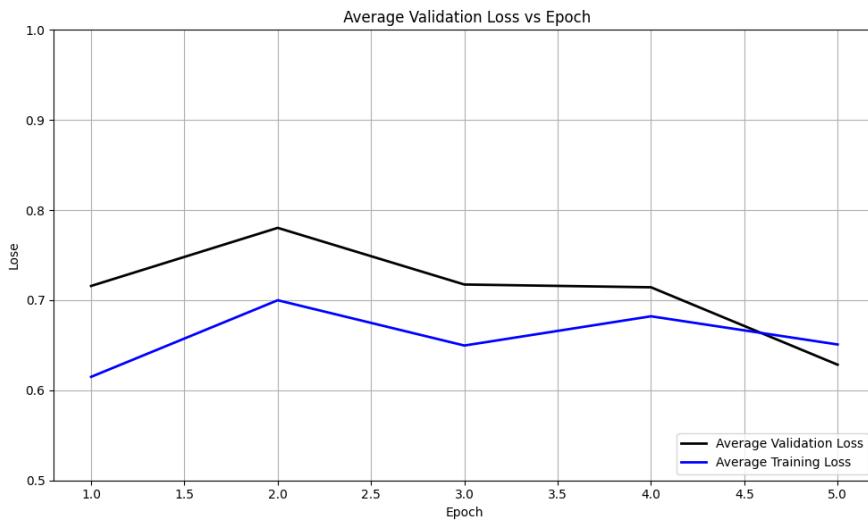


Figure 4.7: Average training and validation loss across folds.

1226 Figure 4.7 shows the average performance of the model in both the training and
 1227 validation in terms of loss across five folds. There is an observable downward trend
 1228 in both the average loss for training and validation. Additionally, the average
 1229 training loss is slightly lower than the average validation loss.

1230 Figure 4.8 shows the confusion matrix for the true class label and predicted class
 1231 label. The matrix shows the correctly predicted male and female samples along
 1232 with their corresponding percentages. There is an observable trend where females
 1233 have slightly higher true positives compared to males in the number and per-
 1234 centages for the correctly classified male and female samples, which are 94 and
 1235 88, corresponding to 74% and 69%, respectively. Additionally, the false classified
 1236 samples were 33 for females and 39 for males, respectively accounting for 26% and

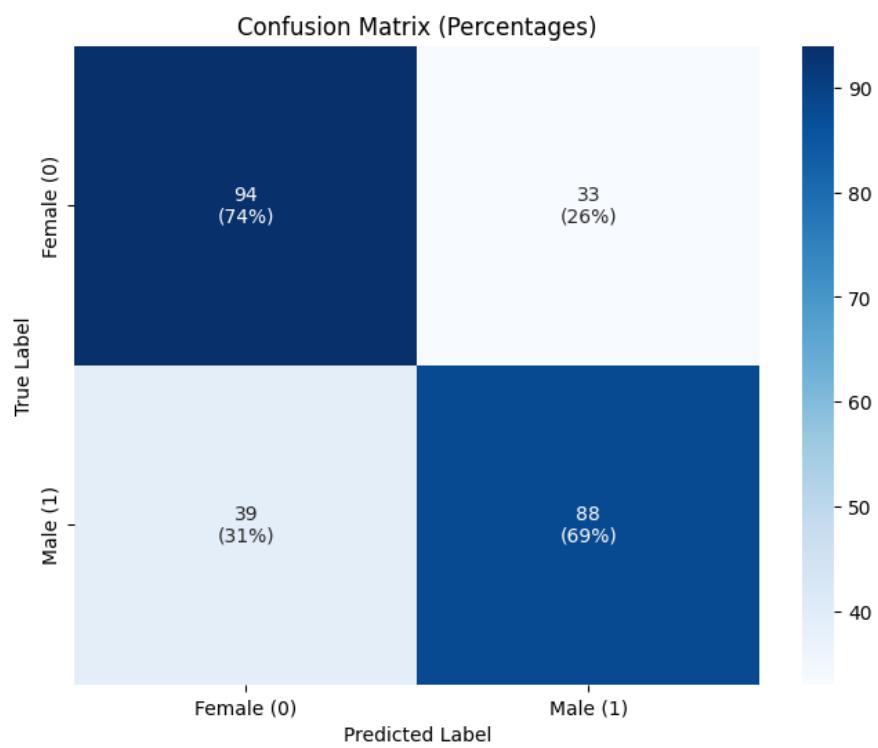


Figure 4.8: Confusion matrix for final model predictions.

₁₂₃₇ 31%.

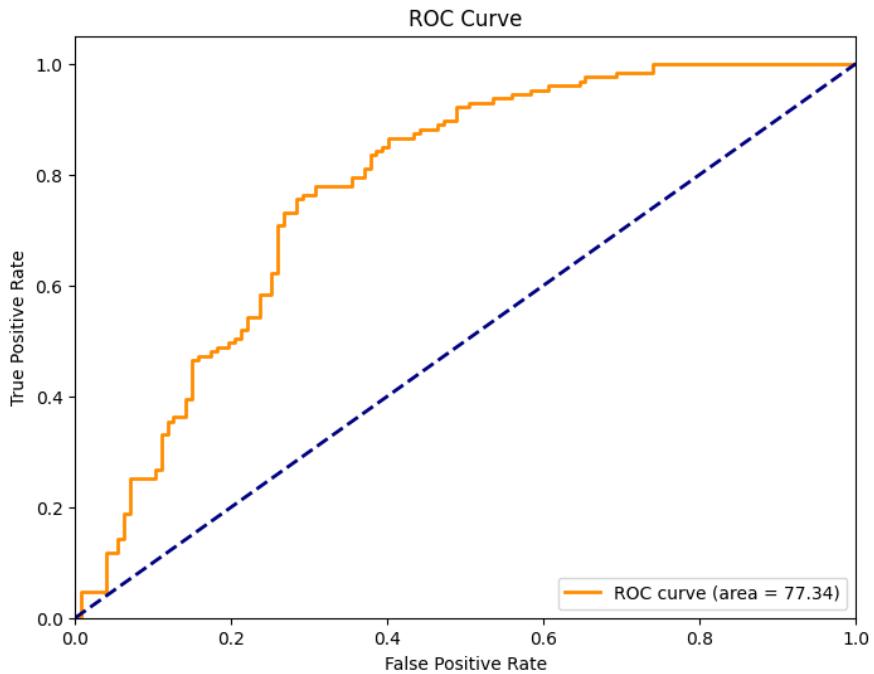


Figure 4.9: ROC curve with AUC score for the proposed model.

₁₂₃₈ Figure 4.9 shows the ROC Curve shows the ability of the proposed model to
₁₂₃₉ correctly identify the true positives, which can help determine the tradeoff between
₁₂₄₀ specificity and sensitivity. It will also determine the validity of the model, that it is
₁₂₄₁ not predicting based only on random chances. The range of AUC ROC is between
₁₂₄₂ 0.5 and 1. The model was able to achieve a score of 0.7734, which is better than
₁₂₄₃ random chances and an indication that the model is performing reasonably.

₁₂₄₄ 4.3 Discussions

₁₂₄₅ This study aimed to develop a non-invasive method for identifying the sex of *T.*
₁₂₄₆ *granosa* using machine learning, deep learning, and computer vision technologies.

1247 The dataset was manually curated by the researchers, including both the linear
1248 measurements and the images captured from six different angles.

1249 The machine learning approach revealed that using five key features, selected
1250 through statistical tests (Mann-Whitney U-test and Kruskal-Wallis test), out-
1251 performed models trained on all 13 features. The K-nearest neighbors (KNN)
1252 classifier, using only these five features, achieved an accuracy of 64.16%, precision
1253 of 64.97%, recall of 64.16%, and an F1-score of 63.57%. These results indicate
1254 that a more focused set of features can enhance model performance, confirming
1255 the potential of non-invasive sex identification using linear measurements.

1256 Further deep learning experiments explored how different image angles impacted
1257 performance. The study found that the Left Lateral view consistently produced
1258 the best results, with an accuracy of 71.68%, precision of 72.52%, recall of 69.29%,
1259 F1-score of 69.12%, and an AUC score of 77.34%. This suggests that optimiz-
1260 ing image angles is crucial, and combining multiple angles did not significantly
1261 improve the model's performance. Data augmentation and regularization tech-
1262 niques, such as early stopping, helped improve the model's generalization and
1263 prevent overfitting.

1264 The findings are significant because they demonstrate the feasibility of a non-
1265 invasive, accurate, and efficient sex identification method for *T. granosa*. This
1266 approach aligns with sustainable aquaculture practices by reducing the need for
1267 harmful physical sex-identifying methods. By integrating machine learning with
1268 deep learning image analysis, this study provides a valuable model for non-invasive
1269 sex identification which could be applied to other species in aquaculture as well.

1270 When compared to similar existing studies such as the gender classification method

1271 for Chinese mitten crab using deep learning CNN (Cui *et al.*, 2020), there are
1272 notable differences in methodology. The crab study used grayscale images and a
1273 CNN with three convolutional layers, achieving 98.90% accuracy. In contrast, this
1274 study utilized a hybrid approach combining machine learning with deep learning
1275 CNNs, trained on RGB images (256×256), and a deeper CNN architecture. De-
1276 spite achieving lower accuracy (71.68%), this variation could be due to the subtler
1277 morphological differences between male and female *T. granosa*, or possibly due to
1278 image quality limitations and sample size.

1279 There are limitations in this study, particularly the size of the dataset (271 sam-
1280 ples) and the reliance on six fixed image angles. These constraints may not fully
1281 represent the morphological variability across different populations or environ-
1282 ments. Despite these limitations, the study successfully demonstrates that com-
1283 bining machine learning and deep learning with computer vision can provide a
1284 reliable and non-invasive solution for sex identification in *T. granosa*.

₁₂₈₅ **Chapter 5**

₁₂₈₆ **Conclusion and
1287 Recommendations**

₁₂₈₈ **5.1 Conclusion**

₁₂₈₉ This study utilized the application of machine learning and deep learning tech-
₁₂₉₀ niques to identify the sex of *T. granosa* based on the morphometric characteristics.
₁₂₉₁ A manually curated dataset was developed, consisting of both linear measurements
₁₂₉₂ and images captured from six different angles. Machine learning methods were
₁₂₉₃ employed to identify statistically significant features, which served as the basis for
₁₂₉₄ deep learning analysis using a 12-layer Convolutional Neural Network (CNN). The
₁₂₉₅ proposed CNN model yielded an average accuracy of 71.68% in the performance
₁₂₉₆ metrics. Overall, this study offers a classification approach which is a viable so-
₁₂₉₇ lution for non-invasive sex identification, providing an in-depth analysis based on
₁₂₉₈ *T. granosa*'s linear measurements and morphological characteristics from different

₁₂₉₉ angles.

₁₃₀₀ Through the availability of the gathered data, trial-and-error experimentation
₁₃₀₁ was conducted by adjusting the number of layers, batch size, epoch, and activa-
₁₃₀₂ tion functions. The different combinations tested provided baseline results that
₁₃₀₃ demonstrate the feasibility of non-invasive sex identification for *T. granosa*.

₁₃₀₄ While the study has made significant progress, challenges were encountered during
₁₃₀₅ CNN training, particularly due to hardware memory limitations. To overcome
₁₃₀₆ these, the researchers utilized synchronous Google Colab with 100 computing
₁₃₀₇ units, requiring subscriptions, repeated retraining, and reconfigurations, which
₁₃₀₈ demanded considerable financial resources and time to optimize the parameters.

₁₃₀₉ Upon comparing the experimental results of model parameters, it was demon-
₁₃₁₀ strated that non-invasive sex identification on *T. granosa* is achievable through
₁₃₁₁ the integration of machine learning and deep learning methods. Machine learn-
₁₃₁₂ ing models based on five statistically selected features had better performances
₁₃₁₃ than those based on all features, with an accuracy of 64.16%, precision of 64.97%,
₁₃₁₄ recall of 64.16%, and an F1-score of 63.57% using K-nearest neighbors (KNN)
₁₃₁₅ classifier. The classification performance was further enhanced by deep learning
₁₃₁₆ models, using Left Lateral image view, achieving an accuracy of 71.68%, precision
₁₃₁₇ of 72.52%, recall of 69.29%, F1-score of 69.12%, and an AUC score of 77.34%.

₁₃₁₈ These findings establish that the CNN model can serve as a baseline for future
₁₃₁₉ studies on non-invasive sex identification of *T. granosa* and potentially other sim-
₁₃₂₀ ilar species. By providing a practical and less harmful alternative to traditional
₁₃₂₁ methods, this research contributes a significant advancement in the field of aqua-
₁₃₂₂ culture and marine biology.

1323 5.2 Recommendations

1324 This special problem entitled Morphometric and Morphological-Based Non-invasive
1325 Sex Identification of *T. granosa* focuses on creating a baseline study that will serve
1326 as a foundation for further studies involving *T. granosa*, blood cockles, using ma-
1327 chine learning, computer vision, and deep technologies in determining the sex of
1328 the samples is a salient need in aquaculture practices. Thus, the proposed rec-
1329 ommendations are the future applications to improve and have detailed analysis,
1330 such as focusing on shape analysis, exploring other state-of-the-art deep learning
1331 techniques, or transfer learning, such as ResNet, SqueezeNet, and InceptionNet,
1332 and comparing the analysis results. Furthermore, the main goal of conducting
1333 this is to have the ability to identify the sex of the samples by taking real-time
1334 angles by rotating from the dorsal, lateral, and ventral.

1335 Due to the time constraints, the researchers were only able to gather a total of
1336 1,626 images with 271 images per angle, and utilized these for model training and
1337 validation. A larger and more diverse collection of images could further improve
1338 the model's generalization. In order to capture more variability, future study
1339 might include expanding the dataset to improve classification performance.

1340 Future studies could also invest in a sturdier and more controlled environment
1341 by using a green background and positioning a fixed camera angle during image
1342 acquisition. In addition, researchers may experiment with other image processing
1343 techniques such as morphological transformations to emphasize features. The
1344 dataset can be utilized for further analysis through advanced deep learning and
1345 computer vision methods to make sense of the images gathered and discern sexual
1346 dimorphism for *T. granosa*.

¹³⁴⁷ Chapter 6

¹³⁴⁸ References

- ¹³⁴⁹ Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*, 5–16. doi: 10.1080/11250000409356545
- ¹³⁵⁰
- ¹³⁵¹
- ¹³⁵² Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. , *10*.
- ¹³⁵³
- ¹³⁵⁴
- ¹³⁵⁵ Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian Sains*, *14*, 14207.
- ¹³⁵⁶
- ¹³⁵⁷ Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021). Data scaling performance on various machine learning algorithms to identify abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi: 10.14710/jtsiskom.2021.14105
- ¹³⁵⁸
- ¹³⁵⁹
- ¹³⁶⁰
- ¹³⁶¹ Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid. *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- ¹³⁶²
- ¹³⁶³ Awan, A. A. (2022, November). *A complete guide to data augmentation*.

- 1364 *tion.* Retrieved from <https://www.datacamp.com/tutorial/complete>
- 1365 **-guide-data-augmentation**
- 1366 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquacul-
1367 ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-
1368 tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on*
1369 *mangrove-friendly aquaculture organized by the seafdec aquaculture depart-*
1370 *ment, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- 1371 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
1372 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
1373 Resources.
- 1374 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of
1375 waste cockle shell (*anadara granosa*) in biodiesel production from palm olein:
1376 Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.
1377 doi: 10.1016/j.fuel.2011.03.002
- 1378 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining*
1379 *mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127
1380 .v1
- 1381 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
1382 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A
1383 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1384 *Molecular Biology and Evolution*, *28*(5), 1645–1659. doi: 10.1093/molbev/
1385 msq345
- 1386 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1387 proaches, challenges and opportunities for aquaculture. *Journal of Marine*
1388 *Science and Engineering*, *3*(2), 329–355. doi: 10.3390/jmse3020329
- 1389 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).

- 1390 Mass mortalities in bivalve populations: A review of the edible cockle
1391 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1392 10.1016/j.ecss.2014.04.011
- 1393 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1394 research in bivalves: Towards the identification of molecular markers of
1395 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1396 10.1016/j.jprot.2012.04.027
- 1397 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data
1398 augmentation for forecasting with deep learning. *ArXiv (Cornell University)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1400 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly
1401 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1402 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1403 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1404 10.1093/icb/ict076
- 1405 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1406 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma*
1407 *edule*) geographical traceability based on shell image analysis. *Ecological
1408 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1409 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1410 for chinese mitten crab using deep convolutional neural network. *Multi-
1411 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1413 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle
1414 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-
1415 posure. *Aquaculture*, 56(2), 151–162. Retrieved from <https://doi.org/>

- 1416 10.1016/0044-8486(86)90024-4 doi: 10.1016/0044-8486(86)90024-4
- 1417 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1418 ageing of eel, anguilla anguilla (l.)? *International Review of Hydrobiology*,
1419 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1420 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and*
1421 *anatomy for young learners*. Rosie Research. Retrieved from <https://rosieresearch.com/clam-dissection-anatomy/>
- 1423 Fao 2024 report: Sustainable aquatic food systems important for global food
1424 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1427 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1428 elemental composition: Complementary tools for stock discrimination of
1429 mulloway (argyrosomus japonicus) in southern australia. *Fish Research*,
1430 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1431 GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*.
1432 Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1434 GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn*. Re-
1435 trieval from <https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/> (Accessed: 2025-04-23)
- 1437 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Black-
1438 well Science.
- 1439 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean*
1440 *Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1441 Hussain, S. S., & Zaidi, S. S. H. (2024). Adaboost ensemble approach with

- 1442 weak classifiers for gear fault diagnosis and prognosis in dc motors. *Applied
1443 Sciences*, 14(7), 3105. doi: 10.3390/app14073105
- 1444 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1445 surface water characterization and heavy metal health risk on cockles in
1446 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1447 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1448 Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a com-
1449 prehensive guide to data rescaling*. DataCamp. Retrieved from [https://www
1451 .datacamp.com/tutorial/normalization-in-machine-learning](https://www
1450 .datacamp.com/tutorial/normalization-in-machine-learning) (Ac-
cessed 2025-04-23)
- 1452 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The ear-
1453 liest example of sexual dimorphism in bivalves—evidence from the astar-
1454 tid *Nicaniella* (lower jurassic, southern germany). *Journal of Paleontology*,
1455 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1456 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1457 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1458 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1459 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1460 learning-based phenotype classification of three ark shells: Anadara kagoshi-
1461 mensis, tegillarca granosa, and anadara broughtonii. *Frontiers in Marine
1462 Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1463 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the
1464 granulated ark, *tegillarca granosa* (linne). *The Korean Journal of Malacol-
1465 ogy*, 13, 55–64.
- 1466 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as
1467 a tool for stock identification of the southern blue whiting, *micromesistius*

- 1468 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1469 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1470 et al. (2016). Identifying blue whiting (*micromesistius poutassou*) stock
1471 structure in the northeast atlantic by otolith shape analysis. *Canadian*
1472 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1473 cjfas-2015-0332
- 1474 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1475 *tegillarca granosa* (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.*
1476 *Arts Sci*, 4.
- 1477 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and
1478 larval rearing of the blood cockle, *tegillarca granosa* (linnaeus, 1758), in the
1479 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2
1480 .001
- 1481 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of
1482 local populations of the common sole *solea solea* (pisces, soleidae) in the nw
1483 mediterranean through otolith morphometrics and shape analysis. *Marine*
1484 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1485 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and
1486 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–
1487 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/
1488 kja.21209
- 1489 Narasimham, K. A. (1988). Taxonomy of the blood clams *anadara* (*tegillarca*)
1490 *granosa* (linnaeus, 1758) and a. (t.) *rhombea* (born, 1780).
- 1491 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1492 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1493 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500

- 1494 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1495 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1496 doi: 10.1002/jmor.10439
- 1497 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1498 morphometrics and machine learning challenge currently accepted species
1499 limits of the land snail *placostylus* (pulmonata: Bothriembryontidae) on the
1500 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1501 doi: 10.1093/mollus/eyz031
- 1502 Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test.*
1503 Retrieved from https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/ (Accessed: 2025-04-23)
- 1505 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,
1506 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1507 carbons from consumption of blood cockle and exposure to contaminated
1508 sediments and water along the klang strait, malaysia. *Marine Pollution
1509 Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1510 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1511 of seasonal variation and anthropogenic stress on blood cockle (*tegillarca*
1512 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-*
1513 *pogenic Stress on Blood Cockle (Tegillarca Granosa) Production Potential*,
1514 44(2), 62–82.
- 1515 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1516 chine learning and its applications to biology. *PLoS Computational Biology*,
1517 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1518 Team, K. (n.d.). *Keras documentation: Reducelronplateau*. Retrieved from
1519 https://keras.io/api/callbacks/reduce_lr_on_plateau/

- 1520 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1521 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1522 & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).
1523 College Park, MD: Maryland Sea Grant.
- 1524 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1525 approach for morphological feature extraction based on variational auto-
1526 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*
1527 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1528 Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in*
1529 *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in_3366.pdf
- 1530 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1531 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972
1532 .5008949
- 1533 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1534 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1535 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-
1536 pacts of four commonly used nanoparticles on the metabolism of a ma-
1537 rine bivalve species, tegillarca granosa. *Chemosphere*, 296, 134079. doi:
1538 10.1016/j.chemosphere.2022.134079
- 1539 Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1540 blood clam tegillarca granosa. *Marine Environmental Research*, 177. Re-
1541 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

1542 Appendix A

1543 Code Snippets

¹⁵⁴⁴ **Appendix B**

¹⁵⁴⁵ **Resource Persons**

¹⁵⁴⁶ **Dr. Firstname1 Lastname1**

¹⁵⁴⁷ Role1

¹⁵⁴⁸ Affiliation1

¹⁵⁴⁹ emailaddr@domain.com

¹⁵⁵⁰ **Mr. Firstname2 Lastname2**

¹⁵⁵¹ Role2

¹⁵⁵² Affiliation2

¹⁵⁵³ emailaddr2@domain.com

¹⁵⁵⁴ **Ms. Firstname3 Lastname3**

¹⁵⁵⁵ Role3

¹⁵⁵⁶ Affiliation3

¹⁵⁵⁷ emailaddr3@domain.net

¹⁵⁵⁸

₁₅₅₉ **Appendix C**

₁₅₆₀ **Data Gathering Documentation**



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



Figure C.2: Sex-Based Separation of *Tegillarca granosa* Samples Post-Spawning



Figure C.3: Sex Identified Female Through Dissection of *Tegillarca granosa*

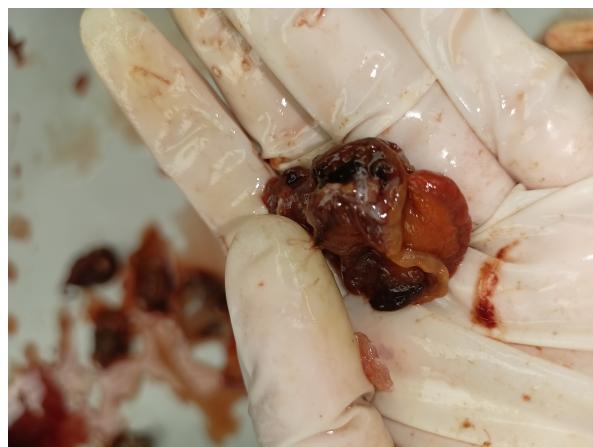


Figure C.4: Sex Identified Male Through Dissection of *Tegillarca granosa*

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 C.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 C.6: Linear Measurements of Male *Tegillarca granosa*

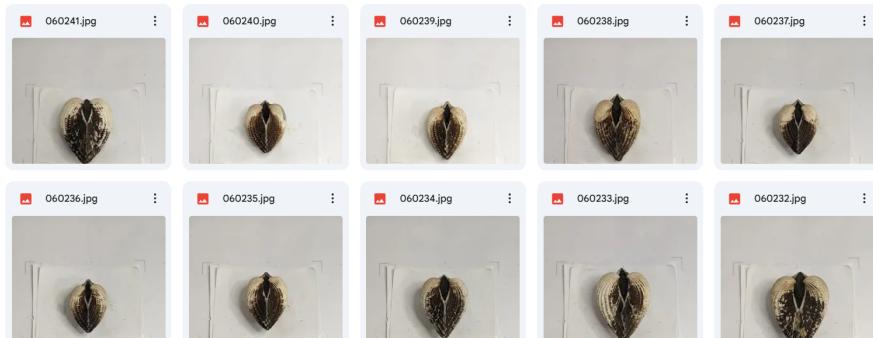


Figure C.7: Captured Images of Female *Tegillarca granosa*

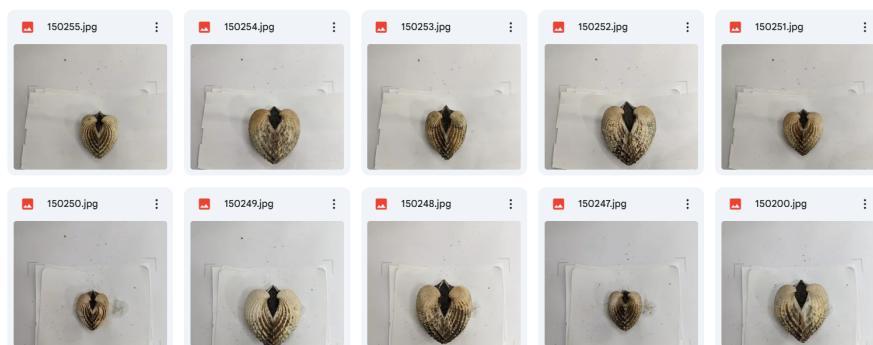


Figure C.8: Captured Images of Male *Tegillarca granosa*