

1 MORPHOMETRIC AND MORPHOLOGICAL-BASED
2 NON-INVASIVE SEX IDENTIFICATION OF BLOOD
3 COCKLE, *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
COCKLE, *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

40

To our family, advisers, and the people of science:

41

A heart full of love,

42

To those who gave wings so we can fly.

43

Stood firm even through moments of doubt.

44

A jovial harmony and warmth that kept us steadfast.

45

A word of thanks is an understatement,

46

To those who cast their light upon our way.

47

A source of wisdom even when the road grew heavy,

48

A north star that guided us through this journey.

49

Immeasurable esteem we offer,

50

To the unsung heroes of science and innovation,

51

Whose drive and dedication uplift and inspire,

52

Changing lives with boundless determination.

53

Acknowledgment

54 **I. General Acknowledgment**

55

56 The researchers would like to extend their heartfelt gratitude to the significant
57 people who were part of this journey. These people extended their expertise,
58 support, and guidance, which made this whole process bearable and fulfilling.

59 First and foremost, to the Almighty God who gives us the strength to never
60 lose hope throughout the challenging moments in our journey; for the wisdom
61 and perseverance He bestowed upon us to accomplish this paper. All for His
62 unconditional love, mercy, and grace. All glory and praise to Him.

63 To our research adviser who provided guidance and support to the researchers,
64 and for the help when things were hard. Dr. Francis Dimzon – thank you for
65 guiding us throughout this journey. We would like to thank you for assuring
66 us, especially when we were about to turn around halfway. We would like to
67 extend our greatest gratitude for the consistent supervision by providing us with
68 valuable feedback and insightful comments that aided us in our methodology and
69 in completing this paper.

70 To our co-adviser who was willing to tour us around the hatchery facility and
71 helped us come up with this study. Thank you, Dr. Victor Marco Emmanuel
72 Ferriols, for sharing your knowledge and wisdom with us researchers. Thank you
73 for answering our questions and extending your time with us. Additionally, the
74 most important part of the research is the samples — the blood cockle samples
75 — thank you for providing them willingly. You served as an inspiration for us to
76 apply solutions to real-world problems, especially in aquaculture.

77 To the research associates and hatchery staff of the UPV Hatchery, Ms. Allena
78 Arteta, Ms. LC May Gasit, Ms. Shiela Untalan, and the hatchery staff, Mr. Paul
79 Andre Lopez and Mr. Joel M. Fabrigas, for assisting us in handling the blood
80 cockles, teaching us how to identify their sex, spawning, and dissection. A big
81 thanks to Ms. Allena Arteta for letting us borrow the camera stand that we used
82 in the entire data gathering process. Your constant warmth and assistance went
83 beyond learning the basics; instead, we learned a whole lot more, especially the
84 importance of creating solutions for aquaculture practices.

85 Lastly, to everyone who directly or indirectly contributed to this thesis, whether
86 through words of encouragement, technical advice, or simply by believing in us,
87 we offer our sincerest gratitude. This study would also not have been possible
88 without you.

89 **i. Adricula's Acknowledgment**

90
91 I would like to express my deepest gratitude to my beloved family who sup-
92 ported me throughout my academic journey. My Mama Madelene and Papa Juvy,
93 who have made countless sacrifices to make me the person I am today. To my
94 sister, Aianna, and cousins Jayson, Joana, and Joel, who have helped me unwind
95 after long and stressful days of academic load. To my aunt, Tita Lybel – thank
96 you for believing in me since day one. You are always there to guide me in every
97 decision I make. And to my Lola Lydia, who has been my inspiration in finish-
98 ing my studies. Your reminders to eat more and stay safe have always kept me
99 grounded and cared for, and I appreciate the extra allowance you always give me
100 whenever I go back to Miagao. Without such a supportive family behind me, I

101 couldn't have done this without you.

102 To my pretty college friends, Arianne, Gliezel, Karielle, Kane, Kzlyr, and
103 Sharah, thank you for making my college experience bearable. The laughter and
104 joy that we shared together lessened the stress I felt. Here's to all the rants,
105 tsismis, and bangs memories we had — look how far we've come. I will forever
106 cherish the memories we made together through ups and downs. Thank you for
107 the encouragement along the way and just for being there by my side.

108 To my co-researchers, Gliezel and Kane, this would not have been possible
109 without you. Your insights and passion for this research inspired me along the
110 way. Thank you for the teamwork, initiatives, and hard work. I will miss the
111 dorm to hatchery to computer lab to dorm adventures with you!

112 To my second home, UPV Balay Gumamela, our dorm manager, and all the
113 staff — thank you for checking up on me, always making me feel welcome, and
114 ensuring that we had a comfortable stay.

115 And lastly, to the people who believed in me and have always been there for
116 me every step of the way, even when life was harsh — I am forever grateful for
117 your unwavering encouragement, support, and just for believing that I can make
118 things possible.

119 **ii. Pajarilla's Acknowledgment**

120
121 To my mother in heaven, Mama Eva, I made it through the last chapter of
122 this paper. I felt your boundless love and support every step of the way. You will
123 always be my source of strength and inspiration, even when hope seemed bleak

124 and weary. The different storms that were pushed my way — I stood because I
125 know you believed in me and wanted the best for me and my three siblings.

126 Greatest love and appreciation to my siblings – my source of boundless joy.
127 Your smiles, laughter, and warmth gave me a boost of energy that adenosine
128 triphosphate can't surpass. Gil Ivan, Gillian Gella, and my small baby brother
129 Gian Godric – no words can express how thankful I am for your existence and
130 love, as it sparked perseverance and grit in my heart. To the people who cared
131 and looked after me, thank you, Tito Rem and Auntie Maricel, for the financial
132 support and encouragement. Thank you for believing in me and the things that
133 I can achieve.

134 My heart is full of gratitude and appreciation to the Gumamela Girls Dormers.
135 I cherish all the laughter, joy, sadness, ennui, anger, fear, embarrassment, anxiety,
136 envy, and disgust. Kidding aside, I am forever thankful for the bond and friendship
137 we made, the spontaneous moments, and baybay sessions, admiring the beauty of
138 nature. Arianne, Briana, Kane, Karielle, Kzlyr, Sharah – thank you for the words
139 of encouragement, for listening, and for your companionship.

140 I would like to extend my greatest appreciation to my co-researchers, Kane and
141 Briana, who trailed the 100 steps from and to the hatchery facility. I appreciate
142 the moments and laughter that we spent together, from ideation, spending almost
143 the whole day in the computer laboratory to train our model, to crafting this
144 final paper. We went through doubts and a series of burnout moments, but your
145 presence made everything lighter.

146 My love and appreciation are offered to my high school and Senior High School
147 friends who listened to my “I can't do this anymore” moments. Thank you,

148 Jara, Francine, Nellen, Mary, and Nyle, for cheering me on and believing in my
149 potential, even when things sometimes got so dim. Thank you for reminding me
150 that there is always light at the end of the tunnel. You all are my support system,
151 cheering squad, and parents I met along the way.

152 My gratitude extends to the people inside the university who served as family.
153 To Ma'am Tess Geonanga, Ma'am Paula Khryss Ushiyama, Balay Gumamela
154 dormer manager and staff, and the UPV Virgils — thank you for the support, for
155 believing in me, and for the nutritional food that helped me brainstorm, conduct,
156 and write this study. To everyone who supported me, thank you so much, and
157 words can't express how grateful I am.

158 **iii. Vito's Acknowledgment**

159
160 To my beloved family – Mama, Papa, Mommy, Lola Jo, Uncle Mike, Tita Emie,
161 Lola Shirley, and Lola Terry – thank you for your unconditional love, patience, and
162 never-ending encouragement. Your strength and support have been the backbone
163 of everything I've achieved. Your belief in me has always kept me grounded and
164 motivated, even when I doubted myself. Thank you for the constant prayers,
165 thoughtful check-ins, and warm words that lifted me during the hardest times.
166 Your financial support has also played an instrumental role in helping me pursue
167 this journey, and I am deeply grateful.

168 To my lovely (but sometimes annoying) siblings – Hariette, Vladimir, Michaela,
169 Regina, and Daniel – words can't fully express how much you mean to me. Thank
170 you for being my constant source of joy, laughter, and sanity. Through all the
171 sibling chaos, inside jokes, and heart-to-heart conversations, you've kept me con-

172 nected to the things that truly matter. Whether it was a random meme shared
173 to make me laugh, a quick call just to check in, or simply being there in quiet
174 support, you all reminded me that no matter where I was, I was never truly alone.

175 To my ever-reliable and ever-beautiful Gumamela Girls – Briana, Gliezel, Ari-
176 anne, Sharah, Kai, and Kzlyr – thank you for the love, the chika, and the comfort.
177 You all made even the most stressful days feel manageable. You've truly been my
178 home away from home. Special mention to Briana and Gliezel, my fellow re-
179 searchers, who stood by me not only as friends but as teammates navigating this
180 academic journey together. Your dedication, the leg-torturing stairs we climbed
181 to the hatchery for data gathering, and late-night grind sessions will always be
182 one of my favorite parts of this experience.

183 To the UPV Virgils, thank you for being the silent heroes behind my survival.
184 Your monthly grocery supplies kept me fed, sane, and functioning. I probably
185 owe half of this thesis to the snacks.

186 To the Balay Gumamela dorm staff, thank you for being my guardians during
187 my stay at the dorm. Your support, care, and constant presence made my time
188 away from home more comfortable and secure. I will always be grateful for the
189 way you looked out for me and made me feel safe.

190 To my very cutesy, very demure, very mindful pet dogs, Chewy and Max
191 – thank you for the joy and companionship you've brought into my life. Max,
192 though you are no longer with us, your memory continues to live on in my heart.
193 Chewy, thank you for being my ever-present companion, sharing both the quiet,
194 loud, and the playful moments.

195 To chonky cat Mother Litob, your cuteness and presence in the hatchery dur-
196 ing our data gathering made the 100-step stairs bearable. My tiredness would
197 instantly vanish whenever I petted you.

198 This journey has been a collective effort – held up by the love, wisdom, laugh-
199 ter, and resilience of those around me. I am endlessly grateful for each of you.

Abstract

201 *Tegillarca granosa*, commonly known as blood cockles, is a significant marine bi-
202 valve species due to its nutritional value and economic importance. Accurate
203 sex identification is crucial for maintaining a balanced male-to-female ratio, sup-
204 porting sustainable harvesting, and improving resource management. However,
205 macroscopically identifying sex through shell morphology is challenging, and there
206 are currently no available technologies for non-invasive sex classification. This
207 study explores the use of machine learning and deep learning techniques to clas-
208 sify the sex of blood cockles based on shell measurements (length, width, height,
209 hinge line distance, umbo distance, and rib count) and images from various an-
210 gles (dorsal, ventral, anterior, posterior, and lateral views). The initial machine
211 learning analysis using K-Nearest Neighbor (KNN) achieved 64.16% accuracy,
212 64.97% precision, 64.16% recall, and 63.75% F1-score. In contrast, deep learning
213 with Convolutional Neural Networks (CNN) achieved 71.68% accuracy, 72.52%
214 precision, 69.29% recall, 69.12% F1-score, and 77.34% AUC score using images
215 captured from the left lateral angle. These results demonstrate the potential of
216 a non-invasive approach to sex identification, supporting sustainable aquaculture
217 practices and offering a baseline for further research using computer vision and
218 machine learning.

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

Contents

²²⁰	1 Introduction	1
²²¹	1.1 Overview	1
²²²	1.2 Problem Statement	2
²²³	1.3 Research Objectives	4
²²⁴	1.3.1 General Objective	4
²²⁵	1.3.2 Specific Objectives	4
²²⁶	1.4 Scope and Limitations of the Research	5
²²⁷	1.5 Significance of the Research	6
²²⁸		
²²⁹	2 Review of Related Literature	9
²³⁰	2.1 Background on <i>T. granosa</i> and Their Importance	10
²³¹	2.2 Sex Identification Methods in <i>T. granosa</i>	14

232	2.3 Machine Learning and Deep Learning in Biology	16
233	2.3.1 Deep Learning for Phenotype Classification in Ark Shells .	17
234	2.3.2 Geometric Morphometrics and Machine Learning for Species	
235	Delimitation	18
236	2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning	18
237	2.3.4 Machine Learning for Shape Analysis of Marine Organisms	19
238	2.3.5 Deep Learning for Landmark-Free Morphological Feature	
239	Extraction	21
240	2.3.6 Machine Learning for Sex Differentiation in Abalone . . .	21
241	2.3.7 Machine Learning for Geographical Traceability in Bivalves	23
242	2.4 Limitations on Sex Identification in <i>T. granosa</i>	24
243	2.5 Chapter Summary	26
244	3 Research Methodology	29
245	3.1 Sample Collection	30
246	3.2 Ethical Considerations	31
247	3.3 Creating <i>T. granosa</i> Dataset	32
248	3.4 Morphometric Data Collection	33
249	3.5 Image Acquisition and Data Gathering	34

CONTENTS xv

250	3.6 Hardware and Software Configuration	35
251	3.7 Machine Learning on Morphometric Data	36
252	3.7.1 Data Preprocessing	37
253	3.7.2 Machine Learning Models Training	39
254	3.8 Deep Learning for Morphological Analysis	40
255	3.8.1 Convolutional Neural Network	42
256	3.8.2 CNN Training	44
257	3.9 Evaluation Metrics	47
258	4 Results and Discussions	49
259	4.1 Machine Learning Analysis	49
260	4.1.1 Data Exploration	50
261	4.1.2 Statistical Analysis	51
262	4.1.3 Feature Importance Analysis	52
263	4.1.4 Performance Evaluation	53
264	4.1.5 Confusion Matrix Analysis	54
265	4.2 Deep Learning Analysis	55
266	4.2.1 Baseline Model	56

267	4.2.2 Comparison of Individual and Combined Angles	56
268	4.2.3 Training Result and Hyperparameter Tuning	57
269	4.2.4 Proposed Model	59
270	4.2.5 Learning Rates and Training Behavior per Fold	60
271	4.2.6 Visualizations	61
272	4.3 Discussions	65
273	5 Conclusion and Recommendations	69
274	5.1 Conclusion	69
275	5.2 Recommendations	71
276	6 References	73
277	A Code Snippets	81
278	B Resource Persons	83
279	C Data Gathering Documentation	85

²⁸⁰ List of Figures

²⁸¹	2.1 Diagram of <i>T. granosa</i> 's external anatomy.	12
²⁸²	3.1 Male and female <i>T. granosa</i> shells.	31
²⁸³	3.2 Different views of the <i>T. granosa</i> shell captured	33
²⁸⁴	3.3 Linear measurements of the <i>T. granosa</i> shell.	34
²⁸⁵	3.4 Image acquisition setup for <i>T. granosa</i> samples.	35
²⁸⁶	3.5 Data preprocessing in machine learning pipeline.	37
²⁸⁷	3.6 Diagram of k-fold cross-validation with $k = 5$	40
²⁸⁸	3.7 Shadows removed from male samples at different angles.	41
²⁸⁹	3.8 Shadows removed from female samples at different angles.	42
²⁹⁰	3.9 Architecture of convolutional neural network (CNN).	42
²⁹¹	3.10 Diagram of stratified k-fold cross-validation with $k=5$	45
²⁹²	3.11 Data augmentation techniques.	46

293	4.1 Heatmap of morphometric correlations with <i>T. granosa</i> sex.	50
294	4.2 Feature importance scores using the Kruskal-Wallis test.	52
295	4.3 KNN confusion matrix for <i>T. granosa</i> sex classification.	54
296	4.4 Training and validation accuracy per fold.	61
297	4.5 Average training and validation accuracy across folds.	62
298	4.6 Training and validation loss per fold.	62
299	4.7 Average training and validation loss across folds.	63
300	4.8 Confusion matrix for final model predictions.	64
301	4.9 ROC curve with AUC score for the proposed model.	65
302	C.1 Sex Identification Through Spawning of <i>T. granosa</i>	85
303	C.2 Sex-Based Separation of <i>T. granosa</i> Samples Post-Spawning . . .	85
304	C.3 Sex Identified Female Through Dissection of <i>T. granosa</i>	86
305	C.4 Sex Identified Male Through Dissection of <i>T. granosa</i>	86
306	C.5 Linear Measurements of Female <i>T. granosa</i>	86
307	C.6 Linear Measurements of Male <i>T. granosa</i>	87
308	C.7 Captured Images of Female <i>T. granosa</i>	87
309	C.8 Captured Images of Male <i>T. granosa</i>	87

List of Tables

311	2.1 Comparison of the methods used in bivalves studies.	27
312	3.1 Architecture of proposed convolution neural network.	44
313	4.1 Mann-Whitney U test results for sex-based feature comparison. .	51
314	4.2 Performance metrics for models with all 13 features.	53
315	4.3 Performance metrics for models with 5 features.	53
316	4.4 Performance metrics for unbalanced vs. balanced datasets (Batch	
317	Size: 16, Epochs: 20).	56
318	4.5 Performance metrics for individual and combined angles (Batch	
319	Size: 16, Epochs: 20).	57
320	4.6 Effect of batch size and epoch values on CNN model performance.	58
321	4.7 Performance metrics for different activation functions (Batch Size:	
322	32, Epochs: 50).	59

323	4.8	Per-fold performance metrics (Batch Size: 32, Epochs: 50, Activation Function: ReLU).	59
324	4.9	Learning rate reductions, early stopping, and best epochs per fold during 5-fold cross-validation.	61
325			
326			

³²⁷ Chapter 1

³²⁸ Introduction

³²⁹ 1.1 Overview

³³⁰ The Philippines is a global center of marine biodiversity and has established aqua-
³³¹ culture as a significant contributor to total fishery production (Aypa & Baconguis,
³³² 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-
³³³ ally and is the 11th largest seafood producer in the world. Aquaculture is deeply
³³⁴ integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-
³³⁵ duction of various aquatic species, including bivalves. Among these, blood cockles
³³⁶ (*Tegillarca granosa*) hold considerable economic and environmental significance,
³³⁷ making it essential to ensure sustainable production and population balance.

³³⁸ Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent
³³⁹ overharvesting and ensure sustainability. An imbalanced ratio can lead to over-
³⁴⁰ exploitation and negatively impact the population's viability. However, there is

³⁴¹ limited literature on *T. granosa* that provides a thorough understanding of its
³⁴² sex-determining mechanisms, particularly regarding sexual dimorphism based on
³⁴³ morphometric and morphological characteristics (Breton, Capt, Guerra, & Stew-
³⁴⁴ art, 2017).

³⁴⁵ Currently, sex determination methods for blood cockles are invasive, including
³⁴⁶ dissection and histological examinations, which often result in the death of the
³⁴⁷ species. While there is growing literature on sex identification in aquaculture
³⁴⁸ commodities using machine learning and deep learning, there is a notable scarcity
³⁴⁹ of research specific to *T. granosa* (Miranda & Ferriols, 2023).

³⁵⁰ This study aims to provide a detailed baseline analysis of blood cockles by lever-
³⁵¹ aging their morphometric and morphological characteristics. Sexual dimorphism
³⁵² in bivalves is often subtle and challenging to establish macroscopically (Karapunar,
³⁵³ Werner, Fürsich, & Nützel, 2021). However, by integrating machine learning and
³⁵⁴ deep learning, the study seeks to identify distinct features that may indicate sexual
³⁵⁵ dimorphism between male and female blood cockles.

³⁵⁶ 1.2 Problem Statement

³⁵⁷ Identifying the sex of *Tegillarca granosa* is important for promoting sustainable
³⁵⁸ aquaculture and biodiversity by maintaining a balanced male-to-female ratio. A
³⁵⁹ balanced ratio helps prevent overharvesting. Although sex identification is crucial
³⁶⁰ for blood cockle population management and sustainable aquaculture, there is a
³⁶¹ notable lack of research on creating non-invasive methods for determining the sex
³⁶² of *T. granosa*. Many recent studies and approaches rely on invasive methods like

³⁶³ dissection or histological analysis, which are impractical for large-scale aquaculture
³⁶⁴ operations focused on conservation.

³⁶⁵ Current methods for determining the sex of *T. granosa* are invasive and involve
³⁶⁶ dissection, which requires cutting open the shell to visually inspect the gonads
³⁶⁷ (Erica, 2018). This procedure can cause harm to the specimens and frequently
³⁶⁸ leads to their death. Another method is histological examination, where tissue
³⁶⁹ samples are analyzed under a microscope (May, Maung, Phy, & Tun, 2021). Both
³⁷⁰ approaches are labor-intensive and time-consuming, and can pose risks to popula-
³⁷¹ tion management, particularly when maintaining a balanced sex ratio for breeding
³⁷² programs is essential. Moreover, these invasive methods require specialized tech-
³⁷³ nical skills for accurate execution. Resource-limited aquaculture operations face
³⁷⁴ significant challenges in accessing the necessary laboratory equipment, such as
³⁷⁵ microscopes and staining tools, complicating the process.

³⁷⁶ A less invasive approach employed by aquaculturists involves monitor spawning
³⁷⁷ behavior, where individuals are separated and stimulated to reproduce in order
³⁷⁸ to determine their sex through the release of gametes (Miranda & Ferriols, 2023).
³⁷⁹ Although this method is indeed less invasive than dissection, it still induces stress
³⁸⁰ in blood cockles and may not be completely effective for fast identification in large
³⁸¹ populations.

³⁸² Given the limitations of both invasive and less invasive methods, there is a clear
³⁸³ need for a more advanced approach. An alternative, non-invasive method involv-
³⁸⁴ ing machine and deep learning technologies could address these issues by provid-
³⁸⁵ ing a fast, accurate, and effective solution without harming or stressing the blood
³⁸⁶ cockles.

³⁸⁷ 1.3 Research Objectives

³⁸⁸ 1.3.1 General Objective

³⁸⁹ The general objective of this study is to develop a non-invasive method for iden-
³⁹⁰ tifying the sex of *Tegillarca granosa* using machine learning and deep learning
³⁹¹ technologies. This method aims to provide accurate and streamlined sex iden-
³⁹² tification without causing harm to the specimens, thus supporting sustainable
³⁹³ aquaculture practices.

³⁹⁴ 1.3.2 Specific Objectives

³⁹⁵ To achieve the overall general objective of developing a non-invasive sex identifi-
³⁹⁶ cation of *T. granosa* using machine learning and deep learning technologies, the
³⁹⁷ following specific objectives have been established:

- ³⁹⁸ 1. to collect and organize a comprehensive dataset of *T. granosa*, which will
³⁹⁹ include linear measurements and images captured from different camera an-
⁴⁰⁰ gles that will serve as the basis for training and evaluating the machine
⁴⁰¹ learning and deep learning models,
- ⁴⁰² 2. to develop and implement machine learning and deep learning models that
⁴⁰³ can classify the sex of *T. granosa* based on the collected linear measurements
⁴⁰⁴ and images of different camera angles of the sample, and determine the best
⁴⁰⁵ performing models, and
- ⁴⁰⁶ 3. to evaluate the model performance using performance metrics such as accu-

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

409 1.4 Scope and Limitations of the Research

410 This study is conducted alongside the ongoing research by the UPV DOST-
411 PCAARRD, titled "Establishment of the Center for Mollusc Research and De-
412 velopment: Development of Spawning and Hatchery Techniques for the Blood
413 Cockle (*Anadara granosa*) for Sustainable Aquaculture." The ongoing research
414 primarily involves the rearing of *Tegillarca granosa* from spat to larvae, feeding
415 experiments, stocking density evaluations, substrate selection, and settlement rate
416 assessments.

417 In contrast, this study mainly focused on developing a non-invasive method for
418 identifying the sex of *T. granosa* using machine learning and deep learning tech-
419 nologies. The goal is to provide an accurate and efficient means of sex identifica-
420 tion without causing harm to the samples, contributing to sustainable aquaculture
421 practices.

422 The researchers worked with 271 blood cockles that had been sex-identified and
423 taken from Panay Island, specifically sourced from Zarraga Iloilo and Ivisan Capiz.
424 These samples, divided between 144 males and 127 females, were obtained through
425 induced spawning via temperature shock and dissection. Data collection was lim-
426 ited to the spawned stage among the five gonadal stages - immature, developing,
427 mature, spawning, and spent stages. The other stages were not preferable due to
428 indistinguishable gonads and their inability to undergo induced spawning (May

429 et al., 2021). Thus, the researchers only focused on the samples undergoing the
430 spawned stage.

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

440 The method developed in this study is specific to *T. granosa* and may not ap-
441 ply to other bivalve species. The model was trained exclusively for *T. granosa*
442 and morphometric and morphological features, which may not be consistent and
443 applicable across other shellfish species.

444 1.5 Significance of the Research

445 This study will give us a significant advancement in non-invasive sex identification
446 methods in *Tegillarca granosa* providing innovative solutions that could solve the
447 challenges in identifying sex and reshape sustainable approaches to aquaculture.

448 The significance of this study extends to the following:

449 *Research Institution.* The result of this study focusing on the sex-identification

450 mechanism of bivalves, specifically *T. granosa*, will provide valuable insights into
451 universities and research centers that focus on fisheries and coastal management,
452 such as the UPV Institute of Aquaculture, that aim to develop sustainable devel-
453 opment and suitable culture techniques.

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

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

463 *Future Researchers.* The result of this study would serve as the basis for studies
464 that involve sex identification in bivalves such as *T. granosa*. Some technologies
465 are yet to be explored in machine learning and deep learning technologies that
466 can lead to higher accuracy and distinguish the presence of sexual dimorphism in
467 the *T. granosa*.

⁴⁶⁸ Chapter 2

⁴⁶⁹ Review of Related Literature

⁴⁷⁰ Aquaculture is the fastest-growing industry in animal food production and has
⁴⁷¹ great potential as a sustainable solution to global food security, nutrition, and
⁴⁷² development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
⁴⁷³ *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
⁴⁷⁴ tegrated into the livelihoods of Filipinos, not only through fish cultivation but
⁴⁷⁵ also through the production of other aquatic species, including mollusks, oysters,
⁴⁷⁶ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
⁴⁷⁷ clams *Tegillarca granosa*, have economic and environmental significance. It has
⁴⁷⁸ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
⁴⁷⁹ vesting and maintain the optimal ratio to preserve the population and production
⁴⁸⁰ of the blood cockles.

⁴⁸¹ The members of the Arcidae Family, including *T. granosa* are important sources
⁴⁸² of food and livelihood. Cockle aquaculture meets rising demands, however, it
⁴⁸³ faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,

484 2023). To solve the problem, researchers exert a considerable amount of effort,
485 developing a broader understanding of bivalves, including their sex-determining
486 mechanism, due to their notable importance in terms of diversity, environmental
487 benefits, and economic and market importance (Breton et al., 2017). Despite the
488 promising idea of identifying sex, there is limited research reported in terms of
489 sexual dimorphism, making it harder to distinguish through its morphological and
490 morphometric characteristics.

491 By addressing the challenges in the sex identification of *T. granosa*, it would be
492 able to address one problem at a time. Currently, there are no recent documented
493 publications that integrate machine learning and deep learning in characterizing
494 sexual dimorphism, reducing complexity, variability in sex determination, and
495 differentiation mechanisms in bivalves, including *T. granosa* specifically.

496 2.1 Background on *T. granosa* and Their Im- 497 portance

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

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

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

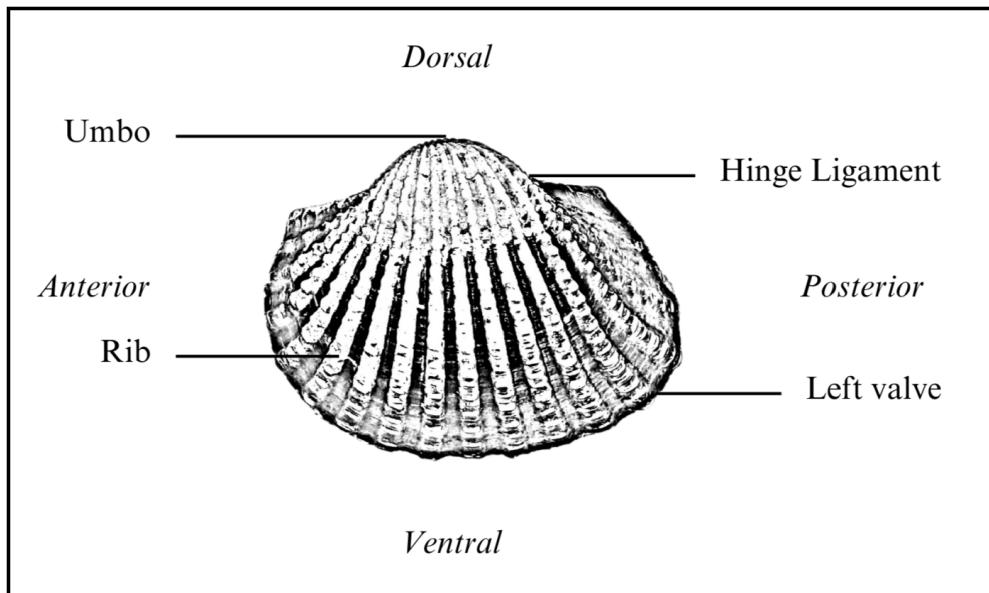


Figure 2.1: Diagram of *T. granosa*'s external anatomy.

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

539 Maniam, Hamid, & Ali, 2011).

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

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

562 2.2 Sex Identification Methods in *T. granosa*

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

567 Induced spawning and larval rearing are considered the less invasive techniques
568 used to study *T. granosa*. In the Philippines, limited research has been done
569 on the *T. granosa* (Linnaeus, 1758), and this study, titled Initial Attempts on
570 Spawning and Larval Rearing of the Blood Cockle, *T. granosa* in the Philippines,
571 was conducted by Miranda and Ferriols (2023). The researchers conducted ex-
572 periments on induced spawning and larval rearing, discovering that the eggs of
573 female *T. granosa* were salmon pink, while the sperm released by males looked
574 milky. After spawning, the researchers successfully generated 6,531,000 fertilized
575 eggs.

576 The researchers highlighted the importance of *T. granosa* and other anadarinids as
577 a food source established worldwide, especially in Malaysia and Korea. However,
578 in the Philippines, the bivalve aquaculture of the clam species is still limited. The
579 experiment, which focused on the culture and rearing of *T. granosa*, was attempted
580 by subjecting the wild broodstocks to a series of temperature fluctuations to
581 induce the spawning of gametes. This is currently the most natural and least
582 invasive method for bivalves (Aji, 2011). The study of Miranda and Ferriols
583 aimed to pave the way for the sustainable production of *T. granosa* seeds for
584 aquaculture and stock enhancement, despite the scarcity of documented hatchery
585 culture of *T. granosa* from larvae to adults in the Philippines.

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

596 The classification of the gonad stages used was by Yurimoto et al. (2014). There
597 are five maturation stages of gonadal development: immature (Stage I), devel-
598 oping (Stage II), mature (Stage III), spawning (Stage IV), and spent (Stage V)
599 stages. The sex of the *T. granosa* was confirmed by the color of the gonad and
600 by conducting a histological examination of the gonads. During the immature
601 stage, sex determination was indistinguishable due to the difficulties of observing
602 the germ cells. In the developing stage, the spermatocytes and a few spermatids
603 can be seen for males, and immature oocytes are attached to the tube wall for
604 the female. In the mature stage, the follicles are full of spermatozoa with their
605 tails pointing towards the center of the tube for the male, and the female is full
606 of mature oocytes that are irregular or polygonal in shape with the oval nucleus.
607 Upon reaching spawning, some spermatozoa are released, causing the empty space
608 in the follicle wall for males and females. There is a decrease in the number of
609 mature oocytes and it exhibits nuclear disappearance due to the breakdown of
610 the germinal vesicle. Lastly, the spent stage is where the genital tube is deformed

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

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

629 **2.3 Machine Learning and Deep Learning in Bi-** 630 **ology**

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

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

2.3.1 Deep Learning for Phenotype Classification in Ark Shells

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

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

656 **2.3.2 Geometric Morphometrics and Machine Learning for**
657 **Species Delimitation**

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

668 **2.3.3 Contour Analysis in Mollusc Shells Using Machine**
669 **Learning**

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

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

684 **2.3.4 Machine Learning for Shape Analysis of Marine Or-
685 ganisms**

686 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*
687 *to Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*
688 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to
689 the traditional method of collecting linear measurements with the application of
690 multivariate statistical methods and outline analysis in recording the structures
691 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and
692 elasmobranchs) with their hard bodies have been used as an indication of age and
693 a determinable time-scale and structure continue to go through life (Arkhipkin,
694 2005; Kerr & Campana, 2014). This study has explored variations in the mor-
695 phometry of recording structures in stock discrimination and systematics. The
696 researchers utilized the principal component analysis rather than the traditional
697 approach, which helps simplify the data without losing important information.
698 They utilized landmark-based geometric morphometrics, which has three differ-
699 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other
700 morphogenetic processes, and lastly, the extremal points are constructed land-

701 marks.

702 Generalized Procrustes Analysis (GPA) is a common superimposition technique in
703 landmark-based geometric morphometrics that aligns landmarks via translation,
704 scaling, and rotation to eliminate non-shape deviations (Zelditch, Swiderski, &
705 Sheets, 2004). However, there is a limit to the amount of smooth areas that may
706 be captured, and it is possible to overlook significant shape details. Utilization
707 of the semi-landmarks enhanced the shape description (Adams, Rohlf, & Slice,
708 2004). The researchers observed that using an outline-based approach would be
709 more effective than using a landmark-based approach.

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

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

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

732 **2.3.5 Deep Learning for Landmark-Free Morphological Fea-
733 ture Extraction**

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

744 **2.3.6 Machine Learning for Sex Differentiation in Abalone**

745 In the study, *Towards Abalone Differentiation Through Machine Learning*, re-
746 searchers identified a problem in abalone farming which is having to identify the

747 sex of abalone to apply measures for its growth or preservation. The researchers
748 classified abalone sex using machine learning. Researchers trained the machine
749 to classify different types of classes which are male, female, and immature. The
750 results demonstrated the effectiveness of utilizing linear classifiers for this task.

751 Similarly, in the study, *Data scaling performance on various machine learning*
752 *algorithms to identify abalone sex*, the researchers of the University of India (2022)
753 focused on the data scaling performance of various machine learning algorithms to
754 identify the abalone sex, specifically using min-max normalization and zero-mean
755 standardization. The different machine learning algorithms are the Supervised
756 Vector Machine (SVM), Random Forest, Naive Bayesian, and Decision Tree. Their
757 study aims to utilize machine learning in terms of identifying the trends and
758 distribution patterns in the abalone dataset. Eight features of the abalone dataset
759 (length, diameter, height, whole weight, shucked weight, viscera weight, shell
760 weight, ring) were used to determine the three sexes of Abalone. Their data has
761 been grouped based on sex which are Female, Male, and Infant. They utilized
762 the Synthetic Minority Oversampling Technique (SMOTE) in data balancing for
763 the preprocessing of the data. Followed by data scaling or normalization where
764 it converts numeric values in a data set to a general scale without distorting
765 differences in the range of values. Then they classified by splitting the data into
766 training and testing sets (Arifin, Ariawan, Rosalia, Lukman, & Tufailah, 2021).

767 The study found that Naive Bayes consistently performed better than other algo-
768 rithms. However, when applied to both min-max and zero-mean normalization,
769 the average accuracies of the algorithms were as follows: Random Forest (62.37%),
770 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-
771 nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with

772 normalization, Random Forest achieved the highest overall metrics, including an
773 average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of
774 83.31%. Liu et al. concluded that Random Forest is highly accurate because it
775 can handle large, complex datasets, run processes in parallel using multiple trees,
776 and select the most relevant features to enhance model performance (Arifin et al.,
777 2021).

778 **2.3.7 Machine Learning for Geographical Traceability in
779 Bivalves**

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

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

791 The researchers developed a non-invasive image-based traceability technique, an
792 alternative to the chemical and biochemical analysis of the bivalves. It was able
793 to incorporate machine learning methods to promote lesser human intervention.
794 The researchers discovered that BivalveNet emerged as the superior model for

795 bivalves with 96.91% accuracy which is comparable to the accuracy of the de-
796 structive methods with 97% and 97.2% accuracy rates. The result of the study
797 aided the researchers in concluding that there is a possibility of on-site evalua-
798 tion of the bivalve through the implementation of a mobile app that would allow
799 the public and official entities to obtain information regarding the provenance of
800 seafood products' traceability because of its non-invasive and image-based aspects
801 (Concepcion et al., 2023).

802 *T. granosa* is known for having no sexual dimorphism. However, through several
803 related studies, the researchers can apply how family shells of *T. granosa* have
804 been identified based on its morphological and morphometric characteristics and
805 the methods used in machine learning in identifying its sex.

806 2.4 Limitations on Sex Identification in *T. gra-* 807 *nosa*

808 To date, no distinction has been made between the male and female *T. granosa*
809 in sexing methodology. In cockle aquaculture without clearly apparent sexual
810 dimorphism, sexing can be performed using invasive methods such as chemical
811 stimulation, dissection, and gonad-stripping. Induced spawning, specifically tem-
812 perature shock, is the most natural and least invasive method for bivalves (Aji,
813 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water
814 from hot to cold with a specific temperature requires deliberate and careful ma-
815 nipulation of the temperature over a specific period and would require constant

816 management and monitoring.

817 Recent studies involved non-invasive methods, with a specific emphasis on mor-
818 phological characteristics as indicators of sex differentiation. However, Tatsuya
819 Yurimoto et al. (2014) stated that the existing methods for determining the sex of
820 bivalves and mollusks in general are somewhat limited (Afiati, 2007). At present,
821 there is no recorded evidence of sexual dimorphism in *T. granosa*. Gonochoristic
822 is the classification given to *T. granosa* (Lee, 1997). However, Lee et al. (2012)
823 reported that the sex ratio varied with shell length, suggesting that sex might
824 alter.

825 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous (syn-
826 chronous or functional) characteristics. Sequential hermaphrodites switch genders
827 after being male or female for one or multiple yearly cycles. (Heller, 1993; Gosling,
828 2004; Collin, 2013). Sex change and consecutive hermaphroditism have been ob-
829 served in different bivalve species, including Ostreidae, Pectinidae, Veneridae,
830 and Patellidae. However, macroscopically differentiating bivalve sex is challeng-
831 ing. The only way it may be identified is through histological analysis of gonad
832 remains but to do so there is an act of killing the organism (Coe, 1943; Gosling,
833 2004). Verification of sex change in bivalves to classify whether male or female
834 while they are alive is challenging since they need to be re-confirmed and re-
835 evaluated to be the same individual after a year.

836 Lee et al. (2012) found out that *T. granosa*, a species in Arcidae, has been dis-
837 covered to be a sequential hermaphrodite, with the sex ratio changing with an
838 increase in the shell size. In bivalves, sex changes usually happen when the gonad
839 is not differentiated between spawning seasons (Thompson, Newell, Kennedy, &

840 Mann, 1996). But in *T. granosa*, after the spawning season, sex changes during
841 its inactive phase. Results showed a 15.1% sex change ratio, with males having
842 a higher sex change ratio (21.2%) than females (6.2%). The 1+ year class had a
843 higher ratio (17.8%) than the 2+ year class (12.1%). Thus, this study indicates
844 that *T. granosa* is a sequential hermaphrodite. The results of the study demon-
845 strated that the bivalve's age affects the sex ratio and degree of sex change, but
846 additional in-depth investigation is required to determine the role that genetic
847 and environmental factors play in these changes.

848 No literature in the study of mollusks specifically addresses the machine learning
849 and deep learning technologies used to determine the sex of *T. granosa* bivalves in
850 various models. Nevertheless, various techniques such as shape analysis, morpho-
851 metric analysis, Wavelet, and Fourier analysis, as well as different deep learning
852 models like VGNet, ResNet, and SqueezeNet in CNN networks, are utilized for
853 phenotype classification, while different machine learning algorithms could serve
854 as the foundation for this research project.

855 2.5 Chapter Summary

856 This section of the paper summarizes the technologies used in the different studies
857 related to the pursuit of the study entitled, Morphometric and Morphological-
858 Based Non-Invasive Sex Identification of Blood Cockle, *Tegillarca granosa* (Lin-
859 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.

860 Recent developments and breakthroughs in machine learning offer promising solu-
861 tions to biological challenges. Research findings indicate that various deep learning
862 techniques — such as convolutional neural networks (CNNs), geometric morpho-
863 metrics, and other machine learning models — are effective in identifying phe-
864 notypes and determining the sex of various aquaculture species, including mol-
865 lusks and abalones. These techniques provide a foundation for developing new,
866 non-invasive methods to differentiate male and female *T. granosa*, potentially ad-
867 dressing the limitations of manual and invasive techniques. Thus, using machine
868 learning to analyze morphological and morphometric features may streamline the
869 process of sex identification.

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

⁸⁷⁵ **Chapter 3**

⁸⁷⁶ **Research Methodology**

⁸⁷⁷ This chapter discusses the materials and methods employed in the study, focus-
⁸⁷⁸ ing on the development requirements, as well as the software and programming
⁸⁷⁹ languages utilized. It also detailed the overall workflow in conducting the study,
⁸⁸⁰ Morphometric and Morphological-Based Non-Invasive Sex Identification of Blood
⁸⁸¹ Cockle, *Tegillarca granosa* (Linnaeus, 1758) using machine learning and deep
⁸⁸² learning technologies.

⁸⁸³ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture, over-
⁸⁸⁴ saw the overall workflow by providing baseline characteristics of the samples that
⁸⁸⁵ the researchers could focus on. Additionally, guidance was offered by the re-
⁸⁸⁶ search associates LC Mae Gasit and Allena Esther Artera. Consequently, the
⁸⁸⁷ entire dataset collection process was conducted at the University of the Philip-
⁸⁸⁸ pines Visayas hatchery facility.

⁸⁸⁹ The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical Con-

siderations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteristics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Software Configuration,(7) Machine Learning on Morphometric Data, (8) Deep Learning for Morphological Analysis, and (9) Evaluation Metrics

3.1 Sample Collection

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

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

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

As part of the data collection process, the researchers utilized induced spawning and dissection to classify the sex of the samples. Induced spawning through temperature fluctuations was the most natural and least invasive method for bivalves compared to other approaches (Aji, 2011). However, since not all samples exhibited gamete release, the researchers also performed dissections, assisted by hatchery staff, to expedite data collection. The sex of the dissected samples was

912 identified based on the coloration of gonad tissue, which varies according to sex
913 and maturity stage. Females exhibited orange-red to pale orange gonads, while
914 males displayed white to grayish-white gonads (May et al., 2021).

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



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

918 3.2 Ethical Considerations

919 The ongoing research project titled "Establishment of the Center for Mollusc Re-
920 search and Development: Development of Spawning and Hatchery Techniques for
921 the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture"—from which
922 the samples used in this study were obtained—was reviewed and approved by the
923 Institutional Animal Care and Use Committee (IACUC) of the University of the

924 Philippines Visayas.

925 3.3 Creating *T. granosa* Dataset

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

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

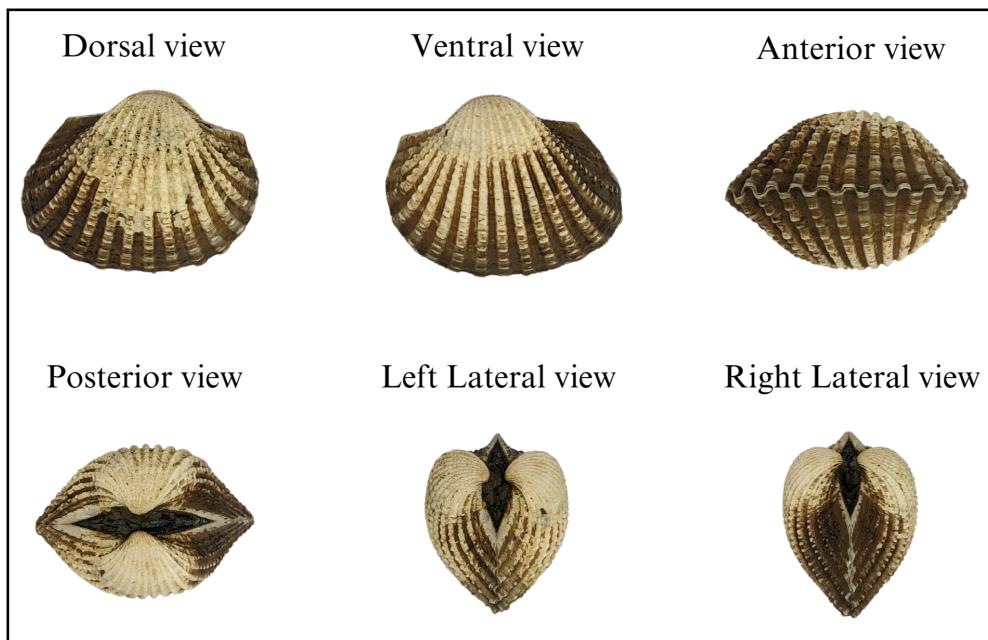


Figure 3.2: Different views of the *T. granosa* shell captured

⁹⁴⁵ 3.4 Morphometric Data Collection

⁹⁴⁶ Morphology refers to biological form and is one of the most visually recognizable
⁹⁴⁷ phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).
⁹⁴⁸ In this study, morphological characteristics describe the structural features of
⁹⁴⁹ *T. granosa*, focusing on measurable attributes such as shape, size, and color.
⁹⁵⁰ Morphometric characteristics, on the other hand, refer to specific quantifiable
⁹⁵¹ features of *T. granosa*, including length, width, height, hinge line length, distance
⁹⁵² between the umbos, and rib count. As stated by the researchers, quantifying and
⁹⁵³ characterizing these traits is essential for understanding and visualizing variations
⁹⁵⁴ in *T. granosa* morphology.

⁹⁵⁵ The researchers measured the height, width, and length of *T. granosa* using a
⁹⁵⁶ Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the

corresponding measurement diagram. Length (A) refers to the distance from the anterior to the posterior of the shell. Width (B) is defined as the widest span across the shell from the left to the right valve. Height (C) measures the distance from the base to the apex of the shell. In addition, the hinge line length (D) near the hinge and the distance between the umbos (E) were recorded.

Reyment and Kennedy (1998) emphasized that including rib count as supplementary information can enhance identification accuracy. Following this insight, the researchers also recorded the rib count for both male and female *T. granosa*, adjusting the values by calculating ratios to account for natural size variation among specimens.

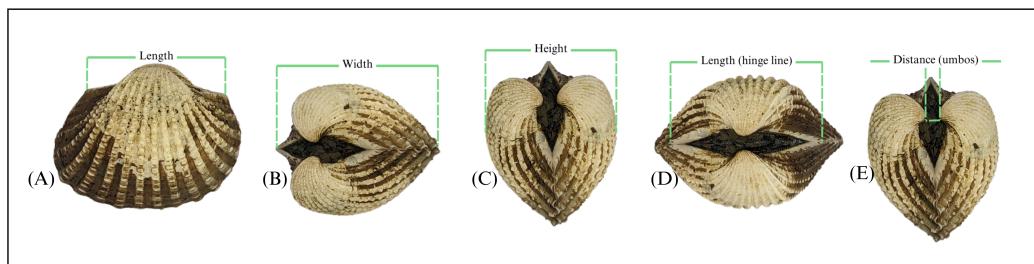


Figure 3.3: Linear measurements of the *T. granosa* shell.

3.5 Image Acquisition and Data Gathering

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

974 box to enhance image quality, eliminate shadows, and ensure clarity of the samples
975 throughout the image acquisition process.

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



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

981 3.6 Hardware and Software Configuration

982 This section of the paper discusses the software, programming languages, and tools
983 used for sex identification. Data collection, preprocessing, and model training
984 were conducted on a Windows 11 operating system using an ACER Aspire 3
985 general-purpose unit (GPU) equipped with an AMD Ryzen 3 7320U CPU with
986 Radeon Graphics (8 cores) @ 2.395 GHz and 8 GB of RAM. Google Colaboratory
987 was utilized for collaborative preprocessing, computer vision tasks, and model

988 training. Image preprocessing was performed using computer vision techniques in
989 Python, while machine learning and deep learning models were developed using
990 Python libraries, including Keras. The results of the gathered measurements were
991 stored and managed using spreadsheet software. GitHub was employed for version
992 control, documentation, and activity tracking throughout the study.

993 **3.7 Machine Learning on Morphometric Data**

994 This section of the paper discusses the machine learning operations that served
995 as a baseline prior to implementing more complex deep learning methods for
996 image classification. The study utilized collected variables including linear mea-
997 surements—length, width, height, hinge line length, distance between the um-
998 bos, and rib count—along with derived features used as predictors. These in-
999 cluded the length-to-width ratio, length-to-height ratio, width-to-height ratio,
1000 umbo distance-to-length ratio, hinge line length-to-length ratio, umbo distance-
1001 to-height ratio, and rib density. The samples were classified by sex, with females
1002 labeled as 0 and males as 1, which served as the response variable.

1003 **3.7.1 Data Preprocessing**

1004 The preprocessing of the dataset involved several essential steps, carried out using
1005 Python in Google Colaboratory, in preparation for machine learning analysis (*see*
1006 *Figure 3.5*).

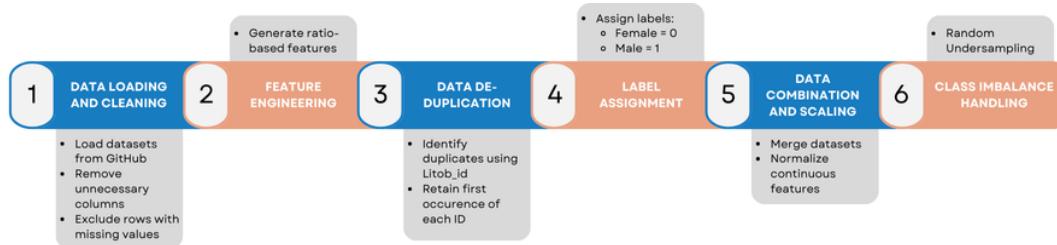


Figure 3.5: Data preprocessing in machine learning pipeline.

1007 **Data Loading and Cleaning**

1008 The process began by loading two separate datasets for male and female *T. granosa*
1009 directly from GitHub using `pd.read_csv()`. Unnecessary columns were removed,
1010 and rows containing missing values were excluded using the `dropna()` function to
1011 ensure data completeness and reliability.

1012 **Feature Engineering**

1013 Additional ratio-based features were generated to augment the existing measurements. These included the length-to-width ratio, length-to-height ratio, width-to-height ratio, hinge line length-to-length ratio, umbos distance-to-length ratio, umbos distance-to-height ratio, and rib density. These derived features aimed to emphasize shape characteristics independent of size, improving the models' ability 1017 to distinguish morphological differences between sexes.

1019 ***Data De-duplication***

1020 To avoid redundancy and ensure each specimen was uniquely represented, the
1021 last three digits of each `Litob_id` were used to identify duplicates. Only the first
1022 occurrence of each unique ID was retained, reducing potential bias caused by
1023 repeated entries.

1024 ***Label Assignment***

1025 A new column labeled `Label` was added to both datasets. Female specimens were
1026 assigned a label of 0, and male specimens a label of 1. This column served as the
1027 target variable for classification.

1028 ***Data Combination and Scaling***

1029 After cleaning and feature engineering, the male and female datasets were merged
1030 into a single DataFrame. The `Litob_id` column was removed post de-duplication.
1031 All continuous numeric features were normalized using `MinMaxScaler` to scale
1032 values to the range [0, 1].

1033 Rib count was excluded from normalization because it is a discrete feature with
1034 biologically meaningful bounds. According to best practices in machine learning,
1035 normalizing discrete or categorical features can distort their meaning and is often
1036 unnecessary (Jaiswal, 2024). In this study, rib count was treated as a categorical
1037 attribute due to its biological significance and finite, non-continuous nature.

1038 ***Class Imbalance Handling***

1039 After normalization, class imbalance was addressed by applying Random Under-
1040 sampling to the male dataset. This technique randomly reduced the number of

1041 male samples to match the number of female samples (127 each), ensuring equal
1042 class representation. By using this approach, model bias was minimized, and the
1043 classification performance became more reliable across both classes.

1044 3.7.2 Machine Learning Models Training

1045 *Model Selection and Hyperparameter Tuning*

1046 To establish a baseline for classification, various models were evaluated: Logis-
1047 tic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest,
1048 AdaBoost, Extra Trees, and Gradient Boosting. Hyperparameter tuning was con-
1049 ducted using `GridSearchCV`, which systematically identified the optimal settings
1050 for each model to enhance accuracy and performance.

1051 *Cross-Validation*

1052 A five-fold cross-validation approach was implemented (*refer to Figure 3.6*). The
1053 dataset was divided into five subsets, with four used for training and one for
1054 testing. This process was repeated five times, with each fold serving as the test set
1055 once. This method ensured that model evaluation was robust and generalizable,
1056 minimizing the bias that may result from a single train-test split. (GeeksforGeeks,
1057 2024)

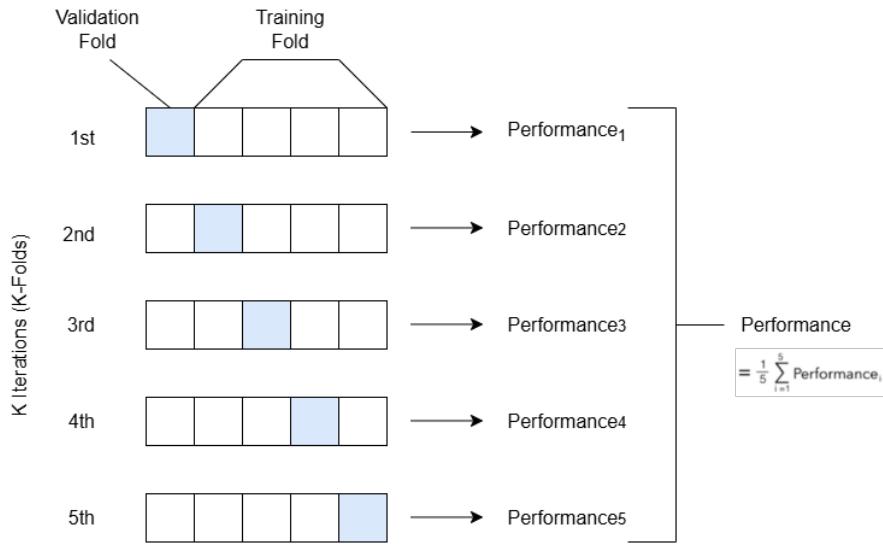


Figure 3.6: Diagram of k-fold cross-validation with $k = 5$.

1058 3.8 Deep Learning for Morphological Analysis

1059 This section outlines the application of deep learning techniques in analyzing the
1060 morphological characteristics of *Tegillarca granosa* to identify their sex based on
1061 shell images. A Convolutional Neural Network (CNN) architecture was imple-
1062 mented and trained on preprocessed images using stratified cross-validation.

1063 *Image Preprocessing*

1064 This subsection details the image processing techniques applied to raw shell images
1065 of *T. granosa* using computer vision methods before training the deep learning
1066 model. The image preprocessing techniques include standardizing input dimen-
1067 sions and removing shadows, background, and noise. Each image underwent data
1068 augmentation to enhance feature visibility for effective learning. Image prepro-
1069 cessing ensures consistent and high-quality input data for model training.

1070 ***Adjusting Dimensions***

1071 All images were resized to a consistent dimension of 256x256 pixels to ensure
1072 uniformity throughout the dataset. This standardization is essential for Convo-
1073 lutional Neural Networks (CNNs), as a consistent input dimension is required.
1074 While resizing, the aspect ratio was maintained to prevent distortion of the mor-
1075 phological features, and padding was added to retain the original format.

1076 ***Background Removal***

1077 Background removal was performed to maintain a consistent white background
1078 throughout the dataset. The tool `rembg` was used to efficiently remove the original
1079 background, retaining the foreground from the raw images. This method resulted
1080 in clear images with a white background, enhancing focus on the morphological
1081 features and defining the shell boundaries.

1082 ***Shadow Removal***

1083 To minimize noise caused by shadows around the shell, HSV thresholding, con-
1084 tours, and morphological thresholds were applied to isolate and remove shadowed
1085 regions. This approach preserved the natural color of the blood cockles and elim-
1086 inated shadows and noise from the surrounding area (*see Figures 3.7 and 3.8*).

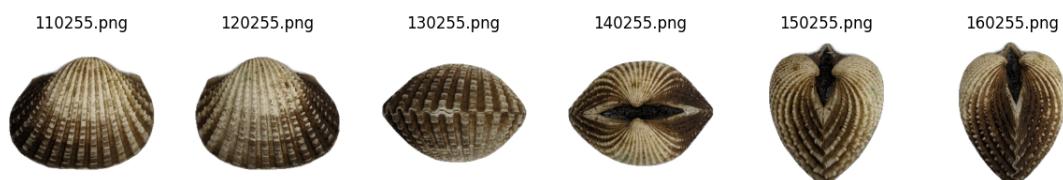


Figure 3.7: Shadows removed from male samples at different angles.

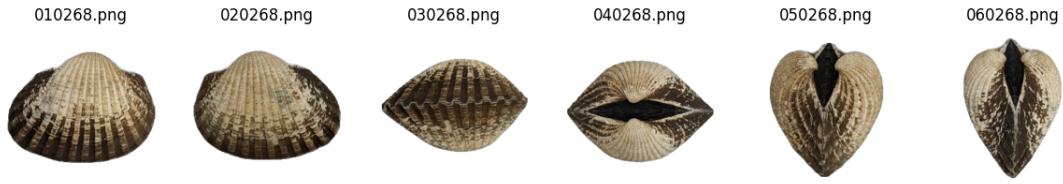


Figure 3.8: Shadows removed from female samples at different angles.

3.8.1 Convolutional Neural Network

Convolutional Neural Networks are the deep learning tool used in image classification, specifically binary classification. CNNs leverage their ability to share weights and use pooling techniques, reducing the number of parameters (Cui, Pan, Chen, & Zou, 2020). The proposed CNN architecture for sex identification of blood cockles employs 5 layers, designed to extract features from the input image with dimensions. The layers consist of three convolution layers, a pooling layer, a flatten layer, dropout, and two dense layers. The CNN framework used in this study was updated from an open source GitHub implementation by Christian Versloot, which focused on K-fold Cross Validation using TensorFlow and Keras, which was customized to align with the objectives of this study. The framework of this study is shown in Figure 3.9.

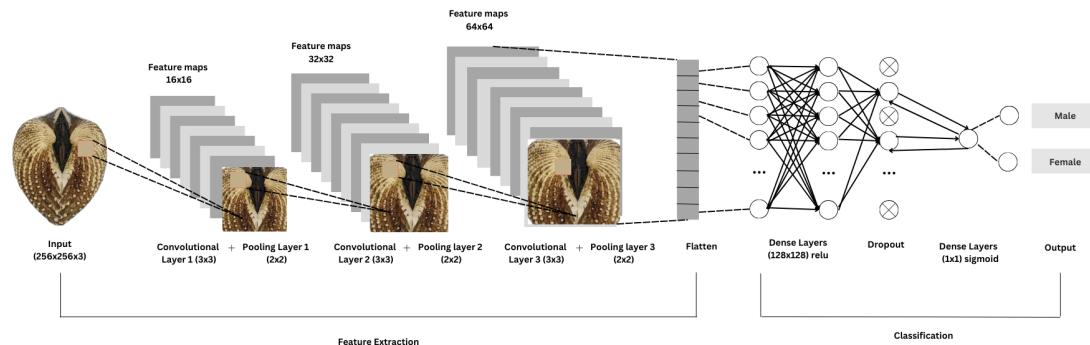


Figure 3.9: Architecture of convolutional neural network (CNN).

1099 ***Convolution Layer***

1100 The convolution layers of CNN extract the features from the input image through
1101 the convolution operation. This study uses three convolution layers with a 3x3
1102 kernel size and filter sizes of 16, 32, and 64 (*refer to Figure 3.1*). The first layer
1103 extracts the low-level features, such as edges, lines, and corners, while the deeper
1104 layers iteratively extract more complex information from these low-level features.
1105 The ReLU activation function is used as the baseline for this model, and experi-
1106 ments are conducted with different activation functions, such as ELU and PReLU,
1107 to evaluate their impact on learning complex patterns within the data.

1108 ***Pooling Layer***

1109 A pooling layer was added after the convolution layer to enhance calculation speed
1110 and prevent overfitting (Cui et al., 2020). In this study, max pooling was applied
1111 with a (3,3) kernel size.

1112 ***Fully Connected and Dropout***

1113 Fully connected layers follow after the convolution and pooling layers. Each neu-
1114 ron connects to all neurons of the previous layer. The output values from the
1115 fully connected layers are sent to an output layer. It was classified using different
1116 sigmoid functions appropriate for binary classification.

1117 A large number of parameters in the training process can lead to overfitting. It
1118 occurs when the model learns the training data too well, including its noise and
1119 irrelevant details. This results in poor performance on unseen data. To mitigate
1120 the overfitting, the dropout layer was employed. Dropout works by temporarily
1121 discarding a portion of the neurons in the network with probability p ($0 < p < 1$).

1122 During this process, these neurons do not participate in the forward propagation
1123 process of CNN and the backward propagation process (Cui et al., 2020).

Layer	Number of Neurons	Stride	Kernel Size	Activation	Parameters
Rescaling					
Convolution	16	1x1	3x3	ReLU	448
Max Pooling		1x1	3x3		
Convolution	32	1x1	3x3	ReLU	4,640
Max Pooling		1x1	3x3		
Convolution	64	1x1	3x3	ReLU	18,496
Max Pooling		1x1	3x3		
Flatten					
Dense	128			ReLU	7,372,928
Dropout					
Dense	1			Sigmoid	129

Table 3.1: Architecture of proposed convolution neural network.

1124 3.8.2 CNN Training

1125 The dataset consists of 1626 images, with 127 samples from females and 144 sam-
1126 ples from males, individually for each angle. Given the minimal class imbalance,
1127 random undersampling was carried out to create a balanced dataset. All images
1128 were resized to 256x256 pixels and normalized using a Rescaling layer, ensuring
1129 pixel values were within the range [0, 1].

1130 *Data Splitting*

1131 Due to the limited dataset size, a traditional train-test split was not adopted.
1132 Instead, a 5-fold stratified cross-validation approach was used to maximize the
1133 use of available data while preserving the class distribution within each fold (*refer*
1134 *to Figure 3.10*). `StratifiedKFold` was applied to ensure that the distribution of
1135 male and female samples remained consistent across all folds, thereby enabling
1136 fair and robust model evaluation (GeeksforGeeks, 2020).

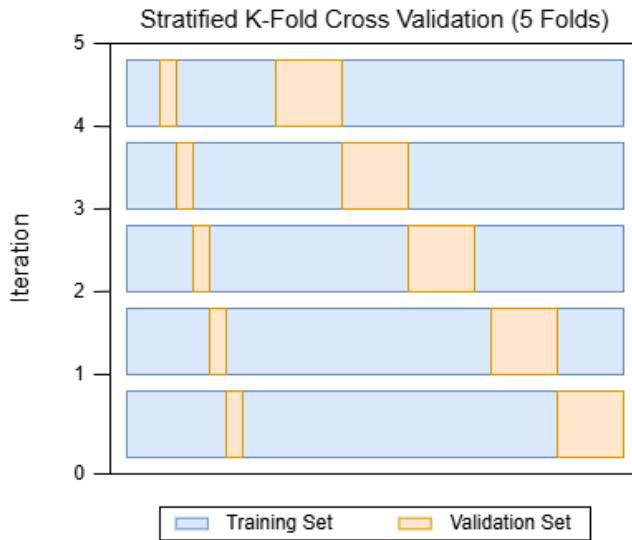


Figure 3.10: Diagram of stratified k-fold cross-validation with $k=5$.

¹¹³⁷ **Data Augmentation**

¹¹³⁸ Before model training, online data augmentation was applied exclusively to the
¹¹³⁹ training data within each fold, creating new data variations on the fly. The aug-
¹¹⁴⁰ mentations included random horizontal flips, slight rotations, and zoom trans-
¹¹⁴¹ formations to enhance data diversity and improve model generalization (Awan,
¹¹⁴² 2022). All augmentation was strictly applied only to the training subset of each
¹¹⁴³ fold to prevent data leakage and maintain the validity of the results (*Figure 3.11*).

¹¹⁴⁴ On-the-fly data augmentation (OnDAT) generates augmented data during each
¹¹⁴⁵ iteration, exposing the model to constantly changing data variations. Augmenting
¹¹⁴⁶ the original data allows better exploration of the underlying data generation pro-
¹¹⁴⁷ cess and has the potential to prevent the model from overfitting spurious patterns,
¹¹⁴⁸ thereby improving performance (Cerqueira, Santos, Baghoussi, & Soares, 2024).

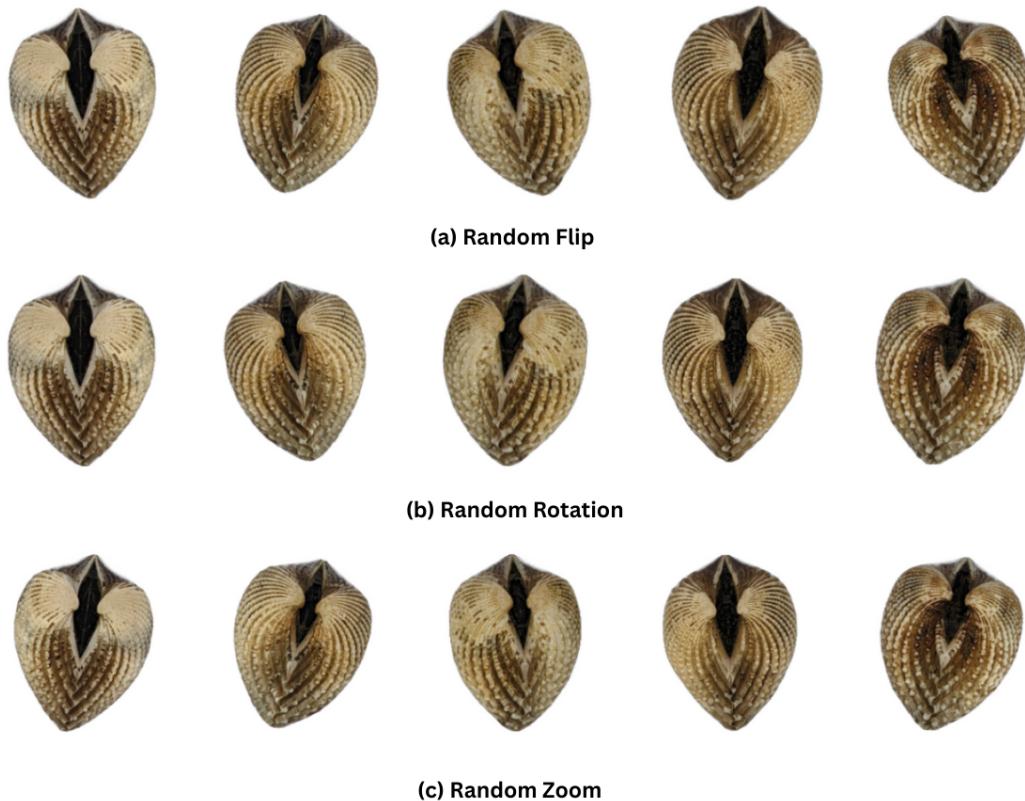


Figure 3.11: Data augmentation techniques.

1149 ***Training Procedure***

1150 During the training process, model performance per fold was carefully monitored.
1151 One important thing to observe is the consistency in the performance, whether
1152 the model is still learning or is at high risk of overfitting. Early stopping was ap-
1153 plied to ensure the stable performance of the model across folds. This technique
1154 allows for monitoring the training of the neural network, stopping when the per-
1155 formance metrics, in this case, validation loss, cease to improve. Furthermore, to
1156 enhance the learning process, `ReduceLROnPlateau` was applied, which decreased
1157 the learning rate if there was no improvement in the model for a specified number
1158 of epochs (Team, n.d.).

1159 The model was trained using the Adam optimization algorithm, with an initial
1160 learning rate of 0.001. Binary cross-entropy, commonly known as the log loss,
1161 was employed as the loss function due to its effectiveness in binary classification
1162 tasks. To reduce the risk of overfitting, a dropout rate of 0.5 was applied, ran-
1163 domly deactivating half of the neurons during the training process to improve
1164 generalization.

1165 3.9 Evaluation Metrics

1166 Evaluating the performance of a binary classification model is essential, and se-
1167 lecting appropriate metrics depends on the specific requirements of the user. The
1168 performance of both supervised machine learning and deep learning models will
1169 be measured using several key metrics, including accuracy, precision, recall, F1
1170 score, and the AUC-ROC score.

1171 Accuracy (ACC) is the ratio of the overall correctly predicted samples to the
1172 total number of examples in the evaluation dataset (Cui et al., 2020). It measures
1173 the overall correctness of the model in predicting both male and female blood
1174 cockles. This metric provides insight into how well the model performs across all
1175 classifications. The formula for accuracy is:

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

1176 Precision (PREC) is the ratio of correctly predicted positive samples to all samples
1177 assigned to the positive class (Cui et al., 2020). This metric helps in evaluating

₁₁₇₈ the fairness of the model and prevents the misclassification of blood cockles as it
₁₁₇₉ identifies potential inaccuracies or biases. The formula for precision is:

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

₁₁₈₀ Recall (REC), also known as sensitivity or the true positive rate (TPR), is the
₁₁₈₁ ratio of correctly predicted positive cases to all the actual positive samples (Cui
₁₁₈₂ et al., 2020). It represents the ability of the model to correctly identify positive
₁₁₈₃ male and female samples. The formula for recall is:

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

₁₁₈₄ The F1 score is the harmonic mean of precision and recall, which penalizes extreme
₁₁₈₅ values of either of the two metrics (Cui et al., 2020). It is particularly useful when
₁₁₈₆ the class distribution is imbalanced. The formula for the F1 score is:

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

₁₁₈₇ The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a
₁₁₈₈ performance measurement for classification problems, particularly used in deep
₁₁₈₉ learning in this study. The ROC curve is a plot of the true positive rate (recall)
₁₁₉₀ against the false positive rate (1 - specificity), and the AUC score quantifies the
₁₁₉₁ overall ability of the model to discriminate between positive and negative classes.
₁₁₉₂ A higher AUC indicates better model performance. (Nahm, 2022)

₁₁₉₃ **Chapter 4**

₁₁₉₄ **Results and Discussions**

₁₁₉₅ This chapter presents the results from the machine learning and deep learning
₁₁₉₆ analyses conducted on the preprocessed dataset. It includes an evaluation of
₁₁₉₇ various machine learning classifiers and the application of deep learning models
₁₁₉₈ for image-based classification. The primary focus is on identifying key predictors
₁₁₉₉ and assessing classification performance for sex identification in *T. granosa*.

₁₂₀₀ **4.1 Machine Learning Analysis**

₁₂₀₁ This chapter outlines the results of preprocessing, training of machine learning
₁₂₀₂ models, and feature importance analysis, all conducted in Google Colab using
₁₂₀₃ Python. The dataset was preprocessed in Colab, and the training and evaluation
₁₂₀₄ of various classifiers were performed entirely within this environment. This part of
₁₂₀₅ the paper includes five subsections: data exploration, statistical analysis, feature
₁₂₀₆ importance analysis, performance evaluation, and confusion matrix analysis.

¹²⁰⁷ 4.1.1 Data Exploration

¹²⁰⁸ Exploratory data analysis was performed to characterize the dataset using visu-
¹²⁰⁹ alizations to understand the patterns and correlations within the data. A corre-
¹²¹⁰ lation heatmap was created to assess the relationship between the predictors and
¹²¹¹ the target variable.

¹²¹² The heatmap (*see Figure 4.1*) revealed three features most correlated with the
¹²¹³ sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos-length ratio (r
¹²¹⁴ = 0.12), and the distance between the umbos ($r = 0.12$). Each of these features
¹²¹⁵ demonstrated a weak positive relationship with the target variable.

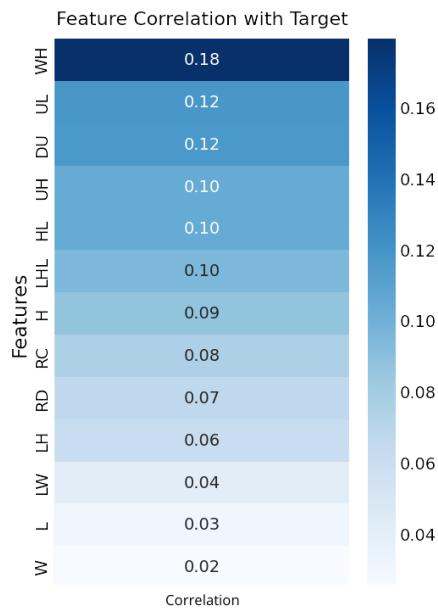


Figure 4.1: Heatmap of morphometric correlations with *T. granosa* sex.

₁₂₁₆ **4.1.2 Statistical Analysis**

₁₂₁₇ As part of the exploratory data analysis, statistical testing confirmed that the
₁₂₁₈ dataset did not follow a normal distribution (*see Table 4.1*). Consequently, the
₁₂₁₉ Mann-Whitney U test was applied with a significance level of $\alpha = 0.05$ to com-
₁₂₂₀ pare male and female samples. Out of thirteen features, five showed statistically
₁₂₂₁ significant differences. These included: distance between umbos ($p = 0.025$),
₁₂₂₂ length-width ratio ($p = 0.011$), umbos-length ratio ($p = 0.019$), width-height
₁₂₂₃ ratio ($p = 0.003$), and umbos-height ratio ($p = 0.036$).

₁₂₂₄ It is important to note that statistical significance does not imply predictive im-
₁₂₂₅ portance. Therefore, further analysis, such as feature importance evaluation, was
₁₂₂₆ 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.

4.1.3 Feature Importance Analysis

1228 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric
 1229 method that is suitable for evaluating differences in distributions across groups
 1230 when the data does not follow a normal distribution. This approach was chosen
 1231 because of the non-normality of the dataset and its robustness in handling con-
 1232 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,
 1233 2024)

1234 The analysis showed that the width-to-height ratio (WH ratio) had the high-
 1235 est importance score, indicating it is the most statistically significant feature for
 1236 distinguishing the sex of *T. granosa*. Other notable features included the length-
 1237 to-width ratio (LW ratio), umbo distance-to-length ratio (UL ratio), distance
 1238 between the umbos, and umbo distance-to-height ratio (UH ratio), all of which
 1239 contributed significantly to the classification task (*refer to Figure 4.2*).

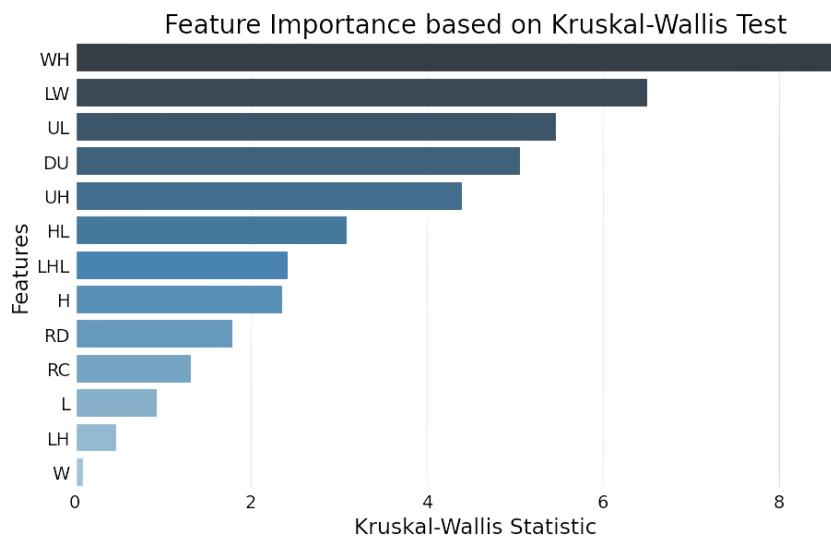


Figure 4.2: Feature importance scores using the Kruskal-Wallis test.

¹²⁴⁰ **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.

¹²⁴¹ Table 4.2 shows the performance metrics of different machine learning models
¹²⁴² trained using all 13 features from the dataset. Among the models, Gradient
¹²⁴³ Boosting achieved the highest accuracy of 61.03%, along with strong precision,
¹²⁴⁴ recall, and F1-score values. AdaBoost also performed competitively, with an ac-
¹²⁴⁵ curacy of 60.63%. These results highlight the effectiveness of ensemble methods
¹²⁴⁶ such as Gradient Boosting and AdaBoost when utilizing the full feature set, likely
¹²⁴⁷ because of their capability to combine multiple weak learners into a more robust
¹²⁴⁸ 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.

¹²⁴⁹ Table 4.3 presents the performance of the same models using only the top five fea-
¹²⁵⁰ tures identified through Kruskal-Wallis feature importance analysis. The selected
¹²⁵¹ features are the distance between the umbos, length-to-width ratio, width-to-

1252 height ratio, umbo distance-to-height ratio, and umbo distance-to-length ratio.

1253 Interestingly, the overall performance of the models improved when using only the
1254 top 5 features compared to using all 13. K-Nearest Neighbors (KNN) achieved the
1255 best results with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and
1256 an F1-score of 63.75%. Gradient Boosting followed closely behind. These find-
1257 ings suggest that reducing the feature set to the most relevant variables helped
1258 simplify the models, improved generalization, and enhanced predictive perfor-
1259 mance—particularly for KNN, which showed a notable improvement over its ear-
1260 lier results with the full feature set.

1261 4.1.5 Confusion Matrix Analysis

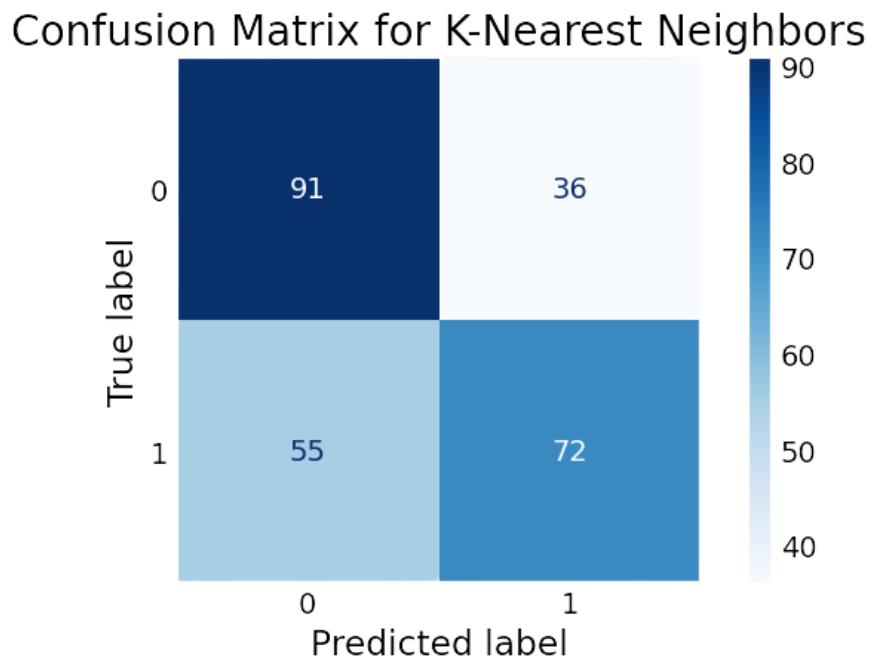


Figure 4.3: KNN confusion matrix for *T. granosa* sex classification.

1262 Figure 4.3 summarizes the performance of the K-Nearest Neighbors model in

1263 classifying *T. granosa* based on their sex, where 0 represents female samples and
1264 1 represents male samples. From the matrix, we observe that out of all the actual
1265 female samples (true label 0), 91 were correctly predicted as female (true positive
1266 for class 0), while 36 were incorrectly classified as male (false negative for class
1267 0). On the other hand, out of all the actual male samples (true label 1), 72 were
1268 correctly predicted as male (true positive for class 1), while 55 were incorrectly
1269 classified as female (false negative for class 1).

1270 4.2 Deep Learning Analysis

1271 This section presents the performance of the Convolutional Neural Network (CNN)
1272 model in classifying the sex of *T. granosa* based on shell morphology. The analysis
1273 evaluates the model's ability to distinguish between male and female shell images
1274 using various evaluation metrics. This part of the paper includes six subsections:
1275 baseline model, comparison of individual and combined angles, training result and
1276 hyperparameter tuning, proposed model, learning rates and training behavior per
1277 fold, and visualizations.

1278 The machine learning analysis (see Figure 4.3) revealed that five of the origi-
1279 nal features produced significant results. The K-Nearest Neighbor (KNN) model
1280 achieved an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1
1281 score of 63.75%. This section compares the model's performance across differ-
1282 ent angles based on the results of the machine learning and feature importance
1283 analysis.

₁₂₈₄ **4.2.1 Baseline Model**

₁₂₈₅ This section presents the baseline model with a batch size of 16 and 20 epochs,
₁₂₈₆ which will serve as the starting point for comparison and provide a guideline for
₁₂₈₇ hyperparameter tuning. The focus will be on one of the angles, specifically the
₁₂₈₈ Left Lateral view, since the feature importance analysis using the Kruskal-Wallis
₁₂₈₉ Test indicated that the width-to-height ratio had the highest importance score,
₁₂₉₀ 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).

₁₂₉₁ The unbalanced dataset, which consisted of 144 male samples and 127 female
₁₂₉₂ samples, achieved an accuracy of 65.27%, precision of 71.82%, recall of 58.99%,
₁₂₉₃ an F1-score of 63.99%, an AUC score of 73.08%, and a loss of 0.6122. However, to
₁₂₉₄ address the class imbalance and enhance model performance, random undersam-
₁₂₉₅ pling was performed. This approach resulted in improved performance metrics for
₁₂₉₆ the balanced dataset, with an accuracy of 67.34%, precision of 69.43%, a recall
₁₂₉₇ of 64.06%, an F1-score of 65.60%, an AUC score of 74.31%, and a lower loss of
₁₂₉₈ 0.5981.

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

₁₃₀₀ Using the same batch size and number of epochs, performance was compared
₁₃₀₁ across all individual angles and the combination of the two highest-performing

₁₃₀₂ angles based on accuracy, using a balanced dataset. For the combined analysis,
₁₃₀₃ samples from the two selected angles were placed side by side, and a new dataset
₁₃₀₄ 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).

₁₃₀₅ Table 4.5 presents the performance metrics for each individual angle and the com-
₁₃₀₆ bination of the two highest-performing angles in terms of accuracy. The Left Lat-
₁₃₀₇ eral view achieved the highest accuracy (67.34%) and precision (69.43%), while the
₁₃₀₈ Dorsal view obtained the highest recall (77.88%) and F1-score (69.96%). Mean-
₁₃₀₉ while, the Ventral view recorded the highest AUC score (74.87%), indicating its
₁₃₁₀ strong ability to distinguish between classes. Combining the Ventral and Left
₁₃₁₁ Lateral views resulted in an overall accuracy of 62.60%, suggesting that while
₁₃₁₂ combined images may provide complementary information, individual angle views
₁₃₁₃ still outperformed the combined views under the current experimental setup.

₁₃₁₄ 4.2.3 Training Result and Hyperparameter Tuning

₁₃₁₅ The Left Lateral angle was selected for further optimization. Several experiments
₁₃₁₆ were conducted by tuning hyperparameters such as batch size, number of epochs,
₁₃₁₇ and activation functions. Each adjustment was compared against the baseline

₁₃₁₈ model to enhance performance and develop a robust CNN for sex classification of
₁₃₁₉ *T. granosa*.

₁₃₂₀ The Left Lateral angle was chosen because it achieved the highest accuracy and
₁₃₂₁ precision among all individual views, and because the Kruskal-Wallis feature im-
₁₃₂₂ portance analysis indicated that the width-to-height ratio, a feature most visible
₁₃₂₃ from the lateral perspective, was the most significant morphological trait for clas-
₁₃₂₄ sification. Therefore, focusing on this view was expected to maximize the model's
₁₃₂₅ learning capacity and improve classification performance.

₁₃₂₆ **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.

₁₃₂₇ Table 4.6 shows the results indicating that a batch size of 32 with 50 epochs
₁₃₂₈ achieved the best overall performance, with an accuracy of 71.68%, a precision of
₁₃₂₉ 72.52%, a recall of 69.29%, an F1-score of 69.12%, and AUC score of 77.34%.

₁₃₃₀ In contrast, increasing the batch size to 64 resulted in lower recall and F1-scores,
₁₃₃₁ suggesting that smaller batch Sizes (16 or 32) are more effective for this dataset.
₁₃₃₂ A moderate batch size of 32 allowed the model to generalize better and maintain
₁₃₃₃ stable learning, while too large batch sizes may have led to underfitting.

¹³³⁴ **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).

¹³³⁵ Table 4.7 the performance of different activation functions applied to the CNN
¹³³⁶ model trained with a batch size of 32 and 50 epochs. Based on the results, the
¹³³⁷ ReLU activation function achieved the best overall performance, with an accu-
¹³³⁸ racy of 71.68%, precision of 72.52%, recall of 69.29%, F1-score of 69.12%, and
¹³³⁹ AUC score of 77.34%, along with the lowest loss at 0.5824. This suggests that
¹³⁴⁰ ReLU remains an effective activation function for the classification of *T. granosa*,
¹³⁴¹ outperforming both ELU and PReLU in this setup.

¹³⁴² **4.2.4 Proposed Model**

¹³⁴³ This section presents the performance evaluation of the proposed Convolutional
¹³⁴⁴ Neural Network (CNN) model, trained with a batch size of 32, 50 epochs, and us-
¹³⁴⁵ ing the ReLU activation function. The model's effectiveness was assessed through
¹³⁴⁶ 5-fold cross-validation to ensure robustness and generalizability across different
¹³⁴⁷ 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).

1348 The proposed model consistently achieved high performance in Folds 1, 3, and
1349 5, with accuracies above 76% and strong recall and AUC scores, demonstrating
1350 its potential for reliable sex identification of *T. granosa*. The slight variation
1351 in performance across folds may be attributed to differences in data distribution,
1352 emphasizing the importance of further data augmentation and balancing for future
1353 work.

1354 4.2.5 Learning Rates and Training Behavior per Fold

1355 This section presents the learning rate adjustments, early stopping events, and
1356 best epoch selections for each fold during the 5-fold cross-validation of the pro-
1357 posed model. During training, the ReduceLROnPlateau callback was employed
1358 to monitor the validation loss and automatically reduce the learning rate when
1359 performance plateaued. Additionally, EarlyStopping was utilized to halt training
1360 once no further improvement was observed after a set patience, and the model
1361 weights were restored from the end of the best-performing epoch to ensure optimal
1362 performance.

1363 The following table summarizes the epochs where learning rate reductions oc-
1364 curred, the adjusted learning rates, the epochs at which early stopping took place,
1365 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.

1366 4.2.6 Visualizations

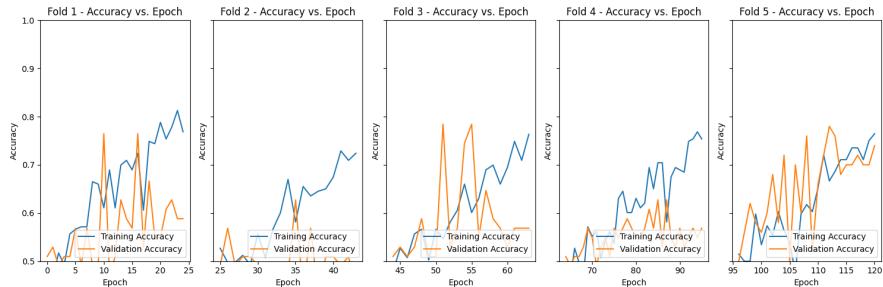


Figure 4.4: Training and validation accuracy per fold.

1367 Figure 4.4 shows the performance of the model in the training and validation in
 1368 terms of accuracy across five folds. The graph across folds displays a consistent
 1369 upward trend for the training accuracy. However, there is an observable change in
 1370 the performance, particularly in Folds 1 and 2, where it shows a slight downward
 1371 trend in the validation accuracy.

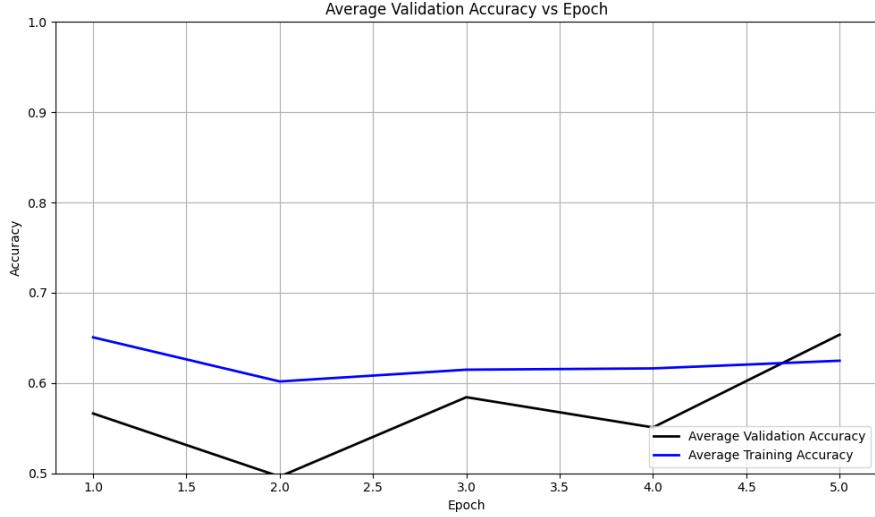


Figure 4.5: Average training and validation accuracy across folds.

1372 Figure 4.5 shows the average performance of the model in both training and accu-
 1373 racy in terms of accuracy across five folds. Similar to the individual performances,
 1374 there is an observable upward trend, which shows that the accuracy score improves
 1375 with the number of folds. The validation accuracy shows a downward and upward
 1376 trend that shows that it gradually improves on later epochs. The accuracy in
 1377 the training is slightly higher than the accuracy when validating the model, it
 1378 indicates that the model learns during training.

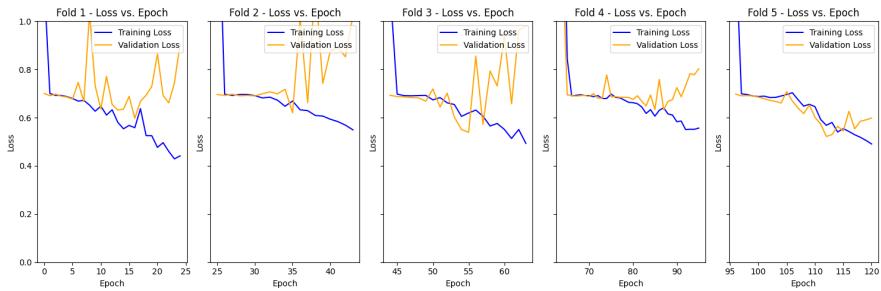


Figure 4.6: Training and validation loss per fold.

1379 Figure 4.6 shows the performance of the model in the training and validation in
 1380 terms of the training and validation loss across five folds. The graph across folds

1381 displays a consistent downward trend for the training loss. On the other hand,
1382 there is an observable change in the performance, especially in Folds 1,2,3, and 4,
1383 where it shows an upward trend in the validation loss. This is an implication for
1384 the learning performance of the model, as it may not be learning effectively.

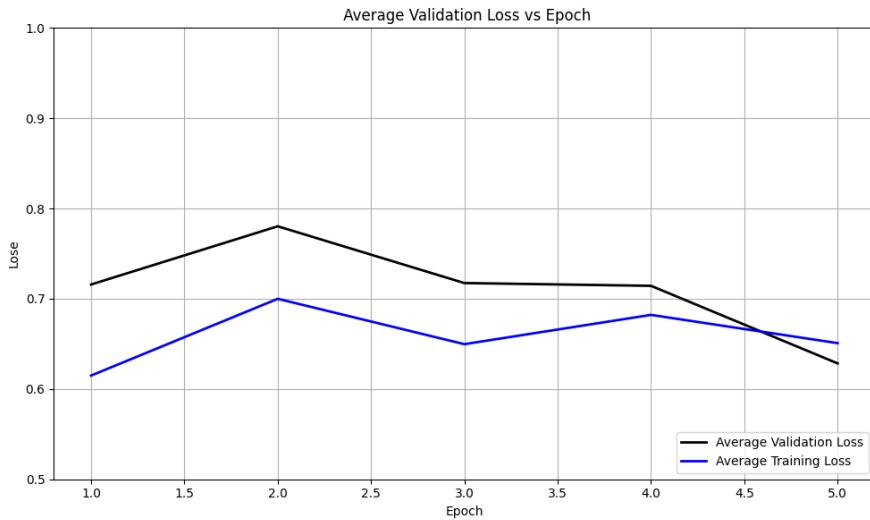


Figure 4.7: Average training and validation loss across folds.

1385 Figure 4.7 shows the average performance of the model in both the training and
1386 validation in terms of loss across five folds. There is an observable downward trend
1387 in both the average loss for training and validation. Additionally, the average
1388 training loss is slightly lower than the average validation loss.

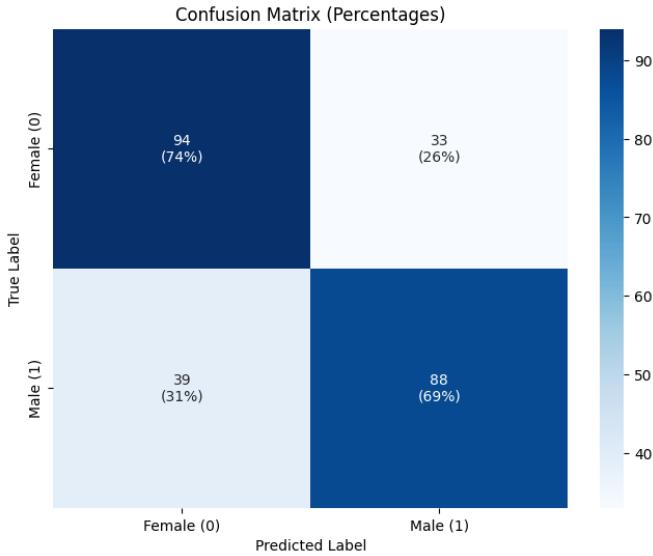


Figure 4.8: Confusion matrix for final model predictions.

1389 Figure 4.8 shows the confusion matrix for the true class label and predicted class
 1390 label. The matrix shows the correctly predicted male and female samples along
 1391 with their corresponding percentages. There is an observable trend where females
 1392 have slightly higher true positives compared to males in the number and per-
 1393 centages for the correctly classified male and female samples, which are 94 and
 1394 88, corresponding to 74% and 69%, respectively. Additionally, the false classified
 1395 samples were 33 for females and 39 for males, respectively accounting for 26% and
 1396 31%.

1397 Figure 4.9 shows the ROC Curve shows the ability of the proposed model to
 1398 correctly identify the true positives, which can help determine the tradeoff between
 1399 specificity and sensitivity. It will also determine the validity of the model, that it is
 1400 not predicting based only on random chances. The range of AUC ROC is between
 1401 0.5 and 1. The model was able to achieve a score of 0.7734, which is better than
 1402 random chances and an indication that the model is performing reasonably.

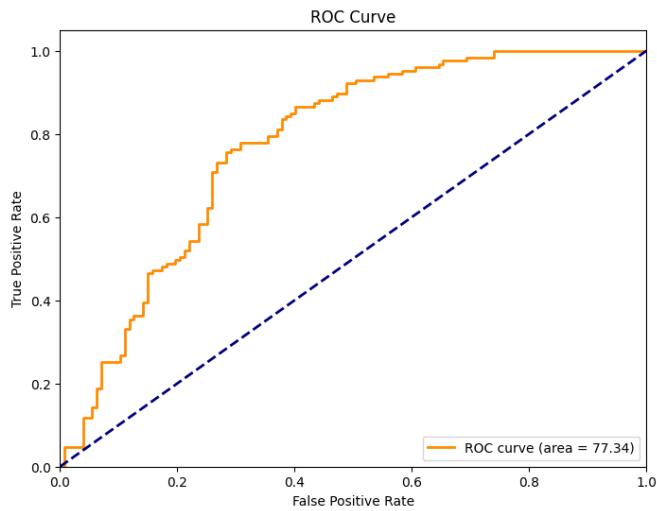


Figure 4.9: ROC curve with AUC score for the proposed model.

1403 4.3 Discussions

1404 This study aimed to develop a non-invasive method for identifying the sex of
 1405 *T. granosa* using machine learning and deep learning techniques. The dataset
 1406 used in the study was manually curated by the researchers, consisting of linear
 1407 measurements and the images captured from six different camera angles.

1408 Initial experiments conducted through machine learning focused on classification
 1409 accuracy, feature selection, and feature importance analysis. The machine learning
 1410 approach revealed that using five key features, selected through statistical tests
 1411 (Mann-Whitney U-test and Kruskal-Wallis test), outperformed models trained on
 1412 all 13 features. The K-nearest neighbors (KNN) classifier, utilizing five significant
 1413 features, achieved an accuracy of 64.16%, a precision of 64.97%, a recall of 64.16%,
 1414 and an F1-score of 63.57%. Feature importance analysis identified the width-
 1415 height ratio as the most discriminative variable, observed in the left lateral view,
 1416 as supported by the feature correlation score of $p=0.18$. These results indicated

1417 that a more focused set of features can enhance model performance, confirming
1418 the potential of non-invasive sex identification using linear measurements.

1419 Subsequent deep learning experiments explored the impact of different image an-
1420 gles and model hyperparameters on classification accuracy. The baseline model
1421 with balanced data outperformed those trained on unbalanced data. Hyperpa-
1422 rameter optimization, involving sets of epochs and batch sizes, further improved
1423 the model performance. Additionally, the influence of different activation func-
1424 tions, such as ReLU, eLU, and pReLU, was evaluated. The study found that the
1425 camera angle from the left lateral view consistently produced the best results, with
1426 an accuracy of 71.68%, precision of 72.52%, recall of 69.29%, F1-score of 69.12%,
1427 and an AUC score of 77.34%. The optimal model convergence was achieved with
1428 50 epochs, a batch size of 32, and ReLU as the activation function.

1429 Performance variations across experiments emphasize the role of fine-tuning pa-
1430 rameters, including image angles, batch sizes, epochs, activation functions, and
1431 the learning rate in influencing model behavior and performance. The progression
1432 of validation loss was closely monitored, as it signifies overfitting or underfitting.
1433 To mitigate overfitting and enhance model generalization, data augmentation,
1434 stratified sampling, and regularization techniques such as early stopping were em-
1435 ployed.

1436 The findings are significant because they demonstrate the feasibility of a non-
1437 invasive, accurate, and efficient sex identification method for *T. granosa*. This
1438 approach aligns with sustainable aquaculture practices by reducing the need for
1439 invasive sex-identifying methods and offering its potential in real-time settings.
1440 By integrating machine learning with deep learning image analysis, this study

¹⁴⁴¹ provides a valuable model for non-invasive sex identification for *T. granosa*.

¹⁴⁴² Compared to similar existing studies, such as the gender classification method for
¹⁴⁴³ Chinese mitten crab using deep learning CNNs (Cui et al., 2020), there are both
¹⁴⁴⁴ methodological similarities and differences. Both studies employed Convolutional
¹⁴⁴⁵ Neural Networks (CNNs) with three convolutional layers, pooling layers, fully
¹⁴⁴⁶ connected layers, and dropout. The crab study used grayscale images resized to
¹⁴⁴⁷ 64×64 pixels, while this study utilized higher-resolution RGB images (256×256).
¹⁴⁴⁸ In terms of architecture, the crab study applied 4, 8, and 16 filters in its con-
¹⁴⁴⁹ volutional layers and 256 neurons in the fully connected layer, achieving a high
¹⁴⁵⁰ accuracy of 98.90%. In contrast, this study used 16, 32, and 64 filters in the convo-
¹⁴⁵¹ lutional layers and 128 neurons in the fully connected layer, reaching an accuracy
¹⁴⁵² of 71.68%. This lower performance may be attributed to the subtler morphological
¹⁴⁵³ differences between male and female *T. granosa*, as well as limitations in image
¹⁴⁵⁴ quality and sample size.

¹⁴⁵⁵ This study acknowledges several limitations, particularly the size of the dataset
¹⁴⁵⁶ (271 samples) and the reliance on six fixed image angles. These constraints may
¹⁴⁵⁷ not fully represent the morphological variability across different populations or
¹⁴⁵⁸ environments. Nevertheless, the results indicate that combining machine learning
¹⁴⁵⁹ and deep learning offers a promising and non-invasive approach for sex identifica-
¹⁴⁶⁰ tion in *T. granosa*.

¹⁴⁶¹ Chapter 5

¹⁴⁶² Conclusion and ¹⁴⁶³ Recommendations

¹⁴⁶⁴ 5.1 Conclusion

¹⁴⁶⁵ This study utilized the application of machine learning and deep learning tech-
¹⁴⁶⁶ niques to identify the sex of *T. granosa* based on the morphometric and morpho-
¹⁴⁶⁷ logical characteristics. A manually curated dataset was developed, consisting of
¹⁴⁶⁸ both linear measurements and images captured from six different angles. Machine
¹⁴⁶⁹ learning methods were employed to identify statistically significant features, which
¹⁴⁷⁰ served as the basis for deep learning analysis using a five-layered Convolutional
¹⁴⁷¹ Neural Network (CNN). The proposed CNN model yielded an average accuracy of
¹⁴⁷² 71.68% in the performance metrics. Overall, this study offers a classification ap-
¹⁴⁷³ proach which is a viable solution for non-invasive sex identification, providing an
¹⁴⁷⁴ in-depth analysis based on *T. granosa*'s linear measurements and morphological

¹⁴⁷⁵ characteristics from different angles.

¹⁴⁷⁶ Through the availability of the gathered data, trial-and-error experimentation
¹⁴⁷⁷ was conducted by adjusting the number of layers, batch size, epoch, and activa-
¹⁴⁷⁸ tion functions. The different combinations tested provided baseline results that
¹⁴⁷⁹ demonstrate the feasibility of non-invasive sex identification for *T. granosa*.

¹⁴⁸⁰ While the study has made significant progress, challenges were encountered during
¹⁴⁸¹ CNN training, particularly due to hardware memory limitations. To overcome
¹⁴⁸² these, the researchers utilized synchronous Google Colab with 100 computing
¹⁴⁸³ units, requiring subscriptions, repeated retraining, and reconfigurations, which
¹⁴⁸⁴ demanded considerable financial resources and time to optimize the parameters.

¹⁴⁸⁵ Upon comparing the experimental results of model parameters, it was demon-
¹⁴⁸⁶ strated that non-invasive sex identification on *T. granosa* is achievable through
¹⁴⁸⁷ the integration of machine learning and deep learning methods. Machine learn-
¹⁴⁸⁸ ing models based on five statistically selected features had better performances
¹⁴⁸⁹ than those based on all features, with an accuracy of 64.16%, precision of 64.97%,
¹⁴⁹⁰ recall of 64.16%, and an F1-score of 63.57% using K-nearest neighbors (KNN)
¹⁴⁹¹ classifier. The classification performance was further enhanced by deep learning
¹⁴⁹² models, using Left Lateral image view, achieving an accuracy of 71.68%, precision
¹⁴⁹³ of 72.52%, recall of 69.29%, F1-score of 69.12%, and an AUC score of 77.34%.

¹⁴⁹⁴ These findings demonstrate that the CNN model can serve as a baseline for future
¹⁴⁹⁵ studies on non-invasive sex identification of *T. granosa*. By providing a practical
¹⁴⁹⁶ and less harmful alternative to traditional methods, this research contributes a
¹⁴⁹⁷ significant advancement in the field of aquaculture and marine biology.

1498 5.2 Recommendations

1499 This special problem entitled Morphometric and Morphological-Based Non-invasive
1500 Sex Identification of *T. granosa* focuses on creating a baseline study that will serve
1501 as a foundation for further studies involving *T. granosa*, blood cockles, using ma-
1502 chine learning and deep technologies in determining the sex of the samples is a
1503 salient need in aquaculture practices. Thus, the proposed recommendations are
1504 the future applications to improve and have detailed analysis, such as focusing
1505 on shape analysis, exploring other state-of-the-art deep learning techniques, or
1506 transfer learning, such as ResNet, SqueezeNet, and InceptionNet, and comparing
1507 the analysis results. Furthermore, the main goal of conducting this is to have the
1508 ability to identify the sex of the samples by taking real-time angles by rotating
1509 from the dorsal, lateral, and anterior.

1510 Due to the time constraints, the researchers were only able to gather a total of
1511 1,626 images with 271 images per angle, and utilized these for model training and
1512 validation. A larger and more diverse collection of images could further improve
1513 the model's generalization. In order to capture more variability, future study
1514 might include expanding the dataset to improve classification performance.

1515 Future studies could also invest in a sturdier and more controlled environment
1516 by using a green background and positioning a fixed camera angle during image
1517 acquisition. In addition, researchers may experiment with other image processing
1518 techniques such as morphological transformations to emphasize features. The
1519 dataset can be utilized for further analysis through advanced deep learning and
1520 computer vision methods to make sense of the images gathered and discern sexual
1521 dimorphism for *T. granosa*.

₁₅₂₂ **Chapter 6**

₁₅₂₃ **References**

- ₁₅₂₄ 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
- ₁₅₂₅
- ₁₅₂₆
- ₁₅₂₇ Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. , *10*.
- ₁₅₂₈
- ₁₅₂₉
- ₁₅₃₀ Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian Sains*, *14*, 14207.
- ₁₅₃₁
- ₁₅₃₂ Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).
- ₁₅₃₃ Data scaling performance on various machine learning algorithms to identify
- ₁₅₃₄ abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:
- ₁₅₃₅ 10.14710/jtsiskom.2021.14105
- ₁₅₃₆ Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.
- ₁₅₃₇ *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- ₁₅₃₈ Awan, A. A. (2022, November). *A complete guide to data augmentation*

- 1539 tion. Retrieved from <https://www.datacamp.com/tutorial/complete>
- 1540 -guide-data-augmentation
- 1541 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquacul-
1542 ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-
1543 tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on*
1544 *mangrove-friendly aquaculture organized by the seafdec aquaculture depart-*
1545 *ment, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- 1546 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
1547 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
1548 Resources.
- 1549 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of
1550 waste cockle shell (*anadara granosa*) in biodiesel production from palm olein:
1551 Optimization using response surface methodology. *Fuel*, 90(7), 2353–2358.
1552 doi: 10.1016/j.fuel.2011.03.002
- 1553 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining*
1554 *mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127
1555 .v1
- 1556 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
1557 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A
1558 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1559 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/
1560 msq345
- 1561 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1562 proaches, challenges and opportunities for aquaculture. *Journal of Marine*
1563 *Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1564 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).

- 1565 Mass mortalities in bivalve populations: A review of the edible cockle
1566 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1567 10.1016/j.ecss.2014.04.011
- 1568 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1569 research in bivalves: Towards the identification of molecular markers of
1570 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1571 10.1016/j.jprot.2012.04.027
- 1572 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data
1573 augmentation for forecasting with deep learning. *ArXiv (Cornell University)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1574 1575 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly
Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1576 1577 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1578 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1579 10.1093/icb/ict076
- 1580 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1581 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma*
1582 *edule*) geographical traceability based on shell image analysis. *Ecological
1583 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1584 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1585 for chinese mitten crab using deep convolutional neural network. *Multi-
1586 media Tools and Applications*, 79(11-12), 7669–7684. doi: [https://doi.org/
1587 10.1007/s11042-019-08355-w](https://doi.org/10.1007/s11042-019-08355-w)
- 1588 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle
1589 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-
1590 posure. *Aquaculture*, 56(2), 151–162. Retrieved from <https://doi.org/>

- 1591 10.1016/0044-8486(86)90024-4 doi: 10.1016/0044-8486(86)90024-4
- 1592 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1593 ageing of eel, anguilla anguilla (l.)? *International Review of Hydrobiology*,
1594 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1595 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and*
1596 *anatomy for young learners*. Rosie Research. Retrieved from <https://rosieresearch.com/clam-dissection-anatomy/>
- 1598 Fao 2024 report: Sustainable aquatic food systems important for global food
1599 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1602 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1603 elemental composition: Complementary tools for stock discrimination of
1604 mulloway (argyrosomus japonicus) in southern australia. *Fish Research*,
1605 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1606 GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*.
1607 Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1609 GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn*. Re-
1610 trieval from <https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/> (Accessed: 2025-04-23)
- 1612 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Black-
1613 well Science.
- 1614 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean*
1615 *Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1616 Hussain, S. S., & Zaidi, S. S. H. (2024). Adaboost ensemble approach with

- 1617 weak classifiers for gear fault diagnosis and prognosis in dc motors. *Applied
1618 Sciences*, 14(7), 3105. doi: 10.3390/app14073105
- 1619 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1620 surface water characterization and heavy metal health risk on cockles in
1621 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1622 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1623 Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a com-
1624 prehensive guide to data rescaling*. DataCamp. Retrieved from [https://www
1626 .datacamp.com/tutorial/normalization-in-machine-learning](https://www
1625 .datacamp.com/tutorial/normalization-in-machine-learning) (Ac-
cessed 2025-04-23)
- 1627 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The ear-
1628 liest example of sexual dimorphism in bivalves—evidence from the astar-
1629 tid *Nicaniella* (lower jurassic, southern germany). *Journal of Paleontology*,
1630 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1631 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1632 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1633 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1634 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1635 learning-based phenotype classification of three ark shells: Anadara kagoshi-
1636 mensis, tegillarca granosa, and anadara broughtonii. *Frontiers in Marine
1637 Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1638 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the
1639 granulated ark, *tegillarca granosa* (linne). *The Korean Journal of Malacol-
1640 ogy*, 13, 55–64.
- 1641 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as
1642 a tool for stock identification of the southern blue whiting, *micromesistius*

- 1643 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1644 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1645 et al. (2016). Identifying blue whiting (*micromesistius poutassou*) stock
1646 structure in the northeast atlantic by otolith shape analysis. *Canadian*
1647 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1648 cjfas-2015-0332
- 1649 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1650 *tegillarca granosa* (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.*
1651 *Arts Sci*, 4.
- 1652 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and
1653 larval rearing of the blood cockle, *tegillarca granosa* (linnaeus, 1758), in the
1654 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2
1655 .001
- 1656 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of
1657 local populations of the common sole *solea solea* (pisces, soleidae) in the nw
1658 mediterranean through otolith morphometrics and shape analysis. *Marine*
1659 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1660 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and
1661 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–
1662 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/
1663 kja.21209
- 1664 Narasimham, K. A. (1988). Taxonomy of the blood clams *anadara* (*tegillarca*)
1665 *granosa* (linnaeus, 1758) and a. (t.) *rhombea* (born, 1780).
- 1666 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1667 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1668 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500

- 1669 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1670 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1671 doi: 10.1002/jmor.10439
- 1672 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1673 morphometrics and machine learning challenge currently accepted species
1674 limits of the land snail *placostylus* (pulmonata: Bothriembryontidae) on the
1675 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1676 doi: 10.1093/mollus/eyz031
- 1677 Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test.*
1678 Retrieved from https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/ (Accessed: 2025-04-23)
- 1680 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,
1681 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1682 carbons from consumption of blood cockle and exposure to contaminated
1683 sediments and water along the klang strait, malaysia. *Marine Pollution
1684 Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1685 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1686 of seasonal variation and anthropogenic stress on blood cockle (*tegillarca*
1687 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-
1688 pogenic Stress on Blood Cockle (Tegillarca Granosa) Production Potential*,
1689 44(2), 62–82.
- 1690 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1691 chine learning and its applications to biology. *PLoS Computational Biology*,
1692 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1693 Team, K. (n.d.). *Keras documentation: Reducelronplateau*. Retrieved from
1694 https://keras.io/api/callbacks/reduce_lr_on_plateau/

- 1695 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1696 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1697 & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).
1698 College Park, MD: Maryland Sea Grant.
- 1699 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1700 approach for morphological feature extraction based on variational auto-
1701 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*
1702 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1703 Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in*
1704 *the laboratory, malaysia.* (Handle.net) doi: 10.3366/in_3366.pdf
- 1705 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1706 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972
1707 .5008949
- 1708 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1709 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1710 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-
1711 pacts of four commonly used nanoparticles on the metabolism of a ma-
1712 rine bivalve species, tegillarca granosa. *Chemosphere*, 296, 134079. doi:
1713 10.1016/j.chemosphere.2022.134079
- 1714 Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1715 blood clam tegillarca granosa. *Marine Environmental Research*, 177. Re-
1716 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

¹⁷¹⁷ Appendix A

¹⁷¹⁸ Code Snippets

¹⁷¹⁹ **Appendix B**

¹⁷²⁰ **Resource Persons**

¹⁷²¹ **Dr. Victor Marco Emmanuel N. Ferriols**

¹⁷²² Provided blood cockles samples used in this study

¹⁷²³ Director, University of the Philippines Institute of Aquaculture

¹⁷²⁴ vnferriols@up.edu.ph

¹⁷²⁵

¹⁷²⁶ **Ms. Allena Esther D. Arteta**

¹⁷²⁷ Performed spawning of blood cockles samples, assisted the researchers with dissection and sex identification.

¹⁷²⁹ Research Associate, Institute of Aquaculture

¹⁷³⁰ adarteta@up.edu.ph

¹⁷³¹

1732 Ms. LC May C. Gasit

- 1733 Performed spawning of blood cockles samples, assisted the researchers with the
1734 dissection and sex identification
1735 Research Associate, Institute of Aquaculture
1736 lcgasit@up.edu.ph

1737

1738 Sheila G. Untalan

- 1739 Performed spawning of blood cockles samples, assisted the researchers with the
1740 dissection and sex identification
1741 Research Associate, Institute of Aquaculture

1742

1743 Joel M. Fabrigas

- 1744 Assisted the researchers with the dissection and sex identification
1745 Hatchery Staff, Institute of Aquaculture

1746

1747 Paul Andre M. Lopez

- 1748 Assisted the researchers with the dissection and sex identification
1749 Hatchery Staff, Institute of Aquaculture

1750

¹⁷⁵¹ Appendix C

¹⁷⁵² Data Gathering Documentation



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



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



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

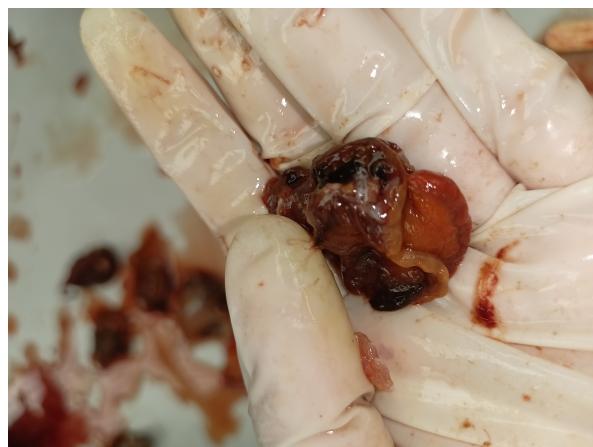
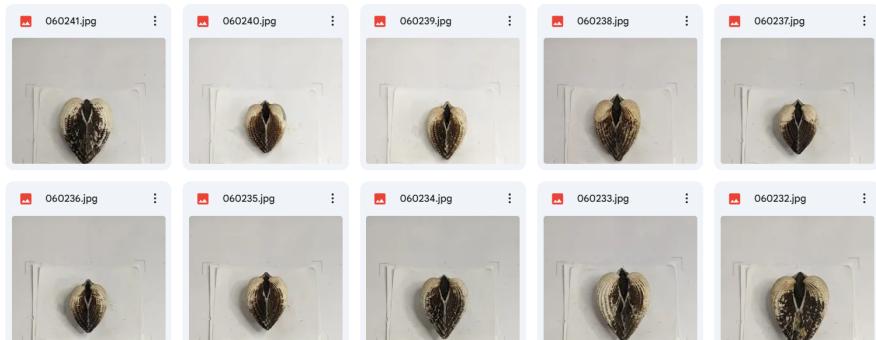
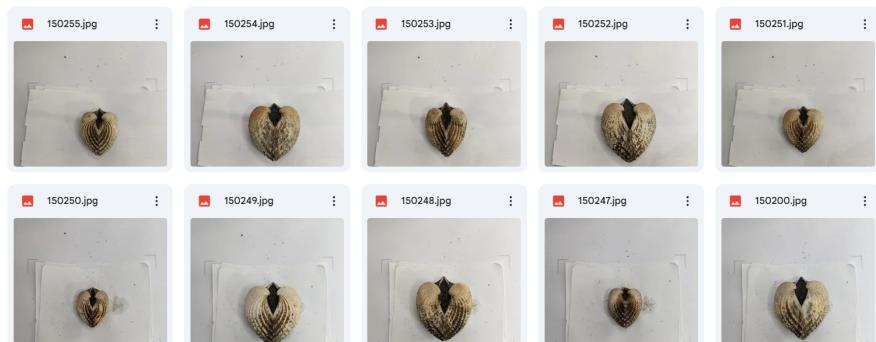


Figure C.4: Sex Identified Male Through Dissection of *T. granosa*

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 *T. 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 *T. granosa*Figure C.7: Captured Images of Female *T. granosa*Figure C.8: Captured Images of Male *T. granosa*