

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 sex
203 identification is crucial for maintaining a balanced male-to-female ratio, support-
204 ing sustainable harvesting, and improving resource management. However, macro-
205 scopically identifying sex through shell morphology is challenging, and there are
206 currently no available technologies for non-invasive sex classification. This study
207 explores the use of machine learning and deep learning techniques to classify the
208 sex of blood cockles based on shell measurements (length, width, height, hinge
209 line length, distance between the umbos, and rib count) and images taken from
210 various angles (dorsal, ventral, anterior, posterior, and lateral views). 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. Moreover, deep learning
213 using 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 view. These results demonstrate the potential
216 of a non-invasive approach to sex identification, supporting sustainable aquacul-
217 ture practices and offering a baseline for further research using computer vision
218 and 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	57
268	4.2.3 Training Result and Hyperparameter Tuning	58
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	66
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	87
279	C Data Gathering Documentation	89

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 that were gathered from the shell of <i>T. granosa</i>	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 Dataset 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.	55
296	4.4 Training and validation accuracy per fold.	62
297	4.5 Average training and validation accuracy across folds.	62
298	4.6 Training and validation loss per fold.	63
299	4.7 Average training and validation loss across folds.	64
300	4.8 Confusion matrix for final model predictions.	65
301	4.9 ROC curve with AUC score for the proposed model.	66
302	A.1 Feature engineering used to create and transform the dataset for 303 machine learning analysis.	81
304	A.2 Feature importance scores derived from the Kruskal-Wallis test to 305 identify the most significant variables.	82
306	A.3 Random undersampling applied in machine learning to address 307 class imbalance.	82
308	A.4 Five-fold cross-validation used to evaluate and tune machine learn- 309 ing model performance.	83
310	A.5 Resizing images to 256x256 pixels for consistent input dimensions.	83

311	A.6 Processing the images to remove the shadows.	84
312	A.7 Random undersampling applied in deep Learning to balance class distribution in the datasets.	85
314	A.8 On-the-fly data augmentation used to create variety of random transformation to increase training images.	85
315		
316	A.9 CNN architecture used for training the image dataset.	86
317		
C.1	Sex identification through spawning of <i>T. granosa</i>	89
318		
C.2	Sex-based separation of <i>T. granosa</i> samples post-spawning.	90
319		
C.3	Sex identified female through dissection of <i>T. granosa</i>	90
320		
C.4	Sex identified male through dissection of <i>T. granosa</i>	90
321		
C.5	Linear measurements of female <i>T. granosa</i>	91
322		
C.6	Linear measurements of male <i>T. granosa</i>	91
323		
C.7	Captured images of female <i>T. granosa</i>	92
324		
C.8	Captured images of male <i>T. granosa</i>	92

List of Tables

326	2.1 Comparison of the methods used in bivalves studies.	27
327	3.1 Architecture of the convolutional neural network used.	44
328	4.1 Mann-Whitney U test results for sex-based feature comparison. .	51
329	4.2 Performance metrics for models with all 13 features.	53
330	4.3 Performance metrics for models with 5 features.	54
331	4.4 Performance metrics for unbalanced vs. balanced datasets (Batch	
332	Size: 16, Epochs: 20).	56
333	4.5 Performance metrics for individual and combined angles (Batch	
334	Size: 16, Epochs: 20).	57
335	4.6 Effect of batch size and epoch values on CNN model performance.	59
336	4.7 Performance metrics for different activation functions (Batch Size:	
337	32, Epochs: 50).	59

338	4.8 Per-fold performance metrics (Batch Size: 32, Epochs: 50, Activation Function: ReLU).	60
340	4.9 Learning rate reductions, early stopping, and best epochs per fold during 5-fold cross-validation.	61
341		

³⁴² Chapter 1

³⁴³ Introduction

³⁴⁴ 1.1 Overview

³⁴⁵ The Philippines is a global center of marine biodiversity and has established aquaculture as a significant contributor to total fishery production (Aypa & Baconguis, 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annually and is the 11th largest seafood producer in the world. Aquaculture is deeply integrated into Filipinos' livelihoods, encompassing fish cultivation and the production of various aquatic species, including bivalves. Among these, blood cockles (*Tegillarca granosa*) hold considerable economic and environmental significance, making it essential to ensure sustainable production and population balance.

³⁵³ Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent overharvesting and ensure sustainability. An imbalanced ratio can lead to overexploitation and negatively impact the population's viability. However, there is

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

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

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

371 1.2 Problem Statement

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

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

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

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

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

402 1.3 Research Objectives

403 1.3.1 General Objective

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

409 1.3.2 Specific Objectives

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

- 413 1. to collect and organize a comprehensive dataset of *T. granosa*, which will
414 include linear measurements and images captured from different camera an-
415 gles that will serve as the basis for training and evaluating the machine
416 learning and deep learning models,
- 417 2. to develop and implement machine learning and deep learning models that
418 can classify the sex of *T. granosa* based on the collected linear measurements
419 and images of different camera angles of the sample, and determine the best
420 performing models, and
- 421 3. to evaluate the model using performance metrics such as accuracy, precision,

422 recall, F1 Score, and AUC-ROC score for deep learning, and improve it by
423 performing hyperparameter optimization.

424 1.4 Scope and Limitations of the Research

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

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

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

⁴⁴⁴ et al., 2021). Thus, the researchers only focused on the samples undergoing the
⁴⁴⁵ spawned stage.

⁴⁴⁶ During the data collection, the researchers personally gathered linear measure-
⁴⁴⁷ ments, including length, width, height, rib count, hinge line length, and distance
⁴⁴⁸ between the umbos through the vernier caliper. The data gathering process was
⁴⁴⁹ supervised by the University Research Associates from the Institute of Aquacul-
⁴⁵⁰ ture, College of Fisheries and Ocean Sciences. Aside from linear measurements,
⁴⁵¹ images were taken from six different angles. The process of linear measurements
⁴⁵² and image collection were non-invasive, considering the blood cockle-built ability
⁴⁵³ to survive in low oxygen environments and naturally inhabit intertidal mudflats
⁴⁵⁴ (Zhan & Bao, 2022).

⁴⁵⁵ The method developed in this study is specific to *T. granosa* and may not ap-
⁴⁵⁶ ply to other bivalve species. The model was trained exclusively for *T. granosa*
⁴⁵⁷ and morphometric and morphological features, which may not be consistent and
⁴⁵⁸ applicable across other shellfish species.

⁴⁵⁹ 1.5 Significance of the Research

⁴⁶⁰ This study will give us a significant advancement in non-invasive sex identification
⁴⁶¹ methods in *Tegillarca granosa*, providing innovative solutions that has the poten-
⁴⁶² tial to address the challenges in identifying sex and reshape sustainable approaches
⁴⁶³ to aquaculture. The significance of this study extends to the following:

⁴⁶⁴ *Research Institution.* The result of this study focusing on the sex-identification

⁴⁶⁵ mechanism of bivalves, specifically *T. granosa*, will provide valuable insights into
⁴⁶⁶ universities and research centers that focus on fisheries and coastal management,
⁴⁶⁷ such as the UPV Institute of Aquaculture, that aim to develop sustainable devel-
⁴⁶⁸ opment and suitable culture techniques.

⁴⁶⁹ *Fishermen.* By developing a non-invasive method in sex identification, this study
⁴⁷⁰ can help long-term harvest efficiency and maintain the ratio of the harvest which
⁴⁷¹ can help prevent exploitation of the *T. granosa*.

⁴⁷² *Coastal Communities.* The result of this study would be beneficial for the coastal
⁴⁷³ communities that are reliant on their source of income with aquaculture com-
⁴⁷⁴ modities like blood cockles. Maintaining the diversity and aspect ratio of male
⁴⁷⁵ and female may increase the market value of blood cockle production since cockle
⁴⁷⁶ aquaculture faces significant obstacles worldwide due to the fluctuating seed sup-
⁴⁷⁷ plies and scarcity of broodstock from the wild.

⁴⁷⁸ *Future Researchers.* The result of this study would serve as the basis for studies
⁴⁷⁹ that involve sex identification in bivalves such as *T. granosa*. Some technologies
⁴⁸⁰ are yet to be explored in machine learning and deep learning technologies that
⁴⁸¹ can lead to higher accuracy and distinguish the presence of sexual dimorphism in
⁴⁸² the *T. granosa*.

⁴⁸³ Chapter 2

⁴⁸⁴ Review of Related Literature

⁴⁸⁵ Aquaculture is the fastest-growing industry in animal food production and has
⁴⁸⁶ great potential as a sustainable solution to global food security, nutrition, and
⁴⁸⁷ development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
⁴⁸⁸ *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
⁴⁸⁹ tegrated into the livelihoods of Filipinos, not only through fish cultivation but
⁴⁹⁰ also through the production of other aquatic species, including mollusks, oysters,
⁴⁹¹ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
⁴⁹² clams *Tegillarca granosa*, have economic and environmental significance. It has
⁴⁹³ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
⁴⁹⁴ vesting and maintain the optimal ratio to preserve the population and production
⁴⁹⁵ of the blood cockles.

⁴⁹⁶ The members of the Arcidae Family, including *T. granosa* are important sources
⁴⁹⁷ of food and livelihood. Cockle aquaculture meets rising demands, however, it
⁴⁹⁸ faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,

⁴⁹⁹ 2023). To solve the problem, researchers exert a considerable amount of effort,
⁵⁰⁰ developing a broader understanding of bivalves, including their sex-determining
⁵⁰¹ mechanism, due to their notable importance in terms of diversity, environmental
⁵⁰² benefits, and economic and market importance (Breton et al., 2017). Despite the
⁵⁰³ promising idea of identifying sex, there is limited research reported in terms of
⁵⁰⁴ sexual dimorphism, making it harder to distinguish through its morphological and
⁵⁰⁵ morphometric characteristics.

⁵⁰⁶ By addressing the challenges in the sex identification of *T. granosa*, it would be
⁵⁰⁷ able to address one problem at a time. Currently, there are no recent documented
⁵⁰⁸ publications that integrate machine learning and deep learning in characterizing
⁵⁰⁹ sexual dimorphism, reducing complexity, variability in sex determination, and
⁵¹⁰ differentiation mechanisms in bivalves, including *T. granosa* specifically.

⁵¹¹ **2.1 Background on *T. granosa* and Their Im- 512 portance**

⁵¹³ *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.
⁵¹⁴ In the Philippines, it is known locally as Litob and Bakalan, a marine bivalve
⁵¹⁵ species from the family Arcidae. Litob is widely distributed in the world including
⁵¹⁶ Southeast Asia. They can be found in the intertidal mudflats adjacent to the
⁵¹⁷ mangrove forest (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With
⁵¹⁸ the intertidal mudflat as *T. granosa*'s habitat, they experience severe hypoxia
⁵¹⁹ or low oxygen levels in the blood tissues during the tidal cycle. The blood clams
⁵²⁰ exhibit a unique red-blood phenotype where it serves two purposes the hemocyte

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

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

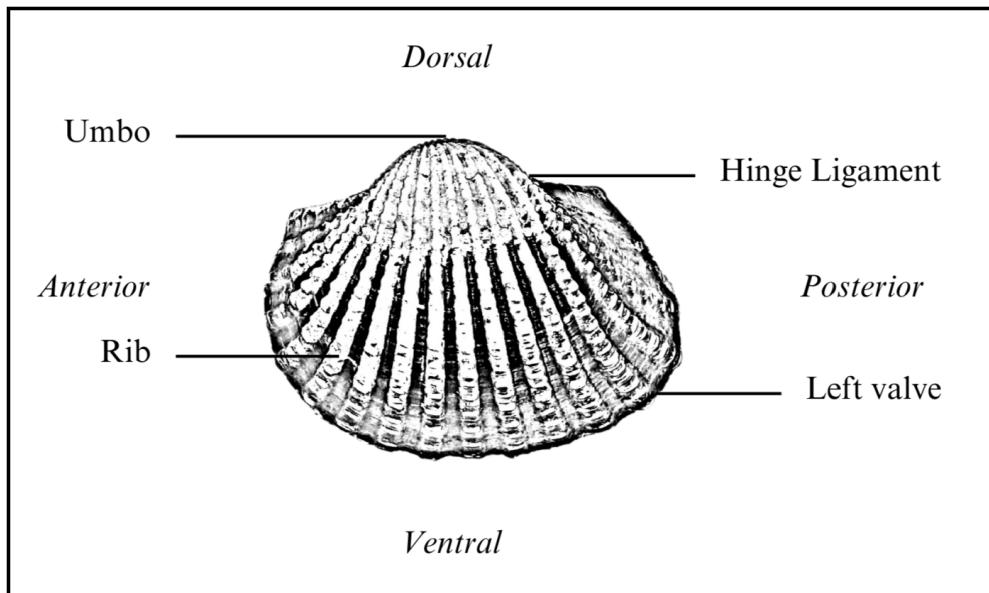


Figure 2.1: Diagram of *T. granosa*'s external anatomy.

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

554 Maniam, Hamid, & Ali, 2011).

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

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

577 2.2 Sex Identification Methods in *T. granosa*

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

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

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

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

611 The classification of the gonad stages used was by Yurimoto et al. (2014). There
612 are five maturation stages of gonadal development: immature (Stage I), devel-
613 oping (Stage II), mature (Stage III), spawning (Stage IV), and spent (Stage V)
614 stages. The sex of the *T. granosa* was confirmed by the color of the gonad and
615 by conducting a histological examination of the gonads. During the immature
616 stage, sex determination was indistinguishable due to the difficulties of observing
617 the germ cells. In the developing stage, the spermatocytes and a few spermatids
618 can be seen for males, and immature oocytes are attached to the tube wall for
619 the female. In the mature stage, the follicles are full of spermatozoa with their
620 tails pointing towards the center of the tube for the male, and the female is full
621 of mature oocytes that are irregular or polygonal in shape with the oval nucleus.
622 Upon reaching spawning, some spermatozoa are released, causing the empty space
623 in the follicle wall for males and females. There is a decrease in the number of
624 mature oocytes and it exhibits nuclear disappearance due to the breakdown of
625 the germinal vesicle. Lastly, the spent stage is where the genital tube is deformed

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

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

2.3 Machine Learning and Deep Learning in Biology

Machine learning has the potential to improve the quality of life of human beings 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 by Kim et al. (2024), the researchers utilized three (3) convolutional neural network (CNN) models: the Visual Geometry Group Network (VGGnet), the Inception Residual Network (ResNet), and the SqueezeNet. 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, Yang, Cha, Jung, & Kim, 2024).

671 **2.3.2 Geometric Morphometrics and Machine Learning for**
672 **Species Delimitation**

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

683 **2.3.3 Contour Analysis in Mollusc Shells Using Machine**
684 **Learning**

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

693 that, the researchers analyzed two geographically and ecologically separated pop-
694ulations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed
695that contour analysis of gastropod shells with enlarged spines can be analyzed
696using both methodologies, but the wavelet analysis provided better local discrim-
697ination. From an ecological perspective, shells with various sizes of spines in both
698areas indicate the broad adaptability of the species.

699 **2.3.4 Machine Learning for Shape Analysis of Marine Or- 700 ganisms**

701 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*
702 *to Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*
703 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to
704 the traditional method of collecting linear measurements with the application of
705 multivariate statistical methods and outline analysis in recording the structures
706 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and
707 elasmobranchs) with their hard bodies have been used as an indication of age and
708 a determinable time-scale and structure continue to go through life (Arkhipkin,
709 2005; Kerr & Campana, 2014). This study has explored variations in the mor-
710 phometry of recording structures in stock discrimination and systematics. The
711 researchers utilized the principal component analysis rather than the traditional
712 approach, which helps simplify the data without losing important information.
713 They utilized landmark-based geometric morphometrics, which has three differ-
714 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other
715 morphogenetic processes, and lastly, the extremal points are constructed land-

⁷¹⁶ marks.

⁷¹⁷ Generalized Procrustes Analysis (GPA) is a common superimposition technique in
⁷¹⁸ landmark-based geometric morphometrics that aligns landmarks via translation,
⁷¹⁹ scaling, and rotation to eliminate non-shape deviations (Zelditch, Swiderski, &
⁷²⁰ Sheets, 2004). However, there is a limit to the amount of smooth areas that may
⁷²¹ be captured, and it is possible to overlook significant shape details. Utilization
⁷²² of the semi-landmarks enhanced the shape description (Adams, Rohlf, & Slice,
⁷²³ 2004). The researchers observed that using an outline-based approach would be
⁷²⁴ more effective than using a landmark-based approach.

⁷²⁵ Another approach is the Fourier analysis which is a curve-fitting approach com-
⁷²⁶ monly used due to its well-known mathematical background and how general
⁷²⁷ functions can be decomposed into trigonometric or exponential functions with
⁷²⁸ definite frequencies. It has two main approaches, namely: Polar Transform (PT)
⁷²⁹ in which it expresses the outline using equally spaced radii, and Elliptical Fourier
⁷³⁰ Analysis (EFA) which separately analyzes the x and y coordinates of the shape.
⁷³¹ The PT works for simple rounded outlines and has the tendency to miss details
⁷³² in more complex shapes, unlike the EFA which can handle complex, convoluted
⁷³³ outlines (Zahn & Roskies, 1972; Doering & Ludwig, 1990; Ponton, 2006). Many
⁷³⁴ researchers view EFA as the most effective Fourier method for providing a compre-
⁷³⁵ hensive and detailed description of recording structures (Mérigot, Letourneau, &
⁷³⁶ Lecomte-Finiger, 2007; Ferguson, Ward, & Gillanders, 2011; Leguá, Plaza, Pérez,
⁷³⁷ & Arkhipkin, 2013; Mahé et al., 2016).

⁷³⁸ Landmark-based methods used in the study showed that there are detectable
⁷³⁹ differences between male and female octopuses. However, the accuracy of deter-

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

747 **2.3.5 Deep Learning for Landmark-Free Morphological Fea-
748 ture Extraction**

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

759 **2.3.6 Machine Learning for Sex Differentiation in Abalone**

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

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

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

782 The study found that Naive Bayes consistently performed better than other algo-
783 rithms. However, when applied to both min-max and zero-mean normalization,
784 the average accuracies of the algorithms were as follows: Random Forest (62.37%),
785 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear kernel
786 (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with nor-

787 malization, Random Forest achieved the highest overall metrics, including an av-
788 erage balanced accuracy of 787%, sensitivity of 66.43%, and specificity of 83.31%.
789 Liu et al. concluded that Random Forest is highly accurate because it can handle
790 large, complex datasets, run processes in parallel using multiple trees, and select
791 the most relevant features to enhance model performance (Arifin et al., 2021).

792 **2.3.7 Machine Learning for Geographical Traceability in
793 Bivalves**

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

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

805 The researchers developed a non-invasive image-based traceability technique, an
806 alternative to the chemical and biochemical analysis of the bivalves. It was able
807 to incorporate machine learning methods to promote lesser human intervention.
808 The researchers discovered that BivalveNet emerged as the superior model for
809 bivalves with 96.91% accuracy which is comparable to the accuracy of the de-

810 structive methods with 97% and 97.2% accuracy rates. The result of the study
811 aided the researchers in concluding that there is a possibility of on-site evalua-
812 tion of the bivalve through the implementation of a mobile app that would allow
813 the public and official entities to obtain information regarding the provenance of
814 seafood products' traceability because of its non-invasive and image-based aspects
815 (Concepcion et al., 2023).

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

820 **2.4 Limitations on Sex Identification in *T. gra-***

821 ***nosa***

822 To date, no distinction has been made between the male and female *T. granosa*
823 in sexing methodology. In cockle aquaculture without clearly apparent sexual
824 dimorphism, sexing can be performed using invasive methods such as chemical
825 stimulation, dissection, and gonad-stripping. Induced spawning, specifically tem-
826 perature shock, is the most natural and least invasive method for bivalves (Aji,
827 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water
828 from hot to cold with a specific temperature requires deliberate and careful ma-
829 nipulation of the temperature over a specific period and would require constant
830 management and monitoring.

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

839 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous (syn-
840 chronous or functional) characteristics. Sequential hermaphrodites switch genders
841 after being male or female for one or multiple yearly cycles. (Heller, 1993; Gosling,
842 2004; Collin, 2013). Sex change and consecutive hermaphroditism have been ob-
843 served in different bivalve species, including Ostreidae, Pectinidae, Veneridae,
844 and Patellidae. However, macroscopically differentiating bivalve sex is challeng-
845 ing. The only way it may be identified is through histological analysis of gonad
846 remains but to do so there is an act of killing the organism (Coe, 1943; Gosling,
847 2004). Verification of sex change in bivalves to classify whether male or female
848 while they are alive is challenging since they need to be re-confirmed and re-
849 evaluated to be the same individual after a year.

850 Lee et al. (2012) found out that *T. granosa*, a species in Arcidae, has been dis-
851 covered to be a sequential hermaphrodite, with the sex ratio changing with an
852 increase in the shell size. In bivalves, sex changes usually happen when the gonad
853 is not differentiated between spawning seasons (Thompson, Newell, Kennedy, &
854 Mann, 1996). But in *T. granosa*, after the spawning season, sex changes during
855 its inactive phase. Results showed a 15.1% sex change ratio, with males having

856 a higher sex change ratio (21.2%) than females (6.2%). The 1+ year class had a
857 higher ratio (17.8%) than the 2+ year class (12.1%). Thus, this study indicates
858 that *T. granosa* is a sequential hermaphrodite. The results of the study demon-
859 strated that the bivalve's age affects the sex ratio and degree of sex change, but
860 additional in-depth investigation is required to determine the role that genetic
861 and environmental factors play in these changes.

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

869 2.5 Chapter Summary

870 This section of the paper summarizes the technologies used in the different studies
871 related to the pursuit of the study entitled, Morphometric and Morphological-
872 Based Non-Invasive Sex Identification of Blood Cockle, *Tegillarca granosa* (Lin-
873 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.

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

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

⁸⁸⁹ **Chapter 3**

⁸⁹⁰ **Research Methodology**

⁸⁹¹ This chapter discusses the materials and methods employed in the study, focus-
⁸⁹² ing on the development requirements, as well as the software and programming
⁸⁹³ languages utilized. It also detailed the overall workflow in conducting the study,
⁸⁹⁴ Morphometric and Morphological-Based Non-Invasive Sex Identification of Blood
⁸⁹⁵ Cockle, *Tegillarca granosa* (Linnaeus, 1758) using machine learning and deep
⁸⁹⁶ learning technologies.

⁸⁹⁷ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture, over-
⁸⁹⁸ saw the overall workflow by providing baseline characteristics of the samples that
⁸⁹⁹ the researchers could focus on. Additionally, guidance was offered by the re-
⁹⁰⁰ search associates LC Mae Gasit and Allena Esther Artera. Consequently, the
⁹⁰¹ entire dataset collection process was conducted at the University of the Philip-
⁹⁰² pines Visayas hatchery facility.

⁹⁰³ 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

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

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

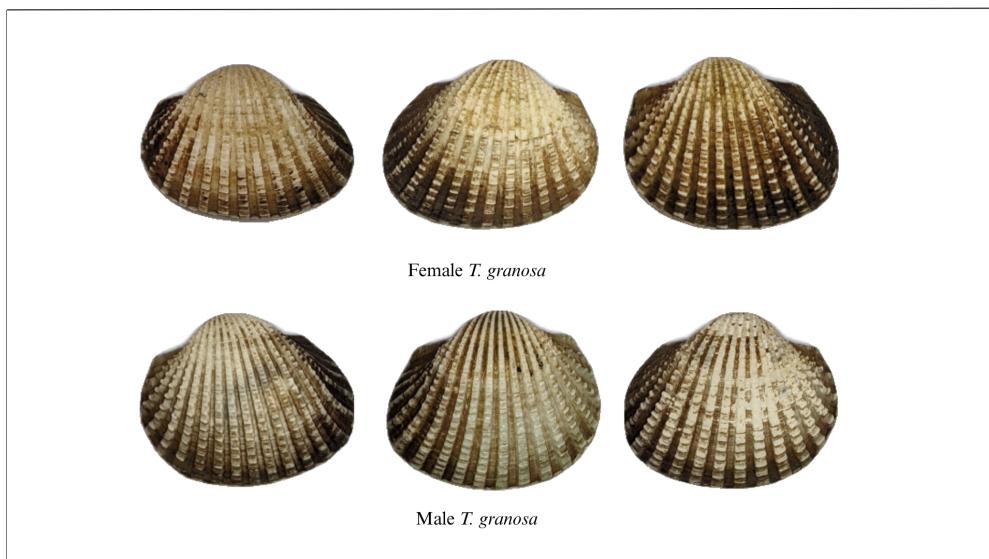


Figure 3.1: Male and female *T. granosa* shells.

932 3.2 Ethical Considerations

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

938 Philippines Visayas.

939 3.3 Creating *T. granosa* Dataset

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

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

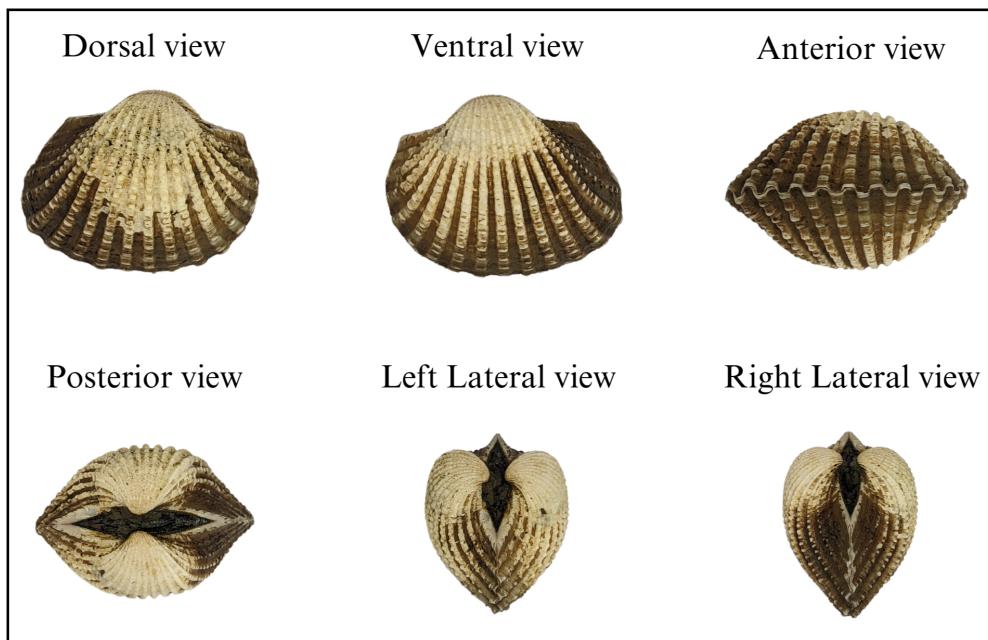


Figure 3.2: Different views of the *T. granosa* shell captured

⁹⁵⁹ 3.4 Morphometric Data Collection

⁹⁶⁰ Morphology refers to biological form and is one of the most visually recognizable
⁹⁶¹ phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).
⁹⁶² In this study, morphological characteristics describe the structural features of
⁹⁶³ *T. granosa*, focusing on measurable attributes such as shape, size, and color.
⁹⁶⁴ Morphometric characteristics, on the other hand, refer to specific quantifiable
⁹⁶⁵ features of *T. granosa*, including length, width, height, hinge line length, distance
⁹⁶⁶ between the umbos, and rib count. As stated by the researchers, quantifying and
⁹⁶⁷ characterizing these traits is essential for understanding and visualizing variations
⁹⁶⁸ in *T. granosa* morphology.

⁹⁶⁹ The researchers measured the height, width, and length of *T. granosa* using a
⁹⁷⁰ 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.

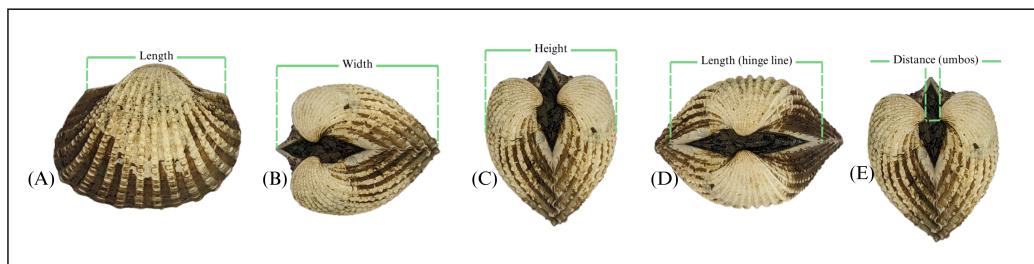


Figure 3.3: Linear measurements that were gathered from the shell of *T. granosa*.

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

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

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



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

995 3.6 Hardware and Software Configuration

996 This section of the paper discusses the software, programming languages, and tools
997 used for sex identification. Data collection, preprocessing, and model training
998 were conducted on a Windows 11 operating system using an ACER Aspire 3
999 general-purpose unit (GPU) equipped with an AMD Ryzen 3 7320U CPU with
1000 Radeon Graphics (8 cores) @ 2.395 GHz and 8 GB of RAM. Google Colaboratory
1001 was utilized for collaborative preprocessing, computer vision tasks, and model

1002 training. Image preprocessing was performed using computer vision techniques in
1003 Python, while machine learning and deep learning models were developed using
1004 Python libraries, including Keras. The results of the gathered measurements were
1005 stored and managed using spreadsheet software. GitHub was employed for version
1006 control, documentation, and activity tracking throughout the study.

1007 **3.7 Machine Learning on Morphometric Data**

1008 This section of the paper discusses the machine learning operations that served
1009 as a baseline prior to implementing more complex deep learning methods for
1010 image classification. The study utilized collected variables including linear mea-
1011 surements—length, width, height, hinge line length, distance between the um-
1012 bos, and rib count—along with derived features used as predictors. These in-
1013 cluded the length-to-width ratio, length-to-height ratio, width-to-height ratio,
1014 umbos distance-to-length ratio, hinge line length-to-length ratio, umbos distance-
1015 to-height ratio, and rib density. The samples were classified by sex, with females
1016 labeled as 0 and males as 1, which served as the response variable.

1017 3.7.1 Data Preprocessing

1018 The preprocessing of the dataset involved several essential steps, carried out using
 1019 Python in Google Colaboratory, in preparation for machine learning analysis (*see*
 1020 *Figure 3.5*).

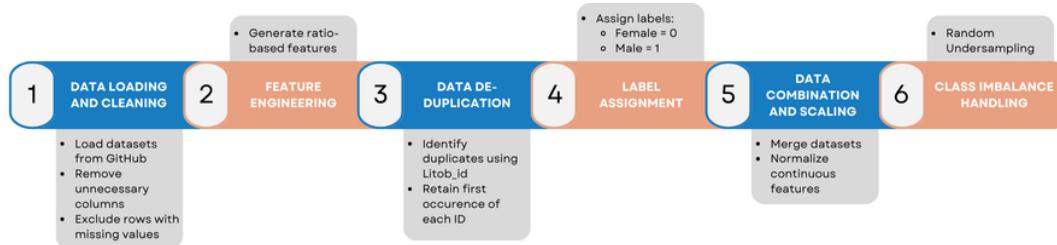


Figure 3.5: Data preprocessing in machine learning pipeline.

1021 *Data Loading and Cleaning*

1022 The process began by loading two separate datasets for male and female *T. granosa*
 1023 directly from GitHub using `pd.read_csv()`. Unnecessary columns were removed,
 1024 and rows containing missing values were excluded using the `dropna()` function to
 1025 ensure data completeness and reliability.

1026 *Feature Engineering*

1027 Additional ratio-based features were generated to augment the existing measurements.
 1028 These included the length-to-width ratio, length-to-height ratio, width-to-height ratio,
 1029 hinge line length-to-length ratio, umbos distance-to-length ratio, umbos distance-to-height ratio,
 1030 and rib density. These derived features aimed to emphasize shape characteristics independent of size, improving the models' ability
 1031 to distinguish morphological differences between sexes.
 1032

1033 ***Data De-duplication***

1034 To avoid redundancy and ensure each specimen was uniquely represented, the
1035 last three digits of each `Litob_id` were used to identify duplicates. Only the first
1036 occurrence of each unique ID was retained, reducing potential bias caused by
1037 repeated entries.

1038 ***Label Assignment***

1039 A new column labeled `Label` was added to both datasets. Female specimens were
1040 assigned a label of 0, and male specimens a label of 1. This column served as the
1041 target variable for classification.

1042 ***Data Combination and Scaling***

1043 After cleaning and feature engineering, the male and female datasets were merged
1044 into a single DataFrame. The `Litob_id` column was removed post de-duplication.
1045 All continuous numeric features were normalized using `MinMaxScaler` to scale
1046 values to the range [0, 1].

1047 Rib count was excluded from normalization because it is a discrete feature with
1048 biologically meaningful bounds. According to best practices in machine learning,
1049 normalizing discrete or categorical features can distort their meaning and is often
1050 unnecessary (Jaiswal, 2024). In this study, rib count was treated as a categorical
1051 attribute due to its biological significance and finite, non-continuous nature.

1052 ***Class Imbalance Handling***

1053 After normalization, class imbalance was addressed by applying Random Under-
1054 sampling to the male dataset. This technique randomly reduced the number of

1055 male samples to match the number of female samples (127 each), ensuring equal
1056 class representation. By using this approach, model bias was minimized, and the
1057 classification performance became more reliable across both classes.

1058 3.7.2 Machine Learning Models Training

1059 *Model Selection and Hyperparameter Tuning*

1060 To establish a baseline for classification, various models were evaluated: Logis-
1061 tic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest,
1062 AdaBoost, Extra Trees, and Gradient Boosting. Hyperparameter tuning was con-
1063 ducted using `GridSearchCV`, which systematically identified the optimal settings
1064 for each model to enhance accuracy and performance.

1065 *Cross-Validation*

1066 A five-fold cross-validation approach was implemented (*refer to Figure 3.6*). The
1067 dataset was divided into five subsets, with four used for training and one for
1068 testing. This process was repeated five times, with each fold serving as the test set
1069 once. This method ensured that model evaluation was robust and generalizable,
1070 minimizing the bias that may result from a single train-test split. (GeeksforGeeks,
1071 2024)

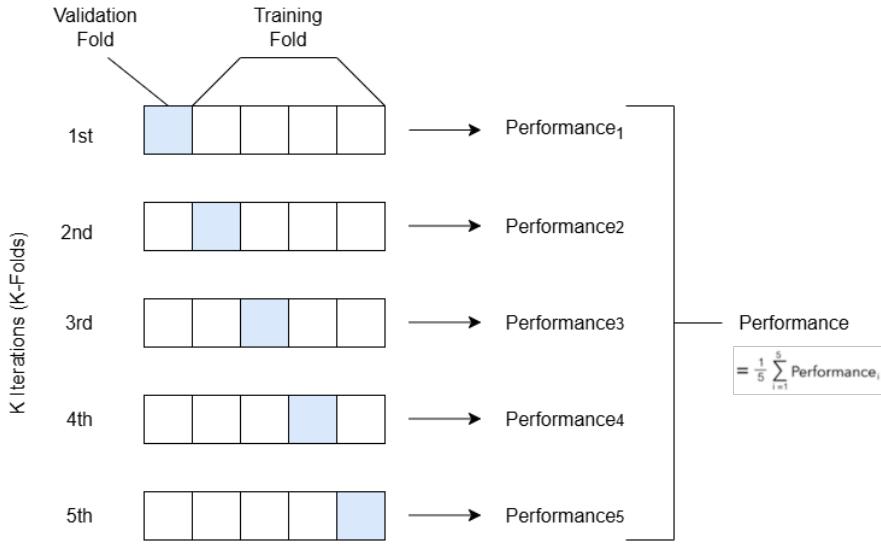


Figure 3.6: Diagram of k-fold cross-validation with $k = 5$.

1072 3.8 Deep Learning for Morphological Analysis

1073 This section outlines the application of deep learning techniques in analyzing the
1074 morphological characteristics of *Tegillarca granosa* to identify their sex based on
1075 shell images. A Convolutional Neural Network (CNN) architecture was imple-
1076 mented and trained on preprocessed images using stratified cross-validation.

1077 ***Image Preprocessing***

1078 This subsection details the image processing techniques applied to raw shell images
1079 of *T. granosa* using computer vision methods before training the deep learning
1080 model. The image preprocessing techniques include standardizing input dimen-
1081 sions and removing shadows, background, and noise. Each image underwent data
1082 augmentation to enhance feature visibility for effective learning. Image prepro-
1083 cessing ensures consistent and high-quality input data for model training.

1084 *Adjusting Dimensions*

1085 All images were resized to a consistent dimension of 256x256 pixels to ensure
1086 uniformity throughout the dataset. This standardization is essential for Convo-
1087 lutional Neural Networks (CNNs), as a consistent input dimension is required.
1088 While resizing, the aspect ratio was maintained to prevent distortion of the mor-
1089 phological features, and padding was added to retain the original format.

1090 *Background Removal*

1091 Background removal was performed to maintain a consistent white background
1092 throughout the dataset. The tool `rembg` was used to efficiently remove the original
1093 background, retaining the foreground from the raw images. This method resulted
1094 in clear images with a white background, enhancing focus on the morphological
1095 features and defining the shell boundaries.

1096 *Shadow Removal*

1097 To minimize noise caused by shadows around the shell, HSV thresholding, con-
1098 tours, and morphological thresholds were applied to isolate and remove shadowed
1099 regions. This approach preserved the natural color of the blood cockles and elim-
1100 inated shadows and noise from the surrounding area (*see Figures 3.7 and 3.8*).

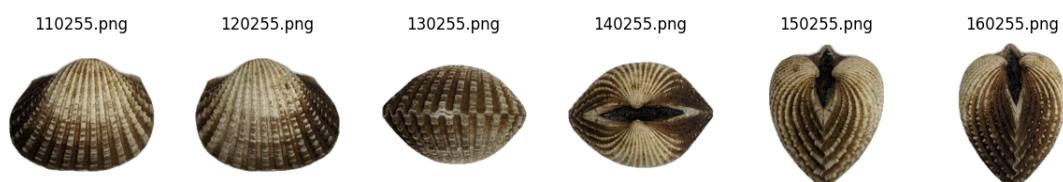


Figure 3.7: Shadows removed from male samples at different angles.

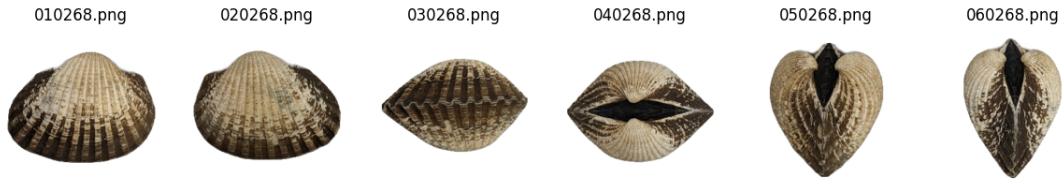


Figure 3.8: Shadows removed from female samples at different angles.

1101 3.8.1 Convolutional Neural Network

1102 Convolutional Neural Networks are the deep learning tool used in image classifica-
1103 tion, specifically binary classification. CNNs leverage their ability to share weights
1104 and use pooling techniques, reducing the number of parameters (Cui, Pan, Chen,
1105 & Zou, 2020). The proposed CNN architecture for sex identification of blood
1106 cockles employs 5 layers, designed to extract features from the input image with
1107 dimensions. The layers consist of three convolution layers, a pooling layer, a flat-
1108 ten layer, dropout, and two dense layers. The CNN framework used in this study
1109 was updated from an open source GitHub implementation by Christian Versloot,
1110 which focused on K-fold Cross Validation using TensorFlow and Keras, which was
1111 customized to align with the objectives of this study. The framework of this study
1112 is shown in Figure 3.9.

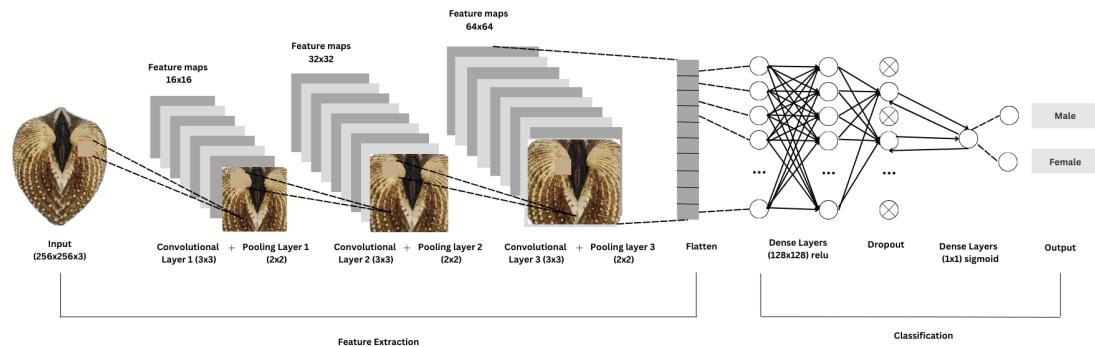


Figure 3.9: Architecture of convolutional neural network (CNN).

1113 ***Convolution Layer***

1114 The convolution layers of CNN extract the features from the input image through
1115 the convolution operation. This study uses three convolution layers with a 3x3
1116 kernel size and filter sizes of 16, 32, and 64 (*refer to Figure 3.1*). The first layer
1117 extracts the low-level features, such as edges, lines, and corners, while the deeper
1118 layers iteratively extract more complex information from these low-level features.
1119 The ReLU activation function is used as the baseline for this model, and experi-
1120 ments are conducted with different activation functions, such as ELU and PReLU,
1121 to evaluate their impact on learning complex patterns within the data.

1122 ***Pooling Layer***

1123 A pooling layer was added after the convolution layer to enhance calculation speed
1124 and prevent overfitting (Cui et al., 2020). In this study, max pooling was applied
1125 with a (2,2) kernel size.

1126 ***Fully Connected and Dropout***

1127 Fully connected layers follow after the convolution and pooling layers. Each neu-
1128 ron connects to all neurons of the previous layer. The output values from the
1129 fully connected layers are sent to an output layer. It was classified using different
1130 sigmoid functions appropriate for binary classification.

1131 A large number of parameters in the training process can lead to overfitting. It
1132 occurs when the model learns the training data too well, including its noise and
1133 irrelevant details. This results in poor performance on unseen data. To mitigate
1134 the overfitting, the dropout layer was employed. Dropout works by temporarily
1135 discarding a portion of the neurons in the network with probability p ($0 < p < 1$).

¹¹³⁶ During this process, these neurons do not participate in the forward propagation
¹¹³⁷ 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	2x2		
Convolution	32	1x1	3x3	ReLU	4,640
Max Pooling		1x1	2x2		
Convolution	64	1x1	3x3	ReLU	18,496
Max Pooling		1x1	2x2		
Flatten					
Dense	128			ReLU	7,372,928
Dropout					
Dense	1			Sigmoid	129

Table 3.1: Architecture of the convolutional neural network used.

¹¹³⁸ 3.8.2 CNN Training

¹¹³⁹ The dataset consists of 1626 images, with 127 samples from females and 144 sam-
¹¹⁴⁰ ples from males, individually for each angle. Given the minimal class imbalance,
¹¹⁴¹ random undersampling was carried out to create a balanced dataset. All images
¹¹⁴² were resized to 256x256 pixels and normalized using a Rescaling layer, ensuring
¹¹⁴³ pixel values were within the range [0, 1].

¹¹⁴⁴ *Data Splitting*

¹¹⁴⁵ Due to the limited dataset size, a traditional train-test split was not adopted.
¹¹⁴⁶ Instead, a 5-fold stratified cross-validation approach was used to maximize the
¹¹⁴⁷ use of available data while preserving the class distribution within each fold (*refer*
¹¹⁴⁸ *to Figure 3.10*). `StratifiedKFold` was applied to ensure that the distribution of
¹¹⁴⁹ male and female samples remained consistent across all folds, thereby enabling
¹¹⁵⁰ fair and robust model evaluation (GeeksforGeeks, 2020).

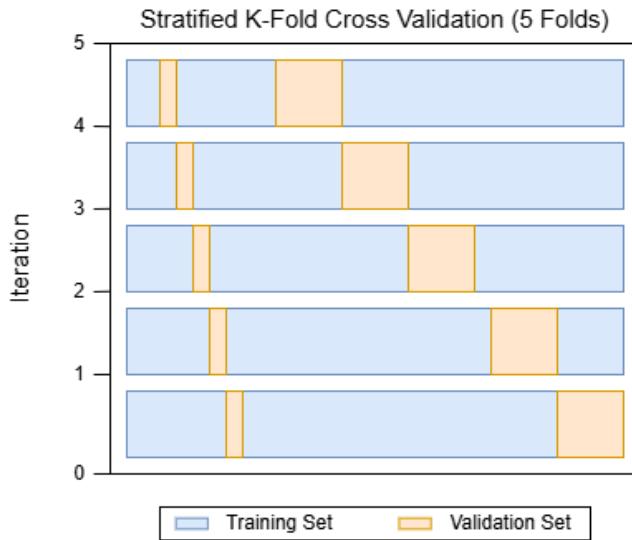


Figure 3.10: Diagram of stratified k-fold cross-validation with $k=5$.

1151 Data Augmentation

1152 Before model training, online data augmentation was applied exclusively to the
 1153 training data within each fold, creating new data variations on the fly. The aug-
 1154 mentations included random horizontal flips, slight rotations, and zoom trans-
 1155 formations to enhance data diversity and improve model generalization (Awan,
 1156 2022). All augmentation was strictly applied only to the training subset of each
 1157 fold to prevent data leakage and maintain the validity of the results (*Figure 3.11*).

1158 On-the-fly data augmentation (OnDAT) generates augmented data during each
 1159 iteration, exposing the model to constantly changing data variations. Augmenting
 1160 the original data allows better exploration of the underlying data generation pro-
 1161 cess and has the potential to prevent the model from overfitting spurious patterns,
 1162 thereby improving performance (Cerqueira, Santos, Baghoussi, & Soares, 2024).

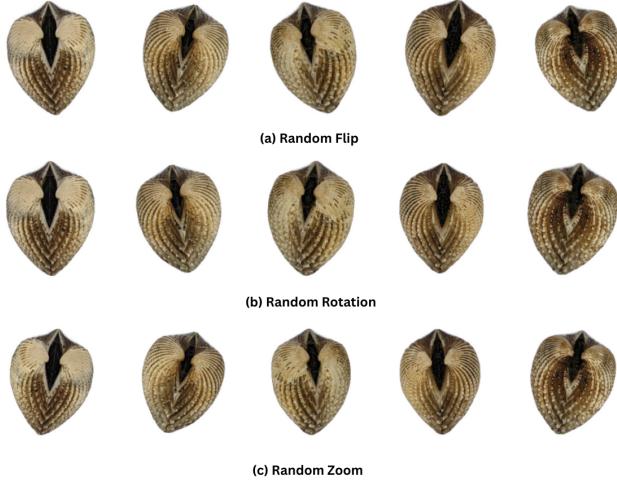


Figure 3.11: Dataset augmentation techniques.

1163 ***Training Procedure***

1164 During the training process, model performance per fold was carefully monitored.
1165 One important thing to observe is the consistency in the performance, whether
1166 the model is still learning or is at high risk of overfitting. Early stopping was ap-
1167 plied to ensure the stable performance of the model across folds. This technique
1168 allows for monitoring the training of the neural network, stopping when the per-
1169 formance metrics, in this case, validation loss, cease to improve. Furthermore, to
1170 enhance the learning process, `ReduceLROnPlateau` was applied, which decreased
1171 the learning rate if there was no improvement in the model for a specified number
1172 of epochs (Team, n.d.).

1173 The model was trained using the Adam optimization algorithm, with an initial
1174 learning rate of 0.001. Binary cross-entropy, commonly known as the log loss,
1175 was employed as the loss function due to its effectiveness in binary classification
1176 tasks. To reduce the risk of overfitting, a dropout rate of 0.5 was applied, ran-
1177 domly deactivating half of the neurons during the training process to improve

₁₁₇₈ generalization.

₁₁₇₉ 3.9 Evaluation Metrics

₁₁₈₀ Evaluating the performance of a binary classification model is essential, and se-
₁₁₈₁ lecting appropriate metrics depends on the specific requirements of the user. The
₁₁₈₂ performance of both supervised machine learning and deep learning models will
₁₁₈₃ be measured using several key metrics, including accuracy, precision, recall, F1
₁₁₈₄ Score, and the AUC-ROC Score.

₁₁₈₅ Accuracy (ACC) is the ratio of the overall correctly predicted samples to the
₁₁₈₆ total number of examples in the evaluation dataset (Cui et al., 2020). It measures
₁₁₈₇ the overall correctness of the model in predicting both male and female blood
₁₁₈₈ cockles. This metric provides insight into how well the model performs across all
₁₁₈₉ classifications. The formula for accuracy is:

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

₁₁₉₀ Precision (PREC) is the ratio of correctly predicted positive samples to all samples
₁₁₉₁ assigned to the positive class (Cui et al., 2020). This metric helps in evaluating
₁₁₉₂ the fairness of the model and prevents the misclassification of blood cockles as it
₁₁₉₃ identifies potential inaccuracies or biases. The formula for precision is:

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

₁₁₉₄ Recall (REC), also known as sensitivity or the true positive rate (TPR), is the
₁₁₉₅ ratio of correctly predicted positive cases to all the actual positive samples (Cui
₁₁₉₆ et al., 2020). It represents the ability of the model to correctly identify positive
₁₁₉₇ male and female samples. The formula for recall is:

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

₁₁₉₈ The F1 Score is the harmonic mean of precision and recall, which penalizes extreme
₁₁₉₉ values of either of the two metrics (Cui et al., 2020). It is particularly useful when
₁₂₀₀ the class distribution is imbalanced. The formula for the F1 Score is:

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

₁₂₀₁ The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a
₁₂₀₂ performance measurement for classification problems, particularly used in deep
₁₂₀₃ learning in this study. The ROC curve is a plot of the true positive rate (recall)
₁₂₀₄ against the false positive rate (1 - specificity), and the AUC score quantifies the
₁₂₀₅ overall ability of the model to discriminate between positive and negative classes.
₁₂₀₆ A higher AUC indicates better model performance. (Nahm, 2022)

₁₂₀₇ **Chapter 4**

₁₂₀₈ **Results and Discussions**

₁₂₀₉ This chapter presents the results from the machine learning and deep learning
₁₂₁₀ analyses conducted on the preprocessed dataset. It includes an evaluation of
₁₂₁₁ various machine learning classifiers and the application of deep learning models
₁₂₁₂ for image-based classification. The primary focus is on identifying key predictors
₁₂₁₃ and assessing classification performance for sex identification in *T. granosa*.

₁₂₁₄ **4.1 Machine Learning Analysis**

₁₂₁₅ This chapter outlines the results of preprocessing, training of machine learning
₁₂₁₆ models, and feature importance analysis, all conducted in Google Colab using
₁₂₁₇ Python. The dataset was preprocessed in Colab, and the training and evaluation
₁₂₁₈ of various classifiers were performed entirely within this environment. This part of
₁₂₁₉ the paper includes five subsections: data exploration, statistical analysis, feature
₁₂₂₀ importance analysis, performance evaluation, and confusion matrix analysis.

1221 4.1.1 Data Exploration

1222 Exploratory data analysis was performed to characterize the dataset using visu-
1223 alizations to understand the patterns and correlations within the data. A corre-
1224 lation heatmap was created to assess the relationship between the predictors and
1225 the target variable.

1226 The heatmap (*see Figure 4.1*) revealed three features most correlated with the
1227 sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos distance-length
1228 ratio ($r = 0.12$), and the distance between the umbos ($r = 0.12$). Each of these
1229 features demonstrated a weak positive relationship with the target variable.

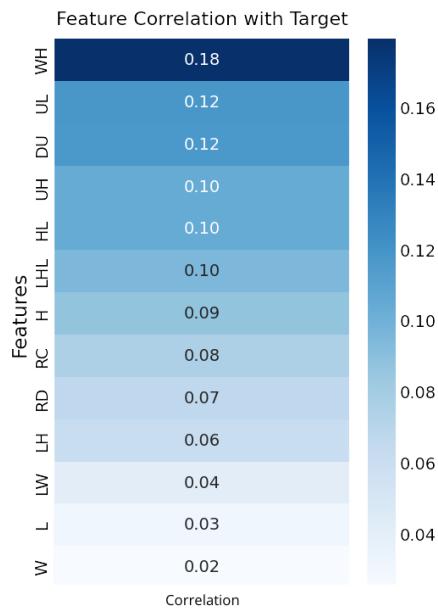


Figure 4.1: Heatmap of morphometric correlations with *T. granosa* sex.

₁₂₃₀ **4.1.2 Statistical Analysis**

₁₂₃₁ As part of the exploratory data analysis, statistical testing confirmed that the
₁₂₃₂ dataset did not follow a normal distribution (*see Table 4.1*). Consequently, the
₁₂₃₃ Mann-Whitney U test was applied with a significance level of $\alpha = 0.05$ to com-
₁₂₃₄ pare male and female samples. Out of thirteen features, five showed statistically
₁₂₃₅ significant differences. These included: distance between umbos ($p = 0.025$),
₁₂₃₆ length-width ratio ($p = 0.011$), umbos distance-length ratio ($p = 0.019$), width-
₁₂₃₇ height ratio ($p = 0.003$), and umbos distance-height ratio ($p = 0.036$).

₁₂₃₈ It is important to note that statistical significance does not imply predictive im-
₁₂₃₉ portance. Therefore, further analysis, such as feature importance evaluation, was
₁₂₄₀ 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.

1241 4.1.3 Feature Importance Analysis

1242 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric
1243 method that is suitable for evaluating differences in distributions across groups
1244 when the data does not follow a normal distribution. This approach was chosen
1245 because of the non-normality of the dataset and its robustness in handling con-
1246 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,
1247 2024)

1248 Kruskal-Wallis test analysis showed that the width-to-height ratio (WH ratio)
1249 had the highest importance score, indicating it is the most statistically significant
1250 feature for distinguishing the sex of *T. granosa*. Other notable features included
1251 the length-to-width ratio (LW ratio), umbos distance-to-length ratio (UL ratio),
1252 distance between the umbos, and umbos distance-to-height ratio (UH ratio), all
1253 of which contributed significantly to the classification task (*refer to Figure 4.2*).

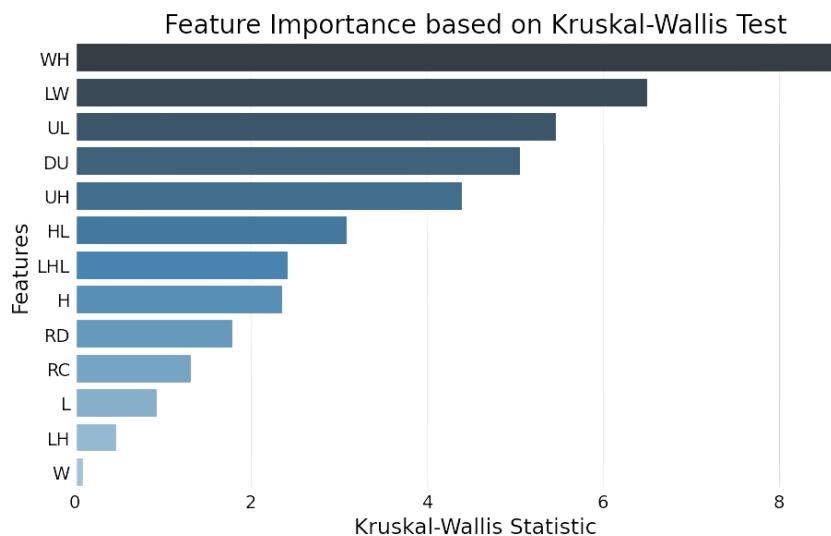


Figure 4.2: Feature importance scores using the Kruskal-Wallis test.

¹²⁵⁴ **4.1.4 Performance Evaluation**

¹²⁵⁵ Table 4.2 shows the performance metrics of different machine learning models
¹²⁵⁶ trained using all 13 features from the dataset. Among the models, Gradient
¹²⁵⁷ Boosting achieved the highest accuracy of 61.03%, along with strong precision,
¹²⁵⁸ recall, and F1 Score values. AdaBoost also performed competitively, with an
¹²⁵⁹ accuracy of 60.63%. These results highlight the effectiveness of ensemble methods
¹²⁶⁰ such as Gradient Boosting and AdaBoost when utilizing the full feature set, likely
¹²⁶¹ because of their capability to combine multiple weak learners into a more robust
¹²⁶² predictive model (Hussain & Zaidi, 2024).

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.

¹²⁶³ Table 4.3 presents the performance of the same models using only the top five fea-
¹²⁶⁴ tures identified through Kruskal-Wallis feature importance analysis. The selected
¹²⁶⁵ features are the distance between the umbos, length-to-width ratio, width-to-
¹²⁶⁶ height ratio, umbos distance-to-height ratio, and umbos distance-to-length ratio.

¹²⁶⁷ Interestingly, the overall performance of the models improved when using only the
¹²⁶⁸ top 5 features compared to using all 13. K-Nearest Neighbors (KNN) achieved the
¹²⁶⁹ best results with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and
¹²⁷⁰ an F1 Score of 63.75%. Gradient Boosting followed closely behind. These find-
¹²⁷¹ ings suggest that reducing the feature set to the most relevant variables helped

1272 simplify the models, improved generalization, and enhanced predictive performance—particularly for KNN, which showed a notable improvement over its earlier results with the full feature set.

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.

1275 4.1.5 Confusion Matrix Analysis

1276 Figure 4.3 summarizes the performance of the K-Nearest Neighbors model in
1277 classifying *T. granosa* based on their sex, where 0 represents female samples and
1278 1 represents male samples. From the matrix, it can be observed that out of all the
1279 actual female samples (true label 0), 91 were correctly predicted as female (true
1280 positive for class 0), while 36 were incorrectly classified as male (false negative
1281 for class 0). On the other hand, out of all the actual male samples (true label
1282 1), 72 were correctly predicted as male (true positive for class 1), while 55 were
1283 incorrectly classified as female (false negative for class 1).

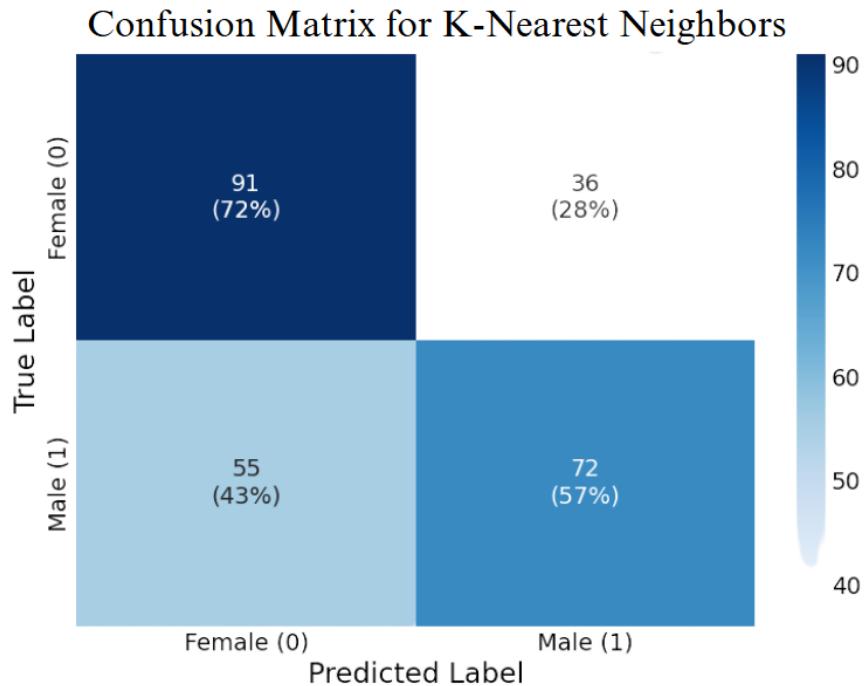


Figure 4.3: KNN confusion matrix for *T. granosa* sex classification.

1284 4.2 Deep Learning Analysis

1285 This section presents the performance of the Convolutional Neural Network (CNN)
1286 model in classifying the sex of *T. granosa* based on shell morphology. The analysis
1287 evaluates the model's ability to distinguish between male and female shell images
1288 using various evaluation metrics. This part of the paper includes six subsections:
1289 baseline model, comparison of individual and combined angles, training result and
1290 hyperparameter tuning, proposed model, learning rates and training behavior per
1291 fold, and visualizations.

1292 The machine learning analysis (see Figure 4.3) revealed that five of the origi-
1293 nal features produced significant results. The K-Nearest Neighbor (KNN) model
1294 achieved an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1

1295 Score of 63.75%. This section compares the model’s performance across differ-
1296 ent angles based on the results of the machine learning and feature importance
1297 analysis.

1298 4.2.1 Baseline Model

1299 This section presents the baseline model with a batch size of 16 and 20 epochs,
1300 which will serve as the starting point for comparison and provide a guideline for
1301 hyperparameter tuning. The focus will be on one of the angles, specifically the
1302 Left Lateral view, since the feature importance analysis using the Kruskal-Wallis
1303 Test indicated that the width-to-height ratio had the highest importance score,
1304 which is most visible from the Left Lateral view.

1305 The unbalanced dataset, which consisted of 144 male samples and 127 female
1306 samples, achieved an accuracy of 65.27%, precision of 71.82%, recall of 58.99%,
1307 an F1 Score of 63.99%, an AUC score of 73.08%, and a loss of 0.6122. However, to
1308 address the class imbalance and enhance model performance, random undersam-
1309 pling was performed. This approach resulted in improved performance metrics for
1310 the balanced dataset, with an accuracy of 67.34%, precision of 69.43%, a recall
1311 of 64.06%, an F1 Score of 65.60%, an AUC score of 74.31%, and a lower loss of
1312 0.5981.

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).

4.2.2 Comparison of Individual and Combined Angles

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

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

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).

1328 4.2.3 Training Result and Hyperparameter Tuning

1329 The Left Lateral angle view was selected for further optimization. Several ex-
1330 periments were conducted by tuning hyperparameters such as batch size, number
1331 of epochs, and activation functions. Each adjustment was compared against the
1332 baseline model to enhance performance and develop a robust CNN for sex classi-
1333 fication of *T. granosa*.

1334 The Left Lateral angle was chosen because it achieved the highest accuracy and
1335 precision among all individual views, and because the Kruskal-Wallis feature im-
1336 portance analysis indicated that the width-to-height ratio, a feature most visible
1337 from the lateral perspective, was the most significant morphological trait for clas-
1338 sification. Therefore, focusing on this view was expected to maximize the model's
1339 learning capacity and improve classification performance.

1340 A. Batch Size and Number of Epochs

1341 Table 4.6 shows the results indicating that a batch size of 32 with 50 epochs
1342 achieved the best overall performance, with an accuracy of 71.68%, a precision of
1343 72.52%, a recall of 69.29%, an F1 Score of 69.12%, and AUC score of 77.34%.

1344 In contrast, increasing the batch size to 64 resulted in lower recall and F1 Scores,
1345 suggesting that smaller batch Sizes (16 or 32) are more effective for this dataset.
1346 A moderate batch size of 32 allowed the model to generalize better and maintain
1347 stable learning, while too large batch sizes may have led to underfitting.

Epoch	Batch Size	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC Score (%)	Loss
20	16	67.34	69.43	64.06	65.60	74.31	0.5981
	32	68.13	72.25	58.95	62.34	74.76	0.6041
	64	56.71	65.96	36.83	41.46	71.28	0.6692
30	16	67.73	70.17	64.06	65.72	75.76	0.5900
	32	71.28	73.17	66.89	68.27	76.76	0.5832
	64	57.95	61.94	48.12	52.66	71.22	0.6241
	16	67.73	70.17	64.06	65.72	75.76	0.5900
	32	71.68	72.52	69.29	69.12	77.34	0.5824
	64	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.

1348 B. Activation Functions

1349 Table 4.7 shows the performance of different activation functions applied to the
 1350 CNN model trained with a batch size of 32 and 50 epochs. Based on the results,
 1351 the ReLU activation function achieved the best overall performance, with an ac-
 1352 curacy of 71.68%, precision of 72.52%, recall of 69.29%, F1 Score of 69.12%, and
 1353 AUC score of 77.34%, along with the lowest loss at 0.5824. This suggests that
 1354 ReLU remains an effective activation function for the classification of *T. granosa*,
 1355 outperforming both ELU and PReLU in this setup.

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).

1356 4.2.4 Proposed Model

1357 This section presents the performance evaluation of the proposed Convolutional
 1358 Neural Network (CNN) model, trained with a batch size of 32, 50 epochs, and us-
 1359 ing the ReLU activation function. The model's effectiveness was assessed through
 1360 5-fold cross-validation to ensure robustness and generalizability across different

1361 data partitions.

1362 The proposed model consistently achieved high performance in Folds 1, 3, and
1363 5, with accuracies above 76% and strong recall and AUC scores, demonstrating
1364 its potential for reliable sex identification of *T. granosa*. The slight variation
1365 in performance across folds may be attributed to differences in data distribution,
1366 emphasizing the importance of further data augmentation and balancing for future
1367 work.

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).

1368 4.2.5 Learning Rates and Training Behavior per Fold

1369 This section presents the learning rate adjustments, early stopping events, and
1370 best epoch selections for each fold during the 5-fold cross-validation of the pro-
1371 posed model. During training, the ReduceLROnPlateau callback was employed
1372 to monitor the validation loss and automatically reduce the learning rate when
1373 performance plateaued. Additionally, EarlyStopping was utilized to halt training
1374 once no further improvement was observed after a set patience, and the model
1375 weights were restored from the end of the best-performing epoch to ensure optimal
1376 performance.

1377 The following table summarizes the epochs where learning rate reductions oc-
1378 curred, the adjusted learning rates, the epochs at which early stopping took place,

¹³⁷⁹ 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.

¹³⁸⁰ 4.2.6 Visualizations

¹³⁸¹ Figure 4.4 shows the performance of the model in the training and validation in
¹³⁸² terms of accuracy across five folds. The graph across folds displays a consistent
¹³⁸³ upward trend for the training accuracy. However, there is an observable change in
¹³⁸⁴ the performance, particularly in Folds 1 and 2, where it shows a slight downward
¹³⁸⁵ trend in the validation accuracy.

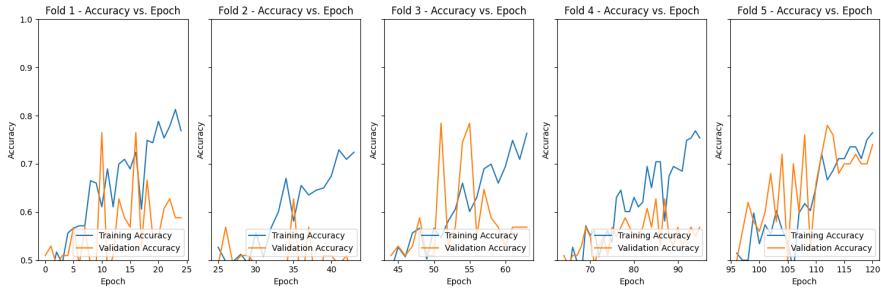


Figure 4.4: Training and validation accuracy per fold.

1386 Figure 4.5 shows the average performance of the model in terms of training and
 1387 validation accuracy across five folds. An upward trend is observable in the training
 1388 and validation accuracy, indicating that the model gradually improves over
 1389 the epochs. While fluctuations or dips can be seen in the validation accuracy,
 1390 the model recovers in later epochs. The training accuracy remains consistently
 1391 higher than the validation accuracy, which is expected behavior, as it learns from
 1392 the training data. Generally, the model demonstrates a gradual improvement in
 1393 learning, as reflected in the average upward trend aggregated across five folds.

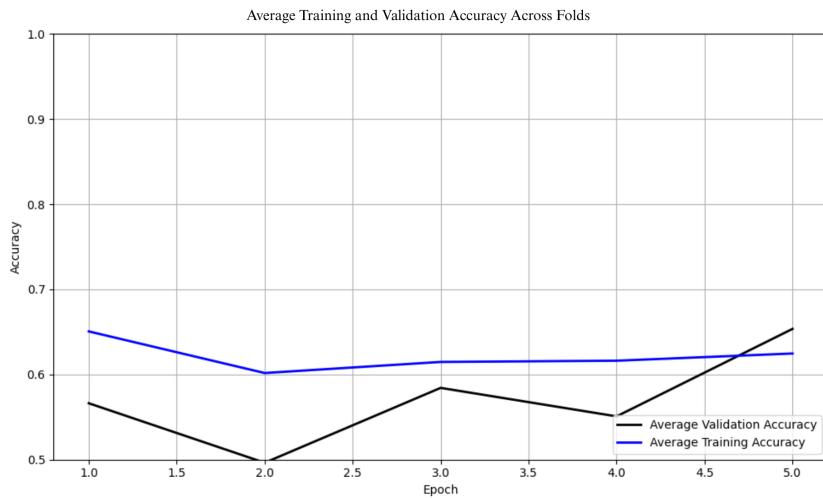


Figure 4.5: Average training and validation accuracy across folds.

1394 Figure 4.6 shows the performance of the model in the training and validation in

1395 terms of the training and validation loss across five folds. The graph across folds
 1396 displays a consistent downward trend for the training loss. On the other hand,
 1397 there is an observable change in the performance, especially in Folds 1,2,3, and 4,
 1398 where it shows an upward trend in the validation loss. This is an implication for
 1399 the learning performance of the model, as it may not be learning effectively.

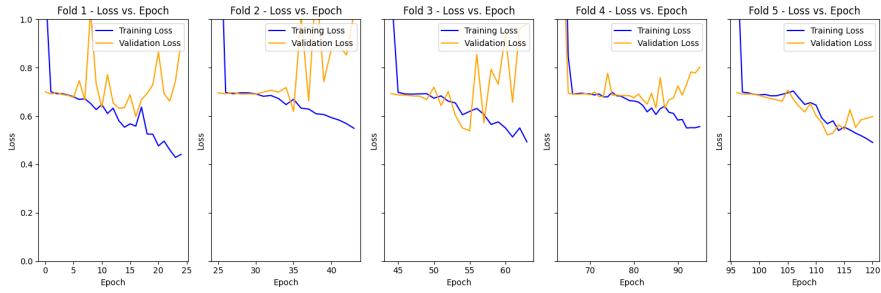


Figure 4.6: Training and validation loss per fold.

1400 Figure 4.7 shows the average performance of the model regarding training and
 1401 validation loss across five folds. A continuous downward trend is observed in
 1402 training and validation accuracy, indicating that the model's loss gradually de-
 1403 creases across epochs. This suggests that the model generalizes better following
 1404 the initial instability in the earlier epoch in the folds. Additionally, the training
 1405 loss consistently remains lower than the validation loss, since the model was di-
 1406 rectly optimized on the training set. Overall, the downward trend in training and
 1407 validation loss signifies that the model is learning and improving across the five
 1408 folds.

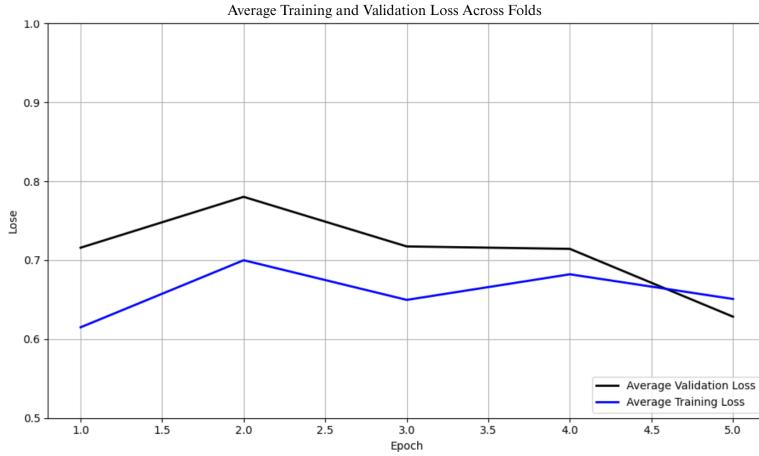


Figure 4.7: Average training and validation loss across folds.

1409 Figure 4.8 shows the confusion matrix for the true class label and predicted class
 1410 label after the training and validation. The matrix shows the correctly predicted
 1411 male and female samples and their corresponding percentages. Females have
 1412 slightly higher true positives compared to males in the number and percentages,
 1413 which are 94 and 88, corresponding to 74% and 69%, respectively. Additionally,
 1414 the falsely classified samples were 33 for females and 39 for males, respectively,
 1415 accounting for 26% and 31%.

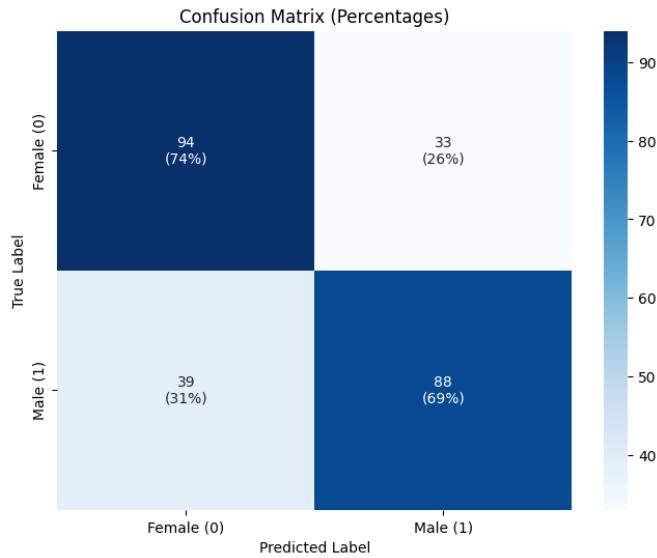


Figure 4.8: Confusion matrix for final model predictions.

1416 Figure 4.9 shows the average ROC curve, showing the proposed model's ability
 1417 to correctly identify the true positives, which can help determine the trade-off
 1418 between specificity and sensitivity. It will also determine the model's validity,
 1419 supporting that it is not being predicted based only on random chance. The
 1420 range of AUC ROC is between 0.5 and 1. The model achieved an average score
 1421 of 0.7734, which is better than random chance and a positive indication that the
 1422 model is performing reasonably.

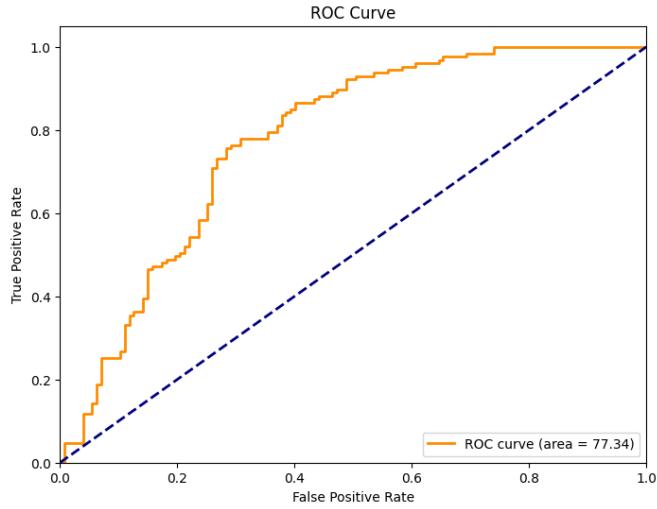


Figure 4.9: ROC curve with AUC score for the proposed model.

1423 4.3 Discussions

1424 This study aimed to develop a non-invasive method for identifying the sex of *T.*
 1425 *granosa* using machine learning and deep learning techniques. The goal of the
 1426 study is to determine the relevance of linear measurements and image data in
 1427 building a classification model that supports sustainable aquaculture practices.

1428 In the initial machine learning experiments, the focus was on classification ac-
 1429 curacy, feature selection, and feature importance analysis. Results revealed that
 1430 models trained on a reduced feature set performed better than those using all
 1431 available data. Specifically, the five key features, selected through statistical tests
 1432 (Mann-Whitney U-test and Kruskal-Wallis test), led to improved classification
 1433 performance. The K-nearest neighbors (KNN) classifier, utilizing five significant
 1434 features, achieved an accuracy of 64.16%, a precision of 64.97%, a recall of 64.16%,
 1435 and an F1 score of 63.57%. Feature importance analysis identified the width-
 1436 height ratio as the most discriminative variable, observed in the left lateral view,

₁₄₃₇ as supported by the feature correlation score of $p = 0.18$. These results indicated
₁₄₃₈ that a more focused set of features can enhance model performance, confirming
₁₄₃₉ the potential of non-invasive sex identification using linear measurements.

₁₄₄₀ Furthermore, deep learning experiments explored the impact of different image
₁₄₄₁ angles and model hyperparameters on classification accuracy. The baseline model
₁₄₄₂ with balanced data outperformed those trained on unbalanced data. Hyperpa-
₁₄₄₃ rameter optimization, involving sets of epochs and batch sizes, further improved
₁₄₄₄ model performance. Additionally, the influence of different activation functions,
₁₄₄₅ such as ReLU, ELU, and PReLU, was evaluated. The study found that the cam-
₁₄₄₆ era angle from the left lateral view consistently produced the best results, with an
₁₄₄₇ accuracy of 71.68%, precision of 72.52%, recall of 69.29%, F1 score of 69.12%, and
₁₄₄₈ an AUC score of 77.34% with a batch size of 32 and epoch of 50. The result sug-
₁₄₄₉ gests that certain morphological features are most visible from the camera angle
₁₄₅₀ taken from the left lateral.

₁₄₅₁ Overall, the key insights from machine learning and deep learning are the effects
₁₄₅₂ of reducing dimensionality through feature selection, supported by statistical ana-
₁₄₅₃ lysis. Additionally, the left lateral angle showed higher performance compared to
₁₄₅₄ other camera angles, which suggests a sex-determining factor that shows potential
₁₄₅₅ for further exploration and analysis.

₁₄₅₆ When compared to similar existing studies—such as the gender classification
₁₄₅₇ method for Chinese mitten crab using deep learning CNNs (Cui et al., 2020)—there
₁₄₅₈ are both methodological similarities and differences. Both studies employed Con-
₁₄₅₉ volutional Neural Networks (CNNs) with three convolutional layers, pooling lay-
₁₄₆₀ ers, fully connected layers, and dropout. The crab study used grayscale images

¹⁴⁶¹ resized to 64×64 pixels, while this study utilized higher-resolution RGB images
¹⁴⁶² (256×256). In terms of architecture, the crab study applied 4, 8, and 16 filters
¹⁴⁶³ in its convolutional layers and 256 neurons in the fully connected layer, achieving
¹⁴⁶⁴ a high accuracy of 98.90%. In contrast, this study used 16, 32, and 64 filters in
¹⁴⁶⁵ the convolutional layers and 128 neurons in the fully connected layer, reaching an
¹⁴⁶⁶ accuracy of 71.68%. This lower performance may be attributed to the subtler mor-
¹⁴⁶⁷ phological differences between male and female *T. granosa*, as well as limitations
¹⁴⁶⁸ in image quality and sample size.

¹⁴⁶⁹ The findings of this study are significant in demonstrating the feasibility of a non-
¹⁴⁷⁰ invasive, accurate, and efficient sex identification method for *T. granosa*. This
¹⁴⁷¹ approach aligns with sustainable aquaculture practices by reducing the need for
¹⁴⁷² invasive sex-identifying methods and offering potential in real-time settings. By
¹⁴⁷³ integrating machine learning with deep learning image analysis, this study pro-
¹⁴⁷⁴ vides a valuable model for non-invasive sex identification for *T. granosa*.

¹⁴⁷⁵ This study acknowledges several limitations, particularly the size of the dataset
¹⁴⁷⁶ (271 samples) and the reliance on six fixed image angles. These constraints may
¹⁴⁷⁷ not fully represent the morphological variability across different populations or
¹⁴⁷⁸ environments. Nevertheless, the results indicate that combining machine learning
¹⁴⁷⁹ and deep learning offers a promising and non-invasive approach for sex identifica-
¹⁴⁸⁰ tion in *T. granosa*.

¹⁴⁸¹ Chapter 5

¹⁴⁸² Conclusion and ¹⁴⁸³ Recommendations

¹⁴⁸⁴ 5.1 Conclusion

¹⁴⁸⁵ This study utilized the application of machine learning and deep learning tech-
¹⁴⁸⁶ niques for non-invasive sex identification of *T. granosa* based on the morphomet-
¹⁴⁸⁷ ric and morphological characteristics. A manually curated dataset was developed,
¹⁴⁸⁸ consisting of linear measurements and images captured from six different angles.
¹⁴⁸⁹ Machine learning methods were first employed to identify statistically significant
¹⁴⁹⁰ features, which served as the basis for deep learning analysis using a five-layered
¹⁴⁹¹ Convolutional Neural Network (CNN). The proposed CNN model yielded an av-
¹⁴⁹² erage accuracy of 71.68%, demonstrating a viable solution for non-invasive sex
¹⁴⁹³ identification of *T. granosa* based on linear measurements and morphological char-
¹⁴⁹⁴ acteristics.

1495 Through the availability of the gathered data, trial-and-error experimentation
1496 was conducted by adjusting the number of layers, batch size, epoch, and activa-
1497 tion functions. While the study has made significant progress, challenges were
1498 encountered during CNN training, particularly due to hardware memory limita-
1499 tions. To overcome these, the researchers utilized synchronous Google Colab with
1500 100 computing units, requiring subscriptions, repeated retraining, and reconfig-
1501 urations, which demanded considerable financial resources and time to optimize
1502 the parameters.

1503 Upon comparing the experimental results of model parameters, it was demon-
1504 strated that non-invasive sex identification on *T. granosa* is achievable through
1505 the integration of machine learning and deep learning methods. Machine learning
1506 models based on five statistically selected features outperformed those using all
1507 features, with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and
1508 an F1 Score of 63.57% using K-nearest neighbors (KNN) classifier. Further im-
1509 provements were attained through deep learning models, using Left Lateral image
1510 angle, achieving an accuracy of 71.68%, precision of 72.52%, recall of 69.29%, F1
1511 Score of 69.12%, and an AUC score of 77.34%.

1512 These results establish the CNN model as a viable baseline for future studies in
1513 non-invasive sex identification of *T. granosa*. By providing a practical and less
1514 invasive alternative to traditional methods, this study contributes a significant
1515 advancement in the field of aquaculture and marine biology.

1516 5.2 Recommendations

1517 This special problem entitled Morphometric and Morphological-Based Non-invasive
1518 Sex Identification of *T. granosa* focuses on creating a baseline study that will serve
1519 as a foundation for further studies involving *T. granosa*, blood cockles, using ma-
1520 chine learning and deep technologies in determining the sex of the samples is a
1521 salient need in aquaculture practices. Thus, the proposed recommendations are
1522 the future applications to improve and have detailed analysis, such as focusing
1523 on shape analysis, exploring other state-of-the-art deep learning techniques, or
1524 transfer learning, such as ResNet, SqueezeNet, and InceptionNet, and comparing
1525 the analysis results. Furthermore, the main goal of conducting this is to have the
1526 ability to identify the sex of the samples by taking real-time angles by rotating
1527 from the dorsal, lateral, and anterior.

1528 Due to the time constraints, the researchers were only able to gather a total of
1529 1,626 images with 271 images per angle, and utilized these for model training and
1530 validation. A larger and more diverse collection of images could further improve
1531 the model's generalization. In order to capture more variability, future study
1532 might include expanding the dataset to improve classification performance.

1533 Future studies could also invest in a sturdier and more controlled environment
1534 by using a green background and positioning a fixed camera angle during image
1535 acquisition. In addition, researchers may experiment with other image processing
1536 techniques such as morphological transformations to emphasize features. The
1537 dataset can be utilized for further analysis through advanced deep learning and
1538 computer vision methods to make sense of the images gathered and discern sexual
1539 dimorphism for *T. granosa*.

₁₅₄₀ **Chapter 6**

₁₅₄₁ **References**

- ₁₅₄₂ 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
- ₁₅₄₃
- ₁₅₄₄
- ₁₅₄₅ Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams *Anadara granosa* (l.) and *Anadara antiquata* (l.) (bivalvia: Arcidae) in central java. *10*.
- ₁₅₄₆
- ₁₅₄₇
- ₁₅₄₈ Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian Sains*, *14*.
- ₁₅₄₉
- ₁₅₅₀ Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).
- ₁₅₅₁ Data scaling performance on various machine learning algorithms to identify
- ₁₅₅₂ abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:
- ₁₅₅₃ 10.14710/jtsiskom.2021.14105
- ₁₅₅₄ Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.
- ₁₅₅₅ *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- ₁₅₅₆ Awan, A. A. (2022, November). *A complete guide to data augmentation*.

- 1557 *tion.* Retrieved from <https://www.datacamp.com/tutorial/complete>
- 1558 **-guide-data-augmentation**
- 1559 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: Mangrove-friendly aquacul-
1560 ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-
1561 tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on*
1562 *mangrove-friendly aquaculture organized by the SEAFDEC aquaculture de-*
1563 *partment, january 11-15, 1999, Iloilo City, Philippines* (pp. 41–56).
- 1564 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
1565 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
1566 Resources.
- 1567 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization
1568 of waste cockle shell (*Anadara granosa*) in biodiesel production from palm
1569 olein: Optimization using response surface methodology. *Fuel*, 90(7), 2353–
1570 2358. doi: 10.1016/j.fuel.2011.03.002
- 1571 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining*
1572 *mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127
1573 .v1
- 1574 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
1575 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A
1576 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1577 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/
1578 msq345
- 1579 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1580 proaches, challenges and opportunities for aquaculture. *Journal of Marine*
1581 *Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1582 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).

- 1583 Mass mortalities in bivalve populations: A review of the edible cockle
1584 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1585 10.1016/j.ecss.2014.04.011
- 1586 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1587 research in bivalves: Towards the identification of molecular markers of
1588 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1589 10.1016/j.jprot.2012.04.027
- 1590 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data
1591 augmentation for forecasting with deep learning. *ArXiv (Cornell University)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1592 1593 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
- 1594 1595 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1596 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1597 10.1093/icb/icb076
- 1598 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1599 Bivalvenet: A hybrid deep neural network for common cockle (*Cerastoderma
1600 edule*) geographical traceability based on shell image analysis. *Ecological
1601 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1602 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1603 for chinese mitten crab using deep convolutional neural network. *Multi-
1604 media Tools and Applications*, 79(11-12), 7669–7684. doi: [https://doi.org/
1605 10.1007/s11042-019-08355-w](https://doi.org/10.1007/s11042-019-08355-w)
- 1606 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle
1607 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-
1608 posure. *Aquaculture*, 56(2), 151–162. Retrieved from <https://doi.org/>

- 1609 10.1016/0044-8486(86)90024-4 doi: 10.1016/0044-8486(86)90024-4
- 1610 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1611 ageing of eel, *Anguilla anguilla* (L.)? *International Review of Hydrobiology*,
1612 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1613 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and*
1614 *anatomy for young learners*. Rosie Research. Retrieved from <https://rosieresearch.com/clam-dissection-anatomy/>
- 1616 Fao 2024 report: Sustainable aquatic food systems important for global food
1617 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1620 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1621 elemental composition: Complementary tools for stock discrimination of
1622 mulloway (*Argyrosomus japonicus*) in southern australia. *Fish Research*,
1623 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1624 GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*. Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1627 GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn*. Retrieved from <https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/> (Accessed: 2025-04-23)
- 1630 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Blackwell Science.
- 1632 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1634 Hussain, S. S., & Zaidi, S. S. H. (2024). Adaboost ensemble approach with

- 1635 weak classifiers for gear fault diagnosis and prognosis in dc motors. *Applied
1636 Sciences*, 14(7), 3105. doi: 10.3390/app14073105
- 1637 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1638 surface water characterization and heavy metal health risk on cockles in
1639 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1640 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1641 Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a com-
1642 prehensive guide to data rescaling*. DataCamp. Retrieved from [https://www
1644 .datacamp.com/tutorial/normalization-in-machine-learning](https://www
1643 .datacamp.com/tutorial/normalization-in-machine-learning) (Accessed 2025-04-23)
- 1645 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The ear-
1646 liest example of sexual dimorphism in bivalves—evidence from the astar-
1647 tid *Nicaniella* (lower jurassic, southern germany). *Journal of Paleontology*,
1648 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1649 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1650 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1651 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1652 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1653 learning-based phenotype classification of three ark shells: *Anadara kagoshi-*
1654 *mensis*, *Tegillarca granosa*, and *Anadara Broughtonii*. *Frontiers in Marine
1655 Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1656 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the
1657 granulated ark, *Tegillarca granosa* (linne). *The Korean Journal of Malacol-
1658 ogy*, 13, 55–64.
- 1659 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as
1660 a tool for stock identification of the southern blue whiting, *Micromesistius*

- 1661 *australis*. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1662 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1663 et al. (2016). Identifying blue whiting (*Micromesistius poutassou*) stock
1664 structure in the northeast atlantic by otolith shape analysis. *Canadian*
1665 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1666 cjfas-2015-0332
- 1667 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1668 *Tegillarca granosa* (linnaeus, 1758) in myeik coastal areas. *J. Myanmar*
1669 *Acad. Arts Sci*, 4.
- 1670 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning
1671 and larval rearing of the blood cockle, *Tegillarca granosa* (linnaeus, 1758),
1672 in the philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023
1673 .36.2.001
- 1674 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of
1675 local populations of the common sole *solea solea* (pisces, soleidae) in the nw
1676 mediterranean through otolith morphometrics and shape analysis. *Marine*
1677 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1678 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and
1679 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–
1680 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/
1681 kja.21209
- 1682 Narasimham, K. A. (1988). Taxonomy of the blood clams anadara (*Tegillarca*
1683 *granosa* (linnaeus, 1758) and a. (t.) *rhombea* (born, 1780).
- 1684 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1685 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1686 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500

- 1687 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1688 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1689 doi: 10.1002/jmor.10439
- 1690 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1691 morphometrics and machine learning challenge currently accepted species
1692 limits of the land snail *placostylus* (pulmonata: Bothriembryontidae) on the
1693 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1694 doi: 10.1093/mollus/eyz031
- 1695 Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test.*
1696 Retrieved from https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/ (Accessed: 2025-04-23)
- 1698 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,
1699 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1700 carbons from consumption of blood cockle and exposure to contaminated
1701 sediments and water along the klang strait, malaysia. *Marine Pollution
1702 Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1703 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1704 of seasonal variation and anthropogenic stress on blood cockle (*Tegillarca
1705 Granosa*) production potential. *Influence of Seasonal Variation and Anthro-
1706 pogenic Stress on Blood Cockle (*Tegillarca Granosa*) Production Potential,*
1707 44(2), 62–82.
- 1708 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1709 chine learning and its applications to biology. *PLoS Computational Biology*,
1710 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1711 Team, K. (n.d.). *Keras documentation: Reducelronplateau.* Retrieved from
1712 https://keras.io/api/callbacks/reduce_lr_on_plateau/

- 1713 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1714 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1715 & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).
1716 College Park, MD: Maryland Sea Grant.
- 1717 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1718 approach for morphological feature extraction based on variational auto-
1719 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*
1720 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1721 Wong, T. M., & Lim, T. G. (2018). *Cockle (Anadara granosa) seed produced in*
1722 *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in_3366.pdf
- 1723 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1724 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972
1725 .5008949
- 1726 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1727 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1728 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-
1729 pacts of four commonly used nanoparticles on the metabolism of a ma-
1730 rine bivalve species, *Tegillarca granosa*. *Chemosphere*, 296, 134079. doi:
1731 10.1016/j.chemosphere.2022.134079
- 1732 Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1733 blood clam *Tegillarca granosa*. *Marine Environmental Research*, 177. Re-
1734 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

¹⁷³⁵ **Appendix A**

¹⁷³⁶ **Code Snippets**

¹⁷³⁷ **i. Machine Learning**

¹⁷³⁸ This section of the paper displays the key steps in the machine learning analysis by
¹⁷³⁹ performing feature engineering to create and transform a new dataset, identifying
¹⁷⁴⁰ the most significant features through the Kruskal-Wallis Test, applying random
¹⁷⁴¹ undersampling to address the minimal imbalance in the dataset, and conducting
¹⁷⁴² five-fold cross-validation to evaluate the model's performance.

```
female_litob['LW_ratio']= female_litob['Length'] / female_litob['Width']
female_litob['LH_ratio'] = female_litob['Length'] / female_litob['Height']
female_litob['WH_ratio'] = female_litob['Width'] / female_litob['Height']
# female_litob['DU_ratio'] = female_litob['Distance Umbos'] / female_litob['Height']
female_litob['UL_ratio'] = female_litob['Distance Umbos'] / female_litob['Length']
female_litob['HL_ratio'] = female_litob['Length (Hinge Line)'] / female_litob['Length']
female_litob['UH_ratio'] = female_litob['Distance Umbos'] / female_litob['Height']
female_litob['Rib Density'] = female_litob['Rib count'] / female_litob['Length']
```

Figure A.1: Feature engineering used to create and transform the dataset for machine learning analysis.

```

sorted_features = feature_importance_scores.sort_values(ascending=False)
colors = sns.color_palette("Blues_d", len(sorted_features))
colors = colors[::-1]
plt.figure(figsize=(10, 6))

# Map codenames to sorted_features.index
sorted_features.index = sorted_features.index.map(codenames)

sns.barplot(x=sorted_features.values, y=sorted_features.index, hue = sorted_features.index, palette=colors,dodge=False,leg
plt.xlabel("Kruskal-Wallis Statistic")
plt.ylabel("Features")
plt.title("Feature Importance based on Kruskal-Wallis Test")
plt.show()

```

Figure A.2: Feature importance scores derived from the Kruskal-Wallis test to identify the most significant variables.

```

from imblearn.under_sampling import RandomUnderSampler

rus = RandomUnderSampler(sampling_strategy=1) # Numerical value
# rus = RandomUnderSampler(sampling_strategy="not minority") # String
X_res, y_res = rus.fit_resample(X, y)

ax = y_res.value_counts().plot.pie(autopct='%.2f')
_ = ax.set_title("Under-sampling")

```

Figure A.3: Random undersampling applied in machine learning to address class imbalance.

1743 ii. Image Processing

- 1744 This section of the paper displays the key steps in the image processing by resizing
- 1745 the images to have similar dimensions of 256x256, and the shadows were removed
- 1746 to improve the image quality, and remove noise before proceeding to the deep
- 1747 learning operations.

```

for model_name, (model, param_grid) in models_and_grids.items():
    print(f"Training {model_name}...")
    grid_search = GridSearchCV(estimator=model, param_grid=param_grid, cv=5, scoring='accuracy', return_train_score=False,
                                grid_search.fit(X, y)

    # Get the best estimator's CV results
    best_index = grid_search.best_index_

    # Get fold scores for the best parameters only
    fold_scores = [
        grid_search.cv_results_['split0_test_score'][best_index],
        grid_search.cv_results_['split1_test_score'][best_index],
        grid_search.cv_results_['split2_test_score'][best_index],
        grid_search.cv_results_['split3_test_score'][best_index],
        grid_search.cv_results_['split4_test_score'][best_index]
    ]

    # Calculate the average score across folds
    avg_score = sum(fold_scores) / len(fold_scores)

    # Convert to percentage by multiplying by 100 for display
    fold_scores_percentage = [score * 100 for score in fold_scores]
    avg_score_percentage = avg_score * 100

    # Round each fold score individually
    fold_scores_rounded = [round(score, 2) for score in fold_scores_percentage]

    # Round the average score
    avg_score_rounded = round(avg_score_percentage, 2)

    # Append the fold scores and average score to the list
    model_scores.append({
        'Model': model_name,
        'Fold 1': fold_scores_rounded[0],
        'Fold 2': fold_scores_rounded[1],
        'Fold 3': fold_scores_rounded[2],
        'Fold 4': fold_scores_rounded[3],
        'Fold 5': fold_scores_rounded[4],
        'Average CV Score': avg_score_rounded,
        'Best Parameters': grid_search.best_params_
    })
)

```

Figure A.4: Five-fold cross-validation used to evaluate and tune machine learning model performance.

```

# Process each image
for img_name in image_files:
    img_path = os.path.join(input_male, img_name)
    output_path = os.path.join(output_male, img_name)

    # Read the image
    image = cv2.imread(img_path)
    if image is None:
        print(f"Skipping invalid image: {img_path}")
        continue

    # Resize and pad the image to 256x256
    resized_image = resize_and_pad(image, 256)

```

Figure A.5: Resizing images to 256x256 pixels for consistent input dimensions.

```

# Convert to HSV and apply threshold
frame_HSV = cv.cvtColor(frame, cv.COLOR_BGR2HSV)
frame_threshold = cv.inRange(frame_HSV, (low_H, low_S, low_V), (high_H, high_S, high_V))

# Filling holes
im_floodfill = frame_threshold.copy()
h, w = frame_threshold.shape[:2]
mask = np.zeros((h+2, w+2), np.uint8)
cv.floodFill(im_floodfill, mask, (0, 0), 255)
im_floodfill_inv = cv.bitwise_not(im_floodfill)
mask = frame_threshold | im_floodfill_inv

# Apply morphological operations
kernel = np.ones((3, 3), np.uint8)
mask = cv.morphologyEx(mask, cv.MORPH_OPEN, kernel, iterations=2)
mask = cv.morphologyEx(mask, cv.MORPH_CLOSE, kernel, iterations=4)

# Find contours
contours, _ = cv.findContours(mask, cv.RETR_EXTERNAL, cv.CHAIN_APPROX_SIMPLE)

if contours:
    # Merge contours using convex hull
    hull = cv.convexHull(np.vstack(contours))

    # Create a mask for the shell
    shell_mask = np.zeros_like(frame)
    cv.drawContours(shell_mask, [hull], -1, (255, 255, 255), -1)

    # Create a white background
    white_background = np.ones_like(frame) * 255

    # Combine the shell with the white background
    result = np.where(shell_mask == 255, frame, white_background)
else:
    result = frame # If no contour is found, return the original image

# Save the processed image
cv.imwrite(output_path, result)

```

Figure A.6: Processing the images to remove the shadows.

1748 **iii. Deep Learning**

1749 This section of the paper displays the key steps in deep learning by implementing
1750 random undersampling in addressing class imbalance, data augmentation to create
1751 variability in the dataset, and the CNN layers comprised of three convolution
1752 layers, a flatten, and 2 dense layers.

```
# Get male and female filenames
male_samples = sorted(os.listdir(male_folder))
female_samples = sorted(os.listdir(female_folder))

# Randomly sample 127 male samples to match female sample size
male_samples_to_keep = random.sample(male_samples, undersample_size)

# Copy the selected male samples to the balanced male directory
for file in male_samples_to_keep:
    shutil.copy(os.path.join(male_folder, file), os.path.join(balanced_male_dir, file))

# Copy all female samples to the balanced female directory (since it's already balanced)
for file in female_samples:
    shutil.copy(os.path.join(female_folder, file), os.path.join(balanced_female_dir, file))
```

Figure A.7: Random undersampling applied in deep Learning to balance class distribution in the datasets.

```
def create_data_augmentation():
    return tf.keras.Sequential([
        layers.RandomFlip("horizontal"),
        layers.RandomRotation(0.05),
        layers.RandomZoom(0.05),
    ])
```

Figure A.8: On-the-fly data augmentation used to create variety of random transformation to increase training images.

```
def create_cnn_model(img_width=256, img_height=256):
    model = Sequential([
        layers.Input(shape=(img_width, img_height, 3)),
        layers.Rescaling(1./255),
        layers.Conv2D(16, (3,3), activation='relu'),
        layers.MaxPooling2D(2,2),
        layers.Conv2D(32, (3,3), activation='relu'),
        layers.MaxPooling2D(2,2),
        layers.Conv2D(64, (3,3), activation='relu'),
        layers.MaxPooling2D(2,2),
        layers.Flatten(),
        layers.Dense(128, activation='relu'),
        layers.Dropout(0.5),
        layers.Dense(1, activation='sigmoid')
    ])
    return model
```

Figure A.9: CNN architecture used for training the image dataset.

¹⁷⁵³ **Appendix B**

¹⁷⁵⁴ **Resource Persons**

¹⁷⁵⁵ This section of the paper presents information about the resource persons who
¹⁷⁵⁶ contributed to and assisted the researchers during the data gathering process.

¹⁷⁵⁷ **Dr. Victor Marco Emmanuel N. Ferriols**

¹⁷⁵⁸ Provided blood cockles samples used in this study

¹⁷⁵⁹ Director, University of the Philippines Institute of Aquaculture

¹⁷⁶⁰ vnferriols@up.edu.ph

¹⁷⁶¹

¹⁷⁶² **Ms. Allena Esther D. Arteta**

¹⁷⁶³ Performed spawning of blood cockles samples, assisted the researchers with dis-
¹⁷⁶⁴ section and sex identification.

¹⁷⁶⁵ Research Associate, Institute of Aquaculture

¹⁷⁶⁶ adarteta@up.edu.ph

¹⁷⁶⁷

1768 Ms. LC May C. Gasit

- 1769 Performed spawning of blood cockles samples, assisted the researchers with the
1770 dissection and sex identification
1771 Research Associate, Institute of Aquaculture
1772 lcgasit@up.edu.ph

1773

1774 Sheila G. Untalan

- 1775 Performed spawning of blood cockles samples, assisted the researchers with the
1776 dissection and sex identification
1777 Research Associate, Institute of Aquaculture

1778

1779 Joel M. Fabrigas

- 1780 Assisted the researchers with the dissection and sex identification
1781 Hatchery Staff, Institute of Aquaculture

1782

1783 Paul Andre M. Lopez

- 1784 Assisted the researchers with the dissection and sex identification
1785 Hatchery Staff, Institute of Aquaculture

1786

¹⁷⁸⁷ **Appendix C**

¹⁷⁸⁸ **Data Gathering Documentation**

¹⁷⁸⁹ This section of the paper presents the data gathering process, including spawning,
¹⁷⁹⁰ dissection, sex identification, collection of linear measurements, and image capture
¹⁷⁹¹ from six different camera angles.



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*.

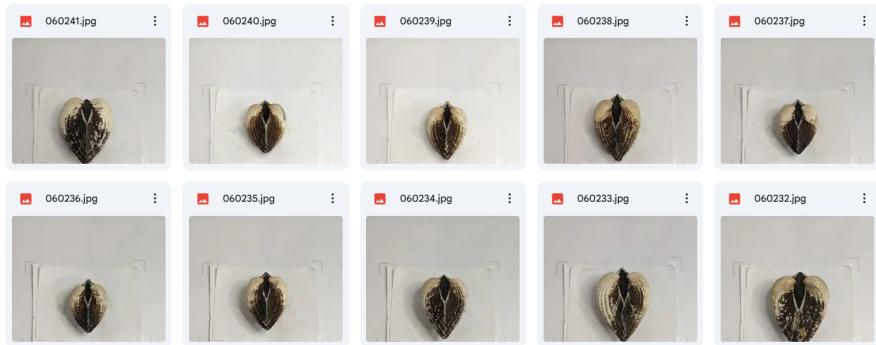


Figure C.7: Captured images of female *T. granosa*.

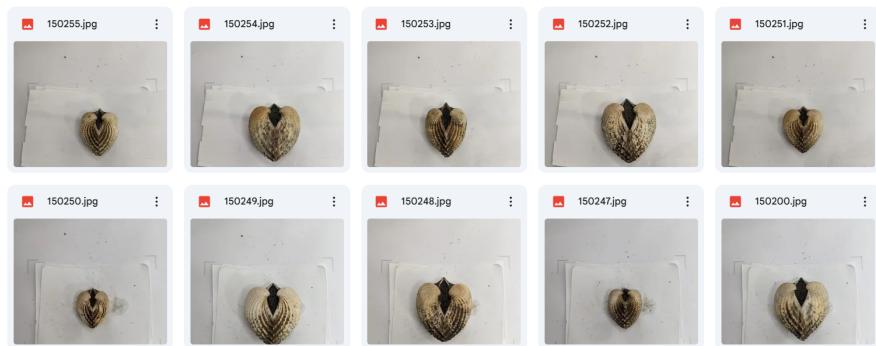


Figure C.8: Captured images of male *T. granosa*.