

¹ MORPHOMETRIC-BASED NON-INVASIVE SEX
² IDENTIFICATION OF BLOOD COCKLES *TEGILLARCA*
³ *GRANOSA* (LINNAEUS, 1758)

⁴ A Special Problem Proposal
⁵ Presented to
⁶ the Faculty of the Division of Physical Sciences and Mathematics
⁷ College of Arts and Sciences
⁸ University of the Philippines Visayas
⁹ Miag-ao, Iloilo

¹⁰ In Partial Fulfillment
¹¹ of the Requirements for the Degree of
¹² Bachelor of Science in Computer Science by

¹³ ADRICULA, Briana Jade
¹⁴ PAJARILLA, Gliezel Ann
¹⁵ VITO, Ma. Christina Kane

¹⁶ Francis DIMZON
¹⁷ Adviser
¹⁸ Victor Marco Emmanuel FERRIOLS
¹⁹ Co-Adviser

²⁰ April 28, 2025

Abstract

22 *Tegillarca granosa* (Linnaeus, 1758), commonly known as blood cockles, is one
23 of the most well-known marine bivalve for its nutritional benefits and economic
24 significance. Determining their sex is essential for maintaining a balanced male-
25 to-female ratio, which is crucial for preventing overexploitation of this shellfish
26 resource. The sex-determining mechanism in the shell morphology of bivalves is
27 challenging macroscopically due to the limited literature regarding this expertise.
28 In addition, no current technologies are employed to classify the sex based on shell
29 morphology. This study proposes a machine learning approach for classifying the
30 sex of blood cockles using various linear measurements (length, width, height,
31 distance between the hinge line, distance between umbos, and rib count) and
32 angles (dorsal, ventral, anterior, posterior, left lateral, and right lateral) collected
33 from male and female specimens. Available machine learning models in MATLAB
34 were trained to discern sexual dimorphism. Among the models, Linear SVM
35 performed best, achieving an accuracy of 69.80%, precision of 69.82%, recall of
36 69.80%, and an F1-score of 69.73%. Feature importance analysis indicated that
37 the distance between the umbos and height were the most significant features.

Keywords: deep learning, supervised machine learning , convolutional
neural network, blood cockle, sex identification, *Tegillarca*
granosa

³⁹ Contents

⁴⁰ 1 Introduction	1
⁴¹ 1.1 Overview	1
⁴² 1.2 Problem Statement	2
⁴³ 1.3 Research Objectives	3
⁴⁴ 1.3.1 General Objective	3
⁴⁵ 1.3.2 Specific Objectives	3
⁴⁶ 1.4 Scope and Limitations of the Research	4
⁴⁷ 1.5 Significance of the Research	5
⁴⁸ 2 Review of Related Literature	6
⁴⁹ 2.1 Background on <i>Tegillarca granosa</i> and Their Importance	7
⁵⁰ 2.2 Current Methods of Sex Identification in <i>Tegillarca granosa</i>	9
⁵¹ 2.3 Machine Learning and Deep Learning in Biological Studies	11
⁵² 2.3.1 Deep Learning for Phenotype Classification in Ark Shells .	12
⁵³ 2.3.2 Geometric Morphometrics and Machine Learning for Species ⁵⁴ Delimitation	12
⁵⁵ 2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning	12
⁵⁶ 2.3.4 Machine Learning for Shape Analysis of Marine Organisms	13

57	2.3.5	Deep Learning for Landmark-Free Morphological Feature Extraction	14
58			
59	2.3.6	Machine Learning for Sex Differentiation in Abalone	15
60	2.3.7	Machine Learning for Geographical Traceability in Bivalves	16
61	2.4	Limitations on Sex Identification in <i>Tegillarca granosa</i>	16
62	2.5	Synthesis of the Study	18
63	3	Research Methodology	21
64	3.1	Sample Collection	21
65	3.2	Ethical Considerations	23
66	3.3	Creating <i>T. granosa</i> Dataset	23
67	3.4	Morphological and Morphometric Characteristics Collection	24
68	3.5	Image Acquisition and Data Gathering	25
69	3.6	Hardware and Software Configuration	26
70	3.7	Morphometric Characteristics Evaluation Using Machine Learning	26
71	3.7.1	Data Preprocessing	27
72	3.7.2	Machine Learning Models Training	28
73	3.8	Morphological Characteristics Evaluation Using Deep Learning	29
74	3.8.1	Convolutional Neural Network	30
75	3.8.2	CNN Training	32
76	3.9	Evaluation Metrics	34
77	4	Results and Discussions	36
78	4.1	Machine Learning Analysis	36
79	4.1.1	Data Exploration	36

80	4.1.2 Statistical Analysis	37
81	4.1.3 Feature Importance Analysis	38
82	4.1.4 Performance Evaluation	39
83	4.1.5 Confusion Matrix Analysis	40
84	4.2 Deep Learning Analysis	41
85	4.2.1 Baseline Model	41
86	4.2.2 Comparison of Individual and Combined Angles	42
87	4.2.3 Training Result and Hyperparameter Tuning	42
88	4.2.4 Proposed Model	44
89	4.2.5 Learning Rates and Training Behavior per Fold	44
90	4.2.6 Visualizations	45
91	4.3 Discussions	49
92	5 Conclusion and Recommendations	51
93	5.1 Conclusion	51
94	5.2 Recommendations	52
95	References	54
96	A Data Gathering Documentation and Supplementary Analysis	59

⁹⁷ List of Figures

98	2.1 Diagram of <i>Tegillarca granosa</i> Anatomy	7
99	3.1 Male and Female <i>Tegillarca granosa</i> shells	22
100	3.2 Different Views of the <i>T. granosa</i> Shell Captured	24
101	3.3 Linear Measurements of <i>Tegillarca granosa</i> shell.	25
102	3.4 Image Acquisition Setup for <i>T. granosa</i> Samples	26
103	3.5 Data Preprocessing Pipeline	27
104	3.6 Diagram of k-fold cross-validation with k = 5	29
105	3.7 Shadows removed from male samples at different angles	30
106	3.8 Shadows removed from female samples at different angles	30
107	3.9 Architecture of Convolutional Neural Network (CNN)	31
108	3.10 Diagram of stratified k-fold cross-validation with k=5	32
109	3.11 Data Augmentation Techniques	33
110	4.1 Correlation heatmap of morphometric features with the sex of <i>T.</i> <i>granosa</i>	37
111		
112	4.2 Feature Importance Scores Using the Kruskal-Wallis Test	38
113	4.3 Feature Importance Scores Using the Kruskal-Wallis Test	40
114	4.4 Training and Validation Accuracy per Fold	45

115	4.5 Average Training and Validation Accuracy Across Folds	46
116	4.6 Training and Validation Loss per Fold	46
117	4.7 Average Training and Validation Loss Across Folds	47
118	4.8 Confusion Matrix for Final Model Predictions	48
119	4.9 ROC Curve and AUC Score	48
120	A.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i>	59
121	A.2 Separating Male and Female Samples After Spawning of <i>Tegillarca</i> <i>granosa</i>	59
122	A.3 Sex Identified Female Through Dissecting of <i>Tegillarca granosa</i>	60
123	A.4 Sex Identified Male Through Dissecting of <i>Tegillarca granosa</i>	60
124	A.5 Linear Measurements of Female <i>Tegillarca granosa</i>	60
125	A.6 Linear Measurements of Male <i>Tegillarca granosa</i>	61
126	A.7 Distribution of the Features of <i>Tegillarca granosa</i>	61
127		

¹²⁸ List of Tables

¹²⁹	2.1 Comparison of the Methods Used in Bivalves Studies	19
¹³⁰	4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison	38
¹³¹	4.2 Performance Metrics for Models with All 13 Features	39
¹³²	4.3 Performance Metrics for Models with 5 Features	39
¹³³	4.4 Performance Metrics for Unbalanced vs. Balanced Datasets (Batch Size: 16, Epochs: 20)	41
¹³⁴		
¹³⁵	4.5 Performance Metrics for Individual and Combined Angles (Batch Size: 16, Epochs: 20)	42
¹³⁶		
¹³⁷	4.6 Effect of Batch Size and Epoch Values on CNN Model Performance	43
¹³⁸		
¹³⁹	4.7 Performance Metrics for Different Activation Functions (Batch Size: 32, Epochs: 50)	43
¹⁴⁰		
¹⁴¹	4.8 Per-Fold Performance Metrics (Batch Size: 32, Epochs: 50, Activation Function: ReLU)	44
¹⁴²		
¹⁴³	4.9 Learning Rate Reductions, Early Stopping, and Best Epochs per Fold During 5-Fold Cross-Validation	45

¹⁴⁴ **Chapter 1**

¹⁴⁵ **Introduction**

¹⁴⁶ **1.1 Overview**

¹⁴⁷ The Philippines is a global center of marine biodiversity and has established aqua-
¹⁴⁸ culture as a significant contributor to total fishery production (Aypa & Baconguis,
¹⁴⁹ 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-
¹⁵⁰ ally and is the 11th largest seafood producer in the world. Aquaculture is deeply
¹⁵¹ integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-
¹⁵² duction of various aquatic species, including bivalves. Among these, blood cockles
¹⁵³ (*Tegillarca granosa*) hold considerable economic and environmental significance,
¹⁵⁴ making it essential to ensure sustainable production and population balance.

¹⁵⁵ Maintaining a balanced male-to-female ratio of blood cockles is crucial to pre-
¹⁵⁶ vent overharvesting and ensure sustainability. An imbalanced ratio can lead to
¹⁵⁷ overexploitation and negatively impact the population's viability. However, there
¹⁵⁸ is limited literature on *T. granosa* that provides a thorough understanding of its
¹⁵⁹ sex-determining mechanisms, particularly regarding sexual dimorphism based on
¹⁶⁰ morphological and morphometric characteristics (Breton, Capt, Guerra, & Stew-
¹⁶¹ art, 2017).

¹⁶² Currently, sex determination methods for blood cockles are invasive, including
¹⁶³ dissection and histological examinations, which often result in the death of the
¹⁶⁴ species. While there is growing literature on sex identification in aquaculture
¹⁶⁵ commodities using machine learning and deep learning, there is a notable scarcity
¹⁶⁶ of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

¹⁶⁷ This study, titled "Morphometric-Based Non-Invasive Sex Identification of

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

175 1.2 Problem Statement

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

184 Current methods for determining the sex of *T. granosa* are invasive and in-
185 volve dissection, which requires cutting open the shell to visually inspect the
186 gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-
187 quently leads to their death. Another method is histological examination, where
188 tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,
189 2021). Both approaches are labor-intensive and time-consuming, and can pose
190 risks to population management, particularly when maintaining a balanced sex
191 ratio for breeding programs is essential. Moreover, these invasive methods require
192 specialized technical skills for accurate execution. Resource-limited aquaculture
193 operations face significant challenges in accessing the necessary laboratory equip-
194 ment, such as microscopes and staining tools, complicating the process.

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

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

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

206 1.3 Research Objectives

207 1.3.1 General Objective

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

213 1.3.2 Specific Objectives

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

- 217 1. To collect and organize a comprehensive dataset of *T. granosa* which will
218 include high-quality images and relevant morphological measurements that
219 will serve as the basis for the machine-learning model.
- 220 2. To develop and implement machine learning models that can classify the
221 sex of *T. granosa* based on the collected linear measurements and images of
222 different angles of the sample.
- 223 3. To evaluate the performance of the models used using performance metrics
224 such as accuracy, precision, recall, and F1-score.
- 225 4. To develop a system that can identify the sex of *T. granosa* based on its
226 morphological characteristics.

227 1.4 Scope and Limitations of the Research

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

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

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

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

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

262 1.5 Significance of the Research

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

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

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

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

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

²⁸⁶ **Chapter 2**

²⁸⁷ **Review of Related Literature**

²⁸⁸ Aquaculture is the fastest-growing industry in animal food production and has
²⁸⁹ great potential as a sustainable solution to global food security, nutrition, and
²⁹⁰ development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*
²⁹¹ *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-
²⁹² tegrated into the livelihoods of Filipinos, not only through fish cultivation but
²⁹³ also through the production of other aquatic species, including mollusks, oysters,
²⁹⁴ clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood
²⁹⁵ clams *Tegillarca granosa*, have economic and environmental significance. It has
²⁹⁶ been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-
²⁹⁷ vesting and maintain the optimal ratio to preserve the population and production
²⁹⁸ of the blood cockles.

²⁹⁹ The members of the Arcidae Family, including *T. granosa* are important
³⁰⁰ sources of food and livelihood. Cockle aquaculture meets rising demands, however,
³⁰¹ it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,
³⁰² 2023). To solve the problem, researchers exert a considerable amount of effort,
³⁰³ developing a broader understanding of bivalves, including their sex-determining
³⁰⁴ mechanism, due to their notable importance in terms of diversity, environmental
³⁰⁵ benefits, and economic and market importance (Breton et al., 2017). Despite the
³⁰⁶ promising idea of identifying sex, there is limited research reported in terms of
³⁰⁷ sexual dimorphism, making it harder to distinguish through its morphological and
³⁰⁸ morphometric characteristics.

³⁰⁹ By addressing the challenges in the sex identification of *T. granosa*, it would be
³¹⁰ able to address one problem at a time. Currently, there are no recent documented
³¹¹ publications that integrate machine learning and computer vision in characterizing
³¹² sexual dimorphism, reducing complexity, variability in sex determination, and

³¹³ differentiation mechanisms in bivalves, including *T. granosa* specifically.

³¹⁴ 2.1 Background on *Tegillarca granosa* and Their ³¹⁵ Importance

³¹⁶ *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.
³¹⁷ In the Philippines, it is commonly known as a Litob, a marine bivalve species from
³¹⁸ the family Arcidae. Litob is widely distributed in the world including Southeast
³¹⁹ Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest
³²⁰ (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat
³²¹ as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the
³²² blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood
³²³ phenotype where it serves two purposes the hemocyte carries oxygen around the
³²⁴ body and strengthens immune defenses. In addition, it possesses a unique ability
³²⁵ to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

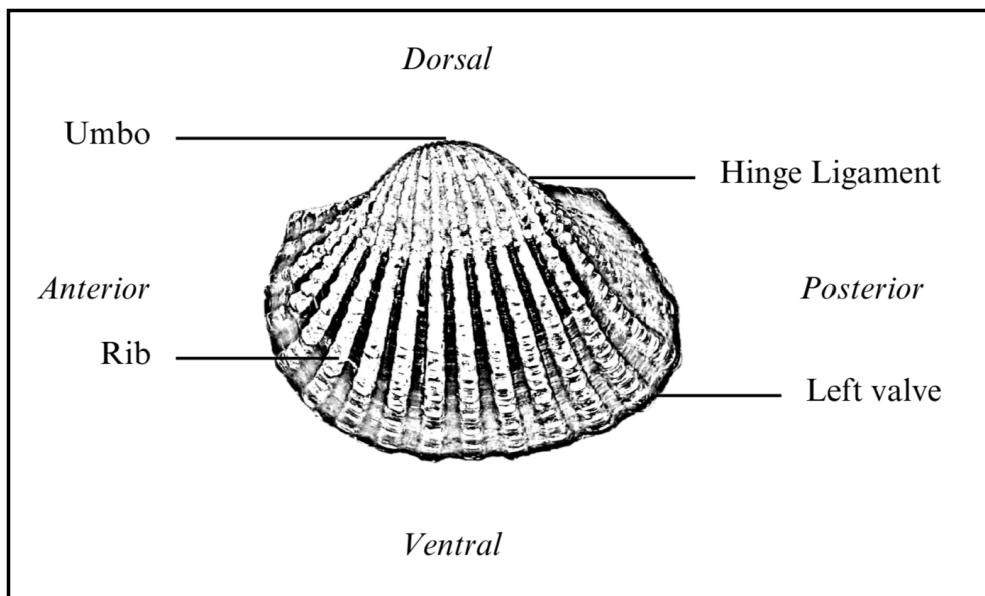


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

³²⁶ *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both
³²⁷ valves being equal in size but asymmetrical from the hinge. The top edge of
³²⁸ the dorsal margin is straight, while the front is rounded and slopes downward,
³²⁹ with its back being obliquely rounded with a concave bottom edge. It has a
³³⁰ narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,
³³¹ although some may be incomplete. The shell's outer layer, or the periostracum, is

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

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

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

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

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

379 **2.2 Current Methods of Sex Identification in *Tegillarca granosa***

380

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

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

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

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

406 evidence from the astartid *Nicanella* (Lower Jurassic, southern Germany)," the
407 researchers utilized Principal Component Analysis and Fourier Analysis as a non-
408 invasive method that investigates sexual expression in the *Nicanella rakoveci*. In
409 the study, researchers discovered that the bivalves with crenulations were found to
410 have a different shell shape, which made them more inflated than those without
411 crenulations. This suggests that when they became females, they adapted to
412 hold more eggs rather than for protection from predators as previously thought.
413 The formation of crenulations is likely part of the genetic process that controls
414 both the sex change and the changes in shell structure (Karapunar et al., 2021).
415 Overall, the findings demonstrate that the genetic mechanisms for sex change and
416 shell morphology in bivalves existed as early as the Early Jurassic, contributing
417 to our understanding of bivalve diversity and evolution. Thus, the researchers
418 concluded that crenulations serve as a morphological marker for identifying the
419 sex and reproductive stage of these bivalves (Karapunar et al., 2021).

420 On the other hand, invasive techniques such as histological analysis offer a
421 more thorough but harmful method for determining the sex of *T. granosa*. A
422 study on the Spawning Period of Blood Cockle *Tegillarca granosa* (Linnaeus,
423 1758) in Myeik Coastal. 240 blood cockle samples were examined for sex and
424 gonad maturity stages using histological examination, with shell lengths ranging
425 from 26-35mm and shell weights from 8.1-33g. For histological analysis, the whole
426 soft tissues were removed from the shell and the flesh containing most parts of
427 the gonads was fixed in formalin, dehydrated in an upgraded series of ethanol,
428 and cleared in xylene. This invasive method allows for precise identification of
429 the gonadal maturation stages based on the cellular and structural changes in the
430 gonads.

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

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

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

464 **2.3 Machine Learning and Deep Learning in Bi- 465 ological Studies**

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

476 **2.3.1 Deep Learning for Phenotype Classification in Ark
477 Shells**

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

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

491 **2.3.2 Geometric Morphometrics and Machine Learning for
492 Species Delimitation**

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

503 **2.3.3 Contour Analysis in Mollusc Shells Using Machine
504 Learning**

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

508 evolutionary influences. The researchers stated that classic morphometric meth-
509 ods may not accurately depict morphological features of the shell, especially when
510 using the angular decomposition of the contour. The current research examined
511 and compared the robustness of the contour analysis using wavelet transformed
512 and Elliptic Fourier descriptors for gastropod shells with enlarged spines. For
513 that, the researchers analyzed two geographically and ecologically separated pop-
514 ulations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed
515 that contour analysis of gastropod shells with enlarged spines can be analyzed
516 using both methodologies, but the wavelet analysis provided better local discrim-
517 ination. From an ecological perspective, shells with various sizes of spines in both
518 areas indicate the broad adaptability of the species.

519 2.3.4 Machine Learning for Shape Analysis of Marine Or- 520 ganisms

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

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

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

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

567 2.3.5 Deep Learning for Landmark-Free Morphological Fea- 568 ture Extraction

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

579 2.3.6 Machine Learning for Sex Differentiation in Abalone

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

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

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

613 **2.3.7 Machine Learning for Geographical Traceability in**
614 **Bivalves**

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

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

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

641 **2.4 Limitations on Sex Identification in *Tegillarca***
642 ***granosa***

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

647 perature shock, is the most natural and least invasive method for bivalves (Aji,
648 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water
649 from hot to cold with a specific temperature requires deliberate and careful ma-
650 nipulation of the temperature over a specific period and would require constant
651 management and monitoring.

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

660 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous
661 (synchronous or functional) characteristics. Sequential hermaphrodites switch
662 genders after being male or female for one or multiple yearly cycles. (Heller,
663 1993; Gosling, 2004; Collin, 2013). Sex change and consecutive hermaphroditism
664 have been observed in different bivalve species, including Ostreidae, Pectinidae,
665 Veneridae, and Patellidae. However, macroscopically differentiating bivalve sex is
666 challenging. The only way it may be identified is through histological analysis of
667 gonad remains but to do so there is an act of killing the organism (Coe, 1943;
668 Gosling, 2004). Verification of sex change in bivalves to classify whether male or
669 female while they are alive is challenging since they need to be re-confirmed and
670 re-evaluated to be the same individual after a year.

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

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

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

690 **2.5 Synthesis of the Study**

691 This section of the paper summarizes the technologies used in the different studies
692 related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex
693 Identification of Blood Cockles *Tegillarca granosa* (Linnaeus, 1758).

Author	Technology / Method Used	Description of Problem	Pros	Cons
D. V. Miranda and V. M. E. N. Ferriols	Temperature shock	No recent studies are available on the production and rearing of <i>T. granosa</i> in the Philippines.	Employed less invasive techniques which minimize the stress in <i>T. granosa</i> and can lead to better survival rates.	Time-consuming as the entire process from fertilization to the spat stage took 120 days.
Karapunar, Baran and Werner, W. and Fürsich, F. T. and Nützel, A.	Morphometric analysis, microscope imaging, principal component analysis (PCA), and Fourier shape analysis	To address the observed shell dimorphism in the Early Jurassic bivalve <i>Nicanella rakoveci</i> , namely the presence or lack of crenulations on the ventral shell margin, and whether these variations represent sexual dimorphism and sequential hermaphroditism.	The methods used reveal significant morphological differences with regard to sexual dimorphism.	There could be misinterpretation of the shape differences of bivalves due to the constraints and resolution of technologies used.
K. May and C. Maung and E. Phyu and N. Tun	Histological examination	The need to understand the reproductive period of <i>T. granosa</i> in Myeik to ensure sustainable aquaculture and to prevent overexploitation.	Method used allows for accurate sex identification based on the histological characteristics and color of the gonads.	Invasive technique used to determine the sex of <i>T. granosa</i> through gonad histological analysis.
E. Kim and S.-M. Yang and J.-E. Cha and D.-H. Jung and H.-Y. Kim	Convolutional neural network (CNN) models, VGGNet, Inception-ResNet, SqueezeNet	Traditional methods of recognizing and classifying ark shell species based on shell traits are time-consuming and inaccurate.	Automated classification of the three ark shells using a deep learning model obtained an accuracy of 92.4%.	Challenges may arise with certain ark shells that share similar morphology.
Mathieu Quemu and S. A. Trewick and F. Brescia and M. Morgan-Richards	Neural network analysis (supervised learning) and Gaussian mixture models (unsupervised learning)	To determine whether the shape and size of the snail's shells can distinguish between two <i>Placostylus</i> species, particularly in groups that appear to be hybrids.	Combining geometric morphometrics and machine learning effectively answers biological issues, providing insights into species classification and possible hybridization.	Difficulty classifying intermediate phenotypes, with potential for overfitting and misclassification in both learning methods.
V. M. Tusset and E. Galimany and A. Farrés and E. Marco-Herrero and J. L. Otero-Ferrer and A. Lombarte and M. Ramón	Wavelet functions and Elliptic Fourier descriptors	Addresses the difficulty of accurately defining phenotypic diversity in gastropod shells.	Advanced contour analysis methods allow accurate differentiation of gastropod shell forms.	Cannot clarify the causes of phenotypic variation in the two populations studied.
Fedor Lishchenko and Jones, J. B.	Landmark- and outline-based Geometric Morphometric methods	To address difficulties in differentiating between stocks of marine organisms to prevent misidentification that could affect conservation and management.	Shape analysis improves taxonomic classification precision and offers close distinction between related species or organisms.	Landmark-based methods can be sensitive to landmark placement.
M. Tsutsumi and N. Saito and D. Koyabu and C. Furusawa	Morphological regulated variational AutoEncoder (Morpho-VAE)	The need for reliable, landmark-free methods, such as a modified variational autoencoder, to extract and decipher complex shapes from image data.	Employs dimension reduction and feature extraction, making it a user-friendly tool for biology non-experts.	Limited sample size in certain families presented challenges.
Barrera-Hernandez, R. and Barrera-Soto, V. and Martinez-Rodriguez, J. L. and Ríos-Alvarado, A. B. and Ortiz-Rodriguez, F.	Machine learning algorithms	Identifying the sex of abalones is challenging for producers applying specific growth or preservation strategies.	Machine learning algorithms accurately classify abalone sex into three categories: male, female, and immature.	Selected features may not fully capture the complexity of abalone morphology.
Concepcion, R. and Guillermo, M. and Tanner, S. E. and Fonseca, V. and Duarte, B.	EfficientNet-Bo, ResNet101, MobileNetV2, InceptionV3	Addresses the difficulty of accurately tracing bivalve harvesting origins using computer vision and machine learning algorithms to enhance seafood traceability and combat food fraud.	Non-invasive, image-based tools for bivalve traceability provide faster, cheaper, and equally accurate alternatives to traditional chemical analysis methods.	Small sample size (only 30 cockles) limits model reliability.

Table 2.1: Comparison of the Methods Used in Bivalves Studies

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

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

⁷⁰⁸ Chapter 3

⁷⁰⁹ Research Methodology

⁷¹⁰ This chapter discussed the materials and methods employed in the study, focusing
⁷¹¹ on the development requirements, as well as the software and programming
⁷¹² languages utilized. It also detailed the overall workflow in conducting the study,
⁷¹³ Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca*
⁷¹⁴ *granosa* (Linnaeus), 1758) using machine learning and deep learning technologies.

⁷¹⁵ Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,
⁷¹⁶ oversaw the overall workflow and conduct of the experiment. The researchers were
⁷¹⁷ also guided by research associates LC Mae Gasit and Allena Esther Artera. Con-
⁷¹⁸ sequently, the entire dataset collection process was conducted at the University of
⁷¹⁹ the Philippines Visayas hatchery facility.

⁷²⁰ The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical
⁷²¹ Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-
⁷²² tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-
⁷²³ ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine
⁷²⁴ Learning, (8) Morphological Characteristics Evaluation Using Deep Learning, and
⁷²⁵ (9) Evaluation Metrics

⁷²⁶ 3.1 Sample Collection

⁷²⁷ The collection of *T. granosa* samples used in this study was part of an ongoing
⁷²⁸ research project by UPV DOST-PCAARRD titled "Establishment of the Center
⁷²⁹ for Mollusc Research and Development: Development of Spawning and Hatchery
⁷³⁰ Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."

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

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

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

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

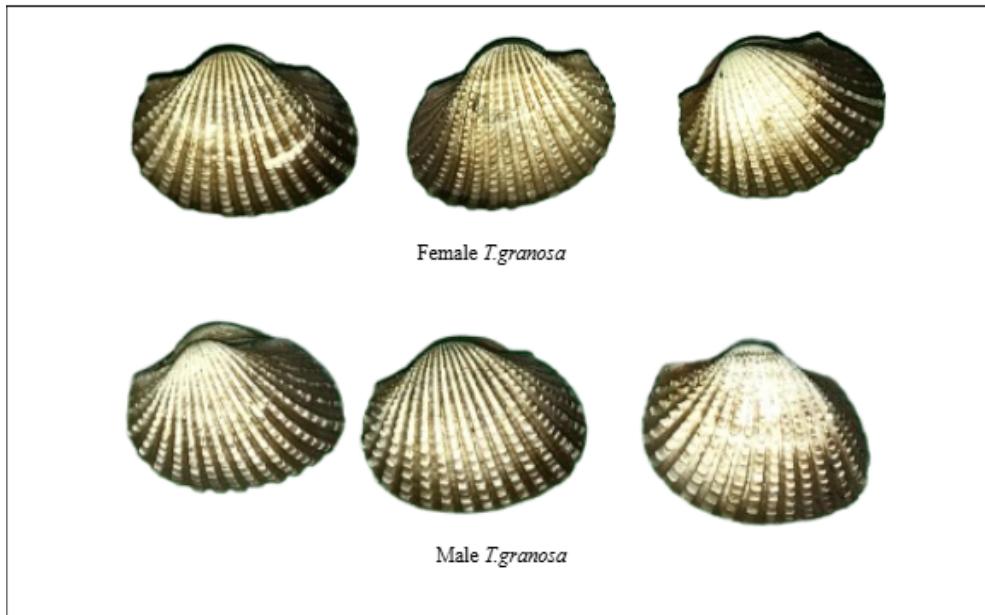


Figure 3.1: Male and Female *Tegillarca granosa* shells

750 3.2 Ethical Considerations

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

757 3.3 Creating *T. granosa* Dataset

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

766 Images of each sample were captured and saved in JPG format using a stan-
767 dardized file naming convention that included the sample's sex, the shell's ori-
768 entation or view, and its corresponding number out of the 271 total samples. File
769 names for female *T. granosa* samples began with "0", while those for male sam-
770 ples began with "1". Each file name also included one of the six captured views:
771 (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right
772 lateral (refer to Figure 3.2), followed by a unique sample number. For exam-
773 ple, "010001" denoted the first female sample taken from the dorsal view, while
774 "110001" represented the first male sample from the same view. This naming
775 convention was implemented to prevent data leakage and ensure accurate labeling
776 of images according to their respective samples.

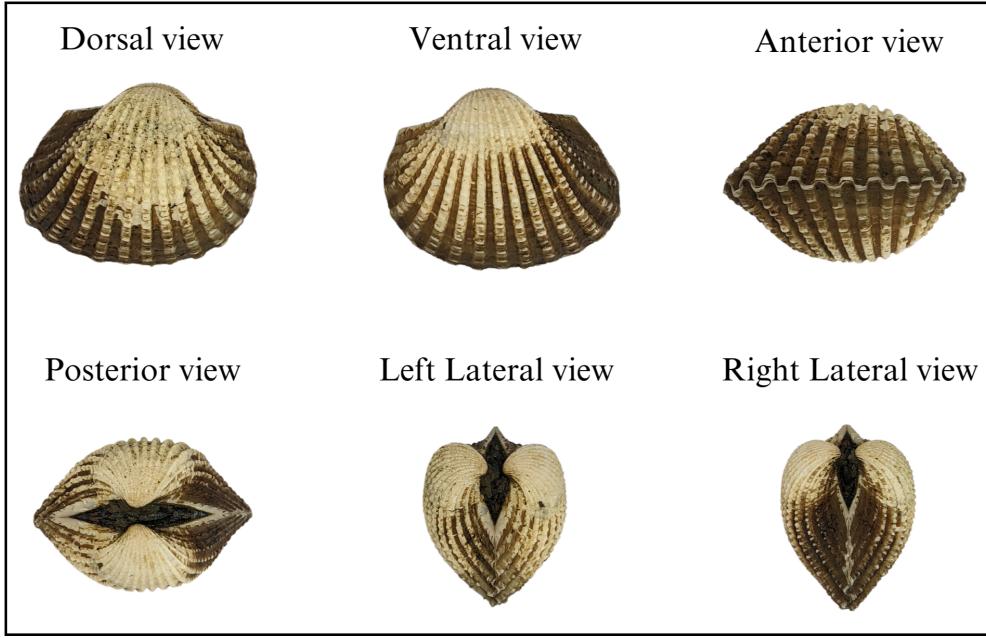


Figure 3.2: Different Views of the *T. granosa* Shell Captured

⁷⁷⁷ 3.4 Morphological and Morphometric Characteristics Collection

⁷⁷⁹ Morphology refers to biological form and is one of the most visually recognizable
⁷⁸⁰ phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).
⁷⁸¹ In this study, morphological characteristics describe the structural features of
⁷⁸² *T. granosa*, focusing on measurable attributes such as shape, size, and color.
⁷⁸³ Morphometric characteristics, on the other hand, refer to specific quantifiable
⁷⁸⁴ features of *T. granosa*, including length, width, height, hinge line length, distance
⁷⁸⁵ between the umbos, and rib count. As stated by the researchers, quantifying and
⁷⁸⁶ characterizing these traits is essential for understanding and visualizing variations
⁷⁸⁷ in *T. granosa* morphology.

⁷⁸⁸ The researchers measured the height, width, and length of *T. granosa* using
⁷⁸⁹ a Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the
⁷⁹⁰ corresponding measurement diagram. Length (A) refers to the distance from the
⁷⁹¹ anterior to the posterior of the shell. Width (B) is defined as the widest span
⁷⁹² across the shell from the left to the right valve. Height (C) measures the distance
⁷⁹³ from the base to the apex of the shell. In addition, the hinge line length (D) near
⁷⁹⁴ the hinge and the distance between the umbos (E) were recorded.

⁷⁹⁵ Reyment and Kennedy (1998) emphasized that including rib count as supple-

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

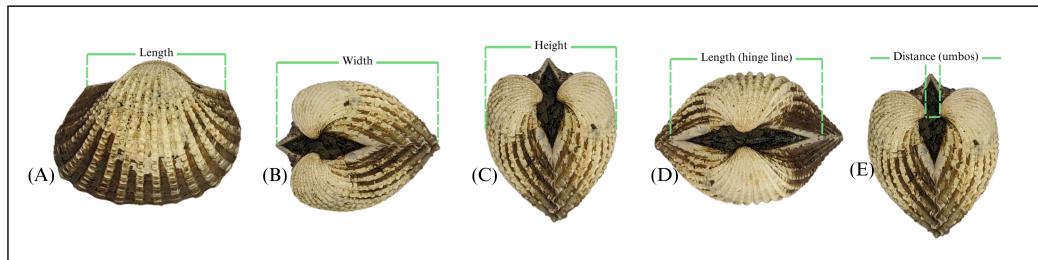


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

3.5 Image Acquisition and Data Gathering

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

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



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

814 3.6 Hardware and Software Configuration

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

826 3.7 Morphometric Characteristics Evaluation Us- 827 ing Machine Learning

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

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

837 **3.7.1 Data Preprocessing**

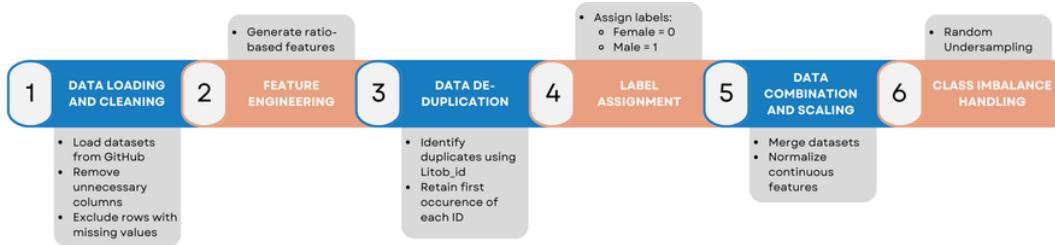


Figure 3.5: Data Preprocessing Pipeline

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

841 ***Data Loading and Cleaning***

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

846 ***Feature Engineering***

847 Additional ratio-based features were generated to augment the existing mea-
848 surements. These included the length-to-width ratio, length-to-height ratio, width-
849 to-height ratio, hinge line length-to-length ratio, umbos distance-to-length ratio,
850 umbos distance-to-height ratio, and rib density. These derived features aimed to
851 emphasize shape characteristics independent of size, improving the models' ability
852 to distinguish morphological differences between sexes.

853 ***Data De-duplication***

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

858 ***Label Assignment***

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

862 ***Data Combination and Scaling***

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

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

872 ***Class Imbalance Handling***

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

878 **3.7.2 Machine Learning Models Training**

879 ***Model Selection and Hyperparameter Tuning***

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

885 ***Cross-Validation***

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

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

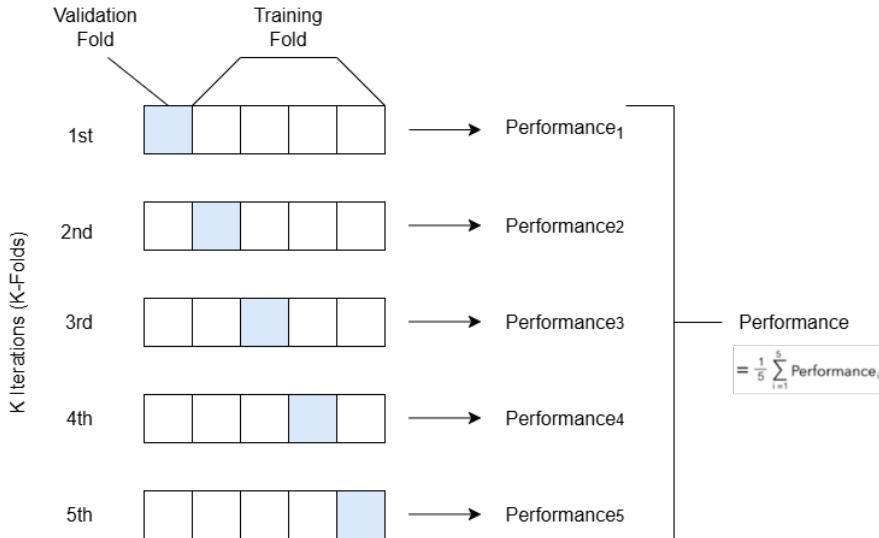


Figure 3.6: Diagram of k-fold cross-validation with $k = 5$

891 3.8 Morphological Characteristics Evaluation Us- 892 ing Deep Learning

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

897 *Image Preprocessing*

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

904 *Adjusting Dimensions*

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

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

910 ***Background Removal***

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

916 ***Shadow Removal***

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

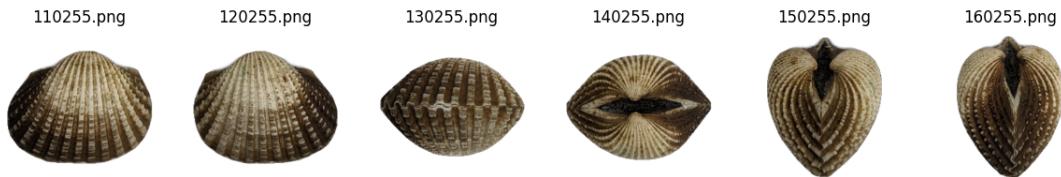


Figure 3.7: Shadows removed from male samples at different angles

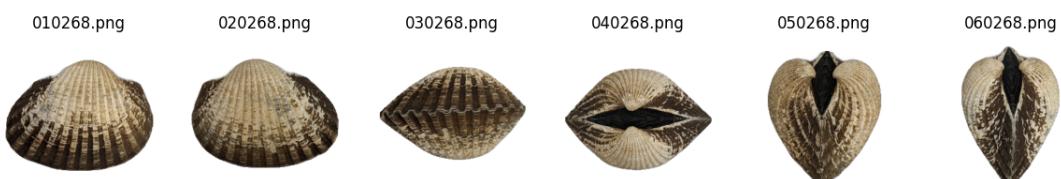


Figure 3.8: Shadows removed from female samples at different angles

921 **3.8.1 Convolutional Neural Network**

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

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

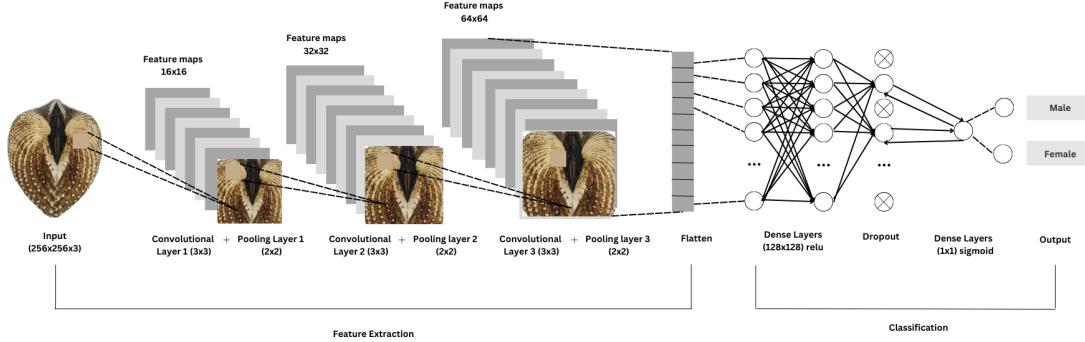


Figure 3.9: Architecture of Convolutional Neural Network (CNN)

930 ***Convolution Layer***

931 The convolution layers of CNN extract the features from the input image
 932 through the convolution operation. This study uses four convolution layers with
 933 a 3x3 kernel size and filter sizes of 16, 32, 64, and 128. The first layer extracts
 934 the low-level features, such as edges, lines, and corners, while the deeper layers
 935 iteratively extract more complex information from these low-level features. The
 936 ReLU activation function is used as the baseline for this model, and experiments
 937 are conducted with different activation functions, such as ELU and PReLU, to
 938 evaluate their impact on learning complex patterns within the data.

939 ***Pooling Layer***

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

943 ***Fully Connected and Dropout***

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

948 A large number of parameters in the training process can lead to overfitting.
 949 It occurs when the model learns the training data too well, including its noise and
 950 irrelevant details. This results in poor performance on unseen data. To mitigate
 951 the overfitting, the dropout layer was employed. Dropout works by temporarily

952 discarding a portion of the neurons in the network with probability p ($0 < p < 1$).
953 During this process, these neurons do not participate in the forward propagation
954 process of CNN and the backward propagation process (Cui et al., 2020).

955 3.8.2 CNN Training

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

961 *Data Splitting*

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

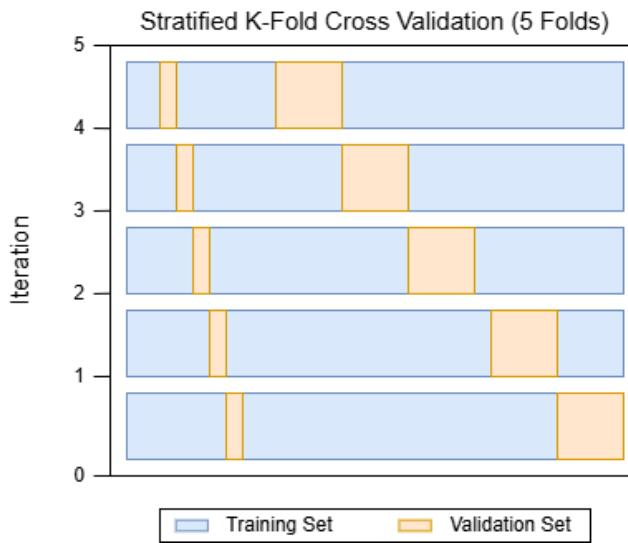


Figure 3.10: Diagram of stratified k-fold cross-validation with k=5

968 *Data Augmentation*

969 Before model training, online data augmentation was applied exclusively to
970 the training data within each fold, creating new data variations on the fly. The

971 augmentations included random horizontal flips, slight rotations, and zoom trans-
972 formations to enhance data diversity and improve model generalization (Awan,
973 2022). All augmentation was strictly applied only to the training subset of each
974 fold to prevent data leakage and maintain the validity of the results (*Figure 3.11*).

975 On-the-fly data augmentation (OnDAT) generates augmented data during
976 each iteration, exposing the model to constantly changing data variations. Aug-
977 menting the original data allows better exploration of the underlying data genera-
978 tion process and has the potential to prevent the model from overfitting spurious
979 patterns, thereby improving performance (Cerqueira, Santos, Baghoussi, & Soares,
980 2024).

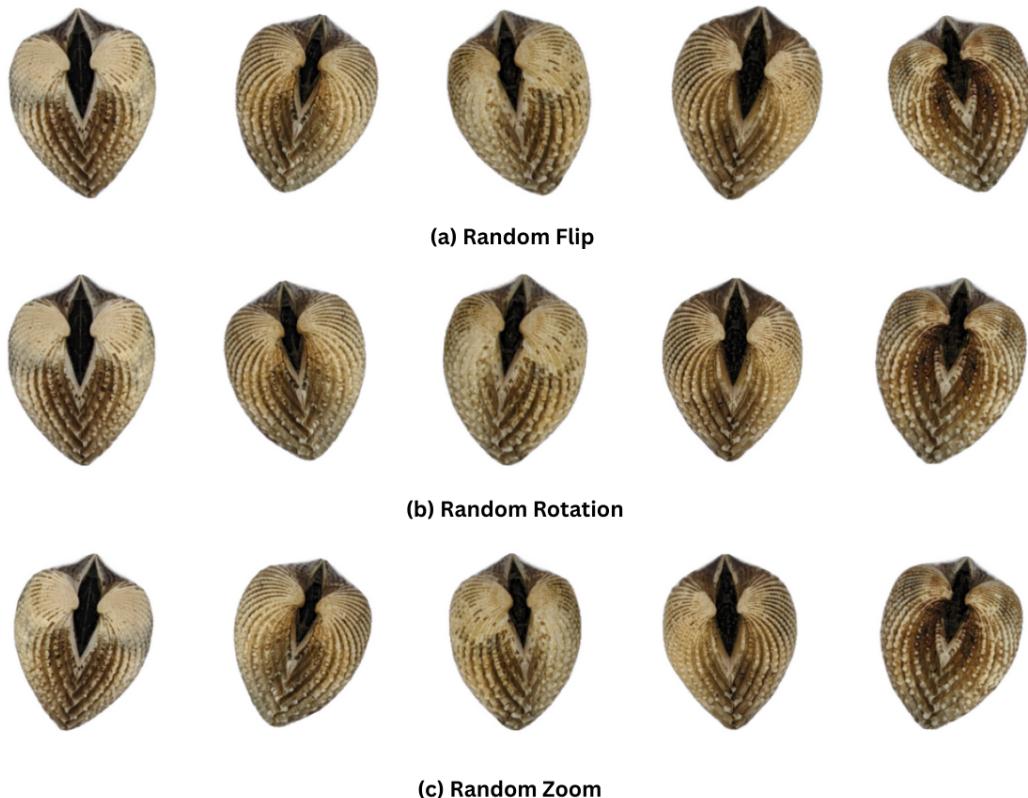


Figure 3.11: Data Augmentation Techniques

981 ***Training Procedure***

982 During the training process, model performance per fold was carefully mon-
983 itored. One important thing to observe is the consistency in the performance,
984 whether the model is still learning or is at high risk of overfitting. Early stopping
985 was applied to ensure the stable performance of the model across folds. This
986 technique allows for monitoring the training of the neural network, stopping when

987 the performance metrics, in this case, validation loss, cease to improve. Furthermore,
988 to enhance the learning process, ReduceLROnPlateau was applied, which
989 decreased the learning rate if there was no improvement in the model for a specified
990 number of epochs (Team, n.d.).

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

997 3.9 Evaluation Metrics

998 Evaluating the performance of a binary classification model is essential, and selecting
999 appropriate metrics depends on the specific requirements of the user. The
1000 performance of both supervised machine learning and deep learning models will
1001 be measured using several key metrics, including accuracy, precision, recall, F1
1002 score, and the AUC-ROC score.

1003 Accuracy (ACC) is the ratio of the overall correctly predicted samples to the
1004 total number of examples in the evaluation dataset (Cui et al., 2020). It measures
1005 the overall correctness of the model in predicting both male and female blood
1006 cockles. This metric provides insight into how well the model performs across all
1007 classifications. The formula for accuracy is:

$$1008 \text{ACC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

1008 Precision (PREC) is the ratio of correctly predicted positive samples to all
1009 samples assigned to the positive class (Cui et al., 2020). This metric helps in
1010 evaluating the fairness of the model and prevents the misclassification of blood
1011 cockles as it identifies potential inaccuracies or biases. The formula for precision
1012 is:

$$1008 \text{PREC} = \frac{\text{True positive samples}}{\text{Samples assigned to positive class}} = \frac{TP}{TP + FP} \quad (3.2)$$

1013 Recall (REC), also known as sensitivity or the true positive rate (TPR), is the

1014 ratio of correctly predicted positive cases to all the actual positive samples (Cui
1015 et al., 2020). It represents the ability of the model to correctly identify positive
1016 male and female samples. The formula for recall is:

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

1017 The F1 score is the harmonic mean of precision and recall, which penalizes
1018 extreme values of either of the two metrics (Cui et al., 2020). It is particularly
1019 useful when the class distribution is imbalanced. The formula for the F1 score is:

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

1020 The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is
1021 a performance measurement for classification problems, particularly used in deep
1022 learning in this study. The ROC curve is a plot of the true positive rate (recall)
1023 against the false positive rate (1 - specificity), and the AUC score quantifies the
1024 overall ability of the model to discriminate between positive and negative classes.
1025 A higher AUC indicates better model performance. (Nahm, 2022)

¹⁰²⁶ **Chapter 4**

¹⁰²⁷ **Results and Discussions**

¹⁰²⁸ This chapter presents the results from the machine learning and deep learning
¹⁰²⁹ analyses conducted on the preprocessed dataset. It includes an evaluation of
¹⁰³⁰ various machine learning classifiers and the application of deep learning models
¹⁰³¹ for image-based classification. The primary focus is on identifying key predictors
¹⁰³² and assessing classification performance for sex identification in *T. granosa*.

¹⁰³³ **4.1 Machine Learning Analysis**

¹⁰³⁴ This chapter outlines the results of preprocessing, training of machine learning
¹⁰³⁵ models, and feature importance analysis, all conducted in Google Colab using
¹⁰³⁶ Python. The dataset was preprocessed in Colab, and the training and evaluation
¹⁰³⁷ of various classifiers were performed entirely within this environment. This part of
¹⁰³⁸ the paper includes five subsections: data exploration, statistical analysis, feature
¹⁰³⁹ importance analysis, performance evaluation, and confusion matrix analysis.

¹⁰⁴⁰ **4.1.1 Data Exploration**

¹⁰⁴¹ Exploratory data analysis was performed to characterize the dataset using visu-
¹⁰⁴² alizations to understand the patterns and correlations within the data. A corre-
¹⁰⁴³ lation heatmap was created to assess the relationship between the predictors and
¹⁰⁴⁴ the target variable.

¹⁰⁴⁵ The heatmap (see Figure 4.1) revealed three features most correlated with the

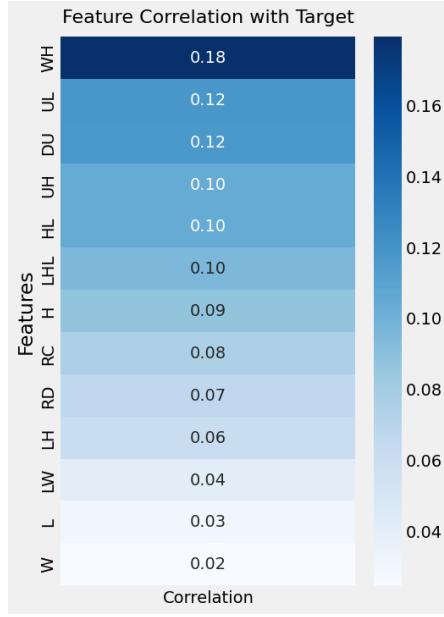


Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

1046 sex of *T. granosa*: the width-height ratio ($r = 0.18$), the umbos-length ratio ($r = 0.12$), and the distance between the umbos ($r = 0.12$). Each of these features
 1047 demonstrated a weak positive relationship with the target variable.
 1048

1049 4.1.2 Statistical Analysis

1050 As part of the exploratory data analysis, statistical testing confirmed that the
 1051 dataset did not follow a normal distribution (see Table 4.1). Consequently, the
 1052 Mann-Whitney U test was applied with a significance level of $\alpha = 0.05$ to com-
 1053 pare male and female samples. Out of thirteen features, five showed statistically
 1054 significant differences. These included: distance between umbos ($p = 0.025$),
 1055 length-width ratio ($p = 0.011$), umbos-length ratio ($p = 0.019$), width-height
 1056 ratio ($p = 0.003$), and umbos-height ratio ($p = 0.036$).

1057 It is important to note that statistical significance does not imply predictive
 1058 importance. Therefore, further analysis, such as feature importance evaluation,
 1059 was performed to identify the most informative predictors for classification.

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

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

1060 4.1.3 Feature Importance Analysis

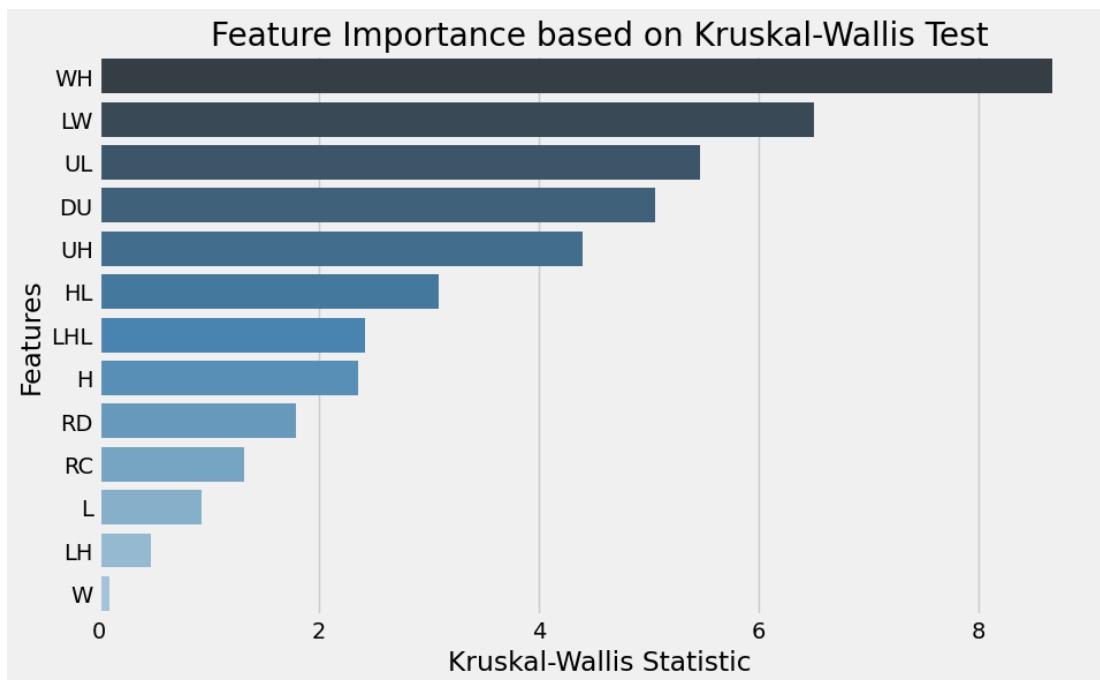


Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

1061 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric
 1062 method that is suitable for evaluating differences in distributions across groups

when the data does not follow a normal distribution. This approach was chosen because of the non-normality of the dataset and its robustness in handling continuous and ordinal data without assuming homogeneity of variances. (Ribeiro, 2024)

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

4.1.4 Performance Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	58.62	58.62	58.62	58.44
Logistic Regression	57.83	57.83	57.83	57.61
K-Nearest Neighbors	51.18	51.31	51.18	50.77
Extra Trees	59.07	59.54	59.07	58.45
Random Forest	59.85	59.99	59.85	59.80
Gradient Boosting	61.03	61.32	61.03	60.81
AdaBoost	60.63	60.98	60.63	60.39

Table 4.2: Performance Metrics for Models with All 13 Features

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

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	63.77	64.47	63.77	63.42
Logistic Regression	63.75	63.87	63.75	63.70
K-Nearest Neighbors	64.16	64.97	64.16	63.75
Extra Trees	61.04	61.68	61.04	60.67
Random Forest	61.01	61.12	61.01	60.91
Gradient Boosting	64.15	64.24	64.15	64.01
AdaBoost	61.02	61.26	61.02	60.82

Table 4.3: Performance Metrics for Models with 5 Features

1082 Table 4.3 presents the performance of the same models using only the top
1083 5 features identified through Kruskal-Wallis feature importance analysis. The
1084 selected features are the distance between the umbos, length-to-width ratio, width-
1085 to-height ratio, umbo distance-to-height ratio, and umbo distance-to-length ratio.

1086 Interestingly, the overall performance of the models improved when using only
1087 the top 5 features compared to using all 13. K-Nearest Neighbors (KNN) achieved
1088 the best results with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%,
1089 and an F1-score of 63.75%. Gradient Boosting followed closely behind. These find-
1090 ings suggest that reducing the feature set to the most relevant variables helped
1091 simplify the models, improved generalization, and enhanced predictive perfor-
1092 mance—particularly for KNN, which showed a notable improvement over its ear-
1093 lier results with the full feature set.

1094

4.1.5 Confusion Matrix Analysis

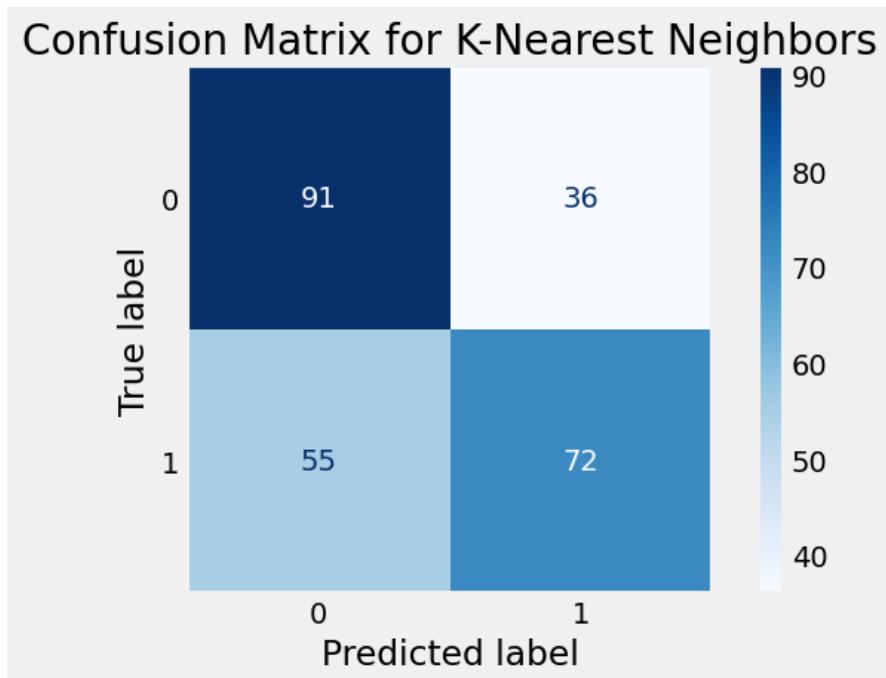


Figure 4.3: Feature Importance Scores Using the Kruskal-Wallis Test

1095 Figure 4.3 summarizes the performance of the K-Nearest Neighbors model in
1096 classifying *T. granosa* based on their sex, where 0 represents female samples and
1097 1 represents male samples. From the matrix, we observe that out of all the actual
1098 female samples (true label 0), 91 were correctly predicted as female (true positive
1099 for class 0), while 36 were incorrectly classified as male (false negative for class

1100 0). On the other hand, out of all the actual male samples (true label 1), 72 were
1101 correctly predicted as male (true positive for class 1), while 55 were incorrectly
1102 classified as female (false negative for class 1).

1103 4.2 Deep Learning Analysis

1104 This section presents the performance of the Convolutional Neural Network (CNN)
1105 model in classifying the sex of *T. granosa* based on shell morphology. The analysis
1106 evaluates the model's ability to distinguish between male and female shell images
1107 using various evaluation metrics. This part of the paper includes six subsections:
1108 baseline model, comparison of individual and combined angles, training result and
1109 hyperparameter tuning, proposed model, learning rates and training behavior per
1110 fold, and visualizations.

1111 The machine learning analysis (see Figure 4.3) revealed that five of the origi-
1112 nal features produced significant results. The K-Nearest Neighbor (KNN) model
1113 achieved an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1
1114 score of 63.75%. This section compares the model's performance across differ-
1115 ent angles based on the results of the machine learning and feature importance
1116 analysis.

1117 4.2.1 Baseline Model

1118 This section presents the baseline model with a batch size of 16 and 20 epochs,
1119 which will serve as the starting point for comparison and provide a guideline for
1120 hyperparameter tuning. The focus will be on one of the angles, specifically the
1121 Left Lateral view, since the feature importance analysis using the Kruskal-Wallis
1122 Test indicated that the width-to-height ratio had the highest importance score,
1123 which is most visible from the Left Lateral view.

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

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

1124 The unbalanced dataset, which consisted of 144 male samples and 127 female
1125 samples, achieved an accuracy of 65.27%, precision of 71.82%, recall of 58.99%,

1126 an F1-score of 63.99%, an AUC score of 73.08%, and a loss of 0.6122. However, to
 1127 address the class imbalance and enhance model performance, random undersam-
 1128 pling was performed. This approach resulted in improved performance metrics for
 1129 the balanced dataset, with an accuracy of 67.34%, precision of 69.43%, a recall
 1130 of 64.06%, an F1-score of 65.60%, an AUC score of 74.31%, and a lower loss of
 1131 0.5981.

1132 **4.2.2 Comparison of Individual and Combined Angles**

1133 Using the same batch size and number of epochs, performance was compared
 1134 across all individual angles and the combination of the two highest-performing
 1135 angles based on accuracy, using a balanced dataset. For the combined analysis,
 1136 samples from the two selected angles were placed side by side, and a new dataset
 1137 folder was created for male and female samples.

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

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

1138 Table 4.5 presents the performance metrics for each individual angle and the
 1139 combination of the two highest-performing angles in terms of accuracy. The
 1140 Left Lateral view achieved the highest accuracy (67.34%) and precision (69.43%),
 1141 while the Dorsal view obtained the highest recall (77.88%) and F1-score (69.96%).
 1142 Meanwhile, the Ventral view recorded the highest AUC score (74.87%), indicat-
 1143 ing its strong ability to distinguish between classes. Combining the Ventral and
 1144 Left Lateral views resulted in an overall accuracy of 62.60%, suggesting that while
 1145 combined images may provide complementary information, individual angle views
 1146 still outperformed the combined views under the current experimental setup.

1147 **4.2.3 Training Result and Hyperparameter Tuning**

1148 The Left Lateral angle was selected for further optimization. Several experiments
 1149 were conducted by tuning hyperparameters such as batch size, number of epochs,

and activation functions. Each adjustment was compared against the baseline model to enhance performance and develop a robust CNN for sex classification of *T. granosa*.

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

A. Batch Size and Number of Epochs

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

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

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

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

B. Activation Functions

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

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

Table 4.7 the performance of different activation functions applied to the CNN model trained with a batch size of 32 and 50 epochs. Based on the results, the

1170 ReLU activation function achieved the best overall performance, with an accu-
1171 racy of 71.68%, precision of 72.52%, recall of 69.29%, F1-score of 69.12%, and
1172 AUC score of 77.34%, along with the lowest loss at 0.5824. This suggests that
1173 ReLU remains an effective activation function for the classification of *T. granosa*,
1174 outperforming both ELU and PReLU in this setup.

1175 4.2.4 Proposed Model

1176 This section presents the performance evaluation of the proposed Convolutional
1177 Neural Network (CNN) model, trained with a batch size of 32, 50 epochs, and us-
1178 ing the ReLU activation function. The model's effectiveness was assessed through
1179 5-fold cross-validation to ensure robustness and generalizability across different
1180 data partitions.

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

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

1181 The proposed model consistently achieved high performance in Folds 1, 3, and
1182 5, with accuracies above 76% and strong recall and AUC scores, demonstrating
1183 its potential for reliable sex identification of *T. granosa*. The slight variation
1184 in performance across folds may be attributed to differences in data distribution,
1185 emphasizing the importance of further data augmentation and balancing for future
1186 work.

1187 4.2.5 Learning Rates and Training Behavior per Fold

1188 This section presents the learning rate adjustments, early stopping events, and
1189 best epoch selections for each fold during the 5-fold cross-validation of the pro-
1190 posed model. During training, the ReduceLROnPlateau callback was employed
1191 to monitor the validation loss and automatically reduce the learning rate when
1192 performance plateaued. Additionally, EarlyStopping was utilized to halt training
1193 once no further improvement was observed after a set patience, and the model
1194 weights were restored from the end of the best-performing epoch to ensure optimal
1195 performance.

1196 The following table summarizes the epochs where learning rate reductions
 1197 occurred, the adjusted learning rates, the epochs at which early stopping took
 1198 place, 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		
	12	0.0005000		
Fold 4	15	0.0002500	32	24
	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

1199 4.2.6 Visualizations

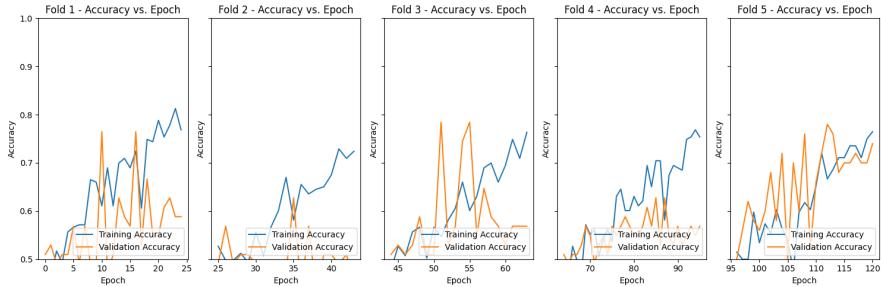


Figure 4.4: Training and Validation Accuracy per Fold

1200 Figure 4.4 shows the performance of the model in the training and validation
 1201 in terms of accuracy across five folds. The graph across folds displays a consistent
 1202 upward trend for the training accuracy. However, there is an observable change in
 1203 the performance, particularly in Folds 1 and 2, where it shows a slight downward
 1204 trend in the validation accuracy.

1205 Figure 4.5 shows the average performance of the model in both training and
 1206 accuracy in terms of accuracy across five folds. Similar to the individual perfor-
 1207 mances, there is an observable upward trend, which shows that the accuracy score
 1208 improves with the number of folds. The validation accuracy shows a downward
 1209 and upward trend that shows that it gradually improves on later epochs. The

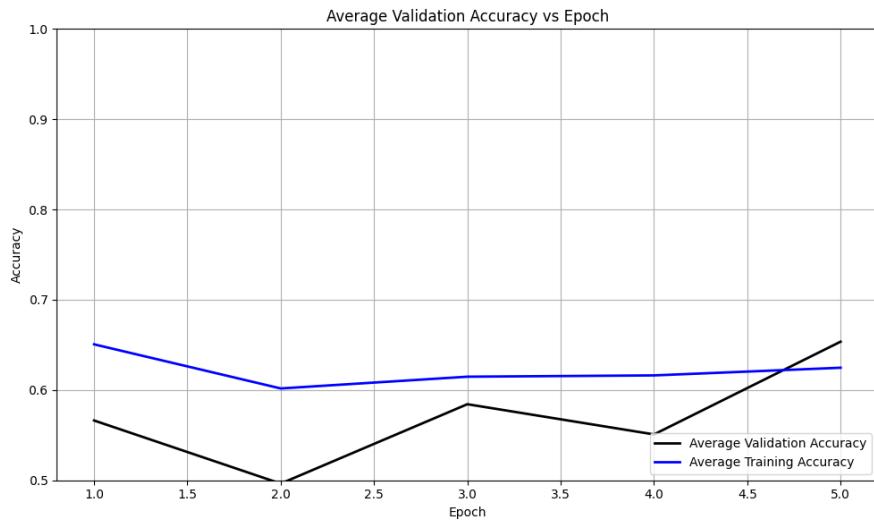


Figure 4.5: Average Training and Validation Accuracy Across Folds

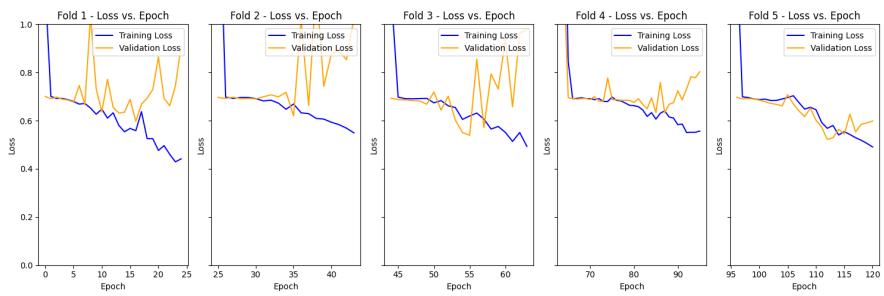


Figure 4.6: Training and Validation Loss per Fold

accuracy in the training is slightly higher than the accuracy when validating the model, it indicates that the model learns during training.

Figure 4.6 shows the performance of the model in the training and validation in terms of the training and validation loss across five folds. The graph across folds displays a consistent downward trend for the training loss. On the other hand, there is an observable change in the performance, especially in Folds 1,3, and 4, where it shows an upward trend in the validation loss. This is an implication for the learning performance of the model, as it may not be learning effectively.

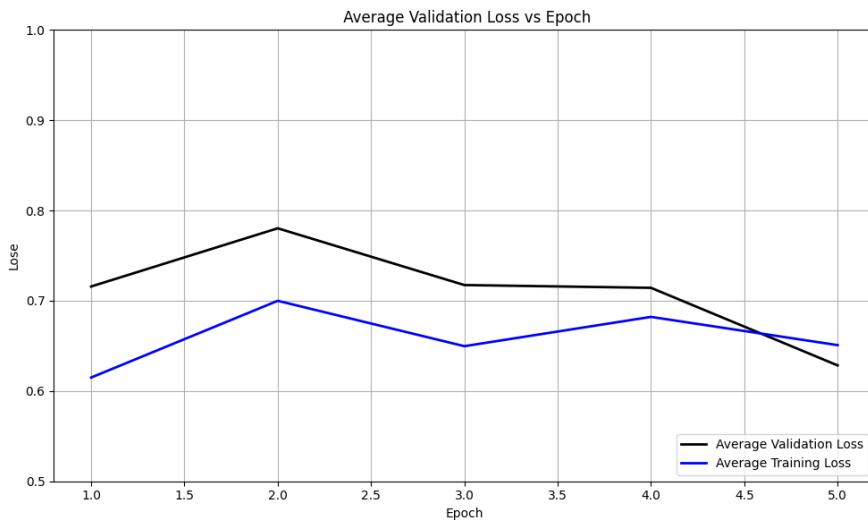


Figure 4.7: Average Training and Validation Loss Across Folds

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

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

Figure 4.9 shows the ROC Curve shows the ability of the proposed model to correctly identify the true positives, which can help determine the tradeoff between accuracy and sensitivity. It will also determine the validity of the model, that it is

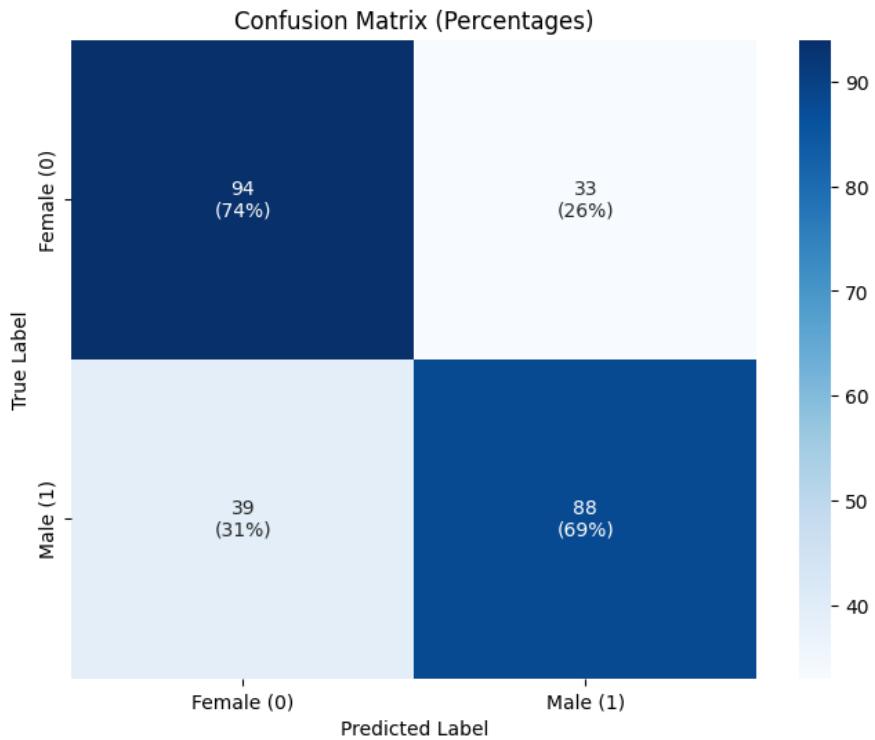


Figure 4.8: Confusion Matrix for Final Model Predictions

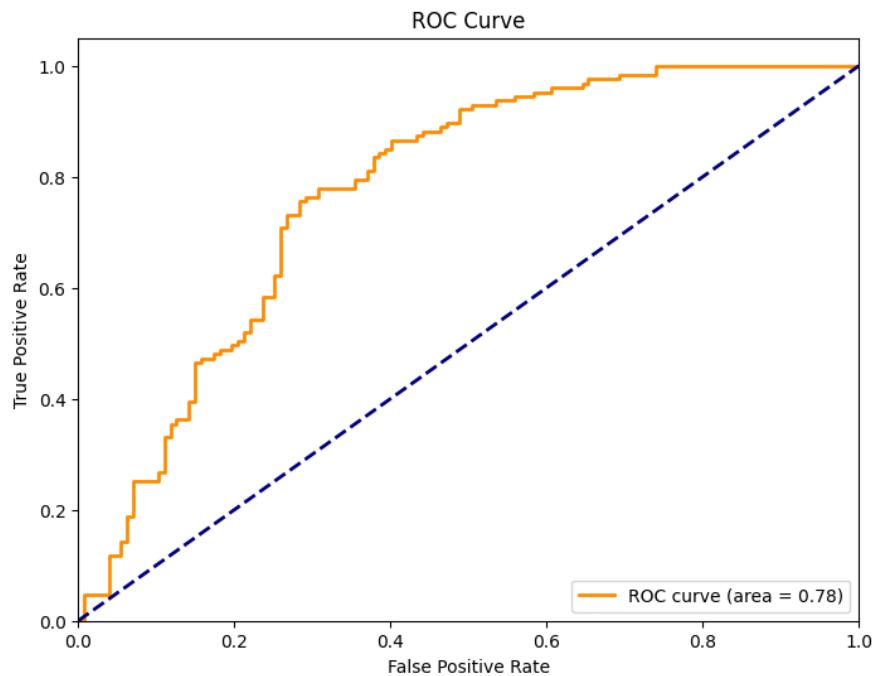


Figure 4.9: ROC Curve and AUC Score

1233 not predicting based only on random chances. The range of AUC ROC is between
1234 0.5 and 1. The model was able to achieve a score of 0.7734, which is better than
1235 random chances and an indication that the model is performing reasonably.

1236 4.3 Discussions

1237 This study aimed to develop a non-invasive method for identifying the sex of *T.*
1238 *granosa* using machine learning, deep learning, and computer vision technologies.
1239 The dataset was manually curated by the researchers, including both the linear
1240 measurements and the images captured from six different angles.

1241 The machine learning approach revealed that using five key features, selected
1242 through statistical tests (Mann-Whitney U-test and Kruskal-Wallis test), out-
1243 performed models trained on all 13 features. The K-nearest neighbors (KNN)
1244 classifier, using only these five features, achieved an accuracy of 64.16%, precision
1245 of 64.97%, recall of 64.16%, and an F1-score of 63.57%. These results indicate
1246 that a more focused set of features can enhance model performance, confirming
1247 the potential of non-invasive sex identification using linear measurements.

1248 Further deep learning experiments explored how different image angles im-
1249 pacted performance. The study found that the Left Lateral view consistently
1250 produced the best results, with an accuracy of 71.68%, precision of 72.52%, recall
1251 of 69.29%, F1-score of 69.12%, and an AUC score of 77.34%. This suggests that
1252 optimizing image angles is crucial, and combining multiple angles did not signif-
1253 icantly improve the model's performance. Data augmentation and regularization
1254 techniques, such as early stopping, helped improve the model's generalization and
1255 prevent overfitting.

1256 The findings are significant because they demonstrate the feasibility of a non-
1257 invasive, accurate, and efficient sex identification method for *T. granosa*. This
1258 approach aligns with sustainable aquaculture practices by reducing the need for
1259 harmful physical sex-identifying methods. By integrating machine learning with
1260 deep learning image analysis, this study provides a valuable model for non-invasive
1261 sex identification which could be applied to other species in aquaculture as well.

1262 When compared to similar existing studies such as the gender classification
1263 method for Chinese mitten crab using deep learning CNN (Cui *et al.*, 2020), there
1264 are notable differences in methodology. The crab study used grayscale images and
1265 a CNN with three convolutional layers, achieving 98.90% accuracy. In contrast,
1266 this study utilized a hybrid approach combining machine learning with deep learn-
1267 ing CNNs, trained on RGB images (256×256), and a deeper CNN architecture.

1268 Despite achieving lower accuracy (71.68%), this variation could be due to the sub-
1269 tler morphological differences between male and female *T. granosa*, or possibly
1270 due to image quality limitations and sample size.

1271 There are limitations in this study, particularly the size of the dataset (271
1272 samples) and the reliance on six fixed image angles. These constraints may not
1273 fully represent the morphological variability across different populations or en-
1274 vironments. Despite these limitations, the study successfully demonstrates that
1275 combining machine learning and deep learning with computer vision can provide
1276 a reliable and non-invasive solution for sex identification in *T. granosa*.

₁₂₇₇ **Chapter 5**

₁₂₇₈ **Conclusion and
Recommendations**

₁₂₈₀ **5.1 Conclusion**

₁₂₈₁ This study utilized the application of machine learning and deep learning tech-
₁₂₈₂ niques to identify the sex of *T. granosa* based on the morphometric characteristics.
₁₂₈₃ A manually curated dataset was developed, consisting of both linear measurements
₁₂₈₄ and images captured from six different angles. Machine learning methods were
₁₂₈₅ employed to identify statistically significant features, which served as the basis for
₁₂₈₆ deep learning analysis using a 12-layer Convolutional Neural Network (CNN). The
₁₂₈₇ proposed CNN model yielded an average accuracy of 71.68% in the performance
₁₂₈₈ metrics. Overall, this study offers a viable solution for non-invasive sex identifi-
₁₂₈₉ cation, providing an in-depth analysis based on *T. granosa*'s linear measurements
₁₂₉₀ and morphological characteristics from different angles.

₁₂₉₁ Through the availability of the gathered data, trial-and-error experimentation
₁₂₉₂ was conducted by adjusting the number of layers, batch size, epoch, and activa-
₁₂₉₃ tion functions. The different combinations tested provided baseline results that
₁₂₉₄ demonstrate the feasibility of non-invasive sex identification for *T. granosa*.

₁₂₉₅ While the study has made significant progress, challenges were encountered
₁₂₉₆ during CNN training, particularly due to hardware memory limitations. To over-
₁₂₉₇ come these, the researchers utilized synchronous Google Colab with 100 comput-
₁₂₉₈ ing units, requiring subscriptions, repeated retraining, and reconfigurations, which
₁₂₉₉ demanded considerable financial resources and time to optimize the parameters.

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

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

5.2 Recommendations

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

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

Future studies could also invest in a sturdier and more controlled environment by using a green background and positioning a fixed camera angle during image acquisition. In addition, researchers may experiment with other image processing techniques such as morphological transformations to emphasize features. The

¹³³⁵ dataset can be utilized for further analysis through advanced deep learning and
¹³³⁶ computer vision methods to make sense of the images gathered and discern sexual
¹³³⁷ dimorphism for *T. granosa* or to serve as a basis for conducting similar studies to
¹³³⁸ other bivalve species.

¹³³⁹ References

- 1340 Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten
1341 years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,
1342 5–16. doi: 10.1080/11250000409356545
- 1343 Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara
1344 granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,
1345 10.
- 1346 Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian
1347 Sains*, *14*, 14207.
- 1348 Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).
1349 Data scaling performance on various machine learning algorithms to identify
1350 abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:
1351 10.14710/jtsiskom.2021.14105
- 1352 Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.
1353 *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- 1354 Awan, A. A. (2022, November). *A complete guide to data augmentation*. Retrieved from <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>
- 1355 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquaculture.
1356 In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Surtida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on mangrove-friendly aquaculture organized by the seafdec aquaculture department, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- 1357 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,
1358 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic
1359 Resources.
- 1360 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of
1361 waste cockle shell (anadara granosa) in biodiesel production from palm olein:
1362 Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.
1363 doi: 10.1016/j.fuel.2011.03.002
- 1364 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining
1365 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127

- 1371 .v1
- 1372 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,
1373 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A
1374 new sex determination system in freshwater mussels (bivalvia: Unionoida)?
1375 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/
1376 msq345
- 1377 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-
1378 proaches, challenges and opportunities for aquaculture. *Journal of Marine
1379 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1380 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).
1381 Mass mortalities in bivalve populations: A review of the edible cockle
1382 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:
1383 10.1016/j.ecss.2014.04.011
- 1384 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic
1385 research in bivalves: Towards the identification of molecular markers of
1386 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:
1387 10.1016/j.jprot.2012.04.027
- 1388 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data
1389 augmentation for forecasting with deep learning. *ArXiv (Cornell Univer-
1390 sity)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1391 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly
1392 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1393 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan
1394 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:
1395 10.1093/icb/ict076
- 1396 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).
1397 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma
1398 edule*) geographical traceability based on shell image analysis. *Ecological
1399 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1400 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method
1401 for chinese mitten crab using deep convolutional neural network. *Multi-
1402 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1403 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle
1404 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-
1405 posure. *Aquaculture*, 56(2), 151–162. Retrieved from [https://doi.org/10.1016/0044-8486\(86\)90024-4](https://doi.org/10.1016/0044-8486(86)90024-4) doi: 10.1016/0044-8486(86)90024-4
- 1406 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect
1407 ageing of eel, *anguilla anguilla* (l.)? *International Review of Hydrobiology*,
1408 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1409 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and
1410 anatomy for young learners*. Rosie Research. Retrieved from <https://>

- 1413 rosiereresearch.com/clam-dissection-anatomy/
- 1414 Fao 2024 report: Sustainable aquatic food systems important for global food
1415 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1416 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and
1417 elemental composition: Complementary tools for stock discrimination of
1418 mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,
1419 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1420 GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*. Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1421 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)
- 1422 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Blackwell Science.
- 1423 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1424 Hussain, S. S., & Zaidi, S. S. H. (2024). Adaboost ensemble approach with
1425 weak classifiers for gear fault diagnosis and prognosis in dc motors. *Applied Sciences*, 14(7), 3105. doi: 10.3390/app14073105
- 1426 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and
1427 surface water characterization and heavy metal health risk on cockles in
1428 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.
1429 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1430 Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a comprehensive guide to data rescaling*. DataCamp. Retrieved from <https://www.datacamp.com/tutorial/normalization-in-machine-learning> (Accessed 2025-04-23)
- 1431 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest
1432 example of sexual dimorphism in bivalves—evidence from the astartid
1433 *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,
1434 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1435 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts
1436 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–
1437 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1438 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep
1439 learning-based phenotype classification of three ark shells: Anadara kagoshimensis,
1440 tegillarca granosa, and anadara broughtonii. *Frontiers in Marine Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1441 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the

- 1455 granulated ark, tegillarca granosa (linne). *The Korean Journal of Malacology*, 13, 55–64.
- 1456
- 1457 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as
1458 a tool for stock identification of the southern blue whiting, micromesistius
1459 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1460 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &
1461 et al. (2016). Identifying blue whiting (micromesistius poutassou) stock
1462 structure in the northeast atlantic by otolith shape analysis. *Canadian
1463 Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/
1464 cjfas-2015-0332
- 1465 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle
1466 tegillarca granosa (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.
1467 Arts Sci*, 4.
- 1468 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and
1469 larval rearing of the blood cockle, tegillarca granosa (linnaeus, 1758), in the
1470 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2
1471 .001
- 1472 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of
1473 local populations of the common sole solea solea (pisces, soleidae) in the nw
1474 mediterranean through otolith morphometrics and shape analysis. *Marine
1475 Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1476 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and
1477 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–
1478 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/
1479 kja.21209
- 1480 Narasimham, K. A. (1988). Taxonomy of the blood clams anadara (tegillarca)
1481 granosa (linnaeus, 1758) and a. (t.) rhombea (born, 1780).
- 1482 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,
1483 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world
1484 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1485 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith
1486 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.
1487 doi: 10.1002/jmor.10439
- 1488 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric
1489 morphometrics and machine learning challenge currently accepted species
1490 limits of the land snail placostylus (pulmonata: Bothriembryontidae) on the
1491 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.
1492 doi: 10.1093/mollus/eyz031
- 1493 Ribeiro, D. (2024, Aug.). *Diogo ribeiro. data science. kruskal-wallis test*.
1494 Retrieved from https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/ (Accessed: 2025-04-23)
- 1495 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,

- 1497 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-
1498 carbons from consumption of blood cockle and exposure to contaminated
1499 sediments and water along the klang strait, malaysia. *Marine Pollution*
1500 *Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1501 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence
1502 of seasonal variation and anthropogenic stress on blood cockle (*Tegillarca*
1503 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-*
1504 *pogenic Stress on Blood Cockle (*Tegillarca Granosa*) Production Potential*,
1505 44(2), 62–82.
- 1506 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-
1507 chine learning and its applications to biology. *PLoS Computational Biology*,
1508 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1509 Team, K. (n.d.). *Keras documentation: Reducelronplateau*. Retrieved from
1510 https://keras.io/api/callbacks/reduce_lr_on_plateau/
- 1511 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-
1512 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,
1513 & A. F. Eble (Eds.), *The eastern oyster *Crassostrea virginica** (pp. 335–370).
1514 College Park, MD: Maryland Sea Grant.
- 1515 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning
1516 approach for morphological feature extraction based on variational auto-
1517 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*
1518 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1519 Wong, T. M., & Lim, T. G. (2018). *Cockle (*Anadara granosa*) seed produced in*
1520 *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in_3366.pdf
- 1521 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.
1522 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972
1523 .5008949
- 1524 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*
1525 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1526 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-
1527 pacts of four commonly used nanoparticles on the metabolism of a ma-
1528 rine bivalve species, *Tegillarca granosa*. *Chemosphere*, 296, 134079. doi:
1529 10.1016/j.chemosphere.2022.134079
- 1530 Zhan, P. L., Zha, S., & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the
1531 blood clam *Tegillarca granosa*. *Marine Environmental Research*, 177. Re-
1532 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

¹⁵³³ **Appendix A**

¹⁵³⁴ **Data Gathering Documentation
and Supplementary Analysis**



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



Figure A.2: Separating Male and Female Samples After Spawning of *Tegillarca granosa*



Figure A.3: Sex Identified Female Through Dissecting of *Tegillarca granosa*



Figure A.4: Sex Identified Male Through Dissecting of *Tegillarca granosa*

Litob_Id	Length	Width	Height	Rib count	Length (Hinge Line)	Distance Umbos
10001	48.05	37.6	32.15	20	33.55	4.1
20001	48.05	37.6	32.15	20	33.55	4.1
30001	48.05	37.6	32.15	20	33.55	4.1
40001	48.05	37.6	32.15	20	33.55	4.1
50001	48.05	37.6	32.15	20	33.55	4.1
60001	48.05	37.6	32.15	20	33.55	4.1
10002	47.4	32.5	32.25	20	33.1	3.05
20002	47.4	32.5	32.25	20	33.1	3.05
30002	47.4	32.5	32.25	20	33.1	3.05
40002	47.4	32.5	32.25	20	33.1	3.05
50002	47.4	32.5	32.25	20	33.1	3.05
60002	47.4	32.5	32.25	20	33.1	3.05
10003	43.3	34.1	31.25	21	32.05	4.5
20003	43.3	34.1	31.25	21	32.05	4.5
30003	43.3	34.1	31.25	21	32.05	4.5
40003	43.3	34.1	31.25	21	32.05	4.5
50003	43.3	34.1	31.25	21	32.05	4.5
60003	43.3	34.1	31.25	21	32.05	4.5
10075	50.05	35.05	32.05	21	30.05	4.1
20075	50.05	35.05	32.05	21	30.05	4.1

Figure A.5: Linear Measurements of Female *Tegillarca granosa*

Litob_id	Length	Width	Height	Rib count	Length (Hinge Line)	Distance Umbos
110004	43.1	33.05	28.15	21	28.5	3.05
120004	43.1	33.05	28.15	21	28.5	3.05
130004	43.1	33.05	28.15	21	28.5	3.05
140004	43.1	33.05	28.15	21	28.5	3.05
150004	43.1	33.05	28.15	21	28.5	3.05
160004	43.1	33.05	28.15	21	28.5	3.05
110005	41.1	31.05	27.6	20	23.05	3.35
120005	41.1	31.05	27.6	20	23.05	3.35
130005	41.1	31.05	27.6	20	23.05	3.35
140005	41.1	31.05	27.6	20	23.05	3.35
150005	41.1	31.05	27.6	20	23.05	3.35
160005	41.1	31.05	27.6	20	23.05	3.35
110006	43.2	33.45	29.35	20	29.35	3.3
120006	43.2	33.45	29.35	20	29.35	3.3
130006	43.2	33.45	29.35	20	29.35	3.3
140006	43.2	33.45	29.35	20	29.35	3.3
150006	43.2	33.45	29.35	20	29.35	3.3
160006	43.2	33.45	29.35	20	29.35	3.3
110007	41.5	32.55	27.7	20	24.1	3.7
120007	41.5	32.55	27.7	20	24.1	3.7

Figure A.6: Linear Measurements of Male *Tegillarca granosa*

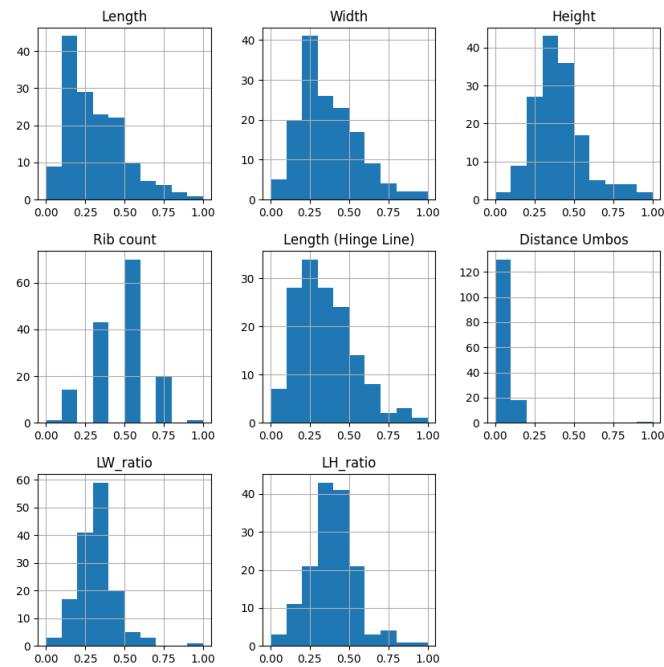


Figure A.7: Distribution of the Features of *Tegillarca granosa*