

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

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

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

<sup>13</sup> ADRICULA, Briana Jade  
<sup>14</sup> PAJARILLA, Gliezel Ann  
<sup>15</sup> VITO, Ma. Christina Kane

<sup>16</sup> Francis DIMZON  
<sup>17</sup> Adviser  
<sup>18</sup> Victor Marco Emmanuel FERRIOLS  
<sup>19</sup> Co-Adviser

<sup>20</sup> 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*

# <sup>39</sup> Contents

<sup>40</sup> <b>1 Introduction</b>	<b>1</b>
<sup>41</sup> 1.1 Overview . . . . .	1
<sup>42</sup> 1.2 Problem Statement . . . . .	2
<sup>43</sup> 1.3 Research Objectives . . . . .	3
<sup>44</sup> 1.3.1 General Objective . . . . .	3
<sup>45</sup> 1.3.2 Specific Objectives . . . . .	3
<sup>46</sup> 1.4 Scope and Limitations of the Research . . . . .	4
<sup>47</sup> 1.5 Significance of the Research . . . . .	5
<sup>48</sup> <b>2 Review of Related Literature</b>	<b>6</b>
<sup>49</sup> 2.1 Background on <i>Tegillarca granosa</i> and Their Importance . . . . .	7
<sup>50</sup> 2.2 Current Methods of Sex Identification in <i>Tegillarca granosa</i> . . . . .	9
<sup>51</sup> 2.3 Machine Learning and Deep Learning in Biological Studies . . . . .	11
<sup>52</sup> 2.3.1 Deep Learning for Phenotype Classification in Ark Shells .	12
<sup>53</sup> 2.3.2 Geometric Morphometrics and Machine Learning for Species <sup>54</sup> Delimitation . . . . .	12
<sup>55</sup> 2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning	12
<sup>56</sup> 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	<b>3</b>	<b>Research Methodology</b>	<b>21</b>
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	<b>4</b>	<b>Results and Discussions</b>	<b>36</b>
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	4.2 Deep Learning Analysis . . . . .	40
84	4.2.1 Baseline Model . . . . .	40
85	4.3 Comparison of Individual and Combined Angles . . . . .	41
86	4.4 Training Result and Hyperparameter Tuning . . . . .	42
87	4.4.1 Batch Size and Number of Epochs . . . . .	42
88	<b>5 Conclusion and Recommendations</b>	<b>43</b>
89	5.1 Conclusion . . . . .	43
90	5.2 Recommendations . . . . .	43
91	<b>References</b>	<b>44</b>
92	<b>A Data Gathering Documentation and Supplementary Analysis</b>	<b>49</b>

# <sup>93</sup> List of Figures

<sup>94</sup>	2.1 Diagram of <i>Tegillarca granosa</i> Anatomy . . . . .	7
<sup>95</sup>	3.1 Male and Female <i>Tegillarca granosa</i> shells . . . . .	22
<sup>96</sup>	3.2 Different Views of the <i>T. granosa</i> Shell Captured . . . . .	24
<sup>97</sup>	3.3 Linear Measurements of <i>Tegillarca granosa</i> shell. . . . .	25
<sup>98</sup>	3.4 Image Acquisition Setup for <i>T. granosa</i> Samples . . . . .	26
<sup>99</sup>	3.5 Data Preprocessing Pipeline . . . . .	27
<sup>100</sup>	3.6 Diagram of k-fold cross-validation with k = 5 . . . . .	29
<sup>101</sup>	3.7 Shadows removed from male samples at different angles . . . . .	30
<sup>102</sup>	3.8 Shadows removed from female samples at different angles . . . . .	30
<sup>103</sup>	3.9 Architecture of Convolutional Neural Network (CNN) . . . . .	31
<sup>104</sup>	3.10 Diagram of stratified k-fold cross-validation with k=5 . . . . .	32
<sup>105</sup>	3.11 Data Augmentation Techniques . . . . .	33
<sup>106</sup>	4.1 Correlation heatmap of morphometric features with the sex of <i>T.</i> <i>granosa</i> . . . . .	37
<sup>108</sup>	4.2 Feature Importance Scores Using the Kruskal-Wallis Test . . . . .	38
<sup>109</sup>	A.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i> . . . . .	49

110	A.2 Separating Male and Female Samples After Spawning of <i>Tegillarca</i>	49
111	<i>granosa</i> . . . . .	
112	A.3 Sex Identified Female Through Dissecting of <i>Tegillarca granosa</i> .	50
113	A.4 Sex Identified Male Through Dissecting of <i>Tegillarca granosa</i> . .	50
114	A.5 Linear Measurements of Female <i>Tegillarca granosa</i> . . . . .	50
115	A.6 Linear Measurements of Male <i>Tegillarca granosa</i> . . . . .	51
116	A.7 Distribution of the Features of <i>Tegillarca granosa</i> . . . . .	51

# <sup>117</sup> List of Tables

<sup>118</sup>	2.1 Comparison of the Methods Used in Bivalves Studies . . . . .	19
<sup>119</sup>	4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison	37
<sup>120</sup>	4.2 Performance Metrics for Models with All 13 Features . . . . .	39
<sup>121</sup>	4.3 Performance Metrics for Models with 5 Features . . . . .	39
<sup>122</sup>	4.4 Performance Metrics for Unbalanced vs. Balanced Datasets (Batch Size: 16, Epochs: 20) . . . . .	40
<sup>124</sup>	4.5 Performance Metrics for Individual and Combined Angles (Batch Size: 16, Epochs: 20) . . . . .	41
<sup>125</sup>	4.6 Effect of Batch Size and Epoch Values on CNN Model Performance	42

<sup>127</sup> **Chapter 1**

<sup>128</sup> **Introduction**

<sup>129</sup> **1.1 Overview**

<sup>130</sup> The Philippines is a global center of marine biodiversity and has established aquaculture as a significant contributor to total fishery production (Aypa & Baconguis, 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annually and is the 11th largest seafood producer in the world. Aquaculture is deeply integrated into Filipinos' livelihoods, encompassing fish cultivation and the production of various aquatic species, including bivalves. Among these, blood cockles (*Tegillarca granosa*) hold considerable economic and environmental significance, making it essential to ensure sustainable production and population balance.

<sup>138</sup> Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent overharvesting and ensure sustainability. An imbalanced ratio can lead to overexploitation and negatively impact the population's viability. However, there is limited literature on *T. granosa* that provides a thorough understanding of its sex-determining mechanisms, particularly regarding sexual dimorphism based on morphological and morphometric characteristics (Breton, Capt, Guerra, & Stewart, 2017).

<sup>145</sup> Currently, sex determination methods for blood cockles are invasive, including dissection and histological examinations, which often result in the death of the species. While there is growing literature on sex identification in aquaculture commodities using machine learning and deep learning, there is a notable scarcity of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

<sup>150</sup> This study, titled "Morphometric-Based Non-Invasive Sex Identification of

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

## 158 1.2 Problem Statement

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

167 Current methods for determining the sex of *T. granosa* are invasive and in-  
168 volve dissection, which requires cutting open the shell to visually inspect the  
169 gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-  
170 quently leads to their death. Another method is histological examination, where  
171 tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,  
172 2021). Both approaches are labor-intensive and time-consuming, and can pose  
173 risks to population management, particularly when maintaining a balanced sex  
174 ratio for breeding programs is essential. Moreover, these invasive methods require  
175 specialized technical skills for accurate execution. Resource-limited aquaculture  
176 operations face significant challenges in accessing the necessary laboratory equip-  
177 ment, such as microscopes and staining tools, complicating the process.

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

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

<sup>186</sup> involving machine and deep learning technologies could address these issues by  
<sup>187</sup> providing a fast, accurate, and effective solution without harming or stressing the  
<sup>188</sup> blood cockles.

## <sup>189</sup> 1.3 Research Objectives

### <sup>190</sup> 1.3.1 General Objective

<sup>191</sup> The general objective of this study is to develop a non-invasive method for iden-  
<sup>192</sup> tifying the sex of *Tegillarca granosa* using machine and deep learning integrated  
<sup>193</sup> with computer vision technologies. This method aims to provide accurate and  
<sup>194</sup> streamlined sex identification without causing harm to the specimens, thus sup-  
<sup>195</sup> porting sustainable aquaculture practices.

### <sup>196</sup> 1.3.2 Specific Objectives

<sup>197</sup> To achieve the overall general objective of developing a non-invasive sex identifi-  
<sup>198</sup> cation of *T. granosa* using machine learning, deep learning, and computer vision  
<sup>199</sup> technologies, the following specific objectives have been established:

- <sup>200</sup> 1. To collect and organize a comprehensive dataset of *T. granosa* which will  
<sup>201</sup> include high-quality images and relevant morphological measurements that  
<sup>202</sup> will serve as the basis for the machine-learning model.
- <sup>203</sup> 2. To develop and implement machine learning models that can classify the  
<sup>204</sup> sex of *T. granosa* based on the collected linear measurements and images of  
<sup>205</sup> different angles of the sample.
- <sup>206</sup> 3. To evaluate the performance of the models used using performance metrics  
<sup>207</sup> such as accuracy, precision, recall, and F1-score.
- <sup>208</sup> 4. To develop a system that can identify the sex of *T. granosa* based on its  
<sup>209</sup> morphological characteristics.

## **210 1.4 Scope and Limitations of the Research**

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

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

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

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

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

## <sup>245</sup> 1.5 Significance of the Research

<sup>246</sup> This study will give us a significant advancement in non-invasive sex identifica-  
<sup>247</sup> tion methods in *T. granosa* providing innovative solutions that could solve the  
<sup>248</sup> challenges in identifying sex and reshape sustainable approaches to aquaculture.  
<sup>249</sup> The significance of this study extends to the following:

<sup>250</sup>        *Research Institution.* The result of this study focusing on the sex-identification  
<sup>251</sup> mechanism of bivalves, specifically *Tegillarca granosa*, will provide valuable in-  
<sup>252</sup> sights into universities and research centers that focus on fisheries and coastal  
<sup>253</sup> management, such as the UPV Institute of Aquaculture, that aim to develop  
<sup>254</sup> sustainable development and suitable culture techniques.

<sup>255</sup>        *Fishermen.* By developing a non-invasive method in sex identification, this  
<sup>256</sup> study can help long-term harvest efficiency and maintain the ratio of the harvest  
<sup>257</sup> which can help prevent overexploitation of the *T. granosa*.

<sup>258</sup>        *Coastal Communities.* The result of this study would be beneficial for the  
<sup>259</sup> coastal communities that are reliant on their source of income with aquaculture  
<sup>260</sup> commodities like blood cockles. Maintaining the diversity and aspect ratio of  
<sup>261</sup> male and female may increase the market value of blood cockle production since  
<sup>262</sup> cockle aquaculture faces significant obstacles worldwide due to the fluctuating  
<sup>263</sup> seed supplies and scarcity of broodstock from the wild.

<sup>264</sup>        *Future Researchers.* The result of this study would serve as the basis for studies  
<sup>265</sup> that involve sex identification in bivalves such as *T. granosa*. Some technologies  
<sup>266</sup> are yet to be explored in machine learning, deep learning, and computer vision  
<sup>267</sup> technologies that can lead to higher accuracy and distinguish the presence of  
<sup>268</sup> sexual dimorphism in the *T. granosa*.

<sup>269</sup> **Chapter 2**

<sup>270</sup> **Review of Related Literature**

<sup>271</sup> Aquaculture is the fastest-growing industry in animal food production and has  
<sup>272</sup> great potential as a sustainable solution to global food security, nutrition, and  
<sup>273</sup> development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*  
<sup>274</sup> *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-  
<sup>275</sup> tegrated into the livelihoods of Filipinos, not only through fish cultivation but  
<sup>276</sup> also through the production of other aquatic species, including mollusks, oysters,  
<sup>277</sup> clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood  
<sup>278</sup> clams *Tegillarca granosa*, have economic and environmental significance. It has  
<sup>279</sup> been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-  
<sup>280</sup> vesting and maintain the optimal ratio to preserve the population and production  
<sup>281</sup> of the blood cockles.

<sup>282</sup> The members of the Arcidae Family, including *T. granosa* are important  
<sup>283</sup> sources of food and livelihood. Cockle aquaculture meets rising demands, however,  
<sup>284</sup> it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,  
<sup>285</sup> 2023). To solve the problem, researchers exert a considerable amount of effort,  
<sup>286</sup> developing a broader understanding of bivalves, including their sex-determining  
<sup>287</sup> mechanism, due to their notable importance in terms of diversity, environmental  
<sup>288</sup> benefits, and economic and market importance (Breton et al., 2017). Despite the  
<sup>289</sup> promising idea of identifying sex, there is limited research reported in terms of  
<sup>290</sup> sexual dimorphism, making it harder to distinguish through its morphological and  
<sup>291</sup> morphometric characteristics.

<sup>292</sup> By addressing the challenges in the sex identification of *T. granosa*, it would be  
<sup>293</sup> able to address one problem at a time. Currently, there are no recent documented  
<sup>294</sup> publications that integrate machine learning and computer vision in characterizing  
<sup>295</sup> sexual dimorphism, reducing complexity, variability in sex determination, and

<sup>296</sup> differentiation mechanisms in bivalves, including *T. granosa* specifically.

## <sup>297</sup> **2.1 Background on *Tegillarca granosa* and Their 298 Importance**

<sup>299</sup> *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.  
<sup>300</sup> In the Philippines, it is commonly known as a Litob, a marine bivalve species from  
<sup>301</sup> the family Arcidae. Litob is widely distributed in the world including Southeast  
<sup>302</sup> Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest  
<sup>303</sup> (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat  
<sup>304</sup> as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the  
<sup>305</sup> blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood  
<sup>306</sup> phenotype where it serves two purposes the hemocyte carries oxygen around the  
<sup>307</sup> body and strengthens immune defenses. In addition, it possesses a unique ability  
<sup>308</sup> to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

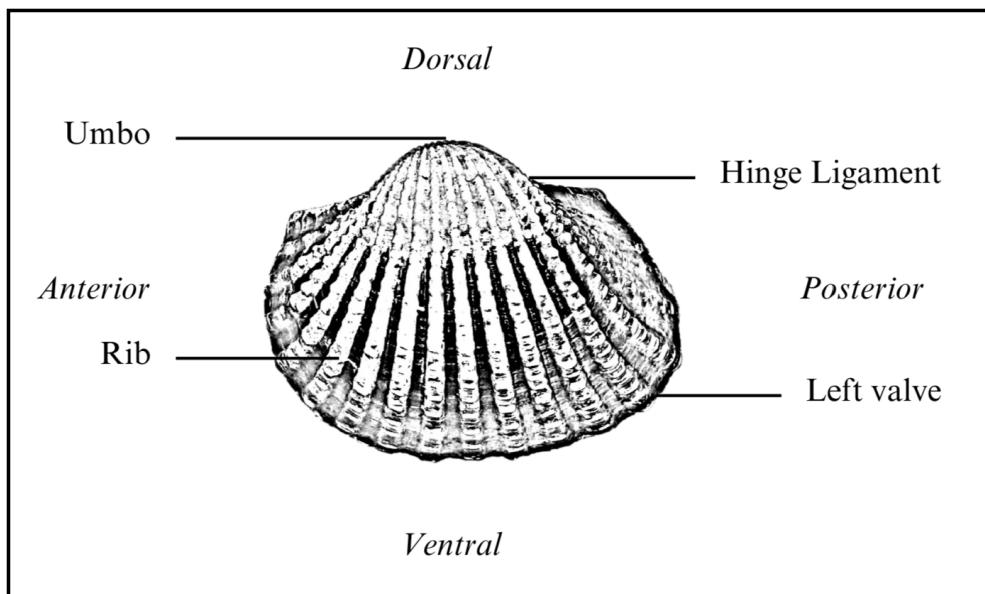


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

<sup>309</sup> *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both  
<sup>310</sup> valves being equal in size but asymmetrical from the hinge. The top edge of  
<sup>311</sup> the dorsal margin is straight, while the front is rounded and slopes downward,  
<sup>312</sup> with its back being obliquely rounded with a concave bottom edge. It has a  
<sup>313</sup> narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,  
<sup>314</sup> although some may be incomplete. The shell's outer layer, or the periostracum, is

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

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

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

<sup>355</sup> males at the time of harvest (Budd, Banh, Domingos, & Jerry, 2015).

<sup>356</sup> Clearly, much more work is required to understand the mechanisms under-  
<sup>357</sup> lying sexual dimorphism in bivalves, specifically *T. granosa*. Just like the other  
<sup>358</sup> aquaculture commodities, sex affects not just reproduction but it can affect mar-  
<sup>359</sup> ket preference and underlying economic value, making the determination of sex  
<sup>360</sup> important for meeting consumer demands. These are the increasing significance  
<sup>361</sup> of the *T. granosa* despite the lack of reviewed articles in the Philippines.

## <sup>362</sup> **2.2 Current Methods of Sex Identification in *Tegillarca granosa***

<sup>363</sup>

<sup>364</sup> The current sex identification methods in *Tegillarca granosa* range from invasive  
<sup>365</sup> histological techniques to less invasive methodologies like temperature-induced  
<sup>366</sup> spawning. Each approach comes with its pros and cons regarding accuracy, feasi-  
<sup>367</sup> bility, and impact on natural populations.

<sup>368</sup> Induced spawning and larval rearing are considered the less invasive techniques  
<sup>369</sup> used to study *Tegillarca granosa*. In the Philippines, limited research has been  
<sup>370</sup> done on the *Tegillarca granosa* (Linnaeus, 1758), and this study, titled Initial At-  
<sup>371</sup> tempts on Spawning and Larval Rearing of the Blood Cockle, *Tegillarca granosa*  
<sup>372</sup> in the Philippines, is conducted by Denise Vergara Miranda and Victor Marco  
<sup>373</sup> Emmanuel Nuestro Ferriols (2023). The researchers conducted experiments on  
<sup>374</sup> induced spawning and larval rearing, discovering that the eggs of female *T. gra-*  
<sup>375</sup> *nosa* were salmon pink, while the sperm released by males looked milky. After  
<sup>376</sup> spawning, the researchers successfully generated 6, 531, 000 fertilized eggs.

<sup>377</sup> They highlighted the importance of *T. granosa* and other anadarinids as a  
<sup>378</sup> food source that was established worldwide, especially in Malaysia and Korea.  
<sup>379</sup> However, in the Philippines, the bivalve aquaculture of the clam species is still  
<sup>380</sup> limited. The experiment which focuses on the culture and rearing of *T. granosa*  
<sup>381</sup> was attempted by subjecting the wild broodstocks to a series of temperature fluc-  
<sup>382</sup> tuations to induce the spawning of gametes. This is currently the most natural  
<sup>383</sup> and least invasive method for bivalves (Aji, 2011). The study of Miranda and  
<sup>384</sup> Ferriols aimed to pave the way to the sustainable production of *T. granosa* seeds  
<sup>385</sup> for aquaculture production and stock enhancement despite the scarcity of docu-  
<sup>386</sup> mented hatchery culture of *T. granosa* from larvae to adults that is available in  
<sup>387</sup> the Philippines.

<sup>388</sup> In the study entitled "The earliest example of sexual dimorphism in bivalves —

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

403 On the other hand, invasive techniques such as histological analysis offer a  
404 more thorough but harmful method for determining the sex of *T. granosa*. A  
405 study on the Spawning Period of Blood Cockle *Tegillarca granosa* (Linnaeus,  
406 1758) in Myeik Coastal. 240 blood cockle samples were examined for sex and  
407 gonad maturity stages using histological examination, with shell lengths ranging  
408 from 26-35mm and shell weights from 8.1-33g. For histological analysis, the whole  
409 soft tissues were removed from the shell and the flesh containing most parts of  
410 the gonads was fixed in formalin, dehydrated in an upgraded series of ethanol,  
411 and cleared in xylene. This invasive method allows for precise identification of  
412 the gonadal maturation stages based on the cellular and structural changes in the  
413 gonads.

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

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

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

## 447 **2.3 Machine Learning and Deep Learning in Bi- 448 ological Studies**

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

459 **2.3.1 Deep Learning for Phenotype Classification in Ark  
460 Shells**

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

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

474 **2.3.2 Geometric Morphometrics and Machine Learning for  
475 Species Delimitation**

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

486 **2.3.3 Contour Analysis in Mollusc Shells Using Machine  
487 Learning**

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

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

### 502 2.3.4 Machine Learning for Shape Analysis of Marine Or- 503 ganisms

504 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*  
505 to *Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*  
506 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to  
507 the traditional method of collecting linear measurements with the application of  
508 multivariate statistical methods and outline analysis in recording the structures  
509 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and  
510 elasmobranchs) with their hard bodies have been used as an indication of age and  
511 a determinable time-scale and structure continue to go through life (Arkhipkin,  
512 2005; Kerr & Campana, 2014). This study has explored variations in the mor-  
513 phometry of recording structures in stock discrimination and systematics. The  
514 researchers utilized the principal component analysis rather than the traditional  
515 approach, which helps simplify the data without losing important information.  
516 They utilized landmark-based geometric morphometrics, which has three differ-  
517 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other  
518 morphogenetic processes, and lastly, the extremal points are constructed land-  
519 marks.

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

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

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

### 550 2.3.5 Deep Learning for Landmark-Free Morphological Feature Extraction

551

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

### 562    2.3.6   Machine Learning for Sex Differentiation in Abalone

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

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

585   The study found that Naive Bayes consistently performed better than other al-  
586   gorithms. However, when applied to both min-max and zero-mean normalization,  
587   the average accuracies of the algorithms were as follows: Random Forest (62.37%),  
588   SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-  
589   nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with  
590   normalization, Random Forest achieved the highest overall metrics, including an  
591   average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of  
592   83.31%. Liu et al. concluded that Random Forest is highly accurate because it  
593   can handle large, complex datasets, run processes in parallel using multiple trees,  
594   and select the most relevant features to enhance model performance (Arifin et al.,  
595   2021).

596 **2.3.7 Machine Learning for Geographical Traceability in**  
597 **Bivalves**

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

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

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

624 **2.4 Limitations on Sex Identification in *Tegillarca***  
625 ***granosa***

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

630 perature shock, is the most natural and least invasive method for bivalves (Aji,  
631 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water  
632 from hot to cold with a specific temperature requires deliberate and careful ma-  
633 nipulation of the temperature over a specific period and would require constant  
634 management and monitoring.

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

643 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous  
644 (synchronous or functional) characteristics. Sequential hermaphrodites switch  
645 genders after being male or female for one or multiple yearly cycles. (Heller,  
646 1993; Gosling, 2004; Collin, 2013). Sex change and consecutive hermaphroditism  
647 have been observed in different bivalve species, including Ostreidae, Pectinidae,  
648 Veneridae, and Patellidae. However, macroscopically differentiating bivalve sex is  
649 challenging. The only way it may be identified is through histological analysis of  
650 gonad remains but to do so there is an act of killing the organism (Coe, 1943;  
651 Gosling, 2004). Verification of sex change in bivalves to classify whether male or  
652 female while they are alive is challenging since they need to be re-confirmed and  
653 re-evaluated to be the same individual after a year.

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

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

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

## 673 **2.5 Synthesis of the Study**

674 This section of the paper summarizes the technologies used in the different studies  
675 related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex  
676 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

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

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

# <sup>691</sup> Chapter 3

## <sup>692</sup> Research Methodology

<sup>693</sup> This chapter discussed the materials and methods employed in the study, focusing  
<sup>694</sup> on the development requirements, as well as the software and programming  
<sup>695</sup> languages utilized. It also detailed the overall workflow in conducting the study,  
<sup>696</sup> Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca*  
<sup>697</sup> *granosa* (Linnaeus), 1758) using machine learning and deep learning technologies.

<sup>698</sup> Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,  
<sup>699</sup> oversaw the overall workflow and conduct of the experiment. The researchers were  
<sup>700</sup> also guided by research associates LC Mae Gasit and Allena Esther Artera. Con-  
<sup>701</sup> sequently, the entire dataset collection process was conducted at the University of  
<sup>702</sup> the Philippines Visayas hatchery facility.

<sup>703</sup> The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical  
<sup>704</sup> Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-  
<sup>705</sup> tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-  
<sup>706</sup> ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine  
<sup>707</sup> Learning, (8) Morphological Characteristics Evaluation Using Deep Learning, and  
<sup>708</sup> (9) Evaluation Metrics

### <sup>709</sup> 3.1 Sample Collection

<sup>710</sup> The collection of *T. granosa* samples used in this study was part of an ongoing  
<sup>711</sup> research project by UPV DOST-PCAARRD titled "Establishment of the Center  
<sup>712</sup> for Mollusc Research and Development: Development of Spawning and Hatchery  
<sup>713</sup> Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."

<sup>714</sup> A total of 271 samples were provided for this study to classify the sex of *T. granosa*.  
<sup>715</sup> The samples, ranging in size from 34 to 61 mm, were sourced from the coastal area  
<sup>716</sup> of Zaraga, Iloilo, and fish markets in Ivisan, Capiz, Philippines (see Figure 3.1).

<sup>717</sup> The research and experimentation were conducted at the University of the  
<sup>718</sup> Philippines Visayas hatchery facility in Miagao, Iloilo, where the samples were  
<sup>719</sup> maintained in 200 L fiberglass-reinforced plastic (FRP) tanks containing filtered  
<sup>720</sup> seawater with 35 ppt salinity (Miranda & Ferriols, 2023).

<sup>721</sup> As part of the data collection process, the researchers utilized induced spawning  
<sup>722</sup> and dissection to classify the sex of the samples. Induced spawning through  
<sup>723</sup> temperature fluctuations was the most natural and least invasive method for bi-  
<sup>724</sup> valves compared to other approaches (Aji, 2011). However, since not all samples  
<sup>725</sup> exhibited gamete release, the researchers also performed dissections, assisted by  
<sup>726</sup> hatchery staff, to expedite data collection. The sex of the dissected samples was  
<sup>727</sup> identified based on the coloration of gonad tissue, which varies according to sex  
<sup>728</sup> and maturity stage. Females exhibited orange-red to pale orange gonads, while  
<sup>729</sup> males displayed white to grayish-white gonads (May et al., 2021).

<sup>730</sup> The methods used for data collection were considered noninvasive, particularly  
<sup>731</sup> given that *T. granosa* are oxygen regulators well adapted to tidal exposure and  
<sup>732</sup> hypoxia (Davenport & Wong, 1986).

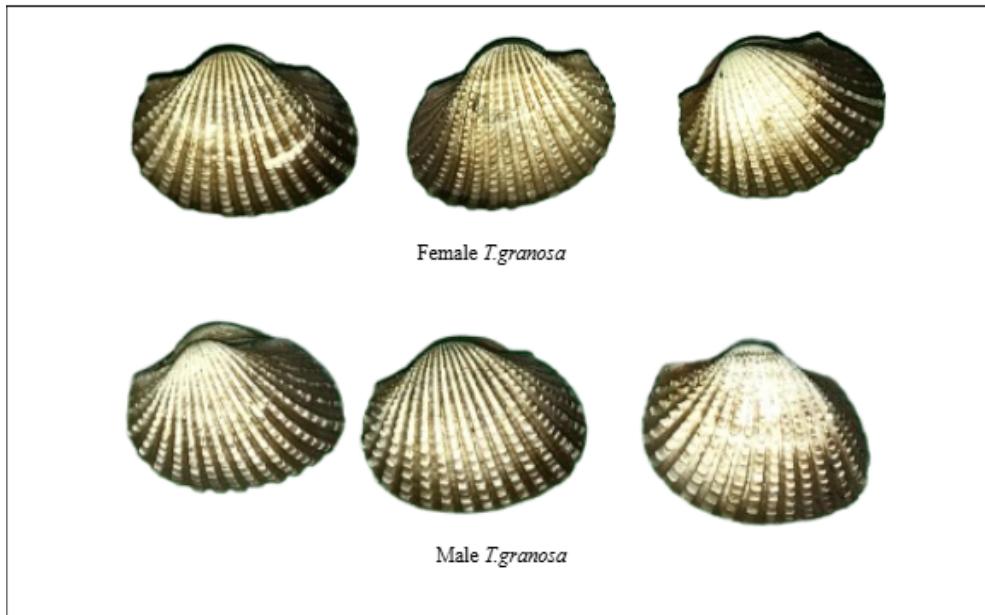


Figure 3.1: Male and Female *Tegillarca granosa* shells

## **733 3.2 Ethical Considerations**

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

## **740 3.3 Creating *T. granosa* Dataset**

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

**749** Images of each sample were captured and saved in JPG format using a stan-  
**750** dardized file naming convention that included the sample's sex, the shell's ori-  
**751** entation or view, and its corresponding number out of the 271 total samples. File  
**752** names for female *T. granosa* samples began with "0", while those for male sam-  
**753** ples began with "1". Each file name also included one of the six captured views:  
**754** (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right  
**755** lateral (refer to Figure 3.2), followed by a unique sample number. For exam-  
**756** ple, "010001" denoted the first female sample taken from the dorsal view, while  
**757** "110001" represented the first male sample from the same view. This naming  
**758** convention was implemented to prevent data leakage and ensure accurate labeling  
**759** of images according to their respective samples.

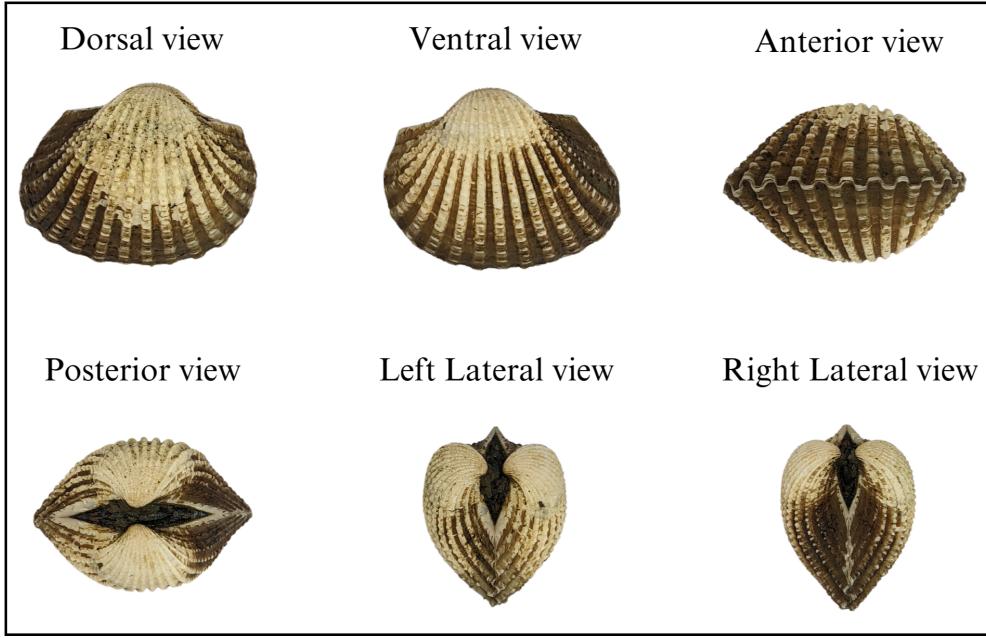


Figure 3.2: Different Views of the *T. granosa* Shell Captured

### <sup>760</sup> 3.4 Morphological and Morphometric Characteristics Collection

<sup>761</sup> Morphology refers to biological form and is one of the most visually recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).  
<sup>762</sup> In this study, morphological characteristics describe the structural features of *T. granosa*, focusing on measurable attributes such as shape, size, and color.  
<sup>763</sup> Morphometric characteristics, on the other hand, refer to specific quantifiable features of *T. granosa*, including length, width, height, hinge line length, distance between the umbos, and rib count. As stated by the researchers, quantifying and characterizing these traits is essential for understanding and visualizing variations in *T. granosa* morphology.

<sup>771</sup> The researchers measured the height, width, and length of *T. granosa* using  
<sup>772</sup> a Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the  
<sup>773</sup> corresponding measurement diagram. Length (A) refers to the distance from the  
<sup>774</sup> anterior to the posterior of the shell. Width (B) is defined as the widest span  
<sup>775</sup> across the shell from the left to the right valve. Height (C) measures the distance  
<sup>776</sup> from the base to the apex of the shell. In addition, the hinge line length (D) near  
<sup>777</sup> the hinge and the distance between the umbos (E) were recorded.

<sup>778</sup> Reament and Kennedy (1998) emphasized that including rib count as supple-

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

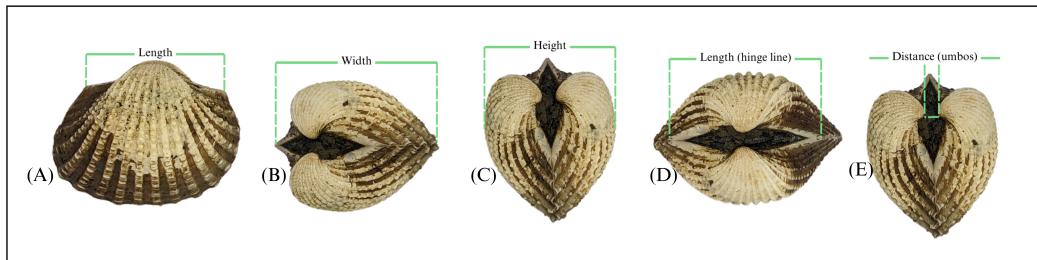


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

### 3.5 Image Acquisition and Data Gathering

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

The images were captured using a Google Pixel 3 XL smartphone, which features a resolution of  $2960 \times 1440$  pixels and a 12.2 MP camera ( $4032 \times 3024$  pixels). Additional camera specifications include an f/1.8 aperture, 28mm wide lens,  $\frac{1}{2.55}$ " sensor size, 1.4 $\mu\text{m}$  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

## 797 3.6 Hardware and Software Configuration

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

## 809 3.7 Morphometric Characteristics Evaluation Us- 810 ing Machine Learning

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

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

820 **3.7.1 Data Preprocessing**

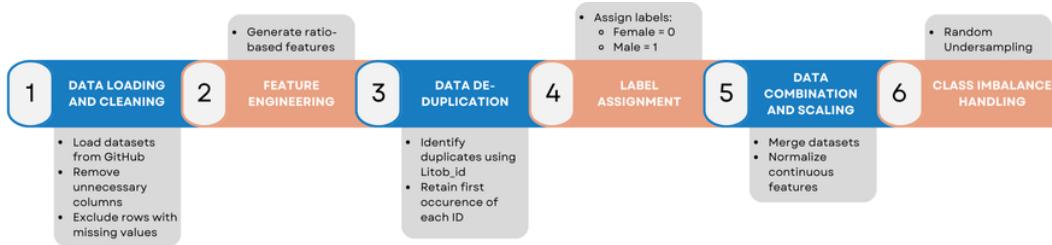


Figure 3.5: Data Preprocessing Pipeline

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

824 ***Data Loading and Cleaning***

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

829 ***Feature Engineering***

830 Additional ratio-based features were generated to augment the existing mea-  
831 surements. These included the length-to-width ratio, length-to-height ratio, width-  
832 to-height ratio, hinge line length-to-length ratio, umbos distance-to-length ratio,  
833 umbos distance-to-height ratio, and rib density. These derived features aimed to  
834 emphasize shape characteristics independent of size, improving the models' ability  
835 to distinguish morphological differences between sexes.

836 ***Data De-duplication***

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

841        ***Label Assignment***

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

845        ***Data Combination and Scaling***

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

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

855        ***Class Imbalance Handling***

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

861        **3.7.2 Machine Learning Models Training**

862        ***Model Selection and Hyperparameter Tuning***

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

868        ***Cross-Validation***

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

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

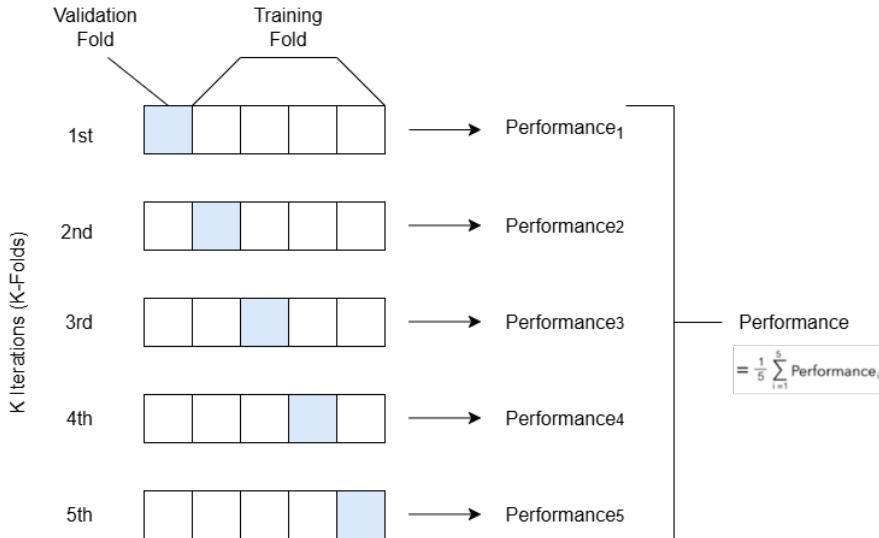


Figure 3.6: Diagram of k-fold cross-validation with  $k = 5$

## 874 3.8 Morphological Characteristics Evaluation Us- 875 ing Deep Learning

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

### 880 *Image Preprocessing*

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

### 887 *Adjusting Dimensions*

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

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

893 ***Background Removal***

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

899 ***Shadow Removal***

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

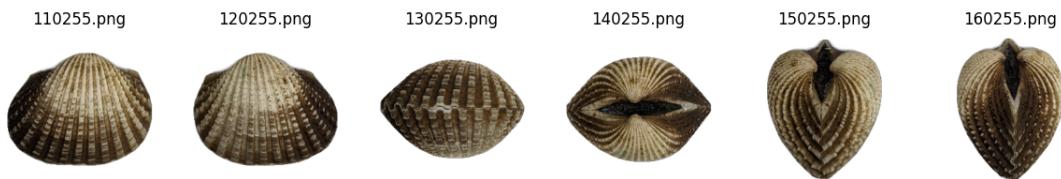


Figure 3.7: Shadows removed from male samples at different angles

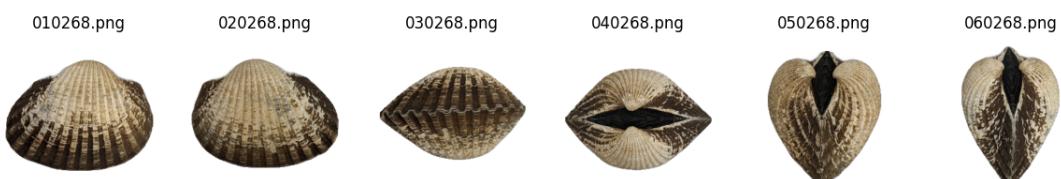


Figure 3.8: Shadows removed from female samples at different angles

904 **3.8.1 Convolutional Neural Network**

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

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

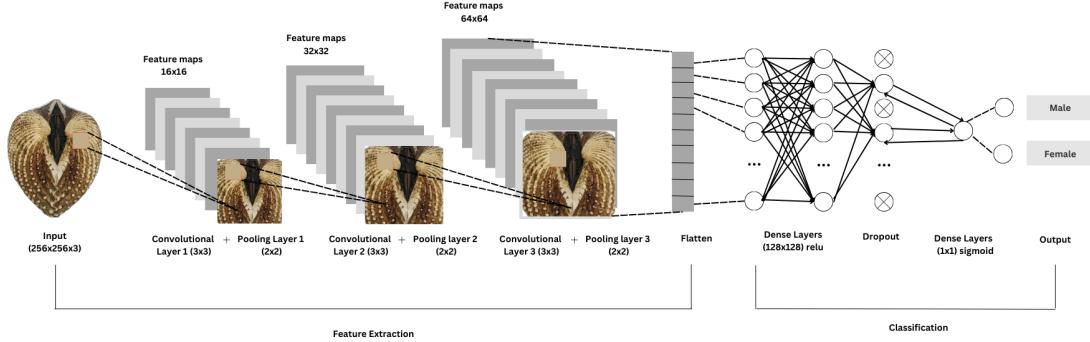


Figure 3.9: Architecture of Convolutional Neural Network (CNN)

### 913 ***Convolution Layer***

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

### 921 ***Pooling Layer***

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

### 925 ***Fully Connected and Dropout***

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

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

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

### 937 3.8.2 CNN Training

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

#### 943 *Data Splitting*

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

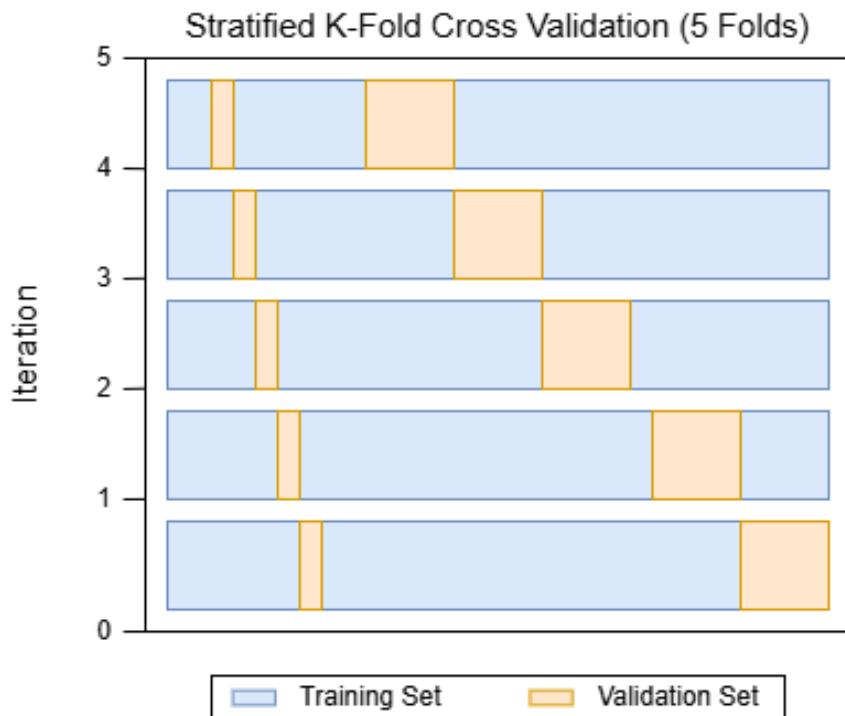


Figure 3.10: Diagram of stratified k-fold cross-validation with k=5

950        ***Data Augmentation***

951        Before model training, online data augmentation was applied exclusively to  
952        the training data within each fold, creating new data variations on the fly. The  
953        augmentations included random horizontal flips, slight rotations, and zoom trans-  
954        formations to enhance data diversity and improve model generalization (Awan,  
955        2022). All augmentation was strictly applied only to the training subset of each  
956        fold to prevent data leakage and maintain the validity of the results. On-the-fly  
957        data augmentation (OnDAT) generates augmented data during each iteration,  
958        exposing the model to constantly changing data variations. Augmenting the orig-  
959        inal data allows better exploration of the underlying data generation process and  
960        has the potential to prevent the model from overfitting spurious patterns, thereby  
961        improving performance (Cerqueira, Santos, Baghoussi, & Soares, 2024).

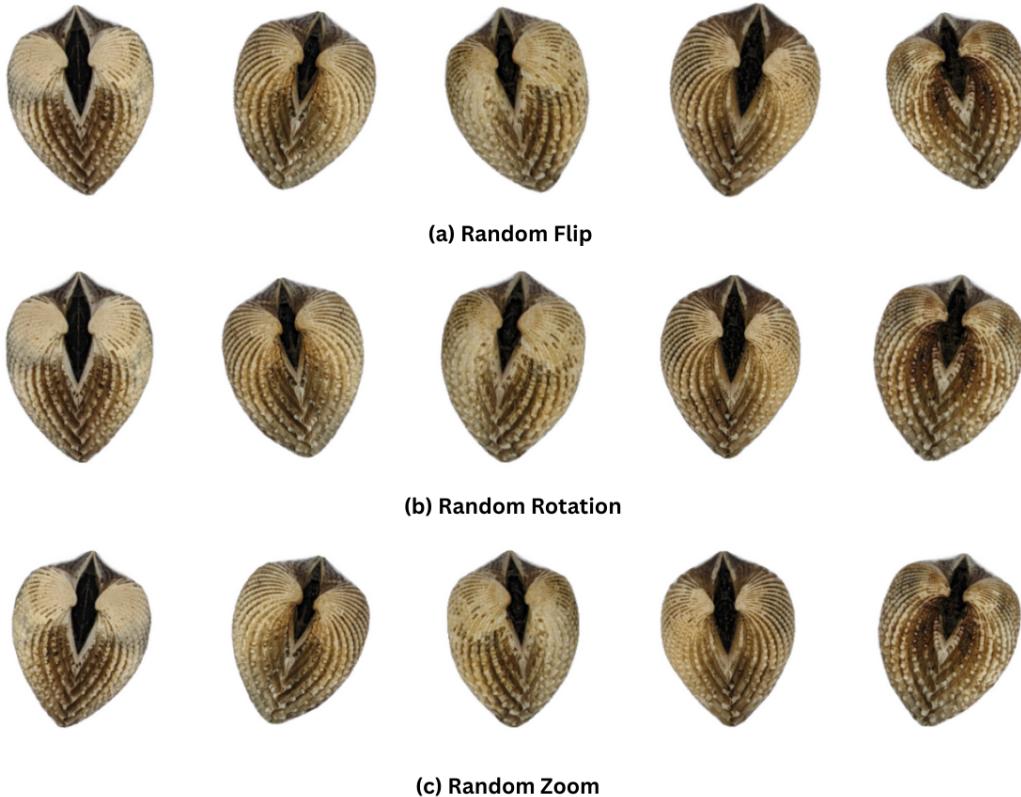


Figure 3.11: Data Augmentation Techniques

962        ***Training Procedure***

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

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

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

### 978 3.9 Evaluation Metrics

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

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

$$984 \text{ACC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

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

$$993 \text{PREC} = \frac{\text{True positive samples}}{\text{Samples assigned to positive class}} = \frac{TP}{TP + FP} \quad (3.2)$$

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

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

998 The F1 score is the harmonic mean of precision and recall, which penalizes  
999 extreme values of either of the two metrics (Cui et al., 2020). It is particularly  
1000 useful when the class distribution is imbalanced. The formula for the F1 score is:

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

1001 The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is  
1002 a performance measurement for classification problems, particularly used in deep  
1003 learning in this study. The ROC curve is a plot of the true positive rate (recall)  
1004 against the false positive rate (1 - specificity), and the AUC score quantifies the  
1005 overall ability of the model to discriminate between positive and negative classes.  
1006 A higher AUC indicates better model performance. (Nahm, 2022)

1007 **Chapter 4**

1008 **Results and Discussions**

1009 This chapter presents the results of the machine learning and deep learning anal-  
1010 yses conducted on the preprocessed dataset. Preprocessing was performed using  
1011 Python in Google Colaboratory. The chapter includes the evaluation of various  
1012 machine learning classifiers, analysis of feature importance, and the application  
1013 of deep learning models for image-based classification. These approaches aim to  
1014 identify key predictors and assess classification performance for sex identification  
1015 in *T. granosa*.

1016 **4.1 Machine Learning Analysis**

1017 **4.1.1 Data Exploration**

1018 Exploratory data analysis was performed to characterize the dataset using visu-  
1019 alizations to understand the patterns and correlations within the data. A corre-  
1020 lation heatmap was created to assess the relationship between the predictors and  
1021 the target variable.

1022 The heatmap (see Figure 4.1) revealed three features most correlated with the  
1023 sex of *T. granosa*: the width-height ratio ( $r = 0.18$ ), the umbos-length ratio ( $r$   
1024 = 0.12), and the distance between the umbos ( $r = 0.12$ ). Each of these features  
1025 demonstrated a weak positive relationship with the target variable.

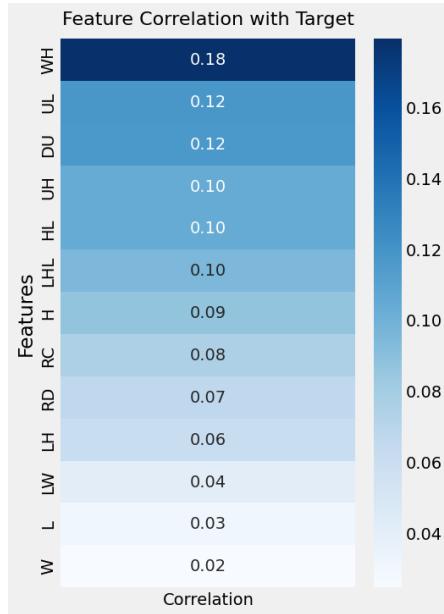


Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

#### 1026 4.1.2 Statistical Analysis

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

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

1027 As part of the exploratory data analysis, statistical testing confirmed that the  
 1028 dataset did not follow a normal distribution. Consequently, the Mann-Whitney  
 1029 U test was applied with a significance level of  $\alpha = 0.05$  to compare male and

1030 female samples. Out of thirteen features, five showed statistically significant dif-  
1031 ferences. These included: distance between umbos ( $p = 0.025$ ), length-width ratio  
1032 ( $p = 0.011$ ), umbos-length ratio ( $p = 0.019$ ), width-height ratio ( $p = 0.003$ ), and  
1033 umbos-height ratio ( $p = 0.036$ ).

1034 It is important to note that statistical significance does not imply predictive  
1035 importance. Therefore, further analysis, such as feature importance evaluation,  
1036 was performed to identify the most informative predictors for classification.

#### 1037 4.1.3 Feature Importance Analysis

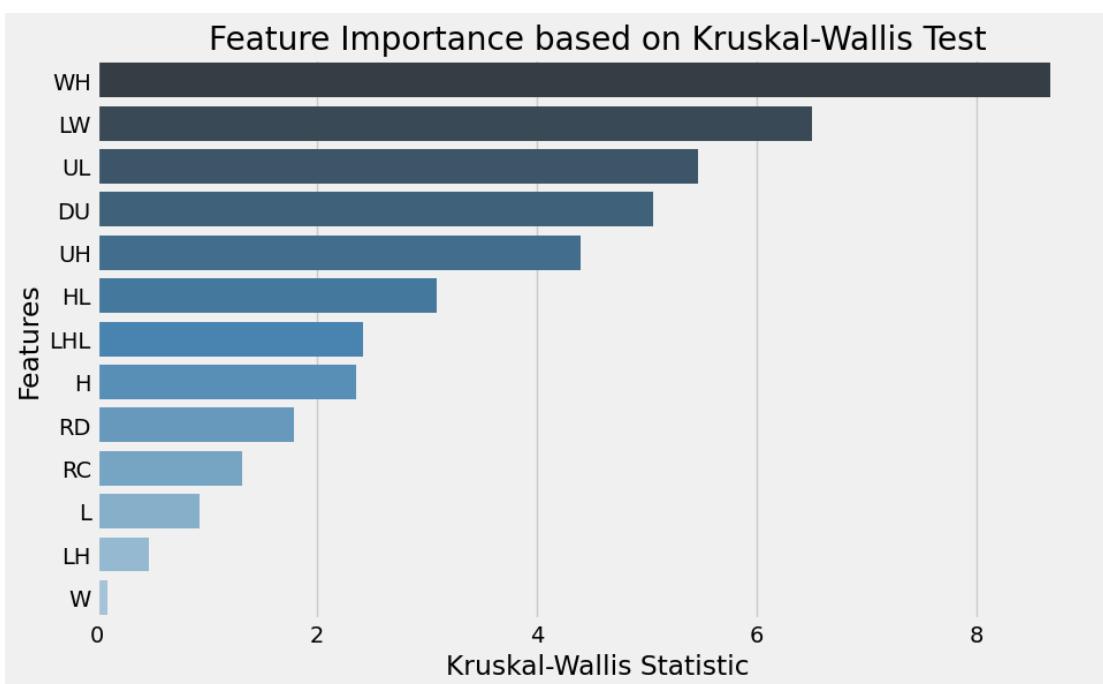


Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

1038 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric  
1039 method that is suitable for evaluating differences in distributions across groups  
1040 when the data does not follow a normal distribution. This approach was chosen  
1041 because of the non-normality of the dataset and its robustness in handling con-  
1042 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,  
1043 2024)

1044 The analysis showed that the width-to-height ratio (WH\_ratio) had the high-  
1045 est importance score, indicating it is the most statistically significant feature for

1046 distinguishing the sex of *T. granosa*. Other notable features included the length-to-width ratio (LW\_ratio), umbos-to-length ratio (UL\_ratio), and the distance between the umbos, all of which contributed significantly to the classification task.  
1047  
1048  
1049

#### 1050 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	60.24	56.98	56.69	56.39
Random Forest	59.07	59.46	59.06	58.74
Gradient Boosting	60.27	60.98	60.27	59.96
AdaBoost	60.63	60.98	60.63	60.39

Table 4.2: Performance Metrics for Models with All 13 Features

1051 In table 4.2, the performance of different machine learning models is presented  
1052 using the full set of 13 features from the dataset. AdaBoost emerges as the  
1053 highest-performing model, with an accuracy of 60.63%, precision of 60.98%, recall  
1054 of 60.63%, and an F1-score of 60.39%. These results suggest that AdaBoost is  
1055 particularly effective when utilizing all available features, likely due to its ability  
1056 to combine multiple weak learners into a more robust model.

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	62.20	59.69	59.08	58.69
Random Forest	62.96	60.10	59.85	59.54
Gradient Boosting	63.39	64.24	64.16	64.04
AdaBoost	61.02	61.26	61.02	60.82

Table 4.3: Performance Metrics for Models with 5 Features

1057 Table 4.3 presents the performance of the same models using only the top 5  
1058 features identified through Kruskal-Wallis feature importance analysis. The top  
1059 5 features selected are distance between the umbos, length-to-width ratio, width-  
1060 to-height ratio, umbos-to-height ratio and umbos-to-length ratio.

1061 Interestingly, the performance of the models improves with the reduced fea-  
1062 ture set. K-Nearest Neighbors (KNN) achieves the highest performance in this

1063 scenario, with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and  
1064 an F1-score of 63.75%. These results suggest that KNN benefits from using only  
1065 the most significant features, showing a notable improvement over its performance  
1066 when all 13 features are used.

## 1067 4.2 Deep Learning Analysis

1068 This section presents the performance of the Convolutional Neural Network (CNN)  
1069 model in classifying the sex of *Tegillarca granosa* based on shell morphology. The  
1070 analysis evaluates the model's ability to distinguish between male and female  
1071 shell images using various evaluation metrics. This part of the paper includes  
1072 six subsections: baseline model, comparison of individual and combined angles,  
1073 training result and hyperparameter tuning, proposed model, learning rates and  
1074 training behavior per fold, and visualizations.

1075 The machine learning analysis (see Figure 4.1.4) revealed that five of the orig-  
1076 inal features produced significant results. The K-Nearest Neighbor (KNN) model  
1077 achieved an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1  
1078 score of 63.75%. This section compares the model's performance across differ-  
1079 ent angles based on the results of the machine learning and feature importance  
1080 analysis.

### 1081 4.2.1 Baseline Model

1082 This section presents the baseline model with a batch size of 16 and 20 epochs,  
1083 which will serve as the starting point for comparison and provide a guideline for  
1084 hyperparameter tuning. The focus will be on one of the angles, specifically the  
1085 Left Lateral view, since the feature importance analysis using the Kruskal-Wallis  
1086 Test indicated that the width-to-height ratio had the highest importance score,  
1087 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)

1088 The unbalanced dataset, which consisted of 144 male samples and 127 female

1089 samples, achieved an accuracy of 65.27%, precision of 71.82%, recall of 58.99%,  
 1090 an F1-score of 63.99%, an AUC score of 73.08%, and a loss of 0.6122. However, to  
 1091 address the class imbalance and enhance model performance, random undersam-  
 1092 pling was performed. This approach resulted in improved performance metrics for  
 1093 the balanced dataset, with an accuracy of 67.34%, precision of 69.43%, a recall  
 1094 of 64.06%, an F1-score of 65.60%, an AUC score of 74.31%, and a lower loss of  
 1095 0.5981.

### 1096 **4.3 Comparison of Individual and Combined An-** 1097 **gles**

1098 Using the same batch size and number of epochs, performance was compared  
 1099 across all individual angles and the combination of the two highest-performing  
 1100 angles based on accuracy, using a balanced dataset. For the combined analysis,  
 1101 samples from the two selected angles were placed side by side, and a new dataset  
 1102 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)

1103 Table 4.5 presents the performance metrics for each individual angle and the  
 1104 combination of the two highest-performing angles in terms of accuracy. The  
 1105 Left Lateral view achieved the highest accuracy (67.34%) and precision (69.43%),  
 1106 while the Dorsal view obtained the highest recall (77.88%) and F1-score (69.96%).  
 1107 Meanwhile, the Ventral view recorded the highest AUC score (74.87%), indicat-  
 1108 ing its strong ability to distinguish between classes. Combining the Ventral and  
 1109 Left Lateral views resulted in an overall accuracy of 62.60%, suggesting that while  
 1110 combined images may provide complementary information, individual angle views  
 1111 still outperformed the combined views under the current experimental setup.

## 1112 4.4 Training Result and Hyperparameter Tun- 1113 ing

1114 The Left Lateral angle was selected for further optimization. Several experiments  
1115 were conducted by tuning hyperparameters such as batch size, number of epochs,  
1116 and activation functions. Each adjustment was compared against the baseline  
1117 model to enhance performance and develop a robust CNN for sex classification of  
1118 *T. granosa*.

1119 The Left Lateral angle was chosen because it achieved the highest accuracy  
1120 and precision among all individual views, and because the Kruskal-Wallis fea-  
1121 ture importance analysis indicated that the width-to-height ratio, a feature most  
1122 visible from the lateral perspective, was the most significant morphological trait  
1123 for classification. Therefore, focusing on this view was expected to maximize the  
1124 model's learning capacity and improve classification performance.

### 1125 4.4.1 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

1126 Table 4.6 shows the results indicating that a batch size of 32 with 50 epochs  
1127 achieved the best overall performance, with an accuracy of 71.68%, a precision of  
1128 72.52%, a recall of 69.29%, an F1-score of 69.12%, and AUC score of 77.34%.

1129 In contrast, increasing the batch size to 64 resulted in lower recall and F1-  
1130 scores, suggesting that smaller batch Sizes (16 or 32) are more effective for this  
1131 dataset. A moderate batch size of 32 allowed the model to generalize better and  
1132 maintain stable learning, while too large batch sizes may have led to underfitting.

<sup>1133</sup> **Chapter 5**

<sup>1134</sup> **Conclusion and  
1135 Recommendations**

<sup>1136</sup> **5.1 Conclusion**

<sup>1137</sup> **5.2 Recommendations**

<sup>1138</sup> This special problem entitled Morphometric-Based Non-invasive Sex Identification  
<sup>1139</sup> of *T. granosa* focuses on creating a baseline study that will serve as a foundation  
<sup>1140</sup> for further studies involving *Tegillarca granosa*, blood cockles using machine learn-  
<sup>1141</sup> ing, deep learning, and computer vision technologies in determining the sex of the  
<sup>1142</sup> samples is a salient need in aquaculture practices. Thus, the proceeding rec-  
<sup>1143</sup> ommendations are the future applications to improve and have detailed analysis  
<sup>1144</sup> such as focusing on shape analysis, exploring other state-of-the-art CNN such as  
<sup>1145</sup> ResNet, SqueezeNet, and InceptionNet, and comparing the analysis result. Fur-  
<sup>1146</sup> thermore, the main goal of conducting this is to have the ability to identify the  
<sup>1147</sup> sex of the samples by taking real-time angles by rotating from the dorsal, lateral,  
<sup>1148</sup> and ventral.

<sup>1149</sup> Future studies could also invest in a much sturdier and more controlled envi-  
<sup>1150</sup> ronment by using a green background and positioning a webcam at a fixed angle.  
<sup>1151</sup> In addition, experiment with other image processing techniques such as scaling,  
<sup>1152</sup> rotating, and augmentation. The dataset can be utilized for further analysis us-  
<sup>1153</sup> ing deep learning and computer vision to make sense of the images gathered and  
<sup>1154</sup> discern sexual dimorphism for *T.granosa* or will serve as the basis for conducting  
<sup>1155</sup> similar studies to other bivalve species.

# <sup>1156</sup> References

- <sup>1157</sup> Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten  
<sup>1158</sup> years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,  
<sup>1159</sup> 5–16. doi: 10.1080/11250000409356545
- <sup>1160</sup> Afiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara  
<sup>1161</sup> granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,  
<sup>1162</sup> *10*.
- <sup>1163</sup> Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian  
1164 Sains*, *14*, 14207.
- <sup>1165</sup> Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).  
<sup>1166</sup> Data scaling performance on various machine learning algorithms to identify  
<sup>1167</sup> abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:  
<sup>1168</sup> 10.14710/jtsiskom.2021.14105
- <sup>1169</sup> Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.  
<sup>1170</sup> *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- <sup>1171</sup> Awan, A. A. (2022, November). *A complete guide to data augmentation*. Retrieved from <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>
- <sup>1172</sup> Aypa, S. M., & Bacongus, S. R. (2000). Philippines: mangrove-friendly aquaculture.  
<sup>1173</sup> 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).
- <sup>1174</sup> BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,  
<sup>1175</sup> Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic  
<sup>1176</sup> Resources.
- <sup>1177</sup> Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of  
<sup>1178</sup> waste cockle shell (anadara granosa) in biodiesel production from palm olein:  
<sup>1179</sup> Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.  
<sup>1180</sup> doi: 10.1016/j.fuel.2011.03.002
- <sup>1181</sup> Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining  
1182 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127

- 1188 .v1
- 1189 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,  
1190 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A  
1191 new sex determination system in freshwater mussels (bivalvia: Unionoida)?  
1192 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/  
1193 msq345
- 1194 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-  
1195 proaches, challenges and opportunities for aquaculture. *Journal of Marine  
1196 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1197 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).  
1198 Mass mortalities in bivalve populations: A review of the edible cockle  
1199 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:  
1200 10.1016/j.ecss.2014.04.011
- 1201 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic  
1202 research in bivalves: Towards the identification of molecular markers of  
1203 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:  
1204 10.1016/j.jprot.2012.04.027
- 1205 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data  
1206 augmentation for forecasting with deep learning. *ArXiv (Cornell Univer-  
1207 sity)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1208 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly  
1209 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1210 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan  
1211 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:  
1212 10.1093/icb/ict076
- 1213 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).  
1214 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma  
1215 edule*) geographical traceability based on shell image analysis. *Ecological  
1216 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1217 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method  
1218 for chinese mitten crab using deep convolutional neural network. *Multi-  
1219 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1220 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle  
1221 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-  
1222 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
- 1223 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect  
1224 ageing of eel, *anguilla anguilla* (l.)? *International Review of Hydrobiology*,  
1225 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1226 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and  
1227 anatomy for young learners*. Rosie Research. Retrieved from <https://>

- 1230                    [rosiereresearch.com/clam-dissection-anatomy/](http://rosiereresearch.com/clam-dissection-anatomy/)  
 1231     *Fao 2024 report: Sustainable aquatic food systems important for global food*  
 1232        *security – european fishmeal.* (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.  
 1233  
 1234  
 1235     Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and  
 1236        elemental composition: Complementary tools for stock discrimination of  
 1237        mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,  
 1238        110, 75–83. doi: 10.1016/j.fishres.2011.03.014  
 1239     GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation.* Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>  
 1240  
 1241  
 1242     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)  
 1243  
 1244  
 1245     Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture.* Oxford: Blackwell Science.  
 1246  
 1247     Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1  
 1248  
 1249     Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and  
 1250        surface water characterization and heavy metal health risk on cockles in  
 1251        kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.  
 1252        263–271). doi: 10.1016/j.sbspro.2016.05.156  
 1253     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)  
 1254  
 1255  
 1256  
 1257     Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest  
 1258        example of sexual dimorphism in bivalves—evidence from the astartid  
 1259        *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,  
 1260        95(6), 1216–1225. doi: 10.1017/jpa.2021.48  
 1261     Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts  
 1262        as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–  
 1263        234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4  
 1264     Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep  
 1265        learning-based phenotype classification of three ark shells: *Anadara kagoshimensis*, *tegillarca granosa*, and *anadara broughtonii*. *Frontiers in Marine Science*, 11. doi: 10.3389/fmars.2024.1356356  
 1266  
 1267  
 1268     Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the  
 1269        granulated ark, *tegillarca granosa* (linne). *The Korean Journal of Malacology*, 13, 55–64.  
 1270  
 1271     Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as

- 1272 a tool for stock identification of the southern blue whiting, micromesistius  
1273 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1274 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &  
1275 et al. (2016). Identifying blue whiting (micromesistius poutassou) stock  
1276 structure in the northeast atlantic by otolith shape analysis. *Canadian*  
1277 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/  
1278 cjfas-2015-0332
- 1279 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle  
1280 tegillarca granosa (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.*  
1281 *Arts Sci*, 4.
- 1282 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and  
1283 larval rearing of the blood cockle, tegillarca granosa (linnaeus, 1758), in the  
1284 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2  
1285 .001
- 1286 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of  
1287 local populations of the common sole solea solea (pisces, soleidae) in the nw  
1288 mediterranean through otolith morphometrics and shape analysis. *Marine*  
1289 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1290 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and  
1291 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–  
1292 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/  
1293 kja.21209
- 1294 Narasimham, K. A. (1988). Taxonomy of the blood clams anadara (tegillarca)  
1295 granosa (linnaeus, 1758) and a. (t.) rhombea (born, 1780).
- 1296 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,  
1297 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world  
1298 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1299 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith  
1300 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.  
1301 doi: 10.1002/jmor.10439
- 1302 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric  
1303 morphometrics and machine learning challenge currently accepted species  
1304 limits of the land snail placostylus (pulmonata: Bothriembryontidae) on the  
1305 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.  
1306 doi: 10.1093/mollus/eyz031
- 1307 Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test*.  
1308 Retrieved from [https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal\\_wallis/](https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/) (Accessed: 2025-04-23)
- 1309 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,  
1310 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-  
1311 carbons from consumption of blood cockle and exposure to contaminated  
1312 sediments and water along the klang strait, malaysian. *Marine Pollution*

- 1314      *Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004

1315      Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence  
1316      of seasonal variation and anthropogenic stress on blood cockle (*tegillarca*  
1317      *granosa*) production potential. *Influence of Seasonal Variation and Anthro-*  
1318      *pogenic Stress on Blood Cockle (Tegillarca Granosa) Production Potential*,  
1319      44(2), 62–82.

1320      Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-  
1321      chine learning and its applications to biology. *PLoS Computational Biology*,  
1322      3(6), e116. doi: 10.1371/journal.pcbi.0030116

1323      Team, K. (n.d.). *Keras documentation: Reducelronplateau*. Retrieved from  
1324      [https://keras.io/api/callbacks/reduce\\_lr\\_on\\_plateau/](https://keras.io/api/callbacks/reduce_lr_on_plateau/)

1325      Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-  
1326      ductive process and early development. In V. S. Kennedy, R. I. E. Newell,  
1327      & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).  
1328      College Park, MD: Maryland Sea Grant.

1329      Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning  
1330      approach for morphological feature extraction based on variational auto-  
1331      encoder: An application to mandible shape. *Npj Systems Biology and Ap-*  
1332      *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6

1333      Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in*  
1334      *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in\_3366.pdf

1335      Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.  
1336      *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972  
1337      .5008949

1338      Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*  
1339      *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.

1340      Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-  
1341      pacts of four commonly used nanoparticles on the metabolism of a ma-  
1342      rine bivalve species, *tegillarca granosa*. *Chemosphere*, 296, 134079. doi:  
1343      10.1016/j.chemosphere.2022.134079

1344      Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the  
1345      blood clam *tegillarca granosa*. *Marine Environmental Research*, 177. Re-  
1346      trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

<sup>1347</sup> **Appendix A**

<sup>1348</sup> **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*

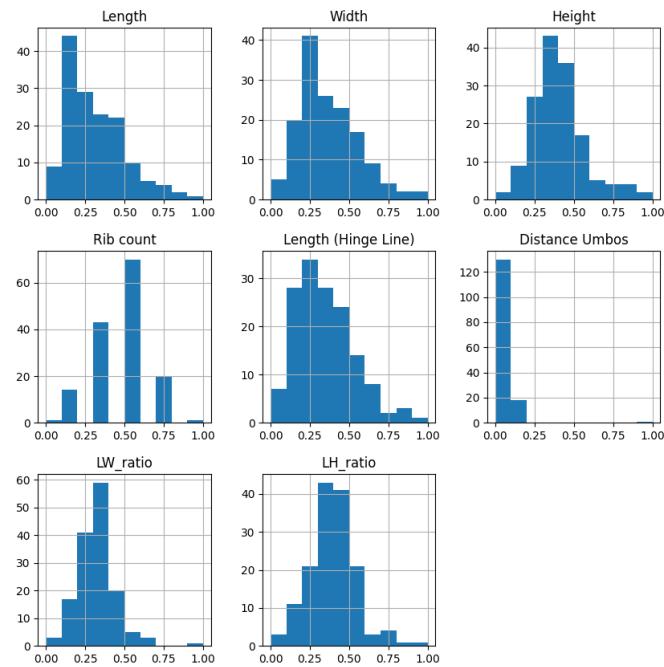


Figure A.7: Distribution of the Features of *Tegillarca granosa*