

<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 23, 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<br><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 |          |                                                                            |           |
| 61 | 2.3.7    | Machine Learning for Geographical Traceability in Bivalves                 | 16        |
| 62 | 2.4      | Limitations on Sex Identification in <i>Tegillarca granosa</i> . . . . .   | 16        |
| 63 | 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 | 3.7.3    | Evaluation Metrics for Machine Learning . . . . .                          | 29        |
|    |          |                                                                            |           |
| 74 | <b>4</b> | <b>Results and Discussions</b>                                             | <b>31</b> |
| 75 | 4.1      | Machine Learning Analysis . . . . .                                        | 31        |
| 76 | 4.1.1    | Data Exploration . . . . .                                                 | 31        |
| 77 | 4.1.2    | Statistical Analysis . . . . .                                             | 32        |
| 78 | 4.1.3    | Feature Importance Analysis . . . . .                                      | 33        |
| 79 | 4.1.4    | Performance Evaluation . . . . .                                           | 34        |

|    |                                                                  |           |
|----|------------------------------------------------------------------|-----------|
| 80 | <b>5 Conclusion and Recommendations</b>                          | <b>36</b> |
| 81 | 5.1 Conclusion . . . . .                                         | 36        |
| 82 | 5.2 Recommendations . . . . .                                    | 36        |
| 83 | <b>References</b>                                                | <b>37</b> |
| 84 | <b>A Data Gathering Documentation and Supplementary Analysis</b> | <b>42</b> |

# <sup>85</sup> List of Figures

|     |                                                                                                        |    |
|-----|--------------------------------------------------------------------------------------------------------|----|
| 86  | 2.1 Diagram of <i>Tegillarca granosa</i> Anatomy . . . . .                                             | 7  |
| 87  | 3.1 Male and Female <i>Tegillarca granosa</i> shells . . . . .                                         | 22 |
| 88  | 3.2 Different Views of the <i>T. granosa</i> Shell Captured . . . . .                                  | 24 |
| 89  | 3.3 Linear Measurements of <i>Tegillarca granosa</i> shell. . . . .                                    | 25 |
| 90  | 3.4 Image Acquisition Setup for <i>T. granosa</i> Samples . . . . .                                    | 26 |
| 91  | 3.5 Data Preprocessing Pipeline . . . . .                                                              | 27 |
| 92  | 3.6 Diagram of k-fold cross-validation with k = 5 . . . . .                                            | 29 |
| 93  | 4.1 Correlation heatmap of morphometric features with the sex of <i>T.</i><br><i>granosa</i> . . . . . | 32 |
| 94  | 4.2 Feature Importance Scores Using the Kruskal-Wallis Test . . . . .                                  | 33 |
| 96  | A.1 Sex Identification Through Spawning of <i>Tegillarca granosa</i> . . . . .                         | 42 |
| 97  | A.2 Separating Male and Female Samples After Spawning of <i>Tegillarca</i><br><i>granosa</i> . . . . . | 42 |
| 99  | A.3 Sex Identified Female Through Dissecting of <i>Tegillarca granosa</i> .                            | 43 |
| 100 | A.4 Sex Identified Male Through Dissecting of <i>Tegillarca granosa</i> . .                            | 43 |
| 101 | A.5 Linear Measurements of Female <i>Tegillarca granosa</i> . . . . .                                  | 43 |
| 102 | A.6 Linear Measurements of Male <i>Tegillarca granosa</i> . . . . .                                    | 44 |



# <sup>104</sup> List of Tables

|                |                                                                  |    |
|----------------|------------------------------------------------------------------|----|
| <sup>105</sup> | 2.1 Comparison of the Methods Used in Bivalves Studies . . . . . | 19 |
| <sup>106</sup> | 4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison | 32 |
| <sup>107</sup> | 4.2 Performance Metrics for Models with 2 Features . . . . .     | 34 |

<sup>108</sup> **Chapter 1**

<sup>109</sup> **Introduction**

<sup>110</sup> **1.1 Overview**

<sup>111</sup> The Philippines is a global center of marine biodiversity and has established aqua-  
<sup>112</sup> culture as a significant contributor to total fishery production (Aypa & Baconguis,  
<sup>113</sup> 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-  
<sup>114</sup> ally and is the 11th largest seafood producer in the world. Aquaculture is deeply  
<sup>115</sup> integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-  
<sup>116</sup> duction of various aquatic species, including bivalves. Among these, blood cockles  
<sup>117</sup> (*Tegillarca granosa*) hold considerable economic and environmental significance,  
<sup>118</sup> making it essential to ensure sustainable production and population balance.

<sup>119</sup> Maintaining a balanced male-to-female ratio of blood cockles is crucial to pre-  
<sup>120</sup> vent overharvesting and ensure sustainability. An imbalanced ratio can lead to  
<sup>121</sup> overexploitation and negatively impact the population's viability. However, there  
<sup>122</sup> is limited literature on *T. granosa* that provides a thorough understanding of its  
<sup>123</sup> sex-determining mechanisms, particularly regarding sexual dimorphism based on  
<sup>124</sup> morphological and morphometric characteristics (Breton, Capt, Guerra, & Stew-  
<sup>125</sup> art, 2017).

<sup>126</sup> Currently, sex determination methods for blood cockles are invasive, including  
<sup>127</sup> dissection and histological examinations, which often result in the death of the  
<sup>128</sup> species. While there is growing literature on sex identification in aquaculture  
<sup>129</sup> commodities using machine learning and deep learning, there is a notable scarcity  
<sup>130</sup> of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

<sup>131</sup> This study, titled "Morphometric-Based Non-Invasive Sex Identification of

<sup>132</sup> Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)," aims to provide a detailed  
<sup>133</sup> baseline analysis of blood cockles by leveraging their morphological and morpho-  
<sup>134</sup> metric characteristics. Sexual dimorphism in bivalves is often subtle and chal-  
<sup>135</sup> lenging to establish mascropically (Karapunar, Werner, Fürsich, & Nützel, 2021).  
<sup>136</sup> However, by integrating machine learning and deep learning, the study seeks to  
<sup>137</sup> identify distinct features that may indicate sexual dimorphism between male and  
<sup>138</sup> female blood cockles.

## <sup>139</sup> 1.2 Problem Statement

<sup>140</sup> Identifying the sex of *T. granosa* is important for promoting sustainable aquacul-  
<sup>141</sup> ture and biodiversity by maintaining a balanced male-to-female ratio. A balanced  
<sup>142</sup> ratio helps prevent overharvesting. Although sex identification is crucial for blood  
<sup>143</sup> cockle population management and sustainable aquaculture, there is a notable  
<sup>144</sup> lack of research on creating non-invasive methods for determining the sex of *T.*  
<sup>145</sup> *granosa*. Many recent studies and approaches rely on invasive methods like dis-  
<sup>146</sup> section or histological analysis, which are impractical for large-scale aquaculture  
<sup>147</sup> operations focused on conservation.

<sup>148</sup> Current methods for determining the sex of *T. granosa* are invasive and in-  
<sup>149</sup> volve dissection, which requires cutting open the shell to visually inspect the  
<sup>150</sup> gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-  
<sup>151</sup> quently leads to their death. Another method is histological examination, where  
<sup>152</sup> tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,  
<sup>153</sup> 2021). Both approaches are labor-intensive and time-consuming, and can pose  
<sup>154</sup> risks to population management, particularly when maintaining a balanced sex  
<sup>155</sup> ratio for breeding programs is essential. Moreover, these invasive methods require  
<sup>156</sup> specialized technical skills for accurate execution. Resource-limited aquaculture  
<sup>157</sup> operations face significant challenges in accessing the necessary laboratory equip-  
<sup>158</sup> ment, such as microscopes and staining tools, complicating the process.

<sup>159</sup> A less invasive approach employed by aquaculturists involves monitor spawning  
<sup>160</sup> behavior, where individuals are separated and stimulated to reproduce in order  
<sup>161</sup> to determine their sex through the release of gametes (Miranda & Ferriols, 2023).  
<sup>162</sup> Although this method is indeed less invasive than dissection, it still induces stress  
<sup>163</sup> in blood cockles and may not be completely effective for fast identification in large  
<sup>164</sup> populations.

<sup>165</sup> Given the limitations of both invasive and less invasive methods, there is a  
<sup>166</sup> clear need for a more advanced approach. An alternative, non-invasive method

<sup>167</sup> involving machine and deep learning technologies could address these issues by  
<sup>168</sup> providing a fast, accurate, and effective solution without harming or stressing the  
<sup>169</sup> blood cockles.

## <sup>170</sup> 1.3 Research Objectives

### <sup>171</sup> 1.3.1 General Objective

<sup>172</sup> The general objective of this study is to develop a non-invasive method for iden-  
<sup>173</sup> tifying the sex of *Tegillarca granosa* using machine and deep learning integrated  
<sup>174</sup> with computer vision technologies. This method aims to provide accurate and  
<sup>175</sup> streamlined sex identification without causing harm to the specimens, thus sup-  
<sup>176</sup> porting sustainable aquaculture practices.

### <sup>177</sup> 1.3.2 Specific Objectives

<sup>178</sup> To achieve the overall general objective of developing a non-invasive sex identifi-  
<sup>179</sup> cation of *T. granosa* using machine learning, deep learning, and computer vision  
<sup>180</sup> technologies, the following specific objectives have been established:

- <sup>181</sup> 1. To collect and organize a comprehensive dataset of *T. granosa* which will  
<sup>182</sup> include high-quality images and relevant morphological measurements that  
<sup>183</sup> will serve as the basis for the machine-learning model.
- <sup>184</sup> 2. To develop and implement machine learning models that can classify the  
<sup>185</sup> sex of *T. granosa* based on the collected linear measurements and images of  
<sup>186</sup> different angles of the sample.
- <sup>187</sup> 3. To evaluate the performance of the models used using performance metrics  
<sup>188</sup> such as accuracy, precision, recall, and F1-score.
- <sup>189</sup> 4. To develop a system that can identify the sex of *T. granosa* based on its  
<sup>190</sup> morphological characteristics.

## **191 1.4 Scope and Limitations of the Research**

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

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

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

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

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

## **226 1.5 Significance of the Research**

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

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

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

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

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

<sup>250</sup> **Chapter 2**

<sup>251</sup> **Review of Related Literature**

<sup>252</sup> Aquaculture is the fastest-growing industry in animal food production and has  
<sup>253</sup> great potential as a sustainable solution to global food security, nutrition, and  
<sup>254</sup> development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*  
<sup>255</sup> *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-  
<sup>256</sup> tegrated into the livelihoods of Filipinos, not only through fish cultivation but  
<sup>257</sup> also through the production of other aquatic species, including mollusks, oysters,  
<sup>258</sup> clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood  
<sup>259</sup> clams *Tegillarca granosa*, have economic and environmental significance. It has  
<sup>260</sup> been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-  
<sup>261</sup> vesting and maintain the optimal ratio to preserve the population and production  
<sup>262</sup> of the blood cockles.

<sup>263</sup> The members of the Arcidae Family, including *T. granosa* are important  
<sup>264</sup> sources of food and livelihood. Cockle aquaculture meets rising demands, however,  
<sup>265</sup> it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,  
<sup>266</sup> 2023). To solve the problem, researchers exert a considerable amount of effort,  
<sup>267</sup> developing a broader understanding of bivalves, including their sex-determining  
<sup>268</sup> mechanism, due to their notable importance in terms of diversity, environmental  
<sup>269</sup> benefits, and economic and market importance (Breton et al., 2017). Despite the  
<sup>270</sup> promising idea of identifying sex, there is limited research reported in terms of  
<sup>271</sup> sexual dimorphism, making it harder to distinguish through its morphological and  
<sup>272</sup> morphometric characteristics.

<sup>273</sup> By addressing the challenges in the sex identification of *T. granosa*, it would be  
<sup>274</sup> able to address one problem at a time. Currently, there are no recent documented  
<sup>275</sup> publications that integrate machine learning and computer vision in characterizing  
<sup>276</sup> sexual dimorphism, reducing complexity, variability in sex determination, and

<sup>277</sup> differentiation mechanisms in bivalves, including *T. granosa* specifically.

## <sup>278</sup> **2.1 Background on *Tegillarca granosa* and Their 279 Importance**

<sup>280</sup> *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.  
<sup>281</sup> In the Philippines, it is commonly known as a Litob, a marine bivalve species from  
<sup>282</sup> the family Arcidae. Litob is widely distributed in the world including Southeast  
<sup>283</sup> Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest  
<sup>284</sup> (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat  
<sup>285</sup> as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the  
<sup>286</sup> blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood  
<sup>287</sup> phenotype where it serves two purposes the hemocyte carries oxygen around the  
<sup>288</sup> body and strengthens immune defenses. In addition, it possesses a unique ability  
<sup>289</sup> to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

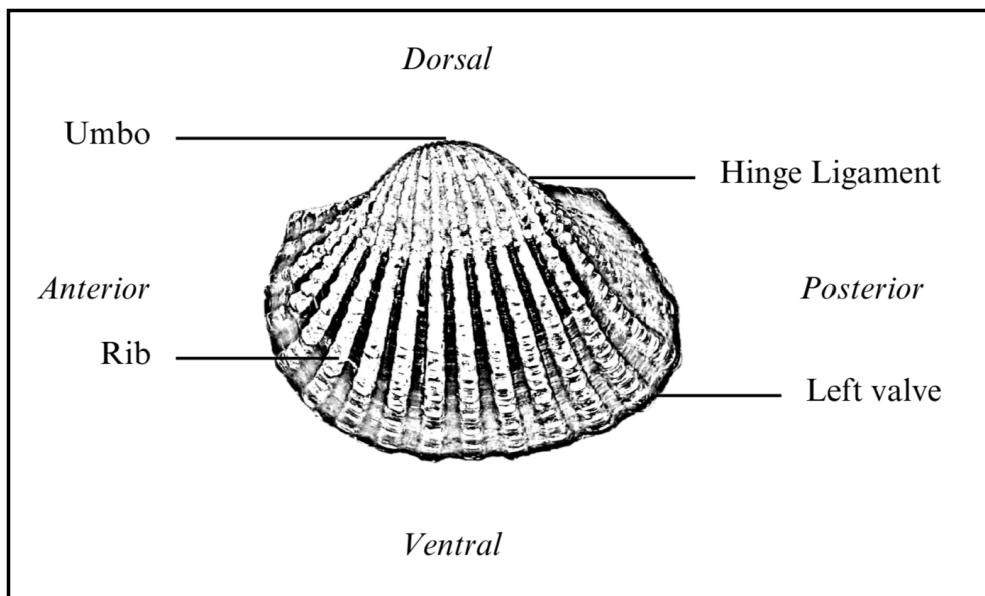


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

<sup>290</sup> *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both  
<sup>291</sup> valves being equal in size but asymmetrical from the hinge. The top edge of  
<sup>292</sup> the dorsal margin is straight, while the front is rounded and slopes downward,  
<sup>293</sup> with its back being obliquely rounded with a concave bottom edge. It has a  
<sup>294</sup> narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,  
<sup>295</sup> although some may be incomplete. The shell's outer layer, or the periostracum, is

296 smooth and brown with a straight hinge line and 40-68 fine short teeth arranged  
297 in a straight line. The beak, or prosogyrate, curves forward, with the shell having  
298 18–21 raised ribs with blunt nodules and spaces between them. The inner shell is  
299 white with crenulations along the valves' ventral, anterior, and posterior margins.  
300 The posterior adductor scar is elongated and squarish, while the anterior adductor  
301 scar is similar but smaller in size. The mantle covering the bulk of *T. granosa*'s  
302 visceral mass is thin but the edges are thick and muscular. It bears the impression  
303 of the crenulated shell edges. Their foot is large with a ventral groove with no byssus  
304 or thread-like attachment. The *T. granosa*'s soft body is blood red (Narasimham,  
305 1988).

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

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

<sup>336</sup> males at the time of harvest (Budd, Banh, Domingos, & Jerry, 2015).

<sup>337</sup> Clearly, much more work is required to understand the mechanisms under-  
<sup>338</sup> lying sexual dimorphism in bivalves, specifically *T. granosa*. Just like the other  
<sup>339</sup> aquaculture commodities, sex affects not just reproduction but it can affect mar-  
<sup>340</sup> ket preference and underlying economic value, making the determination of sex  
<sup>341</sup> important for meeting consumer demands. These are the increasing significance  
<sup>342</sup> of the *T. granosa* despite the lack of reviewed articles in the Philippines.

## <sup>343</sup> **2.2 Current Methods of Sex Identification in *Tegillarca granosa***

<sup>344</sup>

<sup>345</sup> The current sex identification methods in *Tegillarca granosa* range from invasive  
<sup>346</sup> histological techniques to less invasive methodologies like temperature-induced  
<sup>347</sup> spawning. Each approach comes with its pros and cons regarding accuracy, feasi-  
<sup>348</sup> bility, and impact on natural populations.

<sup>349</sup> Induced spawning and larval rearing are considered the less invasive techniques  
<sup>350</sup> used to study *Tegillarca granosa*. In the Philippines, limited research has been  
<sup>351</sup> done on the *Tegillarca granosa* (Linnaeus, 1758), and this study, titled Initial At-  
<sup>352</sup> tempts on Spawning and Larval Rearing of the Blood Cockle, *Tegillarca granosa*  
<sup>353</sup> in the Philippines, is conducted by Denise Vergara Miranda and Victor Marco  
<sup>354</sup> Emmanuel Nuestro Ferriols (2023). The researchers conducted experiments on  
<sup>355</sup> induced spawning and larval rearing, discovering that the eggs of female *T. gra-*  
<sup>356</sup> *nosa* were salmon pink, while the sperm released by males looked milky. After  
<sup>357</sup> spawning, the researchers successfully generated 6, 531, 000 fertilized eggs.

<sup>358</sup> They highlighted the importance of *T. granosa* and other anadarinids as a  
<sup>359</sup> food source that was established worldwide, especially in Malaysia and Korea.  
<sup>360</sup> However, in the Philippines, the bivalve aquaculture of the clam species is still  
<sup>361</sup> limited. The experiment which focuses on the culture and rearing of *T. granosa*  
<sup>362</sup> was attempted by subjecting the wild broodstocks to a series of temperature fluc-  
<sup>363</sup> tuations to induce the spawning of gametes. This is currently the most natural  
<sup>364</sup> and least invasive method for bivalves (Aji, 2011). The study of Miranda and  
<sup>365</sup> Ferriols aimed to pave the way to the sustainable production of *T. granosa* seeds  
<sup>366</sup> for aquaculture production and stock enhancement despite the scarcity of docu-  
<sup>367</sup> mented hatchery culture of *T. granosa* from larvae to adults that is available in  
<sup>368</sup> the Philippines.

<sup>369</sup> In the study entitled "The earliest example of sexual dimorphism in bivalves —

370 evidence from the astartid *Nicanella* (Lower Jurassic, southern Germany)," the  
371 researchers utilized Principal Component Analysis and Fourier Analysis as a non-  
372 invasive method that investigates sexual expression in the *Nicanella rakoveci*. In  
373 the study, researchers discovered that the bivalves with crenulations were found to  
374 have a different shell shape, which made them more inflated than those without  
375 crenulations. This suggests that when they became females, they adapted to  
376 hold more eggs rather than for protection from predators as previously thought.  
377 The formation of crenulations is likely part of the genetic process that controls  
378 both the sex change and the changes in shell structure (Karapunar et al., 2021).  
379 Overall, the findings demonstrate that the genetic mechanisms for sex change and  
380 shell morphology in bivalves existed as early as the Early Jurassic, contributing  
381 to our understanding of bivalve diversity and evolution. Thus, the researchers  
382 concluded that crenulations serve as a morphological marker for identifying the  
383 sex and reproductive stage of these bivalves (Karapunar et al., 2021).

384 On the other hand, invasive techniques such as histological analysis offer a  
385 more thorough but harmful method for determining the sex of *T. granosa*. A  
386 study on the Spawning Period of Blood Cockle *Tegillarca granosa* (Linnaeus,  
387 1758) in Myeik Coastal. 240 blood cockle samples were examined for sex and  
388 gonad maturity stages using histological examination, with shell lengths ranging  
389 from 26-35mm and shell weights from 8.1-33g. For histological analysis, the whole  
390 soft tissues were removed from the shell and the flesh containing most parts of  
391 the gonads was fixed in formalin, dehydrated in an upgraded series of ethanol,  
392 and cleared in xylene. This invasive method allows for precise identification of  
393 the gonadal maturation stages based on the cellular and structural changes in the  
394 gonads.

395 The classification of the gonad stages used was by Yurimoto et al. (2014).  
396 There are five maturation stages of gonadal development: immature (Stage I),  
397 developing (Stage II), mature (Stage III), spawning (Stage IV), and spent (Stage  
398 V) stages. The sex of the *T. granosa* was confirmed by the color of the gonad and  
399 by conducting a histological examination of the gonads. During the immature  
400 stage, sex determination was indistinguishable due to the difficulties of observing  
401 the germ cells. In the developing stage, the spermatocytes and a few spermatids  
402 can be seen for males, and immature oocytes are attached to the tube wall for  
403 the female. In the mature stage, the follicles are full of spermatozoa with their  
404 tails pointing towards the center of the tube for the male, and the female is full  
405 of mature oocytes that are irregular or polygonal in shape with the oval nucleus.  
406 Upon reaching spawning, some spermatozoa are released, causing the empty space  
407 in the follicle wall for males and females. There is a decrease in the number of  
408 mature oocytes and it exhibits nuclear disappearance due to the breakdown of  
409 the germinal vesicle. Lastly, the spent stage is where the genital tube is deformed

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

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

## 428 **2.3 Machine Learning and Deep Learning in Bi- 429 ological Studies**

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

440 **2.3.1 Deep Learning for Phenotype Classification in Ark  
441 Shells**

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

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

455 **2.3.2 Geometric Morphometrics and Machine Learning for  
456 Species Delimitation**

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

467 **2.3.3 Contour Analysis in Mollusc Shells Using Machine  
468 Learning**

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

472 evolutionary influences. The researchers stated that classic morphometric meth-  
473 ods may not accurately depict morphological features of the shell, especially when  
474 using the angular decomposition of the contour. The current research examined  
475 and compared the robustness of the contour analysis using wavelet transformed  
476 and Elliptic Fourier descriptors for gastropod shells with enlarged spines. For  
477 that, the researchers analyzed two geographically and ecologically separated pop-  
478 ulations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed  
479 that contour analysis of gastropod shells with enlarged spines can be analyzed  
480 using both methodologies, but the wavelet analysis provided better local discrim-  
481 ination. From an ecological perspective, shells with various sizes of spines in both  
482 areas indicate the broad adaptability of the species.

#### 483 2.3.4 Machine Learning for Shape Analysis of Marine Or- 484 ganisms

485 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*  
486 to *Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*  
487 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to  
488 the traditional method of collecting linear measurements with the application of  
489 multivariate statistical methods and outline analysis in recording the structures  
490 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and  
491 elasmobranchs) with their hard bodies have been used as an indication of age and  
492 a determinable time-scale and structure continue to go through life (Arkhipkin,  
493 2005; Kerr & Campana, 2014). This study has explored variations in the mor-  
494 phometry of recording structures in stock discrimination and systematics. The  
495 researchers utilized the principal component analysis rather than the traditional  
496 approach, which helps simplify the data without losing important information.  
497 They utilized landmark-based geometric morphometrics, which has three differ-  
498 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other  
499 morphogenetic processes, and lastly, the extremal points are constructed land-  
500 marks.

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

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

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

### 531 **2.3.5 Deep Learning for Landmark-Free Morphological Fea- 532 ture Extraction**

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

### 543    2.3.6 Machine Learning for Sex Differentiation in Abalone

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

550    Similarly, in the study, *Data scaling performance on various machine learning*  
551    *algorithms to identify abalone sex*, the researchers of the University of India (2022)  
552    focused on the data scaling performance of various machine learning algorithms to  
553    identify the abalone sex, specifically using min-max normalization and zero-mean  
554    standardization. The different machine learning algorithms are the Supervised  
555    Vector Machine (SVM), Random Forest, Naive Bayesian, and Decision Tree. Their  
556    study aims to utilize machine learning in terms of identifying the trends and  
557    distribution patterns in the abalone dataset. Eight features of the abalone dataset  
558    (length, diameter, height, whole weight, shucked weight, viscera weight, shell  
559    weight, ring) were used to determine the three sexes of Abalone. Their data has  
560    been grouped based on sex which are Female, Male, and Infant. They utilized  
561    the Synthetic Minority Oversampling Technique (SMOTE) in data balancing for  
562    the preprocessing of the data. Followed by data scaling or normalization where  
563    it converts numeric values in a data set to a general scale without distorting  
564    differences in the range of values. Then they classified by splitting the data into  
565    training and testing sets (Arifin, Ariawan, Rosalia, Lukman, & Tufailah, 2021).

566    The study found that Naive Bayes consistently performed better than other al-  
567    gorithms. However, when applied to both min-max and zero-mean normalization,  
568    the average accuracies of the algorithms were as follows: Random Forest (62.37%),  
569    SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-  
570    nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with  
571    normalization, Random Forest achieved the highest overall metrics, including an  
572    average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of  
573    83.31%. Liu et al. concluded that Random Forest is highly accurate because it  
574    can handle large, complex datasets, run processes in parallel using multiple trees,  
575    and select the most relevant features to enhance model performance (Arifin et al.,  
576    2021).

577 **2.3.7 Machine Learning for Geographical Traceability in**  
578 **Bivalves**

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

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

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

605 **2.4 Limitations on Sex Identification in *Tegillarca***  
606 ***granosa***

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

611 perature shock, is the most natural and least invasive method for bivalves (Aji,  
612 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water  
613 from hot to cold with a specific temperature requires deliberate and careful ma-  
614 nipulation of the temperature over a specific period and would require constant  
615 management and monitoring.

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

624 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous  
625 (synchronous or functional) characteristics. Sequential hermaphrodites switch  
626 genders after being male or female for one or multiple yearly cycles. (Heller,  
627 1993; Gosling, 2004; Collin, 2013). Sex change and consecutive hermaphroditism  
628 have been observed in different bivalve species, including Ostreidae, Pectinidae,  
629 Veneridae, and Patellidae. However, macroscopically differentiating bivalve sex is  
630 challenging. The only way it may be identified is through histological analysis of  
631 gonad remains but to do so there is an act of killing the organism (Coe, 1943;  
632 Gosling, 2004). Verification of sex change in bivalves to classify whether male or  
633 female while they are alive is challenging since they need to be re-confirmed and  
634 re-evaluated to be the same individual after a year.

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

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

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

## 654 **2.5 Synthesis of the Study**

655 This section of the paper summarizes the technologies used in the different studies  
656 related to the pursuit of the study entitled, Morphometric-Based Non-Invasive Sex  
657 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 Quennec and S. A. Trewick and F. Brescia and M. Morgan-Richards                                                   | Neural network analysis (supervised learning) and Gaussian mixture models (unsupervised learning)         | To determine whether the shape and size of the snail's shells can distinguish between two <i>Placostylus</i> species, particularly in groups that appear to be hybrids.                                                                                              | Combining geometric morphometrics and machine learning effectively answers biological issues, providing insights into species classification and possible hybridization. | Difficulty classifying intermediate phenotypes, with potential for overfitting and misclassification in both learning methods.    |
| V. M. Tuset and E. Galimany and A. Farrés and E. Marco-Herrero and J. L. Otero-Ferret and A. Lombarte and M. Ramón        | Wavelet functions and Elliptic Fourier descriptors                                                        | Addresses the difficulty of accurately defining phenotypic diversity in gastropod shells.                                                                                                                                                                            | Advanced contour analysis methods allow accurate differentiation of gastropod shell forms.                                                                               | Cannot clarify the causes of phenotypic variation in the two populations studied.                                                 |
| Fedor Lishchenko and Jones, J. B.                                                                                         | Landmark- and outline-based Geometric Morphometric methods                                                | To address difficulties in differentiating between stocks of marine organisms to prevent misidentification that could affect conservation and management.                                                                                                            | Shape analysis improves taxonomic classification precision and offers close distinction between related species or organisms.                                            | Landmark-based methods can be sensitive to landmark placement.                                                                    |
| M. Tsutsumi and N. Saito and D. Koyabu and C. Furusawa                                                                    | Morphological regulated variational AutoEncoder (Morpho-VAE)                                              | The need for reliable, landmark-free methods, such as a modified variational autoencoder, to extract and decipher complex shapes from image data.                                                                                                                    | Employs dimension reduction and feature extraction, making it a user-friendly tool for biology non-experts.                                                              | Limited sample size in certain families presented challenges.                                                                     |
| Barrera-Hernandez, R. and Barrera-Soto, V. and Martinez-Rodriguez, J. L. and Ríos-Alvarado, A. B. and Ortiz-Rodríguez, F. | Machine learning algorithms                                                                               | Identifying the sex of abalones is challenging for producers applying specific growth or preservation strategies.                                                                                                                                                    | Machine learning algorithms accurately classify abalone sex into three categories: male, female, and immature.                                                           | Selected features may not fully capture the complexity of abalone morphology.                                                     |
| Concepcion, R. and Guillermo, M. and Tanner, S. E. and Fonseca, V. and Duarte, B.                                         | EfficientNet-Bo, ResNet101, MobileNetV2, InceptionV3                                                      | Addresses the difficulty of accurately tracing bivalve harvesting origins using computer vision and machine learning algorithms to enhance seafood traceability and combat food fraud.                                                                               | Non-invasive, image-based tools for bivalve traceability provide faster, cheaper, and equally accurate alternatives to traditional chemical analysis methods.            | Small sample size (only 30 cockles) limits model reliability.                                                                     |

Table 2.1: Comparison of the Methods Used in Bivalves Studies

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

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

# <sup>672</sup> Chapter 3

## <sup>673</sup> Research Methodology

<sup>674</sup> This chapter discussed the materials and methods employed in the study, focusing  
<sup>675</sup> on the development requirements, as well as the software and programming  
<sup>676</sup> languages utilized. It also detailed the overall workflow in conducting the study,  
<sup>677</sup> Morphometric-Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca*  
<sup>678</sup> *granosa* (Linnaeus), 1758) using machine learning and deep learning technologies.

<sup>679</sup> Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,  
<sup>680</sup> oversaw the overall workflow and conduct of the experiment. The researchers were  
<sup>681</sup> also guided by research associates LC Mae Gasit and Allena Esther Artera. Con-  
<sup>682</sup> sequently, the entire dataset collection process was conducted at the University of  
<sup>683</sup> the Philippines Visayas hatchery facility.

<sup>684</sup> The methodology consisted of eight parts: (1) Sample Collection, (2) Ethical  
<sup>685</sup> Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-  
<sup>686</sup> tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-  
<sup>687</sup> ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine  
<sup>688</sup> Learning, and (8) Morphological Characteristics Evaluation Using Deep Learning

### <sup>689</sup> 3.1 Sample Collection

<sup>690</sup> The collection of *T. granosa* samples used in this study was part of an ongoing  
<sup>691</sup> research project by UPV DOST-PCAARRD titled "Establishment of the Center  
<sup>692</sup> for Mollusc Research and Development: Development of Spawning and Hatchery  
<sup>693</sup> Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture."  
<sup>694</sup> A total of 271 samples were provided for this study to classify the sex of *T. granosa*.

695 The samples, ranging in size from 34 to 61 mm, were sourced from the coastal area  
696 of Zaraga, Iloilo, and fish markets in Ivisan, Capiz, Philippines (see Figure 3.1).

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

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

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

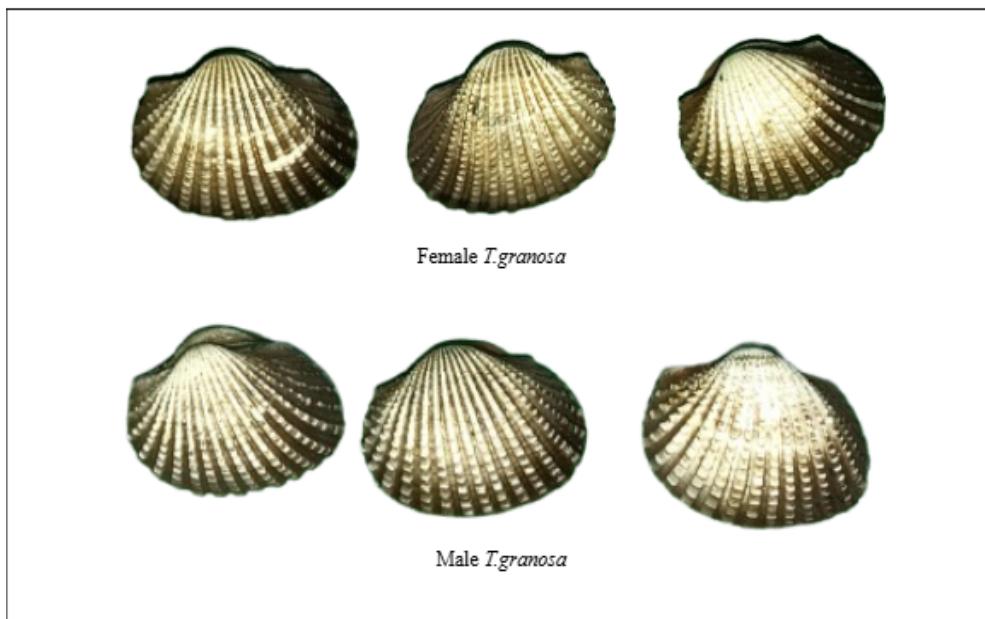


Figure 3.1: Male and Female *Tegillarca granosa* shells

## **713 3.2 Ethical Considerations**

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

## **720 3.3 Creating *T. granosa* Dataset**

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

**729** Images of each sample were captured and saved in JPG format using a stan-  
**730** dardized file naming convention that included the sample's sex, the shell's ori-  
**731** entation or view, and its corresponding number out of the 271 total samples. File  
**732** names for female *T. granosa* samples began with "0", while those for male sam-  
**733** ples began with "1". Each file name also included one of the six captured views:  
**734** (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right  
**735** lateral (refer to Figure 3.2), followed by a unique sample number. For exam-  
**736** ple, "010001" denoted the first female sample taken from the dorsal view, while  
**737** "110001" represented the first male sample from the same view. This naming  
**738** convention was implemented to prevent data leakage and ensure accurate labeling  
**739** of images according to their respective samples.

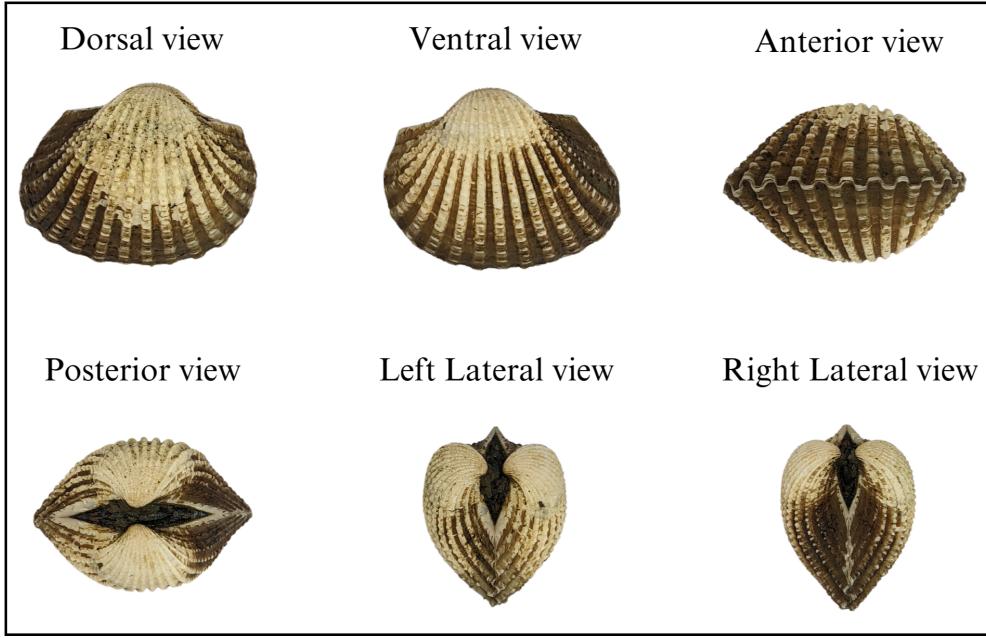


Figure 3.2: Different Views of the *T. granosa* Shell Captured

### <sup>740</sup> 3.4 Morphological and Morphometric Characteristics Collection

<sup>741</sup> Morphology refers to biological form and is one of the most visually recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).  
<sup>742</sup> In this study, morphological characteristics describe the structural features of *T. granosa*, focusing on measurable attributes such as shape, size, and color.  
<sup>743</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>751</sup> The researchers measured the height, width, and length of *T. granosa* using  
<sup>752</sup> a Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the  
<sup>753</sup> corresponding measurement diagram. Length (A) refers to the distance from the  
<sup>754</sup> anterior to the posterior of the shell. Width (B) is defined as the widest span  
<sup>755</sup> across the shell from the left to the right valve. Height (C) measures the distance  
<sup>756</sup> from the base to the apex of the shell. In addition, the hinge line length (D) near  
<sup>757</sup> the hinge and the distance between the umbos (E) were recorded.

<sup>758</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.

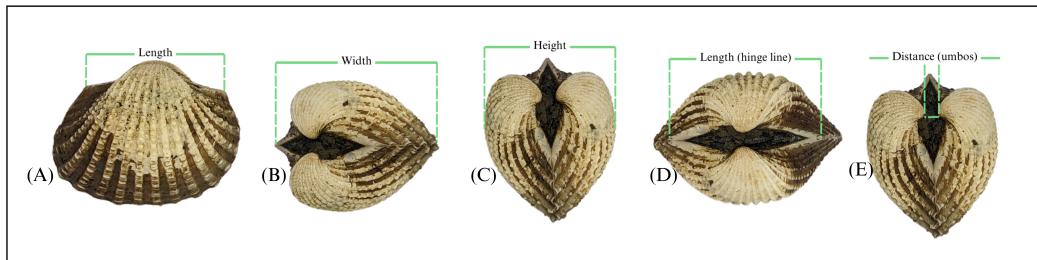


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

### 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

### 777 3.6 Hardware and Software Configuration

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

### 789 3.7 Morphometric Characteristics Evaluation Us- 790 ing Machine Learning

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

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

800 **3.7.1 Data Preprocessing**

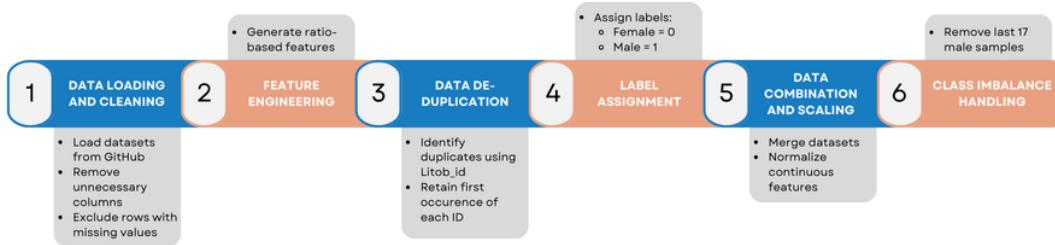


Figure 3.5: Data Preprocessing Pipeline

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

804 ***Data Loading and Cleaning***

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

809 ***Feature Engineering***

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

816 ***Data De-duplication***

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

821        ***Label Assignment***

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

825        ***Data Combination and Scaling***

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

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

835        ***Class Imbalance Handling***

836        After normalization, class imbalance was addressed by removing the last 17  
837        samples from the male dataset. This ensured an equal number of male and female  
838        samples (127 each), helping to prevent model bias and improving the reliability  
839        of classification performance across both classes.

840        **3.7.2 Machine Learning Models Training**

841        ***Model Selection and Hyperparameter Tuning***

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

847        ***Cross- Validation***

848        A five-fold cross-validation approach was implemented. The dataset was di-  
849        vided into five subsets, with four used for training and one for testing. This  
850        process was repeated five times, with each fold serving as the test set once. This  
851        method ensured that model evaluation was robust and generalizable, minimizing

852 the bias that may result from a single train-test split. (GeeksforGeeks, 2024)

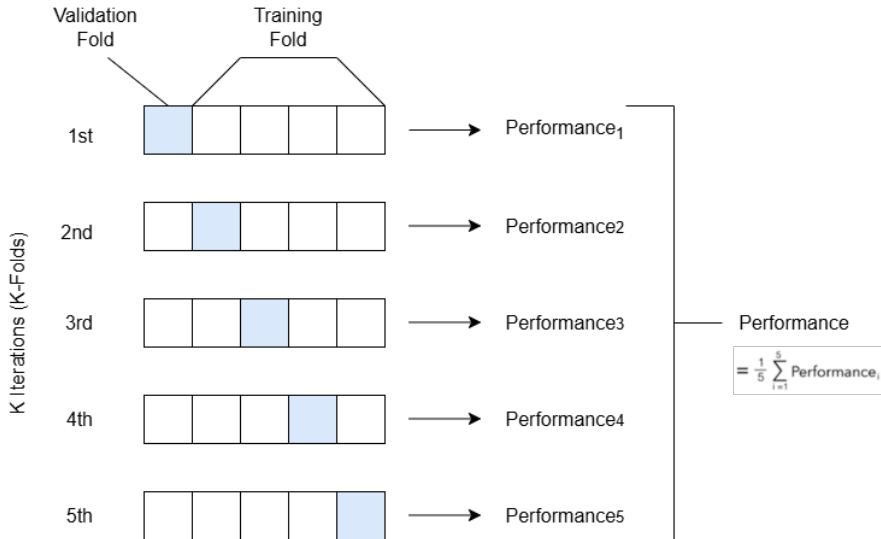


Figure 3.6: Diagram of k-fold cross-validation with  $k = 5$

### 853 3.7.3 Evaluation Metrics for Machine Learning

854 Evaluating the performance of the binary classification model is important as well  
855 as selecting the appropriate metrics that is based on the requirements of the user.  
856 The performance of the supervised machine learning models will be measured  
857 based on four metrics namely: accuracy, precision, recall, and F1 score.

858 Accuracy (ACC) is the ratio of the overall correctly predicted samples to the  
859 total number of examples in the evaluation dataset (Cui, Pan, Chen, & Zou, 2020).  
860 The overall correctness of the model in predicting male and female blood cockles.  
861 This metric could help in understanding how well the model performs across all  
862 classifications. The formula for the accuracy is:

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

863 Precision (PREC) is the ratio between correctly predicted samples in all samples  
864 that are assigned to the positive class (Cui et al., 2020). This metric promotes fair  
865 representation and prevents the misidentification of blood cockles as it identifies  
866 potential inaccuracies or biases. The formula for precision is:

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

867 Recall (REC) is known as the sensitivity or the true positive rate (TPR) which  
 868 is the ratio of the correctly predicted cases from all the samples assigned to the  
 869 actual positive cases (Cui et al., 2020). This metric is the ability of the model to  
 870 correctly identify positive male and female samples. The formula for the recall is:

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

871 F1 score is defined as the mean of the precision and recall in which it penalizes  
 872 the extreme values of either of the two (Cui et al., 2020). The formula for the F1  
 873 is:

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

<sup>874</sup> **Chapter 4**

<sup>875</sup> **Results and Discussions**

<sup>876</sup> This chapter presents the results of the machine learning and deep learning anal-  
<sup>877</sup> yses conducted on the preprocessed dataset. Preprocessing was performed using  
<sup>878</sup> Python in Google Colaboratory. The chapter includes the evaluation of various  
<sup>879</sup> machine learning classifiers, analysis of feature importance, and the application  
<sup>880</sup> of deep learning models for image-based classification. These approaches aim to  
<sup>881</sup> identify key predictors and assess classification performance for sex identification  
<sup>882</sup> in *T. granosa*.

<sup>883</sup> **4.1 Machine Learning Analysis**

<sup>884</sup> **4.1.1 Data Exploration**

<sup>885</sup> Exploratory data analysis was performed to characterize the dataset using visu-  
<sup>886</sup> alizations to understand the patterns and correlations within the data. A corre-  
<sup>887</sup> lation heatmap was created to assess the relationship between the predictors and  
<sup>888</sup> the target variable.

<sup>889</sup> The heatmap (see Figure 4.1) revealed three features most correlated with the  
<sup>890</sup> sex of *T. granosa*: the width-height ratio ( $r = 0.18$ ), the umbos-length ratio ( $r$   
<sup>891</sup> = 0.12), and the distance between the umbos ( $r = 0.12$ ). Each of these features  
<sup>892</sup> demonstrated a weak positive relationship with the target variable.

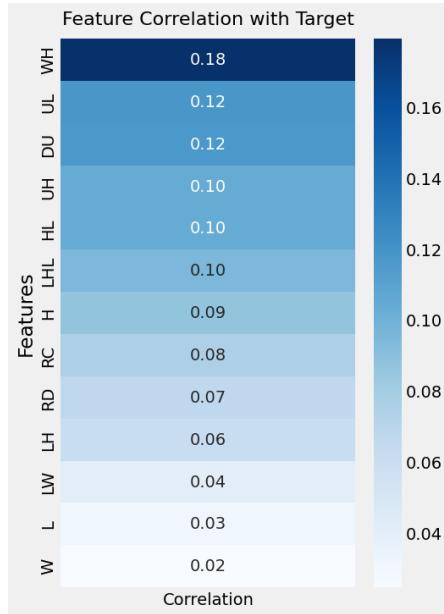


Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

### <sup>893</sup> 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

<sup>894</sup> As part of the exploratory data analysis, statistical testing confirmed that the  
<sup>895</sup> dataset did not follow a normal distribution. Consequently, the Mann-Whitney  
<sup>896</sup> U test was applied with a significance level of  $\alpha = 0.05$  to compare male and

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

901 It is important to note that statistical significance does not imply predic-  
902 tive importance. Therefore, further analysis—such as feature importance evalua-  
903 tion—was performed to identify the most informative predictors for classification.

#### 904 4.1.3 Feature Importance Analysis

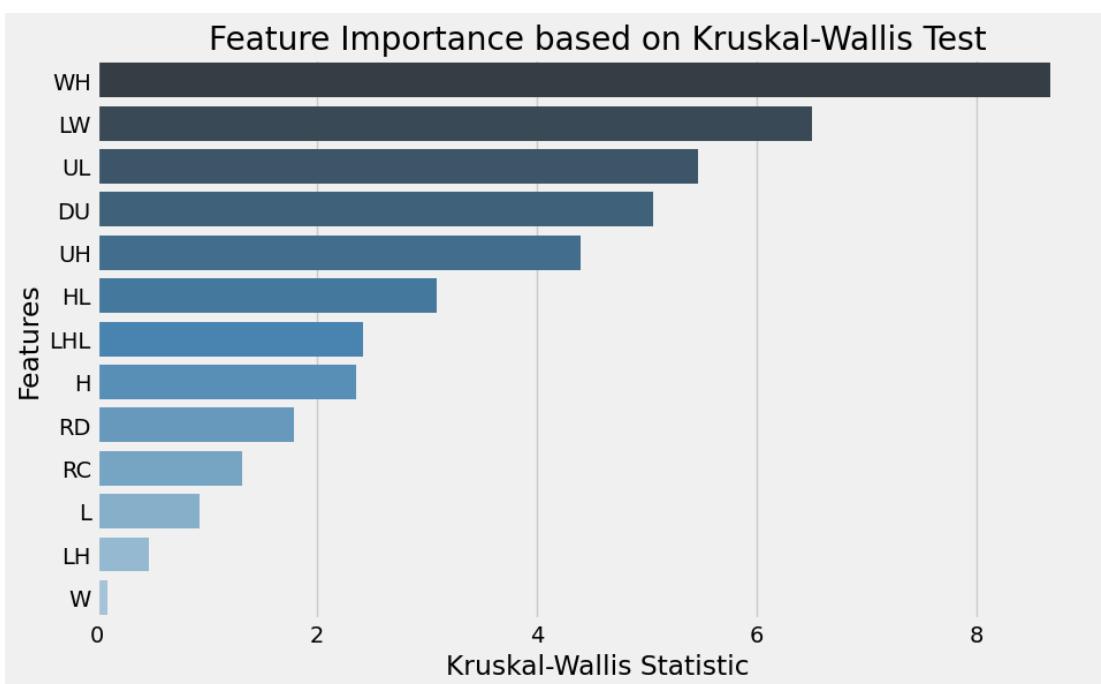


Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

905 Feature importance was assessed using the Kruskal-Wallis test, a non-parametric  
906 method that is suitable for evaluating differences in distributions across groups  
907 when the data does not follow a normal distribution. This approach was chosen  
908 because of the non-normality of the dataset and its robustness in handling con-  
909 tinuous and ordinal data without assuming homogeneity of variances. (Ribeiro,  
910 2024)

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

913 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.

#### 917 4.1.4 Performance Evaluation

| Model                  | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|------------------------|--------------|---------------|------------|--------------|
| Support Vector Machine | 64.89        | 65.88         | 64.89      | 64.22        |
| Logistic Regression    | 64.50        | 65.37         | 64.50      | 63.85        |
| K-Nearest Neighbors    | 64.14        | 64.52         | 64.14      | 63.99        |
| Extra Trees            | 63.75        | 63.22         | 62.96      | 62.84        |
| Random Forest          | 62.96        | 60.81         | 60.59      | 60.38        |
| Gradient Boosting      | 62.56        | 63.26         | 62.56      | 62.22        |
| AdaBoost               | 60.60        | 61.39         | 60.60      | 60.03        |

Table 4.2: Performance Metrics for Models with 2 Features

918 Table 4.2 presents the performance metrics of various machine learning models 919 when trained using only the two most informative features, as identified through 920 feature selection. The evaluation metrics—accuracy, precision, recall, and F1- 921 score—collectively provide a comprehensive view of each model’s ability to 922 correctly classify the sex of *T. granosa* specimens.

923 Interestingly, despite the simplicity of using only two features, this configuration 924 yielded the highest classification performance across all feature sets. Among 925 the models, the Support Vector Machine (SVM) achieved the best results, with 926 an accuracy of 64.89%, precision of 65.88%, recall of 64.89%, and an F1-score 927 of 64.22%. Logistic Regression and K-Nearest Neighbors (KNN) also performed 928 competitively, with F1-scores of 63.85% and 63.99%, respectively.

929 In contrast, ensemble-based models such as Random Forest, Gradient Boosting, 930 and AdaBoost demonstrated slightly lower performance, potentially due to 931 the limited number of features restricting their ability to capture complex patterns 932 and interactions typically leveraged by such models.

933 It is important to note that these machine learning models were also trained 934 using the full set of 13 features and an intermediate set of 5 features, as detailed 935 in Table 4.1. Surprisingly, the two-feature configuration outperformed both larger 936 feature sets across most models. This finding underscores the value of effective 937 feature selection, suggesting that simpler models with fewer inputs can sometimes 938 yield better generalization and performance than more complex configurations.

<sup>939</sup> Moreover, reducing the number of features has practical advantages in terms of  
<sup>940</sup> data collection, model interpretability, and computational efficiency.

941 **Chapter 5**

942 **Conclusion and  
943 Recommendations**

944 **5.1 Conclusion**

945 **5.2 Recommendations**

946 This special problem entitled Morphometric-Based Non-invasive Sex Identification  
947 of *T. granosa* focuses on creating a baseline study that will serve as a foundation  
948 for further studies involving *Tegillarca granosa*, blood cockles using machine learn-  
949 ing, deep learning, and computer vision technologies in determining the sex of the  
950 samples is a salient need in aquaculture practices. Thus, the proceeding rec-  
951 ommendations are the future applications to improve and have detailed analysis  
952 such as focusing on shape analysis, exploring other state-of-the-art CNN such as  
953 ResNet, SqueezeNet, and InceptionNet, and comparing the analysis result. Fur-  
954 thermore, the main goal of conducting this is to have the ability to identify the  
955 sex of the samples by taking real-time angles by rotating from the dorsal, lateral,  
956 and ventral.

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

## 964 References

- 965 Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten  
966 years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,  
967 5–16. doi: 10.1080/11250000409356545
- 968 Afifiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara  
969 granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,  
970 10.
- 971 Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian  
972 Sains*, *14*, 14207.
- 973 Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).  
974 Data scaling performance on various machine learning algorithms to identify  
975 abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:  
976 10.14710/jtsiskom.2021.14105
- 977 Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.  
978 *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- 979 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquacul-  
980 ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-  
981 tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on  
982 mangrove-friendly aquaculture organized by the seafdec aquaculture depart-  
983 ment, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- 984 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,  
985 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic  
986 Resources.
- 987 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of  
988 waste cockle shell (anadara granosa) in biodiesel production from palm olein:  
989 Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.  
990 doi: 10.1016/j.fuel.2011.03.002
- 991 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining  
992 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127  
993 .v1
- 994 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,  
995 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A

- 996 new sex determination system in freshwater mussels (bivalvia: Unionoida)?  
997 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/  
998 msq345
- 999 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-  
1000 proaches, challenges and opportunities for aquaculture. *Journal of Marine  
1001 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1002 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).  
1003 Mass mortalities in bivalve populations: A review of the edible cockle  
1004 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:  
1005 10.1016/j.ecss.2014.04.011
- 1006 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic  
1007 research in bivalves: Towards the identification of molecular markers of  
1008 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:  
1009 10.1016/j.jprot.2012.04.027
- 1010 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly  
1011 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1012 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan  
1013 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:  
1014 10.1093/icb/ict076
- 1015 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).  
1016 Bivalvenet: A hybrid deep neural network for common cockle (cerastoderma  
1017 edule) geographical traceability based on shell image analysis. *Ecological  
1018 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1019 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method  
1020 for chinese mitten crab using deep convolutional neural network. *Multi-  
1021 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1022 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle  
1023 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-  
1024 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
- 1025 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect  
1026 ageing of eel, *anguilla anguilla* (l.)? *International Review of Hydrobiology*,  
1027 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1028 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and  
1029 anatomy for young learners*. Rosie Research. Retrieved from <https://rosieresearch.com/clam-dissection-anatomy/>
- 1030 Fao 2024 report: Sustainable aquatic food systems important for global food  
1031 security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1032 Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and

- 1038 elemental composition: Complementary tools for stock discrimination of  
1039 mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,  
1040 110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1041 GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn*. Re-  
1042 trieval from [https://www.geeksforgeeks.org/cross-validation-using](https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/)  
1043 [-k-fold-with-scikit-learn/](#) (Accessed: 2025-04-23)
- 1044 Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Black-  
1045 well Science.
- 1046 Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean*  
1047 *Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1048 Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and  
1049 surface water characterization and heavy metal health risk on cockles in  
1050 kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.  
1051 263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1052 Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a com-  
1053 prehensive guide to data rescaling*. DataCamp. Retrieved from <https://www>  
1054 [.datacamp.com/tutorial/normalization-in-machine-learning](#) (Ac-  
1055 cessed 2025-04-23)
- 1056 Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The ear-  
1057 liest example of sexual dimorphism in bivalves—evidence from the astar-  
1058 tid *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,  
1059 95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1060 Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts  
1061 as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–  
1062 234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1063 Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep  
1064 learning-based phenotype classification of three ark shells: Anadara kagoshi-  
1065 mensis, tegillarca granosa, and anadara broughtonii. *Frontiers in Marine*  
1066 *Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1067 Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the  
1068 granulated ark, tegillarca granosa (linne). *The Korean Journal of Malacol-  
1069 ogy*, 13, 55–64.
- 1070 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as  
1071 a tool for stock identification of the southern blue whiting, micromesistius  
1072 australis. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1073 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &  
1074 et al. (2016). Identifying blue whiting (micromesistius poutassou) stock  
1075 structure in the northeast atlantic by otolith shape analysis. *Canadian*  
1076 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/  
1077 cjfas-2015-0332
- 1078 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle  
1079 tegillarca granosa (linnaeus, 1758) in myein coastal areas. *J. Myanmar Acad.*

- 1080                  *Arts Sci*, 4.
- 1081    Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and  
1082         larval rearing of the blood cockle, *tegillarca granosa* (linnaeus, 1758), in the  
1083         philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2  
1084         .001
- 1085    Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of  
1086         local populations of the common sole *solea solea* (pisces, soleidae) in the nw  
1087         mediterranean through otolith morphometrics and shape analysis. *Marine  
1088         Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1089    Narasimham, K. A. (1988). Taxonomy of the blood clams *anadara* (*tegillarca*)  
1090         *granosa* (linnaeus, 1758) and a. (t.) *rhombea* (born, 1780).
- 1091    Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,  
1092         M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world  
1093         fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1094    Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith  
1095         shape of different fish species? *Journal of Morphology*, 267(7), 750–757.  
1096         doi: 10.1002/jmor.10439
- 1097    Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric  
1098         morphometrics and machine learning challenge currently accepted species  
1099         limits of the land snail *placostylus* (pulmonata: Bothriembryontidae) on the  
1100         isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.  
1101         doi: 10.1093/mollus/eyz031
- 1102    Ribeiro, D. (2024, Aug). *Diogo ribeiro. data science. kruskal-wallis test*.  
1103         Retrieved from [https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal\\_wallis/](https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/) (Accessed: 2025-04-23)
- 1104    Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,  
1105         & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-  
1106         carbons from consumption of blood cockle and exposure to contaminated  
1107         sediments and water along the klang strait, malaysia. *Marine Pollution  
1108         Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1109    Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence  
1110         of seasonal variation and anthropogenic stress on blood cockle (*tegillarca*  
1111         *granosa*) production potential. *Influence of Seasonal Variation and Anthro-  
1112         pogenic Stress on Blood Cockle (Tegillarca Granosa) Production Potential*,  
1113         44(2), 62–82.
- 1114    Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-  
1115         chine learning and its applications to biology. *PLoS Computational Biology*,  
1116         3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1117    Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-  
1118         ductive process and early development. In V. S. Kennedy, R. I. E. Newell,  
1119         & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).  
1120         College Park, MD: Maryland Sea Grant.
- 1121

- 1122 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning  
1123 approach for morphological feature extraction based on variational auto-  
1124 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*  
1125 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1126 Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in*  
1127 *the laboratory, malaysia.* (Handle.net) doi: 10.3366/in.3366.pdf
- 1128 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.  
1129 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972  
1130 .5008949
- 1131 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*  
1132 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1133 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Impacts of four commonly used nanoparticles on the metabolism of a ma-  
1134 rine bivalve species, tegillarca granosa. *Chemosphere*, 296, 134079. doi:  
1135 10.1016/j.chemosphere.2022.134079
- 1136 Zhan, P. L., Zha, & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the  
1137 blood clam tegillarca granosa. *Marine Environmental Research*, 177. Re-  
1138 trieval from <https://doi.org/10.1016/j.marenvres.2022.105632>
- 1139

<sub>1140</sub> **Appendix A**

<sub>1141</sub> **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*

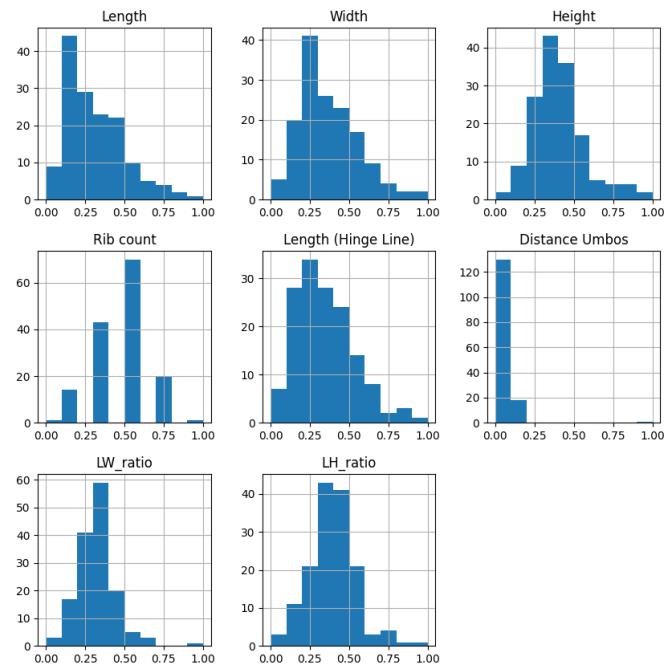


Figure A.7: Distribution of the Features of *Tegillarca granosa*