

1           **MORPHOMETRIC AND MORPHOLOGICAL-BASED**  
2           **NON-INVASIVE SEX IDENTIFICATION OF BLOOD**  
3           **COCKLES *TEGILLARCA GRANOSA* (LINNAEUS,**  
4           **1758)**

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

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

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

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

May 6, 2025

## Abstract

23       *Tegillarca granosa* (Linnaeus, 1758), commonly known as blood cockles, is one  
24 of the most well-known marine bivalves for its nutritional benefits and economic  
25 significance. Determining their sex is essential for maintaining a balanced male-  
26 to-female ratio, which is crucial for preventing the exploitation of this shellfish  
27 resource. The sex-determining mechanism in the shell morphology of bivalves is  
28 challenging macroscopically due to the limited literature regarding this expertise.  
29 In addition, no current technologies are employed to classify the sex based on  
30 shell morphology. This study proposes a machine learning and deep learning ap-  
31 proach for classifying the sex of blood cockles using various linear measurements  
32 (length, width, height, distance between the hinge line, distance between umbos,  
33 and rib count) and angles (dorsal, ventral, anterior, posterior, left lateral, and  
34 right lateral) collected from male and female specimens. Initial machine learning  
35 analysis aimed to determine the best-performing model and the significant fea-  
36 tures. Among the models, K-Nearest Neighbor (KNN) performed best, achieving  
37 an accuracy of 64.16%, a precision of 64.97%, a recall of 64.16%, and an F1-score  
38 of 63.75%. Feature importance analysis indicated that the Width-Height ratio was  
39 the most significant feature. Subsequently, deep learning analysis was conducted  
40 utilizing Convolutional Neural Networks (CNN), using the Left Lateral Angle,  
41 achieving an accuracy of 71.68%, a precision of 72.52%, a recall of 69.29%, an  
42 F1-score of 69.12%, an AUC score of 77.34%. By developing a method to identify  
43 their sex, this study aims to improve the long-term availability of these marine  
44 resources and promote sustainable harvesting.

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

# <sup>46</sup> Contents

<sup>47</sup> <b>1 Introduction</b>	<b>1</b>
<sup>48</sup> 1.1 Overview . . . . .	1
<sup>49</sup> 1.2 Problem Statement . . . . .	2
<sup>50</sup> 1.3 Research Objectives . . . . .	3
<sup>51</sup> 1.3.1 General Objective . . . . .	3
<sup>52</sup> 1.3.2 Specific Objectives . . . . .	3
<sup>53</sup> 1.4 Scope and Limitations of the Research . . . . .	3
<sup>54</sup> 1.5 Significance of the Research . . . . .	4
<sup>55</sup> <b>2 Review of Related Literature</b>	<b>6</b>
<sup>56</sup> 2.1 Background on <i>Tegillarca granosa</i> and Their Importance . . . . .	7
<sup>57</sup> 2.2 Current Methods of Sex Identification in <i>Tegillarca granosa</i> . . . . .	9
<sup>58</sup> 2.3 Machine Learning and Deep Learning in Biological Studies . . . . .	11
<sup>59</sup> 2.3.1 Deep Learning for Phenotype Classification in Ark Shells .	12
<sup>60</sup> 2.3.2 Geometric Morphometrics and Machine Learning for Species <sup>61</sup> Delimitation . . . . .	12
<sup>62</sup> 2.3.3 Contour Analysis in Mollusc Shells Using Machine Learning	12
<sup>63</sup> 2.3.4 Machine Learning for Shape Analysis of Marine Organisms	13

64	2.3.5 Deep Learning for Landmark-Free Morphological Feature Extraction . . . . .	14
65		
66	2.3.6 Machine Learning for Sex Differentiation in Abalone . . . . .	15
67		
68	2.3.7 Machine Learning for Geographical Traceability in Bivalves . . . . .	16
69		
70	2.4 Limitations on Sex Identification in <i>Tegillarca granosa</i> . . . . .	16
71		
72	2.5 Synthesis of the Study . . . . .	18
73		
74	<b>3 Research Methodology</b>	<b>21</b>
75		
76	3.1 Sample Collection . . . . .	21
77		
78	3.2 Ethical Considerations . . . . .	23
79		
80	3.3 Creating <i>T. granosa</i> Dataset . . . . .	23
81		
82	3.4 Morphometric and Morphological Characteristics Collection . . . . .	24
83		
84	3.5 Image Acquisition and Data Gathering . . . . .	25
85		
86	3.6 Hardware and Software Configuration . . . . .	26
87		
88	3.7 Morphometric Characteristics Evaluation Using Machine Learning . . . . .	26
89		
90	3.7.1 Data Preprocessing . . . . .	27
91		
92	3.7.2 Machine Learning Models Training . . . . .	28
93		
94	3.8 Morphological Characteristics Evaluation Using Deep Learning . . . . .	29
95		
96	3.8.1 Convolutional Neural Network . . . . .	30
97		
98	3.8.2 CNN Training . . . . .	32
99		
100	3.9 Evaluation Metrics . . . . .	34
101		
102	<b>4 Results and Discussions</b>	<b>36</b>
103		
104	4.1 Machine Learning Analysis . . . . .	36
105		
106	4.1.1 Data Exploration . . . . .	36
107		

87	4.1.2 Statistical Analysis . . . . .	37
88	4.1.3 Feature Importance Analysis . . . . .	38
89	4.1.4 Performance Evaluation . . . . .	39
90	4.1.5 Confusion Matrix Analysis . . . . .	40
91	4.2 Deep Learning Analysis . . . . .	41
92	4.2.1 Baseline Model . . . . .	41
93	4.2.2 Comparison of Individual and Combined Angles . . . . .	42
94	4.2.3 Training Result and Hyperparameter Tuning . . . . .	42
95	4.2.4 Proposed Model . . . . .	44
96	4.2.5 Learning Rates and Training Behavior per Fold . . . . .	44
97	4.2.6 Visualizations . . . . .	45
98	4.3 Discussions . . . . .	48
99	<b>5 Conclusion and Recommendations</b>	<b>50</b>
100	5.1 Conclusion . . . . .	50
101	5.2 Recommendations . . . . .	51
102	<b>References</b>	<b>53</b>
103	<b>A Data Gathering Documentation</b>	<b>58</b>
104	<b>B Supplementary Analysis</b>	<b>61</b>

# <sup>105</sup> List of Figures

<sup>106</sup>	2.1 Diagram of <i>Tegillarca granosa</i> Anatomy . . . . .	7
<sup>107</sup>	3.1 Male and Female <i>Tegillarca granosa</i> shells . . . . .	22
<sup>108</sup>	3.2 Different Views of the <i>T. granosa</i> Shell Captured . . . . .	24
<sup>109</sup>	3.3 Linear Measurements of <i>Tegillarca granosa</i> shell. . . . .	25
<sup>110</sup>	3.4 Image Acquisition Setup for <i>T. granosa</i> Samples . . . . .	26
<sup>111</sup>	3.5 Data Preprocessing Pipeline . . . . .	27
<sup>112</sup>	3.6 Diagram of k-fold cross-validation with k = 5 . . . . .	29
<sup>113</sup>	3.7 Shadows removed from male samples at different angles . . . . .	30
<sup>114</sup>	3.8 Shadows removed from female samples at different angles . . . . .	30
<sup>115</sup>	3.9 Architecture of Convolutional Neural Network (CNN) . . . . .	31
<sup>116</sup>	3.10 Diagram of stratified k-fold cross-validation with k=5 . . . . .	32
<sup>117</sup>	3.11 Data Augmentation Techniques . . . . .	33
<sup>118</sup>	4.1 Correlation heatmap of morphometric features with the sex of <i>T. granosa</i> . . . . .	37
<sup>119</sup>		
<sup>120</sup>	4.2 Feature Importance Scores Using the Kruskal-Wallis Test . . . . .	38
<sup>121</sup>	4.3 Feature Importance Scores Using the Kruskal-Wallis Test . . . . .	40
<sup>122</sup>	4.4 Training and Validation Accuracy per Fold . . . . .	45

123	4.5	Average Training and Validation Accuracy Across Folds . . . . .	46
124	4.6	Training and Validation Loss per Fold . . . . .	46
125	4.7	Average Training and Validation Loss Across Folds . . . . .	47
126	4.8	Confusion Matrix for Final Model Predictions . . . . .	47
127	4.9	ROC Curve and AUC Score . . . . .	48
128	A.1	Sex Identification Through Spawning of <i>Tegillarca granosa</i> . . . . .	58
129	A.2	Sex-Based Separation of <i>Tegillarca granosa</i> Samples Post-Spawning	58
130	A.3	Sex Identified Female Through Dissection of <i>Tegillarca granosa</i> .	59
131	A.4	Sex Identified Male Through Dissection of <i>Tegillarca granosa</i> .	59
132	A.5	Linear Measurements of Female <i>Tegillarca granosa</i> . . . . .	59
133	A.6	Linear Measurements of Male <i>Tegillarca granosa</i> . . . . .	60
134	A.7	Captured Images of Female <i>Tegillarca granosa</i> . . . . .	60
135	A.8	Captured Images of Male <i>Tegillarca granosa</i> . . . . .	60
136	B.1	Feature Distribution of <i>Tegillarca granosa</i> . . . . .	61
137	B.2	Correlation Matrix of Morphological Variables <i>Tegillarca granosa</i>	62
138	B.3	Feature Maps from First Convolution Layer . . . . .	62
139	B.4	Feature Maps from Second Convolution Layer . . . . .	63
140	B.5	Feature Maps from Third Convolution Layer . . . . .	63

# <sup>141</sup> List of Tables

<sup>142</sup>	2.1 Comparison of the Methods Used in Bivalves Studies . . . . .	19
<sup>143</sup>	4.1 Mann-Whitney U Test Results for Sex-Based Feature Comparison	38
<sup>144</sup>	4.2 Performance Metrics for Models with All 13 Features . . . . .	39
<sup>145</sup>	4.3 Performance Metrics for Models with 5 Features . . . . .	39
<sup>146</sup>	4.4 Performance Metrics for Unbalanced vs. Balanced Datasets (Batch Size: 16, Epochs: 20) . . . . .	41
<sup>147</sup>		
<sup>148</sup>	4.5 Performance Metrics for Individual and Combined Angles (Batch Size: 16, Epochs: 20) . . . . .	42
<sup>149</sup>		
<sup>150</sup>	4.6 Effect of Batch Size and Epoch Values on CNN Model Performance	43
<sup>151</sup>		
<sup>152</sup>	4.7 Performance Metrics for Different Activation Functions (Batch Size: 32, Epochs: 50) . . . . .	43
<sup>153</sup>		
<sup>154</sup>	4.8 Per-Fold Performance Metrics (Batch Size: 32, Epochs: 50, Activation Function: ReLU) . . . . .	44
<sup>155</sup>		
<sup>156</sup>	4.9 Learning Rate Reductions, Early Stopping, and Best Epochs per Fold During 5-Fold Cross-Validation . . . . .	45

<sup>157</sup> **Chapter 1**

<sup>158</sup> **Introduction**

<sup>159</sup> **1.1 Overview**

<sup>160</sup> The Philippines is a global center of marine biodiversity and has established aqua-  
<sup>161</sup> culture as a significant contributor to total fishery production (Aypa & Baconguis,  
<sup>162</sup> 2000; BFAR, 2019). The country produces over 4 million tonnes of seafood annu-  
<sup>163</sup> ally and is the 11th largest seafood producer in the world. Aquaculture is deeply  
<sup>164</sup> integrated into Filipinos' livelihoods, encompassing fish cultivation and the pro-  
<sup>165</sup> duction of various aquatic species, including bivalves. Among these, blood cockles  
<sup>166</sup> (*Tegillarca granosa*) hold considerable economic and environmental significance,  
<sup>167</sup> making it essential to ensure sustainable production and population balance.

<sup>168</sup> Maintaining a balanced male-to-female ratio of blood cockles is crucial to pre-  
<sup>169</sup> vent overharvesting and ensure sustainability. An imbalanced ratio can lead to  
<sup>170</sup> overexploitation and negatively impact the population's viability. However, there  
<sup>171</sup> is limited literature on *T. granosa* that provides a thorough understanding of its  
<sup>172</sup> sex-determining mechanisms, particularly regarding sexual dimorphism based on  
<sup>173</sup> morphometric and morphological characteristics (Breton, Capt, Guerra, & Stew-  
<sup>174</sup> art, 2017).

<sup>175</sup> Currently, sex determination methods for blood cockles are invasive, including  
<sup>176</sup> dissection and histological examinations, which often result in the death of the  
<sup>177</sup> species. While there is growing literature on sex identification in aquaculture  
<sup>178</sup> commodities using machine learning and deep learning, there is a notable scarcity  
<sup>179</sup> of research specifically addressing *T. granosa* (Miranda & Ferriols, 2023).

<sup>180</sup> This study, titled "Morphometric and Morphological-Based Non-Invasive Sex

<sup>181</sup> Identification of Blood Cockles *Tegillarca granosa* (Linnaeus, 1758)," aims to pro-  
<sup>182</sup> vide a detailed baseline analysis of blood cockles by leveraging their morphome-  
<sup>183</sup> tric and morphological characteristics. Sexual dimorphism in bivalves is often  
<sup>184</sup> subtle and challenging to establish macroscopically (Karapunar, Werner, Fürsich,  
<sup>185</sup> & Nützel, 2021). However, by integrating machine learning and deep learning,  
<sup>186</sup> the study seeks to identify distinct features that may indicate sexual dimorphism  
<sup>187</sup> between male and female blood cockles.

## <sup>188</sup> 1.2 Problem Statement

<sup>189</sup> Identifying the sex of *T. granosa* is important for promoting sustainable aquacul-  
<sup>190</sup> ture and biodiversity by maintaining a balanced male-to-female ratio. A balanced  
<sup>191</sup> ratio helps prevent overharvesting. Although sex identification is crucial for blood  
<sup>192</sup> cockle population management and sustainable aquaculture, there is a notable  
<sup>193</sup> lack of research on creating non-invasive methods for determining the sex of *T.*  
<sup>194</sup> *granosa*. Many recent studies and approaches rely on invasive methods like dis-  
<sup>195</sup> section or histological analysis, which are impractical for large-scale aquaculture  
<sup>196</sup> operations focused on conservation.

<sup>197</sup> Current methods for determining the sex of *T. granosa* are invasive and in-  
<sup>198</sup> volve dissection, which requires cutting open the shell to visually inspect the  
<sup>199</sup> gonads (Erica, 2018). This procedure can cause harm to the specimens and fre-  
<sup>200</sup> quently leads to their death. Another method is histological examination, where  
<sup>201</sup> tissue samples are analyzed under a microscope (May, Maung, Phy, & Tun,  
<sup>202</sup> 2021). Both approaches are labor-intensive and time-consuming, and can pose  
<sup>203</sup> risks to population management, particularly when maintaining a balanced sex  
<sup>204</sup> ratio for breeding programs is essential. Moreover, these invasive methods require  
<sup>205</sup> specialized technical skills for accurate execution. Resource-limited aquaculture  
<sup>206</sup> operations face significant challenges in accessing the necessary laboratory equip-  
<sup>207</sup> ment, such as microscopes and staining tools, complicating the process.

<sup>208</sup> A less invasive approach employed by aquaculturists involves monitor spawning  
<sup>209</sup> behavior, where individuals are separated and stimulated to reproduce in order  
<sup>210</sup> to determine their sex through the release of gametes (Miranda & Ferriols, 2023).  
<sup>211</sup> Although this method is indeed less invasive than dissection, it still induces stress  
<sup>212</sup> in blood cockles and may not be completely effective for fast identification in large  
<sup>213</sup> populations.

<sup>214</sup> Given the limitations of both invasive and less invasive methods, there is a  
<sup>215</sup> clear need for a more advanced approach. An alternative, non-invasive method

<sup>216</sup> involving machine and deep learning technologies could address these issues by  
<sup>217</sup> providing a fast, accurate, and effective solution without harming or stressing the  
<sup>218</sup> blood cockles.

## <sup>219</sup> 1.3 Research Objectives

### <sup>220</sup> 1.3.1 General Objective

<sup>221</sup> The general objective of this study is to develop a non-invasive method for iden-  
<sup>222</sup> tifying the sex of *Tegillarca granosa* using machine and deep learning integrated  
<sup>223</sup> with computer vision technologies. This method aims to provide accurate and  
<sup>224</sup> streamlined sex identification without causing harm to the specimens, thus sup-  
<sup>225</sup> porting sustainable aquaculture practices.

### <sup>226</sup> 1.3.2 Specific Objectives

<sup>227</sup> To achieve the overall general objective of developing a non-invasive sex identifi-  
<sup>228</sup> cation of *T. granosa* using machine learning, deep learning, and computer vision  
<sup>229</sup> technologies, the following specific objectives have been established:

- <sup>230</sup> 1. To collect and organize a comprehensive dataset of *T. granosa* which will  
<sup>231</sup> include high-quality images and relevant morphological measurements that  
<sup>232</sup> will serve as the basis for the machine-learning model.
- <sup>233</sup> 2. To develop and implement machine learning and deep learning models that  
<sup>234</sup> can classify the sex of *T. granosa* based on the collected linear measurements  
<sup>235</sup> and images of different angles of the sample.
- <sup>236</sup> 3. To evaluate the performance of the models used using performance metrics  
<sup>237</sup> such as accuracy, precision, recall, and F1-score, and AUC-ROC score for  
<sup>238</sup> deep learning.

## <sup>239</sup> 1.4 Scope and Limitations of the Research

<sup>240</sup> This study is conducted alongside the ongoing research by the UPV DOST-  
<sup>241</sup> PCAARRD, titled "Establishment of the Center for Mollusc Research and De-

242 development: Development of Spawning and Hatchery Techniques for the Blood  
243 Cockle (*Anadara granosa*) for Sustainable Aquaculture." The ongoing research pri-  
244 marily involves the rearing of *T. granosa* from spat to larvae, feeding experiments,  
245 stocking density evaluations, substrate selection, and settlement rate assessments.

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

251 The researchers worked with 271 blood cockles that had been sex-identified  
252 and taken from Panay Island, specifically sourced from Zarraga Iloilo and Ivisan  
253 Capiz. These samples, divided between 144 males and 127 females, were obtained  
254 through induced spawning via temperature shock and dissection. Data collection  
255 was limited to the spawned stage among the five gonadal stages - immature, devel-  
256 oping, mature, spawning, and spent stages. The other stages were not preferable  
257 due to indistinguishable gonads and their inability to undergo induced spawning  
258 (May et al., 2021). Thus, the researchers only focused on the samples undergoing  
259 the spawned stage.

260 During the data collection, the researchers personally gathered linear mea-  
261 surements, including length, width, height, rib count, length of the hinge line,  
262 and distance between the umbos through the vernier caliper. The data-gathering  
263 process was supervised by the University Research Associates from the Institute  
264 of Aquaculture, College of Fisheries and Ocean Sciences. Aside from linear mea-  
265 surements, images were taken from six different angles. The image collection was  
266 non-invasive, considering the blood cockle-built ability to survive in low oxygen  
267 environments and naturally inhabit intertidal mudflats (Zhan & Bao, 2022).

268 The method developed in this study is specific to *Tegillarca granosa* and may  
269 not apply to other bivalve species. The model was trained exclusively for *Tegillarca*  
270 *granosa* and morphological features - such as length, width, height, rib count,  
271 length of the hinge line, and distance between the umbos may not be consistent  
272 and applicable across other shellfish species.

## 273 1.5 Significance of the Research

274 This study will give us a significant advancement in non-invasive sex identifica-  
275 tion methods in *T. granosa* providing innovative solutions that could solve the  
276 challenges in identifying sex and reshape sustainable approaches to aquaculture.

<sup>277</sup> The significance of this study extends to the following:

<sup>278</sup>        *Research Institution.* The result of this study focusing on the sex-identification  
<sup>279</sup> mechanism of bivalves, specifically *Tegillarca granosa*, will provide valuable in-  
<sup>280</sup> sights into universities and research centers that focus on fisheries and coastal  
<sup>281</sup> management, such as the UPV Institute of Aquaculture, that aim to develop  
<sup>282</sup> sustainable development and suitable culture techniques.

<sup>283</sup>        *Fishermen.* By developing a non-invasive method in sex identification, this  
<sup>284</sup> study can help long-term harvest efficiency and maintain the ratio of the harvest  
<sup>285</sup> which can help prevent exploitation of the *T. granosa*.

<sup>286</sup>        *Coastal Communities.* The result of this study would be beneficial for the  
<sup>287</sup> coastal communities that are reliant on their source of income with aquaculture  
<sup>288</sup> commodities like blood cockles. Maintaining the diversity and aspect ratio of  
<sup>289</sup> male and female may increase the market value of blood cockle production since  
<sup>290</sup> cockle aquaculture faces significant obstacles worldwide due to the fluctuating  
<sup>291</sup> seed supplies and scarcity of broodstock from the wild.

<sup>292</sup>        *Future Researchers.* The result of this study would serve as the basis for studies  
<sup>293</sup> that involve sex identification in bivalves such as *T. granosa*. Some technologies  
<sup>294</sup> are yet to be explored in machine learning, deep learning, and computer vision  
<sup>295</sup> technologies that can lead to higher accuracy and distinguish the presence of  
<sup>296</sup> sexual dimorphism in the *T. granosa*.

<sup>297</sup> **Chapter 2**

<sup>298</sup> **Review of Related Literature**

<sup>299</sup> Aquaculture is the fastest-growing industry in animal food production and has  
<sup>300</sup> great potential as a sustainable solution to global food security, nutrition, and  
<sup>301</sup> development (*FAO 2024 Report: Sustainable Aquatic Food Systems Important*  
<sup>302</sup> *for Global Food Security – European Fishmeal*, 2024). Aquaculture is deeply in-  
<sup>303</sup> tegrated into the livelihoods of Filipinos, not only through fish cultivation but  
<sup>304</sup> also through the production of other aquatic species, including mollusks, oysters,  
<sup>305</sup> clams, scallops, and mussels (Breton et al., 2017). Mollusks, particularly blood  
<sup>306</sup> clams *Tegillarca granosa*, have economic and environmental significance. It has  
<sup>307</sup> been a collective effort to maintain an ideal male-to-female ratio to avoid overhar-  
<sup>308</sup> vesting and maintain the optimal ratio to preserve the population and production  
<sup>309</sup> of the blood cockles.

<sup>310</sup> The members of the Arcidae Family, including *T. granosa* are important  
<sup>311</sup> sources of food and livelihood. Cockle aquaculture meets rising demands, however,  
<sup>312</sup> it faces significant challenges due to fluctuating seed supplies (Miranda & Ferriols,  
<sup>313</sup> 2023). To solve the problem, researchers exert a considerable amount of effort,  
<sup>314</sup> developing a broader understanding of bivalves, including their sex-determining  
<sup>315</sup> mechanism, due to their notable importance in terms of diversity, environmental  
<sup>316</sup> benefits, and economic and market importance (Breton et al., 2017). Despite the  
<sup>317</sup> promising idea of identifying sex, there is limited research reported in terms of  
<sup>318</sup> sexual dimorphism, making it harder to distinguish through its morphological and  
<sup>319</sup> morphometric characteristics.

<sup>320</sup> By addressing the challenges in the sex identification of *T. granosa*, it would be  
<sup>321</sup> able to address one problem at a time. Currently, there are no recent documented  
<sup>322</sup> publications that integrate machine learning and computer vision in characterizing  
<sup>323</sup> sexual dimorphism, reducing complexity, variability in sex determination, and

<sup>324</sup> differentiation mechanisms in bivalves, including *T. granosa* specifically.

## <sup>325</sup> **2.1 Background on *Tegillarca granosa* and Their 326 Importance**

<sup>327</sup> *Tegillarca granosa* (Linnaeus, 1758) is also known as blood cockles or blood clam.  
<sup>328</sup> In the Philippines, it is commonly known as a Litob, a marine bivalve species from  
<sup>329</sup> the family Arcidae. Litob is widely distributed in the world including Southeast  
<sup>330</sup> Asia. They can be found in the intertidal mudflats adjacent to the mangrove forest  
<sup>331</sup> (Srisunont, Nobpakhun, Yamalee, & Srisunont, 2020). With the intertidal mudflat  
<sup>332</sup> as *T. granosa*'s habitat, they experience severe hypoxia or low oxygen levels in the  
<sup>333</sup> blood tissues during the tidal cycle. The blood clams exhibit a unique red-blood  
<sup>334</sup> phenotype where it serves two purposes the hemocyte carries oxygen around the  
<sup>335</sup> body and strengthens immune defenses. In addition, it possesses a unique ability  
<sup>336</sup> to absorb oxygen at similar rates in water and air (Zhan & Bao, 2022).

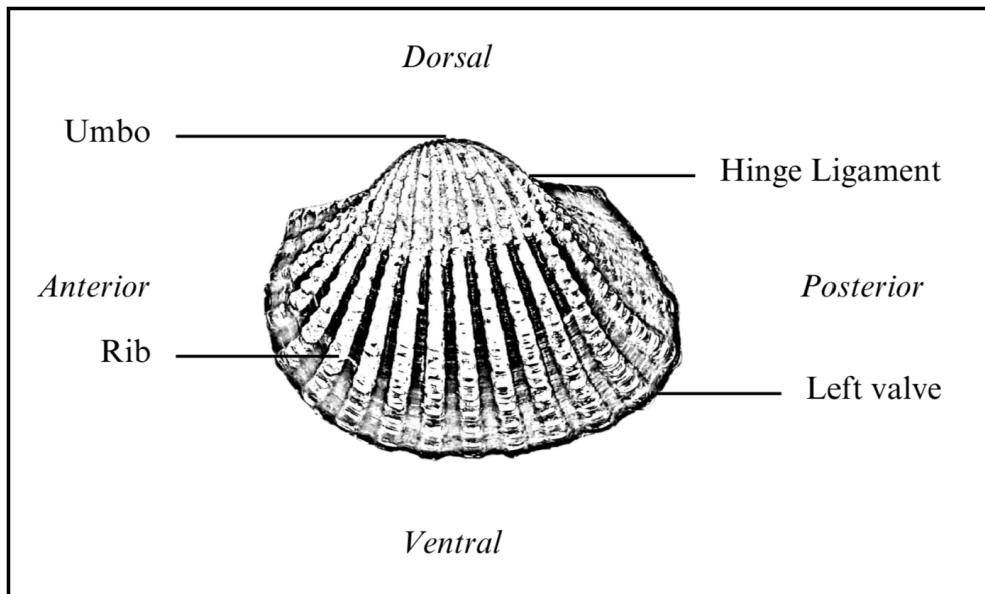


Figure 2.1: Diagram of *Tegillarca granosa* Anatomy

<sup>337</sup> *T. granosa* shell is medium-sized, fairly thick, ovate, and convex, with both  
<sup>338</sup> valves being equal in size but asymmetrical from the hinge. The top edge of  
<sup>339</sup> the dorsal margin is straight, while the front is rounded and slopes downward,  
<sup>340</sup> with its back being obliquely rounded with a concave bottom edge. It has a  
<sup>341</sup> narrow diamond-shaped ligament near the hinge with 3-4 dark chevron markings,  
<sup>342</sup> although some may be incomplete. The shell's outer layer, or the periostracum, is

343 smooth and brown with a straight hinge line and 40-68 fine short teeth arranged  
344 in a straight line. The beak, or prosogyrate, curves forward, with the shell having  
345 18–21 raised ribs with blunt nodules and spaces between them. The inner shell is  
346 white with crenulations along the valves' ventral, anterior, and posterior margins.  
347 The posterior adductor scar is elongated and squarish, while the anterior adductor  
348 scar is similar but smaller in size. The mantle covering the bulk of *T. granosa*'s  
349 visceral mass is thin but the edges are thick and muscular. It bears the impression  
350 of the crenulated shell edges. Their foot is large with a ventral grove with no byssus  
351 or thread-like attachment. The *T. granosa*'s soft body is blood red (Narasimham,  
352 1988).

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

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

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

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

## 390 **2.2 Current Methods of Sex Identification in *Tegillarca granosa***

391

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

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

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

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

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

431 On the other hand, invasive techniques such as histological analysis offer a  
432 more thorough but harmful method for determining the sex of *T. granosa*. A  
433 study on the Spawning Period of Blood Cockle *Tegillarca granosa* (Linnaeus,  
434 1758) in Myeik Coastal. 240 blood cockle samples were examined for sex and  
435 gonad maturity stages using histological examination, with shell lengths ranging  
436 from 26-35mm and shell weights from 8.1-33g. For histological analysis, the whole  
437 soft tissues were removed from the shell and the flesh containing most parts of  
438 the gonads was fixed in formalin, dehydrated in an upgraded series of ethanol,  
439 and cleared in xylene. This invasive method allows for precise identification of  
440 the gonadal maturation stages based on the cellular and structural changes in the  
441 gonads.

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

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

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

## 475 **2.3 Machine Learning and Deep Learning in Bi- 476 ological Studies**

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

487 **2.3.1 Deep Learning for Phenotype Classification in Ark**  
488 **Shells**

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

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

502 **2.3.2 Geometric Morphometrics and Machine Learning for**  
503 **Species Delimitation**

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

514 **2.3.3 Contour Analysis in Mollusc Shells Using Machine**  
515 **Learning**

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

519 evolutionary influences. The researchers stated that classic morphometric meth-  
520 ods may not accurately depict morphological features of the shell, especially when  
521 using the angular decomposition of the contour. The current research examined  
522 and compared the robustness of the contour analysis using wavelet transformed  
523 and Elliptic Fourier descriptors for gastropod shells with enlarged spines. For  
524 that, the researchers analyzed two geographically and ecologically separated pop-  
525 ulations of *Bolinus brandaris* from the NW Mediterranean Sea. Results showed  
526 that contour analysis of gastropod shells with enlarged spines can be analyzed  
527 using both methodologies, but the wavelet analysis provided better local discrim-  
528 ination. From an ecological perspective, shells with various sizes of spines in both  
529 areas indicate the broad adaptability of the species.

### 530 2.3.4 Machine Learning for Shape Analysis of Marine Or- 531 ganisms

532 In the study of Lishchenko and Jones (2021), titled *Application of Shape Analyses*  
533 *to Recording Structures of Marine Organisms for Stock Discrimination and Taxo-*  
534 *nomic Purposes*, they utilized geometric morphometrics (GM) as an approach to  
535 the traditional method of collecting linear measurements with the application of  
536 multivariate statistical methods and outline analysis in recording the structures  
537 of marine organisms. The main taxonomic categories (mollusks, teleost fish, and  
538 elasmobranchs) with their hard bodies have been used as an indication of age and  
539 a determinable time-scale and structure continue to go through life (Arkhipkin,  
540 2005; Kerr & Campana, 2014). This study has explored variations in the mor-  
541 phometry of recording structures in stock discrimination and systematics. The  
542 researchers utilized the principal component analysis rather than the traditional  
543 approach, which helps simplify the data without losing important information.  
544 They utilized landmark-based geometric morphometrics, which has three differ-  
545 ent types, namely: discrete juxtaposition of tissue, maxima or curvature, or other  
546 morphogenetic processes, and lastly, the extremal points are constructed land-  
547 marks.

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

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

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

### 578        2.3.5 Deep Learning for Landmark-Free Morphological Fea- 579        ture Extraction

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

### 590 2.3.6 Machine Learning for Sex Differentiation in Abalone

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

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

613 The study found that Naive Bayes consistently performed better than other al-  
614 gorithms. However, when applied to both min-max and zero-mean normalization,  
615 the average accuracies of the algorithms were as follows: Random Forest (62.37%),  
616 SVM with RBF kernel (59.49%), Decision Tree (57.20%), SVM with linear ker-  
617 nel (56.59%), and Naive Bayes (53.39%). Despite the performance decrease with  
618 normalization, Random Forest achieved the highest overall metrics, including an  
619 average balanced accuracy of 74.87%, sensitivity of 66.43%, and specificity of  
620 83.31%. Liu et al. concluded that Random Forest is highly accurate because it  
621 can handle large, complex datasets, run processes in parallel using multiple trees,  
622 and select the most relevant features to enhance model performance (Arifin et al.,  
623 2021).

624    **2.3.7 Machine Learning for Geographical Traceability in**  
625    **Bivalves**

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

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

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

652    **2.4 Limitations on Sex Identification in *Tegillarca***  
653    ***granosa***

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

658 perature shock, is the most natural and least invasive method for bivalves (Aji,  
659 2011). However, the method (Wong & Lim, 2018) of immersing cockles in water  
660 from hot to cold with a specific temperature requires deliberate and careful ma-  
661 nipulation of the temperature over a specific period and would require constant  
662 management and monitoring.

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

671 Hermaphrodites can exhibit either sequential (asynchronous) or simultaneous  
672 (synchronous or functional) characteristics. Sequential hermaphrodites switch  
673 genders after being male or female for one or multiple yearly cycles. (Heller,  
674 1993; Gosling, 2004; Collin, 2013). Sex change and consecutive hermaphroditism  
675 have been observed in different bivalve species, including Ostreidae, Pectinidae,  
676 Veneridae, and Patellidae. However, macroscopically differentiating bivalve sex is  
677 challenging. The only way it may be identified is through histological analysis of  
678 gonad remains but to do so there is an act of killing the organism (Coe, 1943;  
679 Gosling, 2004). Verification of sex change in bivalves to classify whether male or  
680 female while they are alive is challenging since they need to be re-confirmed and  
681 re-evaluated to be the same individual after a year.

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

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

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

## 701 **2.5 Synthesis of the Study**

702 This section of the paper summarizes the technologies used in the different studies  
703 related to the pursuit of the study entitled, Morphometric and Morphological-  
704 Based Non-Invasive Sex Identification of Blood Cockles *Tegillarca granosa* (Lin-  
705 naeus, 1758).

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

Table 2.1: Comparison of the Methods Used in Bivalves Studies

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

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

<sup>720</sup> **Chapter 3**

<sup>721</sup> **Research Methodology**

<sup>722</sup> This chapter discussed the materials and methods employed in the study, focusing  
<sup>723</sup> on the development requirements, as well as the software and programming  
<sup>724</sup> languages utilized. It also detailed the overall workflow in conducting the study,  
<sup>725</sup> Morphometric and Morphological-Based Non-Invasive Sex Identification of Blood  
<sup>726</sup> Cockles *Tegillarca granosa* (Linnaeus), 1758) using machine learning and deep  
<sup>727</sup> learning technologies.

<sup>728</sup> Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture,  
<sup>729</sup> oversaw the overall workflow and conduct of the experiment. The researchers were  
<sup>730</sup> also guided by research associates LC Mae Gasit and Allena Esther Artera. Con-  
<sup>731</sup> sequently, the entire dataset collection process was conducted at the University of  
<sup>732</sup> the Philippines Visayas hatchery facility.

<sup>733</sup> The methodology consisted of nine parts: (1) Sample Collection, (2) Ethical  
<sup>734</sup> Considerations, (3) Creating *T.granosa* Dataset, (4) Morphological Characteris-  
<sup>735</sup> tics Collection (5) Image Acquisition and Pre-processing, (6) Hardware and Soft-  
<sup>736</sup> ware Configuration,(7) Morphometric Characteristics Evaluation Using Machine  
<sup>737</sup> Learning, (8) Morphological Characteristics Evaluation Using Deep Learning, and  
<sup>738</sup> (9) Evaluation Metrics

<sup>739</sup> **3.1 Sample Collection**

<sup>740</sup> The collection of *T. granosa* samples used in this study was part of an ongoing  
<sup>741</sup> research project by UPV DOST-PCAARRD titled "Establishment of the Center  
<sup>742</sup> for Mollusc Research and Development: Development of Spawning and Hatchery

<sup>743</sup> Techniques for the Blood Cockle (*Anadara granosa*) for Sustainable Aquaculture.”  
<sup>744</sup> A total of 271 samples were provided for this study to classify the sex of *T. granosa*.  
<sup>745</sup> The samples, ranging in size from 34 to 61 mm, were sourced from the coastal area  
<sup>746</sup> of Zaraga, Iloilo, and fish markets in Ivisan, Capiz, Philippines (see Figure 3.1).

<sup>747</sup> The research and experimentation were conducted at the University of the  
<sup>748</sup> Philippines Visayas hatchery facility in Miagao, Iloilo, where the samples were  
<sup>749</sup> maintained in 200 L fiberglass-reinforced plastic (FRP) tanks containing filtered  
<sup>750</sup> seawater with 35 ppt salinity (Miranda & Ferriols, 2023).

<sup>751</sup> As part of the data collection process, the researchers utilized induced spawning  
<sup>752</sup> and dissection to classify the sex of the samples. Induced spawning through  
<sup>753</sup> temperature fluctuations was the most natural and least invasive method for bi-  
<sup>754</sup> valves compared to other approaches (Aji, 2011). However, since not all samples  
<sup>755</sup> exhibited gamete release, the researchers also performed dissections, assisted by  
<sup>756</sup> hatchery staff, to expedite data collection. The sex of the dissected samples was  
<sup>757</sup> identified based on the coloration of gonad tissue, which varies according to sex  
<sup>758</sup> and maturity stage. Females exhibited orange-red to pale orange gonads, while  
<sup>759</sup> males displayed white to grayish-white gonads (May et al., 2021).

<sup>760</sup> The methods used for data collection were considered noninvasive, particularly  
<sup>761</sup> given that *T. granosa* are oxygen regulators well adapted to tidal exposure and  
<sup>762</sup> hypoxia (Davenport & Wong, 1986).

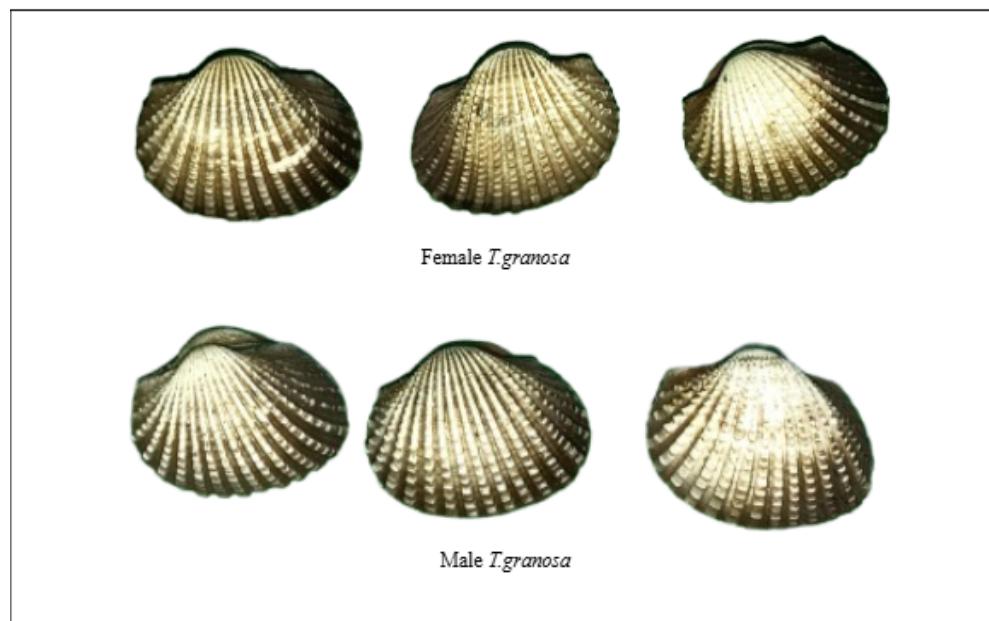


Figure 3.1: Male and Female *Tegillarca granosa* shells

## **763 3.2 Ethical Considerations**

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

## **770 3.3 Creating *T. granosa* Dataset**

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

**779** Images of each sample were captured and saved in JPG format using a stan-  
**780** dardized file naming convention that included the sample's sex, the shell's ori-  
**781** entation or view, and its corresponding number out of the 271 total samples. File  
**782** names for female *T. granosa* samples began with "0", while those for male sam-  
**783** ples began with "1". Each file name also included one of the six captured views:  
**784** (1) dorsal, (2) ventral, (3) anterior, (4) posterior, (5) left lateral, and (6) right  
**785** lateral (refer to Figure 3.2), followed by a unique sample number. For exam-  
**786** ple, "010001" denoted the first female sample taken from the dorsal view, while  
**787** "110001" represented the first male sample from the same view. This naming  
**788** convention was implemented to prevent data leakage and ensure accurate labeling  
**789** of images according to their respective samples.

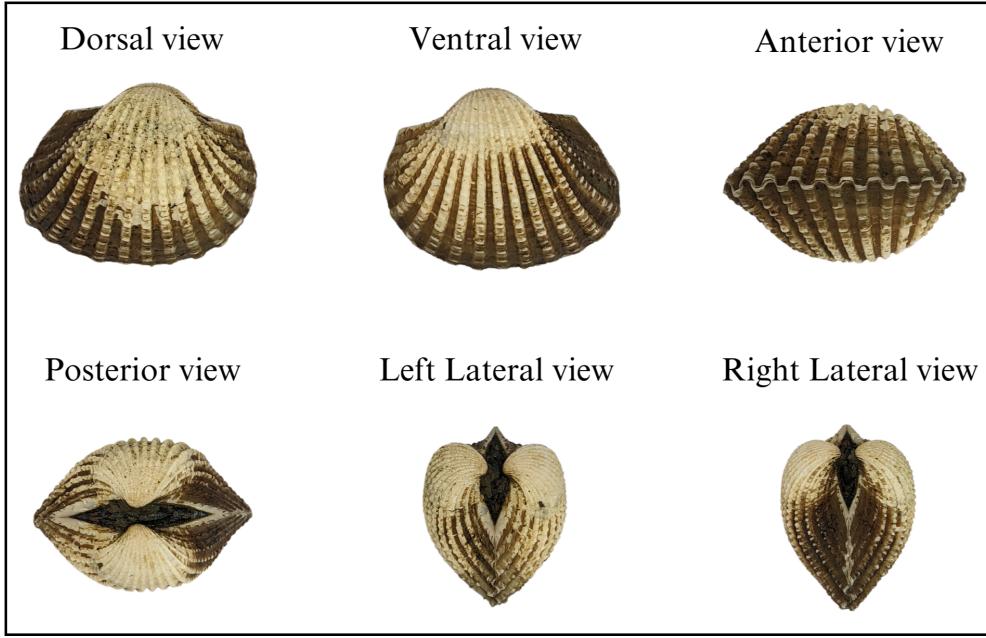


Figure 3.2: Different Views of the *T. granosa* Shell Captured

### <sup>790</sup> 3.4 Morphometric and Morphological Characteristics Collection

<sup>791</sup> Morphology refers to biological form and is one of the most visually recognizable phenotypes across all organisms (Tsutsumi, Saito, Koyabu, & Furusawa, 2023).  
<sup>792</sup> In this study, morphological characteristics describe the structural features of <sup>793</sup> *T. granosa*, focusing on measurable attributes such as shape, size, and color.  
<sup>794</sup> Morphometric characteristics, on the other hand, refer to specific quantifiable <sup>795</sup> features of *T. granosa*, including length, width, height, hinge line length, distance <sup>796</sup> between the umbos, and rib count. As stated by the researchers, quantifying and <sup>797</sup> characterizing these traits is essential for understanding and visualizing variations <sup>798</sup> in *T. granosa* morphology.

<sup>799</sup> The researchers measured the height, width, and length of *T. granosa* using <sup>800</sup> a Vernier caliper with a precision of up to 0.01 mm. Refer to Figure 3.3 for the <sup>801</sup> corresponding measurement diagram. Length (A) refers to the distance from the <sup>802</sup> anterior to the posterior of the shell. Width (B) is defined as the widest span <sup>803</sup> across the shell from the left to the right valve. Height (C) measures the distance <sup>804</sup> from the base to the apex of the shell. In addition, the hinge line length (D) near <sup>805</sup> the hinge and the distance between the umbos (E) were recorded.

<sup>806</sup> Reyment and Kennedy (1998) emphasized that including rib count as supple-

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

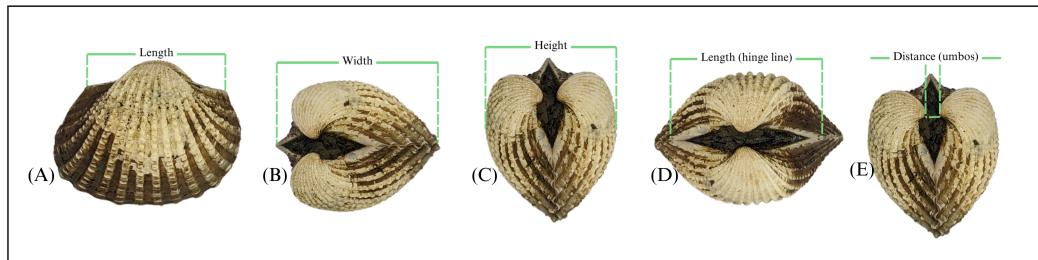


Figure 3.3: Linear Measurements of *Tegillarca granosa* shell.

### 813      3.5 Image Acquisition and Data Gathering

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

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



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

## 827    3.6    Hardware and Software Configuration

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

## 839    3.7    Morphometric Characteristics Evaluation Us- 840        ing Machine Learning

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

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

850 **3.7.1 Data Preprocessing**

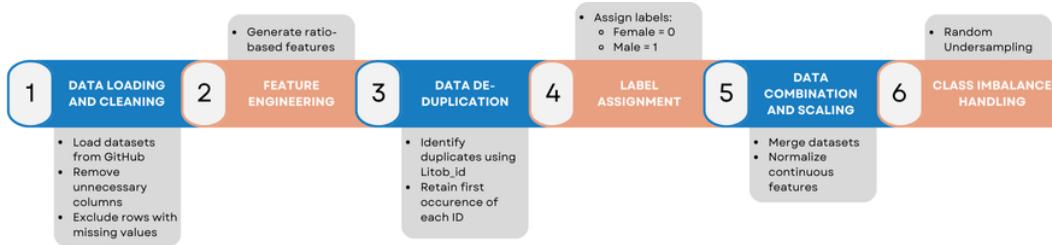


Figure 3.5: Data Preprocessing Pipeline

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

854 ***Data Loading and Cleaning***

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

859 ***Feature Engineering***

860 Additional ratio-based features were generated to augment the existing mea-  
861 surements. These included the length-to-width ratio, length-to-height ratio, width-  
862 to-height ratio, hinge line length-to-length ratio, umbos distance-to-length ratio,  
863 umbos distance-to-height ratio, and rib density. These derived features aimed to  
864 emphasize shape characteristics independent of size, improving the models' ability  
865 to distinguish morphological differences between sexes.

866 ***Data De-duplication***

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

871        ***Label Assignment***

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

875        ***Data Combination and Scaling***

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

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

885        ***Class Imbalance Handling***

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

891        **3.7.2 Machine Learning Models Training**

892        ***Model Selection and Hyperparameter Tuning***

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

898        ***Cross-Validation***

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

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

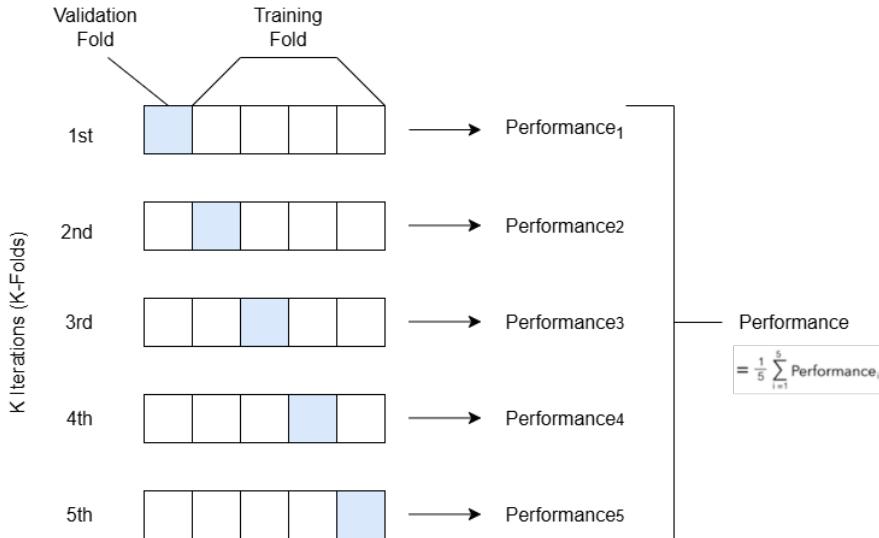


Figure 3.6: Diagram of k-fold cross-validation with  $k = 5$

## 904 3.8 Morphological Characteristics Evaluation Us- 905 ing Deep Learning

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

### 910 *Image Preprocessing*

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

### 917 *Adjusting Dimensions*

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

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

923 ***Background Removal***

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

929 ***Shadow Removal***

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

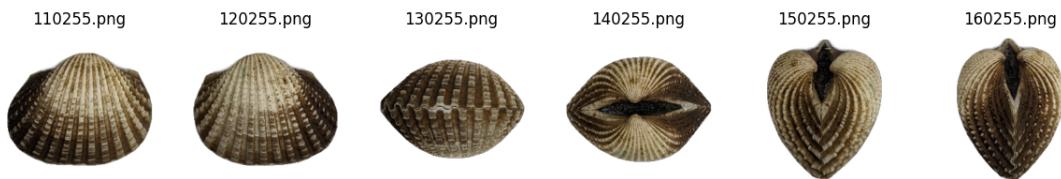


Figure 3.7: Shadows removed from male samples at different angles

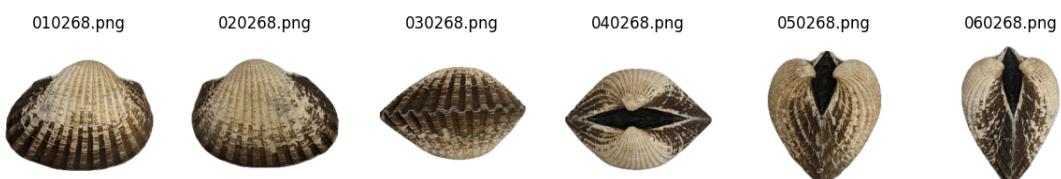


Figure 3.8: Shadows removed from female samples at different angles

934 **3.8.1 Convolutional Neural Network**

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

image with dimensions. The layers consist of four convolution layers, a flatten layer, dropout and two dense layers. The CNN framework used in this study is shown in Figure 3.9.

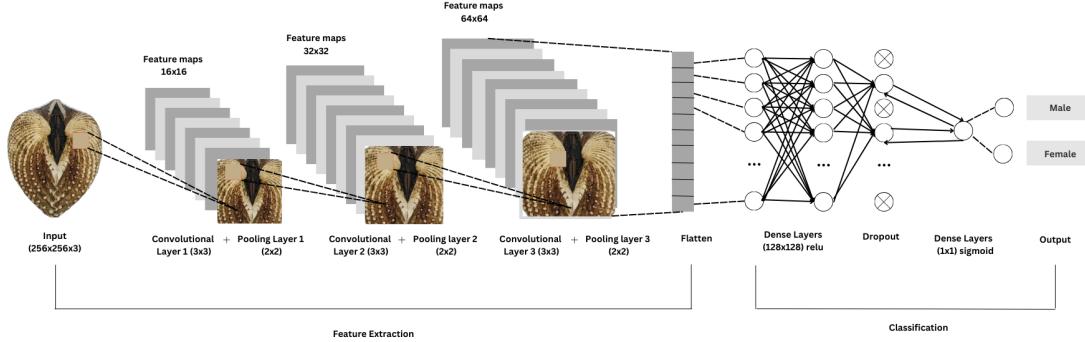


Figure 3.9: Architecture of Convolutional Neural Network (CNN)

### **Convolution Layer**

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

### **Pooling Layer**

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

### **Fully Connected and Dropout**

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

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

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

### 968 3.8.2 CNN Training

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

#### 974 *Data Splitting*

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

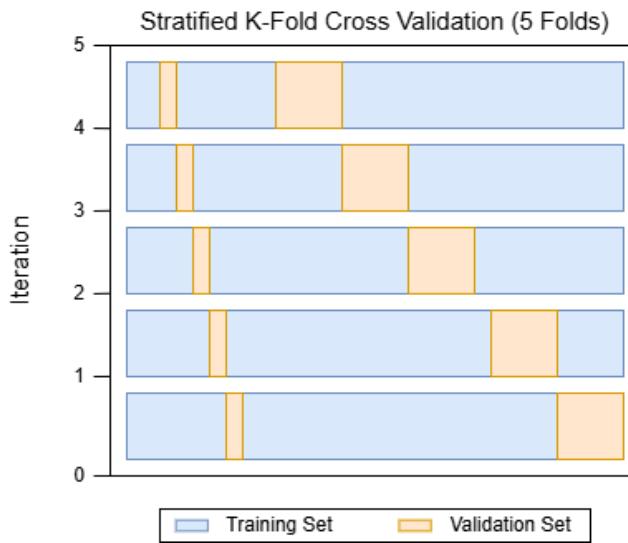


Figure 3.10: Diagram of stratified k-fold cross-validation with  $k=5$

#### 981 *Data Augmentation*

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

984 augmentations included random horizontal flips, slight rotations, and zoom trans-  
985 formations to enhance data diversity and improve model generalization (Awan,  
986 2022). All augmentation was strictly applied only to the training subset of each  
987 fold to prevent data leakage and maintain the validity of the results (*Figure 3.11*).

988 On-the-fly data augmentation (OnDAT) generates augmented data during  
989 each iteration, exposing the model to constantly changing data variations. Aug-  
990 menting the original data allows better exploration of the underlying data genera-  
991 tion process and has the potential to prevent the model from overfitting spurious  
992 patterns, thereby improving performance (Cerqueira, Santos, Baghoussi, & Soares,  
993 2024).

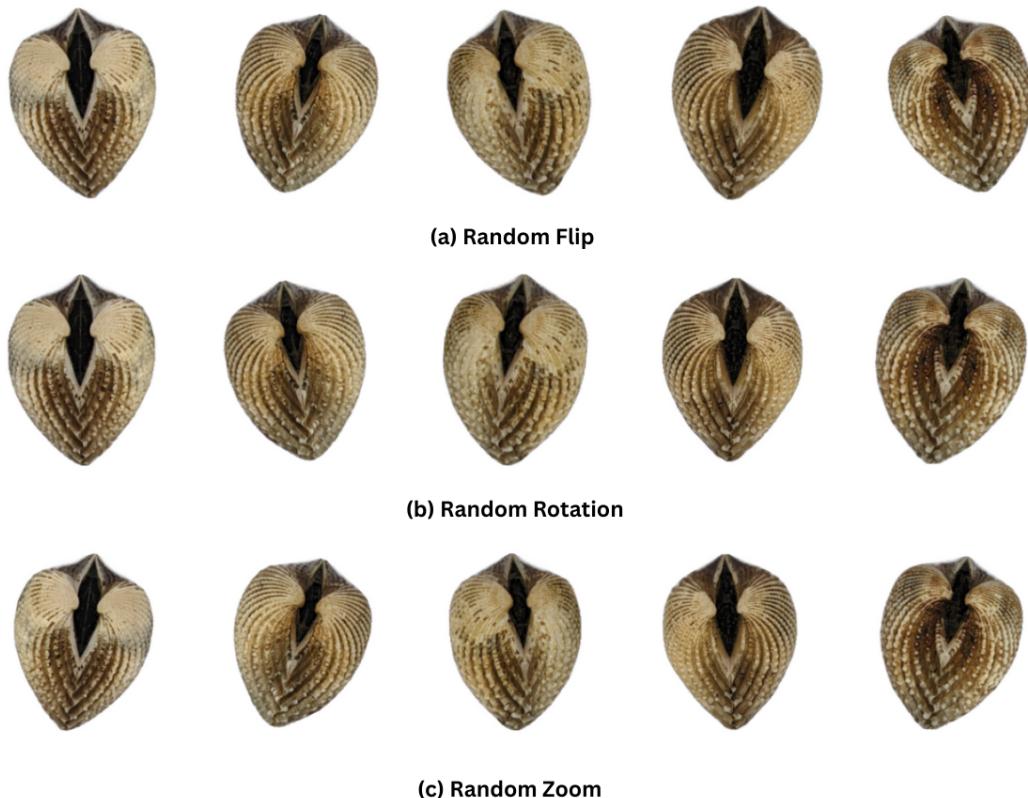


Figure 3.11: Data Augmentation Techniques

994 ***Training Procedure***

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

1000 the performance metrics, in this case, validation loss, cease to improve. Furthermore,  
1001 to enhance the learning process, `ReduceLROnPlateau` was applied, which  
1002 decreased the learning rate if there was no improvement in the model for a specified  
1003 number of epochs (Team, n.d.).

1004 The model was trained using the Adam optimization algorithm, with an initial  
1005 learning rate of 0.001. Binary cross-entropy, commonly known as the log loss,  
1006 was employed as the loss function due to its effectiveness in binary classification  
1007 tasks. To reduce the risk of overfitting, a dropout rate of 0.5 was applied, randomly  
1008 deactivating half of the neurons during the training process to improve  
1009 generalization.

## 1010 3.9 Evaluation Metrics

1011 Evaluating the performance of a binary classification model is essential, and selecting  
1012 appropriate metrics depends on the specific requirements of the user. The  
1013 performance of both supervised machine learning and deep learning models will  
1014 be measured using several key metrics, including accuracy, precision, recall, F1  
1015 score, and the AUC-ROC score.

1016 Accuracy (ACC) is the ratio of the overall correctly predicted samples to the  
1017 total number of examples in the evaluation dataset (Cui et al., 2020). It measures  
1018 the overall correctness of the model in predicting both male and female blood  
1019 cockles. This metric provides insight into how well the model performs across all  
1020 classifications. The formula for accuracy is:

$$1021 \text{ACC} = \frac{\text{Correctly classified samples}}{\text{All samples}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

1022 Precision (PREC) is the ratio of correctly predicted positive samples to all  
1023 samples assigned to the positive class (Cui et al., 2020). This metric helps in  
1024 evaluating the fairness of the model and prevents the misclassification of blood  
1025 cockles as it identifies potential inaccuracies or biases. The formula for precision  
is:

$$1026 \text{PREC} = \frac{\text{True positive samples}}{\text{Samples assigned to positive class}} = \frac{TP}{TP + FP} \quad (3.2)$$

1027 Recall (REC), also known as sensitivity or the true positive rate (TPR), is the

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

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

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

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

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

<sup>1039</sup> **Chapter 4**

<sup>1040</sup> **Results and Discussions**

<sup>1041</sup> This chapter presents the results from the machine learning and deep learning  
<sup>1042</sup> analyses conducted on the preprocessed dataset. It includes an evaluation of  
<sup>1043</sup> various machine learning classifiers and the application of deep learning models  
<sup>1044</sup> for image-based classification. The primary focus is on identifying key predictors  
<sup>1045</sup> and assessing classification performance for sex identification in *T. granosa*.

<sup>1046</sup> **4.1 Machine Learning Analysis**

<sup>1047</sup> This chapter outlines the results of preprocessing, training of machine learning  
<sup>1048</sup> models, and feature importance analysis, all conducted in Google Colab using  
<sup>1049</sup> Python. The dataset was preprocessed in Colab, and the training and evaluation  
<sup>1050</sup> of various classifiers were performed entirely within this environment. This part of  
<sup>1051</sup> the paper includes five subsections: data exploration, statistical analysis, feature  
<sup>1052</sup> importance analysis, performance evaluation, and confusion matrix analysis.

<sup>1053</sup> **4.1.1 Data Exploration**

<sup>1054</sup> Exploratory data analysis was performed to characterize the dataset using visu-  
<sup>1055</sup> alizations to understand the patterns and correlations within the data. A corre-  
<sup>1056</sup> lation heatmap was created to assess the relationship between the predictors and  
<sup>1057</sup> the target variable.

<sup>1058</sup> The heatmap (see Figure 4.1) revealed three features most correlated with the

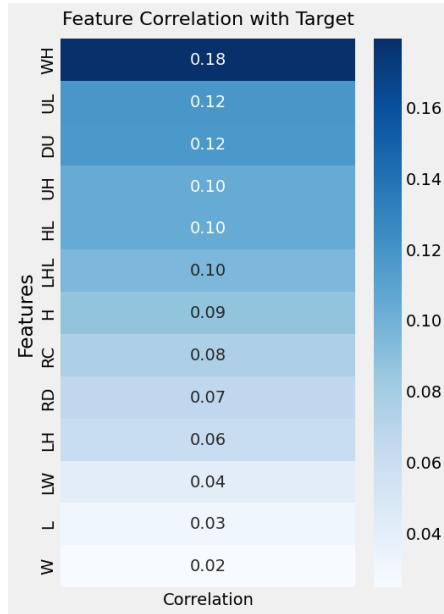


Figure 4.1: Correlation heatmap of morphometric features with the sex of *T. granosa*

1059 sex of *T. granosa*: the width-height ratio ( $r = 0.18$ ), the umbos-length ratio ( $r = 0.12$ ), and the distance between the umbos ( $r = 0.12$ ). Each of these features  
 1060 demonstrated a weak positive relationship with the target variable.  
 1061

#### 1062 4.1.2 Statistical Analysis

1063 As part of the exploratory data analysis, statistical testing confirmed that the  
 1064 dataset did not follow a normal distribution (see Table 4.1). Consequently, the  
 1065 Mann-Whitney U test was applied with a significance level of  $\alpha = 0.05$  to com-  
 1066 pare male and female samples. Out of thirteen features, five showed statistically  
 1067 significant differences. These included: distance between umbos ( $p = 0.025$ ),  
 1068 length-width ratio ( $p = 0.011$ ), umbos-length ratio ( $p = 0.019$ ), width-height  
 1069 ratio ( $p = 0.003$ ), and umbos-height ratio ( $p = 0.036$ ).

1070 It is important to note that statistical significance does not imply predictive  
 1071 importance. Therefore, further analysis, such as feature importance evaluation,  
 1072 was performed to identify the most informative predictors for classification.

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

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

#### <sup>1073</sup> 4.1.3 Feature Importance Analysis

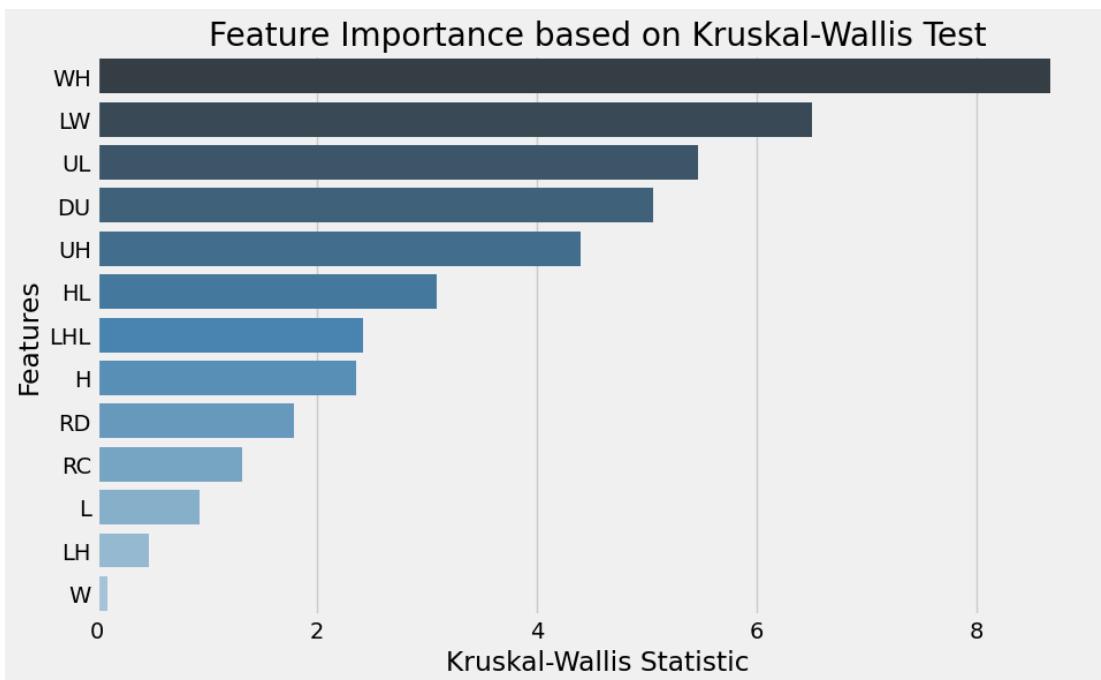


Figure 4.2: Feature Importance Scores Using the Kruskal-Wallis Test

<sup>1074</sup> Feature importance was assessed using the Kruskal-Wallis test, a non-parametric  
<sup>1075</sup> method that is suitable for evaluating differences in distributions across groups

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

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

#### 4.1.4 Performance Evaluation

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

Table 4.2: Performance Metrics for Models with All 13 Features

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

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

Table 4.3: Performance Metrics for Models with 5 Features

1095     Table 4.3 presents the performance of the same models using only the top  
1096     5 features identified through Kruskal-Wallis feature importance analysis. The  
1097     selected features are the distance between the umbos, length-to-width ratio, width-  
1098     to-height ratio, umbo distance-to-height ratio, and umbo distance-to-length ratio.

1099     Interestingly, the overall performance of the models improved when using only  
1100     the top 5 features compared to using all 13. K-Nearest Neighbors (KNN) achieved  
1101     the best results with an accuracy of 64.16%, precision of 64.97%, recall of 64.16%,  
1102     and an F1-score of 63.75%. Gradient Boosting followed closely behind. These find-  
1103     ings suggest that reducing the feature set to the most relevant variables helped  
1104     simplify the models, improved generalization, and enhanced predictive perfor-  
1105     mance—particularly for KNN, which showed a notable improvement over its ear-  
1106     lier results with the full feature set.

1107     **4.1.5 Confusion Matrix Analysis**

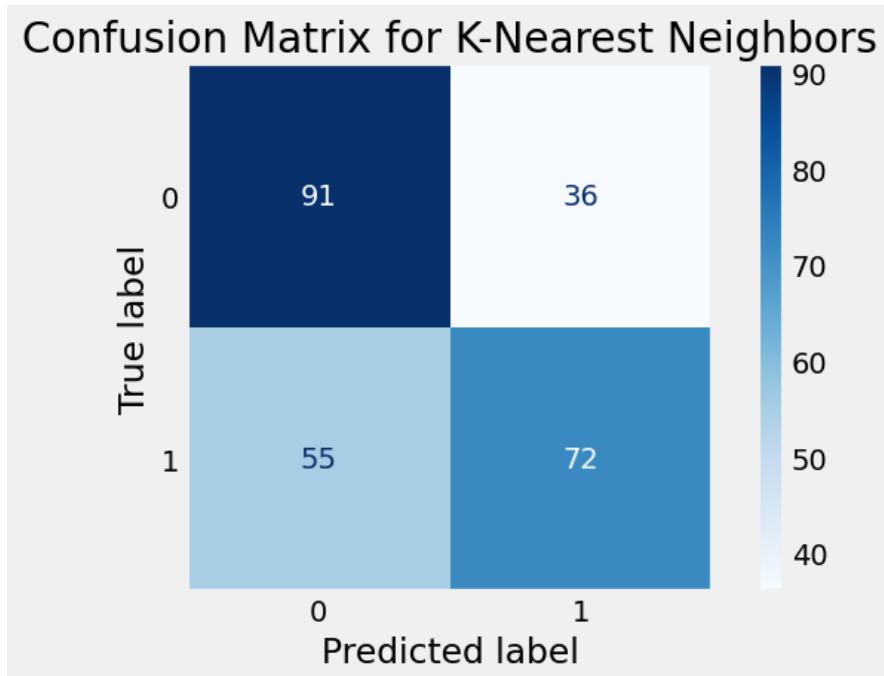


Figure 4.3: Feature Importance Scores Using the Kruskal-Wallis Test

1108     Figure 4.3 summarizes the performance of the K-Nearest Neighbors model in  
1109     classifying *T. granosa* based on their sex, where 0 represents female samples and  
1110     1 represents male samples. From the matrix, we observe that out of all the actual  
1111     female samples (true label 0), 91 were correctly predicted as female (true positive  
1112     for class 0), while 36 were incorrectly classified as male (false negative for class

1113 0). On the other hand, out of all the actual male samples (true label 1), 72 were  
1114 correctly predicted as male (true positive for class 1), while 55 were incorrectly  
1115 classified as female (false negative for class 1).

## 1116 4.2 Deep Learning Analysis

1117 This section presents the performance of the Convolutional Neural Network (CNN)  
1118 model in classifying the sex of *T. granosa* based on shell morphology. The analysis  
1119 evaluates the model's ability to distinguish between male and female shell images  
1120 using various evaluation metrics. This part of the paper includes six subsections:  
1121 baseline model, comparison of individual and combined angles, training result and  
1122 hyperparameter tuning, proposed model, learning rates and training behavior per  
1123 fold, and visualizations.

1124 The machine learning analysis (see Figure 4.3) revealed that five of the origi-  
1125 nal features produced significant results. The K-Nearest Neighbor (KNN) model  
1126 achieved an accuracy of 64.16%, precision of 64.97%, recall of 64.16%, and an F1  
1127 score of 63.75%. This section compares the model's performance across differ-  
1128 ent angles based on the results of the machine learning and feature importance  
1129 analysis.

### 1130 4.2.1 Baseline Model

1131 This section presents the baseline model with a batch size of 16 and 20 epochs,  
1132 which will serve as the starting point for comparison and provide a guideline for  
1133 hyperparameter tuning. The focus will be on one of the angles, specifically the  
1134 Left Lateral view, since the feature importance analysis using the Kruskal-Wallis  
1135 Test indicated that the width-to-height ratio had the highest importance score,  
1136 which is most visible from the Left Lateral view.

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

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

1137 The unbalanced dataset, which consisted of 144 male samples and 127 female  
1138 samples, achieved an accuracy of 65.27%, precision of 71.82%, recall of 58.99%,

1139 an F1-score of 63.99%, an AUC score of 73.08%, and a loss of 0.6122. However, to  
1140 address the class imbalance and enhance model performance, random undersam-  
1141 pling was performed. This approach resulted in improved performance metrics for  
1142 the balanced dataset, with an accuracy of 67.34%, precision of 69.43%, a recall  
1143 of 64.06%, an F1-score of 65.60%, an AUC score of 74.31%, and a lower loss of  
1144 0.5981.

#### 1145 4.2.2 Comparison of Individual and Combined Angles

1146 Using the same batch size and number of epochs, performance was compared  
1147 across all individual angles and the combination of the two highest-performing  
1148 angles based on accuracy, using a balanced dataset. For the combined analysis,  
1149 samples from the two selected angles were placed side by side, and a new dataset  
1150 folder was created for male and female samples.

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

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

1151 Table 4.5 presents the performance metrics for each individual angle and the  
1152 combination of the two highest-performing angles in terms of accuracy. The  
1153 Left Lateral view achieved the highest accuracy (67.34%) and precision (69.43%),  
1154 while the Dorsal view obtained the highest recall (77.88%) and F1-score (69.96%).  
1155 Meanwhile, the Ventral view recorded the highest AUC score (74.87%), indicat-  
1156 ing its strong ability to distinguish between classes. Combining the Ventral and  
1157 Left Lateral views resulted in an overall accuracy of 62.60%, suggesting that while  
1158 combined images may provide complementary information, individual angle views  
1159 still outperformed the combined views under the current experimental setup.

#### 1160 4.2.3 Training Result and Hyperparameter Tuning

1161 The Left Lateral angle was selected for further optimization. Several experiments  
1162 were conducted by tuning hyperparameters such as batch size, number of epochs,

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

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

## A. Batch Size and Number of Epochs

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

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

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

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

## B. Activation Functions

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

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

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

1183 ReLU activation function achieved the best overall performance, with an accuracy  
1184 of 71.68%, precision of 72.52%, recall of 69.29%, F1-score of 69.12%, and  
1185 AUC score of 77.34%, along with the lowest loss at 0.5824. This suggests that  
1186 ReLU remains an effective activation function for the classification of *T. granosa*,  
1187 outperforming both ELU and PReLU in this setup.

#### 1188 4.2.4 Proposed Model

1189 This section presents the performance evaluation of the proposed Convolutional  
1190 Neural Network (CNN) model, trained with a batch size of 32, 50 epochs, and using  
1191 the ReLU activation function. The model's effectiveness was assessed through  
1192 5-fold cross-validation to ensure robustness and generalizability across different  
1193 data partitions.

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

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

1194 The proposed model consistently achieved high performance in Folds 1, 3, and  
1195 5, with accuracies above 76% and strong recall and AUC scores, demonstrating  
1196 its potential for reliable sex identification of *T. granosa*. The slight variation  
1197 in performance across folds may be attributed to differences in data distribution,  
1198 emphasizing the importance of further data augmentation and balancing for future  
1199 work.

#### 1200 4.2.5 Learning Rates and Training Behavior per Fold

1201 This section presents the learning rate adjustments, early stopping events, and  
1202 best epoch selections for each fold during the 5-fold cross-validation of the pro-  
1203 posed model. During training, the ReduceLROnPlateau callback was employed  
1204 to monitor the validation loss and automatically reduce the learning rate when  
1205 performance plateaued. Additionally, EarlyStopping was utilized to halt training  
1206 once no further improvement was observed after a set patience, and the model  
1207 weights were restored from the end of the best-performing epoch to ensure optimal  
1208 performance.

1209        The following table summarizes the epochs where learning rate reductions  
 1210    occurred, the adjusted learning rates, the epochs at which early stopping took  
 1211    place, and the best epochs from which model weights were restored for each fold.

Fold no.	Epoch (LR Reduced)	Learning Rate After Reduction	Early Stopping Epoch	Best Epoch (Restored)
Fold 1	20	0.0005000	25	17
	23	0.0002500		
Fold 2	9	0.0005000	19	11
	14	0.0002500		
	17	0.0001250		
Fold 3	15	0.0005000	20	12
	18	0.0002500		
Fold 4	12	0.0005000	32	24
	15	0.0002500		
	27	0.0001250		
	30	0.0000625		
Fold 5	20	0.0005000	25	17
	23	0.0002500		

Table 4.9: Learning Rate Reductions, Early Stopping, and Best Epochs per Fold During 5-Fold Cross-Validation

#### 1212    4.2.6    Visualizations

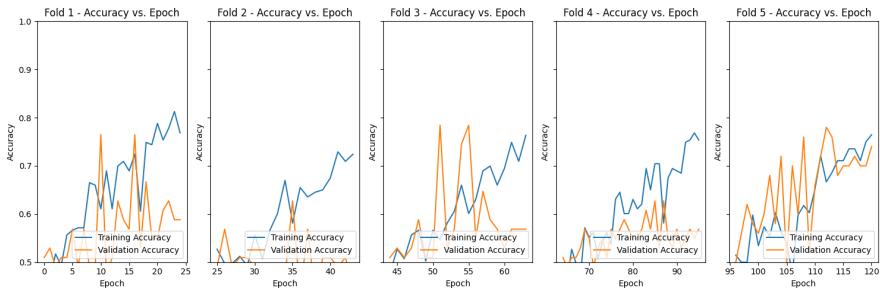


Figure 4.4: Training and Validation Accuracy per Fold

1213        Figure 4.4 shows the performance of the model in the training and validation  
 1214    in terms of accuracy across five folds. The graph across folds displays a consistent  
 1215    upward trend for the training accuracy. However, there is an observable change in  
 1216    the performance, particularly in Folds 1 and 2, where it shows a slight downward  
 1217    trend in the validation accuracy.

1218        Figure 4.5 shows the average performance of the model in both training and

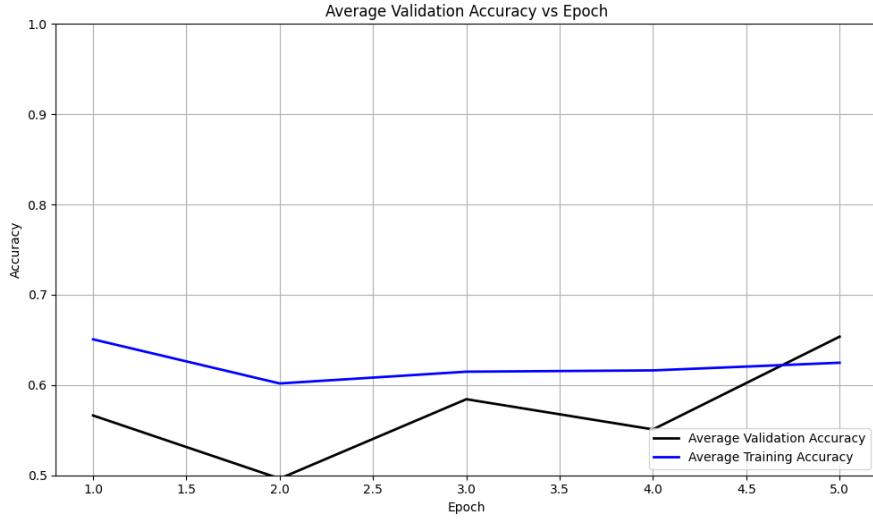


Figure 4.5: Average Training and Validation Accuracy Across Folds

accuracy in terms of accuracy across five folds. Similar to the individual performances, there is an observable upward trend, which shows that the accuracy score improves with the number of folds. The validation accuracy shows a downward and upward trend that shows that it gradually improves on later epochs. The accuracy in the training is slightly higher than the accuracy when validating the model, it indicates that the model learns during training.

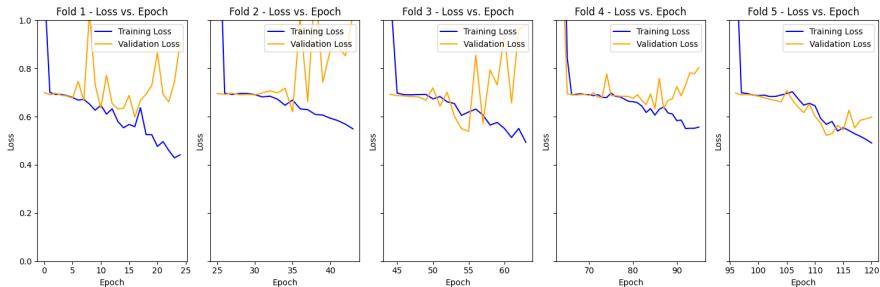


Figure 4.6: Training and Validation Loss per Fold

Figure 4.6 shows the performance of the model in the training and validation in terms of the training and validation loss across five folds. The graph across folds displays a consistent downward trend for the training loss. On the other hand, there is an observable change in the performance, especially in Folds 1,2,3, and 4, where it shows an upward trend in the validation loss. This is an implication for the learning performance of the model, as it may not be learning effectively.

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

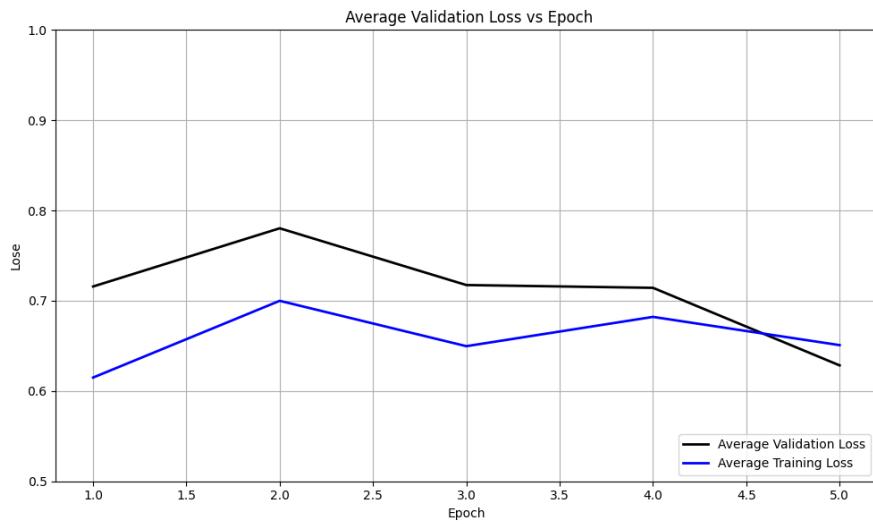


Figure 4.7: Average Training and Validation Loss Across Folds

in both the average loss for training and validation. Additionally, the average training loss is slightly lower than the average validation loss.

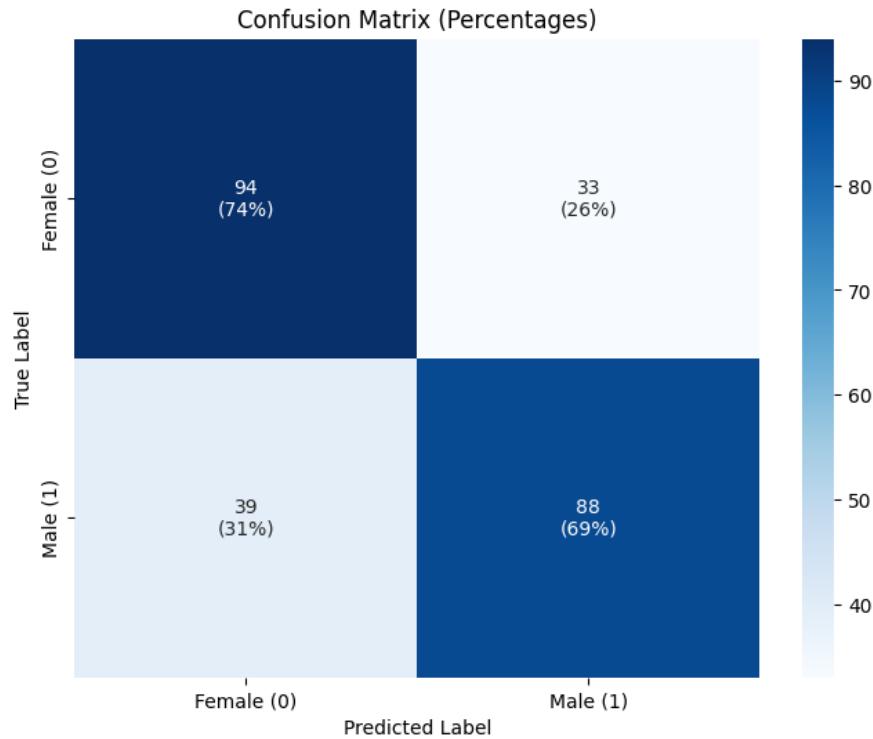


Figure 4.8: Confusion Matrix for Final Model Predictions

1235       Figure 4.8 shows the confusion matrix for the true class label and predicted  
 1236       class label. The matrix shows the correctly predicted male and female samples  
 1237       along with their corresponding percentages. There is an observable trend where  
 1238       females have slightly higher true positives compared to males in the number and  
 1239       percentages for the correctly classified male and female samples, which are 94 and  
 1240       88, corresponding to 74% and 69%, respectively. Additionally, the false classified  
 1241       samples were 33 for females and 39 for males, respectively accounting for 26% and  
 1242       31%.

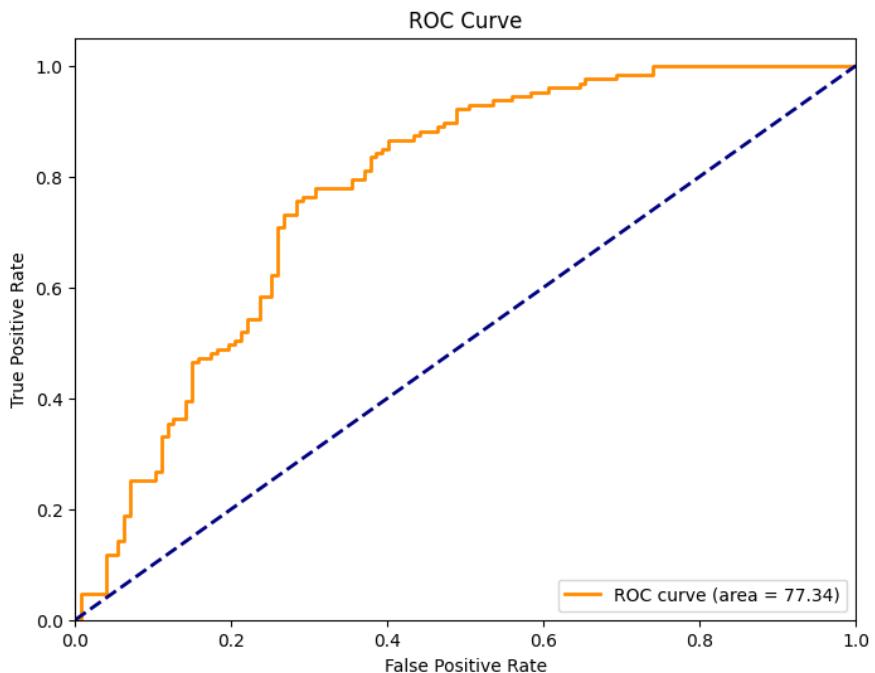


Figure 4.9: ROC Curve and AUC Score

1243       Figure 4.9 shows the ROC Curve shows the ability of the proposed model to  
 1244       correctly identify the true positives, which can help determine the tradeoff between  
 1245       specificity and sensitivity. It will also determine the validity of the model, that it is  
 1246       not predicting based only on random chances. The range of AUC ROC is between  
 1247       0.5 and 1. The model was able to achieve a score of 0.7734, which is better than  
 1248       random chances and an indication that the model is performing reasonably.

### 1249       4.3 Discussions

1250       This study aimed to develop a non-invasive method for identifying the sex of *T.*  
 1251       *granosa* using machine learning, deep learning, and computer vision technologies.

1252 The dataset was manually curated by the researchers, including both the linear  
1253 measurements and the images captured from six different angles.

1254 The machine learning approach revealed that using five key features, selected  
1255 through statistical tests (Mann-Whitney U-test and Kruskal-Wallis test), out-  
1256 performed models trained on all 13 features. The K-nearest neighbors (KNN)  
1257 classifier, using only these five features, achieved an accuracy of 64.16%, precision  
1258 of 64.97%, recall of 64.16%, and an F1-score of 63.57%. These results indicate  
1259 that a more focused set of features can enhance model performance, confirming  
1260 the potential of non-invasive sex identification using linear measurements.

1261 Further deep learning experiments explored how different image angles im-  
1262 pacted performance. The study found that the Left Lateral view consistently  
1263 produced the best results, with an accuracy of 71.68%, precision of 72.52%, recall  
1264 of 69.29%, F1-score of 69.12%, and an AUC score of 77.34%. This suggests that  
1265 optimizing image angles is crucial, and combining multiple angles did not signif-  
1266 icantly improve the model's performance. Data augmentation and regularization  
1267 techniques, such as early stopping, helped improve the model's generalization and  
1268 prevent overfitting.

1269 The findings are significant because they demonstrate the feasibility of a non-  
1270 invasive, accurate, and efficient sex identification method for *T. granosa*. This  
1271 approach aligns with sustainable aquaculture practices by reducing the need for  
1272 harmful physical sex-identifying methods. By integrating machine learning with  
1273 deep learning image analysis, this study provides a valuable model for non-invasive  
1274 sex identification which could be applied to other species in aquaculture as well.

1275 When compared to similar existing studies such as the gender classification  
1276 method for Chinese mitten crab using deep learning CNN (Cui *et al.*, 2020), there  
1277 are notable differences in methodology. The crab study used grayscale images and  
1278 a CNN with three convolutional layers, achieving 98.90% accuracy. In contrast,  
1279 this study utilized a hybrid approach combining machine learning with deep learn-  
1280 ing CNNs, trained on RGB images ( $256 \times 256$ ), and a deeper CNN architecture.  
1281 Despite achieving lower accuracy (71.68%), this variation could be due to the sub-  
1282 tler morphological differences between male and female *T. granosa*, or possibly  
1283 due to image quality limitations and sample size.

1284 There are limitations in this study, particularly the size of the dataset (271  
1285 samples) and the reliance on six fixed image angles. These constraints may not  
1286 fully represent the morphological variability across different populations or en-  
1287 vironments. Despite these limitations, the study successfully demonstrates that  
1288 combining machine learning and deep learning with computer vision can provide  
1289 a reliable and non-invasive solution for sex identification in *T. granosa*.

<sub>1290</sub> **Chapter 5**

<sub>1291</sub> **Conclusion and  
<sub>1292</sub> Recommendations**

<sub>1293</sub> **5.1 Conclusion**

<sub>1294</sub> This study utilized the application of machine learning and deep learning tech-  
<sub>1295</sub> niques to identify the sex of *T. granosa* based on the morphometric characteristics.  
<sub>1296</sub> A manually curated dataset was developed, consisting of both linear measurements  
<sub>1297</sub> and images captured from six different angles. Machine learning methods were  
<sub>1298</sub> employed to identify statistically significant features, which served as the basis for  
<sub>1299</sub> deep learning analysis using a 12-layer Convolutional Neural Network (CNN). The  
<sub>1300</sub> proposed CNN model yielded an average accuracy of 71.68% in the performance  
<sub>1301</sub> metrics. Overall, this study offers a classification approach which is a viable so-  
<sub>1302</sub> lution for non-invasive sex identification, providing an in-depth analysis based on  
<sub>1303</sub> *T. granosa*'s linear measurements and morphological characteristics from different  
<sub>1304</sub> angles.

<sub>1305</sub> Through the availability of the gathered data, trial-and-error experimentation  
<sub>1306</sub> was conducted by adjusting the number of layers, batch size, epoch, and activa-  
<sub>1307</sub> tion functions. The different combinations tested provided baseline results that  
<sub>1308</sub> demonstrate the feasibility of non-invasive sex identification for *T. granosa*.

<sub>1309</sub> While the study has made significant progress, challenges were encountered  
<sub>1310</sub> during CNN training, particularly due to hardware memory limitations. To over-  
<sub>1311</sub> come these, the researchers utilized synchronous Google Colab with 100 comput-  
<sub>1312</sub> ing units, requiring subscriptions, repeated retraining, and reconfigurations, which  
<sub>1313</sub> demanded considerable financial resources and time to optimize the parameters.

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

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

## 5.2 Recommendations

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

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

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

<sub>1349</sub> dataset can be utilized for further analysis through advanced deep learning and  
<sub>1350</sub> computer vision methods to make sense of the images gathered and discern sexual  
<sub>1351</sub> dimorphism for *T. granosa* .

# <sup>1352</sup> References

- 1353 Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten  
1354 years of progress following the ‘revolution’. *Italian Journal of Zoology*, *71*,  
1355 5–16. doi: 10.1080/11250000409356545
- 1356 Afiati, N. (2007, 01). Gonad maturation of two intertidal blood clams anadara  
1357 granosa (l.) and anadara antiquata (l.) (bivalvia: Arcidae) in central java. ,  
1358 *10*.
- 1359 Aji, L. P. (2011). Review: Spawning induction in bivalve. *Jurnal Penelitian  
1360 Sains*, *14*, 14207.
- 1361 Arifin, W. A., Ariawan, I., Rosalia, A. A., Lukman, L., & Tufailah, N. (2021).  
1362 Data scaling performance on various machine learning algorithms to identify  
1363 abalone sex. *Jurnal Teknologi Dan Sistem Komputer*, *10*(1), 26–31. doi:  
1364 10.14710/jtsiskom.2021.14105
- 1365 Arkhipkin, A. I. (2005). Statoliths as ‘black boxes’ (life recorders) in squid.  
1366 *Marine and Freshwater Research*, *56*, 573–583. doi: 10.1071/mf04158
- 1367 Awan, A. A. (2022, November). *A complete guide to data augmentation*. Retrieved from <https://www.datacamp.com/tutorial/complete-guide-data-augmentation>
- 1370 Aypa, S. M., & Baconguis, S. R. (2000). Philippines: mangrove-friendly aquacul-  
1371 ture. In J. H. Primavera, L. M. B. Garcia, M. T. Castaños, & M. B. Sur-  
1372 tida (Eds.), *Mangrove-friendly aquaculture: Proceedings of the workshop on  
1373 mangrove-friendly aquaculture organized by the seafdec aquaculture depart-  
1374 ment, january 11-15, 1999, iloilo city, philippines* (pp. 41–56).
- 1375 BFAR. (2019). *Philippine fisheries profile 2018* (Tech. Rep.). PCA Compound,  
1376 Elliptical Road, Quezon City, Philippines: Bureau of Fisheries and Aquatic  
1377 Resources.
- 1378 Boey, P.-L., Maniam, G. P., Hamid, S. A., & Ali, D. M. H. (2011). Utilization of  
1379 waste cockle shell (anadara granosa) in biodiesel production from palm olein:  
1380 Optimization using response surface methodology. *Fuel*, *90*(7), 2353–2358.  
1381 doi: 10.1016/j.fuel.2011.03.002
- 1382 Breton, S., Capt, C., Guerra, D., & Stewart, D. (2017, June). *Sex determining  
1383 mechanisms in bivalves*. Preprints.org. doi: 10.20944/preprints201706.0127

- 1384 .v1
- 1385 Breton, S., Stewart, D. T., Shepardson, S., Trdan, R. J., Bogan, A. E., Chapman,  
1386 E. G., ... Hoeh, W. R. (2010). Novel protein genes in animal mtDNA: A  
1387 new sex determination system in freshwater mussels (bivalvia: Unionoida)?  
1388 *Molecular Biology and Evolution*, 28(5), 1645–1659. doi: 10.1093/molbev/  
1389 msq345
- 1390 Budd, A., Banh, Q., Domingos, J., & Jerry, D. (2015). Sex control in fish: Ap-  
1391 proaches, challenges and opportunities for aquaculture. *Journal of Marine  
1392 Science and Engineering*, 3(2), 329–355. doi: 10.3390/jmse3020329
- 1393 Burdon, D., Callaway, R., Elliott, M., Smith, T., & Wither, A. (2014, 04).  
1394 Mass mortalities in bivalve populations: A review of the edible cockle  
1395 *Cerastoderma edule* (l.). *Estuarine, Coastal and Shelf Science*, 150. doi:  
1396 10.1016/j.ecss.2014.04.011
- 1397 Campos, A., Tedesco, S., Vasconcelos, V., & Cristobal, S. (2012). Proteomic  
1398 research in bivalves: Towards the identification of molecular markers of  
1399 aquatic pollution. *Proteomic Research in Bivalves*, 75(14), 4346–4359. doi:  
1400 10.1016/j.jprot.2012.04.027
- 1401 Cerqueira, V., Santos, M., Baghoussi, Y., & Soares, C. (2024). On-the-fly data  
1402 augmentation for forecasting with deep learning. *ArXiv (Cornell Univer-  
1403 sity)*. Retrieved from <https://doi.org/10.48550/arxiv.2404.16918>
- 1404 Coe, W. R. (1943). Sexual differentiation in mollusks. i. pelecypods. *The Quarterly  
1405 Review of Biology*, 18, 154–164. doi: 10.1086/qrb.1943.18.issue-2
- 1406 Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan  
1407 sexual systems. *Integrative and Comparative Biology*, 53, 723–735. doi:  
1408 10.1093/icb/ict076
- 1409 Concepcion, R., Guillermo, M., Tanner, S. E., Fonseca, V., & Duarte, B. (2023).  
1410 Bivalvenet: A hybrid deep neural network for common cockle (*cerastoderma  
1411 edule*) geographical traceability based on shell image analysis. *Ecological  
1412 Informatics*, 78, 102344. doi: 10.1016/j.ecoinf.2023.102344
- 1413 Cui, Y., Pan, T., Chen, S., & Zou, X. (2020). A gender classification method  
1414 for chinese mitten crab using deep convolutional neural network. *Multi-  
1415 media Tools and Applications*, 79(11-12), 7669–7684. doi: <https://doi.org/10.1007/s11042-019-08355-w>
- 1417 Davenport, J., & Wong, T. (1986, September). Responses of the blood cockle  
1418 *Anadara granosa* (l.) (bivalvia: Arcidae) to salinity, hypoxia and aerial ex-  
1419 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
- 1421 Doering, P., & Ludwig, J. (1990). Shape analysis of otoliths—a tool for indirect  
1422 ageing of eel, *anguilla anguilla* (l.)? *International Review of Hydrobiology*,  
1423 75(6), 737–743. doi: 10.1002/iroh.19900750607
- 1424 Erica, D. (2018, April 4). *Clam dissection: A first step into dissection and  
1425 anatomy for young learners*. Rosie Research. Retrieved from <https://>

- 1426                    [rosiereresearch.com/clam-dissection-anatomy/](http://rosiereresearch.com/clam-dissection-anatomy/)
- 1427     Fao 2024 report: Sustainable aquatic food systems important for global food  
1428        security – european fishmeal. (2024). <https://effop.org/news-events/fao-2024-report-sustainable-aquatic-food-systems-important-for-global-food-security/>.
- 1431     Ferguson, G. J., Ward, T. M., & Gillanders, B. M. (2011). Otolith shape and  
1432        elemental composition: Complementary tools for stock discrimination of  
1433        mulloway (*argyrosomus japonicus*) in southern australia. *Fish Research*,  
1434        110, 75–83. doi: 10.1016/j.fishres.2011.03.014
- 1435     GeeksforGeeks. (2020, August 6). *Stratified k fold cross validation*. Retrieved from <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
- 1438     GeeksforGeeks. (2024, May). *Cross-validation using k-fold with scikit-learn*. Retrieved from <https://www.geeksforgeeks.org/cross-validation-using-k-fold-with-scikit-learn/> (Accessed: 2025-04-23)
- 1441     Gosling, E. (2004). *Bivalve molluscs: biology, ecology and culture*. Oxford: Blackwell Science.
- 1443     Heller, J. (1993). Hermaphroditism in molluscs. *Biological Journal of the Linnean Society*, 48, 19–42. doi: 10.1111/bij.1993.48.issue-1
- 1445     Hussain, S. S., & Zaidi, S. S. H. (2024). Adaboost ensemble approach with  
1446        weak classifiers for gear fault diagnosis and prognosis in dc motors. *Applied Sciences*, 14(7), 3105. doi: 10.3390/app14073105
- 1448     Ishak, A. R., Mohamad, S., Soo, T. K., & Hamid, F. S. (2016). Leachate and  
1449        surface water characterization and heavy metal health risk on cockles in  
1450        kuala selangor. In *Procedia - social and behavioral sciences* (Vol. 222, pp.  
1451        263–271). doi: 10.1016/j.sbspro.2016.05.156
- 1452     Jaiswal, S. (2024, January 4). *What is normalization in machine learning? a comprehensive guide to data rescaling*. DataCamp. Retrieved from <https://www.datacamp.com/tutorial/normalization-in-machine-learning> (Accessed 2025-04-23)
- 1456     Karapunar, B., Werner, W., Fürsich, F. T., & Nützel, A. (2021). The earliest  
1457        example of sexual dimorphism in bivalves—evidence from the astartid  
1458        *Nicanella* (lower jurassic, southern germany). *Journal of Paleontology*,  
1459        95(6), 1216–1225. doi: 10.1017/jpa.2021.48
- 1460     Kerr, L. A., & Campana, S. E. (2014). Chemical composition of fish hard parts  
1461        as a natural marker of fish stocks. In *Stock identification methods* (pp. 205–  
1462        234). Elsevier. doi: 10.1016/b978-0-12-397003-9.00011-4
- 1463     Kim, E., Yang, S.-M., Cha, J.-E., Jung, D.-H., & Kim, H.-Y. (2024). Deep  
1464        learning-based phenotype classification of three ark shells: Anadara kagoshimensis,  
1465        tegillarca granosa, and anadara broughtonii. *Frontiers in Marine Science*, 11. doi: 10.3389/fmars.2024.1356356
- 1467     Lee, J. H. (1997). Studies on the gonadal development and gametogenesis of the

- 1468 granulated ark, *tegillarca granosa* (linne). *The Korean Journal of Malacology*, 13, 55–64.
- 1469
- 1470 Leguá, J., Plaza, G., Pérez, D., & Arkhipkin, A. (2013). Otolith shape analysis as  
1471 a tool for stock identification of the southern blue whiting, *micromesistius*  
1472 *australis*. *Latin American Journal of Aquatic Research*, 41, 479–489.
- 1473 Mahé, K., Oudard, C., Mille, T., Keating, J., Gonçalves, P., Clausen, L. W., &  
1474 et al. (2016). Identifying blue whiting (*micromesistius poutassou*) stock  
1475 structure in the northeast atlantic by otolith shape analysis. *Canadian*  
1476 *Journal of Fisheries and Aquatic Sciences*, 73, 1363–1371. doi: 10.1139/  
1477 cjfas-2015-0332
- 1478 May, K., Maung, C., Phy, E., & Tun, N. (2021). Spawning period of blood cockle  
1479 *tegillarca granosa* (linnaeus, 1758) in myeik coastal areas. *J. Myanmar Acad.*  
1480 *Arts Sci*, 4.
- 1481 Miranda, D. V., & Ferriols, V. M. E. N. (2023). Initial attempts on spawning and  
1482 larval rearing of the blood cockle, *tegillarca granosa* (linnaeus, 1758), in the  
1483 philippines. *Asian Fisheries Science*, 36(2). doi: 10.33997/j.afs.2023.36.2  
1484 .001
- 1485 Mérigot, B., Letourneur, Y., & Lecomte-Finiger, R. (2007). Characterization of  
1486 local populations of the common sole *solea solea* (pisces, soleidae) in the nw  
1487 mediterranean through otolith morphometrics and shape analysis. *Marine*  
1488 *Biology*, 151(3), 997–1008. doi: 10.1007/s00227-006-0549-0
- 1489 Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and  
1490 practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–  
1491 36. Retrieved from <https://doi.org/10.4097/kja.21209> doi: 10.4097/  
1492 kja.21209
- 1493 Narasimham, K. A. (1988). Taxonomy of the blood clams *anadara* (*tegillarca*)  
1494 *granosa* (linnaeus, 1758) and a. (t.) *rhombea* (born, 1780).
- 1495 Naylor, R. L., Goldburg, R. J., Primavera, J. H., Kautsky, N., Beveridge,  
1496 M. C. M., Clay, J., ... Troell, M. (2000). Effect of aquaculture on world  
1497 fish supplies. *Nature*, 405(6790), 1017–1024. doi: 10.1038/35016500
- 1498 Ponton, D. (2006). Is geometric morphometrics efficient for comparing otolith  
1499 shape of different fish species? *Journal of Morphology*, 267(7), 750–757.  
1500 doi: 10.1002/jmor.10439
- 1501 Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric  
1502 morphometrics and machine learning challenge currently accepted species  
1503 limits of the land snail *placostylus* (pulmonata: Bothriembryontidae) on the  
1504 isle of pines, new caledonia. *Journal of Molluscan Studies*, 86(1), 35–41.  
1505 doi: 10.1093/mollus/eyz031
- 1506 Ribeiro, D. (2024, Aug.). *Diogo ribeiro. data science. kruskal-wallis test.*  
1507 Retrieved from [https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal\\_wallis/](https://diogoribeiro7.github.io/statistics/data%20analysis/kruskal_wallis/) (Accessed: 2025-04-23)
- 1508
- 1509 Sany, S. B. T., Hashim, R., Rezayi, M., Salleh, A., Rahman, M. A., Safari, O.,

- 1510 & Sasekumar, A. (2014). Human health risk of polycyclic aromatic hydro-  
1511 carbons from consumption of blood cockle and exposure to contaminated  
1512 sediments and water along the klang strait, malaysia. *Marine Pollution*  
1513 *Bulletin*, 84(1-2), 268–279. doi: 10.1016/j.marpolbul.2014.05.004
- 1514 Srisunont, C., Nobpakhun, Y., Yamalee, C., & Srisunont, T. (2020). Influence  
1515 of seasonal variation and anthropogenic stress on blood cockle (*Tegillarca*  
1516 *granosa*) production potential. *Influence of Seasonal Variation and Anthro-*  
1517 *pogenic Stress on Blood Cockle (*Tegillarca Granosa*) Production Potential*,  
1518 44(2), 62–82.
- 1519 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Ma-  
1520 chine learning and its applications to biology. *PLoS Computational Biology*,  
1521 3(6), e116. doi: 10.1371/journal.pcbi.0030116
- 1522 Team, K. (n.d.). *Keras documentation: Reducelronplateau*. Retrieved from  
1523 [https://keras.io/api/callbacks/reduce\\_lr\\_on\\_plateau/](https://keras.io/api/callbacks/reduce_lr_on_plateau/)
- 1524 Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Repro-  
1525 ductive process and early development. In V. S. Kennedy, R. I. E. Newell,  
1526 & A. F. Eble (Eds.), *The eastern oyster crassostrea virginica* (pp. 335–370).  
1527 College Park, MD: Maryland Sea Grant.
- 1528 Tsutsumi, M., Saito, N., Koyabu, D., & Furusawa, C. (2023). A deep learning  
1529 approach for morphological feature extraction based on variational auto-  
1530 encoder: An application to mandible shape. *Npj Systems Biology and Ap-*  
1531 *plications*, 9(1), 1–12. doi: 10.1038/s41540-023-00293-6
- 1532 Wong, T. M., & Lim, T. G. (2018). *Cockle (anadara granosa) seed produced in*  
1533 *the laboratory, malaysia*. (Handle.net) doi: 10.3366/in\_3366.pdf
- 1534 Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves.  
1535 *IEEE Transactions on Computers*, C-21, 269–281. doi: 10.1109/tc.1972  
1536 .5008949
- 1537 Zelditch, M., Swiderski, D. L., & Sheets, H. D. (2004). *Geometric morphometrics*  
1538 *for biologists: A primer* (2nd ed.). Waltham: Elsevier Academic Press.
- 1539 Zha, S., Tang, Y., Shi, W., Liu, H., Sun, C., Bao, Y., & Liu, G. (2022). Im-  
1540 pacts of four commonly used nanoparticles on the metabolism of a ma-  
1541 rine bivalve species, *Tegillarca granosa*. *Chemosphere*, 296, 134079. doi:  
1542 10.1016/j.chemosphere.2022.134079
- 1543 Zhan, P. L., Zha, S., & Bao, Y. (2022). Hypoxia-mediated immunotoxicity in the  
1544 blood clam *Tegillarca granosa*. *Marine Environmental Research*, 177. Re-  
1545 trieved from <https://doi.org/10.1016/j.marenvres.2022.105632>

<sub>1546</sub> **Appendix A**

<sub>1547</sub> **Data Gathering Documentation**



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



Figure A.2: Sex-Based Separation of *Tegillarca granosa* Samples Post-Spawning



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



Figure A.4: Sex Identified Male Through Dissection 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*

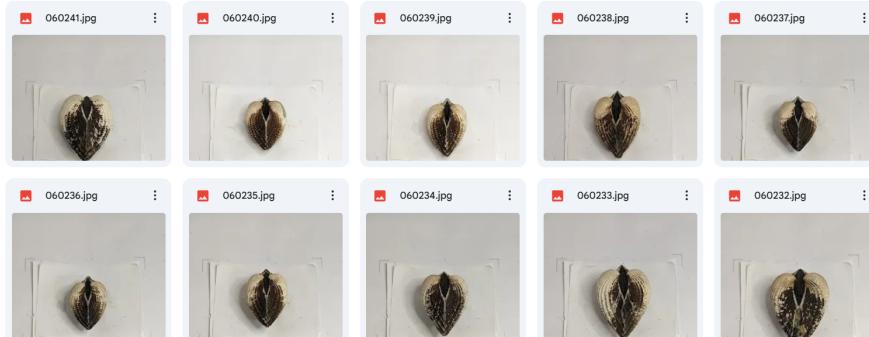


Figure A.7: Captured Images of Female *Tegillarca granosa*

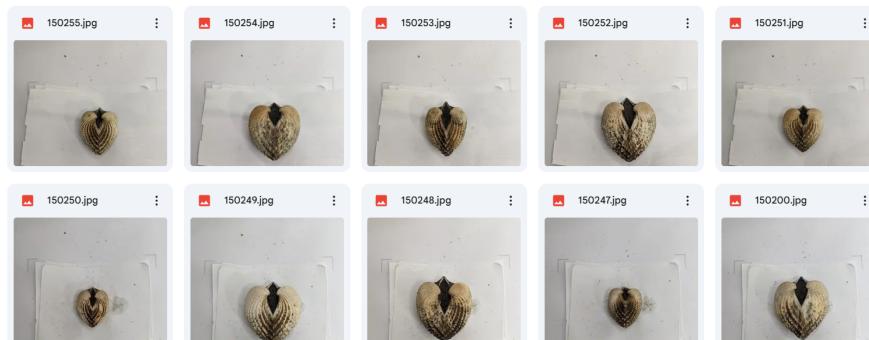


Figure A.8: Captured Images of Male *Tegillarca granosa*

1548 **Appendix B**

1549 **Supplementary Analysis**

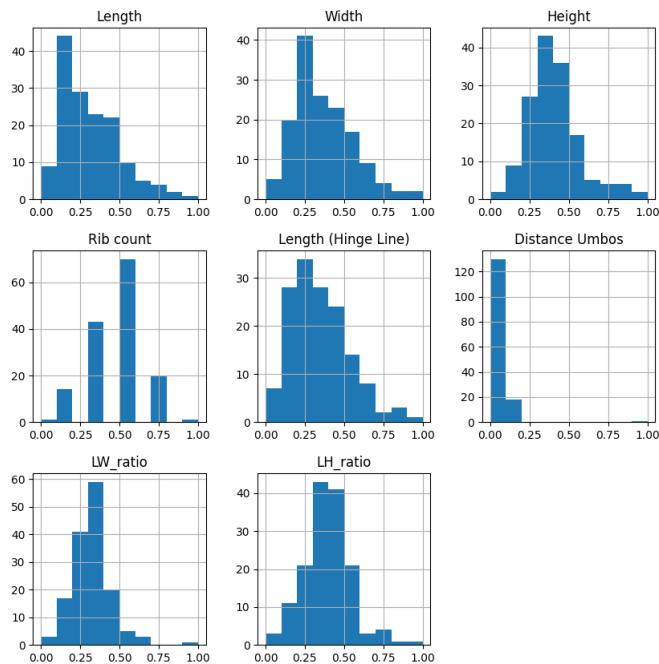


Figure B.1: Feature Distribution of *Tegillarca granosa*

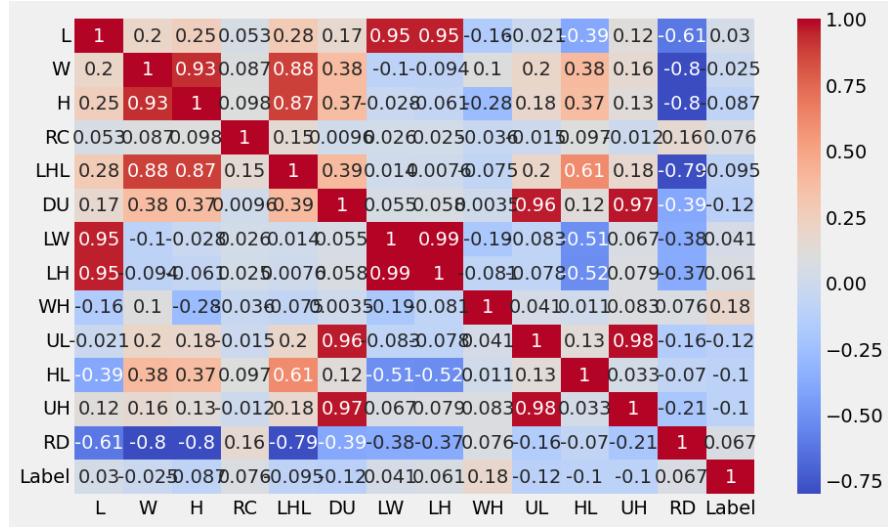


Figure B.2: Correlation Matrix of Morphological Variables *Tegillarca granosa*

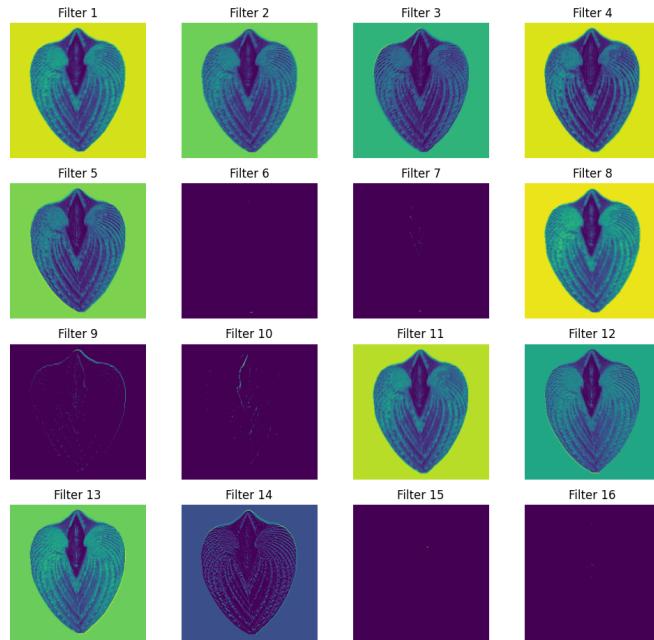


Figure B.3: Feature Maps from First Convolution Layer

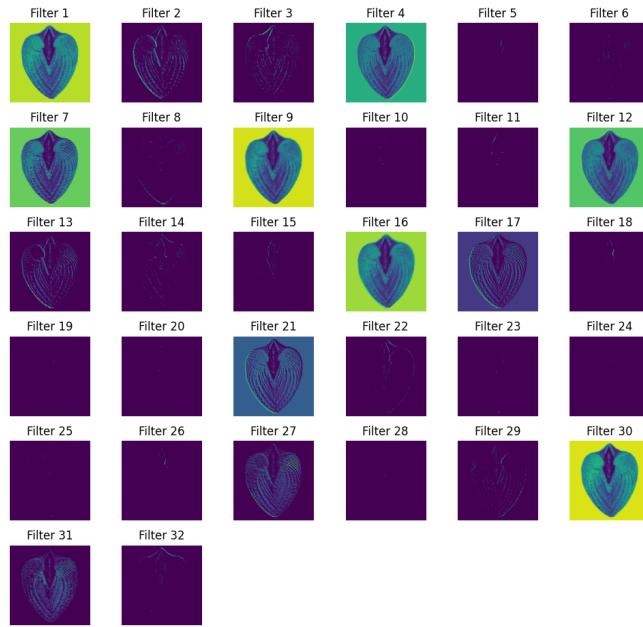


Figure B.4: Feature Maps from Second Convolution Layer

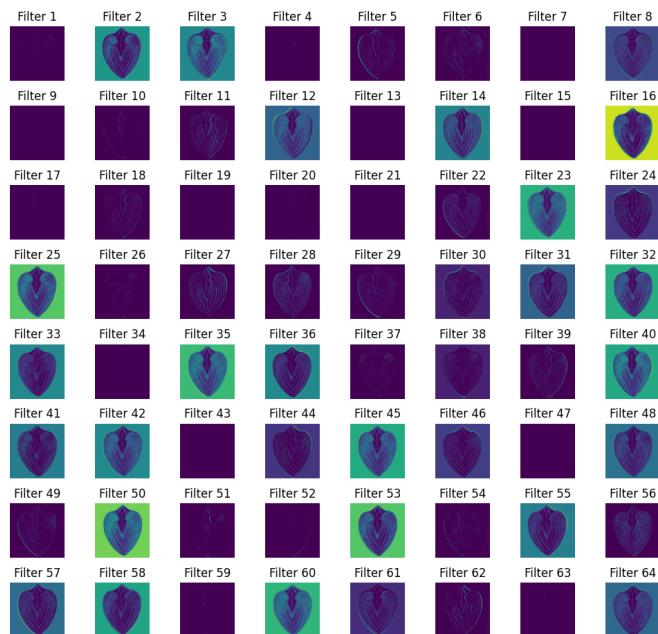


Figure B.5: Feature Maps from Third Convolution Layer