

1 A NON-INVASIVE SEX IDENTIFICATION OF BLOOD
2 COCKLES TEGILLARCA GRANOSA (LINNAEUS, 1758)
3 USING MACHINE LEARNING

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

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

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

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

20 October 22, 2024

Abstract

22 From 150 to 200 words of short, direct and complete sentences, the abstract should
23 be informative enough to serve as a substitute for reading the entire SP document
24 itself. It states the rationale and the objectives of the research. In the final Special
25 Problem document (i.e., the document you'll submit for your final defense), the
26 abstract should also contain a description of your research results, findings, and
27 contribution(s).

28 Suggested keywords based on ACM Computing Classification system can be
29 found at https://dl.acm.org/ccs/ccs_flat.cfm

30 **Keywords:** Keyword 1, keyword 2, keyword 3, keyword 4, etc.

31

Contents

32	1 Introduction	1
33	1.1 Overview	1
34	1.2 Problem Statement	2
35	1.3 Research Objectives	3
36	1.3.1 General Objective	3
37	1.3.2 Specific Objectives	3
38	1.4 Scope and Limitations of the Research	4
39	1.5 Significance of the Research	4
40	2 Review of Related Literature	6
41	2.1 Theme 1 Title	7
42	2.2 Theme 2 Title	7
43	2.3 Chapter Summary	7
44	3 Research Methodology	8
45	3.1 Sample Collection	8
46	3.2 Ethical Considerations	9
47	3.3 Creating <i>T. granosa</i> Dataset	9

48	3.4 Morphological Characteristics Collection	9
49	3.5 Image Acquisition and Pre-Processing	10
50	3.6 Machine/ Deep Learning Technologies	11
51	3.7 Support Vector Machine for Pre-evaluation	11
52	3.8 Deep Learning for Image-Based Classification	12
53	4 Preliminary Results/System Prototype	13
54	References	14
55	A Appendix Title	15
56	B Resource Persons	16

⁵⁷ List of Figures

58 List of Tables

Chapter 1

Introduction

1.1 Overview

The Philippines is a global center of marine biodiversity and has established aquaculture as a significant contributor to total fishery production (Aypa Bacongus, 2000; BFAR, 2019). As the 11th largest seafood producer in the world, the country produces over 4 million tonnes of seafood annually. Aquaculture is deeply integrated into Filipinos' livelihoods, encompassing fish cultivation and the production of various aquatic species, including mollusks. Among these are blood clams (*Tegillarca granosa*) which hold considerable economic and environmental significance.

Maintaining a balanced male-to-female ratio of blood cockles is crucial to prevent overharvesting and ensure sustainable production because an imbalanced ratio can lead to overexploitation and can impact the population's sustainability. However, there is limited literature on *T. granosa* that has a thorough understanding of its sex-determining mechanisms, particularly concerning sexual dimorphism in morphological and morphometric characteristics (Breton et al., 2017).

Currently, sex determination methods for blood cockles are invasive, including dissection, and histological examinations which often result in the death of the specimens. While there is growing literature on aquaculture commodities sex identification using machine learning and deep learning, there is a notable scarcity of research specifically addressing *T. granosa* [citation].

This study, titled "A Non-Invasive Sex Identification of *T. granosa* using Machine Learning," aims to provide a comprehensive analysis of blood cockles by

83 leveraging their morphological and morphometric characteristics. By integrating
84 machine learning and computer vision techniques, the study seeks to identify dis-
85 tinct features that indicate sexual dimorphism between male and female blood
86 cockles.

87 1.2 Problem Statement

88 Accurately identifying the sex of *T. granosa* is important in order to promote sus-
89 tainable aquaculture and biodiversity by maintaining a balanced male-to-female
90 ratio. A balanced ratio helps prevent overharvesting. Although sex identification
91 is important for blood cockle population management and sustainable aquacul-
92 ture, there is a notable lack of research in creating non-invasive methods to identify
93 the sex of *T. granosa*. Many of the latest studies and approaches are based on
94 invasive methods like dissection or histological analysis, which are impractical for
95 large-scale aquaculture operations focused on conservation.

96 The existing invasive methods for identifying the sex of *T. granosa* often re-
97 quire dissection, a technique that involves cutting open the shell to visually inspect
98 the gonads (Erica, 2018). This causes harm and death to the specimens. In some
99 cases, histological examination is used to examine tissue samples through a mi-
100 croscope, leading to further destruction of the organism (May et al., 2021). These
101 methods are time-consuming, labor-intensive, and can pose a threat to population
102 management, especially when it is essential to maintain a balanced sex ratio for
103 breeding programs. Moreover, invasive methods also require technical skills to ex-
104 ecute properly. Aquaculture operations, particularly in resource-limited settings,
105 face challenges in accessing laboratory equipment like microscopes and staining
106 tools which complicates the process.

107 A less invasive approach employed by aquaculturists is to monitor spawning
108 behavior in which individuals are separated and stimulated to reproduce in order
109 to determine their sex through the release of gametes (Miranda Ferriols, 2023).
110 Although it is indeed less invasive than dissection, spawning still involves inducing
111 stress in blood cockles and may not be completely effective for fast identification
112 in large populations.

113 Given the limitations of both invasive and less invasive methods highlight the
114 need for a more advanced approach. An alternative, non-invasive method involv-
115 ing machine and deep learning technologies might solve these issues by providing
116 a fast, accurate, and effective solution without harming or stressing the blood
117 cockles.

118 1.3 Research Objectives

119 1.3.1 General Objective

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

125 1.3.2 Specific Objectives

126 To achieve the general objective of developing a non-invasive sex identification of
127 *T. granosa* using machine and deep learning, the following specific objectives have
128 been established:

- 129 1. To collect and organize a comprehensive dataset of *T. granosa* which will
130 include high-quality images and relevant morphological measurements that
131 will serve as the basis for the machine-learning model.
- 132 2. To preprocess the collected data to perform quality control and consistency
133 checks. This will include techniques such as color thresholding, segmenta-
134 tion, and image hole filling and dilating.
- 135 3. To develop and implement machine learning models that can classify the sex
136 of *T. granosa* based on the collected dataset, implementing algorithms such
137 as support vector machines (SVM) for pre-evaluation, and deep learning
138 models such as Squeezenet and Unet.
- 139 4. To evaluate the performance of the models used using performance metrics
140 such as accuracy, precision, recall, and F1-score to ensure the effectiveness
141 and reliability of the models.
- 142 5. To compare the developed models against existing methods, such as dissec-
143 tion and spawning, and assess their potential for real-world application in
144 aquaculture operations.

1.4 Scope and Limitations of the Research

This study focuses on developing a non-invasive method for identifying the sex of *Tegillarca granosa* using machine learning, deep learning, and computer vision technologies. The goal is to provide an accurate and efficient means of sex identification without causing harm to the specimens, contributing to sustainable aquaculture practices.

The researchers will work with 500 spawned blood cockles taken from Panay island, specifically Zarraga Iloilo and Ivasan Capiz, equally divided between 250 males and 250 females, obtained through temperature shock. The researchers will personally gather linear measurements, including length, width, height, rib count, length of the hinge line, and distance between the umbos using the vernier caliper. Images and corresponding views of the specimens will also be collected by the researchers under the supervision of the University Researchers Associate from the Institute of Aquaculture, College of Fisheries and Ocean Sciences.

Data collection will take place at the hatchery facility of the University of the Philippines Visayas. Data gathering will be conducted in batches, depending on the availability of spawned samples.

The method developed in this study is specific to *Tegillarca granosa* and may not be generalized to other species. The model is trained exclusively for *Tegillarca granosa* and morphological features including length, width, height, rib count, length of the hinge line, and distance between the umbos may not be shared by other shellfish species.

1.5 Significance of the Research

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

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

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

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

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

Chapter 2

Review of Related Literature

This chapter discusses the features, capabilities, and limitations of existing research, algorithms, or software that are related/similar to the Special Problem.

The reviewed works and software must be arranged either in chronological order, or by area (from general to specific). Observe a consistent format when presenting each of the reviewed works. This must be selected in consultation with the adviser.

DO NOT FORGET to cite your references.

A literature review must do these things:

- be organized around and related directly to the thesis or research question you are developing
- synthesize results into a summary of what is and is not known
- identify areas of controversy in the literature
- formulate questions that need further research

A literature review is a piece of discursive prose, not a list describing or summarizing one piece of literature after another. It's usually a bad sign to see every paragraph beginning with the name of a researcher. Instead, organize the literature review into sections that present themes or identify trends, including relevant theory. You are not trying to list all the materials published, but to synthesize and evaluate them according to the guiding concept of your thesis or research question. You should also state the limits or gaps of their researches wherein you will try to fill these gaps in accordance to your research problem and objectives.

214 **2.1 Theme 1 Title**

215 This chapter contains a review of research papers that:

- 216 • Describes work on a research area that is similar or relevant to yours
- 217 • Describes work on a domain that is similar or relevant to yours
- 218 • Uses an algorithm that may be useful to your work
- 219 • Uses a software / tool that may be useful to your work

220 It also contains a review of software systems that:

- 221 • Belongs to a research area similar to yours
- 222 • Addresses a need or domain similar to yours
- 223 • Is your predecessor

224 **2.2 Theme 2 Title**

225 **2.3 Chapter Summary**

226 Should include a table of related studies comparing them based on several criteria.

227 Highlight research gaps and the research problem.

Chapter 3

Research Methodology

This chapter discusses the materials and methods to be employed in the study, focusing on the development requirements and the software and languages utilized. This will also entail the overall workflow in conducting the study, Non-Invasive Methods in Determining the Sex of *Tegillarca granosa* (blood cockles) using machine learning technologies. The different machine/deep learning algorithms will be thoroughly discussed to ensure a comprehensive understanding of the entity of the research endeavor and its processes.

Dr. Victor Emmanuel Ferriols, the director of the Institute of Aquaculture, will oversee the overall workflow and conduct of this experiment. The researchers will also be guided by the research associates, LC Mae Gasit and Allena Esther Artera. Consequently, the whole dataset collection process will be done at the University of the Philippines Visayas hatchery facility.

3.1 Sample Collection

A total of 1000 adult *T. granosa* that have already spawned will be used in this experiment wherein their sex was already classified as male or female. The sample sizes are going to range from 34 to 61 mm and will be sourced from the coastal area in the municipality of Zaraga, Iloilo, Philippines, as well as from fish markets in the municipality of Ivisan, Capiz, Philippines. The research and experimentation will be done at the University of the Philippines Visayas hatchery facility in Miagao, Iloilo, Philippines. The samples will be placed in 200 L fiberglass reinforced plastic (FRP) tanks containing filtered seawater with 35 ppt salinity (Ferriols, Miranda, 2023) and will be subjected to spawning to categorize male from female

252 *T. granosa*. The samples will undergo a series of temperature fluctuations to
253 induce the spawning of gametes as described in the study of Ferriols and Miranda
254 (2023). This method, induced spawning, is the most natural and least invasive
255 method for bivalves compared to other methods (Aji, 2021). Thus, after the
256 spawning, there would be 500 classified males and 500 classified females.

257 3.2 Ethical Considerations

258 Ethical approval was not required for this study involving animals, as per local leg-
259 islation and institutional guidelines, because the experiments were conducted only
260 on species that are commonly used as food and intended for human consumption.

261 3.3 Creating *T. granosa* Dataset

262 For the initial preparation of the experiment, the researchers will collect primary
263 observations for 100 samples of *T. granosa*. For the actual experimentation, the
264 researchers will collect the dataset by batch eventually comprising 1000 samples
265 of *T. granosa*. The images captured for the dataset will be saved in png format
266 with a file naming convention of the sample’s sex, the orientation or view of the
267 shell, and its corresponding number out of the total 1000 samples. Female *T.*
268 *granosa* samples will begin with 0 in their file name, while males will begin with
269 1, followed by the views captured such as (1) dorsal, (2) ventral, (3) anterior,
270 (4) posterior, (5) left lateral, and (6) right lateral, and lastly, a unique sample
271 number. For example, “010001” will be the file name for the first female sample
272 taken from the dorsal view and “110001” for the first male sample also taken from
273 the dorsal view. The dataset will be organized in a CSV file that lists each image’s
274 file name along with their shell’s width, height, length, rib count, length of the
275 hinge line, and distance between their umbos. This dataset will be essential for
276 machine learning model training and testing.

277 3.4 Morphological Characteristics Collection

278 Morphology refers to the biological form and represents one of the most visually
279 recognizable phenotypes across all organisms (Tsutsumi et al., 2023). Morphology
280 is a term that describes structural characteristics by measuring specific compo-
281 nents, namely, dimensions such as shapes, sizes, and colors. As stated by the

researchers, quantifying and characterizing the shape is essential to understanding and visualizing the variations in *T. granosa*'s morphology. In this study, the researchers are going to measure the height, width, and length of *T. granosa*. The dimensions will be recorded using a Vernier caliper to the nearest 0.01 mm. The length of the *T. granosa* refers to the measurement from the anterior to the posterior of the shell, the width will be measured through the shell's widest point from the left to the right valve and lastly, the height will be measured from the base of the shell to the shell's apex. The height of the gap between the valves near the hinge will also be measured. The authors Reymont and Kennedy (1998), indicated that the use of counts of the shell ribs as supplementary information increases identification accuracy. Thus, the researchers will also take into account the difference in the rib count of the male and female *T. granosa* and the ratio will be calculated since the sizes of the blood clams may vary. Sex ratio, size frequency distribution, and relative growth rates were used to investigate sexual dimorphism.

3.5 Image Acquisition and Pre-Processing

In this study, there would be three major phases for the image processing to be employed namely (1) color thresholding, (2) segmentation, and (3) image hole filling and dilating. The researchers constructed a controlled environment for capturing the samples utilizing a box-like structure of (?) meters with a green background surface. This setup was designed to maintain uniform captures of the images, and a consistent measurement between the sample and the camera, fixing the camera at 50 cm above the *T. granosa*. Placing a ring light to the left of the box, and using a camera with flash to ensure the image quality, eliminate shadows and clarity of the sample during the image acquisition process. For color thresholding, the researchers utilized the red, green, blue (RGB), hue saturation value (HSV), luminance, blue chromaticity, red chromaticity (YCbCr), and (Luminance, a, b)** (CIElab) images obtained from the smartphone considering their wide availability across various stages in the bivalve industry using the MATLAB Colour Thresholding Toolbox in determining which among the four-color spectra may generate the cleanest version of the training images with absence of any blobs (Jayasundara et al., 2023). Google Pixel 3 XL will be utilized with the following specifications: 2960 x 1440 for the resolution, 4,032 x 3,024 pixels (12.2 MP) for the dimensions, f/1.8 for the fstop, 28mm (wide), $\frac{1}{2}$.55", 1.4 μ m, dual pixel PDAF, OIS. [insert reference] After thresholding, the lazy snapping technique will be implemented by manually drawing the background and the foreground lines that represent the black pixels and the bivalve pixels. The lazy snapping algorithm will be configured using the 20 000 superpixels which can divide the *T. granosa*'s

320 images into 20, 000 irregularly shaped geometric pixels that will be based on the
 321 CIElab gradients through K-means clustering with $K = 3$. For the last step, the
 322 researchers will perform image hole filling and dilating to ensure that no blobs
 323 are remaining that can contribute to noise which can affect the correctness of the
 324 extracted feature by taking into consideration the 200-pixel blobs that are discon-
 325 nected from the largest object in its binary form. This will result in black pixels
 326 made by binary filling and dilating to remove the blobs. [reference] Image process-
 327 ing will be performed on the MATLAB [version[-]] installed on the [laptop] with
 328 specs. The images will be saved based on how it was stated on the collection of
 329 the image dataset. To ensure consistent comparisons for the analysis, the images
 330 were captured in different angles including dorsal, ventral, lateral, and anterior
 331 and posterior taken in uniform angles to provide visual coverage of the *T. granosa*
 332 sample.

333 **3.6 Machine/ Deep Learning Technologies**

334 This section of the paper will discuss the technologies to be used in training, and
 335 testing the model as well as associated techniques and algorithms. Since obtaining
 336 the induced samples was done per batch, the researchers will conduct an initial
 337 run with a support vector machine before delving into more complex methods
 338 such as deep learning models.

339 **3.7 Support Vector Machine for Pre-evaluation**

340 The shape of recording structures was first analyzed by collecting measurements
 341 of linear distances and applying multivariate statistical methods to these data
 342 (traditional linear measurement method) (Rohlf and Marcus, 1993). Geometric
 343 morphometric (GM) methods are an alternative way of analyzing and quantifying
 344 shape, which in theory retains more detail about the geometry of the structure
 345 than could be obtained from linear measurements (Adams et al., 2004). Machine
 346 learning techniques such as decision tree classification, support vector machines
 347 (SVMs), and artificial neural networks (ANNs) have been applied to the analysis
 348 of bivalve shell geometry and morphology to classify shells based on morpholog-
 349 ical features, including shell shape, size, and texture, among others (Kiel, 2021).
 350 The results of these studies have shown that machine learning algorithms can
 351 accurately classify bivalve shells and provide insights into the relationships be-
 352 tween shell morphology and various environmental factors. Following this, the
 353 researchers are going to conduct a pre-evaluation of the linear measurements for

354 100 samples of *T. granosa* using a Support Vector Machine in order to quantify
355 whether the linear measurements can be a determining factor in determining the
356 sex of the samples before proceeding to more complex methods.

357 **3.8 Deep Learning for Image-Based Classifica-** 358 **tion**

359 After collecting a sufficient number of images and identifying initial patterns,
360 convolutional neural networks (CNNs) will be used. CNNs, models like VGGNet,
361 ResNet, and Inception have been effectively applied in phenotype classification
362 (Kim et al., 2024). In this study, the deep learning model will be specifically
363 adapted for the sex identification of *T. granosa* based on shell images. CNNs
364 will analyze the images and learn important details about their shapes that can
365 help identify whether they are male or female. Unlike the approach of using
366 three models taken by Kim et al. (2024), the researchers will focus on just one
367 model that has shown the best performance in their study which is SqueezeNet.
368 SqueezeNet is particularly advantageous because it reduces the number of pa-
369 rameters and amount of memory required to store the model without sacrificing
370 accuracy (Koonce, 2021; Sayed et al., 2021). Its ability to achieve high accuracy
371 in classifying shell images makes it a suitable choice for distinguishing between
372 male and female *T. granosa*. Python and Keras libraries will be used to train and
373 test the model. The dataset will be divided into training (), validation (), and
374 testing. Performance metrics such as accuracy, precision, recall, and F1-score will
375 be used to evaluate the model’s effectiveness.

376 Chapter 4

377 Preliminary Results/System 378 Prototype

379 This chapter presents the preliminary results or the system prototype of your SP.
380 Include screenshots, tables, or graphs and provide the discussion of results.

References

- Fedkiw, R., Stam, J., & Jensen, H. W. (2001). Visual simulation of smoke. In E. Fiume (Ed.), *Proceedings of siggraph 2001* (pp. 15–22). ACM Press / ACM SIGGRAPH.
- Jobson, D. J., Rahman, Z., & Woodell, G. A. (1995). Retinex image processing: Improved fidelity to direct visual observation. In *Proceedings of the is&t fourth color imaging conference: Color science, systems, and applications* (Vol. 4, pp. 124–125).
- Kartch, D. (2000). *Efficient rendering and compression for full-parallax computer-generated holographic stereograms* (Unpublished doctoral dissertation). Cornell University.
- Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., ... Fulk, D. (2000). The digital michelangelo project. In K. Akeley (Ed.), *Proceedings of siggraph 2000* (pp. 131–144). New York: ACM Press / ACM SIGGRAPH.
- Park, S. W., Linsen, L., Kreylos, O., Owens, J. D., & Hamann, B. (2006, March/April). Discrete sibson interpolation. *IEEE Transactions on Visualization and Computer Graphics*, 12(2), 243–253.
- Parke, F. I., & Waters, K. (1996). *Computer facial animation*. A. K. Peters.
- Pellacini, F., Vidimče, K., Lefohn, A., Mohr, A., Leone, M., & Warren, J. (2005, August). Lpics: a hybrid hardware-accelerated relighting engine for computer cinematography. *ACM Transactions on Graphics*, 24(3), 464–470.
- Sako, Y., & Fujimura, K. (2000). Shape similarity by homotropic deformation. *The Visual Computer*, 16(1), 47–61.
- Yee, Y. L. H. (2000). *Spatiotemporal sensitivitiy and visual attention for efficient rendering of dynamic environments* (Unpublished master’s thesis). Cornell University.

408 **Appendix A**

409 **Appendix Title**

410 **Appendix B**

411 **Resource Persons**

412 **Mr. Firstname1 Lastname1**

413 Role1

414 Affiliation1

415 emailaddr1@domain.com

416 **Ms. Firstname2 Lastname2**

417 Role2

418 Affiliation2

419 emailaddr2@domain.net

420