A Non-Invasive Sex Identification of Blood Cockles Tegillarca Granosa (Linnaeus, 1758) Using Machine Learning

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

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

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

Francis DIMZON
Adviser
Victor Marco Emmanuel FERRIOLS
Co-Adviser

October 22, 2024

Abstract

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

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

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

Contents

List of Figures

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 Baconguis, 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 leveraging their morphological and morphometric characteristics. By integrating machine learning and computer vision techniques, the study seeks to identify distinct features that indicate sexual dimorphism between male and female blood cockles.

1.2 Problem Statement

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

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

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

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

1.3 Research Objectives

1.3.1 General Objective

The general objective of this study is to develop a non-invasive method for identifying the sex of Tegillarca granosa using machine and deep learning integrated with computer vision technologies. This method aims to provide accurate and streamlined sex identification without causing harm to the specimens, thus supporting sustainable aquaculture practices.

1.3.2 Specific Objectives

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

1. To collect and organize a comprehensive dataset of T. granosa which will include high-quality images and relevant morphological measurements that will serve as the basis for the machine-learning model.

2. To preprocess the collected data to perform quality control and consistency checks. This will include techniques such as color thresholding, segmentation, and image hole filling and dilating.

3. To develop and implement machine learning models that can classify the sex of T. granosa based on the collected dataset, implementing algorithms such as support vector machines (SVM) for pre-evaluation, and deep learning models such as Squeezenet and Unet.

4. To evaluate the performance of the models used using performance metrics such as accuracy, precision, recall, and F1-score to ensure the effectiveness and reliability of the models.

5. To compare the developed models against existing methods, such as dissection and spawning, and assess their potential for real-world application in 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.

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

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

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

2.1 Theme 1 Title

This chapter contains a review of research papers that:

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

It also contains a review of software systems that:

- Belongs to a research area similar to yours
- Addresses a need or domain similar to yours
- Is your predecessor

2.2 Theme 2 Title

2.3 Chapter Summary

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

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

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

3.2 Ethical Considerations

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

3.3 Creating T. granosa Dataset

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

3.4 Morphological Characteristics Collection

Morphology refers to the biological form and represents one of the most visually recognizable phenotypes across all organisms (Tsutsumi et al., 2023). Morphology is a term that describes structural characteristics by measuring specific components, 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 Reyment 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", 1.4µm, 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 images into 20, 000 irregularly shaped geometric pixels that will be based on the CIElab gradients through K-means clustering with K = 3. For the last step, the researchers will perform image hole filling and dilating to ensure that no blobs are remaining that can contribute to noise which can affect the correctness of the extracted feature by taking into consideration the 200-pixel blobs that are disconnected from the largest object in its binary form. This will result in black pixels made by binary filling and dilating to remove the blobs. [reference] Image processing will be performed on the MATLAB [version[-] installed on the [laptop] with specs. The images will be saved based on how it was stated on the collection of the image dataset. To ensure consistent comparisons for the analysis, the images were captured in different angles including dorsal, ventral, lateral, and anterior and posterior taken in uniform angles to provide visual coverage of the T. granosa sample.

3.6 Machine/ Deep Learning Technologies

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

3.7 Support Vector Machine for Pre-evaluation

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

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

3.8 Deep Learning for Image-Based Classification

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

Chapter 4

Preliminary Results/System Prototype

This chapter presents the preliminary results or the system prototype of your SP. Include screenhots, 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 sensistivity and visual attention for efficient rendering of dynamic environments* (Unpublished master's thesis). Cornell University.

Appendix A

Appendix Title

Appendix B

Resource Persons

Mr. Firstname1 Lastname1
Role1
Affiliation1
`emailaddr1@domain.com`

Ms. Firstname2 Lastname2
Role2
Affiliation2
`emailaddr2@domain.net`

....