Contents list available at CBIORE journal website

Journal of Emerging Science and Engineering

Journal homepage: https://journal.cbiore.id/index.php/jese/index

Research Article

Shrimp and fish underwater image classification using features extraction and machine learning

Arif Setiawan [1,5*], H. Hadiyanto [2,4], Catur E. Widodo [3,4]

[1]Doctoral Program of Information Systems, School of Postgraduate Studies, Diponegoro University, Jl. Imam Bardjo SH, Semarang, 50241, Indonesia

[2]Center of Biomass and Renewable Energy (CBIORE), Department of Chemical Engineering, Diponegoro University. Jl. Prof. Soedarto SH, Tembalang, Semarang, 50271, Indonesia

[3]Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Jl. Prof. Soedarto, Tembalang, Semarang, 50275, Indonesia

[4]School of Postgraduate Studies, Diponegoro University, Jl. Imam Bardjo SH, Semarang, 50241, Indonesia

[5]Department of Information System, Faculty of Engineering, Muria Kudus University, Jl. Lingkar Utara, Kayuapu Kulon, Gondangmanis, Kudus, 59327, Indonesia

Abstract. Shrimp cultivation is one type of cultivation that has a significant impact on the social status of coastal communities. Shrimp farming traditionally faces several challenges, including water pollution, imbalances in temperature, feed, media, and costs. Monitoring the condition of shrimp in the cultivation environment is very necessary to determine the condition of shrimp in the water. Classification of shrimp and fish is the first step in monitoring the condition of shrimp underwater. This research proposes the development of a method for classifying shrimp and fish underwater using feature extraction and machine learning. The flow of this research is: (1) preparing data from ROI detection results, (2) extraction process of morphometric characteristics P and T, (3) calculating the value of morphometric characteristics P and T, (4) data breakdown for training data and testing data, (5) Model creation process, data training and data testing using SVM, RF, DT, and KNN, (6) Evaluation of classification results using a confusion matrix. From this research, it was found that the Random Forest method obtained the highest accuracy, namely 0.93. From this matrix, the values obtained are True Positive = 349, False Positive = 28, True Negative = 223, False Negative = 0.

Keywords: shrimp, fish, classification, features extraction, machine learning

Received: 5th February 2024; Revised: 29th March 2024; Accepted: 2nd April 2024; Available online: 8th April 2024

1. Introduction

Shrimp cultivation is one type of cultivation that has a significant impact on the social status of coastal communities. Shrimp farmers get more income after they cultivate shrimp. Shrimp cultivation helps increase the economic level from 26% to 36% (Ray et al. 2021). Seeing the very high demand for shrimp, shrimp cultivation is widely implemented to increase mass production needs. Shrimp production takes 3 months for one harvest, so it requires a relatively short time for cultivators (Chaikaew et al. 2019). Seawater shrimp production is based on monoculture, shrimp are usually cultivated in systems in large ponds. The rearing cycle varies from 3 to 4 months, with 10–20 PL (post larvae) stocked per m2 (Valenti et al. 2021). Smart shrimp farming is a new scientific field that aims to optimize the efficient use of resources. Sustainable development in cultivation is carried out through the integration of artificial intelligence technology and other modern information technology. A large amount of data in intelligent farming systems poses various challenges, such as multiple data sources with complex data formats, and various information about equipment, shrimp, environment, and breeding processes. Several data formats include text, image, audio, and video, the complexity of the data comes from different species, water media, and cultivation stages (Yang et al. 2021).

Shrimp farming traditionally faces several challenges, including water pollution, imbalances in temperature, feed, media, and costs. Image processing technology in shrimp cultivation changes manual cultivation systems into sophisticated systems. The data collection system uses cameras and can be analyzed using machine learning models, producing decisions with the help of artificial intelligence (Huang et al. 2019). In shrimp farming, an intelligent system is needed that can monitor shrimp size, shrimp growth, and feeding accurately without catching the shrimp directly. Intelligent systems can help shrimp farmers by reading growth parameters and food requirements promptly to monitor and maintain their quality (Z. Liu, Jia, and Xu 2019). The transformation of the digital sector in the Industrial Revolution 4.0 makes shrimp cultivation a priority. Smart shrimp cultivation is a technological development that emphasizes the use of modern technology (Triantafyllou et al. 2019).

Monitoring the condition of shrimp in the cultivation environment is very necessary to determine the condition of shrimp in the water (H. Liu et al. 2021). Classification of shrimp and fish is the first step in monitoring the condition of shrimp underwater. Digital

* Corresponding author
Email: arif.setiawan@umk.ac.id (A. Setiawan)

https://doi.org/10.61435/jese.2024.e14
ISSN: p-ISSN: 3026-0817; e-ISSN: 3026-0183.The Author(s). Published by CBIORE

image feature extraction is used in the initial process, in recent years digital image processing technology has developed and has achieved high accuracy because it can automatically extract high-dimensional features. This research proposes the development of a new method for classifying shrimp and fish underwater using feature extraction and machine learning.

2. Materials and Methods

Shrimp are a species that breeds in aquatic environments. Shrimp cultivation is a cultivation process that produces high-quality protein for humans (Karim et al. 2021). Shrimp has become one of the most favorite seafood in recent years. As shrimp production grows throughout the world, controlling shrimp quality has become an important task for the seafood industry (Do and Quoc 2022). Monitoring shrimp health is very important to improve shrimp quality and production efficiency. Before shrimp are sent to market, they must be assessed and classified according to established standards (Ray et al. 2021). Extraction of shrimp image features is a search for information on unique characteristics contained in shrimp images. Image characteristics can be taken from color, texture, and shape (Patel 2016). Morphometric characteristics are characteristics of the shape of marine species. Various characteristics are used to identify marine animal species such as morphology, physiology, and behavior. The morphometric characteristics used are shrimp length (L), shrimp height (H), and shrimp width (W), (Ap 2016) (Zhang, Wang, and Duan 2020). In detail, the morphometric characteristics are explained in Table 1.

Table 1
Morphometric features of shrimp

Features	Symbol
Shrimp length	L
Shrimp height	H
Shrimp width	W

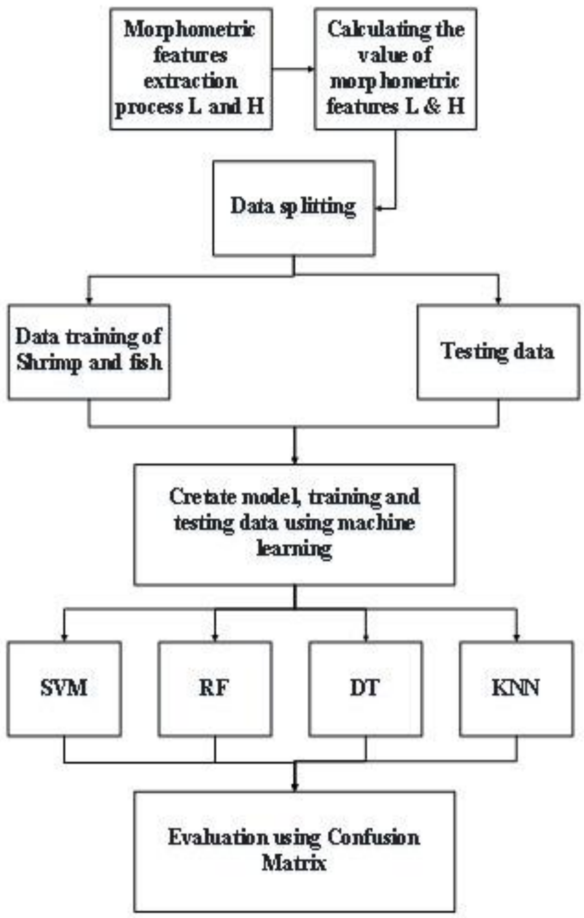


Fig. 1 Research flow diagram

Support Vector Machine (SVM) is a supervised learning method that can be used for regression and classification. SVM has several kernels, including linear kernels, polynomial kernels, and radial basis function (RBF) kernels (Mou et al. 2022). SVM is a method used to find the distance between classes. Random Forest (RF) is a machine-learning algorithm that is a collection of decision trees. RF can be used for regression and classification of big data. RF works by building a decision tree to obtain regression results (Gao and Zhou 2020). RF combines many trees in processing training data, the number of trees used affects the resulting level of accuracy, the more trees used, the higher the accuracy (Yoshikazu, Krecl, and Créso 2022)**.** Decision Tree (DT) is a machine learning algorithm that applies to create decision rules like a tree structure. DT is a supervised learning method that can be used for classification and regression (Ghane et al. 2022). DT uses rules to make decisions such as a tree structure, having branches in decision-making so that the results obtained are maximized (Spirkovska 1993). K-Nearest Neighbor (KNN) is an algorithm that classifies data based on similarities with other data. In the case of KNN regression, it provides recognition of the K nearest neighbors in the regression (Wang et al. 2022). KNN is an algorithm that works using the nearest K, which is used as a reference for determining class.

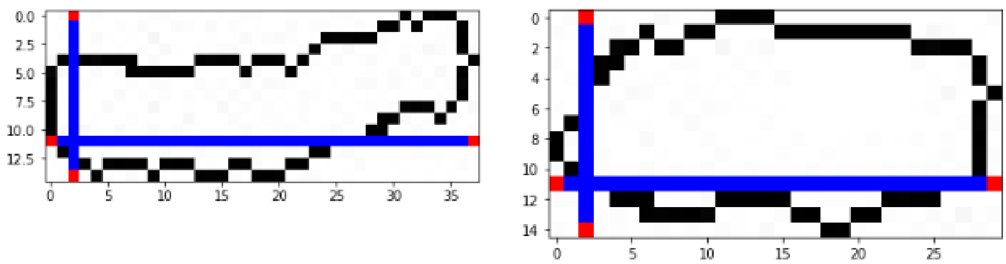
At this stage, the research carried out is to classify shrimp and fish objects underwater. This stage is the fifth stage of research. The flow of this research is: (1) preparing data from ROI detection results, (2) extraction process of morphometric characteristics P and T, (3) calculating the value of morphometric characteristics P and T, (4) data breakdown for training data and testing data, (5)) Model creation process, data training and data testing using Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN) (6) Evaluation of classification results using a confusion matrix. The research flow diagram is explained in Figure 1.

3. Results and Discussion

This research is the fifth stage of research. This research aims to differentiate shrimp and fish objects in water. The machine learning algorithms used in this research are SVM, RF, DT, and KNN (Singla, Garg, and Dubey 2020). This method was chosen because, from several previous studies, this classification method is a method that can obtain good classification accuracy results (Uddin et al. 2022).

3.1 Process of Extracting Morphometric Characteristics of Underwater Objects

The research begins with the process of extracting morphometric characteristics of underwater digital image objects, and the process of calculating morphometric characteristic values using Euclidean Distance. The data used in this research amounted to 2000 data, which had 2 classes, namely shrimp and fish classes. The data used in this research are underwater living creatures, namely shrimp and fish. From each morphometric characteristic of the two types of living creatures, what is similar is the morphometric characteristic of the length (L) and height (H) of the object, so that is used in the classification process is only the morphometric characteristics L and H (Saberioon and Císa 2018). The process of analyzing digital images of shrimp and fish objects underwater is explained in Figure 2.



Extraction of shrimp morphometric features   Extraction of fish morphometric features

Fig. 2 Analysis of digital images of shrimp and fish objects underwater

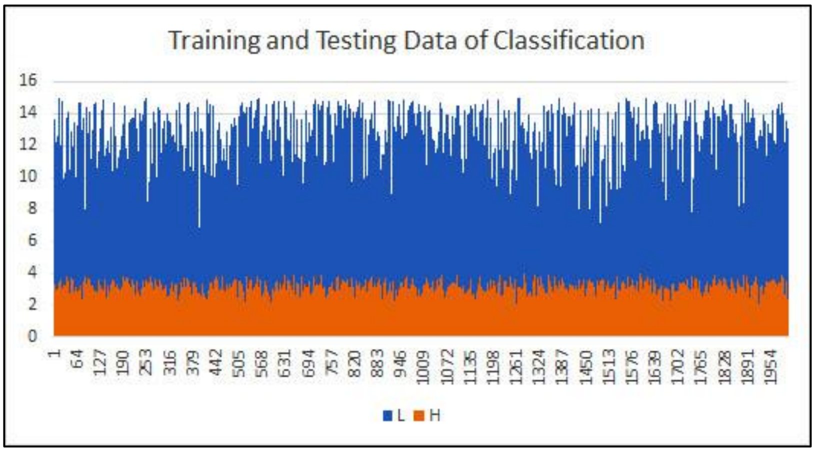


Fig. 3 Training and testing data for classification training and testing

3.2 Training and Testing Data

The machine learning methods used for classification are SVM, RF, DT, and KNN (Uddin et al. 2022). The functions used in Python are SVC.fit(x,y) for the SVM method, RandomForestClassifier.fit(x,y) for the RF method, tree.DecisionTreeClassifier.fit(x,y) for the DT method, and NeighborsClassifier.fit(x,y) for the KNN method. The morphometric characteristics of 2000 training and testing data are shown in Figure 3. Of the total data used in the research, data was split into 70% training data and 30% testing data.

3.3 Analysis Classification to differentiate shrimp from fish

From the results of training and testing data, confusion matrix results were obtained (Isa et al. 2020) as shown in Figure 4, with each accuracy value in detail shown in Figure 5. From the classification process carried out, the results of the RF method obtained the highest accuracy, namely 0.93. The confusion matrix resulting from the RF method is shown in Figure 6. From this matrix, the True Positive value in the lower right quadrant = 349, the False Positive in the upper right quadrant = 28, the True Negative in the upper left quadrant = 223, False Negative in the lower left quadrant = 0.

	precision	recall	f1-score	support
fish	0.69	0.60	0.64	251
shrimp	0.73	0.80	0.76	349
accuracy			0.71	600
macro avg	0.71	0.70	0.70	600
weighted avg	0.71	0.71	0.71	600

SVM

	precision	recall	f1-score	support
fish	1.00	0.89	0.94	251
shrimp	0.89	1.00	0.96	349
accuracy			0.95	600
macro avg	0.96	0.94	0.95	600
weighted avg	0.96	0.95	0.95	600

RF

	precision	recall	f1-score	support
fish	0.49	1.00	0.54	251
shrimp	1.00	0.38	0.81	349
accuracy			0.61	600
macro avg	0.75	0.69	0.60	600
weighted avg	0.81	0.61	0.59	600

DT

	precision	recall	f1-score	support
fish	0.69	0.60	0.64	251
shrimp	0.73	0.80	0.76	349
accuracy			0.71	600
macro avg	0.71	0.70	0.70	600
weighted avg	0.71	0.71	0.71	600

KNN

Fig. 4 Confusion matrix 4 classification methods

The RF method works with many decision trees to improve model performance, several decision trees are trained using the bagging method, namely by combining several decision trees to improve results (Giri et al. 2023), as shown in Figure 6. The RF method selects random samples from the data provided, then this method creates a decision tree for each selected sample, the RF classification process uses the values that appear most frequently, and then the most frequently selected classification results are selected (Zermane et al. 2023).

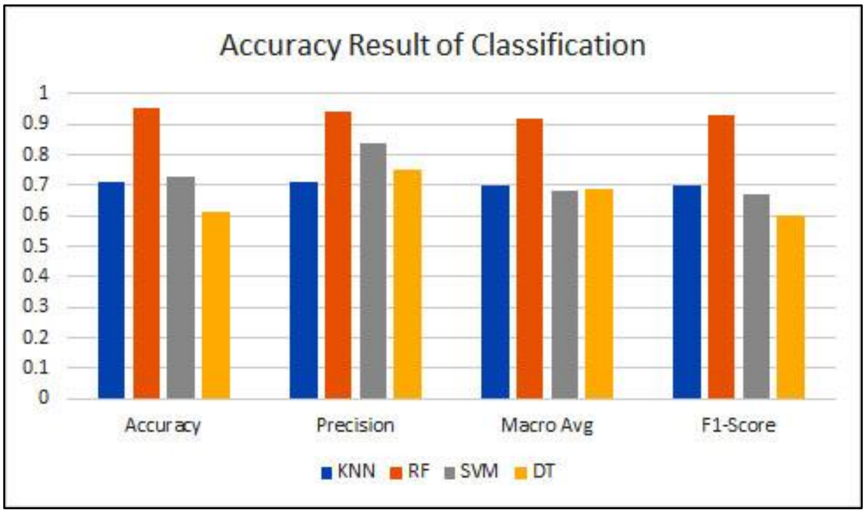


Fig. 5 Accuracy results for shrimp and fish classification

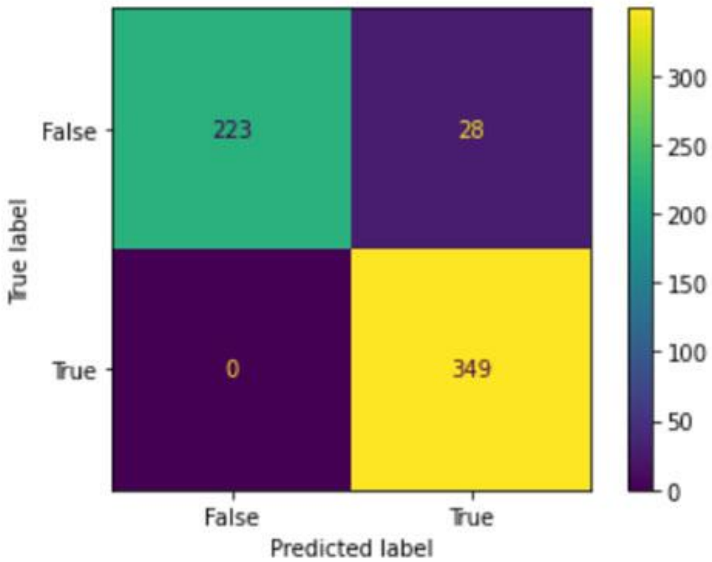


Fig. 6 Confusion matrix of Random Forest in the classification process

6. Conclusion

This research carried out the classification of shrimp and fish that live underwater using feature extraction and machine learning. From the research results, it was concluded that the feature extraction used was the morphometric characteristics of length, height, and width. From the learning machine used, the results of the random forest method were the best, with true positive = 349, true negative = 223, and false positive = 28. Future research will be developed for classification using deep learning

Acknowledgments

The authors are grateful to the Doctoral program of Information Systems, School of Postgraduate Studies, Diponegoro University, the Information Systems Study Program, Faculty of Engineering, Muria Kudus University, and Directorate of Research, Technology and Community Service, Directorate General of Higher Education, Research and Technology, Ministry of Education, Culture, Research and Technology Indonesia, for supporting this study

Author Contributions: All authors have read and agreed to the published version of the manuscript. All authors significantly contributed to this study. Arif Setiawan.: was in charge of collecting, preprocessing, analyzing the process, and compiling data using programming language, followed by testing and writing the research. Hadiyanto: was responsible for supervising the research, resource management, and correcting and approving manuscripts. Catur Edi Widodo: was in charge of analyzing the algorithm, observing the novelty, correcting the grammar, and approving the manuscript. All authors have read and approved the final published manuscript.

Funding: This research was funded by Directorate of Research, Technology and Community Service, Directorate General of Higher Education, Research and Technology, Ministry of Education, Culture, Research and Technology Indonesia.

Conflicts of Interest: The authors state that there are no conflicting problems in this study, professional problems, finances and personal conflicts that may affect the research process.

References

Ap, Dineshbabu. 2016. "MethodologMethodologies for Studying Finfish and Shellfish Biology Methodologies for Studying Finfish and Shellfish Biology Geetha Sasikumar Prathibha Rohities for Studying Finfish and Shellfish Biology Methodologies for Studying Finfish and Shellfish Bio." (July).

Chaikaew, Pasicha, Natcha Rugkarn, Varot Pongpipatwattana, and Vorapot Kanokkantapong. 2019. "Enhancing Ecological-Economic Efficiency of Intensive Shrimp Farm through in-out Nutrient Budget and Feed Conversion Ratio." *Sustainable Environment Research* 1(1): 1–11.

Do, Huu-luat, and Thong Quoc. 2022. "Climate Change Adaptation Strategies and Shrimp Aquaculture : Empirical Evidence from the Mekong Delta of Vietnam." *Ecological Economics* 196(September 2021): 107411. https://doi.org/10.1016/j.ecolecon.2022.107411.

Gao, Wei, and Zhi Hua Zhou. 2020. "Towards Convergence Rate Analysis of Random Forests for Classification." *Advances in Neural Information Processing Systems* 2020-Decem: 103788. https://doi.org/10.1016/j.artint.2022.103788.

Ghane, Mostafa, Mei Choo Ang, Mehrbakhsh Nilashi, and Shahryar Sorooshian. 2022. "Enhanced Decision Tree Induction Using Evolutionary Techniques for Parkinson's Disease Classification." *Biocybernetics and Biomedical Engineering* 42(3): 902–20. https://doi.org/10.1016/j.bbe.2022.07.002.

Giri, Subhasis et al. 2023. "Revealing the Sources of Arsenic in Private Well Water Using Random Forest Classification and Regression." *Science of the Total Environment* 857(October 2022): 159360. https://doi.org/10.1016/j.scitotenv.2022.159360.

Huang, Ing Jer et al. 2019. "The Prototype of a Smart Underwater Surveillance System for Shrimp Farming." *Proceedings of the 2018 IEEE International Conference on Advanced Manufacturing, ICAM 2018*: 177–80.

Isa, Iza Sazanita et al. 2020. "CNN Transfer Learning of Shrimp Detection for Underwater Vision System." *Proceeding - 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering, ICITAMEE 2020*: 226–31.

Karim, Sohail et al. 2021. "IoT Based Smart Fish Farming Aquaculture Monitoring System." *International Journal on Emerging Technologies* 12(2): 45–53. www.researchtrend.net.

Liu, Haoyang et al. 2021. "A High-Density Fish School Segmentation Framework for Biomass Statistics in a Deep-Sea Cage." *Ecological Informatics* 64(July): 101367. https://doi.org/10.1016/j.ecoinf.2021.101367.

Liu, Z., Xiaojun Jia, and Xinsheng Xu. 2019. "Study of Shrimp Recognition Methods Using Smart Networks." *Computers and Electronics in Agriculture* 165.

Mou, Dan, Zhuwen Wang, Xili Tan, and Shaoyun Shi. 2022. "A Variational Inequality Approach with SVM Optimization Algorithm for Identifying Mineral Lithology." *Journal of Applied Geophysics* 204(December 2021): 104747. https://doi.org/10.1016/j.jappgeo.2022.104747.

Patel, Jigisha M. 2016. "A Review on Feature Extraction Techniques in Content Based Image Retrieval." : 2259–63.

Ray, Sunuram et al. 2021. "Role of Shrimp Farming in Socio-Economic Elevation and Professional Satisfaction in Coastal Communities of Southern Bangladesh." *Aquaculture Reports* 20(June 2020): 100708. https://doi.org/10.1016/j.aqrep.2021.100708.

Saberioon, Mohammadmehdi, and Petr Císa. 2018. "Automated within Tank Fi Sh Mass Estimation Using Infrared Re Fl Ection System." 150(September 2017): 484–92.

Singla, Sandeep Kumar, Rahul Dev Garg, and Om Prakash Dubey. 2020. "Ensemble Machine Learning Methods to Estimate the Sugarcane Yield Based on Remote Sensing Information." *Revue d'Intelligence Artificielle* 34(6): 731–43.

Spirkovska, Lilly. 1993. "Three-Dimensional Object Recognition Using Similar Triangles and Decision Trees." *Pattern Recognition* 26(5): 727–32.

Triantafyllou, Anna, Dimosthenis C. Tsouros, Panagiotis Sarigiannidis, and Stamatia Bibi. 2019. "An Architecture Model for Smart Farming." *Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019*: 385–92.

Uddin, Md Galal et al. 2022. "Robust Machine Learning Algorithms for Predicting Coastal Water Quality Index." *Journal of Environmental Management* 321(July): 115923. https://doi.org/10.1016/j.jenvman.2022.115923.

Valenti, Wagner C. et al. 2021. "Aquaculture in Brazil: Past, Present and Future." *Aquaculture Reports* 19(December 2020): 100611. https://doi.org/10.1016/j.aqrep.2021.100611.

Wang, Wenwen et al. 2022. "Improved Depth of Field of the Composite Micro-Lens Arrays by Electrically Tunable Focal Lengths in the Light Field Imaging System." *Optics and Laser Technology* 148(December 2021): 107748. https://doi.org/10.1016/j.optlastec.2021.107748.

Yang, Yu et al. 2021. "Multispectral Image Based Germination Detection of Potato by Using Supervised Multiple Threshold Segmentation Model and Canny Edge Detector." *Computers and Electronics in Agriculture* 182(September 2020): 106041. https://doi.org/10.1016/j.compag.2021.106041.

Yoshikazu, Gabriel, Patricia Krecl, and Admir Créso. 2022. "Science of the Total Environment Fine-Scale Modeling of the Urban Heat Island : A Comparison of Multiple Linear Regression and Random Forest Approaches." 815.

Zermane, Abderrahim et al. 2023. "Predicting Fatal Fall from Heights Accidents Using Random Forest Classification Machine Learning Model." *Safety Science* 159(November 2022): 106023. https://doi.org/10.1016/j.ssci.2022.106023.

Zhang, Lu, Jianping Wang, and Qingling Duan. 2020. "Estimation for Fish Mass Using Image Analysis and Neural Network." *Computers and Electronics in Agriculture* 173(April): 105439. https://doi.org/10.1016/j.compag.2020.105439.