HOKKAIDO UNIVERSITY

Title	Study on applying machine learning into Vietnamese shrimp aquaculture [an abstract of dissertation and a summary of dissertation review]
Author(s)	NGUYEN, KHIEM MINH
Degree Grantor	北海道大学
Degree Name	博士(水産科学)
Dissertation Number	甲第14757号
Issue Date	2022-03-24
Doc URL	https://hdl.handle.net/2115/86092
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	NGUYEN_MINH_KHIEM_abstract.pdf, 論文内容の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文内容の要旨

博士の専攻分野の名称：博士（水産科学）　　氏名：NGUYEN MINH KHIEM

Study on applying machine learning into Vietnamese shrimp aquaculture

(ベトナムのエビ養殖業を対象とした機械学習による魚病と価格形成の予測に関する研究)

1. Introduction

Shrimp aquaculture in Vietnam contributes significantly to national economic growth and improves the livelihoods of local people. However, the spread of various diseases poses a critical threat to shrimp-culture systems, causing huge economic losses for farmers. Shrimp diseases are difficult to treat following outbreak, such that large shrimp farming areas are rapidly affected, with high mortality rates. Shrimp diseases impact local incomes and reduce the volume of shrimp-product exports, increasing the economic risk of shrimp aquaculture in Vietnam. Vietnamese shrimp exports compete fiercely with those of other exporting countries in the global market. Approximately 90% of shrimp produced in Vietnam is destined for export. Vietnamese shrimp export prices are influenced by the competitive prices of other exporters and stringent requirements of importers, including anti-dumping laws and traceability and food-safety requirements that cause barriers to export. Recently, computer-based aquaculture practices have been promoted due to their efficiency. Machine learning (ML) is a powerful computer technique that is well suited to prediction problems. Accurately predicting disease occurrence is essential to minimizing shrimp-farming losses, and predicting shrimp-product export prices is helpful for understanding global market trends and designing export strategies. The application of powerful ML-based prediction models to address these questions will greatly benefit the shrimp aquaculture industry in Vietnam, promoting the economy at the local and national levels.

2. Materials and Methods

This study focused on shrimp farms along the east coast of the Mekong Delta, Vietnam. The study area consisted of four shrimp-producing provinces: Ca Mau, Bac Lieu, Soc Trang, and Tra Vinh. Three diseases that seriously impact farmed shrimp were examined: acute hepatopancreatic necrosis disease (AHPND), white spot

disease (WSD), and disease caused by the parasite *Enterocytozoon hepatopenaei* (EHP). To predict AHPND outbreaks, we used a dataset containing 763 samples, which comprised symptoms, environmental factors, and related farming data. Prediction models were designed to incorporate ML algorithms including logistic-regression, artificial-neural-network (ANN), *k*-nearest-neighbor (KNN), and decision-tree algorithms. AHPND, WSD, and EHP occurrence was also predicted according to geographical factors using a geographic information system (GIS), data collected from 182 farms and 72 hatcheries, and ML algorithms including ANN, logistic-regression, random-forest, and gradient-boosting algorithms. The results of these predictions were used to map areas at risk of infection with AHPND, WSD, and EHP, using GIS.

ML was also used to predict prices of Vietnamese shrimp exported to the US market based on data collected by the United States Department of Agriculture (USDA), the World Trade Organization (WTO), the Food and Agriculture Organization of the United Nations (FAO), and the International Monetary Fund (IMF) between May 1995 and May 2019. The random-forest and gradient-boosting-tree algorithms were used to predict shrimp export prices based on selected periods of historical data (2, 3, 6, or 12 months). To determine the impact of competition on exports, we applied a super-learner algorithm, which combined 10 different ML algorithms to improve the prediction results compared with the single-algorithm ML approach. The dataset included information for six leading shrimp export competitors: China, India, Indonesia, Thailand, Ecuador, and Chile. The relative importance of variables in the predictive model to export prices was interpreted through SHapley Additive exPlanations (SHAP) analysis.

3. Result

ML-based prediction of disease outbreaks in shrimp

ML-based model predictions of AHPND outbreaks showed that logistic regression outperformed all other algorithms, obtaining accuracy rates of 85.50% and 83.04% in the hold-out and cross-validation tests, respectively. Hepatopancreatic atrophy markedly affected model predictions, and high water temperature and high salinity were associated with a higher risk of AHPND outbreak. The ANN performed best among ML algorithms in predicting the occurrence of shrimp disease based on geographical factors, with accuracy rates of 91.88%, 83.78%, and 75.67% for AHPND, WSD, and EHP, respectively. Interpolation using GIS showed that Soc Trang Province was the region most heavily infected with EHP, whereas WSD disease occurred mainly in Bac Lieu and part of Soc Trang, and Tra Vinh and Ca Mau were most heavily infected with AHPND.

ML-based export price prediction

All ML algorithms used to predict shrimp export prices produced mean absolute percentage error (MAPE) rates

< 5.00% and mean squared prediction error (MSPE) rates < 0.04%, although error rates gradually increased from short- to long-term base periods of historical data. The random-forest algorithm performed well for base periods longer than 6 months, whereas the gradient-boosting-tree algorithm performed better for base periods shorter than 6 months. The super-learner algorithm was more accurate and stable than any single candidate algorithm for all base periods. SHAP interpretation showed that Indian, Thai, and Chinese shrimp export prices, WTO membership, and shrimp disease occurrence significantly impacted Vietnamese shrimp export prices, whereas certification requirements were less important.

4. General Discussion

The results of our disease-outbreak-prediction models allowed us to clarify the effects of environmental and geographical factors on the development of disease in shrimp. High water temperature, high salinity, high alkalinity, and shared water sources tended to increase the probability of disease outbreak. Our map of the estimated spread of shrimp diseases will be useful for shrimp farmers to determine suitable locations for new farms. Our analysis of the strengths and weaknesses of Vietnamese shrimp exports elucidated the ability of Vietnam shrimp producers to compete with other exporters in the global market based on various factors, including disease occurrence, global Good Aquaculture Practice (GAP) and Safety Quality Food (SQF) food-safety certifications, and Hazard Analysis and Critical Control Point System (HAACP) product traceability certification. Reducing disease spread and increased certification rates would help to improve product quality and satisfy the requirements of the global market, enhancing the trademark of Vietnamese shrimp worldwide.

A major contribution of this study is the application of advanced computer techniques to the aquaculture industry, both to prevent disease outbreak and boost exports. We also proposed graphic user interfaces (GUIs) to predict shrimp disease outbreaks and Vietnamese shrimp export prices based on input data.

In conclusion, the results of this study will contribute to improving the Vietnamese shrimp aquaculture industry through shrimp disease reduction and increased shrimp product exports in the global market. Minimizing economic losses associated with shrimp diseases will promote the sustainability of local livelihoods and create shrimp export opportunities. In a future study, we will extend our disease-prediction models to other diseases such as yellow head syndrome, white feces syndrome, and opaque muscle disease, and expand our export-prediction model to predict international market trends for other seafood products. We further proposed a blockchain technique for tracing shrimp products from farm to fork, to further enhance the trademark of Vietnamese shrimp.