

DISEASED SHRIMP DETECTION BASED ON TRANSFER LEARNING

Students: Long Phan-Viet, Dinh Pham-Van

18IT5, Faculty of Computer Science; Email: pvlong.18it5@vku.udn.vn, pvdinh.18it5@vku.udn.vn

Supervisor: Ph.D Nga Le-Thi-Thu

Faculty of Computer Science; Email: ltnnga@vku.udn.vn

Abstract: Shrimp farming plays an important role in aquaculture in the central coastal provinces of Vietnam. Shrimp disease is a significant threat to nutritional security and causes considerable economic loss. Identification of infected shrimps in aquaculture remains challenging due to the dearth of necessary infrastructure. The identification timely is an obligatory step to thwart from spread of disease. This paper proposes a technique to detect shrimp diseases based on transfer learning. This work includes three main steps. The first step collects and

preprocess the image dataset of diseased shrimp collected from shrimp farms in Quang Nam province. The second step trains the dataset through three models SVM, VGG16, and the proposed model GonCNN. The third step tests and evaluates the accuracy of these models. Experimental results show that GonCNN has an accuracy of up to 92.93%, while SVM and VGG16 with 75.67% and 86.94% of accuracy, respectively.

Keywords: Shrimp Disease, Identification, Transfer Learning, Deep Learning.

1. Introduction

Shrimp farming plays an important role in aquaculture in the central coastal provinces of Vietnam. Shrimp disease is a significant threat to nutritional security and causes considerable economic loss. Currently, shrimp diseases are still a problem. They are often caused by some virus and the risk of spreading quickly through water to neighboring aquaculture farms. Therefore, it is necessary to diagnose quickly and accurately to control these diseases early.

Identification of infected shrimps in aquaculture remains challenging due to the dearth of necessary infrastructure. Usually, shrimp diseases are analyzed using the experience of shrimp farmers or professionals in fisheries departments. All tasks can be performed manually and observed visually on the surface of the shrimp. This is a large amount of work, requires more time to analyze disease symptoms in shrimp. Because the final conclusion is based on personal skills and experience, the accuracy is not high and the effectiveness is low. Consequently, image-based identification timely is an obligatory step to thwart from spreading of disease, helps to detect diseases quickly, from which early disease treatment plans can be made in time to improve farming productivity and reduce economic losses.

To overcome the above problem, automatic detection and identification of diseased shrimp from digital images has been strongly interested in research and applied. Image-based identification using artificial intelligence has many different approaches. Applying image classification techniques using machine learning, deep learning has been of great interest to many researchers. Duong-Trung et al proposed ImageNet technique for the classification of medicinal plants [1], Luyl-Da Quach show that Support Vector Machines (SVM) achieves the highest classification accuracy (81.27%), TaharatAurpaMd and KalamAzad extracted the involved features to classify the

diseases with the help of SVM algorithm on the salmon fish image dataset to examine the fish disease with 91.42% of accuracy and without augmentation [3]. Bao et al used the Niblack algorithm to detect, identify and eliminate shrimp with YHV yellow head disease [4]. Another work of Ghasemi-Varnamkhasi devoted to WSSV detection in shrimp images using the K-means clustering technique [5]. Some other applications focus on detecting and assessing shrimp freshness, the identification of soft shells, and shrimp sounds [6,7]. These researches have not investigated the transfer learning approach which our work intends to.



Fig. 1. Aquaculture in the central coastal provinces of Vietnam

This paper proposes a technique to detect shrimp diseases based on transfer learning. This work includes three main steps. The first step collects and preprocess the image dataset of diseased shrimp collected from shrimp farms in Quang Nam province. The second step trains the dataset through three models SVM, VGG16, and the proposed model GonCNN. The third step tests and evaluates the accuracy of these models by some evaluation metrics. The achieved results help detect and recognize diseased shrimp quickly and effectively. Early detection and treatment practices could help timely response to emerging shrimp diseases mitigate disease devastation, help farmers have a timely treatment, thus increasing productivity and product quality.

The rest of the paper is organized as follows: Section 2 presents some common diseases in shrimp that this study needs to identify. Section 3 describes identification models used to test and compare, from which to choose the best model to apply to detect diseased shrimp. The proposed identification model is given in Section 4. Experimental results are given in Section 5 and finally the conclusion of the paper.

2. Shrimp Diseases

2.1. Taura syndrome virus/ Taura disease

Taura syndrome is also called the "red tail" disease. When shrimp has this disease, its body and other parts have the color of red or black pink. Shrimp is lazy eating and lethally swimming on water surface or edges of farming ponds. Shrimp hepatopancreas is more yellow than usual with swelling gill. Shrimp usually dies during molting period. The disease is dangerous for shrimp. Its incubation period is short, but can cause 95% of deaths in shrimp. Shrimp dies or sinks at the bottoms of farming ponds, then floats after 2-3 days later. There are many dead shrimps around ponds' edges. The disease is caused by a combination of pathogens, including *Vibrio* Harvae and 3 other viruses. Taura syndrome appears when shrimp is about 2 weeks old until it reaches the mature and molting stages. The disease is potentially acute, making shrimp sick, causing soft shells and destroying shrimp's digestion. The syndrome diffuses and spreads quickly.



Fig. 2. Taura disease

2.2. Yellow head virus/ Yellow head disease

Yellow head disease is a viral infection of shrimp and prawn. The appearance of the disease and mortality often occur within two to four days following an interval of exceptionally high feeding activity that ended in abrupt cessation of feeding. Yellow head disease can be identified by high and rapid mortality in mainly early to late juvenile stages. However, gross signs of infection are not entirely reliable indicators of yellowhead disease and the use of molecular methods for confirmatory diagnosis is required. Because the yellow head disease has caused serious damage and affects international trade, the World Organization for Animal Health (Office International des Epizooties; OIE) puts it on the notifiable list of OIE aquatic animal diseases.



Fig. 3. Yellow head disease

2.3. White spot syndrome virus/ White spot disease

When diagnosing white spot syndrome virus (WSSV) in shrimp, the first step is to determine the source of the sickness so that it may be treated as soon as possible. The WSSV PCR test provides fast and accurate results, and it should be performed as soon as white spots appear on shrimp. Inside the shell, there are numerous white spots ranging in size from 0.5 to 2.0 mm that develop in the cephalothorax, 5th and 6th abdominal segments, and subsequently spread throughout the body. When shrimp are unwell, abnormal behaviors such as loss of appetite and erratic swimming. Most shrimp perish in ponds when white spots appear 3–10 days later. If the PCR result for WSSV is positive, the culture should be harvested right away; if not, the culture can be continued and treatment measures followed as needed.



Fig. 4. White spot disease

3. Identification models

3.1. Support Vector Machine

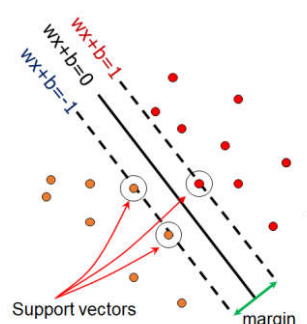


Fig. 5. Support vector machine model

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression. However, it is mostly used in classification problems. In the SVM algorithm, each data item is a point in n -dimensional space with the value of each feature being the value of a particular coordinate. Then, classification is performed by finding the hyper-plane or line that differentiates the two classes very well. Support vectors are simply the coordinates of individual observation. The SVM classifier is a frontier that best segregates the two classes. A discriminative classifier based on optimal separating line, maximize the margin between the positive and negative training examples.

3.2. VGG16

Convolutional Neural Network (CNN) is a neural network consisting of a set of convolution layers that overlap and use nonlinear activation functions such as ReLU and tanh to activate the weights in the nodes. Each class after passing activation functions will generate more

abstract information for the next classes. This is one of the advanced Deep Learning models, helping to build intelligent systems with high accuracy today. CNN is widely used in the problem of object recognition in digital images. During the training of the network, CNN automatically learns the values through the filter classes. In image classification, CNN will try to find the features of the image through the corresponding filters. The identification of image features is usually in the order of point features, edge features, shape features, high-level features, etc., and the last layer is used to classify the image. The CNN model has two characteristics: invariant and associative. With the same object, if this object is projected from different angles (translation, rotation, scaling), the accuracy of the algorithm will be significantly affected.



Fig. 6. VGG16 network architecture

VGG16 is a CNN proposed by K. Simonyan and A. Zisserman. VGG16 network architecture consists of three main components: convolutional layer (conv), pooling layer (pool) and fully connected layer (fc). VGG16 architecture is detailed in Fig6. After training by the VGG16 network, the accuracy is usually above 85%.

3.3. GonCNN

Transfer learning is a new deep learning technique that uses part or all of a pre-trained model. Such pre-trained models are usually trained on one or a few certain datasets, compatible and give high accuracy to one or more tasks (multi-task deep learning) for which it is trained. The use of pre-trained models is a big step forward to inheriting the results of previous models, taking advantage of existing pretrained models to create new models for more specific, applicable target tasks, more practical.

The proposed model GonCNN is based on transfer learning from the above VGG16 model. It reuses convolutional layers in the base model layers and fine-tuning the layers fully connected in the head model layers. Table 1 describes the architecture of GonCNN model

Table 1. Architecture of GonCNN model

Layer (type)	Output Shape	No. of Parameters
image_input (InputLayer)	(None, 128, 128, 3)	0
VGG16 (Functional)	(None, None, None, 512)	14714688
flatten (Flatten)	(None, 8192)	0
fc1 (Dense)	(None, 4096)	33558528
dropout_6 (Dropout)	(None, 4096)	0
fc2 (Dense)	(None, 4096)	16781312
dropout_7 (Dropout)	(None, 4096)	0

predictions (Dense)	(None, 3)	12291
Total params: 65,066819		
Trainable params: 50,352,131		
Non-trainable params: 14,714,688		

4. Diseased Shrimp Identification System

In this paper, the identification of shrimp diseases was tested and compared in parallel on all models SVM, VGG16 and the proposed model GonCNN based on transfer learning from VGG16 to see the optimization of the proposed model. The diseased shrimp identification system is described in Fig 7.

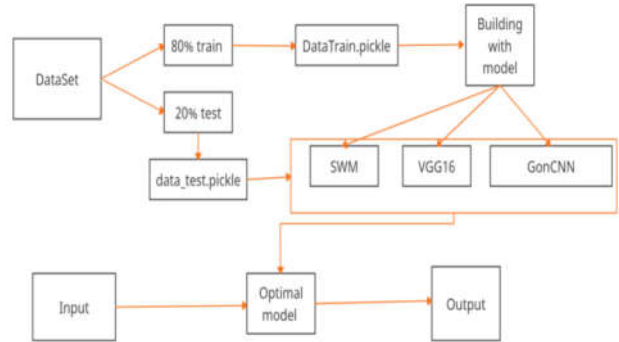


Fig. 7. Shrimp disease identification system

The identification system includes three main parts. The first parts collect and preprocess the image dataset of diseased shrimp collected from shrimp farms. The second step trains the dataset through three models SVM, VGG16 and the proposed model GonCNN. The third step tests and evaluates the accuracy of these models.

The data set after preprocessing to determine shrimp diseases will be divided into two subsets at the rate of 20% (test set) and 80% (train set). To evaluate a model more fully, accurately and objectively, in this paper, the training set is decomposed by k-fold method with k=5. Specifically, the training set data is randomly divided into 5 parts. Then train the model 5 times, each time training will choose 1 part as validation data and the remaining 4 parts as training data. The final model evaluation result will be the average of the results of 5 training times. Next step, three models SVM, Network VGG16 and GonCNN are used to train data, extract features of shrimp images, and compare the achieved results.

GonCNN is based on a transfer learning model in which layers are constructed in the correct order. Conv2D layers retrieved the most accurate information which are the features of the input image. Flatten class is made flat into a vector. Dense is a fully connected layer, all units of the previous layer are connected to all sets of units of the current layer. Some new characteristics of the study include, first, instead of using the sigmoid function, this study uses the softmax function and in the case of analysing diseased shrimp, softmax helps to give results faster, more specific classification and after each class, we will use the dropout function. In plain terms, during random training, there is only one network node. We require dropout because if a fully connected class has too

many parameters and takes up the majority of them during training, the nodes in that layer will become overly reliant on each other, limiting the strength of each node and resulting in overcombining. Another new feature of the research is that instead of giving classification results immediately like previous studies, this research paper added and edited Adam function to give the most optimal function possible, thereby giving more accurate classification results.

5. Experimental results

5.1. Experimental data

To train the recognition system, the paper uses an image dataset of diseased shrimp collected from shrimp farms in Quang Nam province. Data on shrimp diseases are stored in the respective folders, a single disease for each folder. In addition to normal shrimp, the mobile application of this article can detect three common diseases: Taura disease, yellow head disease, and white spot disease.

Table 2. Dataset on shrimp diseases

	Taura	Yellow head	White spot
Train set	578	577	578
Test set	115	114	115

The dataset consists of 1733 images, divided by the ratio of 80% trainset and 20% test set. Then proceed to create a data set and a label from the divided train and test set, save it as a file with the.pickle extension so that it can be reused later. Table 2 describes the dataset in detail.

5.2. Applying the recognition model

SVM model gives an accuracy of 75,67%. Confusion matrix on this model are described in Figure 8.

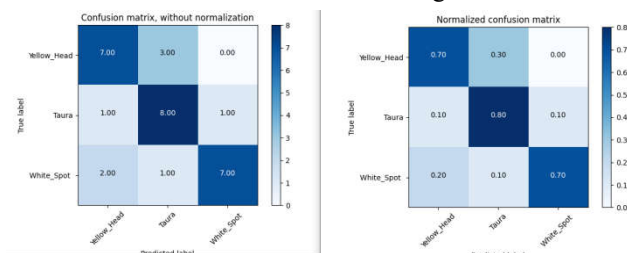


Fig. 8. Confusion matrix on SVM model

VGG16 gives an an ccuracy of 86,94%. The obtained parameters during the training with VGG16 are described in Table 3. The training process of the model built by the VGG16 model is shown in the diagrams in Figure 9.

Table 3. The obtained parameters during the training with VGG16

epochs	accuracy	val_accuracy	loss	val_loss
5	0.7352	0.7579	0.5983	0.5424
10	0.8268	0.7579	0.4049	0.5097
15	0.8571	0.8617	0.3523	0.3220
20	0.8528	0.8732	0.3509	0.3530

25	0.8637	0.9135	0.3306	0.2529
28	0.8694	0.8732	0.3042	0.2953

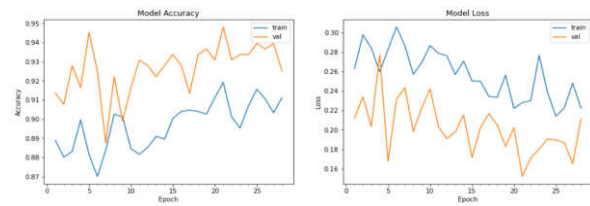


Fig. 9. Accuracy and Loss on VGG16 model

The transfer learning model GonCNN gives an accuracy 92,93%. The obtained parameters during the training with this proposed model are described in Table 4. The training process of the model built by the GonCNN model is shown in the diagrams in Figure 10. Its confusion matrix are described in Figure 11.

Table 4. The obtained parameters during the training with GonCNN

epochs	accuracy	val_accuracy	loss	val_loss
5	0.7583	0.8012	0.5925	0.5288
10	0.8492	0.8703	0.3558	0.3202
15	0.8918	0.9020	0.2730	0.2672
20	0.8824	0.8891	0.2753	0.2273
25	0.9250	0.9395	0.1895	0.1730
28	0.9293	0.8617	0.1801	0.3612

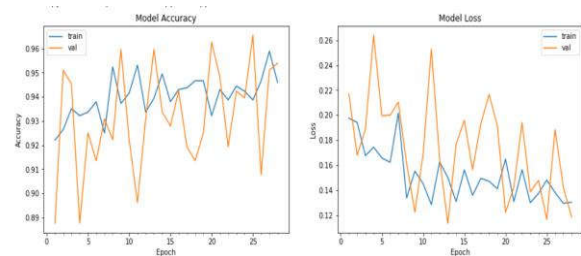


Fig. 10. Accuracy and Loss on GonCNN model

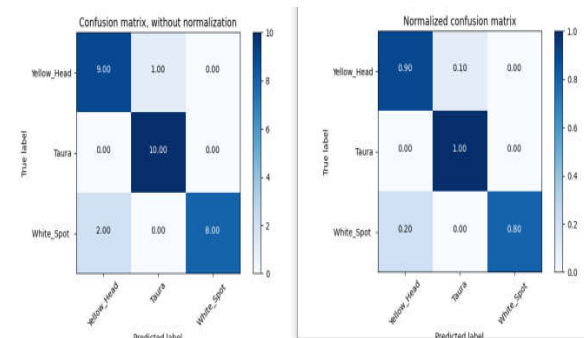


Fig. 11. Confusion matrix on GonCNN model

Recognizing that using SVM model with a accuracy rate is not as good as using the two models VGG16 and GonCNN. With these two CNN models, the GonCNN model outperforms the VGG16 model generated through transfer learning. The table 5 is the results based on the model's accuracy as determined by the above three methods:

Table 5. Accuracy of the models

Model	Accuracy(%)
SVM	75,67
VGG16	86,94
GonCNN	92.93

6. Conclusions

The applications of deep learning model in shrimp disease detection enable quick and accurate disease detection. Three different models (SVM, VGG16, GonCNN) were built, tested, compared, and evaluated. The test results reveal that the proposed model GonCNN outperforms the competition, with a 92.93% percent accuracy rate.

Based on the findings, the model GonCNN was chosen to develop a disease detection system for shrimp using photos captured by a mobile device's camera.

The application can be expanded to include more diseases, allowing the computer to learn, construct, and complete a database of disease information for shrimp. Since then, the application has evolved into a community application that has been deployed in the aquatic system to assist farmers and those interested in diseases of shrimp, as well as other aquatic products, with timely detection so that early decisions can be made in aquaculture care and improvement.

References

- [1]. Duong-Trung, Nghia, Luyl-Da Quach & Chi-Ngon Nguyen. Learning deep transferability for several agricultural classification problems. *Inter. J. of Advanced Computer Science and Applications*, 2019.
- [2]. Luyl-Da Quach, Long Quach Hoang, ghia Duong-Trung, Chi-Ngon Nguyen. Towards machine learning approaches to identify shrimp diseases based on description, FAIR 2019
- [3]. TaharatAurpaMd, KalamAzad. Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture, *Journal of Computer and Information Sciences*, 2021
- [4]. Bao, T.Q., Cuong, T.C., Tu, N.D. & Hieu, L.T. Designing the Yellow Head Virus Syndrome Recognition Application for Shrimp on an Embedded System. *Exchanges: The Interdisciplinary Research Journal*, 6(2), 2019.
- [5]. Ghasemi-Varnamkhasti, M., Goli, R., Forina, M., Mohtasebi, S.S., Shafiee, S. & Naderi-Boldaji, M. Application of image analysis combined with computational expert approaches for shrimp freshness evaluation. *International Journal of Food Properties*,19(10), 2016.
- [6]. G. Nagalakshmi, S. Jyothi, White Spot Syndrome Virus Detection in Shrimp Images using Image Segmentation Techniques, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2013
- [7]. Okpala, C.O.R., Choo, W.S. & Dykes, G.A. Quality and shelf life assessment of Pacific white shrimp (*Litopenaeus vannamei*) freshly harvested and stored on ice. *LWT-Food Science and Technology*, 2014
- [8]. Liu, Z., Cheng, F. & Zhang, W. Identification of soft shell shrimp based on deep learning. In 2016 ASABE Annual International Meeting, American Society of Agricultural and Biological Engineers, 2016
- [9]. Roell, Y. E., Beucher, A., Møller, P. G., Greve, M. B., & Greve, M. H. Comparing a Random-Forest-Based Prediction of Winter Wheat Yield to Historical Yield Potential. *Agronomy*,10(3), 395, 2020