Hue University Journal of Science: Techniques and Technology; pISSN 2588-1175 | eISSN 2615-9732

Vol. 134, No. 2B, 2025, 51–61; DOI: 10.26459/hueunijtt.v134i2B.8147

Fine-tuning deep learning models on microscopic images of liver and intestine cells of shrimps using k-fold cross-validation

Thi Thu Thao KHONG*, Quang Vu BUI

Hue University of Sciences, Hue University, 77 Nguyen Hue, Hue, Vietnam

Abstract. We employ the fine-tuning technique to train compact and efficient deep convolutional neural networks—specifically MobileNet_V2, MobileNet_V3_Small, and MobileNetV3-Large – to classify the nutritional status of farmed shrimp. The classification is based on microscopic images of liver and intestinal cells, enabling rapid and scalable assessment of shrimp health through image-based diagnostics. The experiment was conducted on a dataset comprising 854 cellular images, and used k-fold cross-validation to split the dataset into the training and test sets. The pre-trained MobileNet_V3_Large was fine-tuned on our cellular image dataset using 10-fold cross-validation, achieving the highest mean classification accuracy of 90.89%. This study demonstrates the potential of applying deep learning techniques to the monitoring and nutritional management of farmed shrimp, aiming to enhance productivity in aquaculture operations.

Keywords: Transfer learning, k-fold cross-validation, convolutional neural networks, image classification

1 Introduction

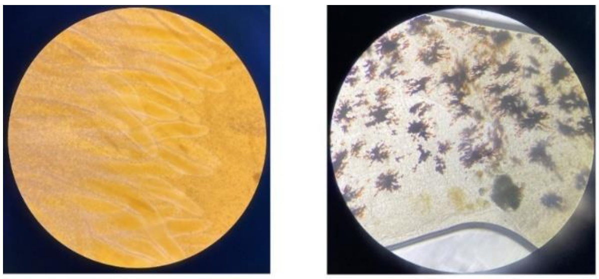
Shrimp farming is an aquaculture sector that plays an important role in the development of Vietnam's fisheries industry. In 2024, Vietnam's shrimp exports reached nearly USD 4 billion, accounting for 39% of the country's total seafood export value. With the policy of promoting the integration of information technology and automation in production, the application of the Internet of Things (IoT) and Artificial Intelligence (AI) has been comprehensively transforming industrial and agricultural production. In the fisheries sector, the application of AI, machine learning techniques, and deep learning models in production and aquaculture has been extensively studied worldwide [1, 2, 3, 4].

In shrimp farming, besides the environment, the feeding process and monitoring of nutritional supply management always play a crucial role in the growth of shrimp. Providing appropriate nutrition for each developmental stage of shrimp will result in high productivity. On the contrary, overfeeding or underfeeding can negatively impact the farming environment, harm shrimp health, and lead to losses in aquaculture productivity. Based on observing the digestive organs, specifically the liver and intestine, farmers can determine whether shrimp are receiving

* *Corresponding: kttthao@hueuni.edu.vn*

Received: 05–12–2025; Accepted: 30–12–2025

adequate, excessive, or insufficient nutrition [5]. With farming experience, shrimp farmers often observe the color of the liver to assess the nutritional status of the shrimp. In addition to color, the shape and fullness of the liver are also important indicators for assessing shrimp health and nutrient absorption efficiency. However, visual observation methods are not highly accurate, so modern shrimp farms often perform dissections and microscopic analysis of liver and intestinal cells (Fig. 1) to assess the nutritional status and health of shrimp more precisely.



(a) (b)

Fig. 1. Microscopic images of liver cells (a) and intestinal cells (b) of normal Litopenaeus vannamei in a shrimp farm

Hu et al [4] and Liu et al. [2] respectively proposed ShrimpNet and Deep-ShrimpNet, which are convolutional neural networks (CNN) to classify shrimp species in the world. In Vietnam, machine learning algorithms, CNN models [6], and the application of transfer learning [7] to shrimp disease classification have been extensively studied. According to domestic and international surveys, we are the first research team to apply CNN to classify microscopic images of liver and intestinal cells of shrimp in order to assess their nutritional status [8].

In this article, we apply transfer learning techniques to state-of-the-art CNNs to classify microscopic images of liver and intestinal cells based on the nutritional status of farmed shrimp. Moreover, k-fold cross-validation is used to estimate the skill of CNNs on our new data. The combination of these two techniques allows us to train models on image datasets of liver and intestinal cells of shrimp more quickly and efficiently. The rest of this paper is ordered as follows. Section 2 introduces related work, including CNN models, transfer learning, and k-fold cross-validation. Section 3 shows the microscopic image dataset of liver and intestinal cells of farmed shrimp. Experimental results are represented in Section 4. The conclusion is in Section 5.

2 Related work

2.1 MobileNet models

To implement in shrimp farms, we prioritize flexible and lightweight CNNs that achieve high accuracy on ImageNet [9] in order to apply transfer learning. We then choose MobileNet architectures. This is a family of lightweight CNNs specifically designed for efficient deployment on edge devices. **MobileNet_V1** [10] introduced depthwise separable convolutions, significantly reducing the number of parameters and computation compared to traditional CNNs. **MobileNet_V2** [11] improved upon V1 by adding inverted residual blocks and linear bottlenecks, enhancing both speed and accuracy. **MobileNet_V3** [12] comes in two variants. **V3_Small is** optimized for low-latency applications on devices with limited resources, and **V3_Large** is designed for higher accuracy while still maintaining efficiency. On ImageNet, the recognition accuracy of MobileNet_V2 and V3 is relatively high despite having a small number of parameters and computational operations. We thus choose MobileNet_V2 and V3 to conduct our experiments.

We employ well-trained models on the ImageNet dataset, with image classification accuracy and parameter counts as described in Table 1.

Table 1. Image classification accuracy (Top-1), number of parameters (Params), and Giga floating point operations per second (GFLOPs) of MobileNet models trained on ImageNet1K_V1 [13].

CNN models	Top-1 (%)	Params (Million)	GFLOPs (Billion)
MobileNet_V2	71.88	3.5	0.30
MobileNet_V3:			
Small	67.67	2.5	0.06
Large	74.04	5.5	0.22

2.2 Transfer learning

Transfer learning enables us to train CNNs more efficiently, with reduced resource consumption and improved training outcomes. By leveraging pre-trained models—often trained on large-scale datasets such as ImageNet—we can significantly accelerate the training process while maintaining high performance. The trained parameters of these models will be used either as a **fixed feature extractor** or as **initialization parameters** for new data [14]. Two types of transfer learning are fine-tuning and feature extraction. Because our new dataset is small and very different from ImageNet, fine-tuning is a technique well-suited for adapting to this dataset. We use pre-trained CNNs on ImageNet and fine-tune them on microscopic images of liver and intestinal cells of shrimp. With fine-tuning, all model parameters are updated slightly to perform well on a new, typically smaller dataset.

2.3 K-fold cross-validation

Cross-validation is a statistical method to estimate the performance of machine learning models on unseen data [15]. Cross-validation ensures that the model performs well on the training data without overfitting. Furthermore, it provides robust evaluation with average results over multiple iterations and efficient utilization of data, especially limited data. In k-fold cross-validation [16], the dataset is divided into *k* subsets of equal size (*k* folds). The model is trained on *k-1* folds and tested on the last fold. This procedure is repeated *k* times, with each fold being used once as the test set. The performance of the model is calculated as the average over *k* iterations.

For example, Fig. 2 shows the training and test set split with 5 folds, and the split process is performed 5 times.

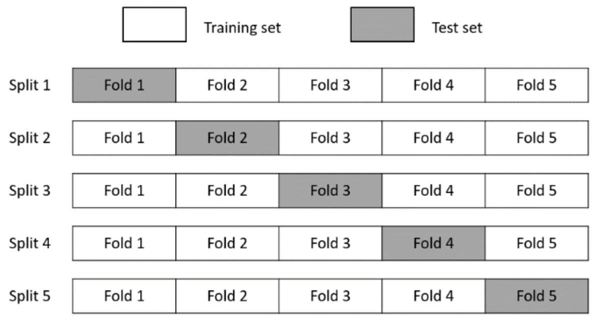


Fig. 2. K-fold cross-validation with k=5

3 Dataset of microscopic images of liver and intestinal cells of farmed shrimp

In order to collect cellular images of the digestive organ in shrimp, aquaculture engineers will divide the rearing ponds into zones and randomly select shrimp samples, then perform surgery to separate the liver and intestines. These aquaculture specialists will extract samples of the liver and midgut for microscopic cellular image analysis.

The 30-day mark is a critical milestone in the shrimp farming process. At this stage, the digestive organs are fully developed, making it a period when shrimp are most susceptible to disease. Therefore, engineers must closely monitor and distribute feed appropriately, as both overfeeding and underfeeding can negatively impact the farming environment and the health of the shrimp. The shrimp samples selected for cellular image analysis were aged above and below 30 days. Based on the expertise of aquaculture specialists, microscopic images of liver and midgut cells will be classified. They are then cropped to remove the excess background. After pre-processing, we label the data as follows:

- *over30_lack_liver, under30_lack_liver:* liver cell images of shrimp aged above and below 30 days showing signs of malnutrition.
- *over30_normal_liver, under30_normal_liver:* liver cell images of shrimp aged above and below 30 days representing normal development.

- *over30_redundant_liver, under30_redundant_liver:* liver cell images of shrimp aged above and below 30 days displaying signs of overnutrition.
- *over30_normal_intestine, under30_normal_intestine:* intestinal cell images of shrimp aged above and below 30 days demonstrating normal development.

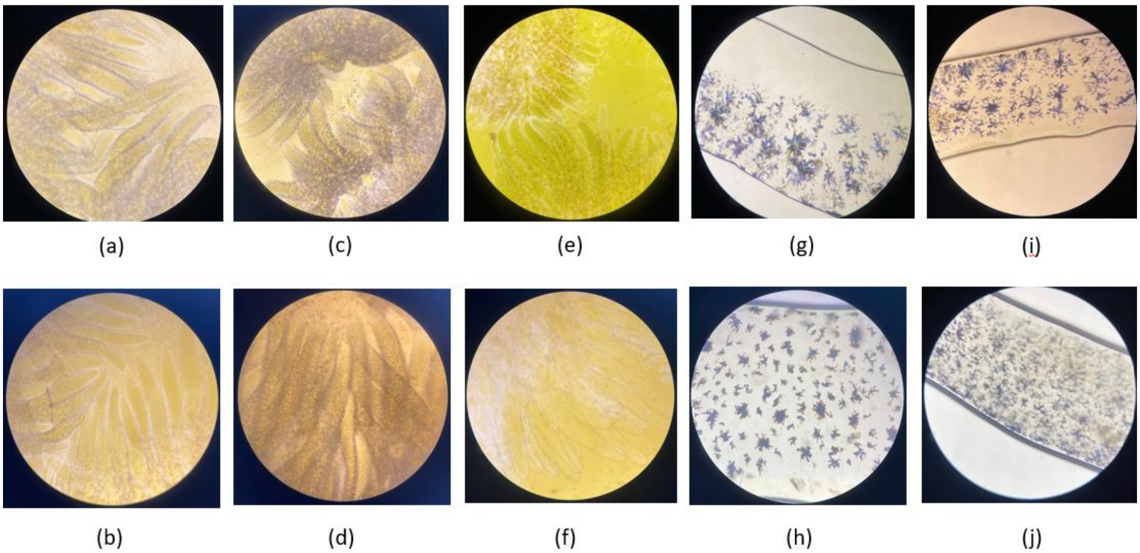


Fig. 3. Description of 10 labels: (a) under30_normal_liver, (b) over30_normal_liver, (c) under30_redundant_liver, (d) over30_redundant_liver, (e) under30_lack_liver, (f) over30_lack_liver, (g) under30_normal_intestine, (h) over30_normal_intestine, (i) under30_abnormal_intestine, (j) over30_abnormal_intestine.

Table 2. Labelling 854 microscopic images of liver and midgut cells from farmed shrimps

Label	Number of images
over30_lack_liver	87
under30_lack_liver	54
over30_normal_liver	119
under30_normal_liver	127
over30_redundant_liver	48
under30_redundant_liver	51
over30_normal_intestine	75
under30_normal_intestine	97
over30_abnormal_intestine	73
under30_abnormal_intestine	123

- *over30_abnormal_intestine, under30_abnormal_intestine:* intestinal cell images of shrimp aged above and below 30 days showing abnormal development.

A dataset of 854 microscopic images of liver and midgut cells, which were randomly collected from farmed shrimps, is labelled as in Table 2. An illustration of the 10 labels indicated in Fig. 3. We apply k-fold cross-validation to split this dataset into the training and test sets.

4 Experiment and Result

4.1 Experiment Setup

We carry out the experiment on an Intel Core i9-10940 CPU and a Quadro RTX8000 GPU. Our program is implemented in the Python programming language on the PyTorch framework.

So as to apply transfer learning, we use models of MobileNet_V2, MobileNet_V3_Small, and MobileNet_V3_Large that were trained on ImageNet1K_V1. Model architectures and pre-trained weights are downloaded from the TorchVision library of PyTorch. Moreover, the ImageNet1K_V1 dataset includes 1000 labels, then the last fully connected layer of these pre-trained models has 1000 outputs for image classification. However, our dataset only has 10 labels, so we changed the size of the last fully connected layers of the three pre-trained models to 10 outputs.

In the experiment, we train each model with fine-tuning and split the dataset with k-fold cross-validation. The choice of k affects the trade-off between computation and variance of performance. If the value of k is small, the computation is faster, but the variance of performance is also higher. On the contrary, with a large k, the variance is lower, but the computation cost is higher. Typically, one executes k-fold cross-validation using $k = 5$ and $k = 10$. Both of these two values are used for our data split to make the comparison, and the choice fitted the model.

4.2 Experiment Results

We sequentially train each model using the fine-tuning method on microscopic images of liver and midgut cells of farmed shrimps. The models include MobileNet_V2, MobileNet_V3_Small, and MobileNet_V3_Large, which were trained on ImageNet1K_V1 as shown in Table 1. We proved [8] that this training process is prone to overfitting, then k-fold cross-validation is applied to separate our dataset into the training and test sets for k training times. Fig. 4 describes training accuracy and loss of MobileNet_V3_Small when applying the fine-tuning technique and 5-fold cross-validation on the dataset of liver and midgut cells of farmed shrimps. The model is configured with a learning rate of 0.001 and trained for 100 epochs. The split 5 achieves better accuracy and loss than the split 1. This indicates that increasing the number of folds leads to higher training accuracy, and fine-tuning also accelerates the training process.

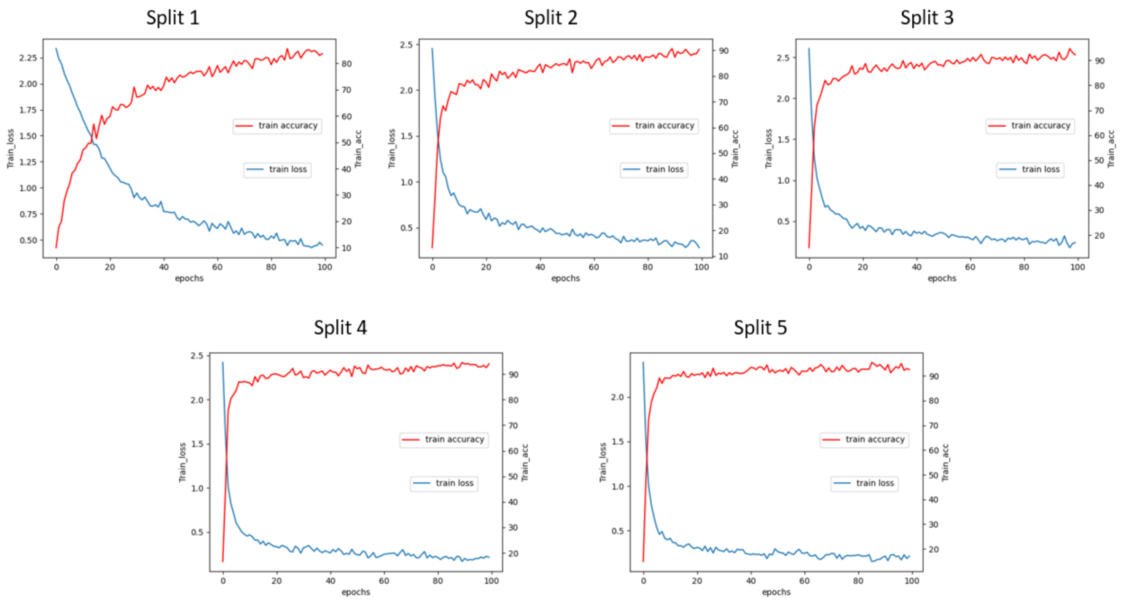


Fig. 4. Training loss and accuracy of MobileNet_V3_Small when using fine-tuning and 5-fold cross-validation

Table 3. Test loss and accuracy of MobileNet_V3_Small when the model is trained with 5-fold cross-validation. Average test loss and accuracy, and training time of 5 times with 100 and 200 epochs once.

Split	Loss		Accuracy (%)	
	100 epochs	200 epochs	100 epochs	200 epochs
1	1.2829	1.0999	60.00	67.25
2	0.7586	0.6326	71.35	78.36
3	0.5554	0.3975	79.53	89.47
4	0.5400	0.3625	82.46	91.23
5	0.4723	**0.2120**	83.63	**93.53**
Average	0.7218	**0.5409**	75.39	**83.97**
Training time of 5 splits (seconds)			**6416**	12801

Table 3 shows the test loss and accuracy, and training time of 5 times of MobileNet_V3_Small when the model is trained in 100 and 200 epochs for each split. The test accuracy of the split 5 is the best, e.g., 83.63% of the model trained 100 iterations for each split, and 93.53% of the model trained 200 iterations for each split. The average test accuracy of 5 splits is 75.39% and 83.97% respectively. The training time of 5 times with 200 epochs for each split is twice the training time of 5 times with 100 epochs for each split.

Table 4 represents the test loss and accuracy of MobileNet_V3_Small when using 10-fold cross-validation for the data division to train and test the model. Compared to Table 3, we notice that the average test accuracy of this model when training 10 times with 100 iterations once decreased by 2.33% compared to the result of the model trained 5 times with 200 iterations once, while the model trained 10 times with only 100 epochs once consumes more training time.

Table 4. Test loss and accuracy of MobileNet_V3_Small when the model is trained with 10-fold cross-validation. Average test loss and accuracy, and training time of 10 times with 100 epochs once.

Split	Loss	Accuracy (%)
1	0.9189	63.95
2	0.8507	74.42
3	0.8623	73.26
4	0.2909	91.86
5	0.5298	81.18
6	0.4286	84.71
7	0.4700	87.06
8	0.6137	80.00
9	0.3607	89.41
10	**0.2824**	**90.59**
Average	**0.5608**	**81.64**
Training time of 10 splits (seconds)		**14417**

Table 5. The average test loss and accuracy, and training time of models on microscopic images of liver and midgut cells of farmed shrimp using k-fold cross-validation with k = 5, 10.

Model	**5-fold**				**10-fold**		**Training time** (seconds)		
	Loss		**Accuracy (%)**		**Loss**	**Accuracy (%)**	**5-fold**		**10-fold**
	100 epochs	200 epochs	100 epochs	200 epochs	100 epochs		100 epochs	200 epochs	100 epochs
MobileNet_V2	0.5155	0.4486	83.97	87.13	0.3581	90.43	6752	13464	13718
MobileNet_V3:									
Small	0.7218	0.5409	75.39	83.97	0.5608	81.64	6416	**12801**	14417
Large	0.3953	0.4474	88.18	87.60	0.2970	**90.89**	6638	13282	14804

In the experiment, we use three pre-trained models on ImageNet1K_V1, e.g., MobileNet_V2, MobileNet_V3_Small, and MobileNet_V3_Large, as shown in Table 1, to apply the fine-tuning technique for training models on our new data. Fine-tuning helps the model achieve better performance and faster convergence on the new data, as seen in Fig. 4. The training of MobileNet_V2 and MobileNet_V3_Large is implemented as MobileNet_V3_Small with the same hyperparameters. The average test loss and accuracy, and training time of models on our dataset are summarized in Table 5. MobileNet_V3_Small is the smallest model; its recognition ability is thus not as good as the other two models, although its training time is the lowest, 12801 seconds. 10-fold with 100 epochs per training time delivers better classification accuracy in MobileNet_V2 and MobileNet_V3_Large. Compared to 5-fold with 200 epochs per training time, the accuracy of the 10-fold method increases by 3.3% in both MobileNet_V2 and MobileNet_V3_Large. Although both models require a significant growth of training time, the improvement in accuracy is substantial. With the large model, such as MobileNet_V3_Large, a boost in the number of iterations and *k* makes the costly computation. Particularly, the training time of MobileNet_V3_Large is about 400 seconds longer than that of MobileNet_V3_Small. However, the accuracy of MobileNet_V3_Large is improved by a mean of 8%. Therefore, the choice of *k* and the number of training epochs must balance between the accuracy and the computation cost. The result of Table 5 shows that the deeper architecture yields higher classification accuracy, and k-fold cross-validation with $k = 10$ also proves to be effective on our dataset. 10-fold cross-validation with 100 epochs per training time ensures a balance between the classification accuracy and the training time for models.

Table 6. Precision, recall, and F1-Score for each class are averaged in 10 folds with MobileNet_V3_Large

Metrics	over30_abnormal_intestine	over30_lack_liver	over30_normal_intestine	over30_normal_liver	over30_redundant_liver	under30_abnormal_intestine	under30_lack_liver	under30_normal_intestine	under30_normal_liver	under30_redundant_liver
Precision	0.9800	0.9292	0.8863	0.8683	0.8327	0.9158	0.9375	0.9325	0.9092	0.9714
Recall	0.9317	0.9218	0.9298	0.8017	0.8133	0.9779	0.9106	0.8468	0.9411	0.9644
F1-score	0.9492	0.9226	0.9030	0.8208	0.8107	0.9422	0.9144	0.8714	0.9207	0.9703

In addition to evaluating classification accuracy, we utilized additional metrics such as precision, recall, and F1-score to further assess the classification performance for each class. Due to the imbalanced dataset, using these additional metrics helps identify if the model is performing well on all categories or just the majority ones. Table 6 demonstrates the average precision, recall, and F1-score for each class of our dataset, which was tested on MobileNet_V3_Large trained in 10 folds with 100 epochs. These results demonstrate that the classification accuracy is well-balanced across all classes. Our method not only achieves high overall accuracy but also ensures robust performance across every category, despite the imbalanced nature of the dataset. In summary, the consistency of these metrics confirms that our method successfully mitigates the impact of data imbalance without sacrificing overall accuracy.

5 Conclusion

In this paper, we collected the dataset of 854 microscopic images of liver and midgut cells of farmed shrimp and applied deep learning for classifying this dataset. We used the fine-tuning technique to train the CNN models more efficiently. Using k-fold cross-validation ensures the model performs well on our new data without overfitting. In particular, we fine-tuned three state-of-the-art CNNs, including MobileNet_V2, MobileNet_V3_Small, and MobileNet_V3_Large, that were well trained on ImageNet1K_V1. 10-fold cross-validation is capable for separating our dataset to train and test the CNN models. The average test accuracy of MobileNet_V3_Large achieves the best result, e.g., 90.89% on the new data with 10-fold cross-validation and 100 epochs per training time. Moreover, our method demonstrates a superior ability to generalize across all categories, making it a highly effective solution for this classification task despite the challenges of the imbalanced data. This is an initial success in implementing deep learning applications for automating the nutritional monitoring process of farmed shrimp in aquaculture farms.

Acknowledgment

This research was supported by the science and technology project of Hue University, DHH2025-01-227.

References

1. Sun M, Yang X, and Xie Y, "Deep learning in aquaculture: A review," J. Comput, vol. 31, no. 1, pp. 294-319, 2020.
2. Liu H, Ma X, Yu Y, Wang L, and Hao L, "Application of deep learning- based object detection techniques in fish aquaculture: A review," Journal of Marine Science and Engineering, vol. 11, no. 4, p. 867, 2023.
3. L. Z, "Soft-shell shrimp recognition based on an improved alexnet for quality evaluations," Journal of Food Engineering, vol. 266, p. 109698, 2020.
4. Hu W.C, Wu H.T, Zhang Y.F, Zhang S.H, and Lo C.H, "Shrimp recognition using shrimpnet based on convolutional neural network," Journal of Ambient Intelligence and Humanized Computing, pp. 1-8, 2020.
5. "Biogency," [Online]. Available: https://biogency.com.vn/yeu-to-anh-huong-den-kha-nang-tieu-hoa-cua-tom/.
6. Quach L.D, Hoang L.Q, Trung N., and Nguyen C.N, "Towards machine learning approaches to identify shrimp diseases based on description," in Kỷ yếu Hội nghị KHCN quốc gia lần thứ XII về nghiên cứu cơ bản và ứng dụng công nghệ thông tin FAIR, Hue City, Vietnam, 2020.

7. Duong-Trung N., Quach L.D., and Nguyen C.N, "Towards classification of shrimp diseases using transferred convolutional neural networks," Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 4, pp. 724-732, 2020.

8. K. T. T. Thảo, H. Đ. Long, và P. H. Phong, "Phân loại dinh dưỡng tôm dựa vào ảnh tế bào gan và ruột sử dụng kỹ thuật học chuyển đổi," in Hội nghị quốc gia lần thứ XXVII về điện tử, truyền thông và công nghệ thông tin REV-ECIT, Ha Noi, Vietnam, 2024.

9. "ImageNet," [Online]. Available: https://image-net.org/index.php.

10. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... and Adam, H., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

11. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C., "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

12. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H., "Searching for mobilenetv3," in Proceedings of the IEEE/CVF international conference on computer vision, 2019.

13. "Pytorch. Models and pre-trained weights.," [Online]. Available: https://pytorch.org/vision/stable/models.html.

14. "Stanford. CS231n Convolutional Neural Networks for Visual Recognition.," [Online]. Available: https://cs231n.github.io/transfer-learning/.

15. D. Berrar, " Cross-validation," 2019.

16. T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," Statistics and Computing, vol. 21, no. 2, pp. 137-146, 2011.