

Behavior Prediction of *Vespa mandarinia* based on Convolutional Neural Networks

Shihao Wu *

School of Business Administration, Guizhou University of Finance and Economics, Guiyang, Guizhou, China

* Corresponding author Email: wish_542@163.com

Abstract: *Vespa mandarinia* poses a significant threat to honey bees and beekeepers. This paper aims to accurately predict the presence and behavior of *Vespa mandarinia* to enable effective control strategies. We develop an "Asian hornet queen prediction model" to forecast the dispersal path and number of queens, identifying their concentration in Bellingham, Washington, and Bellingham, Canada. Using a convolutional neural network, we achieve 96.13% accuracy in identifying *Vespa mandarinia* images. Analyzing confidence levels of unprocessed and unverified labels reveals a significant number of high-confidence samples. Incorporating human control factors into the model, we find 60% human intervention to be most effective in reducing the number of Asian hornet queens. Ultimately, our research highlights the potential for anthropological or ecological measures to eliminate *Vespa mandarinia* populations, aiding in resource allocation for proactive management.

Keywords: Component; Convolutional Neural Networks; Poisson Distribution; *Vespa Mandarinia*.

1. Introduction

Vespa mandarinia, also known as the Asian hornet, is native to southeastern Asia, including Japan, China, and India. *Vespa mandarinia* is a predator of other insects and has a particular affinity for feeding on bees from August to October, causing significant damage to hive and leading to the destruction of entire bee colonies within a short period of time. Consequently, the widespread spread of *Vespa mandarinia* poses a considerable threat to bees and beekeepers. Therefore, it is crucial to accurately identify the presence of *Vespa mandarinia* and further study its physical characteristics and behavioral habits. The objective of this study is to address the challenge of accurately predicting the presence and behavior of *Vespa mandarinia*. To effectively prevent and control the potential impact of *Vespa mandarinia*, it is necessary to assess the reliability of reported information based on photos, videos, and witness comments, enabling the rational allocation of national resources for subsequent investigations.

The significance of studying *Vespa mandarinia*, the Asian hornet, can be understood in three main aspects: pest management strategies, ecological balance, and public safety. Firstly, *Vespa mandarinia* poses a significant threat as a predator to honey bees and other insects. Understanding its distribution range, dispersal rate, and behavioral characteristics is critical for developing effective pest management strategies. Secondly, *Vespa mandarinia* has complex interrelationships with other species, and studying its predatory behavior, feeding habits, and ecological roles can provide insights into food webs and interdependencies in ecosystems. This knowledge is vital for maintaining ecological balance, conserving biodiversity, and ensuring the health of agricultural and natural ecosystems. Lastly, *Vespa mandarinia* is known for its aggressive nature and the potential health risks associated with its stings to humans. By studying the distribution, behavior, and dispersal trends of *Vespa mandarinia*, it is possible to develop public safety measures aimed at reducing human contact and conflicts with this species.

This paper aims to address the challenge of predicting

Vespa mandarinia's presence and behavior using convolutional neural networks (CNNs). The subsequent sections will present the materials and methods employed, the results and discussion of the study, the conclusions drawn from the findings. By accurately predicting *Vespa mandarinia*, this study has the potential to contribute significantly to pest management strategies, ecological balance, and public safety measures.

2. Materials and Methods

To effectively forecast the future spread and reproductive rate of *Vespa mandarinia*, comprehensive information regarding its habits and migration patterns was gathered. This knowledge served as the foundation for the development of the "Asian Giant Hornet Queen Prediction Model," which utilized both the available knowledge and pertinent data. By considering the initial latitude and longitude coordinates, this model offers predictions regarding the temporal changes in the distribution and yearly expansion of *Vespa mandarinia*.

Similarly, in addition to a prediction of the number of *Vespa mandarinia* and their dispersal routes, there is a need to identify *Vespa mandarinia* images for some uncertain and undetected sighting swarms. We originally tried to use a convolutional neural network model for image recognition, but we modified the original model because the Positive ID record was too small and the original model used too many parameters. The incoming parameters and the fit were reduced to make the prediction more accurate.

With the two models we built, we reclassify the more current unidentified, undetermined records to classify the confidence interval and the range of images that may be *Vespa mandarinia*. Using our models to view the growth trend and spread range of *Vespa mandarinia*, the optimal ratio of human intervention is derived, and the number and ratio of *Vespa mandarinia* is reasonably controlled to avoid greater disasters.

2.1. Data Pre-processing

To ensure the accuracy of the results, we pre-processed the data before constructing the model as follows.

Step1: Descriptive statistical analysis [1]. By plotting the

image recognition status, we found that there are more Unprocessed and Unverified image data in the dataset, followed by Negative ID, and Positive ID is relatively few with only 14 valid data.

Step2: General data cleaning [2]. Firstly, the data is aggregated, and after missing values are removed, Lab Status is classified according to attributes to obtain the result of text vectorization.

2.2. Poisson Distribution and Poisson Stochastic Process

After preprocessing the data, we predict the Vespa mandarinia propagation process using a Poisson stochastic process and estimate the prediction accuracy.

If the random variable x takes $0, 1, \dots, n, \dots$. The probability of each value of it is [3]:

$$P\{x = k\} = e^{(-\lambda)} \frac{\lambda^k}{k!} \quad (1)$$

Then x is said to obey a Poisson distribution with parameter λ . The average number of eggs laid by an Asian giant hornet queen each year that can become a queen. In this paper, it is assumed that an Asian giant hornet queen can produce 200 eggs per year, of which 10% can produce queens, since $\lambda = 10$.

The process of generating random numbers using Poisson distribution is as follows [4]:

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad (2)$$

We used the longitude and latitude location of the queen's presence in the first year as the benchmark, and since the queen's activity radius is 30 km, the newly generated hive each year is uniformly distributed within a circle with 30 km radius from its birth point, i.e., the surviving queens after laying eggs are uniformly distributed within 30 km. At the same time, the number of eggs laid was randomly generated with a Poisson distributed random number of the latitude and longitude of the next surviving queen with a radius of 30 km from the latitude and longitude reference value. The average number of eggs laid by Asian giant hornet queens each year that can become Asian giant hornet queens is 10. Formula for calculating the distance between two points using latitude and longitude:

$$L = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{WA-WB}{2}\right) + C}\right) \quad (3)$$

$$C = \cos WA \cos WB \sin^2\left(\frac{JA-JB}{2}\right) \quad (4)$$

$R = 6378137m$, where R is the equatorial radius, WA, WB is the latitude of the two points, JA, JB is the longitude of the two points.

2.3. Convolutional Neural Networks

Since most witnesses mistook other wasps for Vespa mandarinia, we will use convolutional neural networks to build an image recognition model to explore the possibility of predicting misclassification.

When training the image recognition model, the number of positive samples of Vespa mandarinia is small, which leads to a relatively low classification recognition accuracy. Therefore,

we used rotation, enlargement, whitening, width-height displacement transformation, deflation, affine transformation, and mis-cut transformation to increase the number of images, and the data enhancement effect is shown in Figure 1.



Figure 1. Example of data augmentation effect

Since the size of the images of Vespa mandarinia varies and there is no unified specification, it is easy to have odd sample data in the data samples, which affects the training of the network model and may even cause the network model to fail to converge, while the convolutional neural network requires the same size of the input data, so the input images need to be normalized before the network training.

We divide the data set after data enhancement and normalization into training and test sets by 4:1. The number of samples in the training set is 2489 and the number of samples in the validator set is 622 [5].

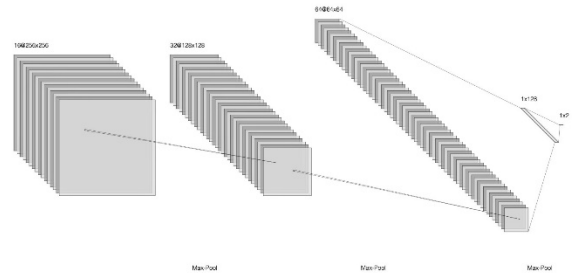


Figure 2. Neural network structure diagram

Convolutional Neural Network (CNN) is a feed-forward neural network, which has artificial neurons that respond locally to the surrounding neurons [6]. Each neuron receives some input and does some dot product computation. It usually contains a convolutional layer, an activation layer, a pooling layer, and a fully connected layer. Our model is shown in figure 2, which can be divided into convolutional layer, pooling layer and dense layer [7]. The convolution formula [8]:

$$(f \otimes g)(n) = \int_{-\infty}^{+\infty} f(\tau)g(n-\tau)d\tau \quad (5)$$

The convolutional layer makes full use of the idea of weight sharing and local correlation to reduce the number of parameters to be trained in the neural network, and the expression of the input-output relationship of the convolutional network [9]:

$$O_j = \sigma\left(\sum_{\substack{dist(i,j) \leq \frac{\sqrt{2}k}{2}}} w_{ij}x_i + b_j\right) \quad (6)$$

Among them, $dist(i, j)$ represents the distance between

i, j, k represents the height and width of the convolutional receptive field (the height and width of this model are equal). w_{ij} is the weight and b_j is the bias. x_i is the pixel value of the corresponding channel.

In the model, three convolutional layers are used, the perceptual field (proper noun) window size is unchanged (3*3), the convolutional kernel is increased exponentially to deepen the extraction of picture features by the network, and the relu function is used for the activation function.

Make the model increase the network training efficiency and model nonlinearity, prevent the gradient disappearance phenomenon in training. The maximum pooling layer exists after each convolution layer, which reduces the feature parameters exponentially and highlights the important feature parameters, and the model is more efficient and accurate. The Dropout layer is introduced to improve the generalization ability of the model and prevent the phenomenon of overfitting. The Dense layer implements the image classification task and the activation function uses the softmax function, and the results are more easily understood.

The optimizer of the model is 'Adam'[10], which is computationally efficient, requires less running memory, and is suitable for solving problems containing very high noise or sparse gradients. The lossy Han formula of the model is a cross-entropy cost function, which makes up for the tendency of saturation of the derivative form of the sigmoid function, while avoiding gradient dispersion. The metric of the model is accurate, which is simple and easy to understand.

3. Results & Discussion

3.1. Distribution Prediction of Vespa Mandarinian Queen

Using the location of Vespa mandarinia in Washington State with Positive ID sightings in 2019 as the reference point, the locations where the Asian giant hornet queen survived were simulated using a Poisson stochastic process approach, we found that the Vespa mandarinia queen's dispersal paths were concentrated in Vancouver, Abbotsford, and Bellingham.

Comparing the predicted distribution of queens from the first to the fourth year, we found that the majority of queens from the first to the second year were distributed in Bellingham, Washington, USA, in the northern part of the state, near the border between the United States and Canada, where the climatic conditions are very suitable for the reproduction of Vespa mandarinia, which can be verified in terms of temperature, precipitation and topography.

While the third to fourth year Asian giant hornet queens are mainly concentrated in Vancouver, Canada, with climatic conditions similar to Bellingham, unlike Bellingham, the area within 3 km of Vancouver is covered by man-made surfaces (100%), the area within 16 km is covered by man-made surfaces (39%) and water (32%), and the 80 km The area within 80 km is covered by trees (44%) and water (32%); trees are mainly located on its fringes, and the fringes of Vancouver are more suitable as a habitat for Vespa mandarinia.

3.2. Convolutional Neural Network Evaluation

The model has overfitting deficiency now, after adjusting the parameters several times, it is found that the questions that the positive sample size is much smaller than the Negative sample, and the Positive sample needs to be collected continuously at a later stage to improve the CNN neural

network recognition ability.

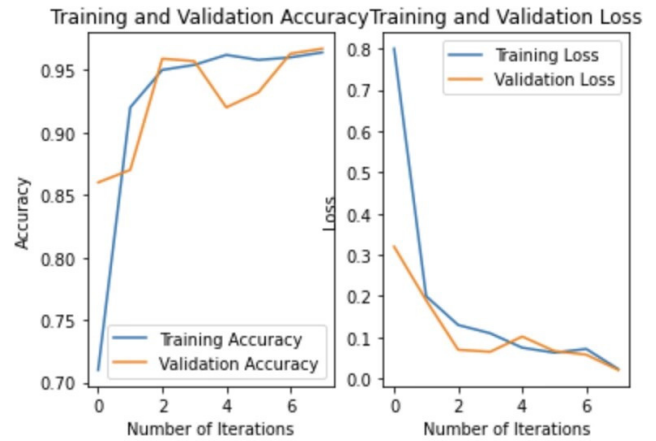


Figure 3. Convolutional neural network model image recognition accuracy vs. loss function graph

We have concluded that as the number of model training increases, the accuracy rate keeps improving and the value of the loss function keeps decreasing.

3.3. Effects of Human Control on the Population Spread of Vespa Mandarinian

On the basis of the first question, we counted the number curve of Vespa mandarinia's natural growth each year without the influence of external conditions. As shown in figure 4, we calculate the growth of each curve, and judge the maximum number of Vespa mandarinias to take measures to limit it. When the number of Vespa mandarinias is small, it is allowed to grow naturally. From this, we can conclude that the period to limit the number of Vespa mandarinias is 2-3 years.

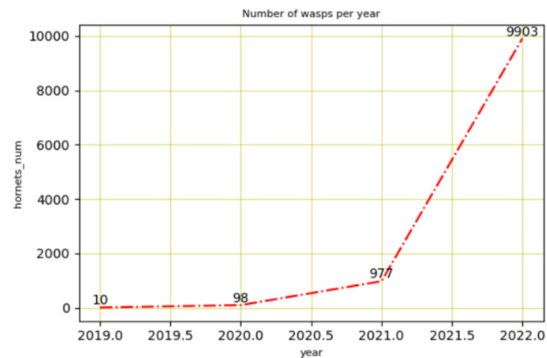


Figure 4. Number curve of natural growth of Vespa mandarinias

Artificial control plays a very important role in the existence of Asian giant hornet queen population. Therefore, based on the number of Vespa s without artificial control for three years, we tested under artificial control conditions of 0%, 40% and 60% respectively, and the results are shown in the figure 5.

From figure 5, It can be seen that under 60% artificial control, number of Asian giant hornet queens has the most significant control effect, and the number of Asian giant hornet queens can be controlled within the range of 40-50. In the absence of artificial control, the number of Asian giant hornet queens was approximately exponential. Vespa mandarinia queen's population grew moderately at a 40% artificial control level.

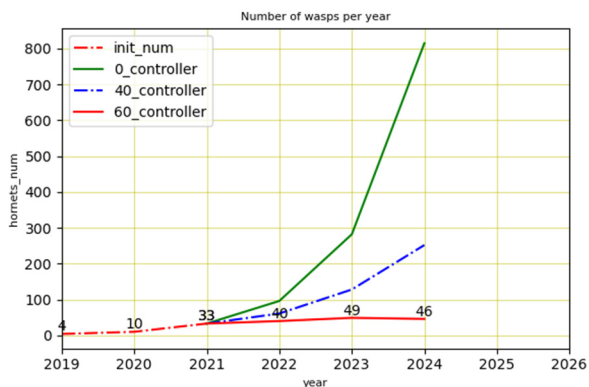


Figure 5. Distribution of Vespa mandarinia queen bee population under different levels of human control

4. Conclusion

we first built the "Asian giant hornet queen prediction model" to predict the spread path and number of Asian giant hornet queens, and found that the queens were mainly concentrated in Bellingham, Washington and Bellingham, Canada; Then, we built a convolutional neural network. After that, we built a convolutional neural network image recognition model to identify images suspected of Vespa mandarinia. The accuracy of image recognition was as high as 96.13% with the increase of model training; Next, we analyzed the confidence level of Unprocess and Unverify labels as Positive labels. The results showed that the confidence level of 489 samples were mainly concentrated in the range of 0.5-0.9, with a higher confidence level. Following that, we added the factor, human control, to the first model. The control effect on the number of Asian giant hornet queens was most significant at 60% human control. Finally, we found that if Vespa mandarinia is controlled anthropogenically or ecologically, its population will be reduced to 0.

References

- [1] M. J. Fisher and A. P. Marshall, "Understanding descriptive statistics," *Australian Critical Care*, vol. 22, no. 2, pp. 93–97, May 2009, doi: 10.1016/j.aucc.2008.11.003.
- [2] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data Cleaning: Overview and Emerging Challenges," in *Proceedings of the 2016 International Conference on Management of Data*, San Francisco California USA: ACM, Jun. 2016, pp. 2201–2206. doi: 10.1145/2882903.2912574.
- [3] H. S. Bakouch, M. Kachour, and S. Nadarajah, "An extended Poisson distribution," *Communications in Statistics - Theory and Methods*, vol. 45, no. 22, pp. 6746–6764, Nov. 2016, doi: 10.1080/03610926.2014.967587.
- [4] D. Johnston, *Random Number Generators—Principles and Practices: A Guide for Engineers and Programmers*. De Gruyter, 2018. doi: 10.1515/9781501506062.
- [5] L. Xu, J. He, and Y. Hu, "Early diabetes risk prediction based on deep learning methods," in *2021 4th international conference on pattern recognition and artificial intelligence (PRAI)*, 2021, pp. 282–286.
- [6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.
- [7] M. Lin and B. Zhao, "The study of molecular ratio trend prediction of electrolytic tank based on deep learning," in *2022 IEEE 5th advanced information management, communicates, electronic and automation control conference (IMCEC)*, 2022, pp. 1457–1460.
- [8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [9] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019, doi: 10.1186/s40649-019-0069-y.
- [10] I. Bello, B. Zoph, V. Vasudevan, and Q. V. Le, "Neural optimizer search with reinforcement learning," in *International Conference on Machine Learning*, PMLR, 2017, pp. 459–468.