Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
C		2122112

A Study on the "Story of Asian Giant Hornets" Based on Intelligence Algorithm

Summary

We conduct a study on the spread pattern of the Asian giant hornets and the "story of Asian giant hornet" based on limited resources, aiming to avoid the negative impacts on the ecological environment of Washington State.

Firstly, the spread pattern of Asian giant hornet populations over time is discussed. Taking the living habits like the queen's nesting range and the workers' maximum motion range as dispersal parameters, the government-controlled number of nest growth as intervention parameters, we improve the traditional cellular automata model. Meanwhile, spread speed and other parameters are considered in the modeling process, in order to facilitate the later modifications based on the number of positive reported sightings. Since Asian giant hornets tend to build their nests in abandoned rodent burrows, we mark the urban areas in Washington State when importing map boundary data, so the probability of Asian giant hornet presence in cities would be reduced. After prediction and validation, the model is proved to have a high precision, and the results can be seen in the Figure.

Secondly, in terms of the scientific classification of hornets, we start with preprocessing the provided data to remove videos, documents and other noises from the image data. For the data enhancement, images with Positive ID and Negative ID are combined, and some images identified as Asian giant hornets are acquired on the website www.alamy.com by using Naver, and then we adopt a random flip method to increase all the Asian giant hornet images, and enhance their details by ESRGAN and CLAHE. Finally, a supervised learning model is built by using convolutional neural network (CNN) approach, and after training and validation, the network is proved to have a low probability of classification error with precision and AUC values of 0.9574 and 0.9546, respectively.

Thirdly, in view of the interpretability of prediction results acquired by CNN, we adopt the LIME (Local Interpretable Model-Agnostic Explanations) method to fit a locally interpretable model to verify the reasonableness of the classification results. In addition, based on the analysis of the living habits of Asian giant hornets and the prediction results of the cellular automata, the reports involved with latitude and longitude at [-122.943,-122.419] and [48.777,49.149], respectively, are most likely to be positive reported sightings in August of each year.

Fourthly, with the newly reported sightings emerging over time, we adopt the no-reference image quality assessment method (NRSS) to remove reports with unclear Asian giant hornet images. NRSS values of the imagines should follow a normal distribution, and reports with NRSS values less than 0.106 are ones to be discarded according to the 3σ principle. The frequency distribution of the reported occurrence time and the prediction results of the cellular automata are counted to derive the mean value of the new report interval, and the training set and model are updated with a frequency of 1.46 days.

Finally, based on the above-mentioned model and negative correlation of intervention time with required intervention intensity, we calculate the control variables for the model, and if the final solutions of the number of nests are convergent, then this pest is proved to be eradicated in Washington State.

key: Cellular Automata; CNN; LIME; NRSS; Data Enhancemen

MEMO

FROM: Team 2122112 , MCM C
To: The Washington State Department of Agriculture
Date: February 8, 2021
Dear Officials:

It is an honor to present to you an analysis of the spread pattern of Asian giant hornets in Washington State. To that end we are writing to you to report the results of our analysis.

Asian giant hornets can wreak havoc on honey bee colonies nearby, as well as harming animals and people. The existing numbers and densities of Asian giant hornets are still low for now, but due to no existence of natural predators yet, Asian giant hornets can become established if not effectively controlled. Our analysis shows that if no effective measures are taken by 2023, there will be more than 20 Asian giant hornet nests and the likelihood of human attacks will increase after the Asian giant hornet infestation.

We build a cellular automata model based on data visualization analysis, and modify it in the later stage by incorporating governmental governance factors to limit the continued growth of the Asian giant hornet population. Through the simulation of the cellular automata, we can provide a minimum threshold of governmental governance, that is, only an eradication of at least 93% of the nests at a time can bring the hornets under control. Also, we deal with the issue of classifying the Asian giant hornets scientifically and effectively by starting with preprocessing the provided data to remove videos, documents and other noises from the image data. In terms of the data enhancement, we adopt a random flip method to increase the number of Asian giant hornet images, and use ESRGAN, CLAHE and other methods to enhance the details of some images. Finally, a supervised learning model is built by using CNN, and after training and validation, the network is proved to have a low probability of classification error with precision and AUC values of 0.9574 and 0.9546, respectively, which shows a bright application prospects. And your side could deploy the model to websites through employing the Django framework to realize online automatic classification. In addition, based on the analysis of the living habits of Asian giant hornets and the prediction results of cellular automata, it is found that the number of workers reach a maximum population in August. Accordingly, reports involved with latitudes and longitudes at [-123.943,-122.419] and [48.777,49.149], respectively, are most likely to be positive sightings in August each year. With the the emergence of newly reported sightings over time, we can adopt a no-reference image quality assessment method to remove reports with unclear images to filter out useless information. The frequency distribution of all reported occurrence times and the prediction results of cellular automata are counted to derive the mean value of the new report interval with an appropriate frequency of 1.46 days to update the training set and model.

Based on the above analysis, we make the following suggestions to control the spread of Asia giant hornets:

- Strengthen publicity efforts to guide residents to identify this species and recognize its hazards. For this purpose we hand-draw the developmental process of the Asian giant hornet that can be used as promotional materials.

- Strict management and control measures should be implemented on logistics and out-of-state imports to prevent Washington State from being invaded by the Asian giant hornets again.
- Establish a mechanism for monitoring, treatment, and eradication. Since complete eradication of the Asian giant hornets cannot be achieved in a short period of time, the government should make a longer-term plan specially for the species, which should include complete monitoring, identification, and eradication of confirmed nests.
- Provide incentives to residents who provide Positive IDs. Local governments should reward residents who provide Positive IDs for their contributions to the control of the Asian giant hornets, which in turn will encourage them to discover more nests of Asian giant hornets.

Sincerely yours

MCM C Team 2122112

Contents

1 Introduction

1.1 Background

The Asian giant hornet is an invasive species from eastern Asia, and a nest was found and destroyed on Vancouver Island in 2019. When winter comes, the current nest is abandoned, all males and workers die out, and only the queens survive. The queens seek out burrows in the dirt, rocks and trees where they will overwinter. And the colder the temperature is, the tighter they will cling. After the natural selection, overwintering queens emerge in the spring, building a nest in some abandoned rodent burrows. Asian giant hornets are less social than ants with its colony naturally disintegrating in the winter. Then in the spring, the queens who have survived the winter begin a new colony, and multiple queens build nests together and lay eggs in the nest chamber, feeding on tree sap during their first delivery. At the same time, these queens fight each other for the highest rank. When dominance hierarchy once is established, the highest-ranking female monopolizes the right to lay eggs. The rest of the females can only be subordinate if they do not fly away, and their eggs can even be eaten by the highest ranking female, which explains why the number of nests established in the following spring is fewer than the number of queens survived in the previous year. In addition to this, the queens can also be unfertilized, as up to 65% of the queens fail to fertilize due to the queens fighting off the males during the mating process of the previous year's departure from the nest.

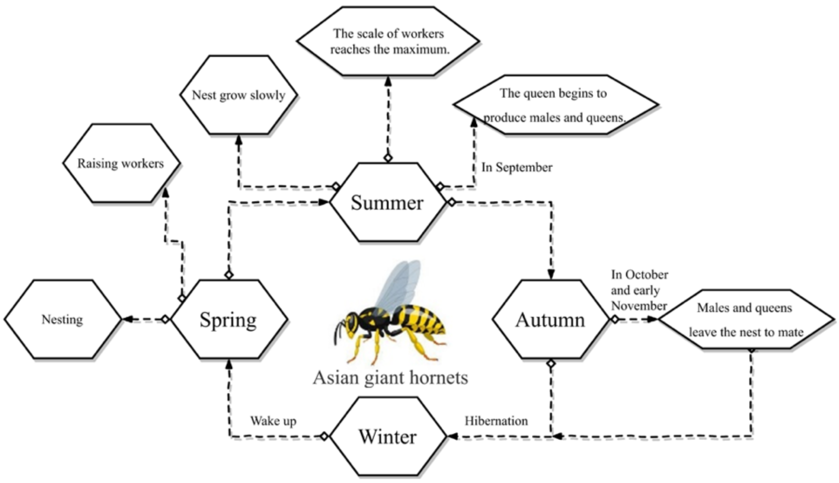


Figure 1: The life cycle of the Asian giant Hornet

As shown in Figure 1, the nests grow slowly through the spring and summer until they reach a maximum population of 100 workers in August. The queen begins producing males and queens in September, and males and queens leave the nest to mate in October and early November. Both fertilized and unfertilized queens overwinter, but only the fertilized queens continue to search for new nests the following year. After the males and queens are produced and begin to leave, the colony is thrown into disarray and they eventually die off as winter approaches.

1.2 Restatement of the Problem

The Asian giant hornet colony was originally discovered on Vancouver Island, British Columbia, Canada, but the nest was destroyed early on and the incident quickly spread throughout neighboring Washington state due to the notorious reputation of "killer hornet". Not all reported sightings are precise, however, there are many mistaken sightings. Due to its severe impact on local honey bee populations, and given that American honey bees do not have the same suicide defense mechanism as Asian honey bees to form a "bee ball", the country sets up helplines and a website for people to report sightings of these hornets in hopes of using limited resources to investigate the spread pattern of the Asian giant hornets, protect honey bees from them, and alert beekeepers as soon as possible.

Therefore, we are facing the following problems:

- To predict the spread pattern of this pest over time based on the available Asian giant hornet data, such as time of sighting, latitude and longitude, and mating patterns, and to assess the precision of the predictions.
- A large number of reported sightings mistake other common hornets for Asian giant hornets. A hornet classification model should be built based on the provided data set files and image files, and the probability of prediction error, sensitivity, and AUC values should be analyzed and discussed.
- To validate the results of model classification to prove that the Positive type in the results is correct, and conclude when and where the predictions are precise and where they are not.
- To optimize the mode, developing a model update mechanism including update methods and update frequency in the case of newly reported sightings emerging over time.
- To find evidence that the Asian giant hornet population in Washington State has declined to a deemed safe range with the mentioned model.

1.3 Our works

First, we investigate the spread pattern of Asian giant hornet populations over time, visualizing reported sightings with the Positive ID on a map of Washington State, which contains the changing status of individuals over time. We also fully research the nesting range and living habits of Asian giant hornets, analyze the geographic features of Washington State, and improve the traditional cellular automata model by employing the provided data and validate the precision of the model.

Second, we preprocess the provided data to remove some noises. For data enhancement, we acquire more labeled Asian giant hornet images, train the convolutional neural network by methods such as random flip and detail enhancement, and evaluate the classification precision of this network using metrics such as AUC, specificity and sensitivity.

Third, we adopt the LIME method to interpret the classification results of the CNN, and discuss the time and location of the most likely positive reported sightings based on the analysis of the living habits of Asian giant hornets and the prediction results of cellular automata model.

Then, in terms of the newly reported sightings that emerge over time, we use the NRSS method to remove the reports with poor image quality of Asian giant hornets. We also analyze the frequency distribution of all reported occurrence times and the prediction results of the cellular automata to discuss the method and frequency of updating the model.

Finally, based on the above model, if the final solutions are convergent, the intervention time as well as the required intervention intensity are negatively correlated. If the relevant data demonstrate that the number of detected nests is below 2 in two consecutive years, then this pest is proved to be eradicated in Washington State.

2 Assumptions and Notations

2.1 Notations

Here are some of the symbols in this article and their meanings.

Symbol	Meaning
S^t	The state of the cell at time t
N^t	The state of the neighboring range of the cell at time t
f	Cellular transform regular function
lr	Learning Rate
$Epoch$	Number of training network iterations
Y_i	Image standardization results
A	Linear transformation parameters of the gamma transform
$H_p(q)$	The value of the loss function of the network
TP	Number of correctly classified positive samples
FP	Number of negative samples that were incorrectly marked as positive
TN	Number of negative samples correctly classified
FN	Number of positive samples incorrectly marked as negative samples
SE	Sensitivity of classification models
SP	Specificity of classification models

2.2 Assumptions

- In the literature provided by the official, it suggests that the maximum motion range of a worker is eight kilometers, so we set the spacing of the new nests to be eight kilometers or more to ensure the adequacy of resources.

- In the model, the probability of a nest presence in the city is low, which is reduced by the queens' habit of nesting in abandoned rodent burrows.
- It is hypothesized that government intervention is the main reason for the decline of hornet colonies. Based on the same literature, it indicates that the Asian giant hornet has no natural predators in North America and will spread rapidly without government intervention. Therefore, the government intervention is incorporated to the model to determine that government intervention is the main cause of hornet colony decline.

3 Model Construction

3.1 Problem One

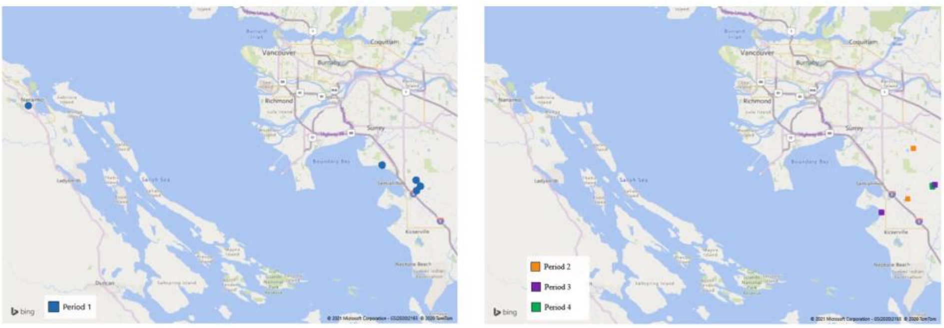
3.1.1 Visualization processing

The Positive ID part of the data was selected, and it was classified according to the year for visual processing. From the picture we can get a lot of information:

After the elimination of some nests in Vancouver, there were no nests in this area in the second year, indicating that the nests would not produce offspring after elimination. And the effect of human intervention on the destruction of nests can be used as one of the rules for the construction of cellular automata.

Since Washington state had four positive sightings in 2019 and nine positive sightings in 2020, this suggests there is governance but not enough.

Washington state officials need to deal with the Asian hornet quickly because it spreads so fast it can easily affect the cities behind it. The graphic periods 1 to 4 in Figure 2 represent the different periods of positive sightings reported.



(a) (b)

Figure 2: The Positive ID map visualization is presented in the data (a)The first year Positive ID map visualization (b)The second year Positive ID map visualization

Table 1 shows the detailed time from Period 1 to Period 4. Since the queen bee is in hibernation, the nest is abandoned and there is no Positive ID in winter, we do not provide a visual map in winter.

Table 1: Legend correspondence table

Legend	Time
Period 1	In the autumn of 2019
Period 2	In the spring of 2020
Period 3	In the summer of 2020
Period 4	In the autumn of 2020

3.1.2 Modeling Cellular Automata for Asian hornet propagation

Cellular Automata (CA) [7] is a grid dynamics model that simulates time and space in discrete state through local interaction. It has strong spatial computing ability and can be used to deduce the changes of the whole region. Cellular automata consists of five main elements: cell, cell space, cell state, cell neighborhood, and transition rules. The basic transformation rule of cellular automata is: the state of the cell at the next time (t+1) is a function of the state of the cell at the last time (t) and the state of the cell in the neighboring range, namely:

$$S^{t+1} = f(S^t, N^t) \tag{1}$$

In formula (1), S^{t+1} is the state of the cell at the next moment, S^t is the state of the cell at the time t, N^t is the state of the neighboring range of the cell at the time t, and F is the transformation rule function, namely, is the local motion rule of the cell.

The main steps are as follows:

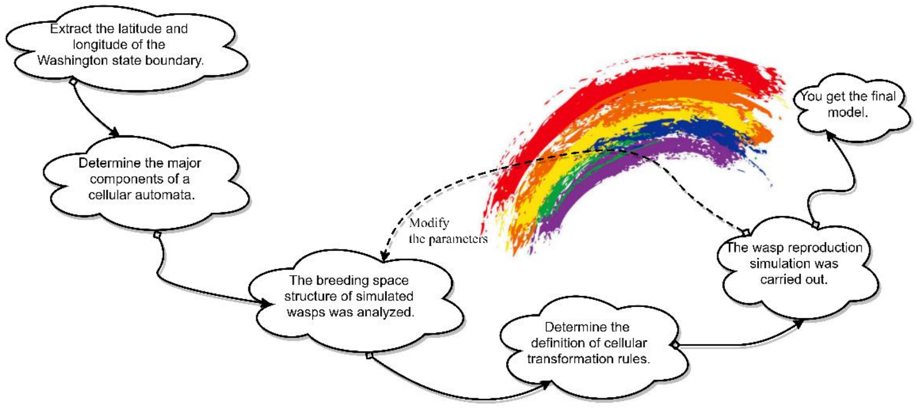


Figure 3: Model Building Steps

1. Extract the latitude and longitude of the boundary of Washington state to get the map of Washington state. At api.map.baidu.com to pick up extraction Washington state boundary latitude and longitude coordinates system, simulate the state of Washington is roughly maps.

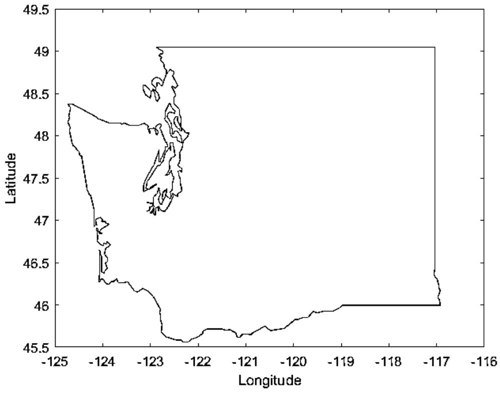


Figure 4: Washington State Map

2. Determine the cell and cell state of the cellular automata. We determined the latitude and longitude of the wasp by screening for positive ID sightings. Then according to the habits of wasps, determine the nearby wasp nest, finally determine the distribution of local wasp nest.Cell states correspond to two states of place: honeycomb free and honeycomb as shown in Table 2.The area state is represented by code.

Table 2: Cellular state

Area state	coding
no nest	0
Have a nest	1

3. The cell neighborhood is selected as the Margolus neighbor. The state of a cell in the CA system at time t+1 is changed from that of other cells in other neighborhoods at time t. Considering the wide range of wasps' activities, the extended form of Margolus neighbor is selected.

4. Determine the parameters of the model: diffusion parameters, propagation parameters and intervention parameters. Our CA model sets the following major parameters to describe wasp reproduction:

- **Spread parameter:** According to the document given by the title, under the rule of natural growth, the range of queen bees' nest is 30km and the maximum range of worker bees' activity is 8km. The spread parameter represents the expansion of the range of honeycomb.
- **Breeding parameters:** under the natural growth rule, the reproduction parameter can be expressed as the probability of survival of the Asian giant hornet by nesting. In the reference materials given in this question, the queen fertilization rate is 35%, but due to the existence of multiple queens nest at the same time and competition, the number of queen bees nests in the second year is less than 35% of the total number of queens.
- **Intervention parameters:** Under the growth rule affected by the intervention, the intervention parameter indicates the probability of survival of the hive under government intervention.

5. Defines the transformation rules for cellular automata. The cellular automata model to simulate wasp reproduction needs to be extended on the traditional model, and the rules are: natural growth principle and growth rule influenced by intervention.

- The principle of natural growth: in the neighborhood of the original wasp community, with the development and reproduction of the community, the community expands outward to generate several new communities, reflecting the agglomeration effect.

$$U(i, j, t+1) = f_1\left[spread, breed, u(i, j, t)\right] \tag{2}$$

- Growth rules affected by intervention: every year, the government takes measures to deal with the existing nests. If the treatment is perfect, the existing hives will be destroyed, otherwise they will not be destroyed.

$$U(i, j, t+1) = f_2[\text{intervene}, u(i, j, t)] \tag{3}$$

3.1.3 Model Calculation and Result Analysis

We have simulated the reproduction of Asian giant hornet for 20 years, and the results are shown in Figure 5:

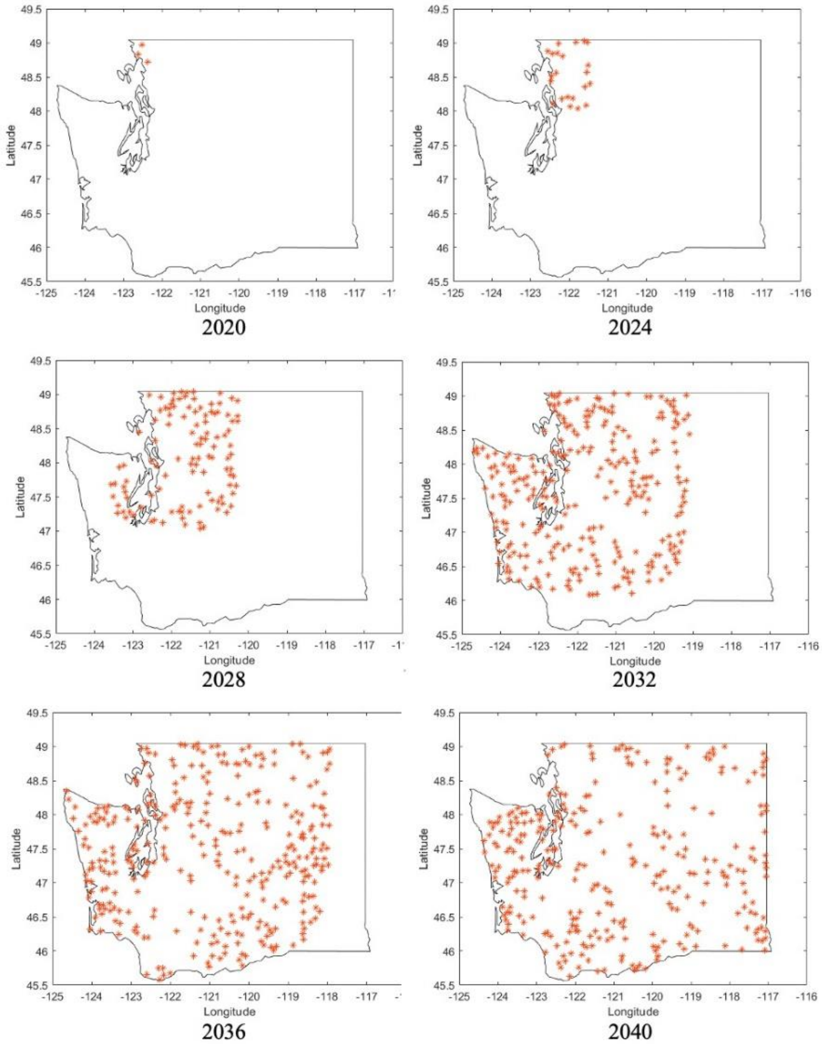


Figure 5: Cellular automata simulation diagram

As shown in Figure 6, the errors of the test set and training set of the model are both small, which meet the error determination criteria and pass the test.

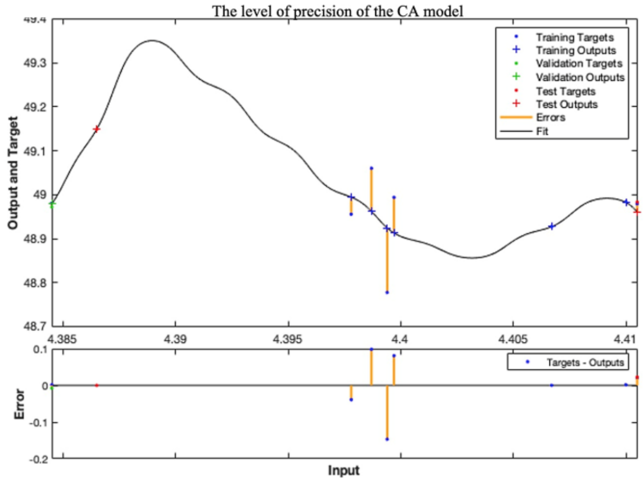


Figure 6: Error between training set and validation set

3.2 Problem Two

3.2.1 Network Architecture

Convolutional neural networks (CNNs) were earlier called neurocognitive machines, a model inspired by the neural mechanisms in the visual system. CNNs is a special feed-forward neural network with features such as weight sharing and local connectivity, where a large number of neurons are organized in a certain way to produce responses to overlapping regions in the visual field. Since the emergence of CNNs in the field of deep learning, they have occupied an important position in large-scale competitions for image recognition and classification, target localization and detection. CNNs are developed from early artificial neural networks and use convolutional operations to solve the disadvantages of artificial neural networks in terms of high computational effort and loss of structural information. To simulate the human visual cognitive function, Fukushima [1] proposed the concept of neurocognitive machine, which is considered as the starting point of CNNs. LeCun [2] built the original LeNet model, which contains convolutional and fully connected layers. After improvement, LeCun [3] proposed the classical LeNet-5 model, which better solved the problem of handwritten digit recognition. This model already contains all the basic modules of modern CNN networks: convolutional layer, nonlinear activation layer, pooling layer, and fully connected layer.

Various CNN-based improvement models have emerged, This experiment uses a more popular building structure for the classification model of Asian giant hornet, as shown in Figure 7. In the order from left to right, the first is the input hornet image, and then the convolutional information is processed by a convolutional layer, and the pooling method is used in this study. The Max Pooling method is used in this study. We do the same process three times, and pass the third processing information into two fully connected neural layers, and finally use a classifier for classification prediction.

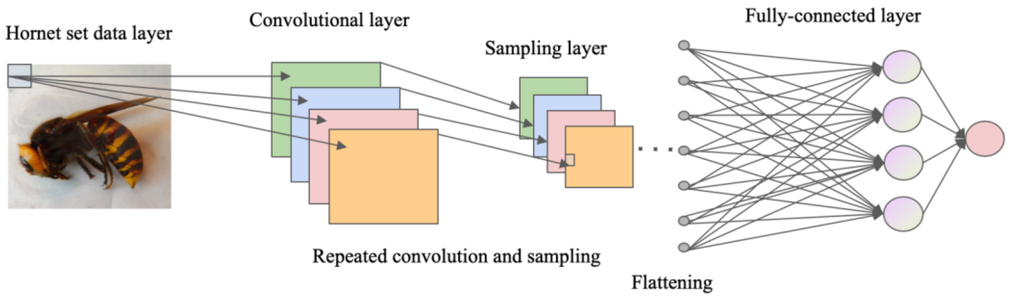


Figure 7: The structure of CNN

3.2.2 Dataset and parameters

Analyzing the dataset provided in the attachment, we found that there were only 14 reports with the attribute Positive ID and 2069 reports with Negative ID. The data of images identified as Asian giant hornets were less, and these data could not guarantee the accuracy of network training, so we used to use the multi-process image finder (Naver) at www.alamy. com website to obtain some images identified as Asian giant hornets, which are high definition with clear outline of Asian giant hornets. the Github address of Naver tool is https://github.com/YoongiKim/AutoCrawler.

Processing the images provided in the attachment, we found that there are also 3 documents of type docx, 8 documents of type pdf, 92 video files and 7 other types of files in these images. There is also a large amount of noise in the remaining images, including images uploaded randomly by users, images of human bodies related to wasps, and images of the external environment with very small wasp size, which will affect the training of CNN. As shown in Figure 8, after removing these noises, we augmented the dataset by randomly rotating it by 90 degrees, 180 degrees, and 270 degrees.

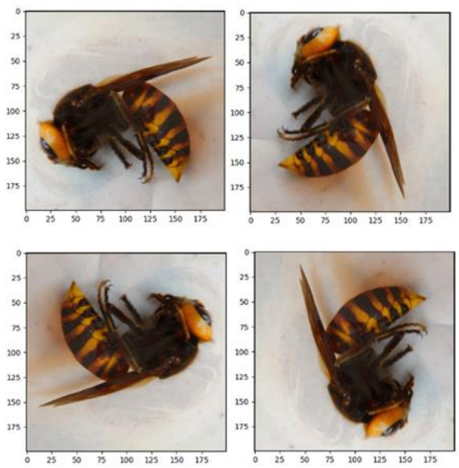


Figure 8: Processing results of random rotation

In Figure 9, there are 1480 other wasp images from the attachment and 370 Asian giant hornets images, both of which become 1480 after data enhancement. Such amount of data is beneficial to improve the accuracy of the supervised learning model.

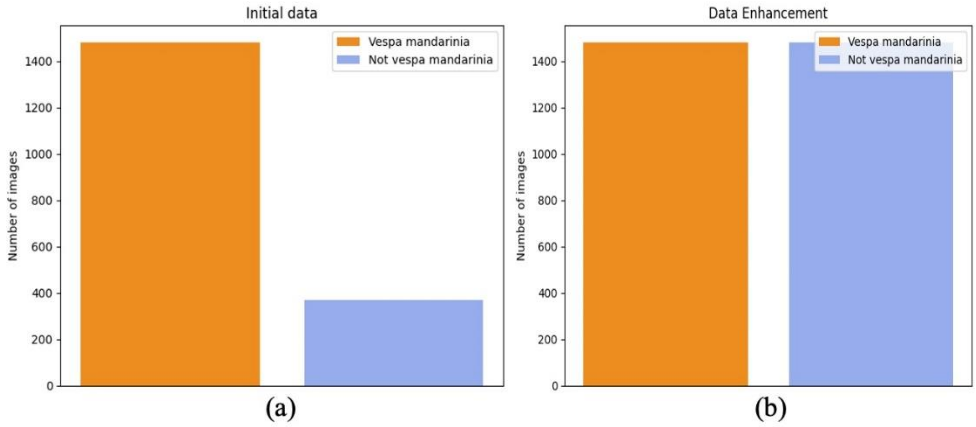


Figure 9: Comparison before and after data enhancement

The experiments were conducted on a platform configured with Intel(R) Xeon(R) Silver 4110 CPU, GPU RTX 2080Ti, RAM 64G, and Ubuntu 18.04, using Python 3.7 + Keras 2.2.5 framework to train the model.

The experiments use Adaptive moment estimation (Adam) as the optimizer when training the model, the learning rate is set to 0.00001, the learning step size (stride_height and stride_width) is 5, the slice size (patch_height and patch_width) is 96, the downsampling factor (subsample) of the output to the input is 500, the number of batches is 25, the number of neurons in the convolutional layer is 128, the total number of convolutional layers is 3, and the final training ends with 8 epochs. Tensorboard and Matplotlib are used to visualize the model training details during the training process.

3.2.3 Image pre-processing

In order to make the neural network model have better learning efficiency and get higher accuracy rate (ACC) at the same time, the image is pre-processed first, and the overall process is referred to Figure 10.

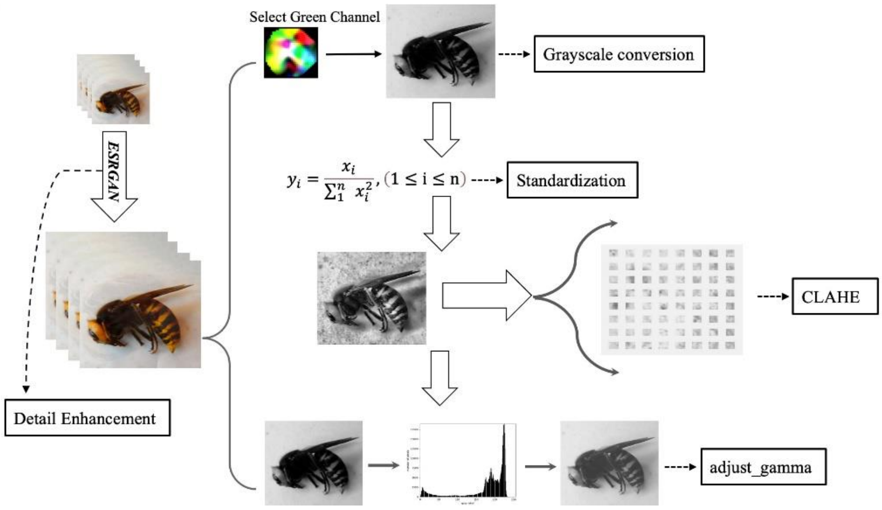


Figure 10: Flow of data processing

Step 1: Detail enhancement. The Asian giant hornet datasets are all RGB images, and each pixel color consists of a mixture of red, green and blue values. To improve the segmentation accuracy, this experiment first processed some of the wasp images using ESRGAN [4] network. After processing, the image contrast increases, the gap between the wasp contour and the background increases, and the noise becomes smaller, which is beneficial to the subsequent classification.

Step 2: Grayscale conversion. The original color wasp images were converted into red (R), green (G) and blue (B) channel images. The contrast between the wasp and the image background in the green channel image is greater and the noise is less, so the data from the green channel is used as the input data.

Step 3: Normalization. The mean and standard deviation of the Asian giant hornet images in the G channel are normalized, and then the forward sequence $x_1, x_2, ..., x_n$ is transformed as follows:

$$y_i = \frac{x_i}{\sum_1^n x_i^2}, (1 \le i \le n) \tag{4}$$

The effect of unit and scale differences between wasp image data features is eliminated by this operation.

Step 4: Contrast-constrained adaptive histogram equalization (CLAHE). This algorithm is used to enhance the contrast of image data. Asian giant hornet images are cut into 64 small blocks of 8 rows and 8 columns, and histogram equalization is used for each image block. CLAHE uses predefined thresholds to crop the histogram for pre-processing purposes, removing image noise and improving image quality.

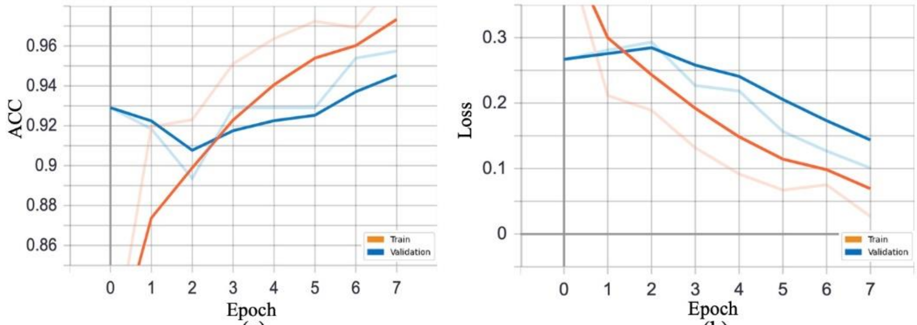
Step 5: Gamma transform. A nonlinear operation used to adjust the intensity of the input wasp image illumination to the input image grayscale values so that the input image grayscale values are exponentially related to the output image grayscale values by:

$$V_{out} = AV_{in}^{\gamma} \tag{5}$$

This processing improves the dark details of the wasp image, through nonlinear transformation, so that the wasp image from the linear response of exposure intensity becomes more similar to the response of the human eye, which is bleached or too dark image, correction.

3.2.4 Hornet classification experiments and model validation

We trained and validated the network in this section, using 70% of the Wasp data for training and 30% for validation. in Figure 11(a), the network achieves an accuracy of 0.9925 on the training set and 0.9574 on the validation set. in Figure 11(b), the network has a Loss of 0.02684 on the training set and a Loss of After the 8th Epoch, the accuracy and loss values level off and the training ends.



(a) (b)

Figure 11: Visualization of training details (a) variation of ACC with increasing Epoch (b) variation of loss with increasing Epoch

The loss function of the training network is chosen as binary cross-entropy:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\big(p(y_i)\big) + (1 - y_i) \cdot \log\big(1 - p(y_i)\big) \tag{6}$$

We substitute the images with attributes Unverified and Unprocessed in the provided dataset into the trained network for testing, and the results are in Figure 12. The network can distinguish the Asian giant hornet from other hornets very well.

Unprocessed				*Not Asian giant hornet*
				*Asian giant hornet*
Unverified				*Not Asian giant hornet*
				*Asian giant hornet*

Figure 12: Some classification results

In Table 3, to test the correctness of the model prediction results, we have selected several commonly used classification performance evaluation metrics:

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$SE = \frac{TP}{TP + FN} \tag{8}$$

$$SP = \frac{TN}{TN + FP} \tag{9}$$

Table 3:Evaluation metrics for hornet classification networks

Method	ACC	AUC	SP	SE
CNN	0.9574	0.9546	0.9026	0.8220

It was tested that the likelihood of model misclassification was low. Figure 13 shows the average ROC curve. The subject operating characteristic curve (ROC) with the false positive rate as the horizontal coordinate and the true positive rate as the vertical coordinate reflects the change in sensitivity and specificity when different thresholds are set for the wasp images, and the average area under the curve (AUC) of the experiment reached 0.9546, indicating that the algorithm classifies better.

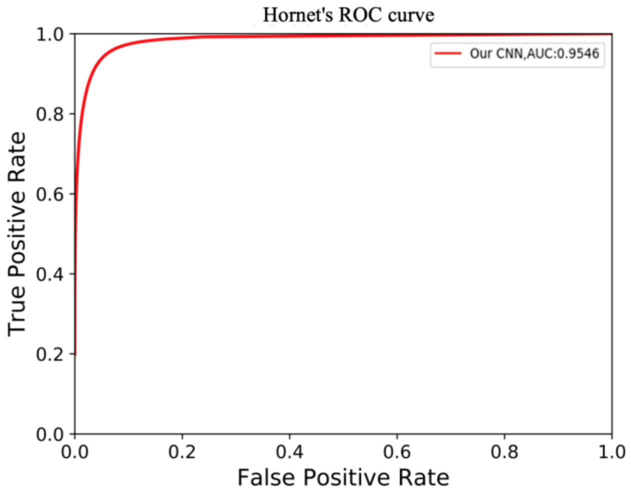


Figure 13: This experiment tests the average ROC curve on the Hornet dataset

3.3 Problem Three

3.3.1 LIME algorithm to explain the CNN model

In order to perceive the CNN model of Asian giant hornet classification, improve the credibility and get the conclusion that the prediction results of the model are generally reasonable, we used LIME (Local Interpretable Model-Agnostic Explanations) [5] method in this section to fit a locally interpretable model to classify the image for interpretation. Equation 10 is the core algorithm description of LIME, where f(z) is the perturbed sample, referring to the predicted value on the d-dimensional space (original features) and taking that predicted value as the target answer, and g(z') is the predicted value on the d'-dimensional space (interpretable features), and then using the similarity as the weight, by linear regression Optimization is performed.

$$\xi(x) = \sum_{z',z\in Z} \pi_x(z)\big(f(z) - g(z')\big)^2 \tag{10}$$

We selected an image of an Asian giant hornet and analyzed it using the above algorithm to interpret the part of the image where the category is most determined. As shown in Figure 14, this image was derived from a pdf type document in the dataset, and the neural network learned the outline of the Asian giant hornet, proving the correctness of the model to classify this image as an Asian giant hornet.

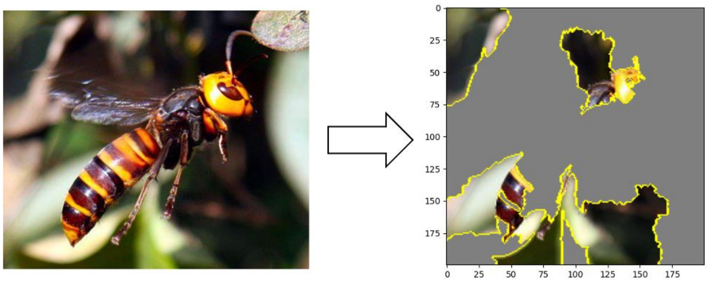


Figure 14: Analysis of locations of interest for hornet classification models based on LIME algorithm

3.3.2 Cellular automata prediction of the likely location of a honeycomb

We imported the geographic location information into the cellular automata and predicted the location of the honeycomb within one to two years. After many predictions to offset the errors, we got a relatively accurate result. The schematic diagram is shown in Figure 5. We chose such a representative figure to illustrate. After calculation, the locations of the honeycomb are within [-122.871,-122.491] and [48.87,49.056].But since the swarm's range is 8 km, the actual sightings should be within [-122.943,-122.419] and [48.777,49.149].

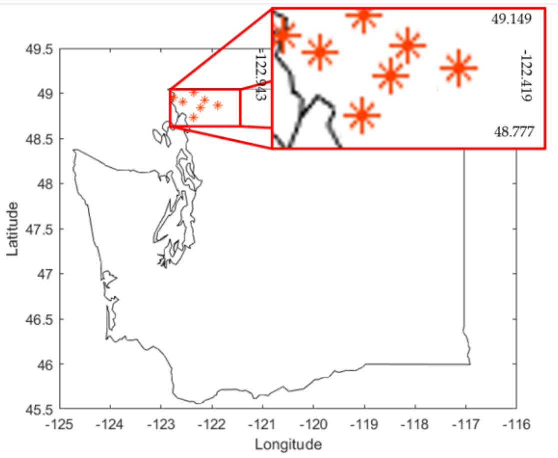


Figure 15: A cellular automaton that predicts the likely location of a honeycomb

3.4 Problem Four

3.4.1 Reference-free image quality evaluation model

Some new reports will appear over time where not all image data are favorable for the Asian giant hornet classification model, and we use a no-reference quality evaluation method (NRSS) [6] for images in this section. Subjective image quality evaluation gives the most correct results though. However, subjective quality evaluation methods require human intervention, are expensive and time consuming, so objective image quality evaluation methods are more suitable for this problem. The methods can be divided into three categories: no reference, partial reference and full reference. The latter two methods require the original image or some features of the original image as a reference to compare with the test image, but in many applications, the original image is not available, so for the Asian giant hornet image quality evaluation problem, we use the NRSS method. In equation 11, the quality of the reference-free structure of the wasp image is calculated based on the N blocks with the richest gradient information found:

$$\mathrm{NRSS} = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathrm{SSIM}(x_i, y_i) \tag{11}$$

As shown in figure 16, we performed NRSS analysis on the 3195 image data in the provided dataset. The quality of these images basically obeys a normal distribution, and according to the 3σ principle, reports with image NRSS values less than 0.106 are rounded off and cannot be used for the training set update of the model.

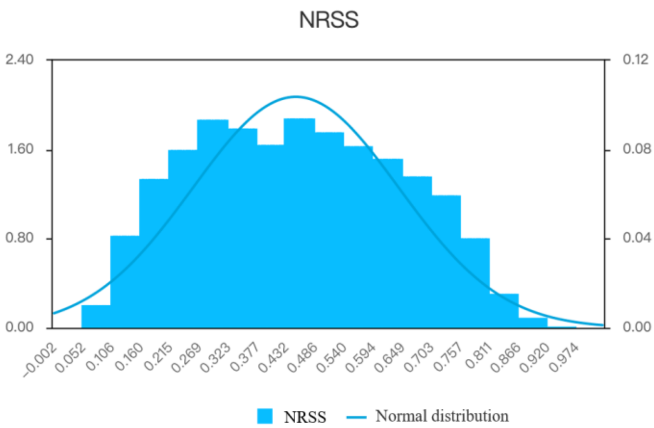


Figure 16: Distribution of NRSS values of Asian giant hornet images

3.4.2Analysis of the temporal frequency distribution of new report

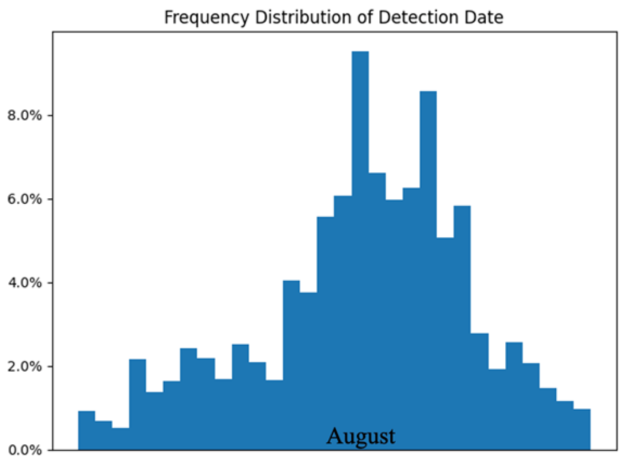


Figure 17: Frequency distribution of the appearance time of new reports

Based on the prediction results of the Cellular Automata, we analyzed the frequency distribution of the arrival of new reports, as shown in Figure 17. The maximum frequency of new reports is 2.4% in August, so we should focus on updating the training set in August. Based on the Markov process, we set the update frequency to the mean value of the new report interval of 1.5 days.

3.5 Problem Five

3.5.1 To solve the train of thought

In this question, we use the model of the first question to solve. The control variable method is used to solve the problem. The control variable method is a method used to reduce variance in the Monte Carlo method [7].This approach turns a multi-factor problem into a multi-factor problem. First, keep other factors unchanged and change one factor, so as to study the influence of this factor on things. Finally, we will work out the solution.

3.5.2 Solving process and conclusion

An analysis of Positive ID suggests that Washington's governance is not so timely. Although Washington has some governance in 2020, things are still not good. First, we fixed the governance year as 2021. As shown in Figure 18, there would be convergence results when the governance efficiency of the government was 0.9 or above. At this point, the control situation is that the pest has been completely eradicated in Washington State.

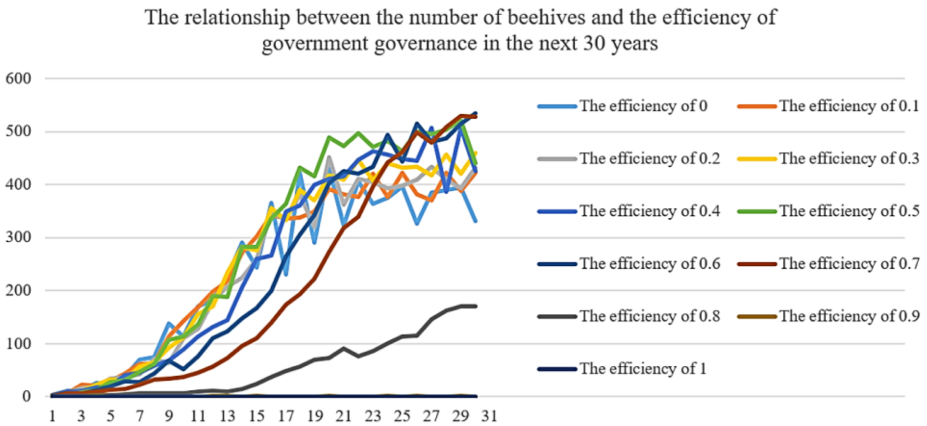


Figure 18: Comparison of the number of beehives in Washington state under different management capacities

However, when the selection of control year is further delayed, the required control capacity should be increased, and the specific convergence year, that is, the year when the pests are eliminated, should be determined by the government control capacity and the government control time. We chose the governance ability as 0.9 and changed the governance year to observe the image changes.

Then we fixed the intensity of government intervention and changed the timing of government intervention. After running the cellular automata model, relevant data are visualized, as shown in Figure 19.

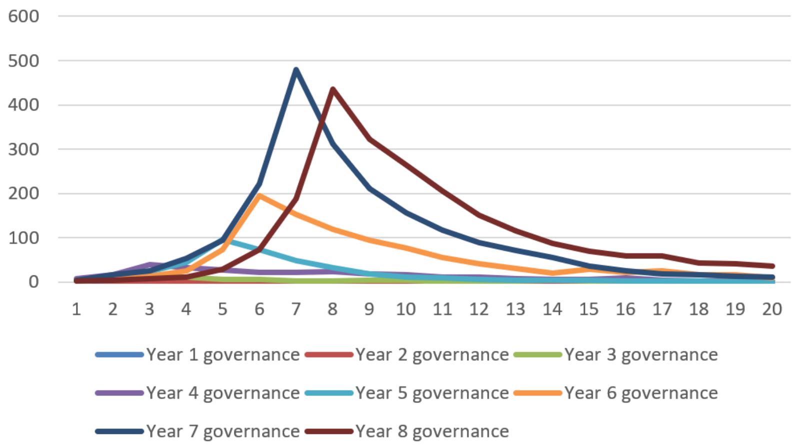


Figure 19: The relationship between the time to start treatment and the number of hives

4 Sensitivity Analysis

We compare the reproduction of wasps under different levels of government intervention by varying the intervention parameters. If we increase the level of government intervention, it will have a limiting effect on wasp reproduction, and the results are in Figure 20. In this section, the stability of the Cellular Automata model is demonstrated.

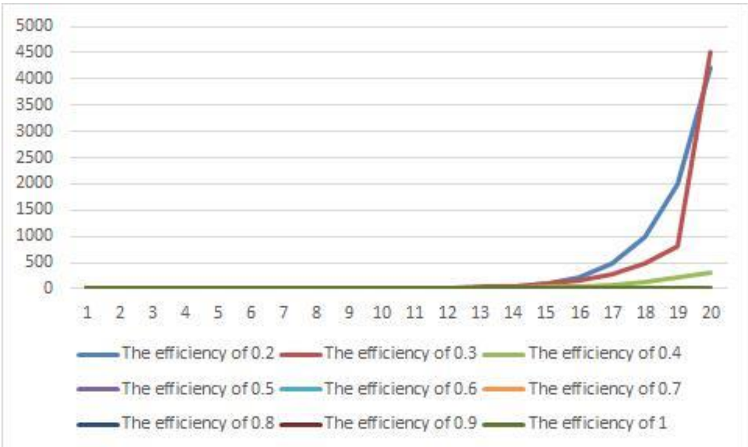


Figure 20: Model sensitivity analysis experiments

When training the neural network, we experimented with the hyperparameters of the model, as shown in Figure 21 below. The orange and blue curves are the best cases for the training and validation sets. Figure 21 (a) and Figure 21 (b) set the Epoch of training to 4, the network is not sufficiently trained and the curves of the training and validation sets do not converge to the level. Figure 21 (c) and Figure 21 (d) are setting the number of convolutional layers to 2. In this case, the ACC and Loss of the training set are better, but the validation set is worse, which is a typical overfitting phenomenon. Figure 21(e) is the curve after modifying the learning rate of training to 0.01, the ACC is lower, and the neural network does not learn the optimal situation at this time because the gradient changes too fast.

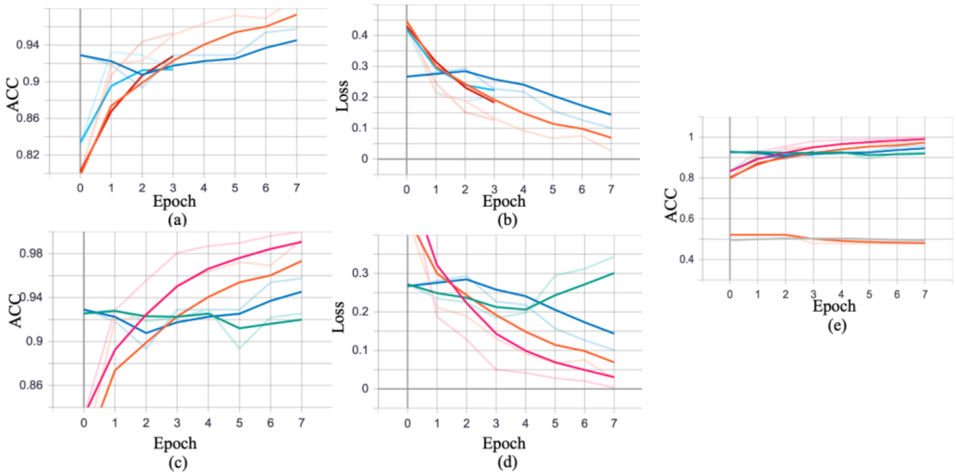


Figure 21: Hyperparametric experiments on classification models of Asian giant hornets

5 Strengths and Weaknesses

5.1 Strengths

- We improved the traditional Cellular Automata model to simulate the reproduction of Asian giant hornet, fully considered the nesting range and living habits of Asian hornets, and analyzed the geographic features of Washington State, thus, the model is more accurate.
- We focused on data enhancement work for image data, including detail enhancement and quantity augmentation of images, and validated the classification results of the model, which has higher AUC values, higher accuracy, specificity and sensitivity, and has good application prospects.
- We used LIME analysis to provide a full explanation of the classification results of the CNN network and improve the confidence level.

5.2 Weaknesses

- In the Cellular Automata model, cellular are relayed according to certain transformation rules. However, in the real world, the reproductive spread of Asian giant hornet does not necessarily spread along its neighbors and may be limited by certain external factors, which leads to less accurate simulations.
- The Asian hornet image data provided in the attachment contains a large amount of noise, such as wasp eggs, natural environment, cell phone screenshots and human images, which will affect the training of the network to some extent. For some wasp images with mutilated or too small targets, the model will some appear to miss detection.

6 Conclusion

An Asian giant hornet colony was discovered on Vancouver Island, British Columbia, Canada, but its nest was destroyed in the early stages. The incident soon spread throughout neighboring Washington State, and there were many reports and images of hornets from residents around the area, but not all sightings were accurate, and there were a large number of false sightings. After our modeling analysis, here are some of our exciting findings:

- We studied the dispersal patterns of Asian giant hornet populations over time, taking into account nesting range, habitat, and geographic factors, and improved the traditional Cellular Automata model to find that the spread of this pest is predictable, our model simulates their spread patterns well in Washington State.
- We built an Asian giant hornet classification model using CNNs, which can effectively distinguish whether an image is an Asian giant hornet or not, the model has high accuracy and can be used on an online website for Asian giant hornet identification. The images with the attribute Unverified in the dataset are substituted into the neural network, A small percentage of images are classified as Asian giant hornet.
- By analyzing the Asian giant hornet habits and Cellular Automata prediction results, reports with latitudes and longitudes at [-122.943,-122.419] and [48.777,49.149], respectively, were most likely to be positive sightings in August of each year.
- Some images in the dataset have low quality and relevance and need to be removed. Also, we counted the frequency distribution of all previous report occurrence times and the prediction results of the Cellular Automata to derive the mean value of the new report interval, we updated the training set and model with a frequency of 1.46 days.
- When the government's treatment efficiency is 0.9 and above, the number of nests of Asian giant hornets solved for has convergence results, after treatment this pest can be considered eradicated in Washington State. To protect bees from it, beekeepers were alerted as soon as possible.

References

[1] FUKUSHIMA K, MIYAKE S.Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1983, SMC-13(5): 267-285.

[2] LECUN Y,BOSER B E, DENKER J S, et al.Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989,1(4): 541-551.

[3] LECUN Y, BOTTOU L, BENGIO Y, etal.Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11): 2278-2324.

[4] Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.

[5] Palatnik de Sousa, I., Maria Bernardes Rebuzzi Vellasco, M., & Costa da

Silva, E. (2019). Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. Sensors, 19(13), 2969.

[6] Xie, X., Zhou, J., & Wu, Q. (2010). No-reference quality index for image blur. Journal of computer applications, 30(4), 921-924.

[7] Wolfram, S. (1983). Cellular automata. Los Alamos Science, 9(2-21), 42.

Appendices for Code

Here is partial code we used in our model, which python and matlab is the main development language.

Appendices A: Asian giant hornet classification using CNN

```
def CNN_model(ft, label):
    ft = ft / 255.0
    feat = ft.shape[1:]
    dense = [1]
    layer = [128]
    convolution = [3] #Number of convolution layers
    epoch = 8 #Number of training iterations
    for i in dense:
        for j in layer:
            for k in convolution:
                name = 'Asian_giant_hornet-{}-conv-{}-layer_size-{}-dense_layer-{}-
Final'.format(k, j, i, int(time.time()))
                tensorboard = TensorBoard(log_dir='log/{}'.format(name))
                cnn = Sequential()# A sequential model for the classification of Asian hornets
                cnn.add(Conv2D(j, (2, 2), input_shape=feat))
                cnn.add(Activation('relu'))#Relu activation function
                cnn.add(MaxPooling2D(pool_size=(2, 2)))
                for x in range(k - 1):
                    cnn.add(Conv2D(j, (2, 2))) #convolution and sampling
                    cnn.add(Activation('relu'))#Relu activation function
                    cnn.add(MaxPooling2D(pool_size=(2, 2)))
                cnn.add(Flatten())#flatten
                for y in range(i):
                    cnn.add(Dense(j))#dense
                    cnn.add(Activation('relu'))#Relu activation function
                cnn.add(Dense(1))
                cnn.add(Activation('sigmoid'))#Secondary Classification
                opt = Adam(lr=0.1)
                cnn.compile(loss='binary_crossentropy', optimizer=opt,
metrics=['accuracy'])#train
                cnn.fit(ft, label, batch_size=32, epochs=epoch,
```

```
callbacks=[tensorboard],validation_split=.1)#Visualization
                    cnn.save(dir_path + '/Models/Save-' + name)
```

Appendices B: Cellular automata model

main.m

```
load InitializationParameter.mat    % Load initialization parameters
load border.mat
breed = 0.35;      % The survival rate of the hive under normal conditions
constraint = 0.07;  % The survival rate of a hive under human disturbance
figure
plot(border(:, 1),border(:, 2), 'color', 'k')
xlabel('Longitude')                 % The X-axis indicates longitude
ylabel('Latitude')                  % The Y-axis indicates latitude
hold on                           % Continue to draw in the diagram
nestLocation = initialPosition;
for year = 1 : 30
    nestLocationAll = [];          % Initialize all nest locations
    for idxNest = 1 : size(nestLocation, 1)
        nestNumNew = sum(rand(40, 1) <  breed);
        nestLocationNew = [unifrnd(nestLocation(idxNest, 1)-0.33,
nestLocation(idxNest, 1)+0.33, nestNumNew, 1) ...
            unifrnd(nestLocation(idxNest, 2)-0.27, nestLocation(idxNest,
2)+0.27, nestNumNew, 1)];
        nestLocationNew = cus_fun(nestLocationNew, 0.07);
        if ~isempty(nestLocationNew)
            flag = inpolygon(nestLocationNew(:,1), nestLocationNew(:, 2),
border(:, 1), border(:, 2));
            nestLocationNew = nestLocationNew(flag, :);
            nestLocationAll = [nestLocationAll; nestLocationNew];
        end end
    if ~isempty(nestLocationAll)
            nestSurviveIdx = rand(size(nestLocationAll, 1), 1) < constraint; %
        nestLocationAll = cus_fun(nestLocationAll, 0.07); % Reproductive
        disp(size(nestLocationAll,1));   % Print the number of hives
    end
    if ~isempty(nestLocationAll)  % Draw the diagram
        nestLocation = nestLocationAll;
        exist p 'var';
        if ans delete(p);end
        p = plot(nestLocation(:, 1), nestLocation(:, 2), '*', 'color', [1 0.25
0]);  pause(1);              % It's updated every second
        drawnow
   end
 end
```