

Asian Giant Hornet Control based on Image Processing and Biological Dispersal

Summary

The Asian giant hornet (AGH) appeared in Washington State appears to have a potential danger of bioinvasion. Washington State has collected public photos and videos of detected insects for verification and further investigation. In this paper, we analyze AGH using data analysis, statistics, discrete mathematics, and deep learning techniques to process the data to control AGH spreading.

First, we visualize the geographical distribution of insects in Washington State. Then we investigate insect populations to varying months of the year and different days of a month. Third, we employ **wavelet analysis** to examine the periodic spread of AGH. Fourth, we apply **ordinary differential equations** to examine AGH numbers at the different natural growth rate and reaction speed and output the potential propagation coefficient. Next, we leverage **cellular automaton** combined with the potential propagation coefficient to simulate the geographical spread under changing potential propagation. To update the model, we use **delayed differential equations** to simulate human intervention. We use the time difference between detection time and submission time to determine the unit of time to delay time. After that, we construct a lightweight CNN called **SqueezeNet** and assess its classification performance. We then relate several **non-reference image quality metrics**, including NIQE, image gradient, entropy, contrast, and TOPSIS to judge the cause of misclassification. Furthermore, we build a **Random Forest** classifier to identify positive and negative samples based on image qualities only. We also display the feature importance and conduct an error analysis. Besides, we present sensitivity analysis to verify the robustness of our models. Finally, we show the strengths and weaknesses of our model and derives the conclusions.

We find meaningful conclusions from our analysis. (1) Most Asian Giant Hornet (AGH) is reported at West Washington from May to December. However, only 14 AGH proves to exist. (2) Local insects have a monthly active cycle of about 7.71 days. AGH breed at a rapid rate every Spring. Meanwhile, the geographical spread of AGH heavily depends on the potential propagation coefficient. (3) Earlier artificial intervention would efficiently control the AGH population and protect other insects. (4) Light-weight convolutional neural network is efficient and robust for classifying AGH. (5) High image Gradient Pictures tend to contain AGH, while high image Contrast and background interference will misguide classification.

We finally provide recommendations based on our analysis result. (1) AGH has a tiny population and is nearly eradicated in Washington State. Therefore is no need for massive panic. (2) AGH has a predicted spread period of approximately one and a half months. When their growth rate is high, we should take earlier actions and eliminate them before expansion. (3) We should prioritize image gradient over image contrast and ignore background noise for subjective classification. (4) We recommend a lightweight CNN for accurate classification.

Key Words: SqueezeNet; Cellular Automaton; Wavelet Analysis; Delay Differential Equation; Random Forest; Non-reference Image Quality Metrics

Contents

1	Introduction	2
1.1	Background	2
1.2	Related Works	2
1.3	Solutions	2
2	Assumptions and Notations	3
3	Data Preprocessing	4
3.1	Data Acquisition	4
3.2	Data Visualization	5
4	Expansion Model	7
4.1	Differential Equations	7
4.1.1	Ordinary Differential Equations	7
4.1.2	Delay Differential Equation	8
4.2	Cell Automaton	9
4.3	Wavelet Analysis	10
5	Classification Model	12
5.1	Data Preparation	12
5.2	Model Construction	13
5.3	Model Evaluation	14
5.3.1	Evaluation Metrics	14
5.3.2	Visual Demonstration	15
5.3.3	Quantitative Comparison	16
5.4	Image Evaluation	16
5.4.1	Image Quality Metrics	16
5.4.2	Model Implementation	17
5.4.3	Feature Importance & Error Analysis	18
6	Sensitivity Analysis	19
6.1	Expansion Model	19
6.2	Classification Model	20
7	Model Reflection	21
7.1	Strengths	21
7.2	Weaknesses	21
8	Memorandum	22

1 Introduction

1.1 Background

Asian Giant Hornet (*Vespa mandarinia*), a hornet species native to eastern Asia's temperate and tropical areas, is a predator to pollinating insects, the world's most enormous hornets.

On Vancouver Island in British Columbia, Canada, Sep 2019, an AGH colony was detected and immediately destroyed. In May 2020, an AGH individual was discovered in Washington state, which means AGH survived a winter in northern America.

In the 21st century, biological invasions have become one of the most significant problems to biodiversity, crop production, and ecological balance. As one kind of social Hymenoptera, AGH perfectly fits in a spreading pattern by human transportation and hidden skills from standard border inspection. They hide in transported goods and can quickly establish a new colony with only one fertilized queen. Furthermore, high fertility, comprehensive diet and habitat, great defense, and competitive abilities also make them dominant in the area [1]. A species under the same genus with AGH, which is called *V. velutina*, invaded throughout the European continent with a rapid speed, and their control costs can be very high [2]. A laboratory researched AGH spreading pattern by building habitat suitability models and dispersal simulations, concluded that AGH has a high possibility of spreading across western North America with no human intervention [3].

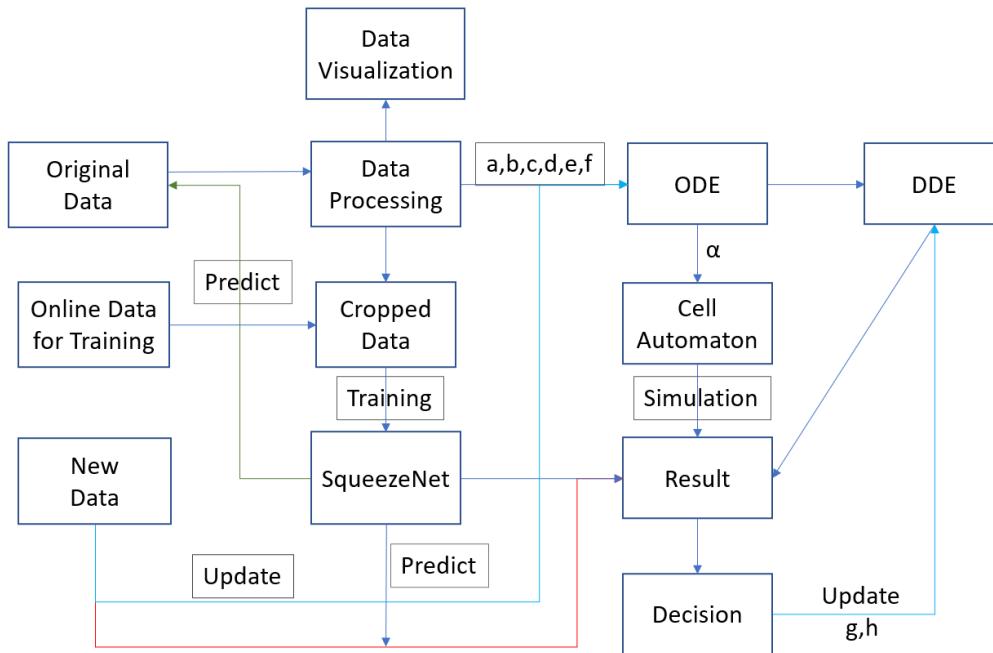
To investigate the current situation of AGH and find the government of Washington State offered a solution defending its bioinvasion, hotlines, and a website for people to report the hornets' emerges in different areas. The public reports need to be processed through multiple analyses for further study and giving solutions.

1.2 Related Works

Convolutional neural network (CNN) has been widely applied to image classification tasks because of their excellent accuracy, capability, and flexibility. Krizhevsky [4] introduces AlexNet, which uses max-pooling layers and drop-out techniques. Simonyan [5], with others in VGG (Visual Geometry Group), designed very deep convolutional networks for large-scale image recognition, which grows deeper layers with a small kernel size of $3 * 3$. Lightweight image classification models have been designed to fit mobile devices that have limited memory capacity. MobileNet [6] uses depth-wise separable convolutions to a trade-off between speed and accuracy efficiently. SqueezeNet [7] adopts small filters, late downsampling, and fire block (squeeze and expand operations) for flexible and efficient model compression.

1.3 Solutions

In this paper, we propose a novel approach for analyzing, predicting, and giving recommendations towards Asia Hornet. Our analysis uses different discrete mathematics, deep learning, and data visualization tools. First, we import and preprocess the data, including mapping images, videos to text, and numbers. Second, we use dispersal models. Our dispersal model mainly includes different equations, cell automation (CA), and wavelet analysis. Third, we bring a CNN for classifying AGH. We leverage statistical evaluation metrics and activation maps to verify our model performance. Fourth, we

Figure 1: **Paper Workflow**

explore the cause for misclassification using a random forest [8] classifier with several non-reference image quality metrics.

The rest of the paper is organized as below. Section 2 introduces basic notations and assumptions. Data acquisition and data visualization are conducted in Section 3. Section 4 and section 5 construct the expansion model and the classification model, respectively. Section 6 perform sensitivity analysis and Section 7 analyzes the strengths and weaknesses. Conclusion and recommendations will be addressed in the Memorandum.

2 Assumptions and Notations

- We assume that AGH only competes with local species and receive minimum impact from altitude, weather, and human activities.
- We assume that the map of Washington State can divide into 128*128 squares.
- We assume that the spread pattern of AGH in Washington State is relatively fixed.
- We assume that the proportional relationship between AGH and other dwelling species is similar to the number of reports submitted within a specific range of latitude and longitude.

Abbreviation	Full form
AGH	Asian Giant Hornet
CNN	Convolutional Neural Network
CA	Cellular Automaton
RF	Random Forest
ODE	Ordinary Differential Equation
DDE	Delay Differential Equation
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
NIQE	Naturalness Image Quality Evaluator
ReLU	Rectified Linear Unit
Conv2d	2D Convolution Layer
OOB	Out Of Bag

Table 1: Abbreviation of Relevant Terms

Symbol	Explanation
A(t)	Amount of AGH at any given time
B(t)	Amount of local species at any given time
a	Natural growth rate of AGH
b	Internal competition coefficient of AGH
c	Aggression coefficient of AGH to local species
d	Natural growth rate of local species
e	Internal competition coefficient of local species
f	Resistance coefficient of local species to AGH
g	Effects of manual intervention on AGH
h	Effects of manual intervention on local species

Table 2: Symbols and Explanations

3 Data Preprocessing

3.1 Data Acquisition

We have several data files for analysis. The first is a pdf file named *Vespa mandarinia*, which we have used in the introduction section. The second is a pair of a spreadsheet. We use GlobalID to combine two spreadsheets. We first convert all images into jpg format. We use the strategy for word and pdf file that contains only images. We then take a screenshot of the video and convert it to jpg images. We take the screenshot at the first explicit appearance of insects.

3.2 Data Visualization

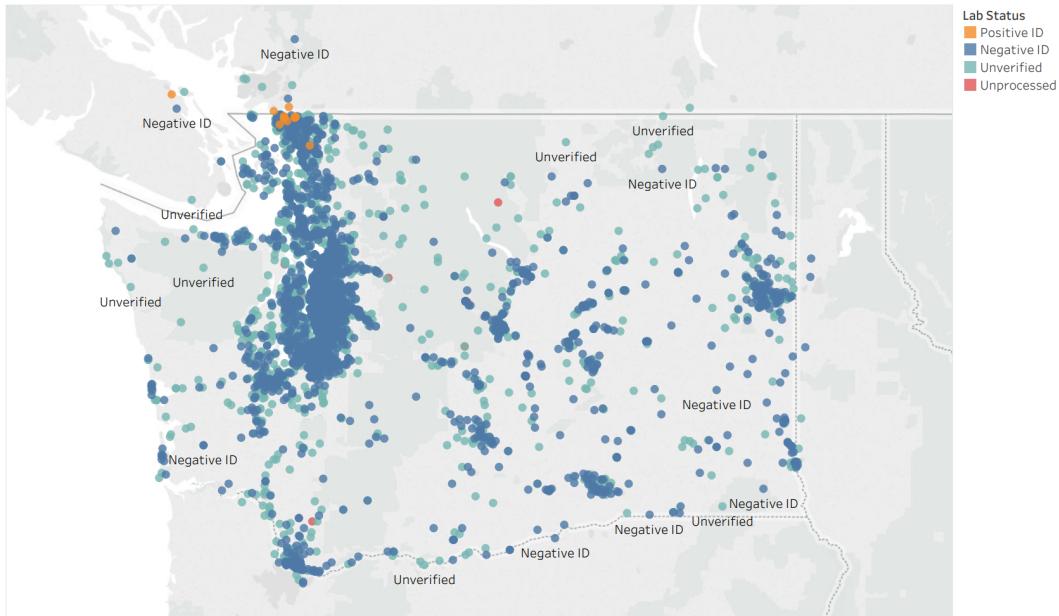


Figure 2: Geographical Visualization

We perform classification using lab comments and lab status. We count each insect species' number under the same lab status in different years by detection date, from January to December, integrating them into one spreadsheet. The Geographical Visualization result in Fig.1 shows most reported cases appear in west Washington. Besides, all positive ID appears in northwest Washington.

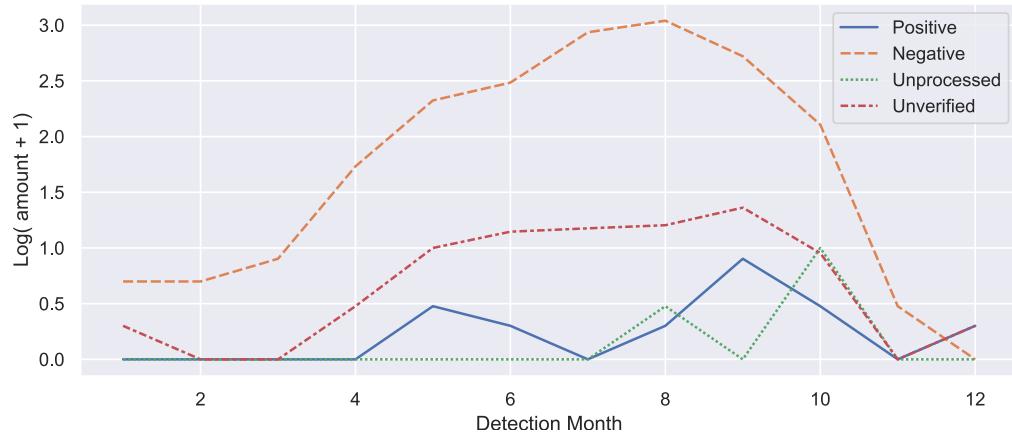


Figure 3: Change of Insect population over Time. X-axis: the month where a insect is observed. Y-axis: Corrected log value of insect amount in a specific month. The category positive, negative, unprocessed, and unverified from lab status data means AGH species insects, non-AGH species insects, unprocessed insects, and unverified insects respectively.

We have several findings from Fig.3. First, each category begins a rapid growth in May and peaked

in around September. Second, the negative category has the highest numbers. Third, the positive and unprocessed category has the least numbers with the similar trend, where as the unverified category has a similar trend with the negative category.

In Fig.4(b), we have categories including symphyta (sawfly), wasp, *Sphex ichneumoneus* (great golden sand digger), *Bombus* (bumblebee) and *Apis* (honey bee), *Dolichovespula maculata* (bald-head hornet) and *Vespa crabro* (European hornet), and other species.

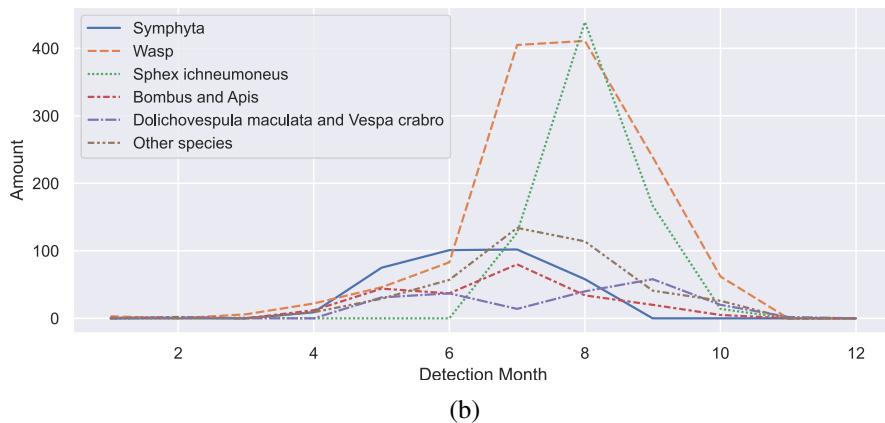
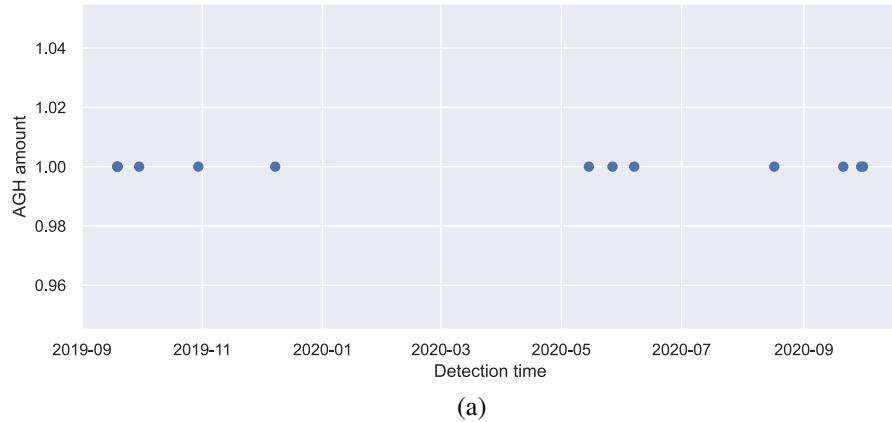


Figure 4: (a): **Change of AGH population over time.** X-axis: the year and month when a AGH is observed. Y-axis: Numbers of AGH at a specific time. (b): **Change of Non-AGH population over time.** X-axis: the month where a insect is observed. Y-axis: the numbers of reported insects that is not AGH. We roughly classified insects into 5 categories and others.

We have several findings from Fig.4. First, there are 14 AGHs detected on specific dates. Second, no AGHs is detected during 2020-01 to 2020-05. Third, only one AGD is reported in one detection. Besides, all non-AGH insects have a rapid increase in June and a significant decline in September. Last, *Sphex ichneumoneus* and wasp have peak values of around 400 during July and August, whereas others category of insects does not.

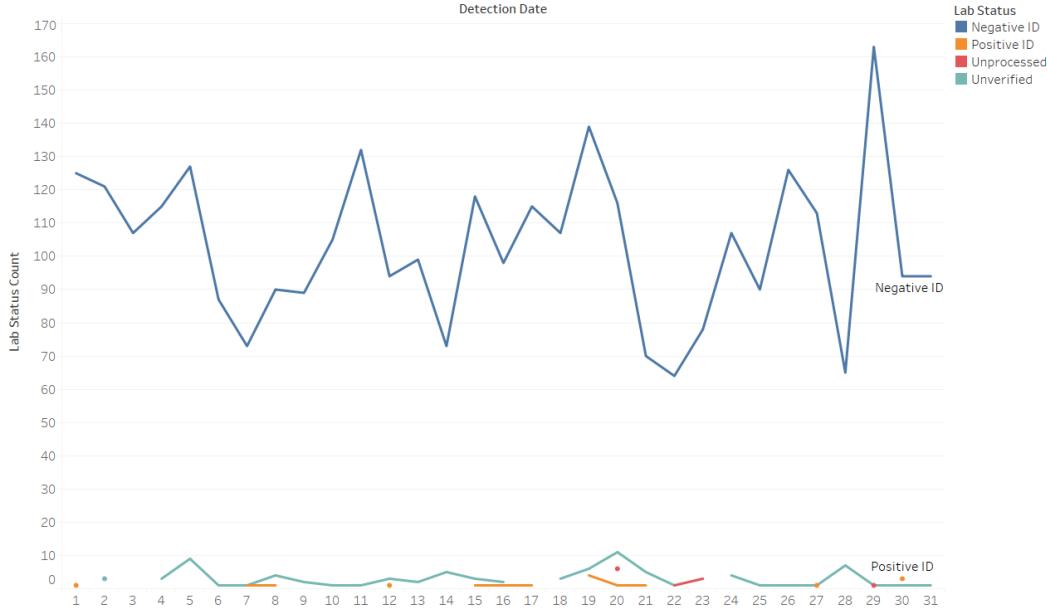


Figure 5: Total Insect Numbers Detected by Day in a Month. X-axis: day in a month, from 1 to 31. Y-axis: lab status comment for indicating insect categories. Legend: different lab status

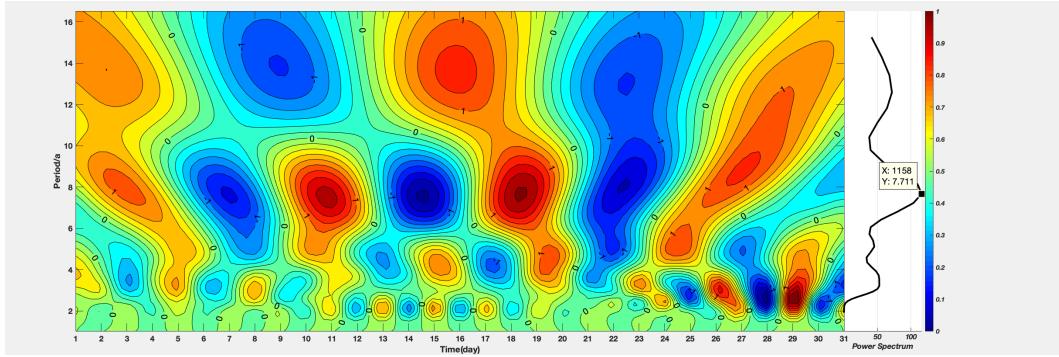


Figure 6: Insect Periodic Activity with Wavelet Analysis. Left X-axis: time in days since simulation. Left Y-axis: cycle time period. Right X-axis: power spectrum frequency. Right Y-axis: intensity

We introduce wavelet analysis [9] in Fig.6 to investigate if the insects have monthly activity patterns. Wavelet analysis figure show that inserts have a 7.71-day period, which explains Washington State insects' monthly activity as 7 days.

4 Expansion Model

4.1 Differential Equations

4.1.1 Ordinary Differential Equations

We use systems of ordinary differential equations (ODE) to simulate how insect populations evolve. Our different equations seek equilibrium points given initial values and constraints. In Fig.7, we

describe AGH as competing species to local species.

We set up a differential equation to study the population change of AGH in Washington. The system of differential equations established is as follows:

$$\begin{cases} \frac{dA(t)}{dt} = aA(t) - bA^2(t) - cA(t)B(t) \\ \frac{dB(t)}{dt} = dB(t) - eB^2(t) - fA(t)B(t) \end{cases} \quad (1)$$

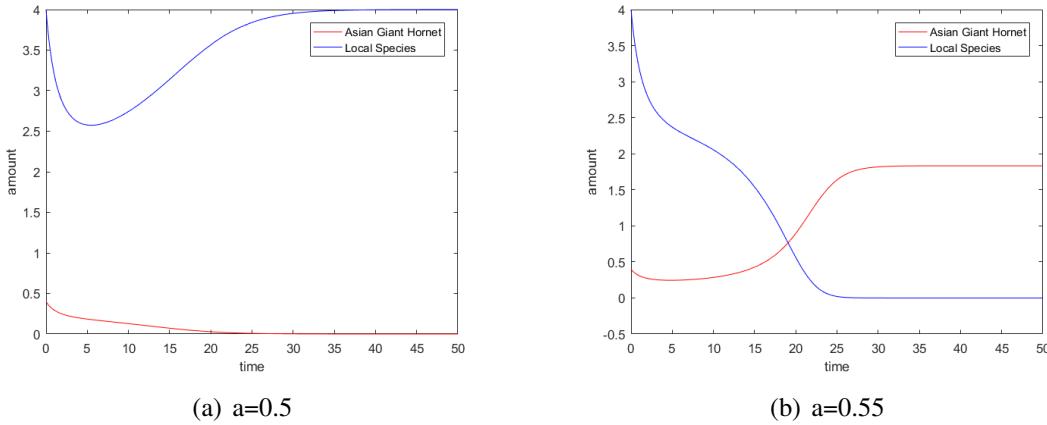


Figure 7: The solution of our ODE systems. (a): Solution when Natural Growth Rate of Asian Giant Hornet (AGH) is 0.5. (b): Solution when growth rate is 0.5. For both conditions, we set $b = 0.3$, $c = 0.2$, $d = 0.4$, $e = 0.1$, and $f = 0.8$

The number of AGH found in northwest Washington is roughly 1:10 for the number of Local species. Therefore, we set $A = 0.4$, $B = 4$ as a starting point. The result of our ODE appears in Fig.7. When the natural growth rate of AGH is 0.5, the amount of AGH decreases steadily from 0.5 to 0, whereas the amount of local species decreases sharply to 2.5 before slowly increasing to 4.0. When the natural growth rate of AGH is 0.55, the amount of AGH increases to 2, whereas the number of local species decreases to 0. When time is greater than 25, they arrive at equilibrium.

4.1.2 Delay Differential Equation

We use systems of delay differential equations (DDE) to analyze the problem further. Unlike ODE, DDE has the derivative of the unknown function in terms of previous time points. The formula of our delayed differential equation is below. Note the last terms at each line comes from an early reaction. Others are the same as the ODE.

We calculated the average difference between image submission time and detection time. The parameters g and h represent the influence of the manual intervention on the population. ($g = 0.05$, $h = 0.001$) Delays in manual intervention can lead to dramatic changes in population evolution.

$$\begin{cases} \frac{dA(t)}{dt} = aA(t) - bA^2(t) - cA(t)B(t) - gA(t-10) \\ \frac{dB(t)}{dt} = dB(t) - eB^2(t) - fA(t)B(t) - hB(t-10) \end{cases} \quad (2)$$

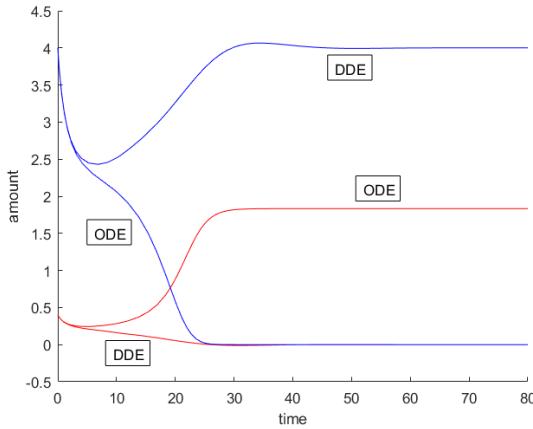


Figure 8: Solution Comparison between ODE and DDE. We set the natural growth rate of AGH as 0.55. We use red for AGH and blur for other species.

Fig.8 shows the result from earlier and on-time actions towards AGH. In the ODE solution, there is quick increase of AGH and a sharp decrease of other species. In the DDE solution, there is quick increase of other species and steadily decline of AGH. Our result demonstrate earlier action help control AGH numbers.

4.2 Cell Automaton

Cellular Automaton (CA) is a discrete computational model in automation theory. It consists of finite-dimensional cells with states (e.g., True/False). A new generation is created according to specific algorithms to iterative update the neighboring cells.

From the the ODE, We get values of $A(t)$ and $B(t)$ with respect to time. Therefore, we define the potential propagation coefficients:

$$\alpha(t) = \frac{A(t)}{A(t) + B(t)} \quad (3)$$

We apply CA to simulate the spread of AGH in Washington and simplify the map into a square, dividing into 128*128 cells. In the upperleft 15*15 grid, we use red to represent AGH existence and black to represent native species. Since most AGH cases is reported in northwest Washington, we set initial probability ratio as 4:1. The probability of each grid occupied by AGH is (if $P < 1$):

$$P(\text{cells} = 1) = \frac{\text{sum}}{2} \cdot \alpha\left(\frac{t}{10}\right) \quad (4)$$

Where $\text{sum}(x,y)$ represents AGH's amount around the grid with coordinate position (x,y) , and t is the time unit of CA. After every 50 units of time, there will be 10 units where AGH cannot reproduce and has a death probability of 0.3:

$$P(\text{cells} = 0) = 0.3 \quad (5)$$

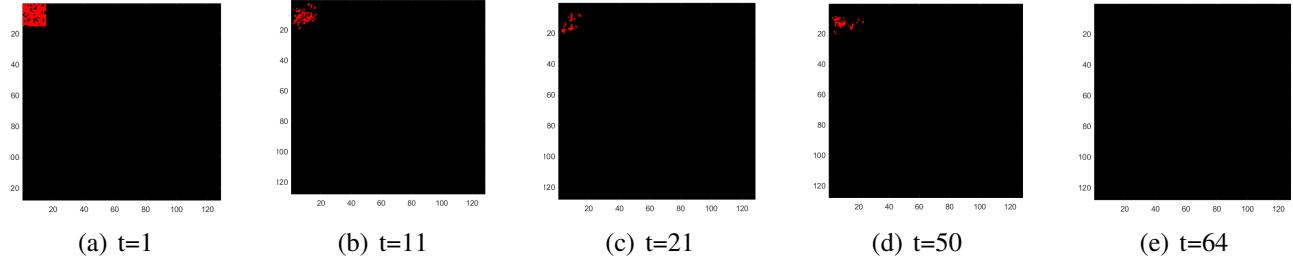


Figure 9: Cellular Automaton Using the First Potential Propagation Coefficient

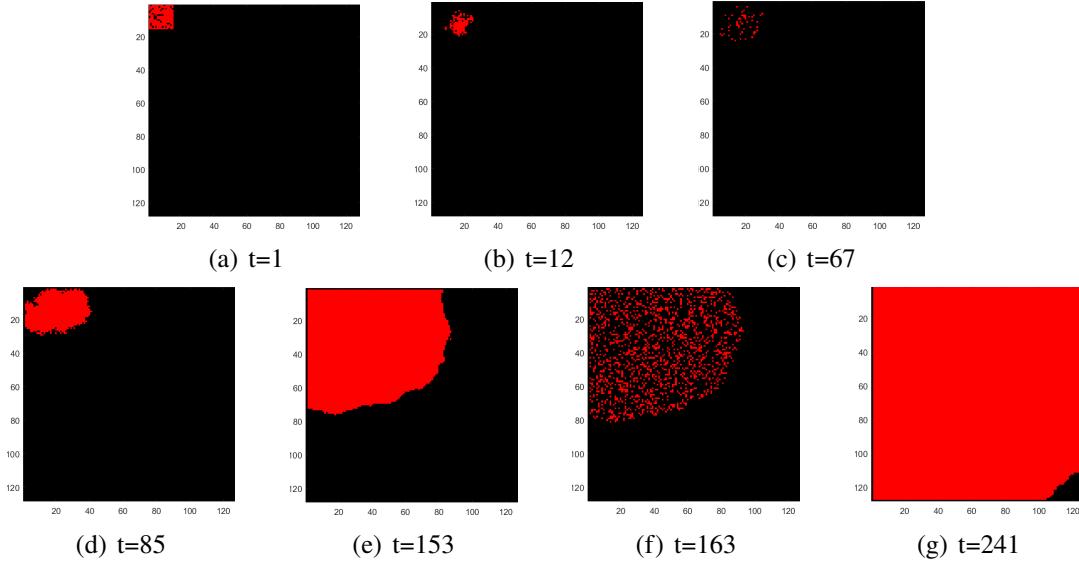


Figure 10: Cellular Automaton Using the Second Potential Propagation

In Fig.9, we visualize AGH's evolution based on the first differential equation, from $t = 1$ to 64, AGH gradually decrease and finally disappear.

In Fig.10, we visualize AGH's evolution based on the first differential equation, from $t = 1$ to 64, AGH gradually decrease and finally disappear. In Fig.10, we visualize AGH's evolution based on the second differential equation. From $t = 1$ to 67, AGH has small and fluctuating numbers. Starting from $t = 85$, the AGH population grow explosively and finally occupy Washington.

4.3 Wavelet Analysis

Wavelet transform is a decomposition of a continuous signal into different scale components. Wavelet analysis is better than Fourier analysis in that it can capture the time information of frequency change. Therefore, wavelet analysis is suitable for complicated movements such as the spread of AGH. Our Morlet wavelet is a continuous plane wave modulated by Gauss function:

$$\Psi(t) = e^{icl} \left(e^{-\frac{l^2}{2}} - \sqrt{2} e^{-\frac{c^2}{4}} e^{-l^2} \right) \quad (6)$$

Its discrete mathematical expression is:

$$W_f(a, b) = |a|^{-\frac{1}{2}} \sum_{i=1}^N f(i\delta t) \psi^* \left(\frac{i\delta t - b}{a} \right) \quad (7)$$

where * represents complex conjugate; a is scale factor (related to period and frequency); b is the translation factor (time position); I is time and position label of data series; f(t) is the variable time series; W is the wavelet coefficient; δt is the time interval of the variable series. Since the wavelet is in complex form, the coefficients after the wavelet transform are also complex.

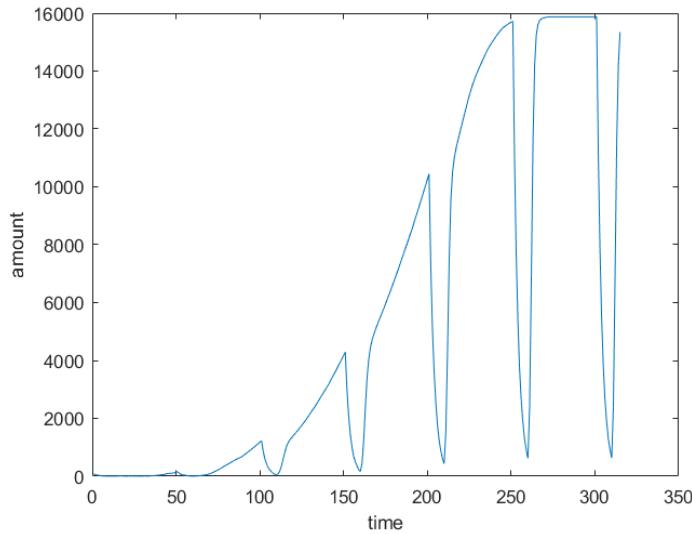


Figure 11: **Red Grid Numbers in Cellular Automaton over Time.** X-axis: Time since starting simulation. Y-axis: Red Grid amount. We select the condition of the second differential equation for plotting. More Grid counts represents more AGH occupancy.

Fig.11 shows a general increasing trend for AGH numbers. However, there is a sharp decrease for after every 50 time units. Therefore, eliminating the Asian bumblebee in winter is a better option.

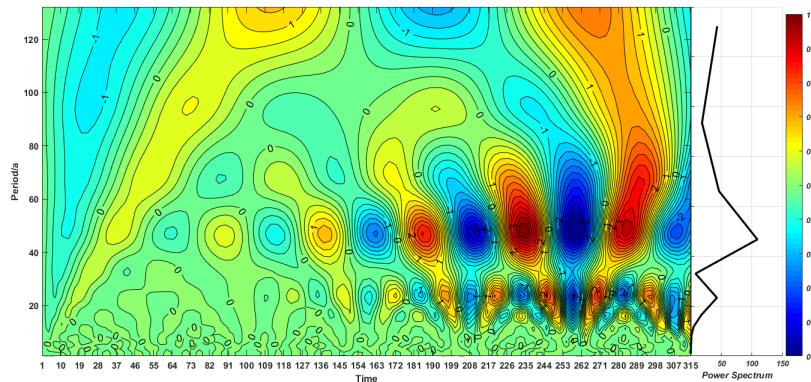


Figure 12: **Wavelet Analysis of Red Grids Numbers.** Left X-axis: Time in days since simulation. Left Y-axis: Cycle Time Period. Right X-axis: Power Spectrum. Right Y-axis: Intensity

Fig.12 shows that the amount of AGH in CA has a period of 46.75 time unit, which meets the stagnation phase of 50 days. Besides, the only difference between two ODE is the reproduction ability of AGH.

The main factors affecting the reproductive ability of AGH are temperature, humidity, and manual intervention. AGH has high fertility, and a little deviation may cause the amount of the AGH has divergent results. Moreover, AGH is intolerant to frigidity; First, AGH almost disappear In the 67th unit which is in winter. However, it has fast outbreak afterwards. Therefore, the Washington Department of Agriculture should pay attention to the number of AGH at next spring. Second, 3 unidentified samples were classified as AGH at spring or summer on the west part of Washington. Therefore, we should monitor the potentially spreading at west Washington.

5 Classification Model

5.1 Data Preparation



Figure 13: Batch of Cropped Dataset

We use an AGH dataset from a Github repository [10] to make 509 positive samples. We then select 569 negative examples from the given dataset (See Fig.13 for a sample batch). The total amount of pictures for the model training and testing is 1078. We first resize the cropped dataset to the shape of 224*224. Then we add Random Horizontal Flip to flip with a probability of 0.5 horizontally. Next, we standardize the data using ImageNet's [11] mean and standard deviation. After that, we split the data into training, validation, and testing set with the ratio of 0.70, 0.15, and 0.15, respectively. We have 757 images for training, 160 for validation, and 161 for testing. Ultimately, we prepare training and testing labels, labeling “1” for positive (AGH) and “0” for negative (non-AGH).

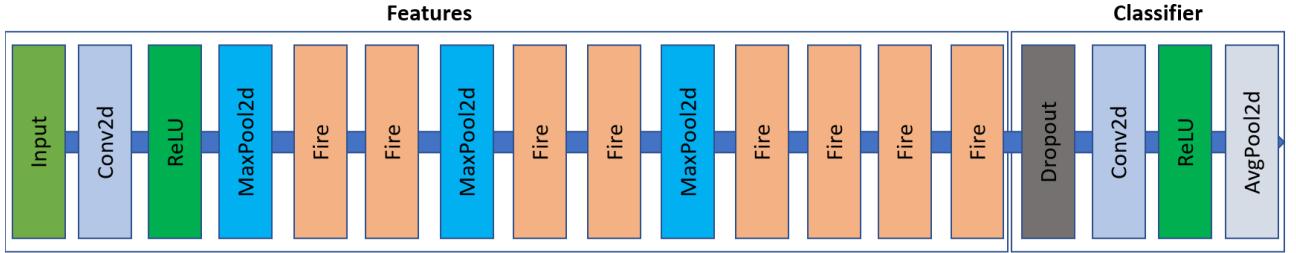


Figure 14: **Model Architecture of SqueezeNet.** SqueezeNet consists of features layers and classifier layers. We use Conv2d, ReLU, MaxPool2d, Fire, Dropout and AvgPool2d for convolutional layer, ReLU activation layer, maximum pooling layer, Fire Block, Dropout layer and average pooling layer, respectively.

5.2 Model Construction

We choose a light-weight model called SqueezeNet [7]. Since our dataset is small (1000 images), using a light-weight CNN can prevent overfitting and achieve better accuracy. We use pretrained weight from ImageNet. We then change the final classification block output from 1000 to 2 for binary classification. We use Dropout [12] with 0.5 probability to prevent overfitting. We also use ReLU [13] activation layers to mitigate vanishing gradient problem.

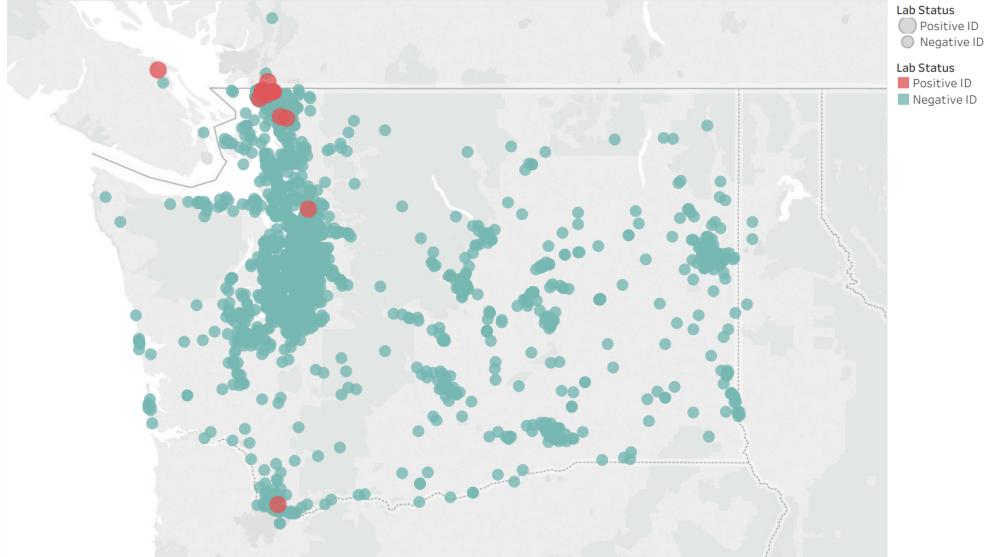


Figure 15: **Geographical Visualization after SqueezeNet classification.** We use red for positive samples and green for negative samples

Fig.15 show most positive samples (AGH) are found at Northwest Washington. Only 1 case is found in Southwest Washington. We use the Pytorch framework for training our model. We set the epoch as 3 and batch size as 8. We adopt cross-entropy loss as the cost function. We utilize stochastic gradient descent with a learning rate of 1e-3 and momentum of 0.9. Since our training only consists of 3 epochs, there is no need for learning rate decay.

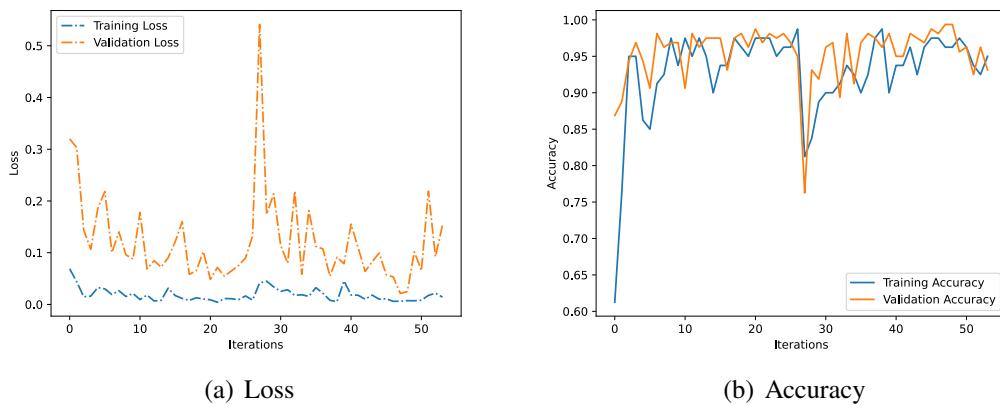


Figure 16: **Loss-Accuracy Plot.** (a): Training and Validation Loss. (b): Training and Validation Accuracy.

Fig.16 shows the loss and accuracy during training and validation. Our training loss quickly converges and remains stable between 0 to 0.03. Our validation loss has a spike at the iteration of 30 but has a general trend of declining from 0.3 to 0.1. Except for the sharp decline in the middle, our training and validation accuracy quickly increase and stay between 0.95 and 1.00. Since both accuracies are similar at a high level, there is neither overfitting nor underfitting. We obtain the final testing accuracy of 0.975.

5.3 Model Evaluation

5.3.1 Evaluation Metrics

We use a group of evaluation metrics to evaluate our models. Specifically, we include loss, total accuracy, class accuracy, recall, specificity, precision, f1 score, and confusion metric for evaluating our image classification model.

Confusion matrix displays a table of true positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). The formula for other evaluation metrics is given as below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$F1\text{ Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

5.3.2 Visual Demonstration

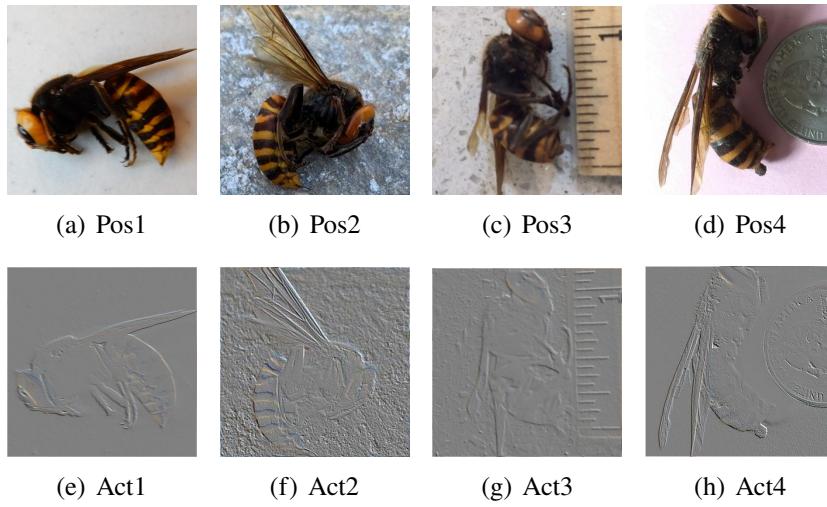


Figure 17: Activation Layers Result with samples classified as positive (AGH). The upper row is the input image. The lower row is the output from the first convolution layer. We use "Pos" for positive samples and "Act" for activation layers, respectively.

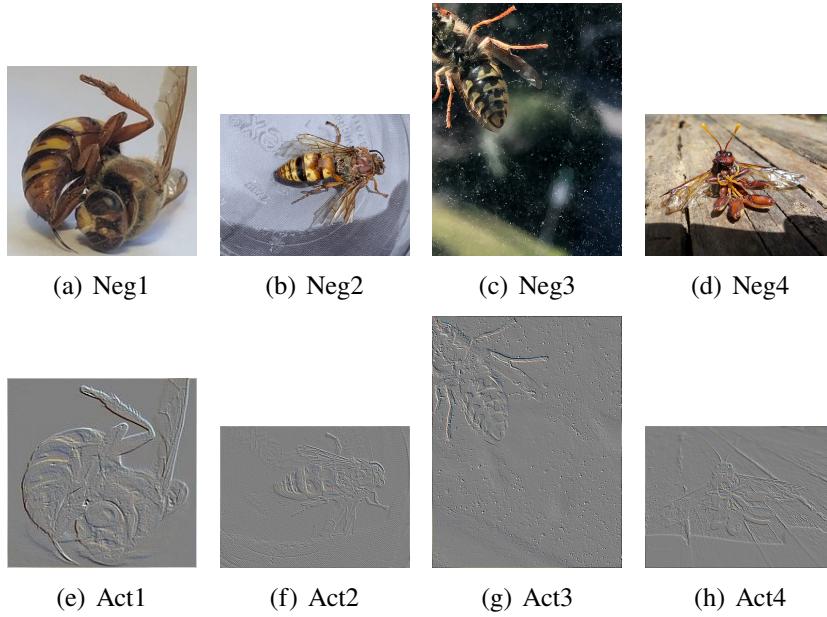


Figure 18: Activation Layers Result with samples classified as Negative (non-AGH). The upper row is the input image. The lower row is the output from the first convolutional layer. We use "Neg" for negative samples and "Act" for activation layers, respectively.

We visualize the first convolutional layer result in Fig.17 and Fig.18. Our rows display the positive sample, positive sample activation result, negative sample, and negative sample activation result, respectively. As shown in the figure, our neural network can filter background interference and separate

the given insects' pattern. Besides, our CNN displays more evident edges for positive samples. That means our model can efficiently capture the distinct feature of the AGH.

5.3.3 Quantitative Comparison

In this part, we quantitatively analyze our model classification performance. We first display the confusion matrix. According to Fig.19, we have 82 true positives, 75 true negatives, 3 false positives, and 1 false negative. We also note from Table 3 that all evaluation metrics have a score greater than 0.96. That means our model has great classification accuracy. Besides, our model has better predictive accuracy at positive samples than negative samples. That is reasonable because the negative examples contain numbers of insects that look similar to AGH (e.g., European Hornet). Table 5 shows the image quality. Positive samples win all evaluation metrics except image contrast. That means image contrast is a misleading metric for classifying AGH.

	Accuracy	Class Accuracy	Recall	Specificity	Precision	F1 Score
Value	0.975	0.965/0.987	0.988	0.962	0.965	0.976

Table 3: **Quantitative result of our image classification model.** Class Accuracy has negative (not AGH) on the left and positive (AGH) on the right.

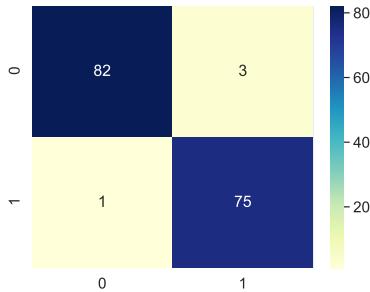


Figure 19: **Confusion Matrix of our image classification model.** Left: Actual Label. Bottom: Predicted Label. from upperleft to lowerright cell: True Positive, False Positive, False Negative and True Negative,

5.4 Image Evaluation

5.4.1 Image Quality Metrics

We use a set of non-reference image quality metrics to understand why people would misclassify other insects as Asian hornets. That group of image quality metrics would also help investigators screen out a more likely picture to contain AGH. Our image quality metrics include Naturalness Image Quality Evaluator (NIQE) [14] , image gradient, entropy, and contrast. We also use TOPSIS to summarize the indicators above. NIQE is a statistical measure derives from natural, undistorted images. Image Gradient is a vectorized change in image intensity. Image Entropy is a statistical measure of randomness

	NIQE	Gradient	Entropy	Contrast	TOPSIS
Negative	6.98	0.057	7.385	3092	0.1912
Positive	10.10	0.075	7.525	2838	0.2655

Table 4: **Image Quality Comparison for positive and negative samples.** We use bold to indicate higher values.

for assessing the texture of an image. Image Contrast is the difference in luminance that makes a picture visually distinguishable. TOPSIS is a multi-criteria decision analysis method that uses weighted and normalized geometric distance for assessing image quality. We use TOPSIS to give a comprehensive evaluation of image quality.

NIQE:

$$D(\nu_1, \nu_2, \Sigma_1, \Sigma_2) = \sqrt{\left((\nu_1 - \nu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\nu_1 - \nu_2) \right)} \quad (13)$$

Image Gradient:

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (14)$$

Entropy:

$$S = k_b \ln \Omega \quad (15)$$

Contrast:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ij} - \bar{I})^2 \quad (16)$$

We acquire the positive and negative ideal points of each indicator in TOPSIS:

$$d_i^* = \sqrt{\sum_{j=1}^4 (z_{ij} - z_j^*)^2}, d_i^- = \sqrt{\sum_{j=1}^4 (z_{ij} - z_j^-)^2} \quad (17)$$

where Z is the weighted evaluation matrix.

5.4.2 Model Implementation

We use four non-reference image quality metrics values as features and the other one as the label. We then split the training and testing data with a train/test ratio of 2:1 and used random forest classifiers to classify positive and negative samples.

Random Forest classifier is an ensembled supervised machine learning algorithm. It is a meta estimator that leverages numbers of decision trees and takes the average of those predictors to improve forecasting accuracy and reduce predictability variability.

We initially set the min sample split as two and the numbers of estimators as 150. We then select four hyperparameters for random grid search cross-validation. Specifically, We run 3-fold cross-validation

	n_estimators	max_depth	min_samples_split	min_samples_leaf
Value	644	110	10	2

Table 5: **Best Hyperparameters of the random forest classifier.** n_estimators: numbers of decision tree in random forest. max_depth: maximum depth of the tree. min_samples_split: minimum samples for splitting internal nodes. min_samples_leaf: minimum samples for a leaf node.

for 100 iterations and find the hyperparameter group that minimizes the loss function. The best hyperparameters summarize in the Table 5. Our Testing Accuracy from the optimized hyperparameters is approximately 0.686.

5.4.3 Feature Importance & Error Analysis

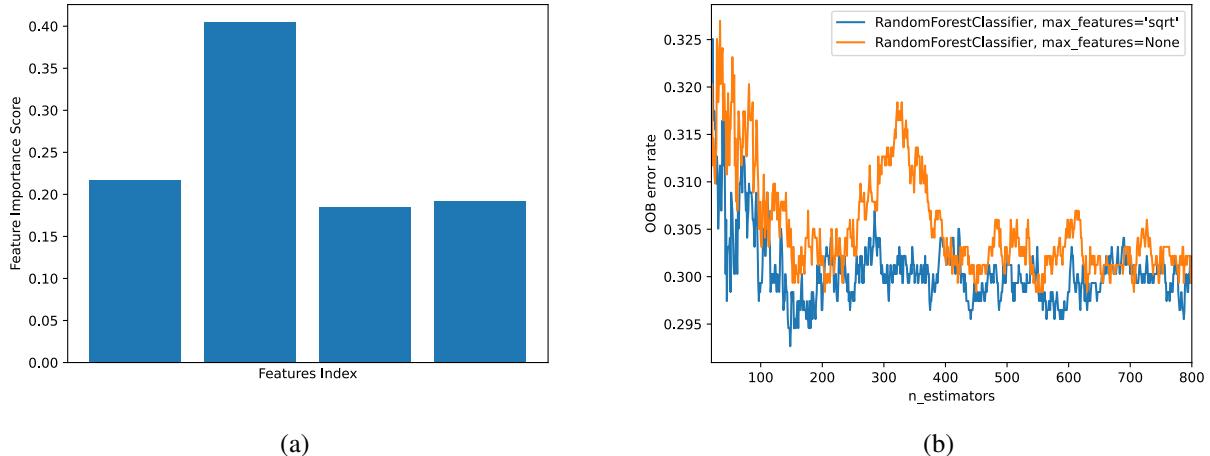


Figure 20: (a): **Feature Importance Score.** X-axis: feature index. From left to right, NIQE, gradient, entropy and contrast. Y-axis: Feature Importance Score, from 0 to 1. (b): **Error Analysis.** X-axis: number of estimators. Y-axis: out-of-bag (OOB) error rate

We visualize the feature importance in the Fig.20(a). The feature importance score describes how significant a specific feature is for image classification. We observe from the plot that the image gradient has dominant importance (around 0.40) for classification, whereas other factors contribute similarly.

We also examine the source of classification errors in Fig.20(b). In Fig.20 (b), we note different types of maximum features: the number of features for calculating the best split. When max features are 'sqrt' (the blue line), the classifier will consider maximum features as the total features' square root. When max features are 'None' (the yellow line), it equates maximum features to total features.

For both max features types, we have decreasing OOB error rate with increasing numbers of features. However, we note that when numbers of estimators grow greater than 600, the OOB error rate remains stable. Therefore, it is efficient to stop at around 600.

6 Sensitivity Analysis

6.1 Expansion Model

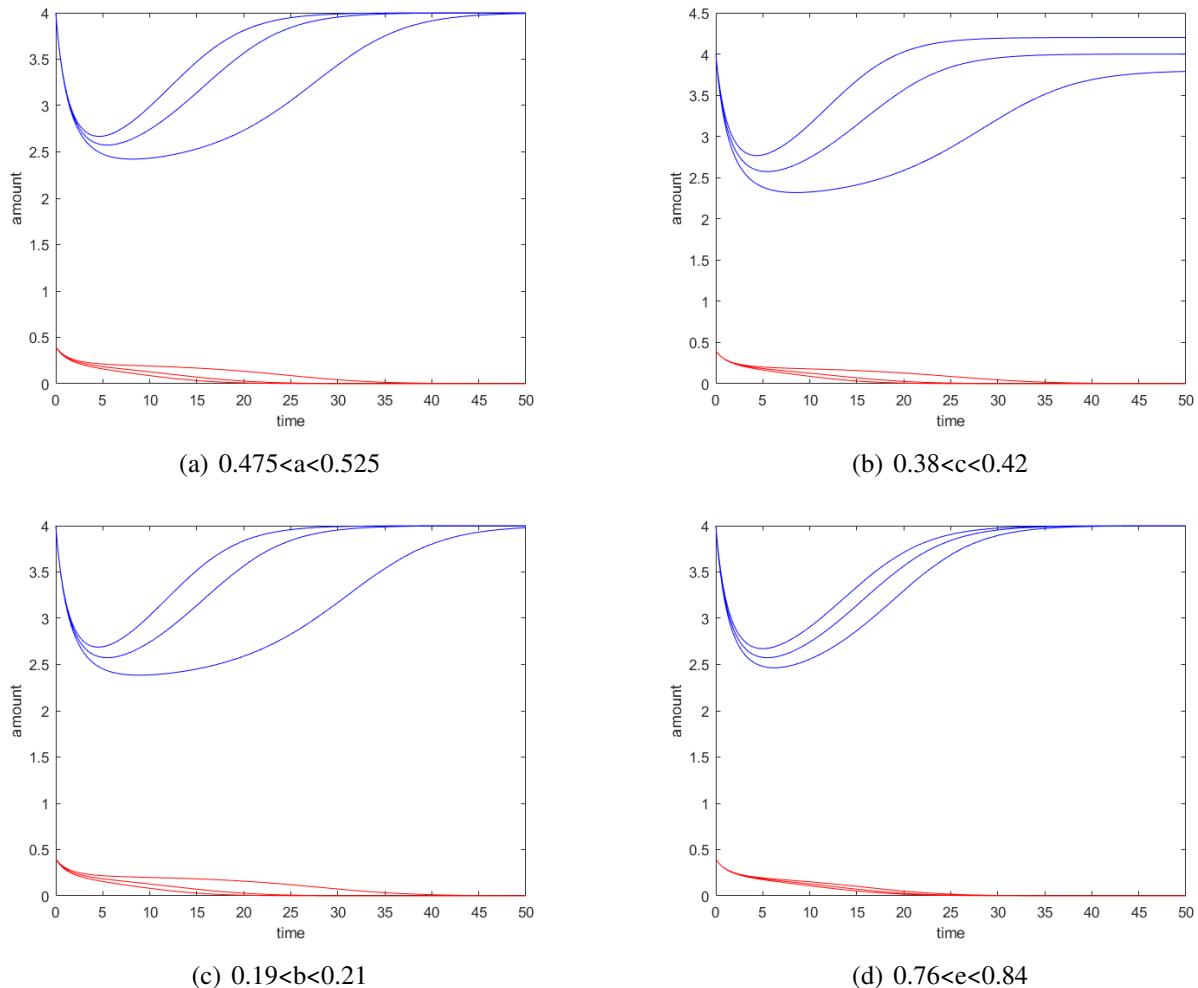


Figure 21: Expansion Model Response to Input Variation

Fig.21 shows how our expansion model responds to input variation. For AGH, Our model has reasonable stability under input variation. For other insects, The internal competition coefficient of AGH results in the most significant fluctuation, whereas the internal competition coefficient of local species leads to the least fluctuation.

6.2 Classification Model

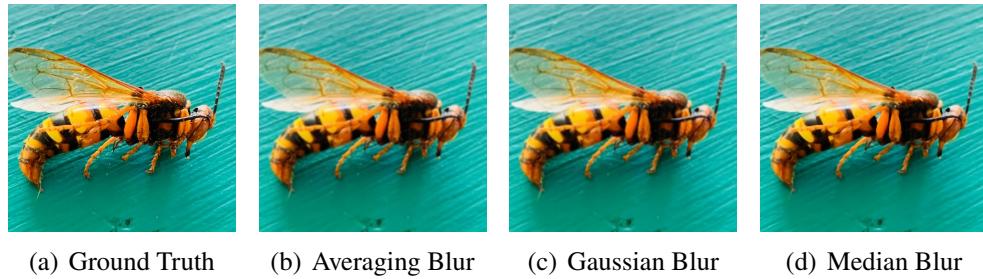


Figure 22: **Clean and Blurred Images. All blur has kernel size of 5.**

In this subsection, we conduct sensitivity analysis to the classification model. We first apply disruptive factors, including averaging, Gaussian Blur, and Median Blur, to the groundtruth. The resulting images are in Fig.22.

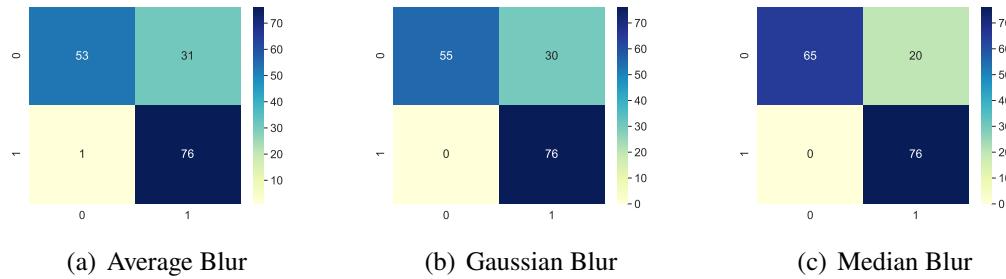


Figure 23: **Confusion matrix under different blur conditions**

In Fig.23, We display the confusion matrix under different blur conditions. Under all blur conditions, our model has 0 or only 1 false negatives. Meanwhile, it has some false positives, but its number is significantly less than true positives and true negatives. That means our model can recognize Asia Wasps under disturbance.

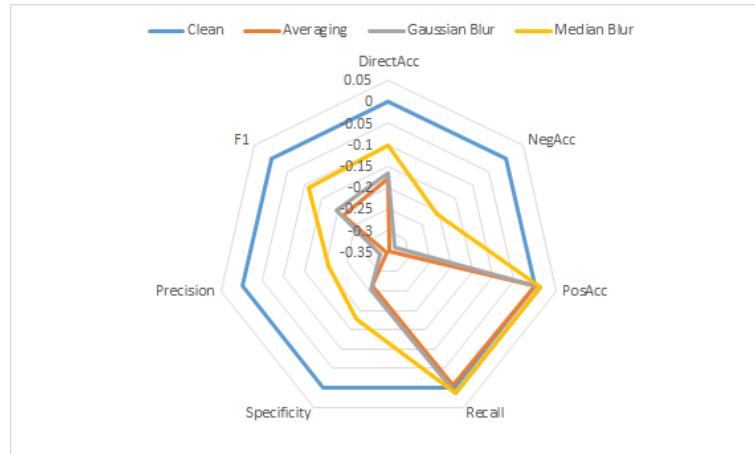


Figure 24: **Radar Plot under different blur conditions**

We then build a radar plot in Fig.24 to assess model performance under disturbance. All blurring has no significant influence on positive accuracy and recall. However, Averaging Blur and Gaussian Blur will slightly decrease direct accuracy, specificity, and f1 score and significantly reduce negative accuracy and precision. Meanwhile, median blur results in less degradation.

7 Model Reflection

7.1 Strengths

- Consistent classification and predictive ability.
- Quick and Convenient Model Update from additional data.
- High Simulation Accuracy with seasonal periodic change, Brownian motion and potential propagation coefficient
- Efficient Analysis of complex images and videos
- Great classification performance and robustness
- Novel binary classification from pure image quality.

7.2 Weaknesses

- Idealize inter-species competition and override human intervention factors.
- Possibly inconsistent potential propagation coefficient update in cellular automation
- Over-Simplified factors others than dual species competition between AGH and Bees.
- Sub-optimal interpretability at intermediate layer's feature processing
- Subjective Image Indicator for binary classification.

8 Memorandum

To: The Washington State Department of Agriculture

From: Team #2116473 of 2021 MCM

Date: Feb 9, 2021

Subject: Recommendations and Solutions to Potential Bioinvasion of Asian Giant Hornet

Honorable Washington State Department of Agriculture,

Recently, Asian giant hornet was discovered around Washington State. According to research, this species of hornet could be extremely dangerous and a serious threat that will probably lead to a biological invasion in a short time. To predict the pest's spread and control it in the future, we make a serial of analysis upon current data.

The report has analyzed the Asian giant hornet using statistics, mathematics, and deep learning tools. First, we visualize the geographical distribution of insects, investigate insect populations of different dates and examine their expansion period. Second, we calculate Asian giant hornet numbers at different conditions and simulate its geographical spread. Third, we build an advanced classification model and discuss the reason for widespread mistaken reporting. Besides, we demonstrate the robustness of our models under various input variations. Finally, we recommend measures for successfully addressing Asian giant hornet.

There are insightful findings from our analysis.

- Most Asian giant hornet is reported at west part of Washington. We use a deep learning model with 0.975 testing accuracy for verification. However, we only found 14 Asian Giant Hornet among 4400+ reports.
- Local insects have a monthly active cycle of about 7.71 days. Asian giant hornet breed at a rapid rate every Spring.
- Asian giant hornet's explosion rely on high potential propagation coefficient, according to our geographical spread analysis. However, AGH are inactive all year along and can hardly reach that condition.
- Asian giant hornet's fragile at early explosion. Timely human interference (about 30 days earlier than spreading) could efficiently control the AGH population and protect other insects.
- Light-weight deep learning models is far more accurate (>0.97) for classifying Asian giant hornet than human observations (<0.01).
- It is most efficient to update our models (with pretrained weight saved) after every winter.
- Reported image with high intensity change is a good metric for subjectively classifying AGH, whereas images with high luminary is misleading.

We finally provide recommendations based on our analysis result.

- Clear Asian giant hornet hives at early spring. You may use fire fighting drones to accelerate that process.
- Timely Classify and Report Asian giant hornet using light-weight convolutional neural network. Once you detect the signs of an outbreak, you must take action within 30 days.
- Ignore reports that has low image intensity change or has noisy backgrounds. They are almost certainly fake Asian giant hornet images. Be careful of misguiding image with high contrast.
- Update classification and explosion models at every winter. At that time, you should have collected data till the end of a year.
- There is no need for massive panic about Asian giant hornet. Even among large numbers eyewitnesses reports over a long period of time, there are few AGH.
- The public should be informed of AGH to reduce unnecessary review time. Meanwhile, you could prioritize images with high TOPSIS score.

We suggest exploiting updating our model in the following ways:

First, you can use the non-reference image quality metrics and Topsis score to determine the order of image audit. Then you can use SqueezeNet to identify if a image contains an Asian giant hornet. For images that are difficult to discriminate, you can use manual screening.

After screening the images, you should collect the data on the adaptation of Asian Giant Hornets in Washington to update the differential equations for simulation. You should also pay attention to the number of Asian giant hornet, at early Spring. Finally, the government should take prompt action when there is a evidence of outbreak.

We hope our analysis and advice will help to control spreading of Asian giant hornets and generating better plans for future issues. You are welcome to contact us at any time.

Yours Sincerely,
Team #2116473 of 2021 MCM

References

- [1] Jacqueline R Beggs, Eckehard G Brockerhoff, Juan C Corley, Marc Kenis, Maité Masciocchi, Franck Muller, Quentin Rome, and Claire Villemant. Ecological effects and management of invasive alien vespidae. *BioControl*, 56(4):505–526, 2011.
- [2] Alberto J Alaniz, Mario A Carvajal, and Pablo M Vergara. Giants are coming? predicting the potential spread and impacts of the giant asian hornet (*vespa mandarinia*, hymenoptera: Vespidae) in the usa. *Pest Management Science*, 77(1):104–112, 2021.
- [3] Gengping Zhu, Javier Gutierrez Illan, Chris Looney, and David W Crowder. Assessing the ecological niche and invasion potential of the asian giant hornet. *Proceedings of the National Academy of Sciences*, 117(40):24646–24648, 2020.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [6] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [7] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- [8] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [9] Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.
- [10] Albert. no. <https://github.com/Albert-1-w/no>, 2021.
- [11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [12] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [13] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.
- [14] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.