Invasive Insect Species Classification and Propagation Prediction Model

1. Conservation Problem

The conservation challenge we aim to address is the significant threat posed by invasive species to ecosystems, biodiversity, and the economy. Invasive species disrupt natural ecosystems by competing with native organisms, altering food webs, and reducing biodiversity. This imbalance weakens ecosystem resilience, diminishes essential services such as water purification and soil fertility, and threatens global efforts to combat climate change and protect endangered species. In the United States alone, invasive species cause an estimated $120 billion in annual damages, with $30 billion attributed to invasive insects. These insects impact over 100 million acres of land, attacking crops, plant species, and soil composition, further exacerbating environmental degradation.

Beyond ecological harm, invasive species impose a heavy economic burden by damaging agriculture, forestry, and fisheries, with escalating costs for control and restoration. Addressing this issue, particularly through identifying and tracking invasive insect species, is critical to preventing irreversible ecological damage, safeguarding biodiversity, protecting agriculture, and maintaining the health and stability of vital ecosystems.

2. AI for Conservation

AI is needed for solving the problem of invasive species identification because it excels in processing large, complex datasets efficiently and with high accuracy. Machine learning models, particularly those in computer vision, are able to recognize patterns in images, sounds, and environmental data that indicate the presence of invasive species, far surpassing the speed and accuracy of manual identification by humans. For instance, algorithms trained on image data from sensors, camera traps, and satellite imagery can quickly analyze and classify species in real time, providing vital information on the spread and behavior of invasive species. This technology can detect changes in ecosystems more quickly than traditional methods, allowing for earlier intervention and more effective management strategies.

In addition to visual data, AI can analyze environmental factors such as temperature, soil composition, and humidity, which may correlate with the presence of invasive species. By integrating these data sources, AI can generate a comprehensive view of species distribution and predict future spread patterns, offering a proactive approach to conservation. The ability to analyze and interpret vast datasets, including historical data, gives AI a unique edge in forecasting the impact of invasive species. AI’s scalability and continuous improvement through exposure to new data ensure that it remains a powerful tool for ongoing monitoring and mitigation, providing critical support to conservationists and decision-makers.

Non-AI alternatives include many different time consuming and labor intensive methods. Physical traps and sampling are when trapping invasive species is used to monitor their prescience. The traps are placed in different environments to capture species for identification. This method has the limitations of the size of the area being monitored and the sample not always being comprehensive. Manual identification and field surveys are when experts in ecology perform surveys to identify invasive species visually. They inspect plants, insects, or other organisms to identify the species. The limitations this method has are that it could be difficult to distinguish from native organisms. Lastly, another example of a non-AI method is chemical control and pesticides which are sometimes used to manage invasive species. Particularly, they are used in agricultural settings which can help reduce the invasive species populations. This can negatively affect the environment with all of the pesticides and chemicals going into the soil.

3. Data

Several datasets are available to develop and test solutions for identifying invasive insect species, each offering valuable resources in different formats and with varying accessibility. EDDMapS (Early Detection & Distribution Mapping System) provides a comprehensive collection of invasive species data across the U.S., with records on species occurrences and distributions. It is available through an API and can be downloaded in CSV, GeoJSON, or ESRI Shapefile formats. Data is contributed by public reports and field surveys, ensuring wide coverage. The platform adheres to the CARE principles by promoting collective benefit through open access to data, allowing multiple stakeholders, including land managers and researchers, to contribute and utilize the information. Contributors retain control over their data, with guidelines ensuring ethical and responsible reporting.

Another thorough and widely accessible option is the iNaturalist dataset. This website is an extremely expansive source, containing information on multitudes of insects across the globe. It is based on community contributions and allows anyone to upload a photo to be reviewed by experts in the community and assigned a verified identification. It is extremely accessible for download, as it has an API that provides RESTful endpoints that enable users to export data using a vast selection of criteria and in a variety of output formats, such as csv, json, KML, or as a direct download into programming environments like R. The dataset's provenance aligns closely with the CARE principles. Collective Benefit is promoted by iNaturalist's open-access policy, allowing researchers, conservationists, and the public to utilize the data for a variety of environmental and conservation purposes. By making the data freely available, iNaturalist contributes to global efforts to monitor biodiversity and invasive species. Regarding Authority to Control, contributors maintain control over the data they submit, deciding whether to share their observations publicly or keep them private. However, iNaturalist encourages transparency and collaboration while respecting contributors' rights. The platform also upholds Responsibility by ensuring data accuracy through community verification, which guarantees the reliability of the data for scientific research and conservation initiatives. Ethics is a key focus for iNaturalist, as it has explicit terms of use and ensures users adhere to those guidelines to prevent misuse or exploitation of data. Furthermore, the platform protects sensitive species location data to prevent exploitation of or harm to vulnerable species as well as abuse of private property. Overall, the iNaturalist dataset embodies the CARE principles by fostering responsible, transparent, and ethical data sharing that benefits personal projects, scientific research, and global conservation efforts.

The iNaturalist dataset was selected for this project because of its extensive collection of species observations, particularly those of invasive insects. This dataset is invaluable for training models to identify species using images and accompanying metadata such as geographical coordinates, species names, and observation dates. It draws from a diverse pool of contributors, including both amateur citizen scientists and professional researchers, making it a rich resource for studying biodiversity and tracking the spread of invasive species. iNaturalist gathers observations through its platform, where users can upload images and data of species they encounter in the wild. These submissions are often verified by a community of experts, ensuring data accuracy and credibility.

4. Annotations

The iNaturalist dataset provides several key annotations that enhance the usefulness of the data for species identification and tracking. It offers basic annotations, such as observation id, description, time zone, url, and other details pertaining specifically to the website’s interpretation of the observation. Additionally, it offers extensive geographical data such as the town, county, state, and country of each observation, in addition to each observation being marked by its latitude and longitude, and sometimes altitude, to enable species distribution to be mapped across landscapes and regions. Additionally, date and time annotations allow for temporal analysis of species occurrences, helping track seasonal patterns or the spread of invasive species over time. Finally, it also provides extensive options for taxonomic data.

Image annotations are another vital feature, as each observation typically includes one or more photos of the species, sometimes annotated with descriptive notes about the species' appearance, behavior, or other features. The dataset also benefits from community verification, where users can help confirm or suggest corrections to species identifications, and observations that achieve consensus are marked as "Research Grade," ensuring reliability. In some cases, habitat and location information is provided, offering context on the environments where species are observed, which can be crucial for understanding ecological preferences or behaviors. These comprehensive annotations contribute to a more detailed, community-verified dataset, making it a powerful tool for biodiversity monitoring and invasive species management.

While there was no explicit need for us to acquire more data, as iNaturalist allowed us to garner a dataset size of 171,545 images for the target invasive species and 67,395 images for the native lookalike species, there is the inherent issue of the imbalance of this data that, while still allows us to achieve a working model, hinders our abilities to achieve absolutely accurate and comprehensive results. This will be discussed in greater depth in the next question.

5. Data Evaluation

The iNaturalist dataset has several statistical properties that make it a valuable resource for addressing the conservation problem of invasive species identification. As an open and participatory platform, the dataset includes millions of observations submitted by users worldwide, and continues to grow rapidly. The observations are distributed across a wide range of species, habitats, and geographical areas. This broad scope helps capture a diverse array of environmental conditions, contributing to the dataset's representativeness of global biodiversity.

In terms of statistical properties, the dataset features observations with varying levels of identification accuracy. Initially, many species are identified at a higher taxonomic level (such as genus or family), but with community verification, the dataset's reliability improves over time. Observations are generally more concentrated in areas with active user participation, such as urban and suburban environments, though more remote locations are also represented. Thus, while iNaturalist users can be found across the globe, there is a natural bias towards those who have access to the iNaturalist application and can have a device that photos can be captured on and uploaded from. This prevents more rural areas from being represented as conveniently and evenly. Temporal data shows seasonal trends in species sightings, which is crucial for tracking the spread of invasive species and understanding their behavior over time.

The dataset's representativeness of the conservation problem is strong but not without limitations. While it includes a broad range of species, invasive insects might not be equally represented across all regions, as some areas have more active monitoring for invasive species than others. Additionally, data availability may vary depending on geographic region, with certain areas having higher numbers of observations due to there being more users in those locations. Due to this inherent imbalance of the dataset and in addition to how it consists of mostly arbitrary contributors photographing insects largely through chance encounter, it is not a truly systematic or organized, and thus accurate, representation of the locations and populations of every species. Despite these limitations, the iNaturalist dataset's broad participation, community-driven data verification, and global scope make it a useful resource for monitoring and identifying invasive species, though supplemental datasets may be needed to fill in gaps or focus on specific regions or species. The iNaturalist dataset has robust statistical properties, with a large and diverse number of observations, but its representativeness for the conservation problem could be enhanced by targeting more specific regions or species with targeted data collection efforts.

6. Baseline & AI Techniques

Our initial baseline algorithm consisted simply of using BioCLIP on datasets of both target invasive species and native lookalike species in the northeast. Our preliminary approach was to use BioCLIP on only the invasive species dataset and on only a dataset of observations in New York State, resulting in a model of limited geographic scope that had only learned the features of the target invasive species and could only recognize whether or not an insect was invasive. However, we then expanded our training to incorporate a dataset of native species with striking similarities to the target invasive species in addition to the original invasive species dataset. Doing so enabled our model to learn features of non-target classes so it could not only recognize invasive species, but also distinguish them from their native lookalikes and in turn reduce false positives and improve generalization in real-world scenarios.

We also expanded our geographic range to the northeast region of the United States in order to obtain a greater picture of the distribution of invasive species across a greater area, and enable the future design of a predictive model. The key steps of our baseline approach included dataset curation, species filtering, then classifier training. First, we harmonized the datasets for the native and invasive species and merged them into a single dataset. We then imposed a ceiling of 500 images per species to balance the dataset and avoid overrepresentation. We defined the target species, then used BioCLIP's CustomLabelsClassifier to process images and predict species labels. These predictions were tracked against ground truth labels for performance evaluation. While we retained this approach of training against both invasive and native lookalike species in our actual implementation, we also used a geospatial model to plot species predictions after being able to identify species and where they were sighted.

In our implementation, we significantly expanded upon the baseline model by incorporating advanced machine learning techniques and adopting a novel hybrid approach. The model utilized a hybrid LSTM-CNN Sequential architecture, combining the strengths of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. The CNN component effectively extracted spatial patterns and features from the sequential data, while the LSTM component captured temporal dependencies, such as seasonal or time-related patterns in insect sightings. Dense layers were employed to model non-linear relationships between features, and regularization techniques like dropout and l2 regularization were incorporated to prevent overfitting and enhance generalization. To prepare the data for modeling, we employed time feature engineering to extract information such as year, month, day, and a normalized time_percentile from timestamps, capturing temporal dynamics effectively. Normalization of both features and targets ensured numerical stability and improved model convergence.

Beyond the model architecture, we implemented advanced error metrics and visualization techniques to evaluate performance. The Geometric Mean Error was introduced to quantify centroid accuracy, while percentile-based error metrics assessed the spread and variability of predictions relative to actual data. To further interpret the results, geospatial visualization techniques were employed. Confidence ellipses were used to visually compare the spread of predictions with actual observations, and bounding boxes based on ellipses dynamically adjusted the map scaling to provide meaningful insights. Our refined model outperformed the baseline in several critical ways. By incorporating native lookalike species and expanding the geographical range, the model demonstrated broader generalization, effectively learning to differentiate invasive species from similar non-target species and significantly reducing false positives. Geospatial evaluation, a feature absent in the baseline, was a notable improvement. Metrics such as geographic error highlighted the accuracy of predicting the centroid of species observations, while percentile-based metrics demonstrated better alignment with the actual spread of the species.

Additionally, visual interpretations using confidence ellipses provided an intuitive and actionable comparison of model predictions versus ground truth data. The inclusion of temporal and geospatial features improved predictive performance and made the model scalable to larger datasets and broader geographic areas. Examples of the output of the model can be seen in the figures displayed at the bottom of the paper.

7. Training, Validation, and Testing

   To train the model effectively, the dataset was split into three parts: training, validation, and testing sets, with a ratio of 60%, 20%, and 20% respectively. This split ensured that the model had sufficient data for learning, hyperparameter tuning, and unbiased performance evaluation. The training set was used to optimize the model weights by minimizing the loss function (mean squared error, MSE). The input features included both spatial data (latitude and longitude) and temporal data (year, month, day). These features allowed the model to capture geographical and seasonal patterns in invasive species sightings. Numerical features were scaled using MinMaxScaler to ensure uniformity. Sequences of 12 time steps were constructed to provide temporal context, creating an input shape suitable for the LSTM-CNN hybrid architecture. The validation set played a crucial role in hyperparameter tuning and model selection. During training, early stopping was used to prevent overfitting, with a patience of 5 epochs. The model saved its best weights based on the lowest validation loss, ensuring optimal generalization capability. Validation loss and Mean Absolute Error (MAE) were tracked during training to monitor the model's performance and detect potential overfitting or underfitting. The testing set was reserved for the final evaluation of the model's performance. Unlike the training and validation sets, the test set was not involved in any model adjustments. The test set provided a realistic measure of how well the model performed on unseen data. Metrics such as Geographic Mean Error and Percentile-Based Error Metrics were calculated to assess both the central accuracy (e.g., centroid prediction) and the spread of the predictions compared to the actual data. The results revealed test losses from 0.97 to 1.83, indicating effective learning. The test evaluation showed mean geographic errors from 7.34 to 87.51 kilometers given the amount of data available for each species, which is acceptable given the small deviations in latitude and longitude (e.g., 2–5 degrees). These metrics highlighted the model's ability to generalize well to new data while maintaining geographical accuracy.

8. Model Application

   The solution provides a significant contribution to addressing the conservation challenge of monitoring and managing invasive species. By accurately predicting the spatial distribution of invasive species, the model supports ecological management efforts in several ways:

   1. **Mapping and Surveillance**: The predictions can be used to create accurate distribution maps, enabling conservationists to target areas at high risk for invasive species spread. This information is crucial for planning resource allocation, such as deploying field surveys or eradication efforts.

2. **Risk Mitigation**: By identifying geographical hotspots, the model can guide policymakers to focus on high-risk zones for implementing biosecurity measures. This proactive approach reduces the likelihood of invasive species causing widespread ecological and economic damage.

3. **Educational Tools**: The results and visualizations, such as confidence ellipses and centroid predictions, can be shared with local communities to raise awareness about invasive species and their impacts.

The main issue with the model is that it is not good at generalizing and cannot capture the variance in longitude and latitude, so the predicted spread ellipses are consistently smaller than the actual spread. However, if deployed, the predicted spread can be scaled up on future data to better capture the variance seen in the actual spread. This is mostly due to potential overfitting and a lack of data. While the solution is reasonably accurate, incorporating human guidance can further improve its effectiveness. For example, expert ecologists can validate the model's predictions and provide additional context about species-specific behaviors or environmental factors that may influence their spread. Regularly updating the training data with new observations will also help the model adapt to changing patterns over time. To address the inherent limitations of a machine learning model, interpretability techniques can be employed. For instance, feature importance analysis could identify the most influential factors driving the predictions, helping conservationists understand the model's decisions. Moreover, coupling the model's outputs with expert opinions ensures that the results are actionable and aligned with on-the-ground realities. Although no model is perfect, the combination of advanced AI techniques, meaningful error metrics, and human oversight makes this solution a powerful tool for managing invasive species in the northeastern United States.

9. SOTA

Invasive insect species pose severe threats to biodiversity, agriculture, and global economies. Effectively managing their spread requires not only identifying species accurately but also predicting their future geographic propagation. Traditional approaches in predicting invasive species distribution have relied heavily on Species Distribution Models (SDMs), which utilize environmental variables and species occurrence data to estimate suitable habitats. However, SDMs often make simplifying assumptions about equilibrium between species and environments, missing the dynamic nature of invasive species propagation. These limitations have driven the exploration of advanced methodologies to predict invasion patterns and classify invasive species. Recent research has improved the prediction of invasive species' spread using machine learning techniques such as decision trees, neural networks, and hybrid models. For example, studies combining remote sensing data with machine learning have classified invasive habitats, such as mapping Kudzu infestations with multispectral data. Network-based models have simulated invasive species dispersal by examining human-mediated pathways and natural spread. While these approaches enhance predictive accuracy, they often focus solely on species propagation without addressing classification. On the classification side, models like BioCLIP excel at identifying invasive insect species using image datasets but are geographically and ecologically constrained, limiting their generalization. Although strides have been made in species classification and propagation modeling independently, few efforts integrate both into a unified framework. Our approach addresses this gap by uniting invasive species classification and propagation prediction into a cohesive framework. Leveraging BioCLIP ensures high classification accuracy for invasive insect species, while a hybrid LSTM-CNN model predicts their geographic spread. The LSTM captures temporal dependencies like seasonal patterns in sightings, while the CNN extracts spatial patterns. This dual-model integration leverages the strengths of both classification and geospatial modeling for a comprehensive solution. Our model incorporates advanced features and metrics that challenge the state of the art. Temporal and geospatial features, including engineered variables like year, month, normalized time_percentile, latitude, longitude, and county/state encodings, enhance the model's understanding of spatial and temporal dynamics. Geospatial metrics like mean geographic error, confidence ellipses, and percentile-based spread evaluations quantify centroid precision and variability, providing insights beyond standard accuracy metrics.

By addressing classification and propagation simultaneously, our model tackles two interconnected challenges that are often studied independently. This integration provides actionable insights into species identity, predicted locations, and potential spread patterns. These advancements enable conservationists and policymakers to make more informed decisions, improving monitoring efforts and mitigation strategies. While prior research has addressed classification and propagation separately, our model bridges this divide, offering a holistic solution that aligns with real-world complexities. By combining advanced machine learning techniques and innovative metrics, our approach not only sets a new benchmark for predictive accuracy but also demonstrates the practical potential of AI in combating invasive insect species.

10. Deployment

One possibility of how our system could be deployed is through being hosted on a cloud service so users can access the system over the internet without needing to maintain their own physical servers or infrastructure. To accommodate this, we would likely need to implement some interface like RESTful APIs. Another way we could deploy it is by constructing a web-based dashboard for conservation agencies so they can visualize real-time species distributions and geospatial predictions, which would assist in their ability to monitor the spread of invasive species and plan for interventions. We could also attempt to expand the model's capacity to process bulk data submissions such as from iNaturalist to improve the model's predictions and geospatial analysis and allow it to be used more flexibly. Similarly, the model can be integrated into existing applications as a feature that plots and displays species observations in real time. A logistical challenge of this is cloud services require regular management and scaling to accommodate increasing data and user demands, which would likely also correspond to increased costs as usage and data storage demands increase. One main challenge of any of these options involving real-time display is that real-time visualization demands continuous data ingestion, so we would need to ensure our program can handle an ever-increasing dataset size and that we are able to manage any latency issues that may be caused as a result of this. Additionally, should data originate from sources other than iNaturalist, we would need to be able to integrate species observations from diverse sources which may have different formats and fields. Thus, we would need to consider developing a data harmonization pipeline that standardizes the formats of incoming data and flags incomplete or inconsistent data for review. Additionally, should this visualization feature reach more widespread applications and originate from more diverse sources, we would have to continuously ensure that geolocation data is not revealing any sensitive information such as valuable species habitats or private, personal locators. To address this challenge, we would have to encrypt sensitive data during transmission and storage and ensure all data is anonymous.

11. Limitations

One issue associated with long-term maintenance of our system is that over time, the environment will change and may be subject to new invasive species, habit shifts, and unpredictable patterns that make our model outdated and less accurate. Additionally, as previously mentioned, hosting our system on a cloud service can incur costs and would require consistent management as usage gets expanded to new applications. Similarly, further development would likely entail expanding the species list and geographical domain to a larger territory, and long-term storage of large datasets can become costly and difficult to manage. This would also entail retraining the model with new datasets and features, and we would need more powerful computation resources in order to process the larger datasets associated with greater species and geographical range. The model's technology would also have to be continuously monitored to ensure its dependencies, such as the python libraries used, do not become obsolete or outdated.

12. Accessibility

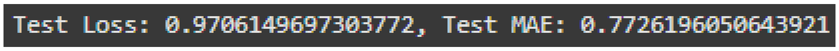
Currently, our data is being sourced from iNaturalist, and our datasets, model, and weights are all being stored in a Github repository. This is in accordance with FAIR principles primarily because all the components of our system are findable. Github is a central, publicly accessible platform, allowing our data and models to be clearly organized and easy to locate. All of our components are also accessible, as hosting on Github ensures public accessibility and supports the open viewing and download of all the resources for our system. It is very interoperable, as it uses a standard CSV format for its datasets and written in Python, ensuring compatibility with various systems. Similarly to how the standard structures and formats of our system make it interoperable, those features also make it very reusable as it is designed and shared in a way that is easily understandable, adaptable, and applicable to other problems. Other developers can use our model to identify different species, retrain it on new datasets, or extend it for different purposes. Finally, iNaturalist itself adheres to the FAIR principles. iNaturalist data is findable, as it provides robust search criteria that allows users to be matched to target species through a variety of queries such as common name, scientific name, geographic location, data, user, and other attributes. It is extremely accessible, as all data is openly accessible via the website and API for public, personal, and academic use. It is also extremely easy to export data from iNaturalist for these purposes. This data is also very interoperable, as exported data is provided in widely used, standard formats such as CSV and JSON which make it easy to integrate with most tools and applications. Finally, iNaturalist data lends itself very well to reuse, as it is bound by Creative Commons licensing that facilitates the ethical reuse of data and is based on a public, community-driven culture that encourages the sharing and use of uploaded observations. Additionally, the extensive metadata it offers per observation allows data to be easily understood and used in various contexts.

13. Ethics

The most glaring ethical, privacy, and security concern of our system is the use of geolocation data. Sharing or storing detailed geolocation could inadvertently expose sensitive habitats to exploitation and reveal private user data. To address this, we would need to anonymize data submission, encrypt sensitive geospatial data during storage and transmission, or ensure only verified users have access to geolocation data. Similarly, to combat the potential for misuse that tracks or harms sensitive wildlife, we would need to again restrict access to sensitive geospatial data or aggregate location information to broader regions, such as at the county level, as opposed to providing exact coordinates. Additionally, misidentification of native species as invasive could lead to unwarranted harm, so it’s crucial to ensure high accuracy and thorough training of the model with diverse datasets. It’s also important to manage bias in the model by ensuring that well-balanced datasets are used throughout its expansion that combat overrepresentation of certain species or geographic areas. One thing to keep in mind about this is that there is often a correlation between lack of technological access and the most underrepresented areas, and so it would be vital when navigating this issue to ensure any data acquisition in these areas does not infringe upon the local or indigenous people inhabiting that land and is done in proper collaboration. Finally, should the model be further deployed, it will be important to ensure there is not an over-reliance on the AI system that excludes ecological experts from the decision-making process, as the model should not be used as a replacement for expert judgement.

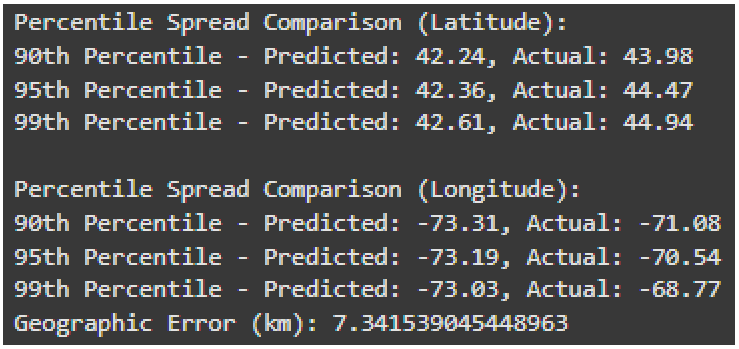
Results for the Asian Lady Beetle (41,803)

Test Loss & Test MAE



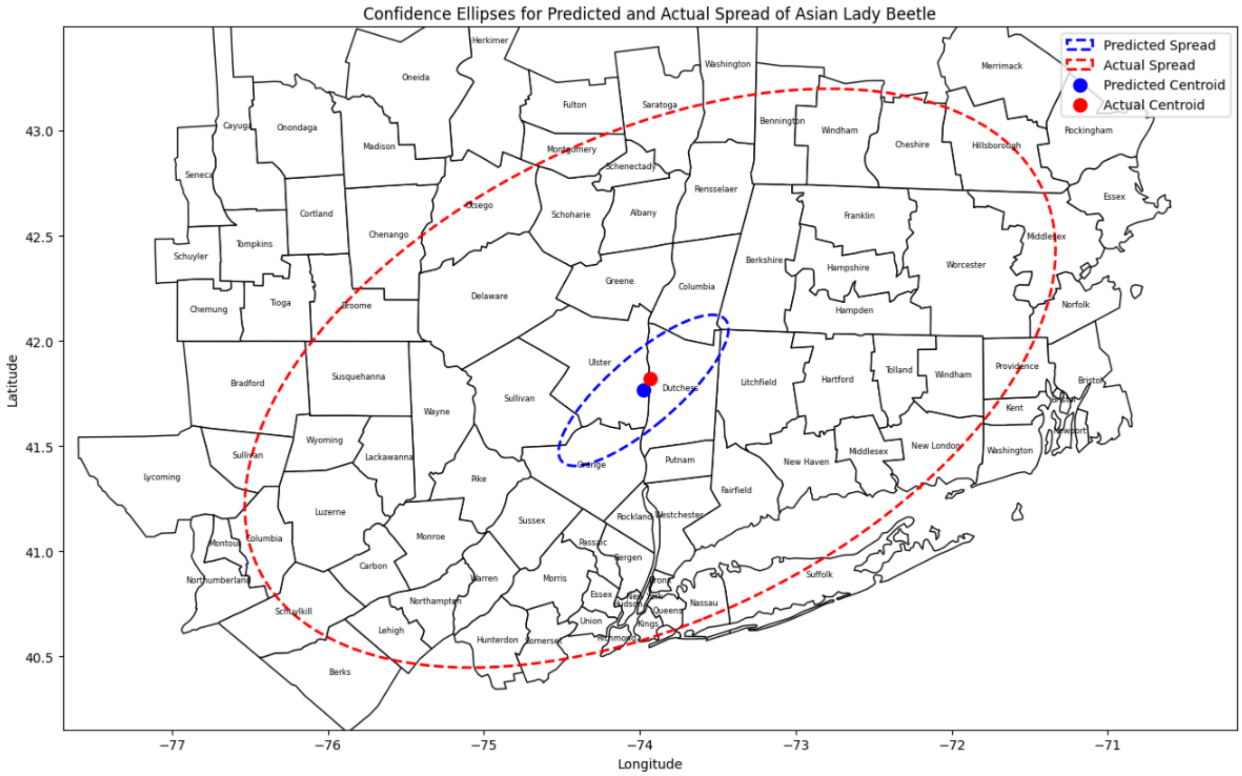
```
Test Loss: 0.9706149697303772, Test MAE: 0.7726196050643921
```

Percentile Spread Comparison in Latitude and Longitude for Error in Variance & Mean Geographic Error for Error in Centroid



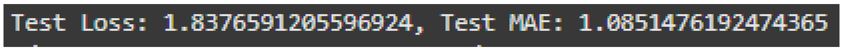
```
Percentile Spread Comparison (Latitude):
90th Percentile - Predicted: 42.24, Actual: 43.98
95th Percentile - Predicted: 42.36, Actual: 44.47
99th Percentile - Predicted: 42.61, Actual: 44.94

Percentile Spread Comparison (Longitude):
90th Percentile - Predicted: -73.31, Actual: -71.08
95th Percentile - Predicted: -73.19, Actual: -70.54
99th Percentile - Predicted: -73.03, Actual: -68.77
Geographic Error (km): 7.341539045448963
```



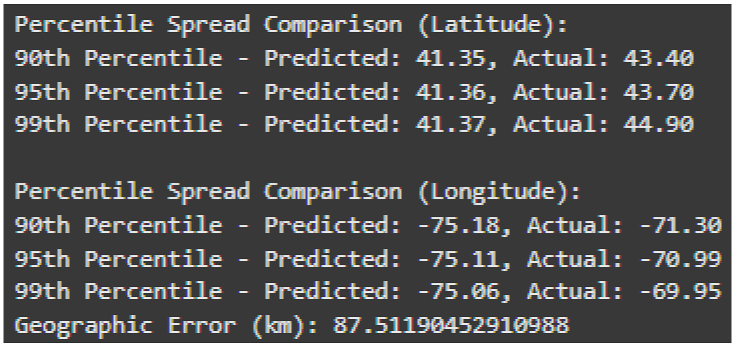
Results for the Emerald Ash Borer (1,272 sightings)

Test Loss & Test MAE



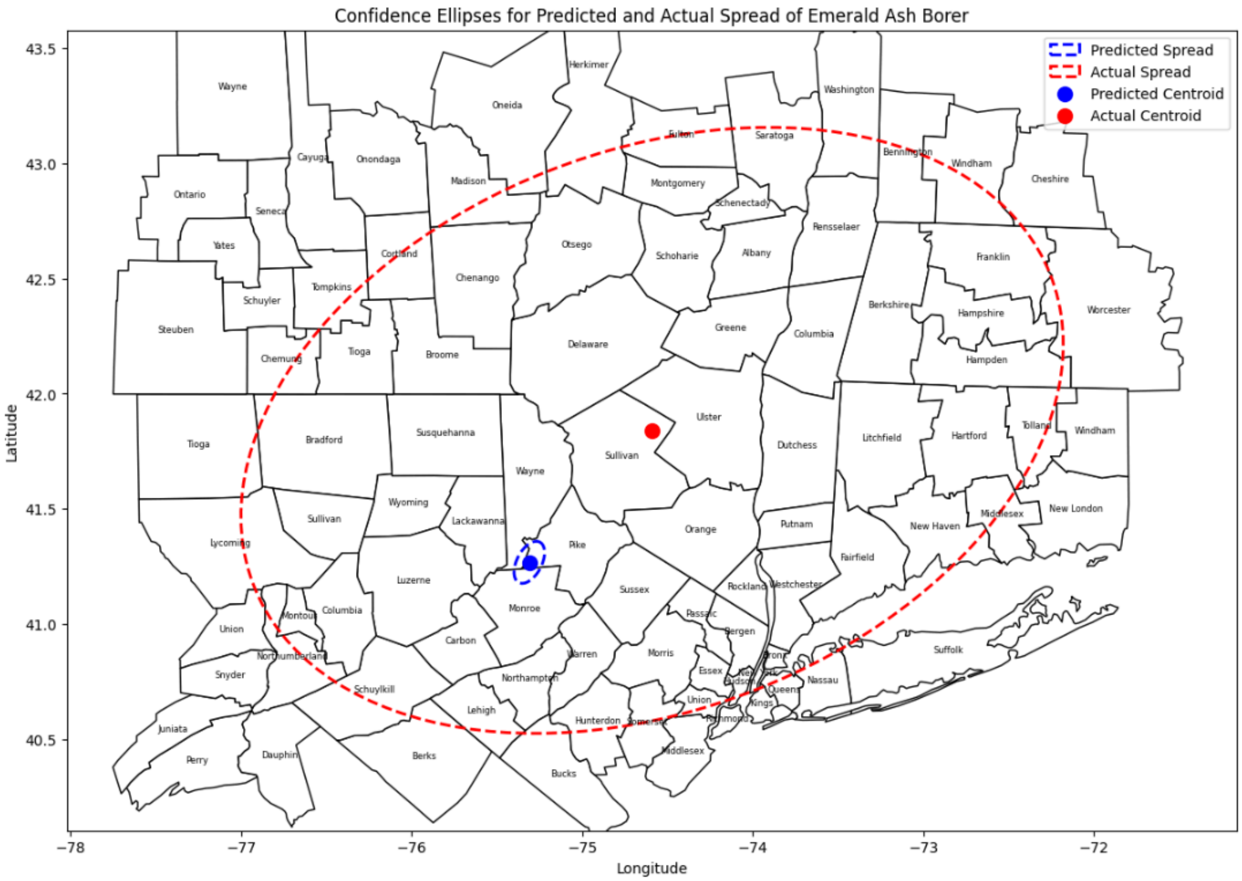
```
Test Loss: 1.8376591205596924, Test MAE: 1.0851476192474365
```

Percentile Spread Comparison in Latitude and Longitude for Error in Variance & Mean Geographic Error for Error in Centroid



```
Percentile Spread Comparison (Latitude):
90th Percentile - Predicted: 41.35, Actual: 43.40
95th Percentile - Predicted: 41.36, Actual: 43.70
99th Percentile - Predicted: 41.37, Actual: 44.90

Percentile Spread Comparison (Longitude):
90th Percentile - Predicted: -75.18, Actual: -71.30
95th Percentile - Predicted: -75.11, Actual: -70.99
99th Percentile - Predicted: -75.06, Actual: -69.95
Geographic Error (km): 87.51190452910988
```


Confidence Ellipses for Predicted and Actual Spread of Emerald Ash Borer
Predicted Spread
Actual Spread
Predicted Centroid
Actual Centroid
Latitude
Longitude
43.5
43.0
42.5
42.0
41.5
41.0
40.5
−78
−77
−76
−75
−74
−73
−72
Wayne
Herkimer
Oneida
Washington
Cayuga
Onondaga
Fulton
Saratoga
Bennington
Windham
Cheshire
Ontario
Madison
Montgomery
Seneca
Schenectady
Yates
Cortland
Otsego
Rensselaer
Franklin
Chenango
Schoharie
Albany
Tompkins
Schuyler
Steuben
Berkshire
Hampshire
Worcester
Greene
Columbia
Chemung
Tioga
Broome
Delaware
Hampden
Ulster
Tolland
Windham
Tioga
Bradford
Susquehanna
Sullivan
Dutchess
Litchfield
Hartford
Wayne
Wyoming
Middlesex
New London
Sullivan
Lackawanna
Putnam
New Haven
Lycoming
Orange
Pike
Fairfield
Luzerne
Sussex
Rockland
Westchester
Montour
Columbia
Monroe
Passaic
Union
Carbon
Bergen
Northumberland
Warren
Morris
Bronx
Suffolk
Snyder
Essex
Nassau
Schuylkill
Northampton
Hudson
Queens
Union
Lehigh
Kings
Hunterdon
Somerset
Richmond
Juniata
Perry
Dauphin
Berks
Middlesex
Bucks