Save Bees Save Planet: Perceptive anticipation to home pollinators in an ideal green playland using machine intelligence

Adithya Ram Ballem
dept. of Computer Science
Georgia State University
Atlanta, GA
aballem1@student.gsu.edu

***Abstract*—A discerning visual system with knowledge engineering of geographical and spatial data of a region opens up new possibilities in agriculture and improving world vegetation. Pollinators are essential ecological survival beings, and humanity cannot live without them. However, our inevitable human activity is accelerating biodiversity loss, and these pollinators are experiencing a decline, including insects like bees and birds. This project presents a novel approach to understanding human activity's effects and geographic location. Analyzing these regions using computational vision and engineering the data of environmental outcomes can allow us to pinpoint definitive green hotspots, homing the pollinator colonies, flourishing their population, and balancing the ecosystem cycle.**

***Index Terms*—pollinators, green, environment, bees, colonies , agriculture, emissions, pesticides, machine learning, crop, area, harvested**

I. Introduction

Bees are responsible for every one-third of the food we consume. They are essential for the people and the planet [3]. The vast majority of plant species, about 90% of the produce, rely on pollinators for plant reproduction. Overall, Bee produce value is estimated at $500 million US Dollars, and the pollination effect of Bees is valued at around $20-$30 Billion US Dollars for added crop value. Bees being a majority contributor out of the pollinators, they are facing a decline due to human activities [5]. Overuse of pesticides and fertilizers in agriculture and other activities like air pollution affects the pollinators' ability to forage [4]. Such pathogens also reduce the lifespan of bees, harming the pollinators and leading to global crop reduction.

Ongoing research in autonomous pollination [2], which utilizes air vehicle pollinators, is essential for preserving and expanding the natural ways of the ecosystem. The research also utilizes computational intelligence of systems machine learning and computer vision to maintain the ecosystem balance using self-sustaining farming methods. However, the concept can be expanded to a vast data set, eliminating the need for autonomous specialties and depending on the world's natural order.

Machine learning is a concept that can allow systems to learn, analyze and adapt according to the situation [1]. It uses different algorithms and statistical learning to analyze and draw inferences from patterns in the data. Similarly, computer vision is a similar concept direction, except that it perceives data in image and video representations and then draws inferences from it.

While the research of autonomous farming uses these concepts of Artificial Intelligence to increase crop production, these systems do not consider the geographical data concerning the terrain, nor do they consider the land trends over the past years. However, by incorporating these geographical and statistical data, pollination can span large areas.

The proposal in this project is to improve crop production and expand the green cover of the environment using computational intelligence, which can predict the region's crop output. Models can utilize the statistical data of land and merge it with geographical data to provide suitable green hotspots favorable for increased and improved farming.

II. Proposed Method

The proposal in this project is to improve crop production and expand the green cover of the environment using computational intelligence, which can predict the region's crop output. Models can utilize the statistical data of land and merge it with geographical data to provide suitable green hotspots favorable for increased and improved farming.

The project to save bees and improve agriculture is a two-phase project. The first statistical approach involves gathering data and numbers on the bee population, and its related data of crop area harvested every year. The data affecting the bees, like pesticides and atmospheric emissions, is also necessary to establish a mathematical relation on how one affects another. The proposed method to improve crop production and increase the pollinator population requires a lot of additional factors, like data, and using this knowledge model that can address these issues and significantly improve pollination.

Initially, it requires analyzing statistical data of the region of land, countries, or states. Assessing information like crop production, air quality, and climate data trends over the years can produce a statistical learning model and provide a reason for the dwindling population of trees and pollinators. The preferable attributes of this data would be to gather all the pollinator-affecting data sets as disassociated as possible. Meaning separated county-wise or other subdivisions of data can inform the geographical mapping in the future than the whole country as a whole.

Understanding data and building machine learning models can emerge from mathematical and statistical proof to understand a region. Once the necessary data is gathered, predictions can be made to understand the correlation between bee population and crop area harvested. The analysis part majorly involves observing the drop or rise in the crop area harvested and appropriate periods of bee population trend.

On the other hand, computer vision can help understand the geographical data of the region, following the observed trend in statistical models. Receding greenery of an area can help us relate to the model predictions with substantial proof for identifying this change in green cover and allow us to act accordingly. With the above statistical discovery, a substantiating increase or decrease in crop area harvested is to be identified by following the trends of the rise or fall of a particular feature. With all the combined information, a new bee colony can be homed at an ideal geo-location.

These AI methods can provide two results. Identify an environmental region's elements and predict the variables that affect them. Secondly, the established regions understood using the combined models can provide hotspots for sustained growth of bee colonies or pollinators and increase the green canopy.

III. Implementation

The research involves a two parts

1) Statistical Approach
2) Geographical Correlation

The statistical approach, as discussed above, involves getting all the data that is affecting the bee population. It also involves observing the effects of this reduced population as a decrease in cultivated crop area harvested over a region.

Geographical correlation is tightly dependent on the statistical approach results. Identifying the reduction in crop area is to be exactly pinpointed on the satellite imagery. Later the relation of the pollinator population with the crop area harvested can be utilized to calculate how much a single pollinator colony or number of bee populations can influence the crop area harvested. With this data, the lost crop harvested area can be recovered by homing a couple of new pollinator colonies. This can save the bees, which are essential beings on this earth and for humanity, and the country can maintain stable food production.

A. Data Gathering

Examining how bee population affects the crop area harvested requires the following data sets from the United States Department of Agriculture(USDA), the Food and Agriculture Organisation of UN(FAO), and World Bank Data [6], [7]. Below is the list of the data sets that have been gathered for the United States Country and aggregated together as a single data set.

- USDAFAO Bees Stock Data
- FAO crop statistics
- FAO Emissions
- FAO pesticide data
- World Bank Data on Honey Produce

B. Data Processing

Data obtained from FAO for crop statistics provide details of all different crop types cultivated in the United States over the years 1965-2021. Many rows have been removed for a single crop type due to the unavailability of the data for many attributes, such as the crop area harvested in a particular year.

There also has been a need for data interpolation for certain attributes of pesticide and bee stock data. While few of the data sets, like crop harvest data, emissions, and bee colonies/produce data, were available between 1965-2020, the other data sets are only available from 1995 to 2020. A total of three different implementations have been done, and results compared for all the cases of linear forward and backward interpolation, spline interpolation, and no interpolation.

Additionally, the data dimensions have also been adjusted for certain data sets as few of them provided data across the whole country while others had state-wise data sets.

- Crop harvest data is available for the United States country as a whole with independent crop data as different rows, and all this data is available over the years. As a result, only bee-pollinated or general pollinated data has been captured based on the economic reports [8]–[11] and aggregated together to display the overall crop area harvested data over the years.
- Emissions data was also available across the country; however, it had subdivisions of different emissions like methane, nitrous oxide, and carbon dioxide emissions. The economic and biological research reports [9]–[11] have been gathered together, and only the emissions responsible for the effect of bee pollination were aggregated and formed a data set.
- The honey data set that was made available by the world bank data, which shows a decline in bee production and bee colony reduction, had a different dimension of the data set. All of the data was represented for every state of the United States over the years. This had to be reduced as entire country data so that an easy uniform aggregation could be made with other data sets.

C. Data Selection

Data that has been aggregated as a single entity for learning contains a variety of sub-datasets. Regarding the issue of the endangered bee population and other effects on these pollinators, appropriate features or data values are selected from the data collection process based on the biological sciences research and the United States Department of Agriculture research reports. [8]–[11]

D. Data Analysis

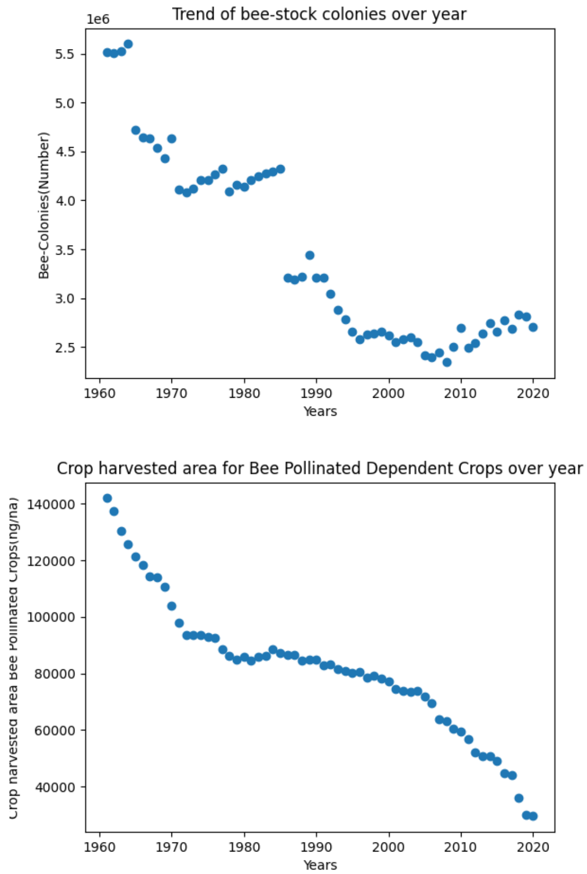

Fig. 1. Data Relations of Bee stock and Crop area harvested over years

On plotting the different data sets over time, the graphs are clear evidence of the endangered bee population. Such a drastic decline had a relative fall in crops harvested for bee-pollinated crops such as rice. Secondly, the bee stock rise between 2010-2020 has also been evident with its relation in bee produce data between these years. The crop area harvested shows a linear decline from 1965 to 2020. Such a change is synonymous with the declining population of bees, substantiating the crop area decrease.

Emissions data, although provided all kinds of emissions, a certain type of emissions such as nitrous dioxide and carbon dioxide data together correlate with declining crop harvested area. The fall in emissions in the year 1990 corresponds to a rise in the bee population. The number of bees increased with the colonies increase.

The pesticide data set also shows a decrease in its usage between 2000-2010. This drop in emissions clearly indicates a rise in bee stock along with an increase in the harvested area of the crops.

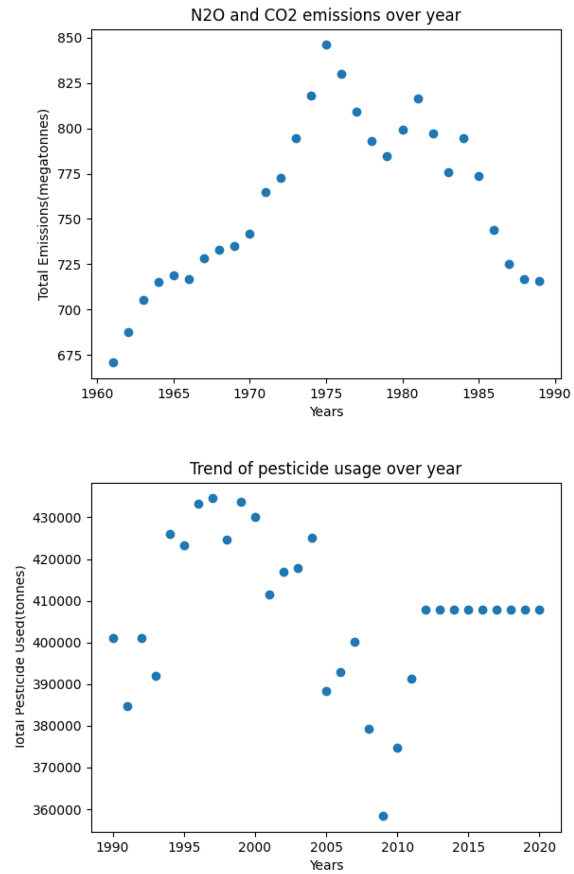

Fig. 2. Data Relations of Emissions and Pesticide usage over years

E. Model Building

The data analysis and other relationships between the data can clearly build a model around it, which can predict the crop area harvested over the future years and analyze the influence of how the bee population is related to the crop area. This is a classical research problem of supervised learning problem where we have the labeled data sets as features, and this problem can lead to a predicted output. There is no classification, as we are not trying to group any kind of data. The regression models are required to output prediction, which is a function of independent variables called features.

The relations are linearly organized by observing the data of how the predictor variable behaves over other features; hence, linear regression models are chosen. After splitting the entire data into training and testing sets, the accuracy or R2 for data can be split into two parts

- 1990-2020
  - Train Data Set - 75%
  - Test Data Set - 70%
- 1960-2020 with interpolation
  - Train Data set - 40 to 50%
  - Test Data set - 40 to 50%

The results of the following learning process are shown below for the 1990-2020 data set without any kind of interpolation.

- The average percentage error stands at 23%
- The coefficients correlation of this linear model shows that the bee population is the highest magnitude value and has more preference over the crop harvested area.
- The shape of the histogram of residuals is shown as a normally distributed gaussian curve, proving that the data set's values support the model to organize linearly.
- The Pearson correlation or the Pearson test values indicate that the p-value is 0.0002, which is less than 0.05, indicating that the values can be modeled linearly. A linear relationship exists between the independent features like bee population, emissions, and pesticide with the dependent feature of crop harvested area.
- The Kolmogorov-Smirnov test also has been performed on the data. As indicated above, the P-Value also stands around 0.000207, substantiating the linear relationship.

IV. Conclusion

The data that was required for this research is best when it was more divided along the state data or even lower division of county data. However, this was different here, and the research modeling went ahead with overall country data. This subdivision is ideal because of the 2nd part of the research, which is a geographical correlation. If data is available as a subdivision, finding the relation of the data set with the satellite imagery of a smaller region is simpler and less complicated than finding the overall agriculture green land reduction over the entire country. But if there is a subdivision, this is possible. With the help of the established model, the bee population can clearly influence the agricultural green harvested area increase. Accordingly, the number of the bee population can be increased with the colonies increase. It allows us to home these new bee colonies at the required locations for observing this result. Additionally, there is also a need for this subdivision for emissions of a region and pesticide usage in that region. Such grid-wise data accounts for this exact locational trend we hope to find.

V. Future Scope

As indicated previously, the first area of improvement is obtaining the ideal data subdivision. Secondly, as the data sets for specific regions are only available for a limited time frame, gathering this is essential for having a good accuracy model. If not possible, different nonlinear interpolation techniques must be incorporated to obtain a better accurate model. Lastly to correlate all this data to the satellite imagery data sets [12], [13].

References

[1] Khaled Ahmed; Ahmed. A. Ewees; Aboul Ella Hassanien, "Prediction and management system for forest fires based on hybrid flower pollination optimization algorithm and adaptive neuro-fuzzy inference system,"

[2] Yi Chen; Yun Li, "Intelligent Autonomous Pollination for Future Farming - A Micro Air Vehicle Conceptual Framework With Artificial Intelligence and Human-in-the-Loop"

[3] UN Environment report on why bees are essential to people and planet, "https://www.unep.org/news-and-stories/story/why-bees-are-essential-people-and-planet"

[4] Research by Liam Jackson on bees ability to forage decreases as air pollution increases"https://www.psu.edu/news/research/story/bees-ability-forage-decreases-air-pollution-increases/"

[5] Food and Agriculture Organization of the United Nations report on declining bee population pose a global threat "https://www.fao.org/news/story/en/item/1194910/icode/"

[6] Food and Agriculture Organization of the United Nations dataset "https://www.fao.org/faostat/en/data/QV/metadata"

[7] World Bank dataset "https://data.worldbank.org"

[8] Wikipedia page for list of crop plants pollinated by bees "https://en.wikipedia.org/wiki/List_of_crop_plants_pollinated_by_bees"

[9] Importance of pollinators in changing landscapes for world crops "https://royalsocietypublishing.org/doi/10.1098/rspb.2006.3721d3e691"

[10] United States Department of Agriculture - Economic Research Report No. (ERR-290) 34 pp "https://www.ers.usda.gov/publications/pub-details/?pubid=101475"

[11] United States Department of Agriculture - Economic Research Report No 290 "https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/ERS-290-09-12-2022.pdf"

[12] Food and Agriculture Organization of the United Nations report on use of geo-spatial data in agriculture statistics "https://www.fao.org/datalab/website/web/use-geo-spatial-data-agriculture-statistics"

[13] Areal lookup system "https://fdotewp1.dot.state.fl.us/AerialPhotoLookUpSystem/"