Hochschule Karlsruhe
University of
Applied Sciences

Institut für
Angewandte Forschung

Bee Activity Prediction and Pattern Recognition in Environmental Data

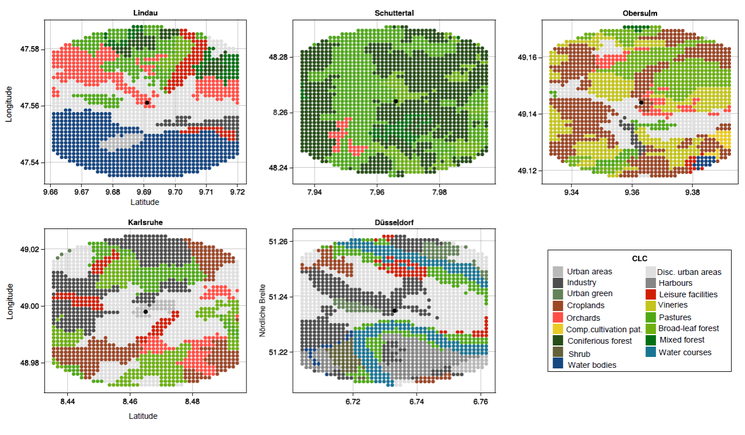

Figure 1: Land Cover Classes around the Beehives

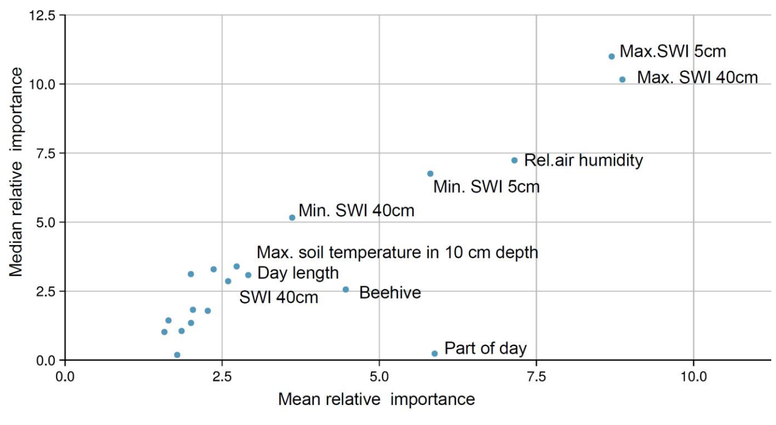

Figure 2: Relative Feature Importance

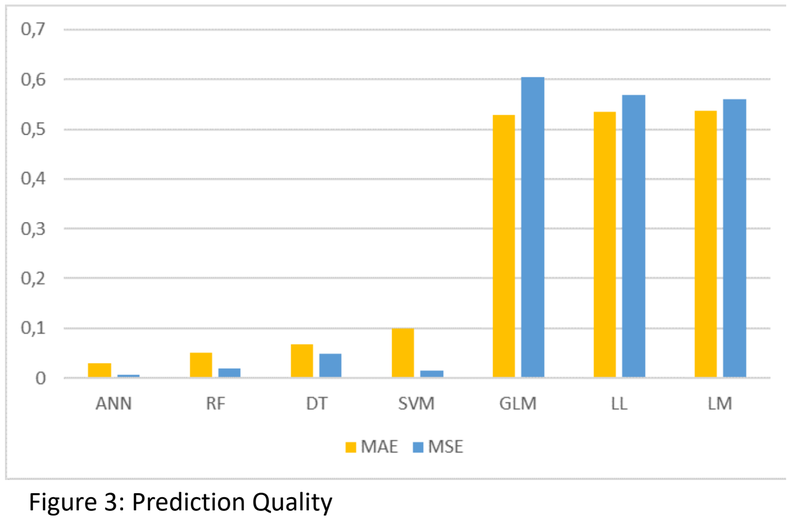

Figure 3: Prediction Quality

Because of climate change, biodiversity is declining rapidly. Many species like insects, especially bees, suffer from changes in temperature and rainfall patterns. Applying machine learning for monitoring and predicting specie's health and life conditions can help understanding and improving biodiversity. In this work we use data collected from cameras and sensors mounted upon beehives together with different other data sources like weather data, information extracted from satellite images and geographical information. We aim at predicting bees' health (measured as their activity) and analyzing influencing environmental conditions. We show that we can accurately predict bees' activity and understand their life conditions by using machine learning algorithms and explainable AI. Understanding these conditions can help to make recommendations on good locations for beehives. This work illustrates the potential of applying machine learning on sensor, satellite and weather data for monitoring and predicting species' health and hence shows the ability for adaptation to climate change and a more accurate species monitoring.

The **main contributions** of our work are:

- Multiple datasets were used, including information from satellite images, environmental sensor data, and data extracted from videos, to predict and understand bee activity.
- Shapley values and **a relative feature importance** score were computed to explain the effects of environmental conditions on bees.
- An **extensive evaluation of machine learning methods**, including both **explainable** and black box **models**, was conducted. Highly predictive models were learned, and interesting new patterns in environmental conditions that affect bees' activity were discovered.

Data

The data was collected during August and September 2021 from 10 beehives in 5 different locations in Germany (figure 1).

In our analysis we used different data sources :

- Weather data, including information on dew point, humidity, precipitation, air pressure, ground temperature, sunshine duration, air temperature and wind.
- Data on greenhouse gases at different altitudes (CO,CO2,N2O, CH4).
- Information on land cover classes (CLC), they represent spatial information on different types of physical coverage of the earth's surface, e.g., forests, grasslands, croplands, lakes, wetlands.
- Information on soil properties like soil temperature and humidity. We used the soil water index (SWI) which is determined by soil temperature and precipitation. SWI is extracted from satellite data.

We extracted 55 features as input to the machine learning models.

Methods & Results

The aim of the study is, to find important environmental factors that have an impact on bee activity as well as to predict the bee activity. To do so, we applied supervised machine learning (regression). First, we focusses on explainable regression algorithms like linear regression (LR), generalized linear model (GLM), lasso model with least angle regression (LL), decision trees (DT), and random forest (RF). Moreover, we applied black-box algorithms like support vector machines for regression (SVM), and feed forward artificial neural networks (ANN) which we assumed to be more accurate in predicting bee activity. Since these models are not interpretable, and different model types are hardly comparable in terms of their **explainable effects**, we computed **Shapley value**s, that represent the contribution of each feature towards a particular prediction. However, when comparing absolute shapely values extracted from different models, they differ strongly. Therefore, it is not possible to decide on the size of an effect over all models. For this reason, we calculated the **relative feature importance** of each feature (figure 2). Deep Learning (ANN) and Random Forest (RF) achieved the most accurate prediction results (figure 3). By using relative feature importance, we were able to detect interesting patterns in environmental data that are crucial for bees' health. To the best of our knowledge, the relevance of variables around soil water index at different depths for bee activity have not been analyzed so far. Understanding these conditions can aid in making recommendations for optimal beehive locations. Strong effects, both negative and positive, of environmental conditions such as weather, time, and land cover class related features, were found to impact bee activity. Therefore, it is crucial to consider the location of beehives to adapt to climate change and maintain their health.

Prof. Dr. Christine Preisach, University of Applied Sciences Karlsruhe, Germany
christine.preisach@h-ka.de
Marius Herrmann, Karlsruhe Institute of Technology, Germany