Contribution ID: **7** Contribution code: **Pres Session 2** Type: **Presentation**

Beehive Health Patterns Using Multi-modal Data Analysis and Unsupervised Machine Learning

Monday 13 October 2025 15:14 (11 minutes)

Agriculture Victoria Research (AVR) undertook a project for multimodal data analysis and anomaly detection for beehive health to address the critical challenge of determining hive health comprehensively in a pollination environment where diverse modalities of data are at interplay. Honeybee populations play an essential role in pollination within Victoria's horticulture industry, making the monitoring and maintenance of hive health crucial for agricultural productivity and pollination sustainability. However, existing methods for monitoring hive health often rely on commercial sensor companies that typically provide inferred hive health matrices using proprietary modelling and algorithms, often lacking transparency and scientific validation. At the same time, commercial monitoring systems heavily rely on IoT sensors alone for measurements and do not take into account other environmental variables at play that cannot be measured via these sensors. This highlights a gap and opportunity for assessing hive health using multimodal data streams and the application of scientifically validated data-driven methodologies for health pattern recognition.

The project collected multimodal data from—IoT sensors (e.g., temperature, humidity, acoustic activity), environmental DNA results, agrochemical records and manual inspections—to demystify interconnected insights into the health of a hive. The data was collected across three different almond farms in the Sunraysia region during 2024 almond pollination. The project aimed to develop a system capable of detecting abnormalities in hive behaviour and determining the overall health status of the hive by cross-checking with chemical sprays and human inspection to demonstrate functional relationships using a data-driven approach.

To address this, the project conducted a comprehensive data analysis on the high-dimensional dataset, including data pre-processing, feature engineering, visualisation and modelling. For modelling, data-driven unsupervised machine learning alongside classical statistical methods was used for anomaly detection. The models, which include Gaussian Mixture Models (GMM), Local Outlier Factor (LOF), and Isolation Forest (IF), were used to flag unusual hive behaviours over the pollination period.

An interactive data dashboard was then created to visualise the results of anomaly detection and cross-validate the identified anomalous behaviours in a particular hive/site with corresponding spray records and eDNA results. The anomaly detection system successfully identified periods of abnormal colony behaviour, which were cross-referenced with chemical records and validated through human inspection data.

These initial findings confirm the scientific base for future research and development in this new domain. For instance, incorporating eDNA data directly into model training would enable the model to capture additional ecological variables that may improve its pattern recognition capability. Similarly, establishing a standardised rubric for human inspections with high accuracy could serve as labels for supervised learning models, enabling the exploration of a broader range of machine-learning approaches. Scaling data collection across more sites and longer periods would further improve model robustness by capturing seasonal trends and environmental influences on hive health that could be made available for industry as a generalisable service.

Primary authors: Dr AGRAHARI BANIYA, Arbind (Agriculture Victoria Research, Department of Energy, Environment and Climate Action, Victorian State Government); Mr CHEN, Jiqiang (John) (University of Melbourne)

Co-authors: MURPHY, Kieran (Agriculture Victoria Research, Department of Energy, Environment and Climate Action, Victorian State Government); Dr THAYALAKUMARAN, Thabo (Agriculture Victoria Research, Department

of Energy, Environment and Climate Action, Victorian State Government)

Presenter: Dr AGRAHARI BANIYA, Arbind (Agriculture Victoria Research, Department of Energy, Environment and Climate Action, Victorian State Government)

Session Classification: Presentations Session 2: Data and Research & Data Science and Data Analysis

Track Classification: SciDataCon Persistent Themes: Data Science and Data Analysis