國立臺灣大學生物資源暨農學院生物機電工程學系

碩士論文

Department of Biomechatronics Engineering

College of Bioresources and Agriculture

National Taiwan University

Master's Thesis

基於多元感測與機器學習之智慧蜂箱健康狀態預測系統

Smart Beehive Health Prediction System Based on Multi-Sensor Data and Machine Learning

程柏勳

Po-Hsun Cheng

指導教授：林達德 博士

Advisor: Ta-Te Lin, Ph.D.

中華民國 114 年 7 月

July 2025

doi:10.6342/NTU202503964

誌謝

時間真的過得好快，去年還看三位學長在寫論文，今年就換我來感謝大家了。

在碩士論文即將完成之際，我衷心感謝這兩年給予我幫助的人，讓我能夠順利完成學業。首先感謝我的口試委員林達德老師、陳世芳老師及楊恩誠老師，感謝你們在論文口試時給予我許多建議，讓我完善我的研究。特別感謝我的指導老師林達德教授，感謝您邀請我加入超棒的 BBLAB，也感謝這兩年來不管是研究還是人生陷入迷茫時，您都為我一一解惑、指點明燈。

謝謝 405 的大家，首先感謝敬恆在最後口試前給予我許多建議，加值我的信心，再來就是易霖、晨宇、JOY 三位學長，有你們在我的碩一才可以過得那麼快樂自在，特別是易霖，我永遠記得我們在鳳凰谷遭遇虎頭蜂你也不怕，帶著膽小的我修蜂箱。還有碩一的各位跟已經出國的化達、凱鈞，謝謝你們在這段時間不吝給我許多幫助，祝福大家在未來一路順風！

接著是我的家人，感謝我的爸媽從小花時間金錢栽培我，並在我陷入低谷時給予我勇氣，還有阿公阿嬤從小照顧我長大，常常帶我到處玩。還有從小到大的死黨們，明明都在忙，半年約一次也能聊得很開心，讓我在研究之餘能聽你們分享趣事。

最後，我想感謝我的女朋友緣莘，謝謝你在我壓力很大的時候聽我說心事，給予我最大的支持，你是我前進的動力，未來也請多多指教。

要謝的人真的太多，沒辦法一一列舉，但我都會記得大家對我的好，謝謝你們！

最後我要說，蜜蜂計畫就交給你啦稚晴！

doi:10.6342/NTU202503964

摘要

本研究開發一套基於智慧蜂箱的蜂群健康監測系統，利用多感測器長期蒐集蜂箱內外部環境資料，結合時間序列預測與異常分類模型，實現蜂群重量、進出流量及花粉採集率的精確預測與健康狀態異常偵測。資料來源為 2024 年 9 月至 2025 年 4 月於台灣雲林兩處場域蒐集的感測數據與蜂群檢查表，蜂農每週填寫健康評估表單以標註蜂群狀態，作為機器學習模型的訓練與驗證依據。模型部分採用 GRU 與 SARIMAX 進行單變量與多變量時間序列預測。結果顯示多變量模型在三項目標變數的預測精度均優於單變量模型，其中 GRU 模型在跨蜂箱驗證下的誤差表現最佳，蜂群重量的平均絕對百分比誤差（MAPE）為 0.6%～2.1%，進出流量的 MAPE 為 19%～24.6%，花粉採集率的對稱平均絕對百分比誤差（sMAPE）為 16.7%～31.3%。透過 SHAP 分析進一步辨識影響預測結果的關鍵特徵，模型未出現預測停滯或漂移現象，具備長期穩定性。GRU 二元分類模型在蜂群健康狀態預測的整體準確率達 94%，能有效偵測失王、蜂巢壅擠等關鍵異常事件。系統能夠提供穩定且可靠的數據化蜂群動態洞察，支援早期健康問題識別與精準養蜂管理，提升蜂群存活率與管理效率。

關鍵詞：機器學習、時間序列、蜂箱健康預測、智慧農業

doi:10.6342/NTU202503964

ABSTRACT

This study presents a smart beehive-based health monitoring system that combines multi-sensor data collection, time-series forecasting, and anomaly detection to monitor honeybee colony health. Data were collected from September 2024 to April 2025 at two apiaries in Yunlin County, Taiwan. Sensor measurements and beekeeper-completed evaluation forms were used to label colony health status for training and validating machine learning models. Two forecasting models, GRU and SARIMAX, were developed to predict hive weight, bee traffic, and pollen collection rate under both univariate and multivariate settings. Experimental results showed that multivariate models consistently outperformed univariate ones, with the multivariate GRU achieving the most stable and generalizable performance. Specifically, during cross-hive validation, the forecasting errors were: a Mean Absolute Percentage Error (MAPE) of 0.6%–2.1% for weight prediction, 19%–24.6% MAPE for bee ingress–egress traffic, and a Symmetric Mean Absolute Percentage Error (sMAPE) of 16.7%–31.3% for pollen collection rate. SHAP-based interpretability analysis further identified the most influential features contributing to model accuracy, and the model exhibited no signs of prediction stagnation or shifting. Additionally, a GRU-based binary classifier for colony health status achieved an overall accuracy of 94%, effectively detecting critical anomalies such as queen loss and hive overcrowding. Overall, the proposed system reliably provides data-driven

doi:10.6342/NTU202503964

insights into colony dynamics, enabling early detection of health issues and supporting precision apiculture practices for improved hive management.

Keywords. machine learning, time series, beehive health prediction, smart agriculture

doi:10.6342/NTU202503964

Tables of Contents

doi:10.6342/NTU202503964

doi:10.6342/NTU202503964

doi:10.6342/NTU202503964

doi:10.6342/NTU202503964

doi:10.6342/NTU202503964

List of Figures

doi:10.6342/NTU202503964

doi:10.6342/NTU202503964

List of Tables

doi:10.6342/NTU202503964

CHAPTER 1

Introduction

Honey bees are indispensable pollinators in global ecosystems, underpinning both agricultural productivity and ecological balance. Roughly one-third of the world's crops—including fruits, vegetables, and nuts, depend on bee pollination and therefore carry considerable economic value. In recent years, however, colonies have suffered increasingly frequent collapses associated with Colony Collapse Disorder (CCD) (Henry et al., 2012), posing serious threats to food production and ecosystem stability. In addition to CCD, beekeepers are increasingly challenged by chronic diseases and stressors—such as mite infestations, viral pathogens, and pesticide exposure—that gradually weaken colonies over time, often leading to irreversible hive collapse. International programs such as the German Bee Monitoring Project and COLOSS (Gray et al., 2019) stress that effective beekeeper management is pivotal in mitigating CCD. Yet conventional hive-health management relies on frequent manual inspections that are time-consuming, labor-intensive, and incapable of providing real-time insight into hive conditions. This limitation is particularly critical for detecting slow-developing health deterioration, where early signs often remain unnoticed until severe losses occur. With the rise of the Internet of Things (IoT), smart agriculture technologies have emerged as a promising remedy. IoT

doi:10.6342/NTU202503964

platforms enable beekeepers to remotely monitor in-hive temperature, humidity, weight, weather conditions, and bee traffic (Dsouza & Hegde, 2023; Ho et al., 2022; T. N. Ngo et al., 2021), thereby overcoming the limitations of conventional approaches and revealing key behavioral patterns. For instance, Meikle et al. (2016) a negative correlation was reported between in-hive temperature amplitude and colony size during harvest, along with links between three-day weight amplitude, yield, and adult mortality. Similarly, Ziegler et al. (2022)found that bee activity peaks when ambient temperature exceeds 25 °C and relative humidity ranges from 60% to 70%.

Despite the rapid advancements in smart beehive technology, achieving fully automated and reliable deployment in real-world settings remains a formidable challenge. Relying solely on individual monitoring parameters often falls short of providing sufficient insight for timely decision-making. Fortunately, recent progress in machine-learning techniques has enabled new solutions that overcome many limitations of traditional rule-based or manually supervised approaches. To address these challenges, machine learning has emerged as a powerful analytical tool in hive-monitoring research, providing enhanced capabilities for identification, classification, and time-series prediction across diverse sensor modalities. For identification, Ngo et al. (2019) applied a convolutional neural network (CNN) to detect bee traffic in images. In classification, Nolasco et al. (2019) used a support-vector machine (SVM) to recognize queen-loss

doi:10.6342/NTU202503964

sound signals, while (Braga et al., 2020) achieved 0.98 accuracy in assessing hive health with a random-forest model. For prediction, Khairul Anuar et al. (2024) employed RNNs and their variants (LSTM, GRU) for multivariate forecasting of hive weight, whereas Kulyukin et al. (2024) showed that ARIMA remains competitive for univariate forecasts of weight, in-hive temperature, and bee traffic while demanding fewer computational resources. However, most existing studies remain limited to single-sensor modalities or isolated hive health indicators. Additionally, while significant efforts have been made to detect acute anomalies such as sudden colony collapse, existing studies rarely focus on developing early-warning mechanisms for chronic health deterioration. These long-term issues—caused by persistent stressors such as mites, pathogens, and pesticides—often progress silently and remain undetected until severe colony losses occur. A fully integrated, predictive framework that could one day incorporate chronic health indicators into model training is still lacking, leaving beekeepers with limited tools for long-term preventive management. Moreover, few studies have addressed the progressive nature of chronic diseases or developed predictive early-warning mechanisms to mitigate long-term colony decline. A fully integrated, field-deployable system that fuses heterogeneous sensor data and concurrently forecasts multiple key metrics - such as weight, bee activity, and pollen collection - is still lacking. This research gap highlights the critical need for a

doi:10.6342/NTU202503964

holistic real-time monitoring and early-warning framework that is both scalable and practical for commercial apiculture applications.

Building on a smart-hive platform previously developed by our team (Liu et al., 2024), this study proposes an integrated system for hive health monitoring and colony-decline early warning. Six hives located across two geographically distinct apiaries were equipped with multi-sensor suites that continuously recorded temperature, humidity, weight, bee traffic, pollen collection, audio signals, and local weather conditions. In the first task, the heterogeneous sensor data were fed into both Gated Recurrent Unit (GRU) and Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) models to forecast three critical health indicators: hive weight, bee traffic, and pollen collection rate. Comparative analyses were conducted to evaluate model performance. In the second task, the same sensor data were combined with beekeeper-provided Healthy Colony Checklist (HCC) forms, which served as labeled annotations to train a GRU-based classifier capable of predicting overall hive health status and issuing early warnings of colony decline. By fusing real-time sensing, expert annotations, and advanced time-series modeling, this research delivers a scalable and practical decision-support tool for modern apiculture, reducing labor demands, improving management quality, and mitigating risks associated with CCD. Unlike conventional monitoring approaches, this system not only detects acute anomalies such as queen loss and

doi:10.6342/NTU202503964

overcrowding but also proactively predicts slow health deterioration caused by chronic diseases, enabling beekeepers to intervene earlier and reduce the risk of long-term colony losses. The following are the specific goals in detail:

1. Establish a smart hive monitoring system that collects multimodal data including hive weight, temperature, humidity, acoustic signals, and bee traffic, and to collaborate with beekeepers in completing Healthy Colony Checklist (HCC) forms to obtain ground-truth health labels.
2. Apply GRU and SARIMAX models for multi-step forecasting of hive weight, bee traffic, and pollen collection rate, and to compare their predictive performance.
3. Integrate sensor data with HCC labels to train a GRU-based health classification model capable of identifying hive health status and providing early warnings of potential colony decline.

doi:10.6342/NTU202503964

CHAPTER 2

Literature Review

2.1 Colony Ecology

2.1.1 Honeybee Social Structure and Worker Division of Labor

As highly social organisms, honey bees rely entirely on cooperation and division of labor within the colony for survival. Their social structure is remarkably organized, with each individual assigned specific roles from birth. This system of task allocation is fundamental to the efficient functioning of the hive. Before conducting in-depth research on honey bees, it is essential to develop a solid understanding of their social organization.

According to Goulson (2002) on eusocial insects, the honey bee society is divided into three main castes: the queen, the workers, and the drones. Each caste has a distinct role and life cycle. The queen is responsible for reproduction and typically lives between two to five years. However, her reproductive capacity declines with age. Worker bees, which constitute the majority of the colony, perform most of the essential tasks such as cleaning cells, foraging, and nursing larvae. Their lifespan is relatively short, lasting only a few weeks. Drones, whose primary function is to mate with the queen, generally die

doi:10.6342/NTU202503964

after mating or are expelled from the hive during the winter. Their lifespan is only a few months (Oldroyd & Thompson, 2006).

Caste differentiation in honey bees is determined at the time of egg laying. Fertilized eggs usually develop into worker bees, while unfertilized eggs give rise to drones. When a new queen is needed, worker bees construct a specialized structure called a queen cell and continuously feed the developing larva with royal jelly, enabling it to develop into a queen.

The division of labor among worker bees is age-dependent. Newly emerged workers are initially assigned tasks such as cleaning brood cells and maintaining hive temperature. As they age, their roles diversify. Around days six to twelve, their mandibular glands become fully developed, and they begin to secrete royal jelly to feed the larvae and queen. From days twelve to eighteen, they start producing wax and constructing combs, while also conducting orientation flights near the hive. These bees are referred to as nurse or house bees and are primarily involved in in-hive duties. After eighteen days, workers become foragers, responsible for collecting nectar, pollen, and water from the external environment. This age-based task allocation ensures the efficient operation of the colony, as each bee performs specific duties aligned with its physiological development. Such a system optimizes labor distribution both inside and outside the hive, enabling the colony

doi:10.6342/NTU202503964

to adapt to external challenges and maintain long-term reproductive success (Johnson, 2008).

2.1.2 Microclimate and Temperature-Humidity Variations

Honey bees exhibit a high sensitivity to internal hive temperature and humidity, as maintaining a stable microclimate is essential for brood development and overall colony health. To achieve this stability, bees constantly regulate the thermal and hygrometric conditions within the hive. In healthy and thriving colonies, the core brood area is typically maintained at a temperature between 33°C and 36°C, a range that is critical for the development of eggs and the physiological health of adult bees (Bujok et al., 2002; Flores et al., 2018).

Humidity plays an especially vital role during the incubation of eggs. Studies have shown that the ideal relative humidity for egg hatching lies between 90 and 95 percent. When humidity falls below this optimal range, hatchability declines significantly. For instance, a relative humidity of 75 percent results in a hatch rate of only 60 percent, and when humidity drops below 30 percent, successful hatching becomes nearly impossible (Al-Ghamdi et al., 2014). In addition, even slight temperature variations within the 32°C to 36°C range can significantly affect the development and behavioral expression of worker bees. Suboptimal temperatures have been linked to wing deformities, which

doi:10.6342/NTU202503964

subsequently impair the flight capability of adult workers (Himmer, 2008; Tan Ken et al., 2005). In contrast, queen bee development is less sensitive to temperature fluctuations, primarily due to the protective insulation provided by the thick walls of the queen cell. Research has shown that colonies tend to construct queen cells closer to the hive periphery during periods of high ambient temperature to prevent overheating, whereas in colder conditions, queen cells are placed near the center of the hive to maintain warmth (Degrandi-Hoffman et al., 1993). This spatial adaptation effectively safeguards queen rearing under varying environmental conditions.

Given the critical role of thermal and hygrometric stability, honey bees have evolved multiple behavioral mechanisms to regulate their internal environment. The hive's spatial organization contributes significantly to this function, with the central region reserved for brood rearing and the outer zones used for storing pollen and nectar. This layout helps buffer temperature fluctuations in the brood area. When internal temperatures drop, nurse bees increase their clustering density in the brood region and generate heat through thoracic muscle vibrations. Conversely, during periods of elevated temperature or humidity, bees enhance airflow by fanning their wings and actively collect water to promote evaporative cooling, thereby reducing both temperature and humidity levels inside the hive (Nicolson, 2009).

doi:10.6342/NTU202503964

2.1.3 Bee Physiology and Its Significance

The wing-buzzing sound produced by honey bees, commonly referred to as buzzing or hive acoustics, serves not only as an indicator of individual physiological and behavioral states but also as a crucial mechanism of intra-colony communication. Different behavioral contexts are often associated with distinct frequency patterns, which play an important role in colony coordination. Research has shown that worker bees, drones, and queen bees each produce characteristic buzzing frequencies that reflect their functional roles and internal conditions (Hord & Shook, 2013).

Among these acoustic signals, the queen bee's buzzing is particularly noteworthy. During her early life stage, the queen emits a unique vocalization that is acoustically distinct from the wing sounds of workers and drones. This sound, described as trumpet-like in quality, consists of both short and long elements and typically begins even before the queen emerges from her queen cell. The acoustic signal can propagate throughout the hive, enabling colony members to accurately identify the queen's location. There are two primary types of queen vocalizations: "quacking," which is emitted by virgin queens still inside their cells and consists of short pulses at a frequency of approximately 300 Hz, and "tooting," which is produced by emerged queens and begins with two long tones followed by several short bursts. Studies have also revealed that queen buzzing frequency increases

doi:10.6342/NTU202503964

with age. Younger queens tend to produce lower-frequency sounds, while queens around 20 days old may reach frequencies as high as 500 Hz (Michelsen et al., 1986).

Queen buzzing plays a critical role in intraspecific competition and reproductive dynamics within the colony. Virgin queens often respond to the calls of rival queens, resulting in a phenomenon known as duetting. These acoustic interactions allow queens to locate potential competitors and prepare for eventual physical confrontation. Such competitive vocal exchanges are vital for colony regulation, as only one dominant queen is typically retained. Furthermore, previous studies have shown that worker bees exhibit a freezing response upon hearing queen buzzing, ceasing their activity temporarily. This behavior promotes order and coordination within the hive (Fletcher, 1978).

Buzzing sounds are not exclusive to queens and are equally significant in other contexts. For instance, worker bees generally produce wing vibrations at around 250 Hz, while drones generate lower-frequency sounds near 200 Hz. When the colony perceives an external threat, workers emit alarm buzzes with frequencies as high as 6000 Hz to alert the colony and signal imminent defensive action (Papachristoforou et al., 2008). This diversity in acoustic signals demonstrates the behavioral complexity of honey bees in responding to both internal and external stimuli.

doi:10.6342/NTU202503964

Table 2-1. Types of Honeybee Buzzing and Corresponding Frequencies

Buzzing Type	Frequency (Hz)
Worker Bee Buzzing	250
Drone Buzzing	200
Alarm Buzzing	6000
Queen Buzzing	340-500

2.1.4 Swarming and Absconding Behavior

Large-scale departure of honeybee colonies typically occurs under two main circumstances: swarming and absconding. Swarming is a natural reproductive process of the colony and generally occurs in spring or autumn when environmental conditions are favorable and colony resources are abundant. During such periods, the hive initiates the construction of queen cells in preparation for the emergence of a new generation of queens (Winston, 1991). As a new queen is about to emerge, the old queen departs from the hive accompanied by a portion of the worker bees. This group leaves to search for a new nesting site, thereby enabling colony expansion. Swarming usually takes place in the warm afternoons, as recorded in the observations of Ferrari et al. (2008).

In the early stage of swarming, the departing bees typically settle temporarily at a site near the original hive for one or two days. During this time, scout bees are dispatched to explore suitable locations for establishing a new nest. Once a favorable site is identified,

doi:10.6342/NTU202503964

the swarm collectively relocates and begins constructing a new hive, completing the colony division process (Hepburn & Radloff, 2013). Within the original hive, multiple queen cells are usually constructed. When the first new queen emerges and establishes dominance, she emits vocal signals to determine whether other rival queens remain unhatched. The new queen may destroy remaining queen cells or engage in physical competition with simultaneously emerged rivals. This competitive interaction sometimes leads to a secondary swarm event (Michelsen et al., 1986).

In contrast, absconding is triggered by external environmental stress rather than internal reproductive needs. Stressors may include prolonged high temperatures, water scarcity, or persistent external threats such as predators or hive disturbances (Schneider, 1990). When the colony perceives the current environment as unsafe or unsustainable, the bees abandon the hive entirely in search of a more suitable habitat. Unlike swarming, which is an intentional and reproductive behavior, absconding is a survival response to adverse conditions. This behavior can cause substantial economic loss to beekeepers, as the productivity of the hive diminishes dramatically once absconding occurs. Therefore, early detection and mitigation of environmental stressors are critical to preventing such events (Hepburn & Radloff, 2013).

doi:10.6342/NTU202503964

2.1.5 Impact of Agricultural Diseases and Pests on Bees

In addition to lethal pesticides, various other environmental factors can negatively impact hive health (Henry et al., 2012). Although these stressors may not result in the immediate death of bees, they can impair the foraging abilities and memory functions of worker bees, ultimately leading to colony decline (Barron, 2015). For instance, neonicotinoids and organophosphate-based miticides have been shown to disrupt neural signal transmission in the bee brain, thereby reducing cognitive performance and memory retention (Palmer et al., 2013). The accumulation of such sublethal effects can severely compromise colony resilience when exposed to multiple stressors (Brown et al., 2000).

Pathogens and parasites also represent significant threats to hive stability, with the parasitic mite Varroa destructor being one of the most prevalent concerns in conventional apiculture. Varroa mites attach to both adult bees and brood, feeding on their hemolymph and thereby weakening the immune system and shortening the lifespan of their hosts (Gregorc & Sampson, 2019). Moreover, Varroa mites serve as vectors for various viral pathogens, which can proliferate rapidly within already weakened colonies. Infected colonies often experience accelerated deterioration due to the compounding effects of mite infestation and viral transmission (Abbo et al., 2017). Without timely control, Varroa infestations have the potential to destroy an entire colony within a single year. Therefore,

doi:10.6342/NTU202503964

regular monitoring of mite populations and prompt implementation of management strategies are essential for sustaining hive health (Gregorc & Sampson, 2019).

2.2 Beehive Monitoring

2.2.1 Analysis of Weight and Temperature-Humidity Data

Monitoring hive weight provides an intuitive indicator of resource storage and colony activity within the hive. However, weight variations are influenced by multiple factors, including daily foraging gains, food consumption, moisture evaporation rates, adult bee population, and mortality (Cardell-Oliver et al., 2022; McLellan, 1977; Meikle et al., 2008). One common analytical method is detrended weight curve analysis, where the first weight measurement of each day is subtracted from subsequent values to reduce the impact of external variation. This normalization facilitates comparison across different days (Meikle & Weiss, 2017). By applying segmented regression modeling to these weight curves, researchers can characterize in-hive activity patterns, such as weight losses during forager departure and gains upon return. These fluctuations can serve as valuable indicators of colony vitality (Meikle et al., 2018).

Although weight data offer important insights, they are not sufficient on their own

doi:10.6342/NTU202503964

to assess hive health. Therefore, temperature and humidity are often used as complementary parameters. The optimal brood-rearing temperature inside the hive ranges from 33 to 36°C, while the ideal humidity lies between 90 and 95 percent (Abou-Shaara et al., 2017; Al-Ghamdi et al., 2014; Li et al., 2022). A consistent internal temperature above 30°C with daily fluctuations of less than 1.5°C typically indicates active brood rearing (Becher et al., 2009). Sensor placement is also crucial to measurement accuracy. Meikle et al. (2016) demonstrated that positioning temperature and humidity sensors directly above the center of the hive frame yields the most reliable data without disrupting hive activity.

Temperature and humidity data can also be used to detect swarming events. During the swarming process, internal hive temperature may rise by 1 to 4°C. This change is often accompanied by noticeable fluctuations in both temperature and humidity, particularly due to the wing-fanning activity of bees prior to departure (Ferrari et al., 2008). Additionally, long-term trends in temperature and weight can reveal important colony dynamics. For example, Meikle et al. (2016) found that brood temperature amplitude during the harvest period was negatively correlated with colony population size, and that three-day weight fluctuations were associated with both yield levels and adult mortality rates.

doi:10.6342/NTU202503964

2.2.2 Analysis of In-Hive Acoustic Data

In-hive audio recordings represent the aggregate of all buzzing sounds generated within the colony. Many studies have sought to extract informative features from these acoustic signals to infer the current state of the hive (Abdollahi et al., 2022). In recent years, research on beehive audio signal analysis has become increasingly sophisticated and has been widely applied to honeybee health monitoring, particularly in the identification of colony behaviors and responses to environmental stressors. Acoustic monitoring technologies have proven effective for detecting various hive-related scenarios, such as the presence of the queen, the prediction of swarming events, and the intrusion of foreign species. These technologies not only assist beekeepers in managing colonies more effectively but also reduce the need for frequent on-site inspections and enable remote surveillance.

Regarding model selection, traditional machine learning algorithms such as Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) are frequently employed. For instance, Nolasco and Benetos (2018) demonstrated the application of SVM and CNN in classifying hive acoustic data. They utilized feature extraction techniques such as Mel Frequency Cepstral Coefficients (MFCC) and Mel spectrograms to detect and distinguish sounds originating from inside and outside the hive. While CNN

doi:10.6342/NTU202503964

models tend to achieve higher accuracy in site-specific scenarios, SVM models exhibit superior generalization performance, particularly when applied to previously unseen hive data, making them suitable for diverse applications.

In terms of feature extraction, MFCC has been extensively used in hive acoustic analysis. Robles-Guerrero et al. (2019) applied MFCC to extract audio features from hive recordings and used Lasso Logistic Regression for feature selection. Their results showed that this approach could effectively identify queenless states. Additionally, singular value decomposition (SVD) clustering analysis revealed that hives in similar physiological states tend to exhibit comparable acoustic patterns, thereby improving the accuracy of hive health assessment.

Zgank (2019) explored an IoT-based monitoring system for colony activity that employed acoustic signals to differentiate between normal activity and swarming conditions. This system extracted MFCC and Linear Predictive Coding (LPC) features and applied Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM) for classification. The resulting performance demonstrated the potential of acoustic monitoring systems in modern beekeeping management.

Moreover, hive acoustic data can be combined with environmental sensor inputs to develop multimodal data fusion models. For example, Imoize et al. (2020) integrated in-

doi:10.6342/NTU202503964

hive temperature and humidity readings with audio features to achieve more accurate behavioral classification. These multimodal approaches significantly enhance the reliability and practicality of real-time colony monitoring.

2.2.3 Analysis of Bee and Pollen Traffic

In recent years, automated monitoring based on image processing has emerged as a mainstream approach in the analysis of hive activity. According to Magnier et al. (2018), monitoring bee traffic at the hive entrance provides critical indicators of colony strength, health status, and structural organization. Their study introduced a computer vision-based system that employed a combination of image transformation and background subtraction techniques to accurately count the number of bees entering and exiting the hive. This method is particularly effective for assessing the impact of environmental stressors such as pesticide exposure, and for tracking the foraging behavior of bees on a daily basis.

In preliminary stages of the present study, a multi-sensor smart beehive monitoring system was developed that incorporates image processing and deep learning-based detection models to monitor bee entry and exit activity. The system quantifies the number of bees entering and leaving the hive each day, as well as the number carrying pollen. The detection achieved a precision of 0.91 and a recall of 0.99, enabling accurate estimation

doi:10.6342/NTU202503964

of daily total traffic and pollen intake. The processed data are exported as several key indicators, including daily pollen traffic counts, total daily pollen weight, pollen collection ratio (total pollen intake divided by total bee entries), and cumulative pollen yield (T. N. Ngo et al., 2021). These metrics provide an intuitive understanding of colony activity levels and serve as valuable indicators for assessing both colony health and productivity.

2.2.4 Effects of Climatic Factors on Bee Colonies

The interaction between the beehive and its surrounding environment significantly affects colony health. Meteorological factors not only influence nectar availability but also affect bee behavior, in-hive microclimate regulation, and the willingness of forager bees to leave the hive. Among these factors, ambient temperature and relative humidity are considered the most influential. When the external temperature exceeds 25°C and relative humidity ranges between 60 and 70 percent, bees exhibit peak activity levels (Ziegler et al., 2022). In contrast, bee activity declines markedly when the relative humidity rises above 80 percent or during rainy weather. This sensitivity to environmental conditions reflects the highly adaptive nature of honey bees as social insects. Their ability to regulate internal conditions is also evident in their microclimate management strategies. Despite continuous changes in external climate, bees are capable of maintaining in-hive

doi:10.6342/NTU202503964

temperatures within a range of 18 to 34°C and relative humidity between 70 and 95 percent by clustering for heat retention or fanning for cooling (Gil-Lebrero et al., 2020). Maintaining a stable internal environment is critical for colony survival and reproduction, particularly in arid or cold environments. These regulatory mechanisms not only help preserve colony health but also enhance productivity, underscoring the important role meteorological conditions play in shaping hive microclimate.

In addition to temperature and humidity, light intensity has also been identified as a significant factor influencing bee foraging behavior. According to Burrill and Dietz (1981), foraging activity shows a linear positive correlation with temperature, while the relationship with light intensity is more complex. They found that although increased light intensity initially promotes bee activity, excessively high levels may actually suppress it. This suggests that the effect of light on bee behavior is not unidirectional. Their study further highlights the interactive influence of temperature and light intensity and suggests that these variables can be used to accurately predict forager bee departure rates.

The profound influence of light on bee behavior is further evident during rare natural phenomena. For example, during solar eclipses, the sudden reduction in light intensity leads to a marked decline in bee activity. This demonstrates that illumination is not only important for stimulating bee foraging but also plays a crucial role in maintaining normal

doi:10.6342/NTU202503964

behavioral rhythms within the colony (Gil-Lebrero et al., 2020). Clarke and Robert (2018) conducted an integrative study incorporating multiple meteorological variables, including illumination, temperature, humidity, wind speed, wind direction, and atmospheric pressure, to model forager behavior. Their results showed that temperature and light intensity were the most dominant predictors of foraging activity and together accounted for up to 78 percent of the observed variance in bee departure rates.

2.3 Studies on Bee Health Status Analysis

By integrating the various sensor-derived features discussed above, it is possible to assess the current health status of a honeybee colony. Among these features, local weather conditions and in-hive temperature and humidity are the most commonly used indicators. Weather patterns can affect bee health through changes in temperature and precipitation. For example, extreme heat can reduce foraging time, while excessive rainfall may dilute nectar concentrations, decreasing the availability of floral resources and increasing the risk of nutritional stress (Mishra et al., 2023; Olate-Olave et al., 2021). In contrast, in-hive temperature and humidity are actively regulated by the bees themselves, making their average values and fluctuation ranges meaningful indicators of colony status (Meikle & Weiss, 2017). Cook et al. (2022) emphasized that maintaining a stable

doi:10.6342/NTU202503964

microclimate is critical for brood development and colony survival, with bees typically regulating hive temperature between 32 and 36°C to ensure proper larval growth.

Edwards-Murphy et al. (2016) incorporated both internal and external environmental parameters—such as hive temperature, humidity, carbon dioxide, and oxygen levels, along with external temperature, rainfall, and light intensity—to identify ten threshold conditions for detecting abnormal hive states. These included large temperature variations greater than 20°C, minimal variation below 5°C, and relative humidity exceeding 95 percent. These conditions provide early warning signs that can assist beekeepers in implementing timely interventions to safeguard colony health.

A review by Hadjur et al. (2022) summarized the current state of automated hive health monitoring. Most existing smart hive systems integrate multiple types of sensors, including visual, acoustic, temperature-humidity, and weight sensors. In a meta-analysis of 40 studies, visual and acoustic monitoring were each employed 13 times, temperature and humidity monitoring appeared in 10 cases, and weight sensing was used in 4 studies. Each sensor modality offers specific advantages: image processing is useful for observing activity levels and external behaviors, acoustic sensing is often applied in queen loss and swarming detection, while hive weight and internal environmental data are used to assess nectar availability and brood conditions.

doi:10.6342/NTU202503964

2.4 Machine Learning Approaches for Beehive Monitoring

2.4.1 Selection and Application of Machine Learning Algorithms

In recent years, various machine learning models have been increasingly applied to extract meaningful information from smart beehive monitoring data. Compared to manual inspection and classification, machine learning offers a faster and more efficient means of analyzing data and making predictions. These models are capable of performing tasks such as identification, classification, and prediction. Based on methodological differences, machine learning applications in beekeeping research can be categorized into three main types: classification models, regression models, and deep learning approaches.

Among deep learning methods, convolutional neural networks (CNNs) have been widely used for image-based analysis tasks. For example, CNNs have been applied to detect the entry and exit of bees at the hive entrance using visual data (T.-N. Ngo et al., 2021). For classification tasks, Nolasco et al. (2019) utilized support vector machines (SVMs) to classify acoustic signals associated with queen loss. In addition, by incorporating beekeeper inspection data and hive health scoring systems such as the Hive Condition Checklist (HCC), hive health status can be formulated as a classification problem. This enables the use of algorithms such as random forest and k-nearest neighbors (KNN) to predict hive health conditions. Braga et al. (2020) demonstrated that

doi:10.6342/NTU202503964

random forest classifiers achieved a high accuracy of 0.98 in predicting hive health categories, validating the feasibility of machine learning for health status analysis in apiculture.

In scenarios involving multi-feature datasets, traditional classification models such as decision trees, random forests, and KNN have also demonstrated strong performance (Braga et al., 2020; Edwards-Murphy et al., 2016). To further enhance classification accuracy, dimensionality reduction and feature extraction techniques such as principal component analysis (PCA) and singular value decomposition (SVD) can be employed. These methods help reduce noise and identify representative features, thereby improving the performance and interpretability of health classification models (Robles-Guerrero et al., 2019).

2.4.2 Colony State Prediction Using Time Series Models

In the context of time series-based hive condition prediction, deep learning models have emerged as powerful tools for monitoring colony health. The study by Andrijević et al. (2022) demonstrated that integrating Internet of Things (IoT) technologies with time series forecasting models can accurately estimate bee traffic patterns and simulate behavioral changes under varying environmental conditions. Their approach combined autoregressive integrated moving average (ARIMA) models with recurrent neural

doi:10.6342/NTU202503964

networks such as LSTM, yielding strong performance in ecological monitoring of beehive systems.

Expanding on this line of research, Truong et al. (2023) introduced a hybrid deep learning architecture that integrates convolutional neural networks (CNN) with gated recurrent units (GRU) to analyze acoustic buzzing signals for colony health assessment. Given the temporal dependencies inherent in audio data, the use of GRU allowed for more effective extraction of time-related features, achieving a high classification accuracy of 0.983. This work highlights the growing importance of audio-based sensing as a non-invasive and highly informative modality in apicultural monitoring.

Complementing this, Ruvinga et al. (2023) investigated the application of acoustic signals for queen detection within hives. By combining CNN and long short-term memory (LSTM) networks and utilizing Mel-frequency cepstral coefficients (MFCC) and short-time Fourier transform (STFT) features, their model was able to accurately distinguish between queenright and queenless conditions. This approach provides practical value by enabling timely intervention when colony reproductive integrity is at risk.

doi:10.6342/NTU202503964

To evaluate the generalizability of time series forecasting approaches across multiple sensing modalities, Kulyukin et al. (2024) conducted a five-month longitudinal study using non-invasive data sources including hive weight, internal temperature, and entrance traffic. They compared the performance of artificial neural networks (ANN), CNN, LSTM, and ARIMA models across multiple prediction horizons ranging from 6 to 24 hours. Their results showed that both ARIMA and deep learning models produced comparable forecasting accuracy, validating the effectiveness of hybrid and classical methods alike. Notably, the study also contributed a FAIR-compliant, high-quality dataset, and emphasized the continued practicality of ARIMA in environments with limited computational resources.

Beyond model architecture, the selection of appropriate input features plays a crucial role in multivariate time series forecasting. In this regard, Wu (2023) examined the influence of climatic variables on bee foraging behavior using GRU and LSTM models. To address multicollinearity, he employed variance inflation factor (VIF) analysis to eliminate redundant yet highly correlated features. This improved the interpretability and predictive accuracy of the models, with R^2 values ranging from 0.84 to 0.91 across seasonal datasets. These results underscore the importance of combining robust modeling with principled feature selection to enhance performance in real-world precision apiculture applications.

doi:10.6342/NTU202503964

Taken together, these studies demonstrate the growing potential of time series models for intelligent beehive monitoring. By integrating multi-sensor data—such as weight, temperature, humidity, acoustic signals, and environmental conditions—into temporal forecasting frameworks, it becomes possible to construct predictive models that reflect the dynamic health status of the colony. In this study, we aim to leverage such models to reduce manual labor demands in beekeeping while enabling real-time anomaly detection and early warning capabilities. This approach not only enhances the precision of hive management but also contributes to the development of scalable and sustainable solutions for modern apiculture.

doi:10.6342/NTU202503964

CHAPTER 3

Methodology

3.1 Smart Multi-Sensor Beehive System

This study developed an integrated smart beehive system that combines sensing, cloud computing, and predictive analytics. Multiple sensors were installed to measure hive weight, temperature, humidity, acoustic signals, bee traffic, and pollen load. Data were transmitted via Raspberry Pi and Jetson TX2 to the AWS cloud platform and visualized in real-time and historically through Amazon QuickSight. During preprocessing, the system performed data imputation, temporal alignment, feature selection, and normalization, with additional log transformation and threshold filtering applied to high-variance variables such as pollen collection rate and bee activity. Hive health status was also annotated using the Hive Colony Checklist (HCC) forms. For forecasting, both traditional SARIMAX and deep learning-based GRU models were employed for univariate and multivariate predictions. The results demonstrated the advantages of GRU in multi-sensor fusion and nonlinear behavior modeling, enabling accurate hive health prediction and timely anomaly alerts for beekeeping decision support. Fig. 3-1 illustrates the overall framework of the study.

doi:10.6342/NTU202503964

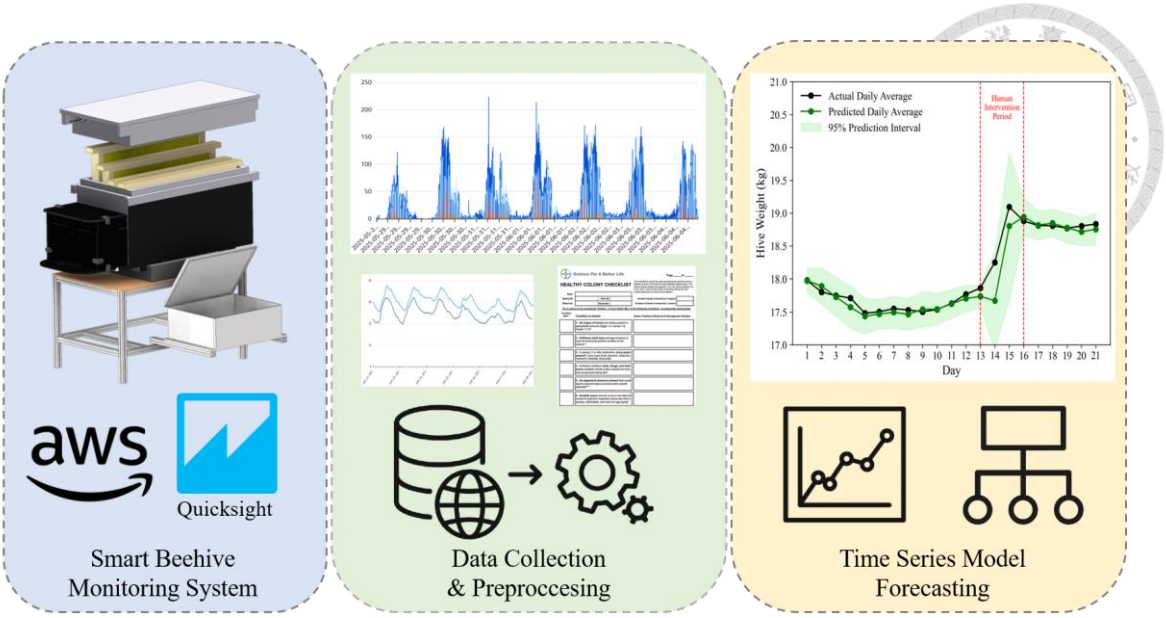


Fig. 3-1 Overall research framework

3.1.1 Hardware Configuration and Sensor Modules

The multi-sensor smart hive architecture, illustrated in Fig. 3-2 (I) and adapted from Liu et al. (2024), comprises two functional subsystems. The first is the environmental sensing module, which collects data on hive weight, temperature, humidity, and acoustic signals. A Raspberry Pi 4 (RPI) serves as the central controller, connecting to each sensor via wired interfaces to ensure stable data transmission. Temperature and humidity are measured using a CMOSens® SHT20 sensor, which is enclosed in a waterproof white casing to protect it from condensation damage. Hive weight is monitored by an EX1704 Power-over-Ethernet load cell, which features a pest-resistant design to effectively block small animals such as geckos. For acoustic monitoring, an omnidirectional lapel USB

doi:10.6342/NTU202503964

microphone is used. It offers a wide frequency response range sufficient to capture the characteristic buzzing sounds of bees.

The second subsystem is responsible for monitoring bee traffic and pollen load. It integrates an acrylic light-shielding tunnel, a camera, and an edge computing processor. The tunnel is mounted at the hive entrance to constrain the bees' flight path and is equipped with red LED strips. Since red light falls outside the bees' visual spectrum, it provides uniform illumination without disturbing colony behavior. Inside the tunnel, a Logitech C922 Pro camera is installed facing the ingress and egress channel and streams video to an NVIDIA Jetson TX2 edge device. This processor runs the YOLOv3 tiny model developed by Ngo et al. (2021) to detect and count bees and pollen. A representative frame from the monitoring system is shown in Fig. 3-2 (II). The research team deployed six hives across two locations, each equipped with multiple key sensors to comprehensively monitor temperature, humidity, weight, bee entry and exit, pollen load, acoustic signals, and local weather conditions. The entire smart beehive system is powered by an AC power supply combined with a smart power outlet. This setup allows remote monitoring and control of the power status in real time, ensuring that the system remains consistently operational over extended periods without manual intervention.

doi:10.6342/NTU202503964

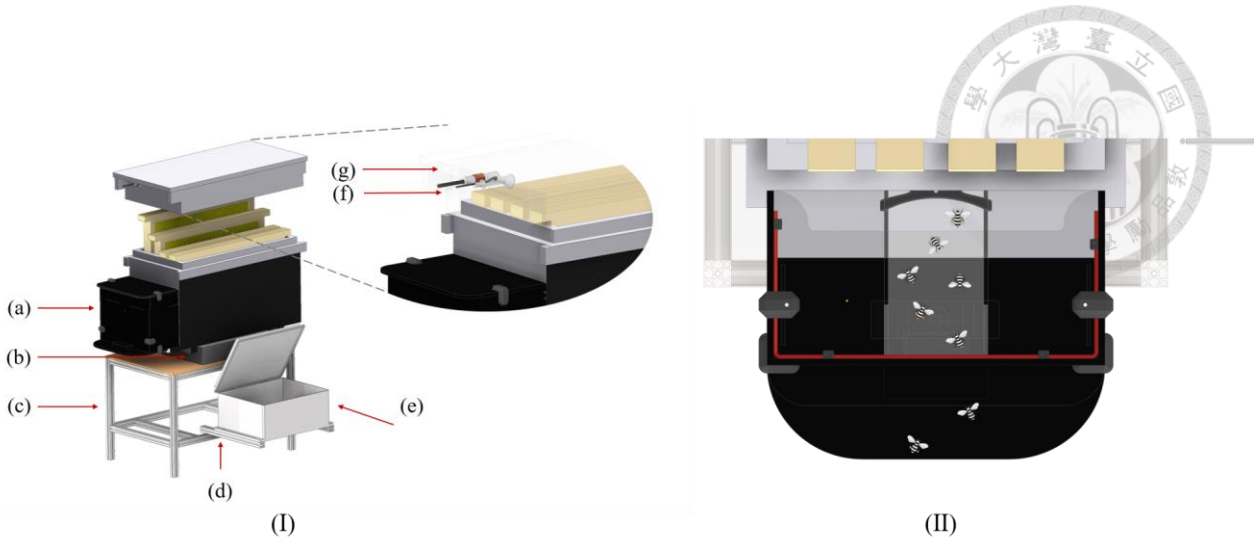


Fig. 3-2 (I) Overall multi-sensor smart beehive system design, (a) bee traffic observation box, (b) weight scale, (c) support frame, (d) sliding rail, (e) waterproof box, (f) microphone, (g) temperature and humidity sensor (SHT20).

(II) Observation of the bee ingress-egress tunnel inside the hive

3.1.2 Cloud-Based Data Management and Real-Time Monitoring

Fig. 3-3 illustrates the overall data transmission workflow from the smart hive to the cloud. Within the hive, the Raspberry Pi and NVIDIA Jetson TX2 transmit lightweight sensor-acquired data in JSON format to AWS IoT Core via Wi-Fi using the MQTT communication protocol. The data are subsequently processed through AWS IoT Analytics for storage and filtering, and are ultimately visualized through Amazon QuickSight, forming a cloud-based monitoring dashboard. Compared to traditional self-hosted servers, AWS provides a more secure data storage environment and a more user-

doi:10.6342/NTU202503964

friendly interface for data visualization. In addition, to further analyze the relationship between internal and external temperature and humidity, as well as the effects of rainfall on bee behavior, the system automatically retrieves real-time environmental data from a government open-data platform.

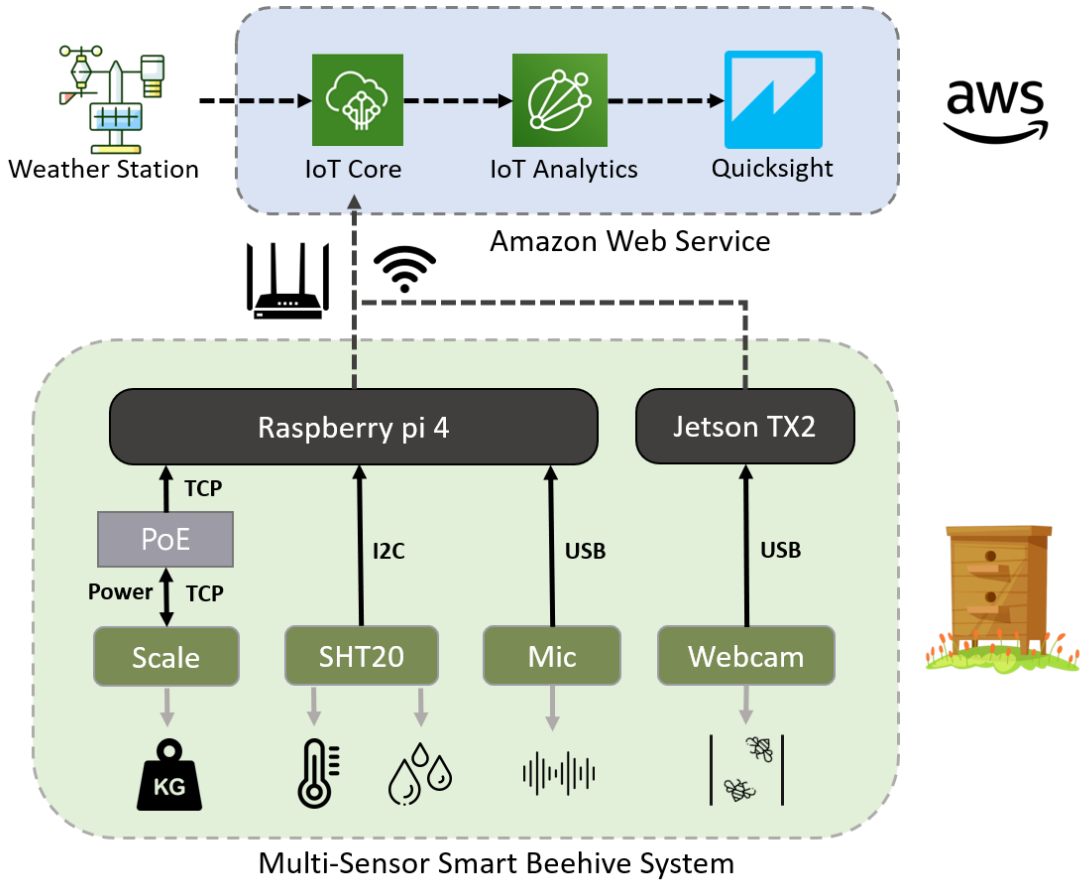


Fig. 3-3 Hardware architecture and sensor data flow through Raspberry Pi and Jetson TX2 to AWS IoT and QuickSight.

As an example, Fig. 3-4 displays the real-time monitoring interface for the first smart hive deployed in Yunlin. The interface is divided into two sections: the left panel shows

doi:10.6342/NTU202503964

real-time data, including bee and pollen traffic, temperature and humidity, and external weather conditions, all presented numerically. The right panel displays historical data trends over the past seven days, including hive weight, bee traffic, and environmental metrics. Through this visualization system, users can quickly identify any anomalies in hive activity, enabling automated monitoring and management of the colony. In the event of a temporary Wi-Fi disconnection, all sensor data are buffered locally within the Raspberry Pi and NVIDIA Jetson TX2. Once the network connection is restored, the system automatically uploads the locally stored data to the AWS IoT Core and clears the cached files to free up storage space. This mechanism guarantees data continuity and prevents information loss during intermittent network disruptions.

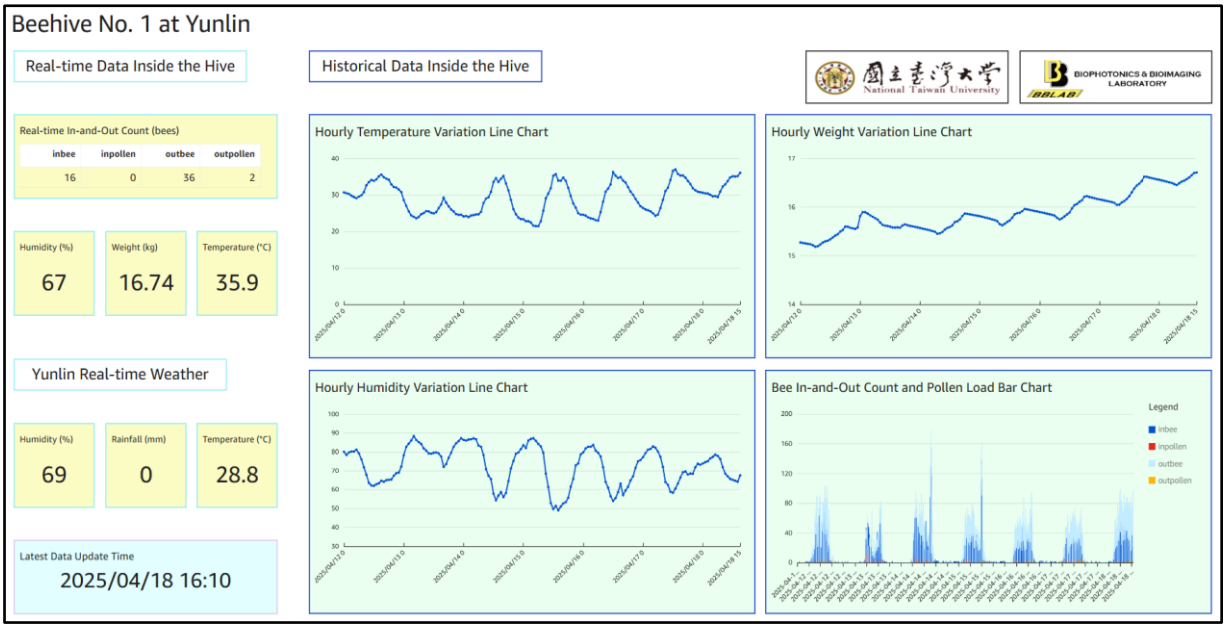


Fig. 3-4 Real-time and historical environmental data dashboard of Yunlin No.1 smart beehive.

doi:10.6342/NTU202503964

3.2 Data Collection

3.2.1 Experimental field

This study established two experimental sites in Yunlin County to observe the operation of smart beehives under different environmental conditions. The first site is located in the Beekeeping Story Park in Gukeng Township, Yunlin County, where four smart beehive systems were installed, as shown in Fig. 3-5 (I). These hives are maintained regularly by park staff and are positioned under tree shade in a demonstration apiary area, though they may be exposed to direct sunlight during midday hours. The second site is situated in the courtyard of a beekeeper's residence in Douliu City, where two additional smart beehive systems were installed, as shown in Fig. 3-5 (II). These hives are placed under a shading canopy, minimizing direct sunlight exposure, and are maintained by the beekeeper and their family.

The beehives at both locations are primarily used for research purposes rather than honey production. The species of bees raised at both sites is the Italian honeybee (Apis mellifera), which is common in Taiwan. Routine beekeeping activities such as sugar water feeding, oxalic acid spraying for mite control, frame reallocation, removal of excess comb, and management of queen cells are carried out to ensure colony health. All maintenance activities are recorded in hive management logs. Each of the six hives consists of four to

doi:10.6342/NTU202503964

six frames and functions as a nucleus (queen-rearing) colony. According to the maintenance records, most hives remained healthy and stable, requiring only occasional frame adjustments. However, Hive 3 experienced queen loss, necessitating the introduction of a new queen to stabilize colony conditions.

Climatically, Gukeng Township has a subtropical monsoon climate, with an average annual temperature of approximately 23.4°C and a summer average of about 28.4°C. The average annual relative humidity is around 73.8%, with higher levels during summer. Annual rainfall ranges from 2000 to 2500 mm, primarily concentrated in the summer and autumn months, with monthly rainfall typically between 300 and 400 mm. The climate of Douliu City is similar, with an average annual temperature of approximately 22.9°C and a summer average of 28.2°C.

(I)

(II)

Fig. 3-5 (I) Smart beehive system at the Bee Museum (II) Smart beehive system in the beekeeper's backyard

doi:10.6342/NTU202503964

3.2.2 Hive Weight Data

Weight data in this study were collected using a PoE-powered scale developed by EXCELL Precision Co., Ltd., model EX1704. This scale supports measurements up to 60 kilograms with a minimum resolution of 0.1 kilograms. It is rated IP54 for water and dust resistance, ensuring stable operation in outdoor environments. In addition, the device features a patented insect-resistant design that prevents interference from large insects or wild animals nesting within the unit. The scale is controlled by a Raspberry Pi 4 (RPI) running a Python program, which collects data at 10-minute intervals. All data are stored simultaneously on both local storage and the cloud. The dataset used in this study consists of weight records collected from each smart hive between August 2024 and March 2025. The scale is shown in Fig. 3-6

Fig. 3-6 PoE-powered weighing scale

doi:10.6342/NTU202503964

3.2.3 Hive Temperature and Humidity

For in-hive temperature and humidity monitoring, this study employed the CMOSens® SHT20 sensor, which transmits data via an I²C communication interface. The sensor supports measurements of relative humidity from 0 to 100% (±3%) and temperatures ranging from -40 to 125 °C (±0.3 °C). To enhance durability and prevent interference from moisture and beeswax, the sensor is enclosed in a waterproof white plastic housing. The sensor is controlled by a Python program running on a Raspberry Pi 4 (RPI), which collects data at 10-minute intervals. All data are stored simultaneously on both local storage and a cloud platform. The dataset used in this study includes sensor data recorded by each smart hive between August 2024 and March 2025.The sensor is shown as Fig. 3-7.

Fig. 3-7 Temperature and humidity sensor

doi:10.6342/NTU202503964

3.2.4 Bee and Pollen Traffic Data

To monitor the inflow and outflow of bees and pollen, this study utilized a bee and pollen traffic monitoring system previously developed by our team (Ngo et al.). The system is powered by the BOXER-8231AI-KIT, an AI module designed by AAEON, featuring an NVIDIA Jetson™ TX2 NX as its central processing unit. The observation box is equipped with red LED strips and a Logitech C922 Pro camera, which is mounted inside a light-shielded enclosure and aligned with the hive entrance. This setup enables real-time detection of both bee traffic and pollen inflow. Data are recorded every five minutes and simultaneously stored on both local and cloud platforms. The dataset used in this study, including records of bee traffic and pollen inflow, was collected from September 2024 to March 2025 by each of the smart beehives.

While the bee ingress–egress and pollen detection system achieved a precision of 0.91 and recall of 0.99, it was developed based on an earlier YOLOv3-tiny architecture. Compared to more recent object detection models, YOLOv3 may have limited accuracy in distinguishing pollen-bearing bees under varying lighting conditions or when multiple bees overlap in the entrance tunnel. These limitations could introduce detection errors and lead to underestimation or overestimation of daily bee traffic and pollen collection rates.

doi:10.6342/NTU202503964

Such factors are considered potential sources of error in the time-series analysis and forecasting results presented in Chapter 4.

3.2.5 Audio Data

For audio data collection, a CKMOVA omnidirectional condenser USB microphone was used. This microphone has a sensitivity of –34 dB ± 3 dB and a frequency response range of 50–20,000 Hz, which sufficiently covers the typical buzzing frequency of bees, predominantly between 100 and 5,000 Hz. Audio was recorded for 30 seconds every 20 minutes. After recording, five major acoustic features were computed: Acoustic Complexity Index, Acoustic Diversity Index, Acoustic Evenness Index, Spectral Entropy, and RMS Energy. All data were simultaneously stored on both local and cloud platforms. The audio dataset used in this study was collected from August 2024 to March 2025 by each of the smart beehives.

3.2.6 Weather Data

For meteorological data, this study utilized the Climate Observation Data Inquire Service (CODiS) provided by the Central Weather Administration under the Ministry of Transportation and Communications. Real-time weather data were obtained from the nearest weather station to the apiary, the Gukeng Flower Center station (Station ID: A2K570). The collected meteorological variables included temperature, relative humidity,

doi:10.6342/NTU202503964

wind speed, wind direction, and precipitation. Data were sampled every 10 minutes and directly uploaded to the cloud.

3.2.7 Honeybee Colony Checklist

The documentation of hive health status serves as a crucial reference for validating the performance and proper functioning of the smart beehive health monitoring system. These records allow for real-time understanding of colony conditions and, when considered alongside the beekeepers' management practices, help elucidate the practical meaning behind the system's recorded data. Moreover, such records provide a reliable foundation for manual labeling required in subsequent machine learning model training.

To assess the actual health conditions of the colonies, local caretakers were asked to regularly complete a health evaluation form, as shown in Table 3-1 Healthy Colony Checklist. Local caretakers were instructed to complete this health evaluation form approximately once per week throughout the monitoring period, ensuring consistent and up-to-date labeling of colony health conditions. The first six indicators in the form are based on the "Healthy Colony Checklist (HCC)" proposed by the Bayer Bee Care Center in the United States (Healthy colony checklist, 2018). The HCC includes six essential criteria for evaluating hive health: colony age structure, brood pattern and development, queen status, nutritional adequacy, external environmental conditions, and space

doi:10.6342/NTU202503964

availability within the hive. Each item is scored as 1 if it meets the standard for a healthy condition, resulting in a maximum score of 6. The HCC form is designed to offer an objective and practical standard for evaluating whether a colony can be considered healthy. According to this definition, a hive is categorized as “apparently healthy” only if all six criteria are met (i.e., a full score of 6). If any criterion is unmet, beekeepers are advised to reconsider their management practices or environmental conditions.

Table 3-1 Healthy Colony Checklist

No.	Evaluation Item	Response
1	Is there a sufficient and seasonally appropriate quantity of brood, larvae, and eggs within the hive?	Yes / No
2	Does the hive contain an adequate and well-balanced age structure of adult bees to support various tasks such as guarding, cleaning, nursing, and foraging?	Yes / No
3	Is there a healthy and young queen (within one year of age) with active reproductive capability?	Yes / No
4	Is there sufficient water and food storage inside and/or outside the hive?	Yes / No
5	Is the colony free from visible external threats (e.g., hornet attacks, parasitic infections)?	Yes / No
6	Is the hive space adequate for colony expansion or food storage, without being excessively crowded or too sparse?	Yes / No

doi:10.6342/NTU202503964

Since the presence of any single abnormal condition in the hive requires immediate intervention by the beekeeper, this study defines only hives with a full score of 6 as "healthy," while those scoring 1 to 5 are categorized as "unhealthy." This binary classification is subsequently used as the labeling criterion for health status classification and model performance evaluation in the machine learning tasks.

3.2.8 Dataset Summary

The study initially deployed multi-sensor monitoring systems across six smart hives. However, due to data interruptions caused by unstable transmission and sensor malfunctions in two hives, only the remaining four hives with complete datasets were included for model training and analysis.

The dataset comprises hive weight, internal temperature and humidity, bee ingress–egress traffic, pollen detection, audio recordings, weather information, and beekeeper-provided Healthy Colony Checklists. Data collection spanned from September 2024 to April 2025. The table below summarizes the final four hives used for analysis:

doi:10.6342/NTU202503964

Table 3-2 Sensor Data Collection Periods and HCC Records for Selected Hives

Hive ID	Sensor and Weather Data	HCC Records
No.2	Oct 2024–Mar 2025	18
No.3	Sep 2024–Mar 2025	20
No.4	Sep 2024–Jan 2025	14
No.6	Nov 2024–Apr 2025	15

3.2.9 Target Variable Definition and Selection

To comprehensively characterize colony health and productivity, this study selected hive weight, bee ingress–egress traffic, and pollen collection rate as the primary forecasting targets.

Hive weight captures the combined mass of bees, stored honey, and food reserves within the hive, serving as a fundamental indicator of colony strength and productivity. Changes in weight also signal seasonal nectar flow, environmental fluctuations, or colony stress events such as queen loss.

Bee traffic, quantified through computer vision, measures the frequency of bees entering and exiting the hive, reflecting foraging activity and overall colony vitality. When colonies experience disease, pesticide exposure, or environmental stress, bee traffic typically declines, making it a critical early-warning signal.

doi:10.6342/NTU202503964

The pollen collection rate, defined as the proportion of returning foragers carrying pollen loads, directly reflects pollination efficiency and floral resource availability; decreases in this metric may indicate limited forage or compromised colony health, affecting agricultural pollination services. Collectively, these three targets describe colony conditions from the perspectives of nutritional reserves, behavioral activity, and pollination efficiency, all of which are highly relevant to practical beekeeping decisions and thus serve as the core objectives of this study's time-series forecasting framework.

3.3 Data Preprocessing

To ensure the quality and consistency of input features used in the forecasting models, a structured data preprocessing and feature engineering pipeline was implemented. This pipeline integrates sensor readings from multiple sources, performs missing value imputation, resamples data to a uniform hourly interval, and applies feature selection techniques to reduce redundancy and improve model robustness. Additionally, normalization procedures are employed to scale variables appropriately for model input. An overview of the full workflow is illustrated in Fig. 3-8 Workflow of missing value handling, feature selection, and normalization.

.

doi:10.6342/NTU202503964

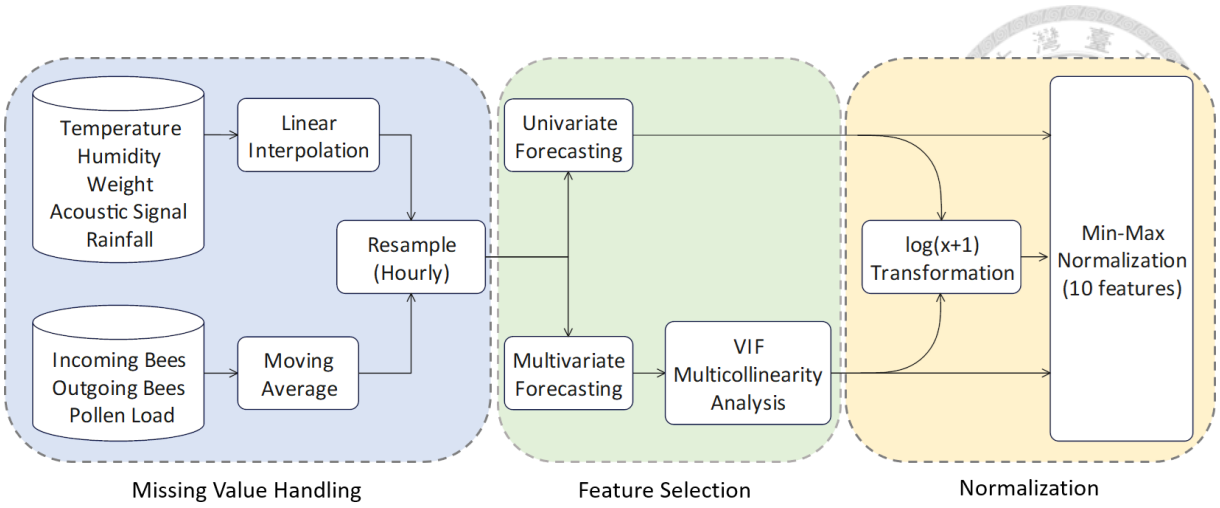


Fig. 3-8 Workflow of missing value handling, feature selection, and normalization.

3.3.1 Missing Value Handling

To ensure data completeness and consistency across different sensor types, this study adopts specific imputation methods according to the nature of each variable. For count-based features such as the number of incoming bees, outgoing bees, and pollen loads, a moving average is applied to interpolate missing values and smooth short-term fluctuations. For continuous variables, including temperature, humidity, and hive weight, linear interpolation is used. In addition, any segment with more than three consecutive hours of missing data is removed to avoid introducing significant bias into the forecasting models. After completing the imputation process, all data are resampled to a uniform hourly interval and aligned in time. Two additional features are then computed: bee traffic, calculated as the total number of incoming and outgoing bees, and pollen collection rate, defined as the ratio of pollen-bearing bee entries to the total number of incoming bees.

doi:10.6342/NTU202503964

3.3.2 Normalization and Outlier Handling

Before feeding the data into the forecasting models, normalization is applied to ensure that all input variables are on a consistent scale. Special treatment is given to the pollen collection rate, which typically falls below 0.5. However, when only a small number of bees return with pollen, the ratio may be disproportionately inflated due to a small denominator. To address this issue, the pollen collection rate is set to zero whenever the number of incoming bees falls below the 20th percentile of hive-specific bee traffic and the timestamp falls within nighttime hours (from 6:00 PM to 8:00 AM).

Subsequently, a log(x+1) transformation is applied to the pollen collection rate to further suppress the influence of extreme values. The same transformation is also applied to the number of incoming bees, outgoing bees, and total bee traffic, as these variables span several orders of magnitude. This transformation enhances the model's ability to capture patterns across a wide range of values while mitigating the impact of outliers. Finally, all features are scaled using Min-Max normalization to map their values into the [0, 1] range, ensuring numerical consistency and accelerating convergence during model training.

doi:10.6342/NTU202503964

3.4 Feature Selection

During the feature selection process illustrated in Fig. 3-8, this study aims to strike a balance between information completeness and statistical independence among input variables. The procedure begins with a correlation analysis to examine the linear relationships between features, serving as an initial basis for variable screening. Subsequently, a Variance Inflation Factor (VIF) analysis is performed to identify and eliminate variables exhibiting high multicollinearity. This approach helps reduce model instability and the risk of overfitting caused by collinearity, thereby enhancing the interpretability and generalization performance of the forecasting models. The theoretical foundations of correlation analysis and VIF are detailed in Sections 3.4.1 and 3.4.2, respectively.

3.4.1 Correlation Analysis

The Pearson Correlation Coefficient is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables. It is denoted by r and ranges from –1 to 1. A value of r close to 1 indicates a strong positive linear relationship, whereas a value close to –1 indicates a strong negative linear relationship. A value near 0 suggests little to no linear correlation.

doi:10.6342/NTU202503964

The Pearson coefficient between variables X and Y is calculated using the following formula:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad \text{3-1}$$

where $\bar{X}$ and $\bar{Y}$ are the sample means of X and Y, respectively. This formula effectively standardizes the covariance of the two variables by the product of their standard deviations, enabling comparison across different scales.

In this study, the Pearson Correlation Coefficient is employed during the feature selection process to detect pairs of features exhibiting strong linear relationships. Specifically, when the absolute value of the correlation coefficient between two variables exceeds 0.7 (Graham, 2003), one of the variables is considered for removal. The decision on which variable to exclude is based on domain relevance and the variable's contribution to the model's predictive performance. This step is crucial to reduce redundancy among input features, mitigate the risk of multicollinearity, and improve the interpretability and stability of the forecasting models.

doi:10.6342/NTU202503964

3.4.2 Variance Inflation Factor

Although some variables exhibit weak pairwise linear correlations in the Pearson analysis, they may still participate in multicollinearity through complex joint interactions with multiple other variables. Therefore, the use of Variance Inflation Factor (VIF) analysis is essential as a complementary step to detect hidden redundancies that cannot be captured by simple correlation measures.

In the feature selection process for multivariate forecasting, this study adopts the Variance Inflation Factor (VIF) (Akinwande et al., 2015) as a key metric to assess multicollinearity among input variables. VIF is a statistical indicator used to quantify the degree to which the variance of a regression coefficient is inflated due to linear relationships between a given predictor and the other predictors in the model. The VIF for a given variable X_j is defined as:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2} \tag{3-2}$$

where

VIF_j = variance inflation factor of the j-th variable

R_j^2 = coefficient of determination obtained by regressing X_j on all other explanatory variables

doi:10.6342/NTU202503964

When a variable is highly linearly correlated with other variables, the resulting R_j^2 approaches 1, making the denominator close to zero and causing the VIF value to grow substantially. A high VIF indicates that the estimated coefficient of the variable may be unstable and highly sensitive to small changes in the model, which may hinder interpretability and prediction performance. In this study, VIF values were calculated for all continuous predictor variables. Variables exhibiting excessive multicollinearity were removed to ensure that the selected features were not only informative but also statistically independent, thereby enhancing the performance and robustness of the forecasting model.

3.5 Time Series Model

Time series modeling plays a critical role in analyzing and forecasting sequential data collected over time, particularly in domains such as climate monitoring, energy management, agriculture, and economic analysis. These datasets are often characterized by temporal dependencies, seasonal patterns, and the presence of exogenous influences. Accurate modeling of such time-dependent behavior enables better planning, decision-making, and real-time response in dynamic systems. Traditional statistical models, such as ARIMA and its seasonal variant SARIMA, have been widely used for time series

doi:10.6342/NTU202503964

forecasting due to their interpretability and solid theoretical foundation. However, their performance may be limited when handling complex nonlinear patterns or when incorporating multiple correlated features. In recent years, the advent of deep learning techniques has introduced powerful alternatives for time series forecasting. Recurrent neural networks (RNNs), especially Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) models, have demonstrated superior capabilities in capturing nonlinear temporal dynamics and long-range dependencies. These models are particularly advantageous in multivariate settings, where multiple time series variables interact over time.

3.5.1 Gated Recurrent Unit

The Gated Recurrent Unit (GRU), proposed by (Cho et al., 2014), is a variant of recurrent neural networks (RNNs) that was developed to mitigate the vanishing gradient problem and enhance the modeling of long-term temporal dependencies. Compared to Long Short-Term Memory (LSTM) networks, GRUs require fewer parameters while maintaining similar performance in many sequence modeling tasks.

doi:10.6342/NTU202503964

The GRU architecture incorporates two gates: the update gate and the reset gate, which together regulate information flow and memory update at each time step. The computations of GRU at time step t are defined as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad 3\text{-}3$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad 3\text{-}4$$

$$\widetilde{h_t} = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad 3\text{-}5$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h_t} \quad 3\text{-}6$$

In the GRU formulation, x_t denotes the input vector at time step t, while h_t represents the hidden state at time t, and $\widetilde{h_t}$ is the candidate hidden state. The update gate z_t determines how much of the previous hidden state should be retained, and the reset gate r_t controls how much of the past hidden information should be forgotten. The sigmoid function $\sigma(\cdot)$ is used to compute the gating values, and the hyperbolic tangent function $\tanh(\cdot)$ is used to generate the candidate hidden state. The symbol $\odot$ indicates element-wise (Hadamard) multiplication. W_* and U_* refer to the learnable weight matrices, while b_* denotes the corresponding bias vectors, all of which are optimized during training.

doi:10.6342/NTU202503964

In this study, we trained a multi-layer Gated Recurrent Unit (GRU) model using a sliding window approach, with historical time-series data as input. The dataset was sampled at hourly intervals, and the model was tasked with forecasting the target variable over the next 24 hours. The data were chronologically split into 80% for training and 20% for testing.

The input sequence length (i.e., the number of past hours used for prediction) was treated as a tunable hyperparameter. For the prediction of hive weight, bee traffic, and pollen collection rate, the optimal input length was selected based on the mean absolute error (MAE) evaluated on the validation set. For hive health status classification, the binary cross-entropy loss with logits (BCE With Logits Loss) was used instead. In addition, we examined the impact of hidden size, number of layers, and dropout rate on the forecasting performance.

All hyperparameters, including input sequence length, were jointly optimized. The detailed configurations evaluated in this study are summarized in Table 3-3.

doi:10.6342/NTU202503964

Table 3-3 Hyperparameter search space for the GRU model

Hyperparameters	**Values**
Input Length (hours)	24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 264, 288, 312, 336, 360
Batch Size	16, 32
Hidden Size	64, 128, 256, 512
Number of GRU Layers	1, 2, 3
Dropout Rate	0, 0.1, 0.3
Learning Rate	0.001, 0.0005, 0.0001
Weight Decay	0, 0.001, 0.0001, 0.00001

3.5.2 SARIMAX

The SARIMAX (Seasonal Auto Regressive Integrated Moving Average with eXogenous variables) model extends the classical ARIMA framework by incorporating both seasonal components and exogenous inputs. Seasonality is addressed through the inclusion of seasonal autoregressive (SAR), seasonal differencing (SD), and seasonal moving average (SMA) terms, while external factors influencing the target variable are modeled using additional exogenous regressors.

doi:10.6342/NTU202503964

The general form of the SARIMAX model is given by:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{P} \Phi_i y_{t-is} + \sum_{i=1}^{Q} \Theta_i \varepsilon_{t-is} + \beta^{\mathrm{T}} x_t + \varepsilon_t \quad 3\text{-}7$$

In this equation, y_t is the target variable at time t, and ε_t is a white noise error term. The vector $x_t \in R^k$ represents the exogenous inputs at time t, and β denotes the associated regression coefficients. The orders p, d, q correspond to the non-seasonal autoregressive, differencing, and moving average terms, respectively, while P, D, Q refer to their seasonal counterparts. The seasonal period is denoted by s, such as $s = 24$ for hourly data with daily seasonality.

For compact notation, the model is written as:

$$SARIMAX(p, d, q)(P, D, Q)_s + \mathrm{X_t} \quad 3\text{-}8$$

To ensure stationarity, differencing operations of order d and D are applied to the original time series. The selection of model orders is guided by empirical diagnostics using the autocorrelation function (ACF) and partial autocorrelation function (PACF), while final model evaluation is performed using information criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

doi:10.6342/NTU202503964

In multivariate forecasting, the availability of future exogenous variables becomes a critical consideration, as these values are required at each forecasted time step. There are several strategies to address this challenge. One approach is to use lagged versions of the exogenous inputs, treating past values as a proxy for future conditions. Another method involves training individual SARIMA models to forecast each exogenous variable independently, and then using the predicted values as inputs to the SARIMAX model. A third strategy assumes that the true future values of the exogenous variables are known in advance, which may be feasible in simulation studies but is often impractical in real-world applications.

In this study, we adopt the second approach to reflect practical deployment scenarios and to ensure a fair comparison with the GRU model. Specifically, each exogenous variable is forecasted using a separate SARIMA model, and the resulting predictions are used as the input to the multivariate SARIMAX model. The SARIMAX and SARIMA models are trained using historical time-series data, sampled at one-hour intervals. A sliding window method is employed for both single-variable and multivariate forecasting tasks. The forecasting objective is to predict the target variable over the next 24 hours (multi-step forecast). The dataset is chronologically split into 80% for training and 20% for testing. The parameter search ranges and information criteria used for model selection are summarized in.

doi:10.6342/NTU202503964

Table 3-4 Hyperparameter search space and modeling criteria for the SARIMAX model

Category	Parameter / Setting	Search Range / Description
Non-seasonal orders	p, d, q	$p, q \in [0,3];\ d \in [0,2]$
Seasonal orders	P, D, Q, s	$P, Q \in [0,2];\ D \in [0,1];\ s = 24$
Model selection criteria	Evaluation metric	Akaike Information Criterion (AIC)
Exogenous variables	Input strategy	Forecasted individually using SARIMA

3.6 Model Interpretation and Shifting Detection

3.6.1 SHAP-Based Feature Interpretation

To evaluate the contribution of each input feature in the GRU-based multi-step forecasting model, this study employed SHAP (SHapley Additive exPlanations) as the interpretability method. SHAP is based on the concept of Shapley values from cooperative game theory. It quantifies the marginal effect of each feature by averaging its contribution over all possible feature subsets. The formal definition of the Shapley value is given as follows:

$$\phi_i(f, x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} \left[f_{S \cup \{i\}}(x) - f_S(x) \right] \qquad 3\text{-}8$$

doi:10.6342/NTU202503964

Here, $\phi_i(f, x)$ denotes the SHAP value of feature i for input sample x, N is the set of all features, S is a subset that excludes i, and $f_S(x)$ represents the model prediction using only the features in subset S. This method satisfies important theoretical properties such as fairness, consistency, and local accuracy, making it widely applicable in the interpretation of machine learning and deep learning models (Lundberg & Lee, 2017).

In this study, the SHAP DeepExplainer module was applied to interpret the daily prediction outputs of the trained GRU forecasting model. To construct the background dataset required by SHAP, a small subset of test samples was randomly selected to simulate the model's baseline behavior across various feature combinations. To reduce computational cost, only a representative sample was used rather than the full test set, a strategy shown in prior research (Lundberg & Lee, 2017) to maintain accurate estimation while significantly improving efficiency.

After establishing the background dataset, another random set of test samples was selected for explanation. SHAP values were computed for each feature based on the daily aggregated predictions. The absolute values of SHAP were used throughout the analysis to avoid cancellation between positive and negative contributions, providing a fair estimation of feature importance.

doi:10.6342/NTU202503964

This analysis focused on comparing feature contributions across different beekeeping sites, including hives placed in backyard environments and hives located at the Honey Museum. The comparison was conducted under three forecasting targets: bee in–out volume, pollen collection rate, and total pollen load. By examining the variation in SHAP distributions across these sites and prediction targets, we aimed to reveal how differences in environmental conditions and management practices influence the model's dependence on specific input features.

3.6.2 Stagnation and Forecast Shifting Detection

To evaluate whether a multi-step time series forecasting model exhibits stagnation or forecast shifting, this study proposes a data-driven detection method based on two dynamic indicators: output amplitude and relative change rate. These indicators are further complemented by a behavioral diagnosis using error comparison against baseline models. For each forecast window, the model generates a sequence of h-step predictions denoted as $(\widehat{y_1}, \widehat{y_2}, \dots, \widehat{y_h})$. Two unitless indicators are defined to characterize the dynamics of this sequence.

The first indicator, called the **Forecast Range Ratio**, measures the normalized amplitude of the forecast series. It is computed by dividing the difference between the

doi:10.6342/NTU202503964

maximum and minimum predicted values by the absolute mean of the sequence, formulated as

$$R = \frac{\max(\widehat{y_1}, \ldots, \widehat{y_h}) - \min(\widehat{y_1}, \ldots, \widehat{y_h})}{\left|\frac{1}{h}\sum_{t=1}^{h} \widehat{y_t}\right| + \varepsilon} \quad \text{3-9}$$

where ε is a small constant added to prevent division by zero. A low range ratio indicates that the output sequence has minimal fluctuation, which may suggest a degenerate prediction.

The second indicator is the **Relative Consecutive Difference**, which quantifies the average magnitude of changes between adjacent predictions, again normalized by the mean of the sequence. It is calculated using the formula:

$$D = \frac{\frac{1}{h-1}\sum_{t=2}^{h} |\widehat{y_t} - \widehat{y_{t-1}}|}{\left|\frac{1}{h}\sum_{t=1}^{h} \widehat{y_t}\right| + \varepsilon} \quad \text{3-10}$$

This indicator reflects the smoothness of the forecast. When both the range ratio and consecutive difference are small, the forecast sequence is likely stagnated—exhibiting little change over time, even if the input dynamics are nontrivial.

In addition to dynamic profile analysis, the method includes a comparison with a shifting baseline model to detect possible forecast shifting. The shifting baseline assumes

doi:10.6342/NTU202503964

that each predicted value simply repeats the most recent observed value. The performance of this baseline is measured using its mean absolute error (MAE), defined as:

$$\mathrm{MAE}_{\mathrm{shift}} = \frac{1}{h-1}\sum_{t=2}^{h}|y_t - y_{t-1}| \qquad 3\text{-}11$$

By comparing the model's actual MAE to that of the shifting baseline, we assess whether the model is merely replicating the time series' persistence without learning meaningful temporal patterns. If the two errors are closely aligned, it indicates that the model may be suffering from forecast shifting, producing outputs that behave no differently from a lagged copy of the original signal.

This diagnostic framework requires no access to internal model parameters, making it broadly applicable to both deep learning models such as GRUs and traditional statistical models like SARIMAX. It provides a post hoc evaluation mechanism to detect symptoms of forecast degradation, such as over-smoothing or temporal copying, and enhances the interpretability and reliability of time series models in practical applications.

doi:10.6342/NTU202503964

CHAPTER 4

Results and Discussions

4.1 Feature Selection Results

To improve the performance and interpretability of the forecasting models, we implemented a two-stage feature selection strategy tailored to each prediction target: hive weight, bee traffic, and pollen collection rate. In the first stage, we calculated Pearson correlation coefficients between all variables to identify any strong pairwise linear associations. While this step provided a useful preliminary overview, it was not used as the primary selection criterion. Some variables exhibited low correlation coefficients when considered in isolation but could still contribute to multicollinearity when interacting with other predictors. Therefore, the second stage relied primarily on the Variance Inflation Factor (VIF) to eliminate such redundant variables.

In the VIF-based selection stage, variables were iteratively removed in descending order of their VIF values until all remaining predictors had VIF scores below the commonly accepted threshold of 5. This method ensures that the final input variables are free from serious multicollinearity, which can otherwise distort model coefficients and reduce generalizability(Akinwande et al., 2015).

doi:10.6342/NTU202503964

For the multivariate hive weight forecasting model, the selected features included hive weight itself, bee inflow, bee outflow, pollen inflow, temperature, humidity, acoustic signal amplitude (RMS), and rainfall. These variables collectively captured the internal and external conditions influencing hive weight dynamics. In the bee traffic prediction task, the final feature set comprised traffic volume, hive weight, temperature, humidity, acoustic data, and rainfall. This configuration reflects both the internal hive state and environmental conditions that affect foraging activity and movement patterns. For forecasting pollen collection rate, the selected predictors were pollen rate, hive weight, temperature, humidity, RMS amplitude, rainfall, and bee inflow. These variables were chosen for their relevance in modeling the physiological and environmental drivers behind pollen foraging behavior.

Through this feature selection framework, each forecasting model was provided with an optimized subset of explanatory variables, ensuring low multicollinearity while preserving the predictive power essential for robust time-series modeling.

4.2 Hive Weight Prediction Results

Monitoring and forecasting hive weight play a crucial role in assessing colony health and productivity. A significant drop in hive weight may indicate abnormal events such as

doi:10.6342/NTU202503964

swarming or the onset of Colony Collapse Disorder (CCD), both of which pose serious threats to colony sustainability. By constructing predictive models that can anticipate future weight changes, beekeepers are equipped with early warning mechanisms that support timely interventions. These predictive insights enable better hive management strategies, reduce the risk of colony loss, and ultimately contribute to sustainable apiculture.

4.2.1 Univariate Hive Weight Prediction Results

Fig. 4-1 presents the univariate forecasting results for hive weight using the SARIMA model (I) and the GRU model (II). Both models were trained using a sliding window approach to generate 24-step hourly forecasts, which were subsequently aggregated into daily averages for comparison with the actual daily hive weight measurements. In each figure, the black line represents the actual daily average hive weight, the green line shows the model's predicted values, and the shaded green region denotes the 95% prediction interval.

For model configuration, the SARIMA model achieved the best performance with the orders $(p, d, q) = (1,1,1)$ and seasonal orders $(P, D, Q)_s = (1,0,1)_{24}$, successfully capturing both non-seasonal and 24-hour seasonal autoregressive components. The GRU

doi:10.6342/NTU202503964

model employed a two-layer architecture with 128 hidden units per layer, a batch size of 16, a learning rate of 0.001, a dropout rate of 0.3, and no weight decay. These settings were selected to ensure sufficient model capacity while mitigating the risk of overfitting.

In terms of forecasting trends, both models were able to capture the main patterns of hive weight dynamics, including the steady decline in the early period and the sharp increase occurring between Days 14 and 15. This abrupt rise was primarily due to manual hive interventions performed by the beekeeper, which presents a practical challenge for time series forecasting models. During the intervention period, the predictions of both models deviated from the actual values. In contrast, under normal, non-intervened conditions, the GRU model provided forecasts that more closely followed the actual hive weight trajectory. The SARIMA model, while stable, exhibited wider prediction intervals, indicating its limitations in issuing timely alerts in field applications.

In terms of forecasting accuracy, the SARIMA model achieved a mean squared error (MSE) of 0.074 kg^2, a root mean square error (RMSE) of 0.272 kg, a mean absolute error (MAE) of 0.167 kg, and a mean absolute percentage error (MAPE) of 0.9%. In comparison, the GRU model achieved lower error metrics, with an MSE of 0.057 kg^2, RMSE of 0.238 kg, MAE of 0.130 kg, and MAPE of 0.7%. Overall, while SARIMA performed more consistently during periods of regular variation, the GRU model

doi:10.6342/NTU202503964

demonstrated better predictive accuracy and stronger responsiveness to irregular changes, making it more suitable for anomaly detection tasks in practical beekeeping scenarios.

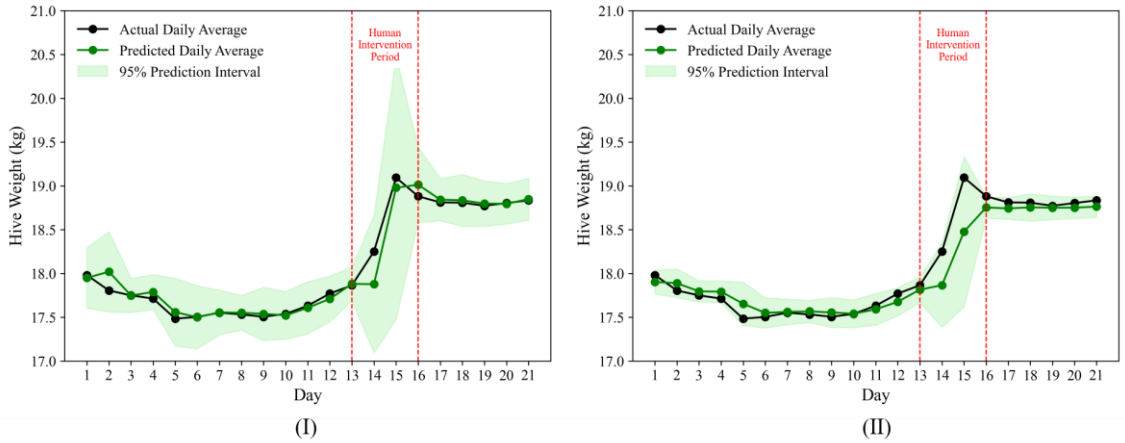


Fig. 4-1 Univariate hive weight forecasting using SARIMA and GRU models. (I) SARIMA model prediction; (II) GRU model prediction.

4.2.2 Multivariate Hive Weight Prediction Results

Fig. 4-2 presents the multivariate forecasting results for hive weight using the SARIMAX model (I) and the GRU model (II). In contrast to the univariate setting, both models were trained not only on historical hive weight data but also on a set of seven carefully selected exogenous variables. These variables—including incoming bee volume, outgoing bee volume, pollen collection count, internal hive humidity, internal hive temperature, acoustic signal amplitude (RMS), and rainfall—were chosen based on variance inflation factor (VIF) multicollinearity analysis to avoid redundancy and ensure

doi:10.6342/NTU202503964

the relevance of input features. The incorporation of these external indicators aimed to enhance each model's ability to reflect the multifactorial nature of hive weight changes.

The SARIMAX model adopted the optimal parameters $(p, d, q) = (1,1,1)$ and seasonal orders $(P, D, Q)_s = (1,0,1)_{24}$, maintaining consistency with the configuration in the univariate experiment. The GRU model employed a one-layer architecture with 128 hidden units, a batch size of 16, a learning rate of 0.001, and no weight decay.

As illustrated in Fig. 4-2, both models effectively captured the overall dynamics of hive weight changes, including the steady decline during the early days and the sharp increase between Days 14 and 15, which was primarily caused by beekeeper intervention. After incorporating multivariate inputs, the GRU model was able to generate more appropriate prediction intervals, avoiding the issue observed in the univariate case where actual values frequently approached the boundaries of the confidence band under normal conditions. In contrast, although the SARIMAX model provided generally reasonable forecasts, its wider prediction intervals indicated lower sensitivity to abrupt deviations, potentially limiting its responsiveness for real-time anomaly detection.

With regard to forecasting accuracy, the SARIMAX model achieved a mean squared error (MSE) of 0.067 kg², a root mean square error (RMSE) of 0.260 kg, a mean absolute error (MAE) of 0.150 kg, and a mean absolute percentage error (MAPE) of 0.8%. The

doi:10.6342/NTU202503964

GRU model yielded improved performance across all metrics, with an MSE of 0.055 kg², RMSE of 0.235 kg, MAE of 0.118 kg, and MAPE of 0.6%. Overall, both models demonstrated enhanced forecasting accuracy compared to their univariate counterparts, confirming the benefit of incorporating relevant exogenous variables. Notably, the GRU model consistently outperformed SARIMAX, highlighting its suitability for multivariate time series forecasting in complex hive monitoring applications.

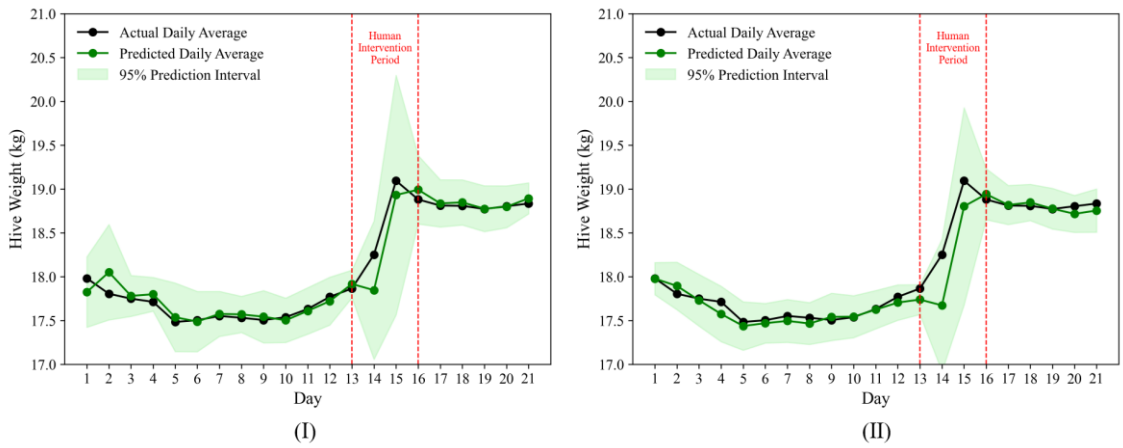


Fig. 4-2 Multivariate hive weight forecasting using SARIMA and GRU models.

(I) SARIMA model prediction; (II) GRU model prediction.

4.2.3 Performance Evaluation of Long-Horizon Multi-Step Forecasting

Using hive weight forecasting as an example, Fig. 4-3 illustrates the results of predicting the next three days of hive weight based on seven days of multivariate sensor data in a single multi-step prediction. As the forecast horizon increases, the 95%

doi:10.6342/NTU202503964

prediction interval (PI) widens, indicating growing uncertainty, and the predicted values gradually deviate from the actual observations. The Mean Absolute Errors (MAE) for Day 1, Day 2, and Day 3 are 0.307 kg, 0.387 kg, and 0.452 kg, respectively, demonstrating the common issue of error accumulation in long-horizon forecasting. In contrast, adopting a one-day-ahead forecasting strategy—where the model is updated daily to predict the next day—achieves a significantly lower MAE of 0.118 kg. Considering model performance and practical deployment, this study ultimately adopts a rolling one-day forecasting approach, ensuring high prediction accuracy while effectively capturing multi-day hive weight trends.

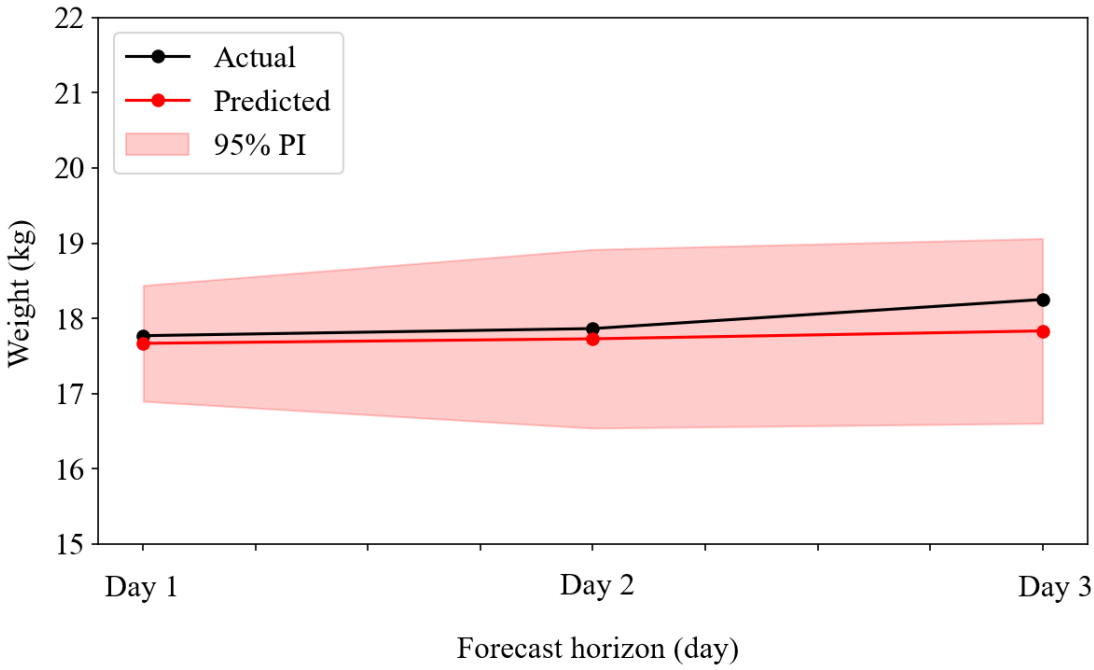


Fig. 4-3 Three-day hive weight forecasting trend line chart

doi:10.6342/NTU202503964

4.2.4 Effects of Human Interference on Hive Weight

Human intervention was identified as a critical factor contributing to increased prediction errors, especially during hive maintenance events. As shown in Figure 4 and Figure 7, a notable deviation between predicted and actual hive weight values occurred during Days 14 to 15, which coincided with manual operations conducted by the beekeeper. These nonstationary disturbances introduced by external human activities posed a significant challenge to both statistical and neural network forecasting models.

To quantitatively assess the impact of such interventions, we evaluated model performance under two conditions: including and excluding the human-intervention period. Table 4-1 Comparison of Prediction Errors with and without Human Interference summarizes the changes in mean absolute error (MAE) and mean absolute percentage error (MAPE) across both univariate and multivariate models. In the multivariate setting, the SARIMAX model's MAE dropped from 0.150 kg to 0.107 kg—a 28.8% reduction—when the intervention period was excluded, and its MAPE decreased from 0.8% to 0.6%. The GRU model showed a similar trend, with MAE decreasing from 0.118 kg to 0.081 kg (a 31.5% improvement) and MAPE dropping from 0.6% to 0.5%. Comparable reductions in prediction errors were also observed in the univariate models.

doi:10.6342/NTU202503964

These findings highlight the disruptive effect of human interference on hive weight predictability. Such events introduce sudden shifts in the data that violate the stationarity assumptions underlying traditional time series models and exceed the generalization capacity of neural networks trained on regular patterns. To mitigate this issue in real-world deployments, future forecasting systems may benefit from explicitly identifying and accounting for human interventions. This could be achieved by incorporating them as auxiliary input features or by masking their timestamps during model training to prevent overfitting to abnormal transitions. Doing so may enhance model robustness and yield more reliable performance under practical field conditions.

Table 4-1 Comparison of Prediction Errors with and without Human Interference

Setting	Human Interference	SARIMA (MAE/MAPE)	GRU (MAE/MAPE)
Univariate	Included	0.167 / 0.910	0.130 / 0.712
	Removed	0.121 / 0.677	0.085 / 0.470
Multivariate	Included	0.149 / 0.821	0.118 / 0.646
	Removed	0.106 / 0.596	0.080 / 0.450

4.3 Bee Traffic Prediction Results

Accurate forecasting of bee traffic is crucial for assessing colony activity, foraging behavior, and overall hive health. The number of bees entering and exiting the hive

doi:10.6342/NTU202503964

reflects both internal biological status and external environmental conditions. Sudden changes in traffic patterns, such as significant drops in foraging or prolonged imbalance between ingress and egress, may serve as early warning signs of stress, resource shortages, predator disturbances, or even colony collapse. Predictive modeling of bee traffic supports continuous behavioral monitoring and enables timely interventions to mitigate potential risks. This section presents the results of both univariate and multivariate models used to predict bee traffic, evaluating their ability to capture temporal patterns and respond to influencing factors.

4.3.1 Univariate Bee Traffic Prediction Results

Fig. 4-4 presents the univariate prediction results of daily bee ingress and egress, defined as the total number of bees entering and exiting the hive, using the SARIMA model (I) and the GRU model (II). The gray bars represent the actual daily traffic volumes, while the red dots indicate the predicted values. The red error bars show the 95% prediction intervals, providing an estimate of model uncertainty over time.

Both models were configured with optimal hyperparameters identified during the model selection process. The SARIMA model used orders $(p, d, q) = (2,0,1)$ and seasonal orders $(P, D, Q)_s = (1,0,1)_{24}$. The GRU model was constructed with two layers,

doi:10.6342/NTU202503964

each containing 64 hidden units, and trained with a batch size of 32, a learning rate of 0.0001, a dropout rate of 0.3, and a weight decay of 0.0001.

As shown in the figure, both models successfully captured the overall temporal trend of bee traffic. However, their performance differed in terms of accuracy and prediction stability. The SARIMA model responded smoothly to fluctuations and exhibited narrower prediction intervals, but it notably underestimated bee traffic on several days, leading to significant deviations from actual values. In contrast, the GRU model produced predictions that more closely matched the observed values across most days. Its prediction intervals better reflected the dynamic variability in bee activity, indicating a stronger ability to learn and adapt to nonlinear and time-varying patterns.

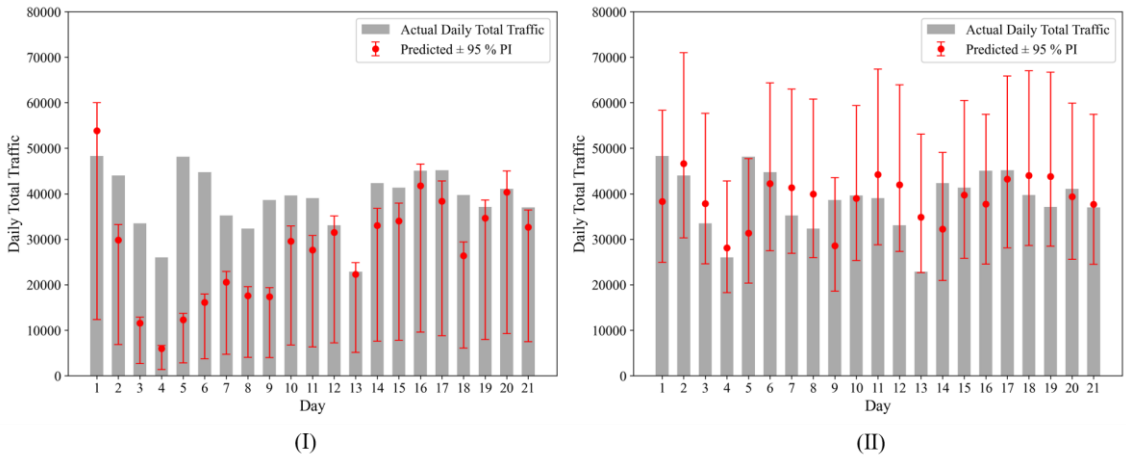


Fig. 4-4 Univariate daily bee traffic prediction results using SARIMAX and GRU models. (I) SARIMAX model; (II) GRU model.

doi:10.6342/NTU202503964

Regarding forecasting accuracy, the SARIMA model achieved a mean absolute error (MAE) of 15,480 bees and a mean absolute percentage error (MAPE) of 40.2%. The GRU model outperformed SARIMA, with an MAE of 7,296 bees and a MAPE of 20.1%, effectively reducing the error by nearly half. These results highlight the GRU model's superior capacity for modeling complex and volatile bee traffic patterns under univariate settings.

4.3.2 Multivariate Bee Traffic Prediction Results

To evaluate the added value of incorporating external information, this section presents multivariate forecasting results for bee ingress and egress using the SARIMAX model (Figure 6a) and the GRU model (Figure 6b). Unlike the univariate setup, these models utilized not only historical bee traffic data but also a set of exogenous variables selected via variance inflation factor (VIF) analysis to minimize multicollinearity. The final inputs included hive weight, internal temperature, internal humidity, acoustic signal amplitude (RMS), and rainfall. These variables were chosen to reflect both internal hive conditions and external environmental influences on bee activity.

Model configurations were kept consistent with the univariate setup, with SARIMAX using $(p, d, q) = (2,0,0)$ and seasonal orders $(P, D, Q)_s = (1,0,1)_{24}$. The

doi:10.6342/NTU202503964

GRU model retained its two-layer architecture with 64 hidden units per layer, a batch size of 32, a learning rate of 0.0001, dropout of 0.3, and weight decay of 0.0001.

Figure 6 shows that both models successfully tracked the overall fluctuation in bee traffic over the 21-day period. However, their behavior differed notably from the univariate results discussed in Section 4.2.1. The inclusion of exogenous variables helped reduce the overall forecast bias, particularly for the GRU model, which produced predictions that were generally closer to the actual daily totals. The GRU model also better accommodated fluctuations in bee activity and maintained tighter prediction intervals compared to its univariate counterpart. In contrast, while the SARIMAX model demonstrated improved performance over its univariate version, its forecasts still lagged behind those of the GRU model. The SARIMAX prediction intervals were wider, and discrepancies from actual values were more pronounced during periods of rapid change.

Quantitative results further underscore this difference. The SARIMAX model achieved a mean absolute error (MAE) of 13,503 bees and a mean absolute percentage error (MAPE) of 35.5%. In comparison, the GRU model achieved a substantially lower MAE of 7,113 bees and a MAPE of 19.0%. These findings confirm that integrating context-rich environmental features enhances the forecasting of complex and dynamic bee movement patterns, and that GRU-based models are particularly effective in learning such multivariate relationships.

doi:10.6342/NTU202503964

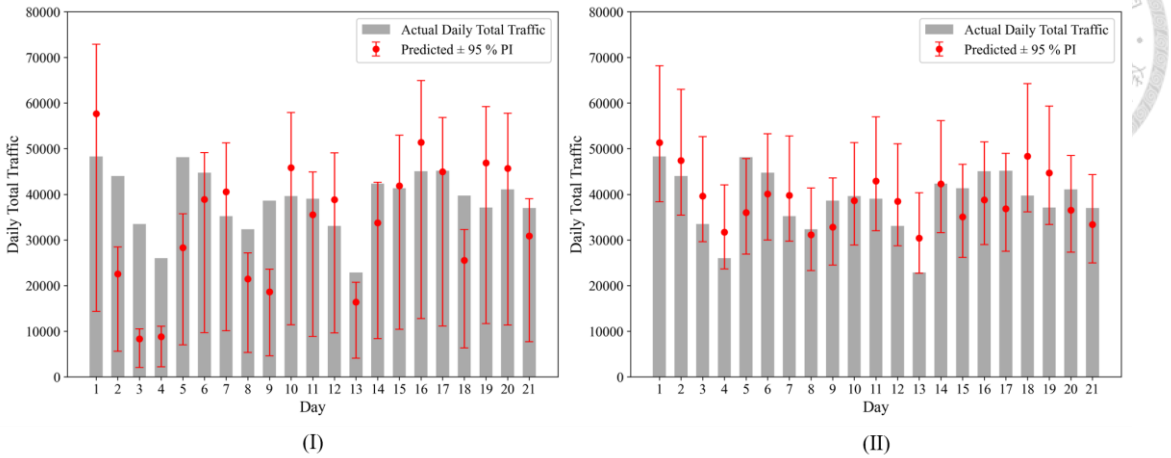


Fig. 4-5 Multivariate daily bee traffic prediction results using SARIMAX and GRU models. (I) SARIMAX model; (II) GRU model.

4.4 Pollen Collection Rate Prediction Results

Pollen collection activity serves as a direct indicator of a colony's foraging efficiency, nutritional intake, and overall vitality. Unlike general traffic volume, the pollen collection rate specifically reflects the success of foragers in locating and returning with floral resources, which is critical for brood development and colony growth. Sudden drops or irregularities in pollen collection may signal unfavorable environmental conditions, floral scarcity, or health issues within the hive. Consequently, the ability to accurately predict pollen collection rates enables early detection of potential stressors and supports data-driven decision-making in apicultural management. This section presents the results of

doi:10.6342/NTU202503964

both univariate and multivariate forecasting models applied to pollen collection rate data, evaluating their capacity to anticipate changes in foraging productivity under various influencing factors.

4.4.1 Univariate Pollen Collection Rate Prediction Results

Fig. 4-6 illustrates the univariate prediction results for pollen collection rate using the SARIMA model (Figure 6a) and the GRU model (Figure 6b). The pollen collection rate is defined as the number of pollen loads collected per hour divided by the number of incoming bees, representing the efficiency with which the colony gathers pollen relative to its foraging activity. This metric is particularly sensitive to changes in weather conditions, floral resource availability, and bee behavior, which makes it inherently difficult to forecast with high accuracy.

In the figure, the black line indicates the actual observed pollen collection rate, while the red line shows the predicted values. The x-axis represents time in days, and the y-axis denotes the normalized pollen collection rate. For the SARIMA model, the optimal configuration included orders $(p, d, q) = (3,0,0)$ and seasonal orders $(P, D, Q)_s = (1,0,1)_{24}$. The GRU model was trained with a hidden size of 128, two layers, a batch size of 32, a learning rate of 0.005, a dropout rate of 0.1, and a weight decay of 1×10^{-4}

doi:10.6342/NTU202503964

In terms of forecasting performance, both models were able to follow the general daily cycles and sharp transitions in pollen collection rate, but their effectiveness varied. The SARIMA model exhibited reasonable performance during periods of moderate fluctuation but failed to capture the sudden spikes and fell short particularly during nighttime hours, when it often predicted non-zero values despite the expected absence of pollen foraging. In contrast, the GRU model demonstrated improved responsiveness to both abrupt changes and low-activity periods. Its predictions more closely followed the actual values and successfully reflected the sharp rises and near-zero troughs that characterize daily pollen collection dynamics.

Quantitatively, the SARIMA model achieved a mean absolute error (MAE) of 0.072 and a symmetric mean absolute percentage error (SMAPE) of 25.0%. The GRU model outperformed SARIMA with an MAE of 0.070 and a SMAPE of 18.8%. While the difference in MAE was marginal, the substantial improvement in SMAPE highlights the GRU model's superior ability to maintain proportional accuracy across both high and low collection intervals. Overall, GRU provided more realistic and temporally consistent forecasts without the night-time prediction failures observed in SARIMA, demonstrating its potential for modeling time-sensitive, nonlinear patterns in pollen foraging behavior.

doi:10.6342/NTU202503964

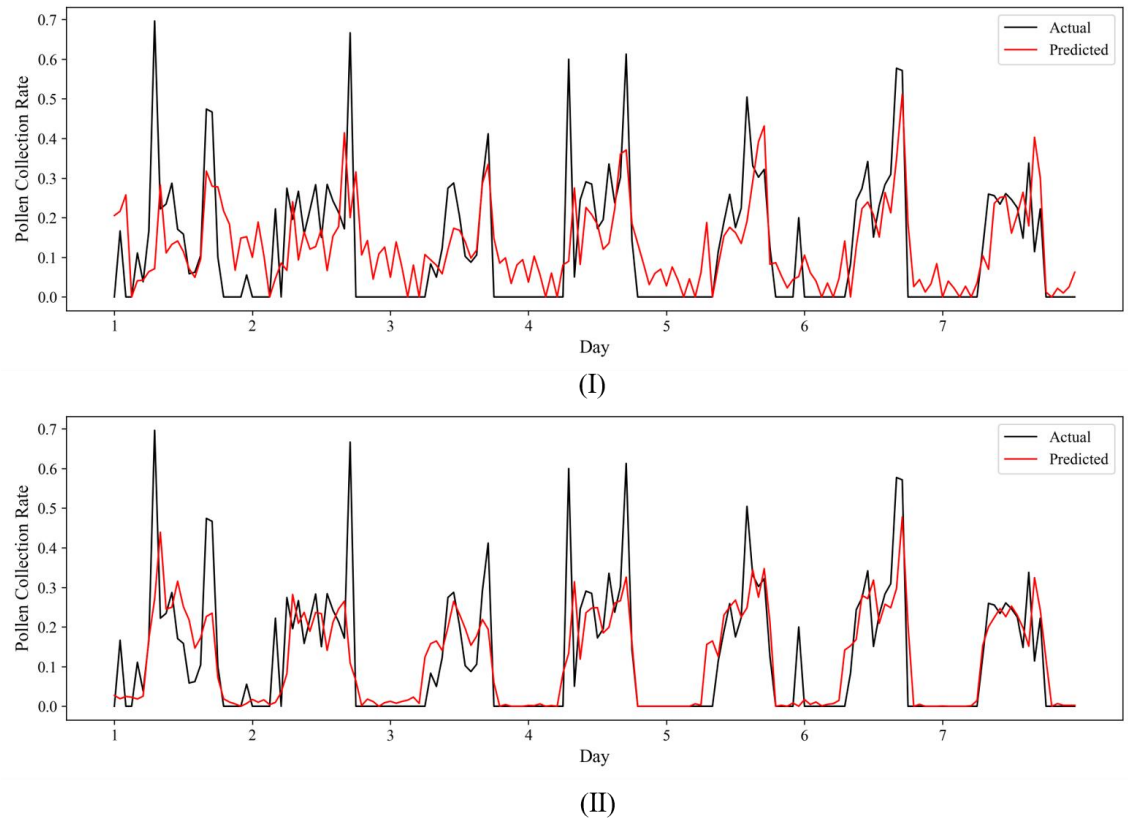


Fig. 4-6 Univariate prediction results of pollen collection rate using SARIMA and GRU models. (I) SARIMA model; (II) GRU model.

4.4.2 Multivariate Pollen Collection Rate Prediction Results

Building upon the univariate analysis, this section presents multivariate prediction results for pollen collection rate using the SARIMAX model and the GRU model. In this extended setting, both models incorporated six exogenous variables selected through variance inflation factor (VIF) multicollinearity analysis: hive weight, internal hive temperature, internal hive humidity, acoustic signal amplitude (RMS), rainfall, and

doi:10.6342/NTU202503964

outgoing bee volume. These variables were introduced to better capture both internal colony dynamics and environmental factors that influence foraging behavior.

The SARIMAX model was configured with optimal parameters of $(p, d, q) = (1,0,0)$ and seasonal orders $(P, D, Q)_s = (1,0,1)_{24}$. The GRU model utilized a single-layer architecture with 256 hidden units, a batch size of 32, a learning rate of 0.001, a dropout rate of 0.1, and a weight decay of 0.0001.

The prediction results reveal distinct characteristics between the two approaches. The GRU model demonstrated a strong ability to track actual fluctuations in pollen collection rate, accurately capturing both sharp increases and near-zero troughs. Its forecasts were temporally stable and well aligned with observed trends, even during rapidly changing periods. In contrast, the SARIMAX model, while generally following the overall trend, exhibited more pronounced fluctuations in its predictions. This made it harder to match the shape and timing of the actual curve, particularly during short-term transitions, suggesting reduced reliability in modeling fine-grained behavioral shifts.

In terms of quantitative performance, the SARIMAX model achieved a mean absolute error (MAE) of 0.071 and a symmetric mean absolute percentage error (SMAPE) of 22.0%. The GRU model outperformed SARIMAX, achieving an MAE of 0.069 and a SMAPE of 16.7%. Compared to their univariate counterparts, both models benefited from

doi:10.6342/NTU202503964

the inclusion of multivariate information, but the GRU model showed more consistent gains. These results highlight the advantage of deep learning models in capturing complex interactions in foraging dynamics when supported by a diverse set of contextual features.

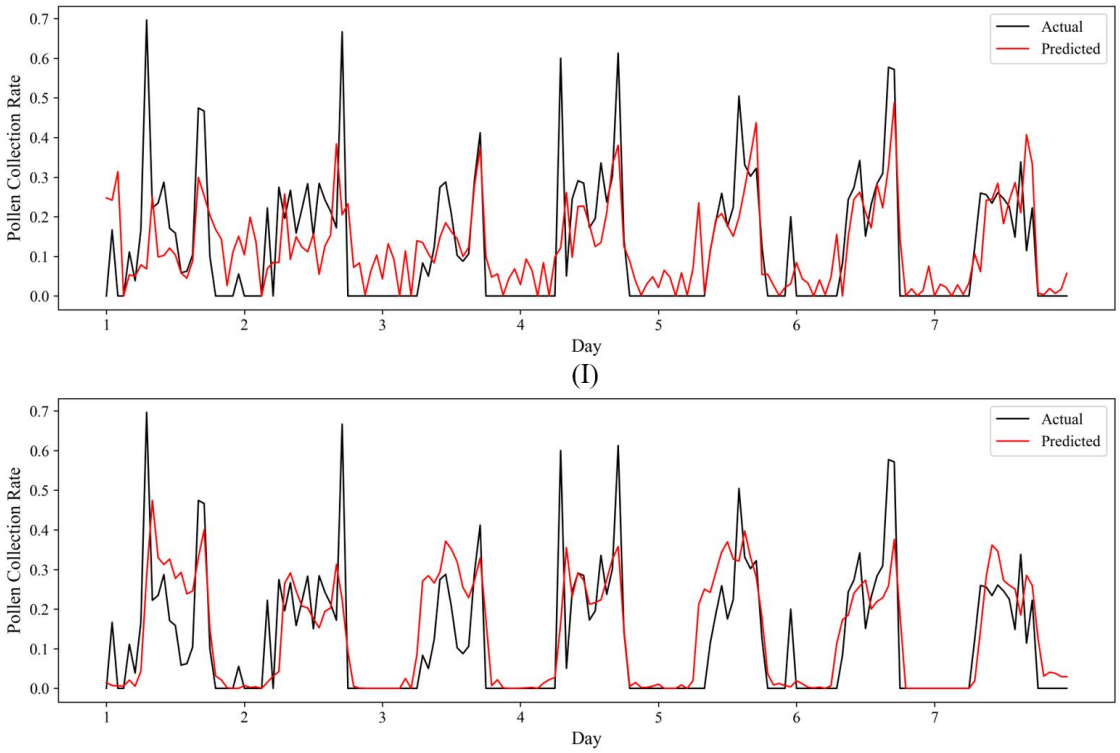
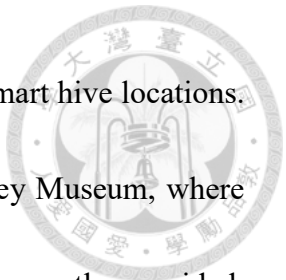


Fig. 4-7 Multivariate prediction results of pollen collection rate using SARIMA and GRU models. (I) SARIMA model; (II) GRU model.

4.5 Cross-Location Model Comparison and Generalization

To better understand how model design and environmental context affect forecasting accuracy, this section compares the performance of univariate and multivariate

doi:10.6342/NTU202503964



SARIMAX and GRU models trained on datasets from two distinct smart hive locations. Table 4-2 shows the results obtained from a hive placed at the Honey Museum, where nearby floral resources were limited and artificial feeding was frequently provided, resulting in higher levels of human intervention. In contrast, Table 4-3 presents results from a hive located in a private residential courtyard, characterized by richer natural forage availability and minimal human disturbance.

Section 4.4.2 shifts the analysis toward generalization performance. Here, multivariate GRU models are independently trained and tested on four different hives, each representing a distinct biological and spatial context. The objective is to examine how well the GRU framework maintains predictive performance across varying colonies and to determine whether model stability persists when applied to different real-world deployment scenarios.

4.5.1 Comparative Analysis of Univariate and Multivariate Models Across Two Hive Locations

To better understand how model design and environmental context affect forecasting accuracy, this section compares the performance of univariate and multivariate SARIMAX and GRU models trained on datasets from two distinct smart hive locations.

doi:10.6342/NTU202503964

Table 4-2 shows the results obtained from a hive placed in a private residential courtyard with minimal human disturbance, no supplemental feeding, and relatively sparse floral availability. In contrast, Table 4-3 presents results from a hive located at the Honey Museum, where nearby floral resources were abundant and artificial feeding was provided regularly.

Across both sites, models were evaluated on three forecasting targets: hive weight, bee traffic, and pollen collection rate. Each model was assessed under both univariate and multivariate configurations, allowing for a systematic comparison of architecture (SARIMAX vs. GRU), variable input (univariate vs. multivariate), and environmental complexity (courtyard vs. museum).

At both locations, the GRU model consistently outperformed SARIMAX in nearly all settings and targets. In the courtyard site (Table 4-2), GRU achieved lower MAE and MAPE (or SMAPE) values for all three targets. For example, in multivariate bee traffic prediction, GRU reduced MAE from 13,503 to 7,113 and MAPE from 35.5% to 19.0%, showing nearly 50% improvement over SARIMAX. Similar trends were observed at the Honey Museum site (Table 4-3), where environmental conditions were more dynamic. Here, GRU again achieved markedly lower error metrics; in particular, pollen collection rate prediction under multivariate GRU reached a SMAPE of 16.6%, compared to 22.0% for SARIMAX.

doi:10.6342/NTU202503964

Comparing the two sites, all models performed slightly better in the courtyard, where environmental noise and foraging complexity were lower. However, the relative advantage of GRU remained stable or even increased in the more complex setting of the Honey Museum, indicating its robustness to external variability. The use of multivariate input further enhanced performance across both model types, but especially for GRU, whose deep learning structure allowed it to better leverage contextual features.

Overall, these results demonstrate that multivariate GRU is the most accurate and reliable approach among the tested configurations. Its consistent advantage across targets and environments highlights its suitability for real-world hive monitoring applications, particularly in complex or highly variable settings.

Table 4-2 Univariate and Multivariate Model Performance – No.4 Hive

Setting	Target Variable	Metric	SARIMA	GRU
Univariate	Hive Weight	MAE	0.167	0.130
		MAPE	0.9	0.7
	Bee Traffic	MAE	15480	7296
		MAPE	40.2	20.1
	Pollen Collection Rate	MAE	0.047	0.047
		SMAPE	23.7	21.7
Multivariate	Hive Weight	MAE	0.149	0.118
		MAPE	0.8	0.6
	Bee Traffic	MAE	13503	7113
		MAPE	35.5	19.0
	Pollen Collection Rate	MAE	0.047	0.042
		SMAPE	23.6	20.9

doi:10.6342/NTU202503964

Table 4-3 Univariate and Multivariate Model Performance – No.2 Hive

Setting	Target Variable	Metric	SARIMA	GRU
Univariate	Hive Weight	MAE	0.266	0.238
		MAPE	1.2	1.1
	Bee Traffic	MAE	5047	2132
		MAPE	54.0	26.3
	Pollen Collection Rate	MAE	0.072	0.070
		SMAPE	25.0	18.8
Multivariate	Hive Weight	MAE	0.246	0.184
		MAPE	1.1	0.9
	Bee Traffic	MAE	3553	1872
		MAPE	41.0	24.6
	Pollen Collection Rate	MAE	0.071	0.069
		SMAPE	22.0	16.6

4.5.2 Cross-Hive Evaluation of Multivariate GRU Models

This section further explores the generalization capability of the multivariate GRU model across four individual hives deployed in two distinct environments. Specifically, Hive No.2 and Hive No.6 were located at the Honey Museum, a site characterized by limited natural forage, frequent supplemental feeding, and higher human disturbance. In contrast, Hive No.3 and Hive No.4 were situated in a private courtyard with abundant floral resources and minimal human interference. These colonies differed not only in location but also in colony strength and behavioral patterns, offering a realistic scenario

doi:10.6342/NTU202503964

to evaluate the robustness of forecasting models under diverse biological and environmental conditions.

Table 4-4 summarizes the prediction performance for three target variables—hive weight, bee traffic, and pollen collection rate—using the multivariate GRU model. All models were independently trained and tested for each hive using the same input features and hyperparameter tuning strategy. Overall, Hive No.4 demonstrated relatively lower prediction errors across all metrics, suggesting that the GRU model performs better in more stable environments. However, Hive No.3, which also resided in a stable setting, showed less accurate predictions. This is largely attributed to a queen loss event during the data collection period, which destabilized the colony. As a result, the beekeeper made several manual interventions, including rearranging the hive structure, leading to greater inconsistencies in the weight data and reduced prediction accuracy. Among the three forecasting targets, hive weight was the most predictable, with MAPE values ranging from 0.6% to 2.1%, reflecting its relatively stable temporal trends. In contrast, bee traffic and pollen collection rate posed greater challenges due to their higher behavioral variability. For instance, Hives No.3 and No.6 exhibited SMAPE values above 30% for pollen collection rate, and bee traffic MAE values were generally high across all hives.

In summary, while the multivariate GRU model demonstrated stable predictive capabilities and reasonable generalization across different hive conditions, its accuracy

doi:10.6342/NTU202503964

was still influenced by the complexity of each variable and environmental dynamics. Variables like hive weight, which follow more regular patterns, are easier to forecast, whereas behavior-driven indicators such as traffic and pollen collection are more susceptible to external and biological fluctuations.

Table 4-4 Multivariate GRU Performance Across Hives

Target Variable	Metric	No.2	No.3	No.4	No.6
Hive Weight	MAE	0.184	0.308	0.118	0.091
	MAPE	0.9	2.1	0.6	0.8
Bee Traffic	MAE	1872	4802	7113	7344
	MAPE	24.6	24.3	19.0	24.3
Pollen Collection Rate	MAE	0.069	0.053	0.042	0.041
	SMAPE	16.7	31.3	20.9	31.3

4.6 SHAP Analysis and Forecast Shifting Detection

4.6.1 Feature Importance in Hive Weight Forecasting

To interpret the model behavior for hive weight forecasting, SHAP-based feature importance was computed for two representative hives under the multivariate GRU prediction setting. The left panel of Fig. 4-8 shows results for Hive No.2, located at the Honey Museum, while the right panel corresponds to Hive No.4, located in a beekeeper's

doi:10.6342/NTU202503964

private backyard. The bar plots illustrate the normalized daily-level SHAP values for each input feature, reflecting their relative contribution to the predicted daily hive weight.

In both hives, historical hive weight emerged as the most dominant feature, contributing over 25% to the prediction. This aligns with expectations, as past weight values inherently contain strong temporal continuity. However, notable differences in the secondary features reflect the contrasting environmental conditions between the two locations.

At Hive No.2 (Honey Museum), the model assigned greater importance to outgoing bee volume and temperature, whereas pollen inflow had a comparatively lower contribution. This result may be influenced by the artificial feeding practices and lower surrounding nectar density at the site, prompting the model to rely more on behavioral indicators and environmental signals for short-term weight fluctuations.

Conversely, Hive No.4 (backyard) showed a higher contribution from pollen inflow and humidity, while the importance of inbee and outbee features decreased slightly. The presence of richer and more stable floral resources at this location likely made pollen-related features more informative. In addition, the increased influence of humidity suggests that ambient moisture may interact more predictably with natural weight dynamics in this less-disturbed environment.

doi:10.6342/NTU202503964

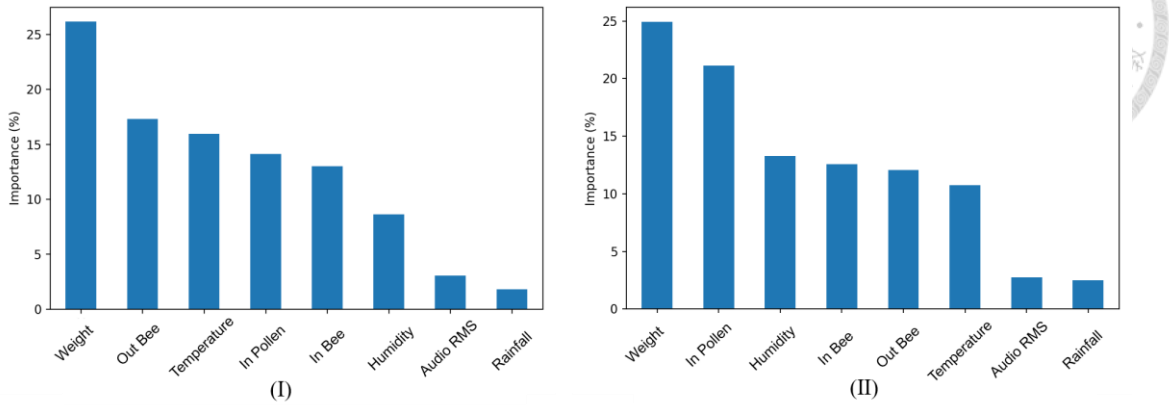


Fig. 4-8 Daily SHAP-based feature importance for hive weight forecasting at (I)Hive No.2 and (II)Hive No.4.

4.6.2 Feature Importance in Bee Traffic Forecasting

This section analyzes how the multivariate GRU model allocates feature importance when forecasting bee in-out traffic, using SHAP-based explanations for Hive No.2 and Hive No.4. Fig. 4-9 displays the normalized daily SHAP values for both sites, revealing the relative contribution of each input variable.

In both cases, the historical traffic volume was the overwhelmingly dominant predictor, contributing over 70% of the model's decision. This indicates the strong autoregressive nature of bee foraging behavior, where past activity levels are highly informative for the next day's forecast.

doi:10.6342/NTU202503964

However, the secondary features differed between sites due to variations in hive conditions and surrounding environments. Hive No.2, located at the Honey Museum, is partially shaded by overhead trees. This natural canopy helps buffer direct sunlight and rainfall, resulting in increased reliance on temperature and humidity as internal environmental cues. Other features such as audio RMS, hive weight, and rainfall showed relatively minor influence, suggesting the microclimate under tree cover played a larger role than broader weather conditions.

In contrast, Hive No.4 is situated in an open backyard environment and is equipped with a shade net to protect against excessive heat and rainfall. The colony was also larger, with a higher number of bees. In this setting, the model assigned slightly more weight to rainfall and temperature, indicating that even with artificial shielding, external weather conditions still influenced bee activity—possibly through changes in ambient climate or group dynamics. The greater bee population in Hive No.4 may have amplified the sensitivity of the colony to subtle environmental shifts, increasing the predictive value of these variables. Overall, while historical traffic data consistently dominated the model's predictions, the influence of environmental features was context-dependent. The model adaptively adjusted its weighting based on hive structure, shading, and colony scale, demonstrating ecological responsiveness in feature attribution.

doi:10.6342/NTU202503964

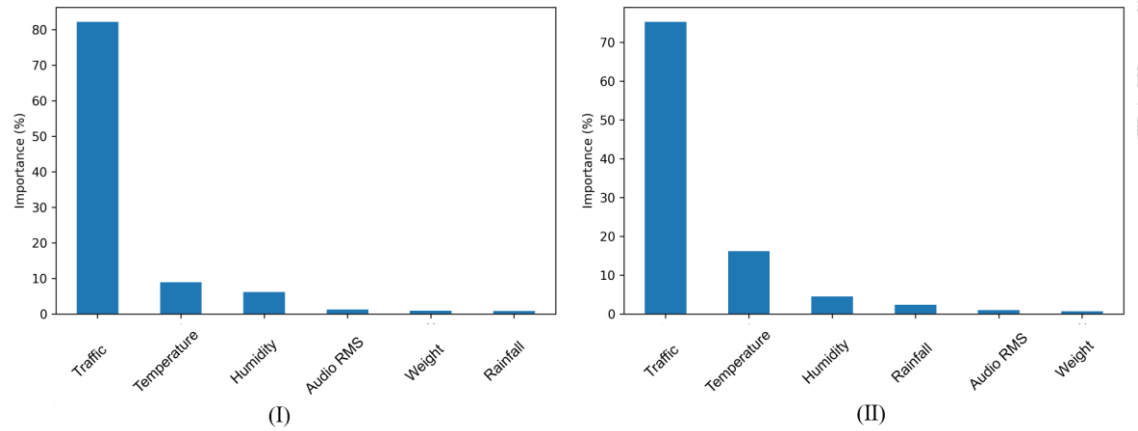


Fig. 4-9 Daily SHAP-based feature importance for bee traffic forecasting at (I)Hive No.2 and (II)Hive No.4

4.6.3 Feature Importance in Pollen Collection Rate Forecasting

This section analyzes the feature contributions for forecasting pollen collection rate, comparing the SHAP-based interpretations between two hives located in distinct environments. Figure 4-9 displays the daily-level SHAP importance for Hive No.2 (left, located at the Honey Museum) and Hive No.4 (right, situated in a private residential courtyard with richer floral resources).

In Hive No.2, the model primarily relied on the target variable itself (pollen collection rate) and the number of outgoing bees, which together contributed over 70% of the total SHAP values. This indicates that under conditions with limited floral resources and potential disruptions due to supplemental feeding or weather variations, the model

doi:10.6342/NTU202503964

depended heavily on past collection behavior and current foraging activity. Climatic features such as temperature, rainfall, and humidity played secondary roles in the model's decision-making process.

In contrast, Hive No.4 exhibited a markedly different contribution profile. Outgoing bee traffic emerged as the most influential factor, with a SHAP value of 39.1%—substantially higher than all other features. This highlights the importance of foraging activity in natural environments with abundant nectar sources and optimal conditions. When colony size is large and floral availability is high, successful foraging becomes strongly linked to the sheer number of bees exiting the hive. Thus, the model's predictions for pollen collection rate in Hive No.4 were more dependent on the "Out bee" feature.

Additionally, rainfall showed a higher contribution in Hive No.4. This could be attributed to the lack of overhead tree cover in this location, where only a shade net was present. As a result, rainwater could directly impact the hive entrance, potentially influencing foraging behavior and making rainfall a more relevant predictor. In contrast, Hive No.2 was sheltered under a tree canopy, reducing the direct effect of precipitation on the hive entrance.

Overall, the GRU model demonstrated strong adaptability in adjusting feature importance based on environmental context. These results confirm not only the model's

doi:10.6342/NTU202503964

forecasting accuracy but also its explainability and robustness across diverse field settings.

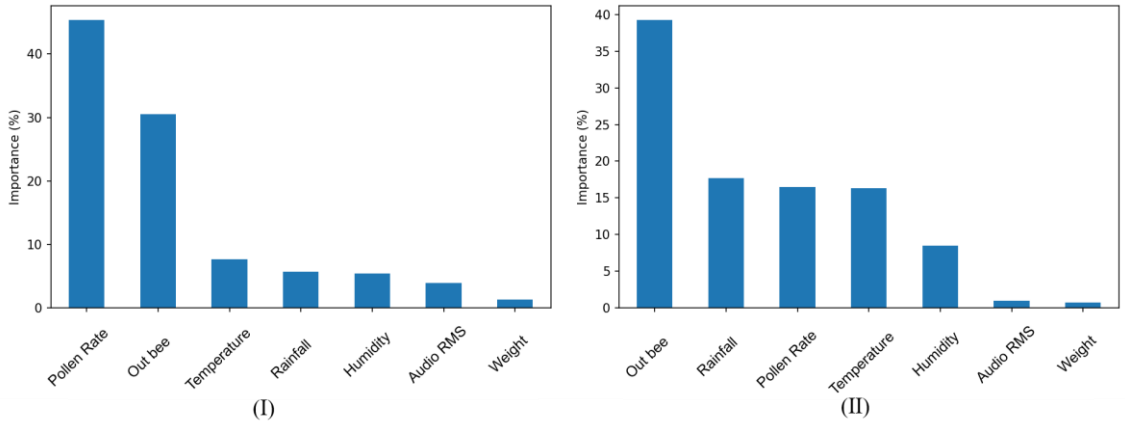


Fig. 4-10 SHAP-based feature importance for pollen collection rate forecasting in (I)Hive No.2 and (II)Hive No.4

4.6.4 Forecast Stability Evaluation

To ensure the robustness of hive weight forecasting in practical deployment scenarios, we conducted a post-hoc analysis to examine two types of undesirable model behaviors: stagnation and shifting. Stagnation refers to forecast sequences that lack sufficient temporal variation, resulting in overly flat or static predictions. Shifting, on the other hand, describes cases where the model fails to generate meaningful forecasts and instead reproduces recent observations, offering little predictive value.

doi:10.6342/NTU202503964

The detection process was based on three diagnostic indicators computed for each forecast window. The first indicator, Range Ratio (RR), measures the amplitude of the predicted sequence and reflects whether the forecast has sufficient variation. The second, Relative Consecutive Difference (RD), evaluates the degree of dynamic responsiveness by quantifying the average change between consecutive prediction steps. The third indicator assesses how the model's MAE compares to that of a naive shifting baseline, where the last observed value is repeated as the forecast.

For both RR and RD, the 5th percentile of all forecast windows was used as the detection threshold. This percentile-based thresholding strategy ensures that only the most stagnant or least dynamic predictions are flagged, while remaining insensitive to outliers. It offers a conservative yet effective standard for diagnosing forecasting degradation, and has been supported in previous literature as a practical alternative to fixed absolute thresholds.

Taking the weight prediction model as an example (Table 4-5), all models produced outputs with RR and RD values exceeding the defined thresholds. For instance, the multivariate GRU model achieved an average RR of 0.0161, which is safely above the 0.0100 threshold, confirming sufficient amplitude. Moreover, the model's MAE was consistently lower than that of the shifting baseline, suggesting that it was not simply mimicking prior values but rather learning meaningful temporal patterns.

doi:10.6342/NTU202503964

These results confirm that the developed models do not suffer from stagnation or shifting. The forecasts exhibit adequate temporal dynamics and maintain predictive novelty across evaluation windows, reinforcing their suitability for long-term beehive monitoring.

Table 4-5 Diagnostics of Stagnation and Shifting in Weight Forecasting

Model	Avg RR	RR 5th pct	Avg RD	RD 5th pct	Naïve	Model	Shifting
SARIMA	0.0056	0.0025	0.0004	0.0001	0.1976	0.1931	0.0366
SARIMAX	0.0129	0.0059	0.0010	0.0005	0.2008	0.1763	0.0366
Univariate GRU	0.0044	0.0008	0.0004	0.0001	0.1838	0.1812	0.0366
Multivariate GRU	0.0161	0.0100	0.0014	0.0010	0.2060	0.1569	0.0366

4.7 Colony Health Status Prediction Results

This section presents the results of colony health status prediction using a GRU-based binary classification model. The model aims to predict the colony status on the eighth day using time-series data from the previous seven days. The health status is categorized as either “Healthy” or “Unhealthy,” based on the Hive Condition Checklist (HCC): a perfect score of 6 is labeled as “Healthy,” while scores between 1 and 5 are categorized as “Unhealthy.”

doi:10.6342/NTU202503964

The GRU model architecture consists of two layers with a hidden size of 64. Each input sequence covers 168 hours (7 days), and the prediction target is the colony status 24 hours ahead. To address class imbalance, a weighted sampling strategy was applied during training. All features were normalized using MinMax scaling. The decision threshold for classification was optimized based on the receiver operating characteristic (ROC) curve. A sliding window approach was used to generate the samples, with the window moving forward one day at a time. Consequently, each sample corresponds to a daily prediction, and the confusion matrix values reflect the number of days that were correctly or incorrectly classified.

Fig. 4-11 Confusion matrices of GRU-based colony health classification for (I)Hive No.2 and (II)Hive No.3.Fig. 4-11and Fig. 4-12 show the confusion matrices for four individual hives. In Fig. 4-11, the left panel corresponds to Hive No.2 with an AUC of 0.82. The model correctly identified 17 abnormal days and 6 healthy days, with only 2 false positives. The right panel represents Hive No.3 with an AUC of 0.81. Although it also detected 17 abnormal days, the model misclassified 3 abnormal days as normal and 2 normal days as anomalies, resulting in slightly lower precision and recall.

doi:10.6342/NTU202503964

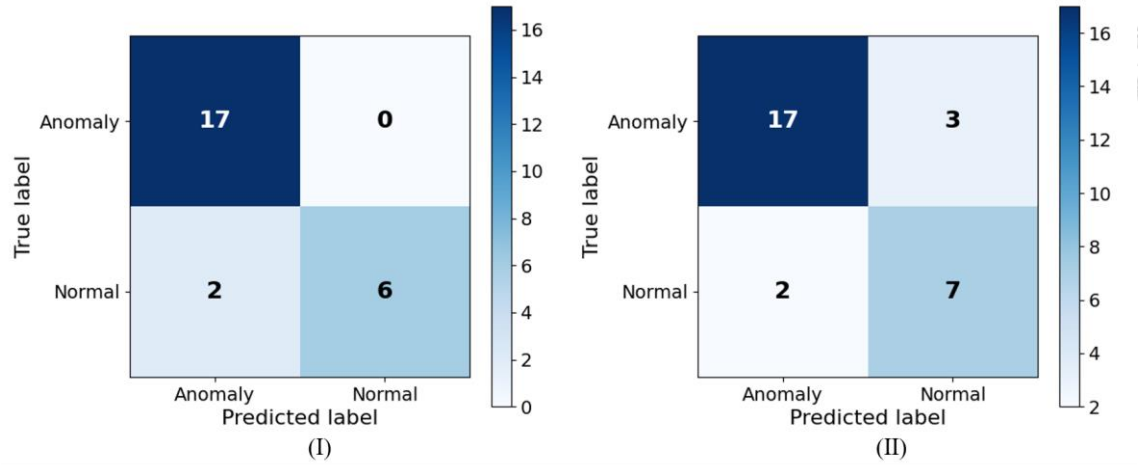


Fig. 4-11 Confusion matrices of GRU-based colony health classification for (I)Hive No.2 and (II)Hive No.3.

In Fig. 4-12, the left panel shows results for Hive No.4, which achieved the highest AUC of 0.98. The model accurately classified 5 abnormal and 9 healthy days, with only 1 false positive. The right panel corresponds to Hive No.6, which yielded an AUC of 0.95. It correctly detected 14 abnormal and 16 healthy days, with only two misclassifications.

doi:10.6342/NTU202503964

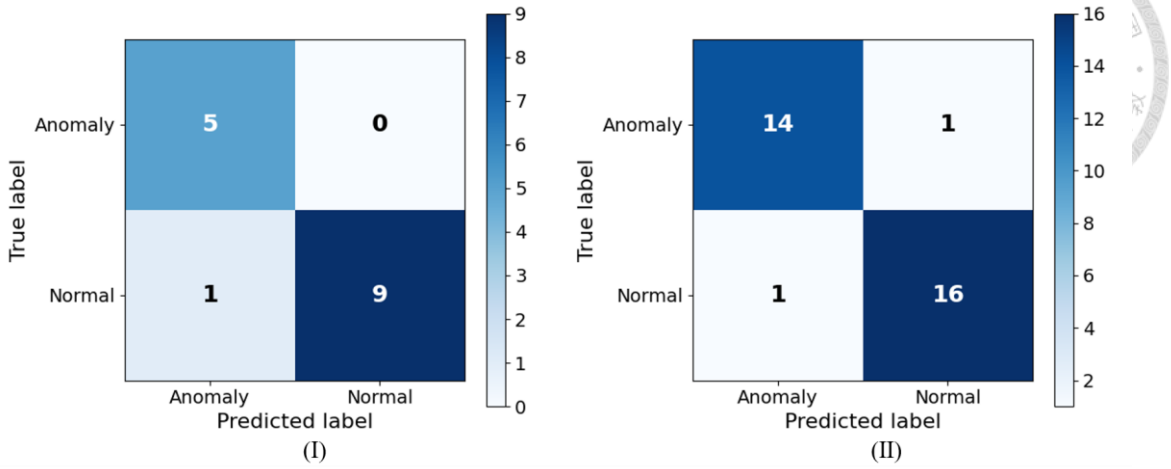


Fig. 4-12 Confusion matrices of GRU-based colony health classification for (I)Hive No.4 and (II)Hive No.6.

Overall, the GRU-based model demonstrated strong and stable predictive performance across all four hives, particularly in identifying abnormal conditions. A detailed summary of accuracy and AUC values for all four hives is provided in Table 4-6. To understand the underlying causes of "Unhealthy" labels between October 2024 and March 2025, we reviewed the HCC records and found that most low scores stemmed from overcrowding. Limited internal hive space restricted colony expansion and food storage, leading to erratic comb construction that interfered with inspections and increased the risk of swarming. In other cases, queen loss contributed to low HCC scores (typically 3–4), as it disrupted colony organization and behavior. Notably, the GRU model successfully captured both types of abnormalities—overcrowding and queen loss—demonstrating its

doi:10.6342/NTU202503964

sensitivity to physiological and environmental changes in colony conditions. This predictive capacity enables timely alerts to beekeepers, prompting inspections and interventions that could mitigate colony degradation or mismanagement.

Table 4-6 Performance Metrics of GRU-based Health Classification Across Individual Hives

	No.2	No.3	No.4	No.6
Accuracy	0.92	0.83	0.93	0.94
AUC	0.82	0.81	0.98	0.95

doi:10.6342/NTU202503964

CHAPTER 5

Conclusions

5.1 Conclusions

This study developed a multivariate time-series forecasting and anomaly detection system tailored for smart beekeeping applications. The system integrates sensor data acquisition, feature selection, GRU and SARIMAX model training, SHAP-based interpretability analysis, and stability diagnostics to achieve real-time colony health monitoring and early-warning capabilities. Data were collected from September 2024 to April 2025, focusing on three key indicators—hive weight, colony traffic, and pollen collection rate—to build multistep forecasting models and evaluate the predictive effectiveness of different feature combinations in practical monitoring scenarios.

Experimental results demonstrated that both GRU and SARIMAX models improved forecasting accuracy when trained with multivariate inputs, with the GRU model showing superior stability and generalizability. Specifically, during cross-hive validation, the multivariate GRU model achieved a Mean Absolute Percentage Error (MAPE) of 0.6%–2.1% for hive weight prediction, 19%–24.6% MAPE for colony traffic, and a Symmetric Mean Absolute Percentage Error (sMAPE) of 16.7%–31.3% for pollen collection rate.

doi:10.6342/NTU202503964

SHAP-based interpretability analysis identified the key contributing features influencing these predictions, and prediction consistency checks confirmed the absence of forecast shifting or stagnation issues. These results validate that the proposed forecasting system delivers stable, interpretable, and quantitatively reliable outputs suitable for field deployment.

For colony health status classification, a binary GRU model was trained to predict the next-day health condition based on the past seven days of sensor data. Labels were derived from beekeeper-provided Healthy Colony Checklist (HCC) forms, with a perfect score of 6 considered "Normal" and any score between 1 and 5 categorized as "Anomaly." The classification model achieved robust predictive performance across multiple hives, with an overall accuracy of 94%. It successfully detected abnormal events such as queen loss and hive overcrowding, enabling timely alerts to assist beekeepers in proactive management and reduce potential colony losses.

While previous research has primarily focused on detecting acute anomalies, a persistent gap remains in addressing early detection of chronic health deterioration. The present system mainly focuses on immediate forecasting and anomaly detection; however, its modular and scalable architecture provides a foundation for future model extensions to incorporate chronic health factors. With additional research and long-term sensor data reflecting chronic stressors, the proposed framework could evolve into a more

doi:10.6342/NTU202503964

comprehensive decision-support tool, empowering beekeepers with proactive, long-term preventive capabilities to enhance colony resilience and support sustainable apiculture and pollination security.

5.2 Suggestions

1. Optimization of Bee and Pollen Image Recognition System

The current system employs a YOLOv3-tiny model for bee traffic and pollen detection, which is outdated and may lack sufficient accuracy. Future work could explore using advanced models such as YOLOv11, YOLOv12, or other lightweight yet high-performance architectures to reduce detection errors and improve data quality.

2. Enhanced Feature Selection via Interpretability Analysis

SHAP analysis in this study has identified features with limited predictive contribution. Future research can leverage this information to refine feature selection, removing irrelevant variables to simplify the model architecture while improving predictive accuracy and computational efficiency.

3. Expansion of Health Status Dataset with Seasonal Diversity

The current health classification dataset spans approximately five months, primarily covering the October–April period. To ensure robustness across diverse conditions, future

doi:10.6342/NTU202503964

studies should collect year-round data encompassing various seasons and colony health states to validate the model's adaptability.

4. Integration of Chronic Pesticide Toxicity Detection

Chronic pesticide poisoning poses a major threat to modern beekeeping. Future research could focus on long-term data collection from pesticide-affected hives and incorporate these data into the health classification model to strengthen its capability in early warning of pesticide-related colony decline.

5. Multimodal Data Fusion and Long-Term Behavioral Modeling

Although the current system integrates multiple sensors, future enhancements could involve fusing multimodal data sources, including in-hive acoustic signals, external floral condition imagery, and weather forecasts. Advanced architectures such as Transformers or hybrid deep learning models could then be employed to capture long-term behavioral patterns, improving chronic disease detection and colony activity forecasting.

doi:10.6342/NTU202503964

References

Abbo, P. M., Kawasaki, J. K., Hamilton, M., Cook, S. C., DeGrandi-Hoffman, G., Li, W. F., Liu, J., & Chen, Y. P. (2017). Effects of Imidacloprid and Varroa destructor on survival and health of European honey bees, Apis mellifera. *Insect science*, *24*(3), 467-477. https://doi.org/10.1111/1744-7917.12335

Abdollahi, M., Giovenazzo, P., & Falk, T. H. (2022). Automated Beehive Acoustics Monitoring: A Comprehensive Review of the Literature and Recommendations for Future Work. *Applied Sciences*, *12*(8).

Abou-Shaara, H., Owayss, A., Ibrahim, Y., & Basuny, N. (2017). A review of impacts of temperature and relative humidity on various activities of honey bees. *Insectes Sociaux*, *64*. https://doi.org/10.1007/s00040-017-0573-8

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics*, *Vol.05No.07*, 14, Article 62189. https://doi.org/10.4236/ojs.2015.57075

Al-Ghamdi, A., Abou-Shaara, H., & Mohamed, A. (2014). Hatching rates and some characteristics of Yemeni and Carniolan honey bee eggs. *JOURNAL OF ENTOMOLOGY AND ZOOLOGY STUDIES*, *2*, 6-10.

Andrijević, N., Urošević, V., Arsić, B., Herceg, D., & Savić, B. (2022). IoT Monitoring and Prediction Modeling of Honeybee Activity with Alarm. *Electronics*, *11*(5).

Barron, A. B. (2015). Death of the bee hive: understanding the failure of an insect society. *Current Opinion in Insect Science*, *10*, 45-50. https://doi.org/10.1016/j.cois.2015.04.004

Becher, M., Scharpenberg, H., & Moritz, R. (2009). Pupal developmental temperature and behavioral specialization of honeybee workers (Apis mellifera L.). *Journal of*

doi:10.6342/NTU202503964

comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology, *195*, 673-679. https://doi.org/10.1007/s00359-009-0442-7

Braga, A. R., Gomes, D. G., Rogers, R., Hassler, E. E., Freitas, B. M., & Cazier, J. A. (2020). A method for mining combined data from in-hive sensors, weather and apiary inspections to forecast the health status of honey bee colonies. *Computers and Electronics in Agriculture*, *169*, 105161. https://doi.org/10.1016/j.compag.2019.105161

Brown, M. J. F., Loosli, R., & Schmid-Hempel, P. (2000). Condition-Dependent Expression of Virulence in a Trypanosome Infecting Bumblebees. *Oikos*, *91*(3), 421-427. http://www.jstor.org/stable/3547517

Bujok, B., Kleinhenz, M., Fuchs, S., & Tautz, J. (2002). Hot spots in the bee hive. *Die Naturwissenschaften*, *89*, 299-301. https://doi.org/10.1007/s00114-002-0338-7

Burrill, R. M., & Dietz, A. (1981). THE RESPONSE OF HONEY BEES TO VARIATIONS IN SOLAR RADIATION AND TEMPERATURE [10.1051/apido:19810402]. *Apidologie*, *12*(4), 319-328. https://doi.org/10.1051/apido:19810402

Cardell-Oliver, R., Anwar, O., Keating, A., Datta, A., & Putrino, G. (2022). WE-Bee: Weight Estimator for Beehives Using Deep Learning.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. https://doi.org/10.48550/arXiv.1406.1078

Clarke, D., & Robert, D. (2018). Predictive modelling of honey bee foraging activity using local weather conditions. *Apidologie*, *49*(3), 386-396. https://doi.org/10.1007/s13592-018-0565-3

Cook, D., Tarlinton, B., McGree, J. M., Blackler, A., & Hauxwell, C. (2022). Temperature

doi:10.6342/NTU202503964

Sensing and Honey Bee Colony Strength. *Journal of Economic Entomology*, *115*(3), 715-723. https://doi.org/10.1093/jee/toac034

Degrandi-Hoffman, G., Spivak, M., & Martin, J. H. (1993). Role of Thermoregulation by Nestmates on the Development Time of Honey Bee (Hymenoptera: Apidae) Queens. *Annals of the Entomological Society of America*, *86*(2), 165-172. https://doi.org/10.1093/aesa/86.2.165

Dsouza, A., & Hegde, S. (2023). HiveLink, an IoT based Smart Bee Hive Monitoring System. *arXiv preprint arXiv:2309.12054*. https://doi.org/10.48550/arXiv.2309.12054

Edwards-Murphy, F., Magno, M., Whelan, P. M., O'Halloran, J., & Popovici, E. M. (2016). b+WSN: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring. *Computers and Electronics in Agriculture*, *124*, 211-219. https://doi.org/10.1016/j.compag.2016.04.008

Ferrari, S., Silva, M., Guarino, M., & Berckmans, D. (2008). Monitoring of swarming sounds in bee hives for early detection of the swarming period. *Computers and Electronics in Agriculture*, *64*(1), 72-77. https://doi.org/10.1016/j.compag.2008.05.010

Flores, J. M., Gil-Lebrero, S., Gamiz, V., Rodriguez, M. I., Ortiz, M., & Quiles Latorre, F. (2018). Effect of the climate change on honey bee colonies in a temperate Mediterranean zone assessed through remote hive weight monitoring system in conjunction with exhaustive colonies assessment. *Science of The Total Environment*, *653*. https://doi.org/10.1016/j.scitotenv.2018.11.004

Gil-Lebrero, S., Navas González, F. J., Gámiz López, V., Quiles Latorre, F. J., & Flores Serrano, J. M. (2020). Regulation of Microclimatic Conditions inside Native Beehives and Its Relationship with Climate in Southern Spain. *Sustainability*, *12*(16).

doi:10.6342/NTU202503964

Goulson, D. (2002). Effects of Introduced Bees on Native Ecosystems. *Annu. Rev. Ecol. Evol. Syst*, *34*, 1-26. https://doi.org/10.1146/annurev.ecolsys.34.011802.132355

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*(11), 2809-2815. https://doi.org/10.1890/02-3114

Gray, A., Brodschneider, R., Adjlane, N., Ballis, A., Brusbardis, V., Charrière, J.-D., Chlebo, R., F. Coffey, M., Cornelissen, B., & Amaro da Costa, C. (2019). Loss rates of honey bee colonies during winter 2017/18 in 36 countries participating in the COLOSS survey, including effects of forage sources. *Journal of Apicultural Research*, *58*(4), 479-485. https://doi.org/10.1080/00218839.2019.1615661

Gregorc, A., & Sampson, B. (2019). Diagnosis of Varroa Mite (Varroa destructor) and Sustainable Control in Honey Bee (Apis mellifera) Colonies—A Review. *Diversity*, *11*(12).

Hadjur, H., Ammar, D., & Lefèvre, L. (2022). Toward an intelligent and efficient beehive: A survey of precision beekeeping systems and services. *Computers and Electronics in Agriculture*, *192*, 106604. https://doi.org/10.1016/j.compag.2021.106604

Henry, M., Béguin, M., Requier, F., Rollin, O., Odoux, J.-F., Aupinel, P., Aptel, J., Tchamitchian, S., & Decourtye, A. (2012). A Common Pesticide Decreases Foraging Success and Survival in Honey Bees. *Science (New York, N.Y.)*, *336*, 348-350. https://doi.org/10.1126/science.1215039

Hepburn, H. R., & Radloff, S. E. (2013). Honeybees of Africa.

Himmer, A. (2008). Die Temperaturverhältnisse bei den sozialen Hymenopteren. *Biological Reviews*, *7*, 224-253. https://doi.org/10.1111/j.1469-185X.1962.tb01042.x

Ho, I.-C., Lai, Y.-J., Chiang, P.-N., Chen, Y.-F., & Lin, T.-T. (2022). Integration of Multiple Sensors for Beehive Health Status Monitoring and Assessment. 2022

doi:10.6342/NTU202503964

ASABE Annual International Meeting,

Imoize, A. L., Odeyemi, S. D., & Adebisi, J. A. (2020). Development of a low-cost wireless bee-hive temperature and sound monitoring system. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *8*(3), 476-485. https://doi.org/10.52549/ijeei.v8i3.2268

Johnson, B. R. (2008). Within-nest temporal polyethism in the honey bee. *Behavioral Ecology and Sociobiology*, *62*, 777-784. https://doi.org/10.1007/s00265-007-0503-2

Khairul Anuar, N. H., Md Yunus, M. A., Baharudin, M. A., Ibrahim, S., & Sahlan, S. (2024). Deep Learning-Driven Predictive Modelling for Optimizing Stingless Beekeeping Yields. *Journal of Computing Research and Innovation*, *9*(2), 244-252. https://doi.org/10.24191/jcrinn.v9i2.451

Kulyukin, V. A., Coster, D., Kulyukin, A. V., Meikle, W., & Weiss, M. (2024). Discrete Time Series Forecasting of Hive Weight, In-Hive Temperature, and Hive Entrance Traffic in Non-Invasive Monitoring of Managed Honey Bee Colonies: Part I. *Sensors*, *24*(19).

Li, L., Lu, C., Hong, W., Zhu, Y., Lu, Y., Wang, Y., Xu, B., & Liu, S. (2022). Analysis of temperature characteristics for overwintering bee colonies based on long-term monitoring data. *Computers and Electronics in Agriculture*, *198*, 107104. https://doi.org/10.1016/j.compag.2022.107104

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Magnier, B., Ekszterowicz, G., Laurent, J., Rival, M., & Pfister, F. (2018). Bee Hive Traffic Monitoring by Tracking Bee Flight Paths. VISIGRAPP,

McLellan, A. R. (1977). Honeybee Colony Weight as an Index of Honey Production and Nectar Flow: A Critical Evaluation. *Journal of Applied Ecology*, *14*(2), 401-408.

doi:10.6342/NTU202503964

https://doi.org/10.2307/2402553

Meikle, W. G., Holst, N., Colin, T., Weiss, M., Carroll, M. J., McFrederick, Q. S., & Barron, A. B. (2018). Using within-day hive weight changes to measure environmental effects on honey bee colonies. *PLoS One*, *13*(5), e0197589. https://doi.org/10.1371/journal.pone.0197589

Meikle, W. G., Rector, B. G., Mercadier, G., & Holst, N. (2008). Within-day variation in continuous hive weight data as a measure of honey bee colony activity [10.1051/apido:2008055]. *Apidologie*, *39*(6), 694-707. https://doi.org/10.1051/apido:2008055

Meikle, W. G., & Weiss, M. (2017). Monitoring colony-level effects of sublethal pesticide exposure on honey bees. *JoVE (Journal of Visualized Experiments)*(129), e56355.

Meikle, W. G., Weiss, M., & Stilwell, A. R. (2016). Monitoring colony phenology using within-day variability in continuous weight and temperature of honey bee hives. *Apidologie*, *47*(1), 1-14. https://doi.org/10.1007/s13592-015-0370-1

Michelsen, A., Kirchner, W. H., Andersen, B. B., & Lindauer, M. (1986). The tooting and quacking vibration signals of honeybee queens: a quantitative analysis. *Journal of Comparative Physiology A*, *158*(5), 605-611. https://doi.org/10.1007/BF00603817

Mishra, M., Bhunia, P., Shubham, S., Sen, R., Bhunia, R., Gulia, J., Mondal, T., Zidane, Z., Saha, T., Chacko, A., Raj, R., S, V., & Kaushal, L. (2023). The Impact of Weather Change on Honey Bee Populations and Disease. *Journal of Advanced Zoology*, *44*, 180-190. https://doi.org/10.17762/jaz.v44iS7.2755

Ngo, T.-N., Rustia, D. J., Yang, E.-C., & Lin, T.-T. (2021). Honey Bee Colony Population Daily Loss Rate Forecasting and an Early Warning Method Using Temporal Convolutional Networks. *Sensors*, *21*(11).

Ngo, T. N., Rustia, D. J. A., Yang, E.-C., & Lin, T.-T. (2021). Automated monitoring and

doi:10.6342/NTU202503964

analyses of honey bee pollen foraging behavior using a deep learning-based imaging system. *Computers and Electronics in Agriculture*, *187*, 106239. https://doi.org/10.1016/j.compag.2021.106239

Ngo, T. N., Wu, K.-C., Yang, E.-C., & Lin, T.-T. (2019). A real-time imaging system for multiple honey bee tracking and activity monitoring. *Computers and Electronics in Agriculture*, *163*, 104841. https://doi.org/10.1016/j.compag.2019.05.050

Nicolson, S. W. (2009). Water homeostasis in bees, with the emphasis on sociality. *Journal of Experimental Biology*, *212*(3), 429-434. https://doi.org/10.1242/jeb.022343

Nolasco, I., & Benetos, E. (2018). To bee or not to bee: Investigating machine learning approaches for beehive sound recognition. *arXiv preprint arXiv:1811.06016*. https://doi.org/ https://doi.org/10.48550/arXiv.1811.06016

Nolasco, I., Terenzi, A., Cecchi, S., Orcioni, S., Bear, H. L., & Benetos, E. (2019). Audio-based identification of beehive states. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Olate-Olave, V., Verde, M., Vallejos, L., Raymonda, L., Cortese, M., & Doorn, M. (2021). Bee Health and Productivity in Apis mellifera, a Consequence of Multiple Factors. *Veterinary Sciences*, *8*, 76. https://doi.org/10.3390/vetsci8050076

Oldroyd, B., & Thompson, G. (2006). Behavioural Genetics of the Honey Bee Apis mellifera. *Advances in insect physiology*, *33*, 1-49. https://doi.org/10.1016/S0065-2806(06)33001-9

Palmer, M. J., Moffat, C., Saranzewa, N., Harvey, J., Wright, G. A., & Connolly, C. N. (2013). Cholinergic pesticides cause mushroom body neuronal inactivation in honeybees. *Nature Communications*, *4*(1), 1634. https://doi.org/10.1038/ncomms2648

Robles-Guerrero, A., Saucedo-Anaya, T., González-Ramírez, E., & De la Rosa-Vargas, J.

doi:10.6342/NTU202503964

I. (2019). Analysis of a multiclass classification problem by Lasso Logistic Regression and Singular Value Decomposition to identify sound patterns in queenless bee colonies. *Computers and Electronics in Agriculture*, *159*, 69-74. https://doi.org/10.1016/j.compag.2019.02.024

Ruvinga, S., Hunter, G., Duran, O., & Nebel, J.-C. (2023). Identifying Queenlessness in Honeybee Hives from Audio Signals Using Machine Learning. *Electronics*, *12*(7).

Schneider, S. S. (1990). Nest characteristics and recruitment behavior of absconding colonies of the African honey bee,Apis mellifera scutellata, in Africa. *Journal of Insect Behavior*, *3*(2), 225-240. https://doi.org/10.1007/BF01417914

Tan Ken, T. K., Bock, F., Fuchs, S., Streit, S., Brockmann, A., & Tautz, J. (2005). Effects of brood temperature on honey bee Apis mellifera wing morphology.

Truong, T. H., Nguyen, H. D., Mai, T. Q. A., Nguyen, H. L., Dang, T. N. M., & Phan, T.-T.-H. (2023). A deep learning-based approach for bee sound identification. *Ecological Informatics*, *78*, 102274. https://doi.org/10.1016/j.ecoinf.2023.102274

Winston, M. L. (1991). *The biology of the honey bee*. Harvard University Press.

Wu, V. (2023, 25-28 July 2023). Development of a Predictive Model of Honey Bee Foraging Activity Under Different Climate Conditions. *2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, IEEE.

Zgank, A. (2019). Bee swarm activity acoustic classification for an IoT-based farm service. *Sensors*, *20*(1), 21. https://doi.org/10.3390/s20010021

Ziegler, C., Ueda, R. M., Sinigaglia, T., Kreimeier, F., & Souza, A. M. (2022). Correlation of Climatic Factors with the Weight of an Apis mellifera Beehive. *Sustainability*, *14*(9).

doi:10.6342/NTU202503964