Smart Beehive Monitoring for Remote Regions

Omar Anwar

THE UNIVERSITY OF WESTERN AUSTRALIA

This thesis is presented for the degree of
Doctor of Philosophy of Electronic Engineering of
The University of Western Australia,
July, 2022

Dedicated to the one who endured all the stings.

Thesis Declaration

I, Omar Anwar, certify that:

This thesis has been substantially accomplished during enrollment in the degree.

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

The work described in this thesis was funded by The Cooperative Research Centre for Honey Bee Products, by The University of Western Australia through SIRF scholarship, and by The Department of Jobs, Tourism, Science and Innovation, Govt. of WA through Science Industry PhD Fellowship.

This thesis contains published work and/or work prepared for publication, some of which has been co-authored.

Signature: [redacted]

Date: 13/07/2022

Abstract

Honeybees play a vital role in sustaining our agricultural economy and maintaining the ecosystem. A healthy and well spread bee population is crucial for better pollination of local crops as well as non-agricultural flora. The decline in global bee population and increased instances of Colony Collapse Disorder (CCD) have drawn attention of researchers all over the world. Recent technological advancements have impacted the bee-keeping industry in numerous ways, and electronic beehive monitoring has significantly improved over the past few years. Monitoring systems have been developed to observe temperature, humidity and acoustics inside the hive, overall weight of the hive and outgoing/incoming bee traffic to gauge the health of beehives. These monitoring systems aided by various wireless communication technologies make it possible for the beekeepers to monitor a large number of hives continuously, simultaneously, from a distance, and only intervene when required.

The most important characteristic of a monitoring system is the set of parameters used for monitoring. Each commercially available solution makes use of its own set of parameters to determine the health of bees. Most of the research carried out in this area focuses on a small set of two to three sensors in each study, rather than examining a bigger set for its collective usefulness. For communication, the monitoring systems rely on either 3G/4G or WiFi networks which are not accessible everywhere, or on satellite communication which can be very expensive. Despite having a high price tag, most of the monitoring systems provide beekeepers with just the raw data from sensors without any analysis on bee health. Proposed systems in the literature have also not been able to make the most of deep learning algorithms, mostly because the data used for training is collected over a short period of time, and from hives with little geographic diversity. Use of such small datasets with limited variations often leads to inconclusive and unreliable results. Beekeepers, in particular from Australia, have not been able to take full advantage of these electronic monitoring systems because of the aforementioned limitations. The vast landscape with no cellular coverage, and the high associated costs of using such monitoring systems are the major challenges faced by the local honeybee industry.

This work addresses the design and development of a beehive monitoring system capable of long range communication with low power consumption. Appropriate sensors for the proposed system are selected after an extensive review of literature. This selection is based on the relevance of sensor with bee health/activity, suitability for long distance transmission over low capacity channels, and optimal use of power. Extraction of appropriate features from sensor data is the key requirement for remote deployment. Different experiments were performed to evaluate various sensors and their features for their importance, and viability for hive deployment. A total of eight sensor systems were deployed in multiple hives, at different locations, and in varying environmental conditions over a 12 month period. During these deployments, Narrow Band Internet of Things (NB-IoT) was thoroughly tested for its communication feasibility from remote sites. Based on the findings, use of NB-IoT is proposed for low cost and reliable communication from remote beehives. The design of this system has also been made available for other researches to use and improve upon.

The aim of sensor deployments in this study is not only to test different sensors and communication for beehive monitoring, but also to build a quality sensor dataset from beehives deployed at different sites. Beehive data collection is a slow process based on the natural activity and life cycle of honeybees. The harsh environment of remote sites, sensor failures, and communication issues make it a very challenging task. A dataset of 2,170 days of beehive sensor data, weather data, and seasonal information has been collected during this study. The resolution of 144 data points per day in this dataset provides a good picture of daily bee activity, and facilitates the use of machine learning in beehive health monitoring. Random forests are used to evaluate the contribution of different sensors in this dataset, as well as of the performance of monitoring system.

Daily hive weight variations are a crucial aspect of hive health and bee activity. Hive weight is affected by multiple complex internal and external factors. Traditionally, an expensive and difficult to deploy weighing scale is used to monitor the hive weight. This is the first work to propose the use of machine learning for beehive weight estimation. Latest machine learning algorithms were tested for their suitability with beehive monitoring and weight estimation, and modified to make most of the information available in beehive sensor data. This work presents two deep learning models for beehive weight estimation, *WE-Bee* and *Apis-Prime*. The features for training and testing these models were selected after an in-depth study of bee behaviour, and the impact of environment on bee foraging activity. *WE-Bee* uses Long Short Term Memory (LSTM) encoders and decoders with temporal attention, whereas *Apis-Prime* uses self-attention encoders for the same task. These models were tested on sensor systems and hives which were not part of the training set. The promising results validate the good performance of both networks for unseen data. The hives used for the data collection were allowed their natural variations in colony strengths and forager activity, and were moved to sites at a significant distance from each other to collect geographically diverse data. The diversity of the training data played a significant role in the quality of estimations. Use of these machine learning models has the potential to eliminate expensive beehive weighing scales, and reduce the cost of beehive monitoring systems by more than half.

Evaluation of sensors and contribution of features towards a specific task is important for improving and fine-tuning the design of monitoring systems. This work proposes the use of attention weights of self-attention encoders to evaluate sensors and sensor features, as well as to identify the times of day when sensor data carries most information. This enables a significant reduction in the number of features used for estimation. The equally good results of weight estimation with reduced features signify the usefulness of self-attention encoders for feature selection. These findings not only help assess the bee health/activity remotely, but also significantly reduce the monitoring costs. The estimates about hive weight variations using machine learning provide the beekeepers with important information about the hive without using an expensive weighing scale. The promising weight estimates indicate that the proposed system collects important data from the hive, which can also be utilized for a variety of beehive health monitoring tasks.

Acknowledgments

And your Lord inspired the bees: "Make your homes in the mountains, the trees, and in what people construct, and feed from the flower of any fruit you please and follow the ways your Lord has made easy for you." From their bellies comes forth liquid of varying colours, in which there is healing for people. Surely in this is a sign for those who reflect.

Quran, Chapter 16 (The Bee), Verse 68-69

I thank Allah Almighty for giving me the ability to think, and to deeply reflect upon one of His fascinating creation, *the honeybee*. Working with honeybees has been a beautiful experience, one which has not only shaped the last three and a half years of my life but will continue to impact what I do in the future. I would like to thank my parents, who wanted me to do a PhD ever since I was in pre-school, and kept pushing me during my 10 year break from studies after my Masters degree. It is only because of there motivation, support and prayers that I have reached this milestone.

My PhD supervisors made this journey a memorable one. It was because of their constant guidance that I never felt lost. They had a plan for me when I didn't, and that made a huge difference. Someday if I get a chance to supervise my own PhD students, I hope I will be able to do a job as good as they have done. I also want to express my gratitude for the UWA staff, specifically those at CSSE, EECE, GRS and UWA finance for all the administrative support they have provided.

A big thank you to my wife for her continuous involvement in every step of this journey. To my daughter who can still think of questions about honeybees which I cannot answer. To my brothers, sisters and in-laws who never doubted my abilities. And to all other extended family members for their constant encouragement. I know how proud you all are of my small accomplishments during this journey. Thank you for all the faith you have shown in me and I hope to continue to be a reason for your happiness. Also to all the friends in Perth, all these years with no possibility of international travel, you have been a family away from home.

I would like to thank The Cooperative Research Centre for Honey Bee Products for all the financial as well as non-financial support they have provided to carry out this research. And my thanks to The University of Western Australia for providing me a Scholarship for International Research Fees, and to the Department of Jobs, Tourism, Science and Innovation, Govt. of WA for supporting this project through Science Industry

PhD fellowship. I am also thankful to all the beekeepers, who have assisted with the experiments by helping to test the monitoring systems, collect hive sensor data and perform hive health assessments.

Authorship Declaration: Co-Authored Publications

This thesis contains work that has been published and/or prepared for publication. My contribution in all the research publications included in this thesis is 85% or above. I envisioned the research ideas, refined and implemented them, thoroughly validated them through experimentation and drafted the papers. My co-authors helped me through useful discussions and feedback, and also reviewed my papers to improve the quality of writing and analysis.

Publications Included in This Thesis

1- Anwar, O., Keating, A., Cardell-Oliver, R., Datta, A., & Putrino, G. (2022, February).
WE-Bee: Weight Estimator for Beehives Using Deep Learning.
Accepted in AAAI Conference on Artificial Intelligence 2022: 1st International Workshop on Practical Deep Learning in the Wild.
Available online `https://practical-dl.github.io/long_paper/18.pdf`
Location in thesis: Chapter 4
Student contribution to work: More than 85%
Co-author signatures and dates

Adrian Keating: [redacted]

Rachel Cardell-Oliver: [redacted]

Amitava Datta: [redacted] 7/12/2022

Gino Putrino: [redacted] 7/12/2022

2- **Design and development of low-power, long-range data acquisition system for beehives - BeeDAS**

Authors: Anwar, O., Keating, A., Cardell-Oliver, R., Datta, A., & Putrino, G.

Submission under review in "Computers and Electronics in Agriculture - Journals | Elsevier"

Location in thesis: Chapter 3

Student contribution to work: More than 85%

Co-author signatures and dates

Adrian Keating: [redacted]/12/2022

Rachel Cardell-Oliver: [redacted]22

Amitava Datta: [redacted] 7/12/2022

Gino Putrino: [redacted] 7/12/2022

3- **Apis-Prime: Self-Attention Encoders for Beehive Weight Estimation and System Evaluation**

Authors: Anwar, O., Keating, A., Cardell-Oliver, R., Datta, A., & Putrino, G.

Submission under review in "Applied Soft Computing - Journals | Elsevier"

Location in thesis: Chapter 5

Student contribution to work: More than 85%

Co-author signatures and dates

Adrian Keating: [redacted] 7/12/2022

Rachel Cardell-Oliver: [redacted]22

Amitava Datta: [redacted] 7/12/2022

Gino Putrino: [redacted]/12/2022

Student signature: [redacted]

Date: 7/12/2022

Principal and Coordinating Supervisor Signature

I, Adrian Keating certify that the student statements regarding their contribution to the works presented in this thesis are correct.

Principal and coordinating supervisor signature: [redacted]

Date: 7/12/2022

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discusses the vital role of honeybees and the importance of Beehive Monitoring Systems (BMS). The use of BMS in Australia is also discussed with a focus on the advantages they bring and the major issues associated with their usage. Based on these issues, three research questions are identified. The last section outlines the layout of the thesis chapters that follow.

1.1 Significance of Honeybees

Food is one of the fundamental physiological human needs and is the basic requirement for humans to survive [1]. All the other needs are secondary in nature until the basic human physiological needs are met. With the constant increase in global human population, it is a necessity to expand the agricultural land as well as to enhance the agricultural productivity. Cross-pollination is a crucial factor that determines the agricultural productivity and directly impacts the quantitative as well as qualitative outcome of around 70% of world's crops [2]. It is estimated that around 50% of all the cross pollination is carried out by honeybees [3], which accounts for more than 30% of global food production. In Australia alone, the crops such as *almonds, apples, pears* and *cherries* are totally dependent on European honeybees or *Apis Mellifera*[1] [4] for pollination, and crops such as *orange, plum, apricot, soybean*, and *canola* partially benefit from the pollination by honeybees. Honeybees are also vital for the pollination of *alfalfa*, a widely used nutrition-rich livestock feed [5]. Agricultural farmers are well aware of the impact of bee pollination and often purchase their own honeybee hives, or pay beekeepers to provide pollination services to their crops/orchids during flowering season to maximize the pollination and production.

The value created for horticulture and the Australian economy from paid pollination services has been estimated above AU$5 billion annually, with some estimates exceeding AU$14 billion [6]. Whereas gross value of honey and beeswax production was estimated at AU$147 million in 2019 [6]. Taking these figures into account, honey is just a by-

[1] Throughout this thesis, bee(s) or honeybee(s) refer to *Apis Mellifera*, unless otherwise specified.

product of what honeybees actually contribute towards the Australian economy. The impact of honeybees on local ecology is far greater than the economy. A wide range of non-agricultural flora depends on native honeybees for their survival, as the majority of pollination in forests is carried out by native bees. This flora then provides food and refuge to a wide range of insects, birds and animals [7]. The honeybees have a strong impact on the quality, quantity and the variety of food available to us. Moreover, the influence of honeybees on the survival of our ecological system is far greater than what is normally perceived.

1.2 Importance of Beehive Monitoring Systems

Figure 1.1: A hive frame with honeybees, some capped honey and a sensor board of designed beehive monitoring system on one side. The sensor board is covered using shade cloth to prevent bees from making direct contact with the sensors. Image captured in November 2020 at Capel WA.

In order to cope with the ever increasing need of agriculture-based food for humans and domestic livestock, the global bee population must grow as well. Honeybees are extremely social and they need a good hive to thrive [8]. A well maintained beehive or colony is fundamental for healthy bee population. Honeybee colonies are often considered as super-organisms, where each bee plays its role in the survival of the colony [9]. Unfortunately, over the last three decades, unusual high rates of decline in the bee colonies are reported globally [10]. The major causes for this decline are the widespread use of pesticides in industrial agriculture, parasites, and climate change [11]. The use of strong pesticides intended to protect crops, often result in the decline of bee population, making it a counterproductive measure for agricultural growth. The global climate change and resulting temperature variations, as well as bush fires also add to the problem, as honeybee colonies find it hard to adjust to the weather extremes.

Figure 1.2 illustrates the life cycle of a worker honeybee and the impact of environment on this cycle. Queen is the single most important bee in a hive, and her only job in 5 to 7 years of life is to breed. Eggs are laid by the queen in hive cells, and each egg passes through larva and pupal stages, before emerging out of the cell as a young bee, also known as a hive bee. This cycle from egg to larva, larva to pupa, and then to hive bee takes three weeks and during this time the juvenile are referred to as brood. For the final brood stage, the larva

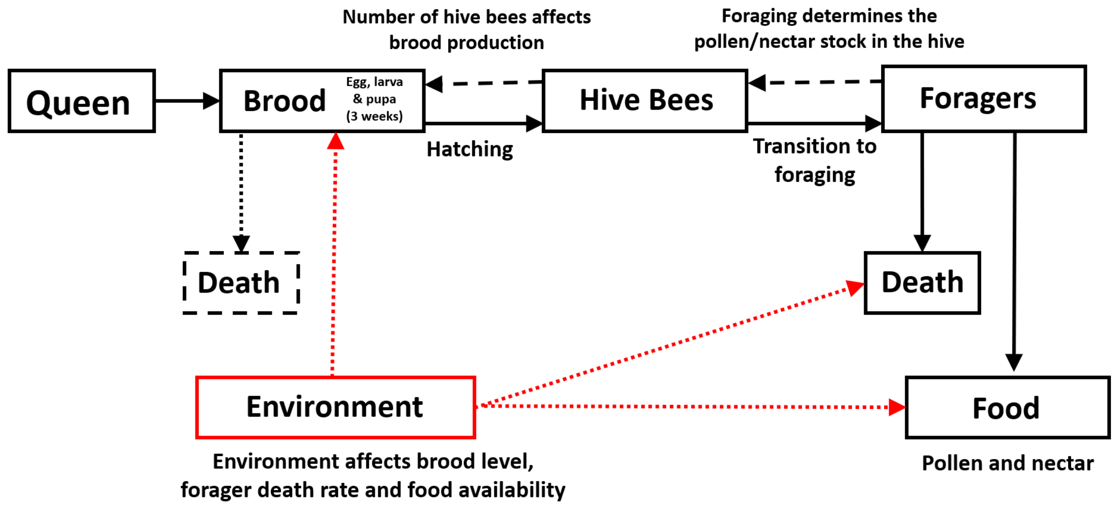

Figure 1.2: Population dynamics of worker honeybees [12].

spins a cocoon around it to pupate, and hive bees cap the cell. This final brood stage is commonly referred to as capped brood. Brood rearing is the primary responsibility of the hive bees, and the balance between the numbers of hive bees and brood is critical. Hive bees transition to forager bees after staying in the hive for about three to four weeks. The forager bees gather the food for the hive, which comprises honey and pollen. After about a couple of weeks of foraging, the foragers die their natural death. In a healthy beehive, the death rates at other stages of the life cycle of bees are insignificant. The balance between forager and hive bees is also maintained through social inhibition. Foragers are the only bees in the colony with direct access to food, illustrated with a solid arrow. The food they bring back is used to feed larvae, queen, worker bees, and is also stored for use in the winter season. Figure 1.3 shows stored pollen, nectar, as well as brood on a single hive frame. The environment through the effects of seasonal variations, climate and weather affects the brood production [12]. Excessive use of pesticides can increase the forager death rates and decrease the food availability in the hive.

In most cases, a colony in distress shows signs of weakness over a period of time before eventually collapsing [13]. If a beekeeper is able to identify a colony in distress and intervene in a timely and adequate manner, there is a good chance that the colony will recover and survive. In a case where the colony is infested with a disease which is hard to recover from, the beekeeper can isolate this colony from the rest of the colonies to stop any further spread of the disease. Either way, timely intervention from the beekeeper is the key. However, this timely intervention is not always possible because of multiple factors:

1. Scale of operations: When honeybees are used in large scale operations for honey production or pollination services, the number of hives at a single site can reach up to several hundreds. This requires a lot of man power to manually inspect the hives, which in itself is a laborious and time consuming process.

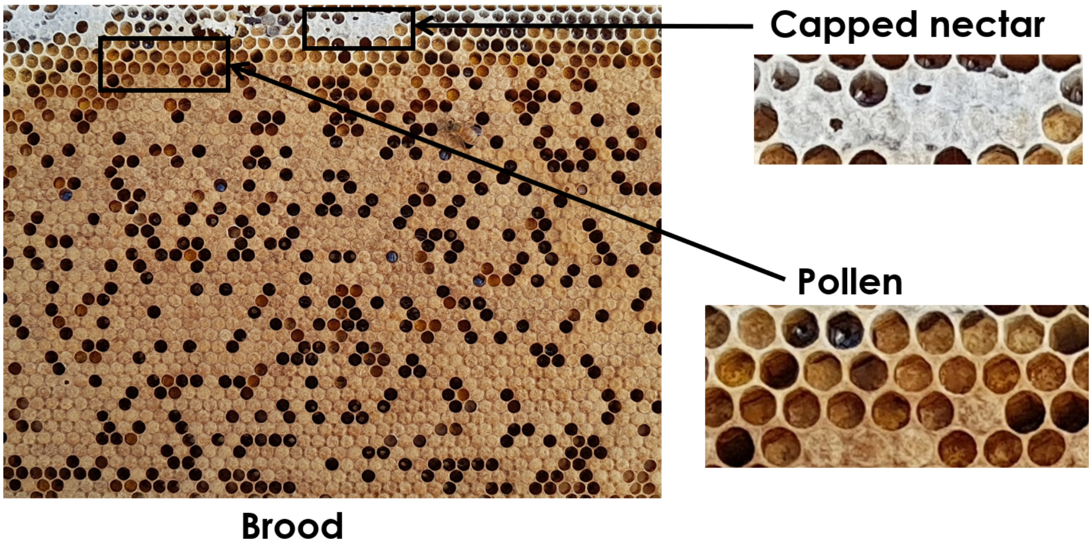

Figure 1.3: A frame with capped nectar/honey (white cells), pollen (yellow cells), and brood (brown cells). Image captured in May 2019 at UWA Campus.

2. Remote location of hive sites: In order to maximize the honey production, hives are often moved from one location to another, either to ensure availability of nectar near the hive site, or to avoid extreme weather conditions. Pollination services also require hives to be moved from one site to another. These sites at times can be hundreds of kilometres away from inhabitable areas, and because of logistics, keeping a manual check on hives can be a difficult and costly process.

3. Minimizing the interference: Honeybees do not like frequent interruptions caused by hive inspections, as this disrupts their hive activity. Each time a hive is opened for inspection [14], it sets back bees on their progress. After the inspection, bees have to first fix the damage caused to their hive wax structure before they can get back to gathering pollen and nectar. This forces the beekeepers to keep their interventions to a reasonable level. On one hand minimal intervention allows the honeybees to maximize their productivity, but also increases the chances of beekeepers missing out on observing any signs of disease or weakness in the colony.

Beehive monitoring systems offer a solution to all three major issues discussed above. Figure 1.4 shows the basic flow of information in a remote beehive monitoring system. BMS use electronic sensors to gather bee health related data from the hives, use a communication medium to transmit this data to a storage. This data is pro-

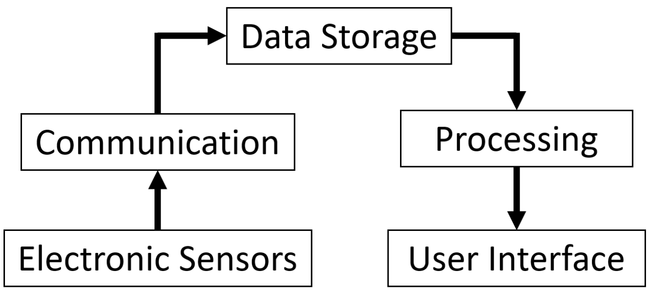

Figure 1.4: Flow of information from sensors to a beekeeper in a remote beehive monitoring system.

cessed to extract information, and beekeeper can access this information using an interface. Figure 1.1 shows an image of a hive frame with designed BMS, captured while beekeeper was inspecting the hive during a test deployment. BMS can be deployed in numbers, thus reducing the human dependency to keep a manual check on the hives. Decision making based on the sensor data can provide an automated alert to beekeeper when intervention is needed. These systems can be used in remote areas and can constantly monitor the state of the hive without the need of beekeeper to open the hives for inspection on regular basis.

1.3 Beehive Monitoring Systems in Australia

One of the biggest problems faced by Australian beekeepers in the monitoring of beehives is the vast agricultural landscape in Australia. Historically, beehive sites were mostly located in close proximity to human settlements. But over last several decades, the area and spread of beehive sites has increased many folds. Unlike regular crops where farmers can arrange for the necessities such as pesticides and fertilizers to make cultivation fruitful in the area of their choosing, beekeepers have to relocate hives to sites where conditions are favourable [15]. This makes beekeeping a challenge, especially if the bee colonies have to be spread over a huge area to ensure a good supply of pollen and nectar for each bee colony. Human monitoring of bee hives on regular basis means putting in a lot of labour, which impacts the cost-effectiveness of honeybee products. In contrast, decreasing the frequency of monitoring puts the health of bee population at risk which is also not desirable.

To elaborate on this using an example, majority of the commercial beekeepers in Perth region move their hives to northern sites during the winter, and towards south during the summer. Beekeepers Nature Reserve is one of the favoured beekeepers site in the north and is more than 200 km away from Perth city, with a travel time of approximately 3 hours. A visit to hive site in this reserve requires a total of 6 hours of travel each day, and many simultaneous days of visits are required to inspect 100s of hives. If beekeepers want to minimise this travel, they arrange for an accommodation in the nearest town for several days while they complete the hive inspections. Multiple beekeepers are required for such inspections, and paying for their wages, accommodation and food can be costly, especially if the inspection is just to check upon bee health. Human involvement is necessary for operations such as replacing the queens and extraction of honey. Whereas regular health inspections can be easily replaced by electronic monitoring systems.

One of the advantages of BMS is that they can be deployed in *remote areas* to monitor hives spread over a large area with minimal human resources. However, for real time monitoring, the majority of such systems rely on either 3G/4G cellular networks [16, 17] or satellite communication [18]. Beekeepers in urban areas make use of WiFi [19] as well as Bluetooth [20], but this is not a practical solution for the majority of beekeepers who operate in rural areas. In Australia, the problem is a bit more complicated as cellular coverage is

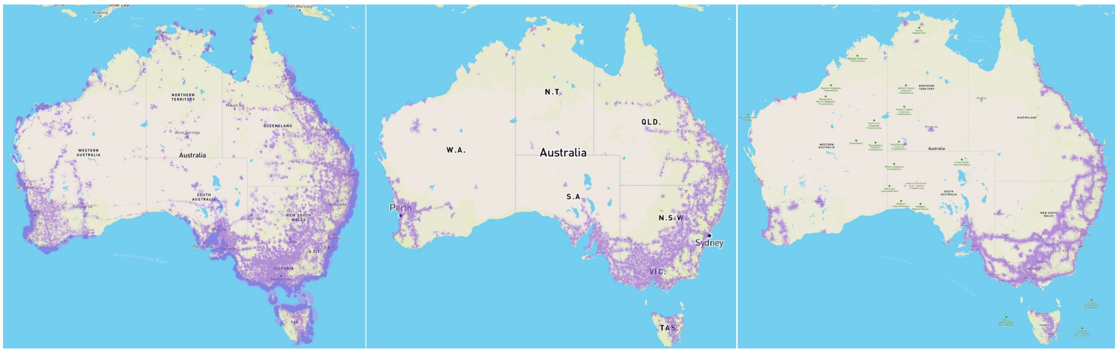

Figure 1.5: Coverage map of 3G/4G network in Australia for Telstra (left), Optus (middle) and Vodafone (right). [21]

available mostly in populated areas as seen in Figure 1.5. Beekeepers move their hives from one site to another based on the availability of nectar and favourable weather conditions. If the site does not have cellular coverage, an alternate solution is to use satellite-based BMS, which are significantly expensive to operate and not always feasible.

Minimizing physical interventions is another key advantage offered by beehive monitoring systems. This is only possible if the monitoring systems are *energy efficient*, thus able to run on battery power for longer periods of time. This is highly dependent on the mode of communication employed by the monitoring systems, as communication can be a very power hungry process, especially if sensor data is to be transmitted frequently. None of the commercially available monitoring systems [19, 20, 18] use the latest *long-range* and *low-power* communication technologies such as LoRaWAN [22] or NB-IoT [23]. These communication technologies are especially designed for low data-rate and long-range communication needs. Remote beehive monitoring systems fit this profile very well, and using such communication systems can significantly reduce the power consumption, thus increasing the duration of battery charge cycle of the BMS.

However, the amount of sensor data generated by current monitoring systems is not appropriate for use with LoRaWAN and NB-IoT, which are both designed for low data-rate communication. Beekeepers in Australia need monitoring systems which are capable of long-range and low-power data transmission. BMS with long range capabilities will not only allow the beekeepers to move their hives further into the remote forests but also reduce the need of frequent visits to hive sites to replace the batteries.

Another problem faced by beekeepers is the poor *cost-effectiveness* of monitoring systems. This becomes a major concern for large scale deployment, as equipping each hive with a monitoring system can be very expensive. Moreover, the beekeepers have to find a balance between the number of monitoring systems they can deploy and the profits they can generate. Apart from the initial cost of purchasing the monitoring systems, there is often an annual subscription fee for unrestricted access to the sensor data or analytical tools. The high costs associated with electronic monitoring of beehives are a big factor for the majority of beekeepers who decide against the use of monitoring systems.

Beehive monitoring systems have not been able to capitalize on the recent advances in the field of machine learning. There is a significant potential of improving the accuracy of predictions about bee health/activity, honey productivity, and bee diseases using machine learning. Soft sensing can make use of machine learning to reduce the cost of monitoring systems. The electronic sensors used in the monitoring systems can also be evaluated for their contribution for specific tasks using machine learning. This is an area that needs more focus of researchers working in the field of electronic beehive monitoring.

The challenges faced by beekeepers in Australia show that issues related with beehive monitoring are complex, diverse and also region specific [24]. The commercially available monitoring systems are yet to come up with a widely acceptable solution. Based on problems faced by beekeepers in general, and the communication problems specific to Australian beekeeping industry, this work aims to answer following three research questions.

- Which sensors should be used to design a low-power and long-range beehive monitoring system for remote regions?
- Is it possible to reduce the system design cost by using soft sensor prediction to replace expensive/difficult to use sensors such as weighing scale?
- Can machine learning algorithms assist in the selection of sensors for specific tasks, and help fine tune the design of beehive monitoring systems?

1.4 Thesis Contributions

The primary focus of this thesis is the design of a cost effective, long range electronic sensor system for remote beehive monitoring. The three research questions detailed above focus on achieving this by improving different aspects of the system design. To investigate the first research question, an in-depth analysis of sensors used in beehive monitoring has been conducted to guide the choice of the best sensors for the proposed system. While one or more sensors can be used for a specific task, not all of them may be appropriate for practical deployment given considerations of the power consumption, size, cost, amount of data generated, or computational complexities involved. Machine learning applied to this design problem required significant consideration within this thesis but is not in itself, the primary focus. Rather it is used as a tool to help find answers to these research questions, leveraging the thesis contributions in the area of sensor integration into the hive.

To this end, the first research question is used to evaluate the contribution of different sensors that have been used in the design. Use of machine learning is not very common in this application, but this work shows that it is a very effective approach to benchmark the performance of different sensors in the system for a given task. The main task it is applied to in this work is to estimate the daily weight change of the hive, which as will be argued, is a difficult and expensive parameter to measure. Different sensors capture varying amount of information contributing towards this daily weight change. Knowledge gaps filled in this work include determination of the correct selection of sensors combined with prudent application of machine learning to understand the impact in minimizing the data generated by the system. Data minimization is an essential requirement to employ NB-IoT or LoRaWAN for long range communication.

Considering the second research question, different methods are explored to reduce the system design cost. Soft sensing using machine learning has shown significant potential and this work shows good accuracy associated with this technique to estimate the hive weight. Importantly, it is the first work to use soft sensing for beehive weight estimation. The first research question uses the daily net change of hive weight as a benchmark, whereas in the second question these estimates are generated as a time series of the weight variations of hive throughout the day. The hive weight variations within a day carry very useful information for the beekeepers and help them access the bee activity in great detail.

To address the third research question, we explore the use of machine learning not only for soft sensing the hive weight, but also to identify the major contributors for these daily weight estimations. This evaluation is not just limited to the contribution by different sensors and different features within sensor data, but also across different times for the day. The time based attention maps generated using machine learning can be used to devise an optimal data collection schedule from the hives. These maps also show the potential of machine learning in studying the activity of honeybees, and the impact of different environmental and weather conditions on the bee activity. The contribution map of different sensors can be used to eliminate less important sensors from the design, as well as improve the feature extraction process by focusing more on features with higher contribution. It is hoped this will reduce the system design cost, minimize the data generated by the system, as well as optimize the power consumption. Such explainability provided by machine learning is a means to advance the electronic system design, improve our comprehension of the bees, as well as help beekeepers better understand and adapt this technology to their needs.

1.5 Thesis Layout

Beehive health monitoring is a multi-disciplinary area and the researchers working in this area have proposed various solutions. Most of these solutions target different aspects of monitoring, based on the area of interest and expertise of the researchers. Chapter 2 examines a range of parameters that impact the electronic monitoring of honeybees, and a wide variety of beehive monitoring systems proposed in the literature. Based on the analysis of proposed monitoring systems, and some commercial beehive monitoring systems, three research questions are identified. To address the core issues related to the system design and sensor datasets, a new system design for remote monitoring is proposed. Chapter 3 discusses the design and development of this monitoring system in detail. The emphasis of this design is on low power consumption and long range transmission to facilitate remote deployment. The senors used in the system are also evaluated for their effectiveness.

Chapter 4 first elaborates the dataset collected in this study using deployment of multiple sensor systems from different geographic locations over a six month period. This chapter then discusses the use of deep neural networks for estimating the daily weight variations of beehive. This is the first work to estimate the daily beehive weight variations using machine learning. Chapter 5 proposes the use of self-attention encoders for the task of beehive weight estimation, using a bigger dataset collected over a 12 month period. The attention weights of self attention encoders are also used to evaluate the contribution of different sensors and features, as well as different time periods of the day towards daily weight variation estimations. Based on these findings, this chapter also describes the changes made to the design of the monitoring system. The final chapter of this thesis concludes our research, and discusses the limitations and future work.

Chapter 2

Background on Beehive Sensors and Monitoring Systems

This chapter provides a background on the sensors used in beehives, and analyses some recent monitoring systems proposed in the literature. The discussion starts with parameters which either trigger a change in the bee behaviour, or are an indicator of change in the status of bee health or activity. An understanding of these parameters is important as they play a significant role in the selection of sensors for the monitoring systems. These parameters can be divided into different categories based on their relevance to bee health and activity. The factors associated with the use of these sensors are also discussed in detail. The is followed by an analysis of the several monitoring systems proposed by researchers for their choice of sensors, design and methodology, data collection techniques, communication, remote deployability, costs and data processing capabilities. A brief review of some of the commercially available BMS is also presented. All of this analysis is then used to identify and discuss major research gaps in this area.

2.1 Sensors and Parameters for Beehive Monitoring

Honeybees live in a very complex social structure, formed by thousands of bees, where each bee has a specific role to play to ensure the survival of the colony. The internal dynamics of the hive such as presence of a queen, it's age, egg laying ability of the queen, ratio between worker bees and the brood as well as forager bees are critical in determining the future population size of the colony. If a bee colony is able to increase in population, or at the very least maintain a population over a period of time, it is defined as a healthy colony. But if the queen is not able to lay enough eggs, or there are not enough worker bees to maintain the hive and take care of the brood, or there are insufficient foragers to support the nectar and pollen needs of the entire colony, the colony will not be able to maintain a healthy population. The number of bees in such a weak colony will decrease and if it continues for a long period of time, the colony may collapse. The availability of food, weather conditions, and threats including diseases, predators and exposure to pesticides are major factors which impact the health of a colony.

A good monitoring system should be able to sense variations in the internal as well as the external conditions, but many of these variations are difficult to detect. While some sensing techniques face technological limitations, others are not practical. For example, even with significant technological advancements over the last 20 years, it is not yet possible to detect the presence of a primitive stage virus in the colony. Only when the virus starts impacting the performance of bees, the abnormal trends and variations become obvious. Also, to inspect the state of the bees in the hive, use of cameras is not practical. Theoretically, it's conceivable to use thermal cameras, infrared cameras or low light cameras inside the hive but such systems will only monitor part of a hive, never providing the complete picture, all at the expense of valuable space inside the hive. Honeybees use propolis (the bee glue) [25] to cover the exposed sensors and circuitry of these monitoring systems which impacts their functionality. This means that a direct assessment of some of the key parameters is only possible with a physical inspection by the beekeeper.

Some researchers have tested commercially available environmental monitoring systems for beehive health monitoring [26]. These systems can measure temperature, humidity, various gas contents in the air such as Molecular Oxygen, Carbon Dioxide, Nitrogen Dioxide, Ethanol, Hydrogen Sulphide, Isobutene, Toluene, Ammonia, Carbon Monoxide, and Methane. This methodology has its drawbacks as many features used for monitoring the environment do not overlap with those required for monitoring the beehives (discussed later in this chapter). From a beekeeper's point of view, every inch of space inside the hive is precious, thus the bulky environmental monitoring systems are considered a poor fit. Most of the environmental monitoring systems intended for use in air or soil are also not packaged in a suitable manner for use with honeybees.

This background study presents different types of parameters, and the important factors to consider while selecting the sensors for use in BMS. The first and foremost factor is the *relevance* of the parameter to beehive health and state. This relevance should always be given the highest preference in deciding upon the use of a parameter in BMS. Based on the relevance criteria, the parameters are divided in four different categories with relevance as *Very High, High, Medium* and *Low or Unknown*. Preferences of other factors largely depend upon the priorities of the designers and the needs of intended users. Some may desire a very low cost system whereas some may want it to be extremely accurate. However, for an optimal design, all of the factors should be given appropriate consideration. This chapter will evaluate the following 6 factors for each sensing parameter.

1. Cost effectiveness: this includes the cost of sensor, any associated equipment or hardware used with the sensor, and the cost of sensor data storage and/or transmission. Low net cost has a higher rating.

2. Accuracy: the accuracy of the sensor, and the impact of the deployment methodologies of the sensor on it's accuracy. This also includes the accuracy of algorithms which extract the information from the sensor dataset. Higher accuracy has a higher rating.
3. Energy efficiency: this factor is the reciprocal of power consumed by the sensor, and the power used in processing and transmitting the data generated by the sensor. Lower power consumption leads to better efficiency and a higher rating.
4. Computationally simple: this is the reciprocal of total amount of computing resources required to acquire and process the sensor data. Lower computational complexity has a higher rating.
5. Deployablity: this is the measure of the ease with which sensors/systems can be deployed, and the ease with which the beekeeper can work with this sensor. Greater ease in deployability has a higher rating.
6. Invasiveness: this is the measure of the impact of sensors and related equipment on the normal activities of honeybees. Lesser impact on honeybees has a higher rating.

The quantification of each factor in these tables as number of '$\star$', ranging from 1 to 5 is relative to the same factor of the other sensors in the tables. For example, the size of temperature sensors is smaller when compared to the size of most the other sensors, while the size of weighing equipment is larger than any other equipment discussed in this review. The quantification of these factors is also dependent on the user's preferences, understanding, experiences and expectations from the sensor. For example, in case of absolute power consumption, some may consider consumption in the milli-watt range as low enough, while others may argue that power consumption should be down in the micro-watt range for a power efficient BMS. Moreover, sensing a particular parameter may have a different power consumption rating based on usage, operating conditions and mode of operation. More than often, different manufacturers also have different power ratings for a similar sensor which makes it difficult to use the absolute numbers for ratings. So instead of using the absolute values provided in datasheets, we use our understanding and knowledge of these sensors and $\star$ratings to quantify each factor relatively. The last factor of *Overall Usability* in the tables is the linear average of all 6 factors mentioned above, rounded to the nearest integer.

There are many parameters that can be used for direct or indirect assessment of the health/activity of the bees [27]. However the following sections will discuss a total of 14 parameters for their relevance, and the factors which are associated with the potential use of respective sensors in beehive monitoring systems.

2.1.1 Temperature

Temperature plays a crucial role in determining honeybee activity and health. Honeybees stop flying when the temperature outside the hive drops below 10°C [28]. At lower

Table 2.1: Parameters of very high relevance to bee health/activity and the factors which determine the usefulness and practicality of the associated sensing equipment.

Parameter	Relevance to Bees	Cost Effective	Accurate	Energy Efficient	Computationally Simple	Deployable	Non-Invasive	Overall Usability
Temperature	Very High	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★
Weight	Very High	★	★★★★	★★★★	★★★★	★★★	★★★★★	★★★★
Int. Imaging	Very High	★★	★	★	★	★	★	★

temperatures, bees stay inside the hive forming clusters, and use their bodies to generate essential heat. However, this metabolic activity requires them to consume the stored food at a much higher rate. With enough stores of pollen and nectar, strong honeybee colonies can survive the winter season in sub-zero conditions. However at the other extreme, honeybees cannot survive at temperatures above 45°C [29, 30]. Figure 2.1 shows the temperature data collected using a beehive monitoring system in different environments. The later portion of the graph shows the thermoregulation of a strong bee colony. More details about studies on thermoregulation of beehives are provided in the Appendix A.

Given that temperature provides very important information about the status and health of bees, temperature sensors are one of the most commonly used components in the hive monitoring systems. It can be observed from Table 2.1 that temperature sensors are an automatic selection for any beehive health monitoring system as they have an excellent profile. They are very easy to use in circuits, supported by a wide range of micro-controllers, and are cheap to buy. Despite being small in size and very energy efficient, these sensors can measure temperature with high accuracy of about 0.1-0.2 °C. They have a quick response time of less than 1 second and converting voltage output to a temperature value requires minimal amount of processing. All of these factors make the use of temperature sensors for beehive health monitoring a very logical choice.

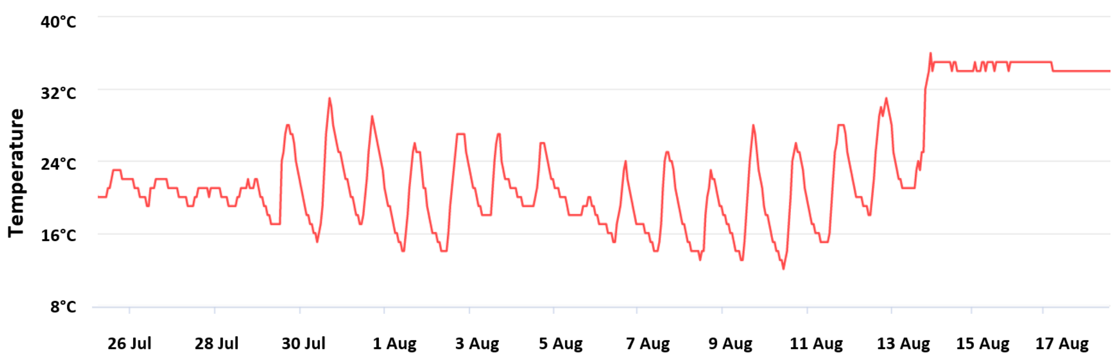

Figure 2.1: Difference between temperature variations indoor, outdoor, and inside a hive located at Yanchep, Western Australia. Temperature readings from a BeeBot sensor [19], inside the lab between 25th July to 29th July 2019, outdoor between 29th July to 13th August 2019, and inside a brood chamber of a hive from 13th August 2019 onwards during a relatively cold and rainy period.

2.1.2 Weight

Weight of a hive and its variations are the most obvious indicators of the status, health and activity of honeybees. A good healthy colony grows in numbers in favorable conditions and also collects honey and pollen for later consumption. This results in a net increase in the weight of the hive over a period of time during spring and summer. More details about daily weight variations of a hive and the factors involved are provided in Appendix B. If the weight of a hive is monitored accurately with a good resolution and at an appropriate interval, the variation in weight data can be used to extract information about bee colony strength and foraging activity [32]. The initial studies about the weight of hives and the factors impacting it were conducted using a mechanical balance in 1925 [33]. Since then, the weight measuring methods and techniques have improved significantly and a substantial amount of research has been reported on how to use the weight of a hive to determine the health status of the honeybees [34]. Figure 2.2 shows the structure of a langstroth hive, where the lower most compartment (called chamber/super) is usually reserved for the queen and the brood, and the remaining chambers are used for honey. A queen excluder restricts the queen to the bottom chamber, whereas the smaller hive bees can pass through and access all parts of the beehive. Weighing scales are usually placed underneath the hive stand to measure the weight of entire hive.

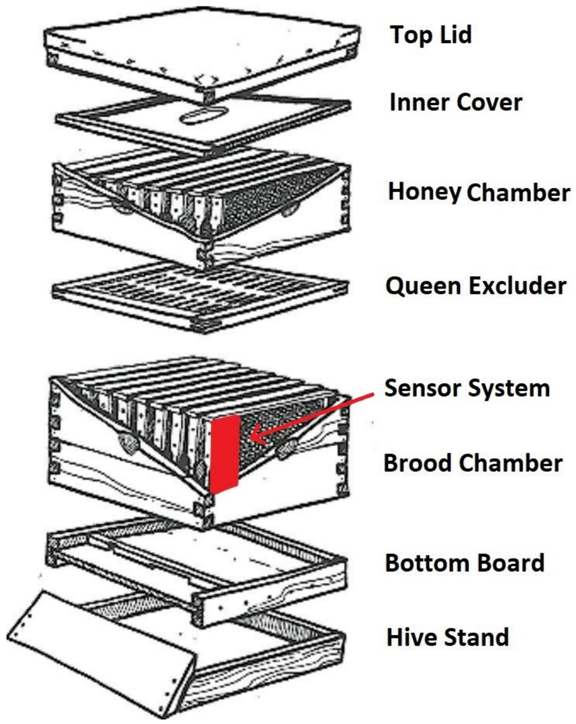

Figure 2.2: Basic structure of a Langstroth Beehive. The designed monitoring system (BeeDAS) is placed towards the edge of outer most frame in brood chamber, marked as red. Note: The image is adapted from [31].

During honey flow season (when nectar is in good supply), the beekeepers regularly replace the honey filled chambers with empty chambers to allow bees more space for honey storage. For accurate, reliable measurements and for ease of use, the weight sensor or weighing scale should be placed under a hive, where each hive can comprise multiple supers/chambers. A single honey chamber can weigh up to 30 kgs, which means that a hive with three honey chambers and a brood chamber can weigh up to 120 kgs. The weighing scale needs to be strong enough to handle this much weight, and sensitive enough to detect daily changes in the weight of hive with a resolution of few grams.

Another question about the hive weight scales is whether these scales should be a stand-alone device or integrated with the monitoring system inside the hive. Weight measuring instruments are very different in nature, size and hive placement when compared to most of the other sensors used in the monitoring systems. All other components of a monitoring system are usually enclosed in a single box or package, residing inside the hive. Weighing scale is often the only major component sitting outside the hive. It can either be designed to draw power from the monitoring system inside the hive and relay sensor data back to the monitoring system using wired connections, or it can have its own power supply, microcontroller and communication system so that it can work independently. Drawing power from the internal system and communicating data using wired connections is cost effective but the external wiring of such setup hinders the hive inspections and transportation. On the other hand, a stand-alone system provides a lot of ease-of-use but equipping each hive with such a scale can be prohibitively expensive.

2.1.3 Internal Imaging

Researchers have used imaging equipment inside the hive to capture photographs and videos [35], in order to monitor the activity of honeybees. Imaging is one of the most effective methods to gauge the bee population inside the hive, which is an important indicator of overall beehive health. Moreover, presence of queen, brood, food, and disease are also easier to identify using visual equipment. Figure 2.3 shows how easy it is to visually differentiate between capped brood and empty cells in a hive frame. However the use of imaging in beehive monitoring and its practical applications are a topic of much debate.

One of the fundamental requirements for imaging is the presence of light, which is scarce inside the hive. Having an artificial light source inside the hive is invasive as the heat generated from the light source can disturb the bee controlled thermoregulation. The light source and cameras being external objects are also subjected to propolisation by bees. This propolis blocks any light emitting from the source as well as the view of the camera. Opening the hive for capturing the images or video manually is only possible once a week at most, as anymore interference will have negative impact on the productivity of honeybees [14, 36]. A few studies have used transparent hives and external imaging equipment [37], but only a part of such hive can be observed and the information collected is incomplete. The cost of equipment used for imaging is high and the amount of processing required to extract the information is far greater than any other

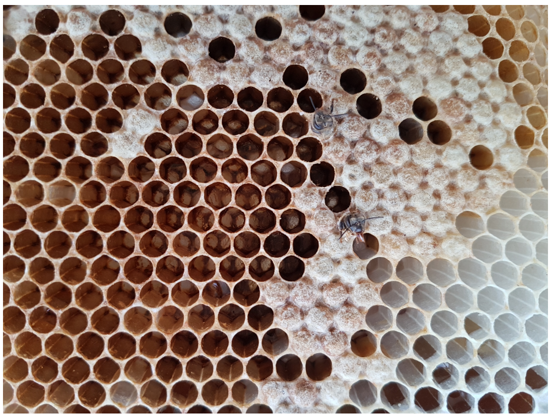

Figure 2.3: Two freshly hatched honeybees emerging from the brood cells. Capped brood and empty cells can also be observed in this frame. Image captured in May 2019 at UWA Campus.

Table 2.2: Parameters of high relevance to bee health/activity and the factors which determine the usefulness and practicality of the associated sensing equipment.

Parameter	Relevance to Bees	Cost Effective	Accurate	Energy Efficient	Computationally Simple	Deployable	Non-Invasive	Overall Usability
R. Humidity	High	⋆⋆⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆⋆
Acoustics	High	⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆	⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆
Ext. Imaging	High	⋆⋆	⋆⋆⋆	⋆	⋆	⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆

sensor used for beehive monitoring. The data generated by imaging systems is huge, hard to store and process within the hive, and very costly to transmit from a remote site in real-time. All these factors are reflected by low ratings for internal imaging in Table 2.1.

2.1.4 Relative Humidity

Relative humidity is the amount of water vapour present in the air compared to maximum amount of water air can hold at that temperature, expressed as a percentage. Experiments have shown that less than 40% relative humidity can dry the eggs and this results in significant reduction in numbers of hatching eggs [38], so nurse bees cover the brood area to decrease the loss of moisture [39]. Literature suggests that honeybees maintain relative humidity levels above 50% in the brood area [40]. Honeybee larvae are fed with a jelly excreted by nurse bees, and total composition of this jelly is around 67% water. If the relative humidity in the brood area falls too low or gets too high, the growth of larvae is adversely affected [41]. The relative humidity in healthy hives is maintained during the brood rearing [42], however during brood-less periods, the relative humidity levels of the beehive can fluctuate significantly. Figure 2.4 shows the difference in humidity levels inside and outside a hive. More details on factors impacting the humidity levels inside the hive and how honeybees regulate the hive humidity are available in Appendix C.

The humidity sensors are easy to use, small in size, and have low power consumption.

Figure 2.4: Difference between humidity variations indoor, outdoor, and inside a hive located at Yanchep, Western Australia. Humidity readings from a BeeBot [19] sensor, indoor between 25$^{\text{th}}$ July to 29$^{\text{th}}$ July 2019, outdoor between 29$^{\text{th}}$ July to 13$^{\text{th}}$ August 2019, and inside a brood chamber of a hive from 13$^{\text{th}}$ August 2019 onwards during a relatively cold and rainy period.

Inexpensive humidity sensors do not measure humidity with a high degree of accuracy but are still a good fit for beehive monitoring systems, where strict measurement of the relative humidity is not required.

2.1.5 Acoustics

Acoustic data from beehives is rich with information about the hive activity [43], and has been studied in detail by researchers. Literature suggests that audio frequencies inside a hive between 100 Hz and 1 KHz carry most of the information [44, 45]. Some beekeepers can assess the state of a hive just by listening to the bee buzz, as the magnitude of bee buzz is directly proportional to the bee activity. Bees generate different acoustic frequencies during specific events, such as swarming, or when alarmed [46]. Swarming is an event where the colony's queen leaves the hive with a part of the bee population from an existing colony to start a new colony. The queen is the single most important bee in the hive and audio data can also be used to determine the absence/presence of queen in a hive [47]. Some of the earlier work on beehive acoustics is shared in Appendix D.

Bees are a very social species and regularly communicate with each other inside the hive using bee dance (discussed later), audible frequencies generated through wings (bee buzz) and using chemical signals also known as pheromones. Bees inside the hive generate buzz dominated by a different set of frequencies depending upon the state of the hive or mood of the honeybees. These sounds vary in amplitude and have reported frequencies from sub 100 Hz to 3600 Hz [48]. However, most of the studies focus on frequencies lower than 1 KHz as this band carries most of the information [45, 44]. Research in beehive acoustics has shown that bee sounds can be used to detect pre-swarming, swarming, death of the colony, excitement level of bees and to some extent flight activities. However, the acquisition of sound is highly dependent on the location and placement of microphone(s) inside or outside the hive. It is speculated that if utilized properly, the acoustics can provide a lot of crucial information about the status of a hive such as being queenless, brood-less and lacking in food. This makes acquisition and processing of acoustics a very interesting and promising prospect for beehive monitoring.

The usage of acoustics in beehive monitoring has its own challenges which range from the placement and protection of microphone to dealing with high quantity of audio data and significant processing power required to process this data. Microphones placed outside the hive, near the entrance are able to pick up the sounds of forager bees going in and out of a hive, but are unable to detect sounds inside the hive. They are also exposed to weather conditions and environmental noise. Microphones inside the hive can detect internal sounds with good detail but not that of forager traffic. The internal microphones also need to be enclosed properly to avoid propolisation from bees. This greatly attenuates the sound reaching the microphone and forces the use of sensitive microphones and amplifiers.

Continuous recording of audio becomes a problem as it generates a lot of data, especially when sampled at high rates to analyse higher frequencies. Storing and/or processing this data locally using the onboard microcontroller is only possible if recorded in small bursts, after predetermined intervals of time. This reduces the space and power requirements but also creates the risk of missing out on significant audible events. To extract useful information from audio data, significant computational power is required, and catering for it onboard the monitoring system means using bigger, more expensive micro-controllers/processors resulting in increased power usage. If a large quantity of audio data is to be transmitted to a server for processing, this brings power hungry and costly communication into the equation. Thus finding the right balance for using acoustics in beehive monitoring is critical. This is why most of the work in this domain is observed in experimental setups rather than in commercial monitoring systems. However the promising potential of beehive acoustic data makes a strong case for it to be part of beehive monitoring systems.

2.1.6 External Imaging

Imaging of bees outside a beehive also provides useful information about health of a bee colony. Researchers have used standard imaging equipments to capture still photographs and videos of hives [49], mostly targeting the hive entrance to track the foraging activities and estimating the bee population. Unlike in-hive imaging, external imaging is non-invasive and bees do not try to cover the imaging equipment with propolis. The issue with imaging external activities and/or the entrance to the hive is that it provides only information about forager traffic, which comes at a great computational expense as flight tracking and bee identification algorithms are quite complex [50]. The amount of data generated from external imaging systems is also huge and requires high bandwidth and computational complexity for transmission and extracting the information respectively. Researchers have also used embedded systems with 256-core low power GPU [51] for tracking the bee activity at hive entrance in real time with good accuracy, but the employed hardware is expensive. The power consumed by these GPUs is significantly less than standard GPUs but for constant functioning throughout the day, these GPUs still demand power levels not practical for remote deployments. Despite significant reduction in the cost of electronics, high quality imaging equipment is still quite expensive, and most of the proposed imaging based BMS use a single camera per hive [52]. For a large scale operation where the number of hives can exceed thousand, this is not a practical approach. Safety and security of expensive imaging equipment placed outside the hive when deployed in remote areas also adds to the worry of beekeepers.

With so many practical limitations associated with the use of imaging in beehive health monitoring, there is room for substantial research and innovation in this domain. Recent research focused on using high speed cameras to track honeybees has shown promising results [53] but it is again too expensive for practical use. As the technology progresses, the algorithms improve on their accuracy and cost of imaging equipment gets lower, external imaging will become more deployable for beehive health monitoring. Despite current limitations in using the imaging systems for beehive health monitoring, external imaging is placed in Table 2.2 with other parameters of high relevance.

2.1.7 Accelerometer

As discussed in section 2.1.5, bees communicate with each other through audible frequencies or bee buzz, which is generated through wings. Before swarming, the bees inside the hive communicate using 500 Hz - 600 Hz frequency band instead of normal 100 Hz - 300 Hz frequency band [44, 54]. The bees also dance to communicate, which is referred to as a waggle dance [55]. This dance is mostly associated with communication related to source of pollen and nectar [56]. The waggle dance creates vibrations that travel through the hive. The frequencies of these dance vibrations are usually less than 300 Hz and honeybees use special receptors in their legs to receive these vibration signals [57, 58]. More details about honeybee waggle dance and the generated vibrations are provided in Appendix F.

Accelerometers are good at picking up vibrations, and can be used to collect information about the pollen and nectar availability in the hive surroundings. However, honeybees and their communication using vibration has evolved for use on wax made comb. In commercial beehives, wax comb is present only in the middle of a wooden/plastic frame, which has a two sided hanging connection with the hive structure. Hence, the signals of these vibrations attenuate a great deal while travelling from one frame to another [57]. The impact of this attenuation on communication of bees has not been studied in detail, but this attenuation makes the measurement of vibration using accelerometers very difficult. However once detected, these vibrations carry information about peak foraging hours and pre-swarming states. It has also been suggested to use vibrations generated by an electronic device embedded in the hive, to disrupt the pre-swarming communication and cause confusion among the honeybees [24]. This will cause a delay in swarming and the beekeeper will have time to intervene. Other than vibrations, accelerometers can also detect a change of position, or the movement of the hive, which can be used to generate an alarm to indicate unauthorized opening/movement of hive. Given this relevance between vibrations and hive state, accelerometers have found a place in Table 2.3 with medium relevance parameters.

Table 2.3: Parameters of medium relevance to bee health/activity and the factors which determine the usefulness and practicality of the associated sensing equipment.

Parameter	Relevance to Bees	Cost Effective	Accurate	Energy Efficient	Computationally Simple	Deployable	Non-Invasive	Overall Usability
Accelerometer	Medium	★★★★	★★★★★	★★★★	★★★	★★★	★★★★★	★★★★
Gas Contents	Medium	★★★	★★★	★★★★	★★★★★	★★★★	★★★★	★★★★
Counters	Medium	★★★	★★★	★★★★	★★★★★	★★★	★★	★★★
Ther. Imaging	Medium	★★	★★★★	★★★	★★★	★★★	★★★★★	★★★

2.1.8 Gas Content

Gas emissions inside the hive are a good indicator of hive health [59, 60]. Some bee diseases produce distinct smells [61], which if detected can help beekeepers either take steps to cure the disease, or to prevent its spread to other hives. Honeybees like all other living organisms, use respiration to move oxygen into the body and remove carbon dioxide from the body. Since a bee colony consists of thousands of bees packed in a very close space, the composition of air inside the hive is different from the air outside. In a healthy hive, the levels and the fluctuations of CO_2 are correlated with the metabolic activity of bees [62] inside the hive. See appendix E for some details on the work done on determining the composition of gases inside a hive.

Just like temperature and humidity, honeybees also regulate the gas contents inside the hive, especially the CO_2. Excessive CO_2 triggers the fanning activity of bees which allows the flow of fresh air into the hive. This ventilation with periodic fanning first allows air current to move out of the hive, and then the fanning of bees at the entrance causes an influx of fresh air into the hive [63]. This regulates the CO_2 levels between 0.1% and 4.25%, with large colonies able to control this with greater precision [64]. The respiration in bee colony decreases over night and this results in different levels of O_2 consumption between day and night. It has been studied that honeybees control and bring the O_2 to lower levels between 15% to 7.5% inside the hive during winter and introduce a bee-induced hypoxia to reduce the metabolic rate. With these low metabolic rates, bees are able to conserve energy and consume less food and water, which allows them to last long winter seasons with limited food stores [65]. Studies have also shown that different types of honeybees have different metabolic rates, resulting in different rates of O_2 consumption [66].

This shows that CO_2, O_2 and other organic compounds inside the hive atmosphere carry information about the bee health and status. A steady level of CO_2 is an indicator of proper air ventilation inside the hive which is a sign of healthy and active bee colony. The sensors for measuring gas contents have improved in their accuracy and have become much smaller over the years. However those with high accuracy are still expensive and need a controlled flow of air for proper measurements. The sensors deployed in BMS are designed to stay in hive for longer periods of time, however a lot of gas sensors require regular calibration in order to maintain their accuracy. Most of the gas sensors are relatively much more power hungry as compared to temperature and humidity sensors. Also, the large variety of gas sensors required to pick up specific gas emissions need detailed justification given the complexity and cost considerations. This means that the use of gas sensors in BMS is not easy, but these sensors cannot be completely ignored given their relevance.

2.1.9 Counters

Counters provide an alternate to imaging for estimating the bee population in a colony. One of the earlier solution for counting bees using photoelectric counter was proposed in 1969 [67]. In this very limited setup, glass tubes were used to separate outgoing bees from incoming bees for a single hive frame. Two digital counters kept track of bees passing through each tube and a camera was used to periodically record the digital values of the counters. Since then, the basic concept of digital counting has remained the same, but different methods have been tested to make it non-invasive. Newly proposed microprocessor-based systems allow multiple entrances with limited obstacles for bees and use algorithms to improve the counting precision [68]. Such systems can also be integrated with other monitoring systems used for measuring the internal parameters of the hive [69].

The traffic at the hive entrance is a good indicator of peak colony activity, preferred flight timings, swarms, robbing and percentage of forager return. Foragers are a key link between colony and food, which is essential for colony survival and honey production. Forager activity is also an indirect measure of bee pollination around the hive site. However the use of counters often restricts the movement of foragers, blocks the colony entrance which impacts the thermoregulation and CO_2 levels inside the colony. Entrances of hives are well maintained by honeybees but the use of counters and narrow channels makes it difficult for the bees to clean them and restricts the disposal of waste from the hive. These channels at times have to be manually cleaned to allow clear passage to the honeybees.

Some natural activities of honeybees offer challenges in counting the forager traffic. Guard bees which are present at the hive entrance, often in varying numbers, can block such counters. These bees often move around and can trigger multiple false counts. A similar event can happen during the fanning activity when a large number of hive bees (different from forager bees) move outside the hive and use their wings to generate a fresh flow of air into the hive. The electronics of these counters is very sound but because of complex bee behaviour, these counters are unable to differentiate between a forager and a hive/guard bee, which compromises the accuracy of these counters. The restrictions caused by such counters at the entrance are invasive which makes them not a preferred choice for most beekeepers, as reflected in Table 2.3.

2.1.10 Thermal Imaging

Thermal imaging of beehives is another way to estimate the number of bees in the colony [70]. Some studies have also used infrared imaging [71] which is similar in nature. Information available through thermal/infrared imaging is not as relevant to bee health as other imaging techniques discussed in section 2.1.3 and section 2.1.6, but is still significant. Using readily available low resolution thermal imaging devices, it is quite easy to observe the relative size of a colony, also known as colony strength. It can be seen from Figure 2.5 that each hive has its own cluster of bees, at different locations inside each hive. The size and intensity of yellow glow of each bee cluster is directly proportional to the number of bees in that cluster. It can be observed that strength of bee colony in the hive on right is lower than the other two hives. However thermal imaging has certain limitations which are discussed in Appendix G.

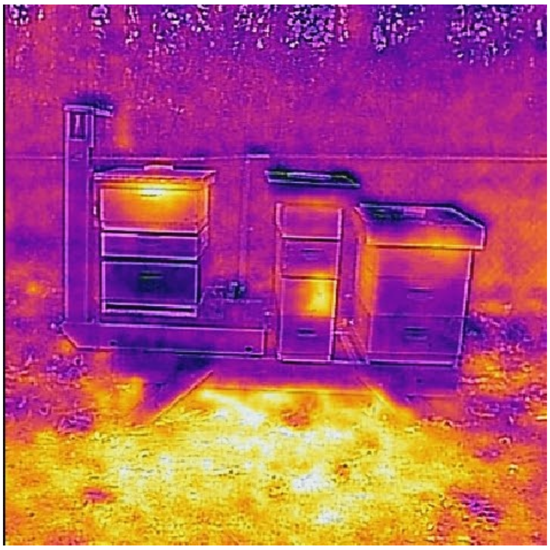

Figure 2.5: Thermal image of three beehives side by side showing the bee cluster as yellow glow. The hive on the right appears to be weak compared to the hives in the middle and the left. Image credits: Foxhound bee company [72]

Thermal cameras with low resolution are relatively affordable and easy to use with microprocessors. They have reasonable accuracy and energy consumption is not very high. However they still need some computing power to extract information but because of low resolution and less detailed nature of thermal imaging, this computation is much less complex when compared to other types of imaging. Thermal cameras have similar issues as external cameras when it comes to deployability because equipment used outside the hive means more work required by beekeeper in managing the placements and connections. External equipment also obstructs the movement of workers and vehicles during replacement of the full honey supers with empty ones. And when BMS using thermal or infrared cameras are deployed at remote location, they add to the security concerns of the beekeeper as sensors outside the hive are more prone to theft.

2.1.11 Global Positioning System

Monitoring systems equipped with Global Positioning Systems (GPS) are able to keep track of their location. Commercial hives are moved from one location to another based on availability of floral resources and weather conditions [73]. This tracking is important for supply chain management and quality assurance as these locations are often used to

Table 2.4: Parameters of low or unknown relevance to bee health and the factors which determine the usefulness and practicality of the associated sensing equipment.

Parameter	Relevance to Bees	Cost Effective	Accurate	Energy Efficient	Computationally Simple	Deployable	Non-Invasive	Overall Usability
GPS	Low	★★★★	★★★★★	★★	★★★★★	★★★★	★★★★★	★★★★
RFID*	Low	★	★★★★	★★★★	★★★★	★	★★	★★★
Pressure	Unknown	★★★★★	★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★
Magnetic-Remanence	Unknown	★★★★	★★	★★★★	★★★★★	★★★★★	★★★★★	★★★★

* RFID tags can only be used with individual bees. Rest of the parameters / phenomena in this table are for entire colony.

validate the floral sources of the honey, which vary from location to location. Monitoring systems equipped with GPS also relay the current location of hives to the servers, which uses the weather conditions at the hive location along with sensor data to estimate a more accurate status of hive. However with no direct impact on health assessment of bees, GPS is a low relevance parameter for beehive health monitoring.

2.1.12 Radio Frequency Identification

Radio Frequency Identification (RFID) [74] has been widely used in experiments related to honeybees in recent years [75]. Tiny RFID tags can be placed on the back of honeybees without significantly disturbing their regular activities. Each tag is uniquely identifiable with the help of a RFID tag reader. Scientists and researchers place these readers around the hive and in areas of interest, or move around with one in hand, and each time a tagged bee passes by the RFID reader, the time and location are automatically recorded [76]. This allows researchers to find out the flight track and pattern of bees in a vast area [77]. More details on use of RFID in honeybee research can be found in Appendix H.

RFID has been helpful in studies involving honeybees but for long term monitoring, tagging thousands of foragers in a hive is not feasible. Even if a few foragers are tagged to get an estimate of flight patterns, these tags will have to be regularly put on new forager bees as the life span of foragers is only a couple of weeks. Bees can travel upto a few kilometers in search for pollen and nectar, thus equipping such large area around a hive site with RFID readers is also not feasible. With applications limited to bee behaviour research only, RFID is a parameter with low relevance.

2.1.13 Atmospheric Pressure

Atmospheric pressure and its impact on honeybees has been part of very few studies [29]. These studies did not find any direct relationship between atmospheric pressure and bee health/activity, primarily because the experimental setups did not significantly vary the atmospheric pressure. To properly test the impact of pressure on bee health and activity, identical experiments need to be performed at different altitudes to vary the atmospheric pressure. Some research indicates that for predicting bee foraging activity using local weather conditions, use of atmospheric pressure can slightly improve the performance of predictive models [78]. However, major change in atmospheric pressure is only observed with a change in altitude, and significant change in the altitude also results in a change in temperature and oxygen levels. For any study that aims to find the impact of major change in altitude/pressure on bees, negating the impact of temperature variations caused by altitude would be required. The atmospheric pressure sensors are low cost, small in size, easy to use and have a decent accuracy, making them very usable. But because of their unknown relevance, they have been placed in Table 2.4.

2.1.14 Magnetic Remanence

The earliest work suggesting that honeybees may have magnetic remanence [79] was published in 1978. Magnetic remanence is the ability of a material to retain magnetization after being exposed to magnetic field(s). During the experiment, researchers applied a strong field of about 700 gauss to 18 honeybees for a brief period. Subsequent measurements found a permanent remanence induced in 15 of the honeybees, with an average strength of 2.7×10^{-6} emu. Earth's magnetic field is about 0.5 gauss, and any magnetic field induced in honeybees because of earth's magnetic field will be significantly smaller, most likely by a factor of 1,400 compared to what the authors observed in the honeybees after the experiment. Some studies [80, 81] have focused on the impact of magnetic fields on the activities and flights of honeybees. However the magnetic remanence has never been used to study bee health. It will be interesting to observe if easily available, small and relatively cheap magnetometers are able to detect the tiny magnetic fields generated by honeybees. This is why magnetic remanence is placed in Table 2.4 with unknown relevance.

2.2 Analysis of Beehive Monitoring Systems in Literature

This section will analyse some of the beehive monitoring systems that have been proposed in the literature. Many such systems have been proposed over the years but this discussion will be limited to only the shortlisted systems which satisfy the following:

1. Experimental prototype(s) designed and tested
2. Evaluated parameter(s) with significant relevance to bee health/activity
3. Published as peer-reviewed work

These three criteria allow for more realistic comparisons and assessments. They allow the discussion to be focused on systems which not only propose the design of a complete practical setup, but also develop and deploy prototype(s) to collect the experimental data. Recent technological advancements have helped reduce the cost, size and power consumption of many electronic sensors, which make them very suitable for use in BMS. This discussion will be limited to prominent systems proposed between 2012 and 2021 in this comparative study. Older systems have a distinct disadvantage in comparison because of the rapid advancement of tools, techniques and sensors. *Ferrari et al.* [44] is the only exception of an older work to be included, as this work from 2008 is a very comprehensive study on honeybee swarming detection system, and is included in this review.

With multiple disciplines involved in the design and development of BMS, it is not possible for any proposed system to be comprehensive in all of the areas. So the aim is to discuss different aspects of each shortlisted system, in order to come up with a set of guidelines for designing a monitoring system. The selected beehive monitoring systems are analysed on the basis of the following:

- Selection of parameters
- Design and methodology
- Data storage/communication
- Remote deployability
- Cost
- Data processing

We will perform a subjective analysis of the selected systems in this section. Some particular details about these systems, with respect to the analysis criteria mentioned above, are provided in Appendix I.

2.2.1 Selection of Parameters

As discussed in the previous section, selection of parameters for the assessment of bee health is the most important aspect in determining the overall effectiveness of a monitoring system. There are many drivers which form the basis of this selection, most common of

Table 2.5: Shortlisted beehive monitoring systems from literature, along with the parameters used in their design.

	Very High Relevance	High Relevance	Medium Relevance
Howard et al. [37]	Temperature, Weight	Humidity, Acoustics, External Imaging	Gas Contents
Tashakkori et al. [82]	Temperature, Weight	Humidity, Acoustics, External Imaging	-
Murphy et al. [26]	Temperature	Humidity	Accelerometer, Gas Contents
Edwards et al. [59]	Temperature	Humidity	Accelerometer, Gas Contents
Murphy et al. [71]	Internal Imaging	Acoustics	Accelerometer, Thermal Imaging
Gil-Lebrero et al. [83]	Temperature, Weight	Humidity	-
Ferrari et al. [44]	Temperature	Humidity, Acoustics	-
Anuar et al. [84]	Temperature, Weight	Humidity	InfraRed Counters
Anand et al. [85]	Temperature, Weight	Humidity, Acoustics	-
Konig et al. [86]	Temperature, Weight	Humidity, Acoustics	Gas Contents (VOC)
Kulyukin et al. [87]	-	Acoustics, Video and images	-
Kridi et al. [88]	Temperature	-	-
Chen et al. [35]	-	-	Infrared Imaging

these are the research interests of those proposing the system, the intended objectives, and the availability of required tools and other resources. Table 2.5 provides a summary of parameters used by the systems included in this analysis. None of the selected systems use any of the low relevance parameters from Table 2.4, so this category is omitted from Table 2.5.

The purpose of this evaluation is to assist the sensor selection process for a beehive monitoring system suitable for remote deployments. From Table 2.5, it is evident that the majority of the researchers agree on the importance of parameters such as temperature, weight, humidity and acoustics. As discussed in previous section, imaging systems do not fit the profile of a remote beehive monitoring system because of high power consumption and the amount of data they generate. Counters at the hive entrance are intrusive, so they should also be excluded from the design. Accelerometers and gas sensors have a medium relevance with bee health/activity, but their better usability makes them a good candidate for inclusion in the design of a remote beehive monitoring system. From sensors with unknown relevance, atmospheric pressure sensors, and magnetometers also have a suitable profile.

Based on relevance with bee activity/health and suitability for remote deployment, following sensors are shortlisted for our design of the beehive monitoring system, namely:

1. Temperature
2. Humidity
3. Atmospheric pressure
4. Acoustics
5. Accelerometer
6. Gas sensor(s)
7. Weight

The functionality of these sensors in designed monitoring system is discussed in detail in the following chapters.

2.2.2 Design and Methodology

This section analyses the proposed designs of beehive monitoring systems, along with different methodologies used by the researchers during sensor deployment and testing. The objective is to identify a design which is most likely to work in a remote environment, and can be relied upon to collect data from beehives. An analysis of different methodologies by different researchers will also help guide the formulation of a methodology for the experimental deployment of our proposed system. Researchers in this work [37] designed and developed their own sensor system including the weighing scales, and used a commercial beehive weighing scale only to validate the results. Use of commercial equipment in the experimental setup significantly limits the ability to customize the design. Designing a system from scratch, including the weighing scale is not easy, but it allows an absolute control over components used in the data collection process which is important. However, the authors collected sensor data at such a high rate that it could not be transmitted in real time. Onboard storage was used to save this data, which had to be manually retrieved every few days. This approach of manual data retrieval does not work well for remote deployments. Hence, the data generated by the system to be designed needs to be small enough to be transmitted in real time.

Researchers in this work [82] opted for a plug and play design, which allowed for easy replacement of components. For a prototype/test-system, this is a very desirable feature which allows rapid replacement, facilitates testing, and changes in the design. In the system proposed by researchers, a client could easily relay commands from a terminal allowing control over the system. This bi-directional communication becomes even more useful in remote deployment, where the beehive monitoring system can be configured remotely. Authors in [26] explored the use of a dedicated transmission system to pool the data from local monitoring systems, and transmit it over long distance. They also explored the use of dedicated sensor systems with different roles, such as one for gas detection inside the hive.

Sensor placement in the roof of the hive resulted in significant variations in the recorded data because of the high influence of external hive conditions. This signifies that the selection of location for placement of sensors inside the hive is very important. Also, using a variety of systems for different roles not only increases the overall complexity of the design, but can also create bottlenecks. In case the system responsible for long-range communication fails, none of the working systems at that particular site will be able to transmit the data. Having each system as a stand-alone device increases the cost per system, but also improves the overall reliability of the system. The authors also used official meteorological data for their hive state classification, which is a good approach to avoid setting up dedicated weather stations for this data.

Infrared imaging has also been explored for its use inside the hive [71], however it was only tested in a hive without bees. In the same setup, the authors used low sampling rates of 100 Hz to sample the audio data to minimize the data rates. This low sampling rate can only monitor bee buzz up to 50 Hz, and fails to collect any information in higher audio bands. For remote deployment of a monitoring system, the reduction of data for communication is important, but it should not cause a significant loss of information. Use of multiple temperature sensors inside a single beehive was also explored in a study [83]. Given the importance of temperature inside the hive, multiple sensors can collect more information about the thermoregulation of the bee colony compared to a single temperature sensor. However from a practical aspect, this can create difficulties for the beekeeper. The extra wiring inside the hive, with multiple frames connected to a temperature sensor each, make the regular hive inspections as well as swapping of frames difficult. The authors also used a single load cell to measure the weight of the hive with a resolution of 100 grams. This resolution is enough for tracking the weight change of a hive over multiple days, but cannot be used to monitor minor weight variations within a day. However this monitoring system was used in a promising 20 bee colonies, which is significantly more than the number of hives used by other systems analysed in this study.

An important study on early detection of swarming [44] used temperature, humidity and acoustic data. The main finding from this study was that bee noise generated during the swarming had a much higher Power Spectral Density (PSD) when compared with the PSD of standard bee buzz, specially in the 500 Hz to 1000 Hz band. To monitor the frequencies of up to 1000 Hz, the sampling rate of the audio data should be at least 2 KHz. This study [85] also used the audio data from inside the hive to detect swarming. The authors used a 256 point Fast Hartley Transform (FHT) to extract frequency components from the audio data. This is a good approach as the audio data with high sampling rate is difficult to transmit from remote sites. The extracted frequency components are much smaller in size and easy to transmit. For a better resolution, a 2048 point transform can be used given the microcontroller unit within hive is able to handle the computations involved.

Authors in this study [84] collected data from a hive with a 7 second interval. This generated 12,200 data points in an experiment of 36 hours in duration. The natural activity of honeybees and the variations within hive micro-climate change with different seasons. Data collected with a larger interval between samples, and over a longer period of time most likely carries much more value. But within each sample, the data such as audio and vibrations should be collected with appropriate sampling rates, and for reasonable durations to capture adequate amount of information. This however is limited by the amount of computing resources available in the monitoring system.

Work of [86] focused on Varroa mite infestation. Until 2022, Australia was the only country in the world free of Varroa mites, but first cases of this mite were reported in New South Wales in June 2022. The authors in their design used Volatile Organic Compounds (VOC) gas sensor in an attempt to pick up the Varroa infestation. Use of gas sensors is quite challenging, which has also been discussed in Section 2.1.8. However if used correctly, the gas sensors can add a value to monitoring systems which other sensors cannot.

Custom monitoring system BeePi [87] is one of the few works where the collected dataset has been shared publicly. Reproducibility and replicability were two fundamental objectives of the BeePi project, so that others may repeat the experiments and replicate the design at minimum costs. Also, the system was designed such that sensors do not interfere with the natural activity of honeybees. For this reason, the microphones were deployed outside the hives, above the landing area of foragers bees. This setup maximizes the audio data collection from forager traffic, but is less sensitive to bee buzz inside the hive. External microphones are also prone to picking up noise from hive surroundings. Authors monitored hives with different species of honeybees, an aspect which is often overlooked by other researchers. As the bee research community develops, it is access to complete datasets such as this, that provide significant value to data analysis and research efforts towards smart beehive monitoring systems.

For a lightweight, low energy and an economical solution, a data collection platform was developed with very limited computational capability [88]. This platform was deployed in two hives and recorded the inside temperature for multiple days. Authors used k-means clustering algorithm on the collected temperature data from the healthy hives to identify 6 temperature patterns for daily variations based on hourly readings. These patterns were used to define a threshold for acceptable micro-climate inside the hive. This approach of classifying the hive state using the deployed sensor system, and communicating just the state is very interesting. However given these states were determined using limited amount of data, it begs the question about the accuracy of this approach for hives from different geographic locations.

In a very unique experiment [35], authors used an infrared CCD camera, and bees tagged with different characters to observe foraging patterns. This approach is only feasible for a short term study on forager activity. In practical systems, tagging bees to monitor their activity is not possible, as the number of foragers inside a hive can be in thousands, and foragers die after about two weeks of foraging activity. Also, such setups force the bees to use specific passages at the hive entrance, which is intrusive in nature and slows the natural activity of the bee colony.

2.2.3 Data Storage/Communication

This section analyses the data storage and communication capabilities of the shortlisted systems. Onboard SD cards were used by [37] to store small amounts of sensor data, and an external hard disk drive for large audio and video data. The collected data had to be manually transferred every few days to make space for new data. Such a setup is very demanding even for research purposes as it requires constant involvement of personal to retrieve the data. For any system collecting data from a remote site hundreds of kilometers away, this becomes almost impossible. The data communication system should allow for stand-alone operations without human involvement. Authors in this work [82] also used SD card on raspberry Pi to temporarily store the high bandwidth audio and video data. The sensor data from low bandwidth sensors was transmitted using Message Queuing Telemetry Transport (MQTT) [89] protocol. MQTT is a very suitable option for data communication in low data rate systems. In most cases, beehive monitoring systems transmit data from machine to machine, which lightweight MQTT protocol can handle very efficiently.

These two studies [26, 59] utilized low power RF modules for communication between two hive systems and a third base station system. The base station acted as a 3G radio bridge for long distance communication for the hive systems, and combined their data into a single file to upload it to a server. All the systems were designed to transmit once per day which reduced the power consumption, but also limited the real time flow of information to the beekeepers. For the purpose of data collection from beehives, real time communication has a low priority. Pooling the data for transmission can conserve the power, as overheads of establishing a connection and the handshaking protocols can be minimised. This can become very useful when system has a low battery and conserving the power becomes a high priority. In another study by same authors [71], the system was designed with cellular communication capability without any bridge. This system had an advantage of ultra-low power operation and allowed remote deployment of beehives depending upon the cellular coverage in the area.

For data transmission from hives, this work [83] used low-rate wireless personal area network. Data requests were periodically broadcast by server to collect data from hives in a synchronised manner. A database was used to store the beehive sensor data, with backups regularly made. Beekeepers were able to access the hive data using internet. This is one of the most complete data communication and storage systems analysed in this chapter. Some researchers have also explored the use of high gain antenna [88] to establish a wireless link between monitoring systems and the base station. Use of directional or high gain antennas can help improve the coverage area, but these antennas are mostly mounted externally on the hive structures. This becomes a problem for beekeepers when they decide to move the hives from one site to another. The rectangular structure of hives allows for very compact placement on the trucks. With external antennas, the hives cannot be stacked right next to each other. Disconnecting the external antennas before hive transportation and connecting them again at the new site is time consuming. Use of high gain antennas inside the hives is only possible if antennas do not take up a lot of space. Directional antennas inside the hive are difficult to manage because majority of beekeepers prefer certain orientations with respect to the sun while placing the hives. Adjusting the direction of the internal antennas at every new site again adds to the complexity, and decreases the user friendliness.

2.2.4 Remote Deployability

Remote deployability of a system can be gauged using two attributes; ability of the system to collect and transmit the data without human involvement, and its ability to perform with limited power resources. The communication capability of systems has already been discussed in previous section. All the systems which used local SD cards and external hard drives to collect the data without any data transmission have poor deployability. The dependence of a system on short range communication such as WiFi also hinders its deployability in remote areas, unless a dedicated WiFi gateway is put in place for the internet connectivity. This section only analyses the shortlisted systems which were designed to perform with limited power resources.

Authors in this design [26] included solar panels for energy harvesting. The sensor systems used 6.5 V, 205 mA solar panels with 6600 mAh batteries, whereas the base station featured a sightly bigger solar panel with 7 V, 500 mA rating. The gas sensors however were very power hungry and the solar panels were inadequate to power these systems. An improved version [59] used the same old setup with better deployability by reducing the number of samples of gas collection from 6 to 3 per day. This provided a significant improvement in battery performance as the installed solar panels were able to replenish this smaller battery drain. Solar panels have significantly improved their efficiency since these studies were conducted in 2015-16, and small solar panels with 16cm x 8cm dimensions are now able to generate up to 2000 mA of current with a 5-6 volt output. This has enabled beehive monitoring systems to use a larger number of sensors, and collect data more frequently. Authors in another work [71] implemented energy harvesting through solar panels using a 1000 mAh battery, which would last several days even without sunshine. These systems compared to their previous setups were much more energy efficient, and the total energy requirement for each system could be met using the solar panels. However, this setup generated a lot of image data, and authors did not test the working of 3G network to transmit this data to server. Communication system is one of the most power hungry aspect of monitoring systems, and a system is only considered self reliant if it can support all of its power needs.

In this work [83], authors claimed that their proposed system is compatible with solar panel for charging the batteries but it was not tested. Batteries and external power supply were used for the systems. The designed system was analysed and found to be energy efficient with data transmission as the biggest consumer of energy. The batteries could last 3 days without charging, which is a reasonable period if solar harvesting is available. Some researchers have also used power packs [88] composed of AA batteries, however no details were provided on how long these batteries were able to power the monitoring system.

Authors in this work [87] experimented with three types of power supplies: solar, battery, and AC power. A single system required 440 mA of current, primarily because of high power consumption of the Raspberry Pi and its camera. Authors found it difficult to sustain this much power usage using solar panels, especially in winter when days were short and often cloudy. The batteries also did not last long in colder temperatures and replacing them frequently at remote sites over a sustained period became logistically very difficult. For these reasons, authors eventually used AC power for all of their systems.

2.2.5 Cost

From a beekeeper's perspective, cost of a beehive monitoring system is one of the most important factor. Commercial beekeepers only invest in monitoring systems if they are convinced that they will get a good value for their money. This section discusses the proposed designs which have provided a cost analysis. For any system designed for long term deployment, there are running costs involved as well. This includes cost of electricity if powered using AC lines, battery replacements, data communication/internet costs, and those associated with human involvement for data retrieval if applicable. These costs are hard to estimate for research systems, and are not discussed. Authors for their design [37] provided a cost analysis of the components and it varied based on the parameters used for monitoring. For a complete system which also included video monitoring equipment, the proposed system cost around GBP 190 whereas the Arnia system without any video monitoring capability cost around GBP 300. The proposed system in this setup included a weighing scale, whereas it had to be purchased separately for the Arnia costing an extra GBP 700. This shows that the proposed system had a very cost effective design when compared to commercially available Arnia system.

This system [82] was designed with an objective to be cost effective. The hardware cost of a single unit is reported at USD 106 by the authors. This includes Raspberry Pi 3 Model B with its power supply, 16 GB SD Card, Raspberry Pi Camera, USB microphone and the cost of materials for printing the 3D case. However, the authors did not report the cost of sensors and the weighing scales.

These two studies [26, 59] used Libelium Waspmote and other commercially available Libelium sensors, all of which come with a high price tag of AUD 800 or more. Using a few Libelium systems for experiments may be a feasible option, but using a large number of such for experiments is impractical. The Libelium sensors are designed for use in environment which requires a low density of deployment, whereas beehive health monitoring systems require a much higher density of deployment within a single hive site. High costs of these environmental sensors also make them a poor fit for commercial beekeeping. The cost of equipment used in this work [71] is lower than the previous setups [26, 59] by same researchers, as this new setup is based on raspberry Pi and other readily available sensors. Waspmote was used only to control the boot cycle of Pi boards, which helped reduce the power costs.

Some authors have used a modular approach [86] where different components can be added to the system. The very basic version of this system consisting of raspberry Pi Zero, temperature, humidity and MEMS microphone cost less than USD 30 for components. However the addition of modules for weighing scale and gas sensor will significantly increase the cost. Authors in this work [88] estimated that the cost of a minimum viable product for their proposed system would be between USD 150 to 200.

2.2.6 Data Processing

Data processing is a very important aspect of modern beehive monitoring systems. The ability of a system to decide upon the accurate state, predict the future state of a hive, and to provide valuable insights about bee health is what makes beekeepers invest in these systems. For this reason, many researchers are working on improving the decision making aspect of beehive monitoring systems. This section looks into the data processing capabilities of systems included in this analysis.

Applications and web tools were developed by researchers to capitalise on rich audio and video data recorded by Beemon [82]. The authors briefly discussed the image processing (BeeVee), audio processing (BeePhon) and streaming (BeeStream) tools in their work. BeeVee uses object detection and tracking techniques on the video stream obtained from the hive entrance to detect the honeybees and provides an estimate of the forager traffic. This tool can be used to automatically analyse the bee data and generate alerts when needed. BeePhon allows visualization of the honeybee audio data, and can be used to analyze audio data for a specific time period. A web interface for the Beemon project provides access to the BeeStream component of the project, where users and researchers can live stream video and audio from beehives, and view the video analysis data related to forager traffic superimposed on temperature and humidity data. Users and researchers can also use the available tools to label the collected data. The labeling/annotation of data is a very important step which allows the training of supervised machine learning algorithms, and the development of smart beehive monitoring systems. Machine learning is a branch of Artificial Intelligence (AI) which uses data and algorithms to mimic another process, or imitate a human trait.

Researchers used multiple experimental setups [59, 71, 26] to collect a substantial amount of beehive sensor data. They were also able to incorporate a classification decision tree for beehive states using this data, and used text messages to alert the beekeeper in case of specific events. A total of ten classes were used for the identification of crucial colony states, including healthy and unhealthy conditions. Authors expressed their desire to work towards decision making capability of monitoring systems in the future. Machine learning has shown a lot of promise in many areas but requires extensive amount of data and labelling to generate quality results. Future endeavours of authors [90] show that they collected more data through sensor deployments, and used big data and machine learning to develop a commercial product.

An efficient, low power algorithm was designed and implemented on the monitoring system within the hive [88] to monitor the hive temperature. The system generated alerts for the beekeeper only when the hive temperature or micro-climate deviated from the recommended temperature patterns, which allowed for significant reduction in communication data. These patterns were determined by collecting temperature data from healthy hives during different periods of time. However, the data used for determining these patterns was collected only from two hives at the same site, over a 40 day period. This raises questions about applicability of the proposed algorithm in other regions where environmental temperature variations may be very different, with different hives with different micro-climate patterns. But the overall approach of the authors and the experimental results hold a lot of promise.

Researchers have also used Circular Hough Transform object detection [91] to locate individual honeybees in video frames captured using IR CCD camera [35]. Bees from the hive under investigation were marked with unique circular tags placed on bee's thorax. A positioning dot on these tags was used by the authors to help identify of the orientation of characters on the tag. The extracted characters from tags were then segmented and classified using a support vector machine (SVM). This identification of characters on bee tag allowed the identification of each bee and the orientation of the tag allowed the authors to distinguish between outgoing and incoming bees.

Authors in this work [86] used the vibration and sounds inside the hive for classification of different states of hives. For this classification authors used Mel-Frequency Cepstral Coefficients (MFCC) [92] feature space plots on acoustic hive data acquired through MEMS microphone. The classification results with high accuracy show the aptness of low-cost MEMS microphone and electronics used in the system.

This work [87] performed a comprehensive analysis of data collected from beehives. Each system saved a 30-sec audio from inside the hive every 15 min, and these files were later split into 28 audio clips of 2-sec duration each, with a 1 second overlap. Authors obtained the ground truth by listening and manually labeling these 2-sec audio samples. Three human listeners divided these recordings into three categories i.e. bee buzzing, cricket chirping, and ambient noise. Classification was performed on raw audio data using different machine learning techniques. The authors provided results for these classifications, with Random Forest generally providing good accuracies. They also evaluated the classification of these datasets using automated feature engineering. These features preformed differently on different datasets, the reasons for which were not discussed by the authors. One reason could be that the audio data from hives changes significantly over different seasons, and different feature extraction techniques suit different seasons, hence providing the variation in results. The authors also tested different models for classification of images as bees, shadows of bees and no bees. This required extensive manual labeling of image data. From 2020 onwards, the authors also analysed the traffic of honeybee using video data, and classified it as incoming, outgoing, or lateral. This work also investigated the possible correlations between audio and video bee traffic and weather conditions at the hive site.

Majority of the systems analysed in this study have not used machine learning extensively, as the emphasis of most researchers was on developing a system and collecting the data. However some studies have used machine learning on existing beehive datasets for various tasks. One of the recent work on swarm detection [93] uses Deep Recurrent Autoencoders to detect the anomalies in beehive data. The dataset used in this study was collected from multiple hives using multiple temperature sensors per hive, along with weight, humidity and carbon dioxide sensors. Using Long Short Term Memory (LSTM) based Autoencoders, authors tested different sensor configurations, and also investigated the impact of sensor placement on swarm detection. The authors conclude that Autoencoders can be used to detect other anomalies in the hive as well. Later chapters of this thesis will discuss how existing machine learning models designed for various tasks can be adapted for tasks related to beehive monitoring.

2.3 Commercially Available Beehive Monitoring Systems

This section will analyse a number of commercially available beehive monitoring systems. These monitoring systems use a variety of sensors, different placements within the hive, various modes of communication, measuring frequencies which at times are configurable, and have different power options. Some of these systems offer modular approach where the user can opt for different modules to have different features, which to a large extent determines the cost of the overall system. A few of these systems (i.e. BeeBot, Broodminder, BeeMate) were tested to a limited extent in our hives, but it was not possible to practically test all of the commercially available systems. So we rely on the information provided by the manufacturers for this analysis. Some manufacturers do not provide the technical details of these systems on their websites, and some require a request to quote the price. Arnia is a popular beehive monitoring system, however we were unable to get a response from the manufacturers about the technical specifications and a cost breakdown of the system.

Bee Hive Monitoring [16] is a Slovakia based manufacturer of monitoring systems and they offer three different modules. First module which comprises three sensors is called ‘Heart of Hive’, which is placed on top of the honey frames. This module costs US$ 56 and can measure the inside temperature, humidity, and sound. This battery powered module can run upto one year on a single fully charged cell. Second module, the hive scale, comprises two wooden planks with load sensors inside for weight measurements. This scale is capable of measuring upto 200 kg in weight. These two planks are connected with each other using a removable cable which allows for ease of use as the distance between planks can be adjusted as per width of the hive. This also makes the weighing scale relatively cheaper, costing around US$ 100, but results in a low measuring accuracy of ± 1 kg. However it has a long battery life of 2 to 5 years. Both of these modules are equipped with a short range communication module (not specified, but likely bluetooth), which enables connection with a smart phone to read data directly from modules. Using the data collected from the sensors and hive scales, the system can generate alerts for events such as loss of the queen bee and robbing. Manufacturers claim to be able to predict swarming weeks in advance, alongside estimating the strength of the colony during winter. Third module offered is the GSM based gateway, which can be used to transmit the data to a server using a cellular network. This module costs US$ 101 and enables access of hive data using the internet.

Table 2.6: Feature summary of some commercially available beehive health monitoring systems, along with their cost in USD.

Commercial System	Temperature	Humidity	Weight	Acoustics	Other	Communication	Cost
Bee Hive Monitoring	✓	✓	✓	✓	GPS Tracker	GSM	$313
BeeBot	✓	✓	✗	✓	-	WiFi	$189
Broodminder	✓	✓	✓	✗	-	Bluetooth	$240
Hivemind	✓	✓	✓	✗	Bee Counter	Satellite	$1,140
EyesOnHive	✗	✗	✗	✗	Video Monitor	WiFi	$425
BeeMate	✓	✓	✓	✓	Video Monitor	WiFi	$580

The battery life of this module depends upon the measuring frequency of sensors, quality of cellular network, and can vary between 3 to 12 months. Manufacturers also offer a GPS tracker for US$ 56 to track the hives in case of theft.

BeeBot [19] from Pollenity is another European system which uses a single module for the monitoring of hives. It is capable of measuring temperature, humidity and records the buzzing frequencies of bees inside the hive. It is packaged in a wooden casing to make it less intrusive for honeybees, and its dimensions allow for an easy fit to a honey frame. It transmits the sensor data to the cloud using local WiFi network, and the measuring frequency can be configured as per user needs to optimize power usage. It comes with a rechargeable battery, and can run on a single charge for 6 to 8 months. It costs US$ 189, however the manufacturers provided us 10 of these beebots, free of cost for research purposes. These systems and the online 'BBoard dashboard' for data visualization have been insightful for our understanding of some of the honeybee behaviour. Figure 2.1 and Figure 2.4 show temperature and humidity data acquired through BeeBots placed in our hives in 2019.

Broodminder [20] is a beehive monitoring system from USA. The BroodMinder Citizen Science package includes three modules in total, two for measuring temperature and humidity each, and the third for measuring the weight of the hive. The sensors are placed either at the top, or at the bottom of the brood chamber and are accurate upto 0.5 degree Fahrenheit and 3% relative humidity. The weighing scale can be placed either at the front, or at the rear end of the hive. One of these systems was tested in our hive, and to keep the hive balanced, a support of same height had to be added under the rear end of the hive. In case of a perfectly balanced hive, the scale effectively measures half of the hive weight. The weight scale can support a total hive weight of upto 180 kg, with a resolution of 5 grams. This temperature compensated module is powered using a coin cell battery, which lasts for six months. This complete package costs US$ 240 and uses bluetooth to connect to the user's smart phone. Establishing bluetooth connections with the device/scale using the android application was found to be troublesome. To allow for data transmission to the server and for continuous monitoring, Broodminder also offers 'BroodMinder-HUB' with two variants, Cellular and WiFi, which was not tested. The Cellular version uses a mobile network to connect to the internet and costs US$ 299 plus US$ 99 for cellular subscription and MyBroodMinder dashboard premium. The WiFi version costs US$ 299 plus MyBroodMinder dashboard and makes use of local WiFi network to connect to the internet. The subscription for MyBroodMinder dashboard costs US$ 54 to US$ 108 for a single Cellular/WiFi hub per year.

Hivemind [18] is the brainchild of Brush Technologies, a New Zealand based internet-of-things design company. In contrast to other monitoring systems, Hivemind uses a satellite hub to transfer hive data to the server. This allows for remote deployability of this system in areas with no cellular coverage. The sensing system 'Hive strength monitor' can measure temperature and humidity with a resolution of 0.1 °C and 1% respectively, and has an accuracy of 0.2°C for temperature and 3% for humidity. It also reports the number of bees entering or leaving the hive every 3 hour periods, which can be adjusted for different durations. This monitor costs US$ 210 and is powered by 2 AA 1.5V replaceable Lithium or Alkaline batteries with a battery life of up to 1 year. It communicates the data to a satellite hub, which has to be within a proximity of 10 to 50 meters. The hive weight scale [94] has a very similar design to that of Bee Hive Monitoring (discussed above), and requires the two parts of weighing scale to be placed under the hive for measurements. This scale costs US$ 300 and has a measuring resolution of 100 grams, with an accuracy of $\pm$1% kg + 2% of hive weight. Maximum measuring weight of the scale is 300 kg and is powered by 2 AA 1.5V replaceable Lithium or Alkaline batteries, which can last up to 1 year. This module reports data to server via satellite hub, or can act as a stand-alone device by speaking the hive weight at the push of a button. The satellite hub comes with a built in GPS, has a default rate of transmitting 4 reports per day, and can connect with upto 25 devices. It is powered by 4 AA 1.5V replaceable Lithium batteries and costs US$ 630.

EyesOnHives Scout B [52] is an imaging based monitoring system developed by Keltronics, USA. This system uses a camera and a computer, packed together inside a single module, placed one to two feet away from the hive and pointing at the hive entrance. It counts the bees going in and out of the hive, alongside measuring the flight behaviour of the hive. Colony activity patterns are tracked over time using computerized video analysis and hours of video is translated into an easy to understand, visual, real-time summary for beekeepers. However this system requires 120V power for the computationally hungry hardware and WiFi connectivity to transmit the data to the user. Despite having limited deployability for remote monitoring, this system costs US$ 425 excluding tax.

BeeMate [95] is an Australian beehive monitoring system powered by Artificial Intelligence (AI). The primary functionality of this system is video monitoring, but can be purchased with the add-ons of temperature, humidity, sound and weight sensors. The video data from each hive is analysed using deep neural networks to estimate the forager traffic. This information along with other sensor data is available to the users using a web dashboard. Users can also view the live streaming of video from the hive entrance. The system uses WiFi connection to transmit the high bandwidth data to the web server. It also requires access to 220 volt AC power to operate. Two of these systems were tested in our hives, and the forager traffic analysis using the dashboard was useful in validating some of the data collected using our own sensor systems. However connecting WiFi based devices to University WiFi network was challenging. BeeMate is also part of The Sentinel Hive Network, which is deploying over 1,000 AI powered Smart Beehives globally by the end of 2022 in locations of strategic bio-security importance.

To summarise the major aspects of commercial beehive monitoring systems, most of the commercial systems deploy sensors in a single package inside the brood chamber. The hive scales are consistently the most expensive component of monitoring systems, and their placement under hive adds to the complexity of the design. Most systems only have short range communication capability, with a separate WiFi to 4G gateway required at the hive site for long range communication. But these gateways are expensive to purchase, require additional monthly/annual subscriptions, and can only support a limited number of monitoring systems at the hive site. And 4G gateways are very dependent on cellular coverage, which is a problem in remote regions of Australia. For power, most systems use replaceable batteries, with some relying on AC power. None of the discussed commercial systems use solar panels to power the beehive monitoring systems.

2.4 Research Gaps

This section discusses the gaps in existing knowledge on beehive monitoring. First and foremost is the use of a small subset of parameters for experimentation in this area. Different researchers focus on different parameters of hive monitoring, based on their area of expertise. Little attention is given to experiments which evaluate a wider set of parameters and their various combinations for different tasks. It is not practically feasible to test a huge number of parameters in a single study, but it is required to be as thorough as possible. The complex behaviour of honeybees demand the use of multiple parameters in BMS to get an accurate picture. Each parameter can help identify and/or predict only limited aspects of bee health. Using multiple parameters not only gives a more detailed picture about the bee health and activity, but also helps decrease the impact of inaccuracies in the measurements of different parameters. Proper evaluation of sensors for different tasks of beehive monitoring is very important aspect of system design, and largely missing from the literature.

While it is important to investigate different sensors for use in beehive monitoring system, it is also vital to improve the cost effectiveness of the system. The cost of monitoring system increases with the increase in the number of sensors. The low cost systems proposed in literature often compromise on the number of parameters which are monitored. One of the most important parameters of a beehive is its weight, which requires an expensive scale. It is very difficult to add a reliable weighing scale to a cost effective design, as beehive scales usually cost more than the rest of the monitoring system put together. Up until now, no work has been done towards estimating the weight of the hive using other not-so-expensive sensors. A system capable of estimating the hive weight with reasonable accuracy will significantly reduce the purchase costs of monitoring systems.

Most of the experimental setups reviewed in this study used only a few hives, and at times only a single hive to collect the beehive sensor data. We understand and have experienced the issues involved in experimenting with multiple hives, especially those related to beekeeping. These issues are almost, if not always, outside the expertise of researchers. The time required in the field to manage the hives, maintain the monitoring systems and the logistic complexity increases multiple folds with the increase in the number of hives. However, the data collected from experiments is not reliable unless it is from multiple hives deployed in varying conditions. Deploying multiple hives for experimentation is only possible if there is a strong collaboration between beekeepers and the researchers. Both working closely together is the only way to setup experiments and collect reliable data on a wide scale.

Another major problem in this area is the absence of any standard framework for experimentation and data collection. Every experimental setup analysed had its own desired placement of sensors, number of sensor systems, and the frequency of reading the sensor data, based on available resources and personal preferences. Some setups mentioned the details of equipment used, the environmental conditions, location of experiment, time of the year and some didn't. This means that the data collected from one experimental setup cannot be merged with data collected from another. Most of the researchers in this area do not make their experimental data publicly available. A few datasets which are publicly available are very different in composition from each other. This makes it almost impossible for anyone to use the data from multiple experiments and come up with a bigger and better picture about bee health/activity. There is a need to devise standard practices for collection of data and its labelling, so that data shared by different researchers can be used together. This will improve the quality of research in this area and speed up the progress towards better beehive monitoring systems.

The problems discussed above lead to another major issue faced by researchers working towards smart beehive monitoring systems. One of the aims of smart beehive monitoring is to generate assessment about the state of the hive, similar to that of a good beekeeper. For this, a supervised machine model requires the assessment of hives from experienced beekeepers. The model then uses the sensor data from hive, and learns to relate this data to the assessment of the hives provided by the beekeepers. Supervised machine learning requires a good quality and quantity of labelled data to train a model capable of classification of current hive state and/or the prediction of future hive state(s). The role of beekeepers in providing the assessment of hive, also known as ground truth or data label, is very important. Without this ground truth, even a sensor dataset of high quality and quantity cannot be used to train a machine model for hive state classification. And for a well trained machine model, the ground truth must come from multiple beekeepers, variety of hives in different states of bee health, spread over a large area to add the impact of various environmental conditions on the hives. Collaboration between researchers and beekeepers is essential for the use of advanced machine learning tools and algorithms for beehive monitoring.

Beehive monitoring systems have not been investigated properly for the identification of the bee disease(s). Most of the work in literature is on classification of different states of hives. When it comes to bee diseases, the current focus of monitoring systems is on detecting the symptoms of diseases. Once these symptoms are detected by the system, only then the beekeepers are alerted so they can manually intervene and identify/verify the disease, and eliminate if it is not too late already. This is a desired feature to have but definitely not the end goal. There is a need for further research to identify the disease inside the hive in its very early stages, or preferably predict it in advance with the help of machine learning. Authors in this detailed survey [96] also emphasize on the integration of efficient, operational, and deployed AI models. This will enable beekeepers to intervene in a timely and effective manner and avoid the loss.

Most of the gaps in the research of beehive monitoring are not easy to fill. For example, any data collection from beehives with disease has to be carried out under strict bio-protocols [97], for which the involvement of experts/researches of honeybee biology is required. Placing sick beehives indoor in a controlled environment alters the bee behaviour, and the data collected by electronic sensors in such setting is biased. Out in the open, honeybees visit a vast area around the hive to collect pollen and nectar. And it is very likely that bees from multiple hives will visit the same resource. Thus, the chances of bee disease spreading from one hive to another are very high. For collection of data from a sick hive in the open, it is important to place that hive where chances of bees crossing the path of other domestic bees are minimum. For this, one has to survey the area to ensure that other beekeepers have not placed their hives near by. Still, there is a chance of infecting the native bees present in that area, and some bee diseases can be devastating for native bee population.

Disease such as 'American Foulbrood' [98] has no cure available, and the bio-protocols require that the infected hive, with all the bees and hive components should be destroyed [99], with burning as a standard practice. If a monitoring system is used to collect data from such a hive, that monitoring system should also be burned, a sacrifice most of the researchers are not willing to make. Most of the experiments with sick hives need multiple repetitions of experiments, over different seasons, and different environment conditions for reliable collection of data. Thorough study and data collection of a single bee disease requires multiple years of carefully designed experimentation and data collection. Moreover, the working methodologies and objectives of researchers working on biology of bees, those working on electronics of beehive monitoring, and data scientists are very different. And then the professional beekeepers operate in a manner which is very different from scientific community. For a comprehensive study of bee diseases using electronic monitoring systems, a great level of willingness to collaborate is required among all the stake holders.

Addressing all these limitations and research gaps in a single study is not possible, and some core problems were identified for investigation in this work. These problems are related to the design of monitoring systems, as progress towards better monitoring systems will also create pathways towards solving other problems that have been identified. This work on the electronic beehive monitoring system aims to answer these three questions.

- Which sensors should be used to design a low-power and long-range beehive monitoring system for remote regions?
- Is it possible to reduce the system design cost by using soft sensor prediction to replace expensive/difficult to use sensors such as weighing scale?
- Can machine learning algorithms assist in the selection of sensors for specific tasks, and help fine tune the design of beehive monitoring systems?

To answer these questions, we design, develop and thoroughly test a monitoring system, capable of collecting data from beehives located at remote sites. The sensors for this system are selected after a careful study of systems proposed in the literature, and those commercially available. The goal is to use sensors which measure hive parameters of significant relevance, as well as facilitate the collection of data in a reliable manner. The system design is scalable to allow collection of data from multiple hives, and supports deployment in remote areas to collect geographically diverse data. A sensor dataset of high quality and resolution is collected using deployments of multiple sensor systems. The design of this system, the data collection methodologies, details of experimental setups, as well as the collected dataset are shared with the research community to facilitate reproducibility.

To test the effectiveness of designed system, sensors, and the collected data, machine learning tools and techniques are utilised in this work. As discussed in previous sections, one of the most expensive and difficult to measure parameter of a beehive is its weight, yet the weight variations are a very good indicator of honeybee activity and hive health. In this work, deep machine learning models are employed to estimate the daily weight variations of hive using the data collected from sensors deployed inside the hive. The proposed models are capable of estimating the daily weight change/variations of hive with good accuracy. The machine models are also used to identify the most important sensors for weight estimation, as well as the times of day which are crucial for data collection. The monitoring of different hives states, or those with disease is beyond the scope of this work. However the promising results of weight estimation indicate that the designed system can be used to collect information rich data from beehives, and help advance the research of beehive health monitoring.

Chapter 3

Design and Development of Low-Power, Long-Range Data Acquisition System for Beehives - BeeDAS

Chapter 2 evaluated different sensors used in beehive monitoring and analysed some monitoring systems from the literature. It also examined some commercially available beehive monitoring systems. Based on that evaluation and analysis, this chapter discusses the design, development and deployment of a multi-sensory, remote data acquisition system for beehives (BeeDAS). The focus of this system is on low-power consumption and long-range communication. The proposed system enables collection of data from beehives at remote locations and harsh environment. Results of field deployments elucidate the effectiveness of various sensors which measure temperature, humidity, atmospheric pressure, CO_2, acoustics, vibrations and the weight of a hive in hostile environment. This chapter addresses the design challenges associated with such systems and highlight the critical issues that need consideration such as sensor placement, power optimization, sleep intervals, noise filtering, calibration, feature extraction from sensor data and the data storage. These findings will help improve all kinds of data acquisition systems designed for remote deployment. The collected dataset and the system design files are also made public in a repository for the scientific community to build upon this work [100].

This chapter also uses random forest regression to evaluate the feature importance for the task of estimating the daily hive weight change. This importance is evaluated for hive sensors, environmental variables such as temperature, humidity, rain, wind speed, and the information related to seasons, on a dataset comprised of 1,250 days of sensor recordings. The protocol designed for communication using Narrow Band Internet of Things (NB-IoT) is also evaluated. The primary objective of this system is to collect data which can be used to improve the decision making capability of beehive monitoring system. Decision making capability of any system is highly dependent upon the quality and quantity of training data. Most of the data acquisition systems designed by researchers are designed to collect data through a small set of sensors, and capable of short-term data collection, with deployments restricted to sites with access to power and short-range communication. Data collected by such systems lack both temporal and spatial diversity. This hinders the development of a dataset that can be used to effectively train machine learning models. This chapter not only investigates an appropriate design for monitoring systems, but also provides a platform to collect reliable data from beehives.

3.1 Background

Data acquisition systems are a common tool for monitoring a wide range of phenomena in industry, agriculture, healthcare, entertainment, transportation and sports. Automated decision making is facilitating many aspects of human life [101, 102, 103], and the data acquisition systems play a fundamental role in collection of the essential data for the training of machine learning algorithms. However, deployment of these systems in remote areas is a big challenge because of power, maintenance and communication constraints [104]. Inaccurate sensor data and unreliable communication from remote sites make precise and timely decision making very difficult. In this chapter, the design of a multi-sensor data acquisition system for beehives (BeeDAS) is presented as a platform to consider these constraints. A total of 8 data acquisition systems were deployed at different locations to test the reliability, power consumption and communication aspects of the designed system.

BeeDAS is capable of stand-alone operation without human intervention. The power efficient and robust design of the system is complemented with long-range communication capabilities using Long Range Wide Area Network (LoRaWAN) and Narrow Band Internet of Things (NB-IoT), allowing round the year collection of sensor data from remote regions. Round the clock monitoring capability of such systems is important because some bee diseases can spread very quickly from hive to hive [105]. The proposed system transmits the data as it is collected, and can alert beekeeper about major hive events within minutes.

This chapter also uses machine learning to evaluate the performance of BeeDAS. The daily weight change of a hive provides a very good valuation of beehive strength and bee activity [32]. This daily weight change is used as a benchmark to evaluate the in-hive sensors of BeeDAS on the basis of their contribution towards predicting/estimating this change. The *sensor feature importance* towards this estimation is directly proportional to the usefulness of that sensor for beehive monitoring. Random forest [106] is used to evaluate this importance. Random forest is a supervised learning algorithm which makes use of ensemble learning, where multiple learning algorithms are used in parallel to obtain a better prediction. Random forest is an ensemble of decision trees [107], usually provide high accuracy, allow the assessment of importance, and are resistant to over-fitting. This makes random forest a good candidate for evaluating sensor feature importance for the task of estimating the daily weight change.

This chapter aims to:

- Design, develop and deploy a low power beehive data acquisition system.
- Investigate a unique set of sensors and efficient feature extraction techniques to generate minimal amount of data to reduce transmission costs.
- Evaluate the sensor feature importance towards beehive weight estimation using random forest, on the collected weather and sensor data .
- Test the performance of NB-IoT for long range data communication from remote hive sites.

3.2 Design Requirements

There are several challenges specific to designing of a beehive data acquisition system as discussed in previous chapter. Improper deployment of electronics in a hive can result in an altered behaviour of the bee colony, thus resulting in collection of biased data. The basic structure of the most commonly used Langstroth hive consists of several chambers, with multiple frames in each chamber as shown in Figure 2.2. The frames inside a hive are stacked both vertically and horizontally, which combined with the complexity of the beehive, makes the placement of sensors difficult without input from experienced beekeepers.

The relative humidity inside the hive is usually above 50% [40], which not only impacts the readings of some sensors but can also cause corrosion in the long-term. Honeybees have a tendency to cover any alien objects in the hive such as exposed sensors with bee glue i.e. propolis [108], which impacts the performance of sensors. A completely sealed system to protect sensors from bees and propolis would fail to collect accurate temperature, humidity and gas data from the beehive. Any attempt to partially cover the sensors results in hive pests, such as small cockroaches, finding a safe refuge under this cover and using it as a breeding ground. From a wireless communication point of view, the honey in the hive attenuates the radio signals, limiting the placement options for systems with internal antennas.

Beekeepers move their hives frequently to different sites in search of honey, or to provide pollination services. A lot of these sites are remote, with no power and communication available. Since the use of long-range communication technologies such as LoRa and/or NB-IoT limit the amount of data that can be transmitted, it is not possible to use sensors which generate large quantities of data. For example, a video camera is very useful for hive monitoring, but is power hungry and also generates a large amount of data. Transmitting such high bandwidth video data from a remote site can often exceed the channel capacity. The alternative of onboard processing to extract features from video data requires significant power and computational resources, making video cameras generally a poor fit for remote, power limited monitoring systems.

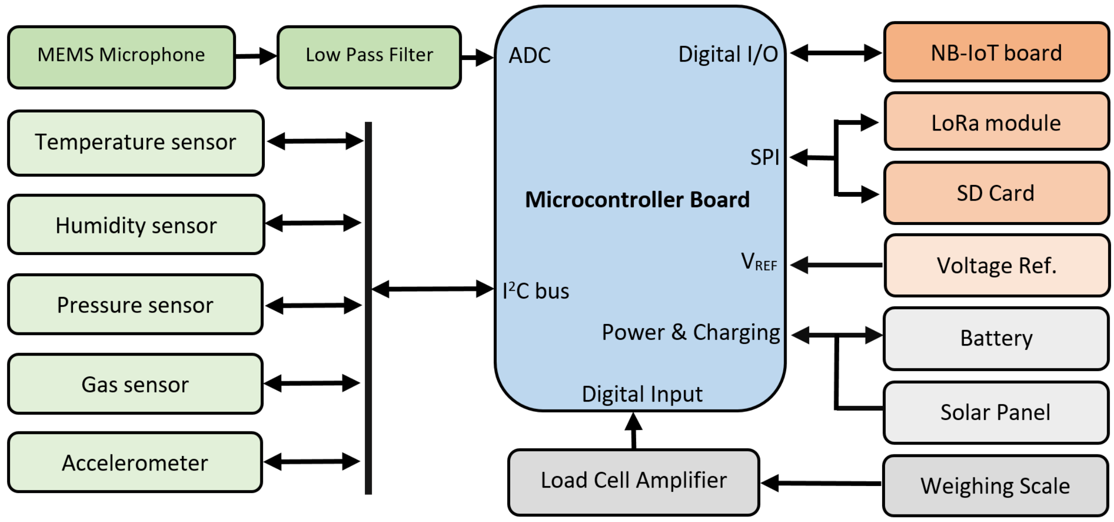

Figure 3.1: Block-level diagram of BeeDAS, with all the sensors, communication boards and their respective interfaces with the microcontroller board.

3.3 System and Sensors

BeeDAS comprises 6 internal hive sensors (sound, temperature, humidity, pressure, gas, vibration) and an external weighing scale as shown in block-level system diagram in Figure 3.1. BeeDAS uses a Sparkfun Redboard Turbo board equipped with a 32-bit/48MHz ARM Cortex-M0+ Micro Controller Unit (MCU). This MCU comes with a 12-bit Analog to Digital Converter (ADC) and 32 kB of SRAM, and provides enough computational power to extract features within hive. An I^2C bus is used as a primary communication interface for most of the digital sensors in the system. BeeDAS is configured to collect data every 10 minutes. A two-layer interconnecting PCB was designed and fabricated for interfacing the MCU board, sensor boards and the communication boards. The approximate cost for a single BeeDAS unit is around 200 AUD, which includes NB-IoT board and LoRa module. The weighing scale costs an extra 500 AUD to manufacture. The PCB schematic and board files of BeeDAS, along with the sensor dataset are available as a github repository [100]. The online links of relevant datasheets are also available in the README of this repository.

3.3.1 Temperature

Modern temperature sensors are small, accurate, and very power efficient, which makes their use in monitoring systems much easier. The temperature in brood chamber varies between different frames. Honeybees are able to regulate the temperature of frames in the middle of the brood chamber much better compared to the frames towards the outer edges [109], where the outside temperature variations have a greater impact. Also, the area inside the hive where bees can maintain the temperature around 35°C is directly proportional to colony strength (number of bees in the hive). Most commercial monitoring systems use a single temperature sensor towards the middle of the hive. Such placement for bulky systems can be intrusive, and adversely impacts the brood rearing of hive. A good strong hive can raise brood on all frames, whereas a weak colony only raises brood in the few middle frames of brood chamber. This means that a temperature sensor in the middle of brood chamber is unable to provide information about the number of brood frames.

The design used a BME280 by Bosch, which has temperature, relative humidity and atmospheric pressure sensors of 16-bit resolution in a single package. The temperature sensor can measure from -40 to 85°C, with a tolerance of ±0.5 °C at 25°C. As shown in Figure 2.2, BeeDAS is deployed at the edge of outer most frame in the brood chamber, to avoid the use of precious brood space in the middle of the chamber. But at this position, the temperature readings are also influenced by the temperature outside the hive. Through experimental deployments, it was observed that the temperature variations at this position decrease in magnitude as the brood area inside the hive expands towards the outer frames. Using the magnitude of these temperature variations and the external temperature as a reference, number of brood frames in the hive can be estimated.

Challenges: Deploying extra sensors to collect the external data adds to the cost, complexity and size of the monitoring system. We record the external temperature, temperature feel, rain, humidity and wind-speed data reported by the Bureau of Meteorology (BOM) Australia [110], using an Application Programming Interface (API). This weather data is reported every 15 minutes, and linear interpolation is used to re-sample it with a 10 minute interval to align it with sensor data from hives. As this weather data is available only for each weather station in Australia, it can lack accuracy for locations far away from these stations. If weather data is required for a large number of BeeDAS deployed at a particular site, having a dedicated local weather station may improve the overall performance.

3.3.2 Relative Humidity

Relative humidity is the percentage of moisture in the air, against the maximum possible level of moisture at that temperature [111]. Even though humidity sensors are less accurate compared to temperature sensors, they are still good enough for monitoring beehives as honeybees can tolerate humidity variations much better than temperature variations. As discussed in the previous section, BME280 is used as a humidity sensor which can measure relative humidity from 0 to 100%, with a ±3% accuracy tolerance.

Configuration: The humidity sensor of BME280 has a built-in IIR filter which can be used to reduce noise when data is sampled at a faster rate such as every second. Since BeeDAS has been designed to be power efficient, the forced mode is used where the BME280 sleeps without collecting any data unless forced to do so. With a 10 minute interval for data collection, we do not use the built-in filter as it results in loss of information.

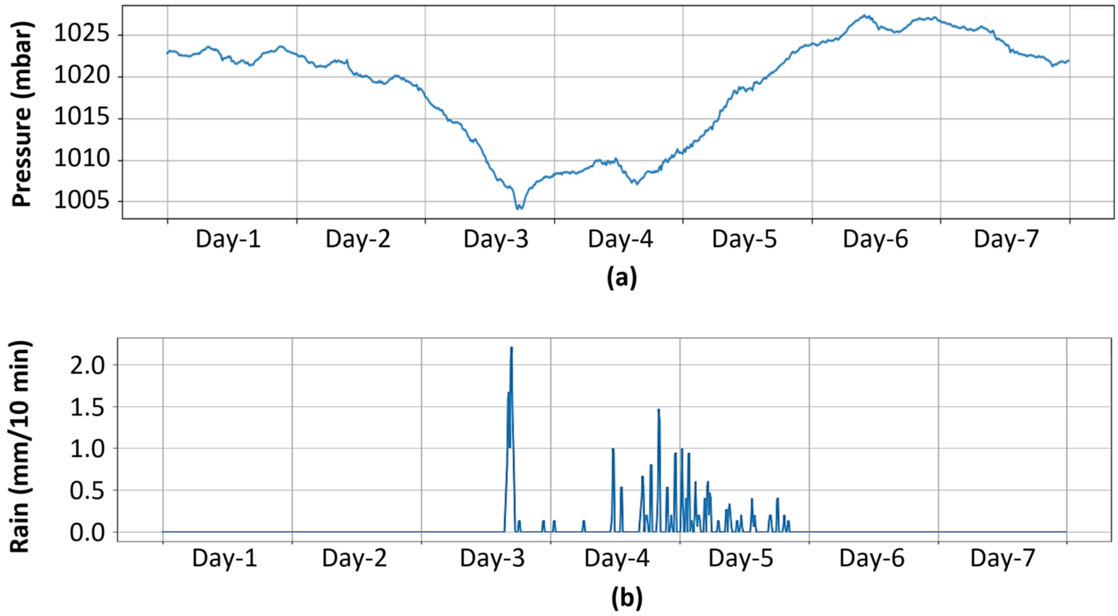

Figure 3.2: (a) The atmospheric pressure data recorded from a hive located at The University of Western Australia (UWA), Crawley campus from 20^{th} to 26^{th} of May, 2021, where each tick on the x-axis of graph represents midnight. (b) Recorded rain in Crawley for the same duration shows that the rainy days coincide with low atmospheric pressure.

3.3.3 Atmospheric Pressure

The pressure sensor in BME280 can measure atmospheric pressure between 300 mbar (milli-bar) and 1100 mbar, with RMS noise of 2 mbar. The temperature offset coefficient for this sensor is ± 15 mbar/°C. Our initial aim was to use the pressure sensor to determine the altitude of hive. To precisely determine the altitude, a reference pressure reading is required at the sea level, which is not always available. Due to the lack of a pressure reference, we did not include the impact of altitude in this analysis.

Observations: Low atmospheric pressure readings from hive sensors are often accompanied by heavy rain, as seen in Figure 3.2. Bees stop foraging activity during rain, which means that pressure has an indirect relationship with the bee activity. However, light rain is not always marked by low atmospheric pressure.

3.3.4 Acoustics

BeeDAS collects audio data using an analog MEMS microphone ADMP401, manufactured by Analog Devices. This microphone has a flat frequency response from 100 Hz to 15 kHz, and a signal-to-noise ratio (SNR) of 62 dBA. The microphone board comes with an amplifier which provides a gain of 66 to the audio signal. Despite this gain, the audio signal does not utilize all of the ADC bits during normal bee activity. An amplifier with a bigger gain can be used to utilise the ADC to its maximum, however that will create problems when the beehive goes into an agitated state. The amplitude of bee buzz increases multiple folds in

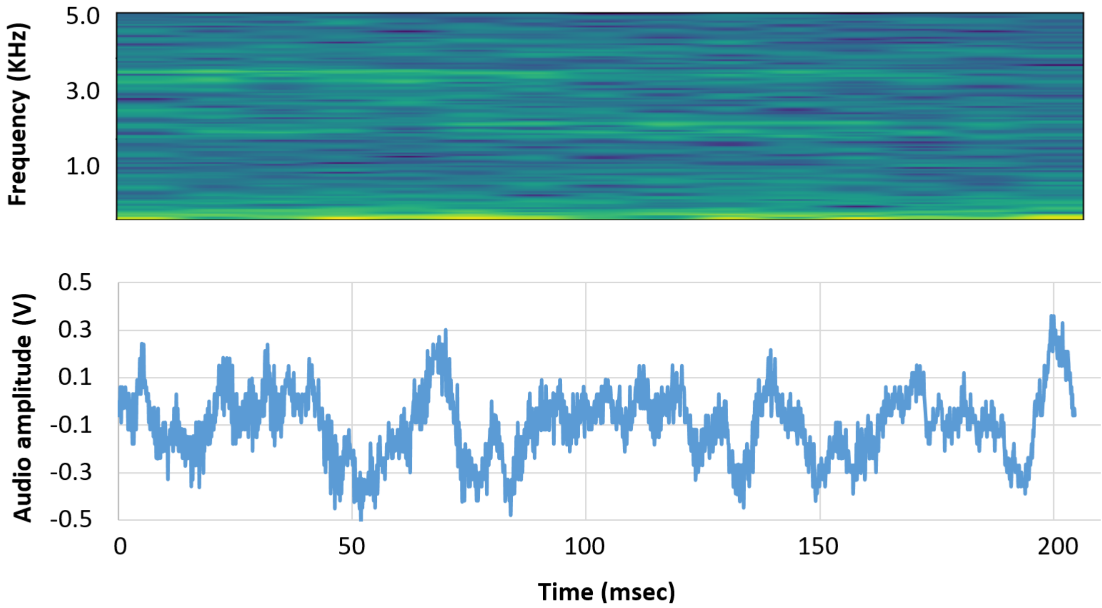

Figure 3.3: An example of noisy bee-buzz, sampled at 10 kHz using the built-in voltage reference of microcontroller for the ADC.

such a state and the high gain of amplifier can result in clipping of the audio signal. It is important to select an appropriate gain for audio signal to ensure the signal does not change its shape during infrequent events.

The 12-bit ADC of the MCU was used to sample the audio data. A key to high accuracy ADC sampling is a good, stable reference voltage. The built-in voltage reference for the ADC was observed to induce excessive low-frequency noise as shown in Figure 3.3. As a

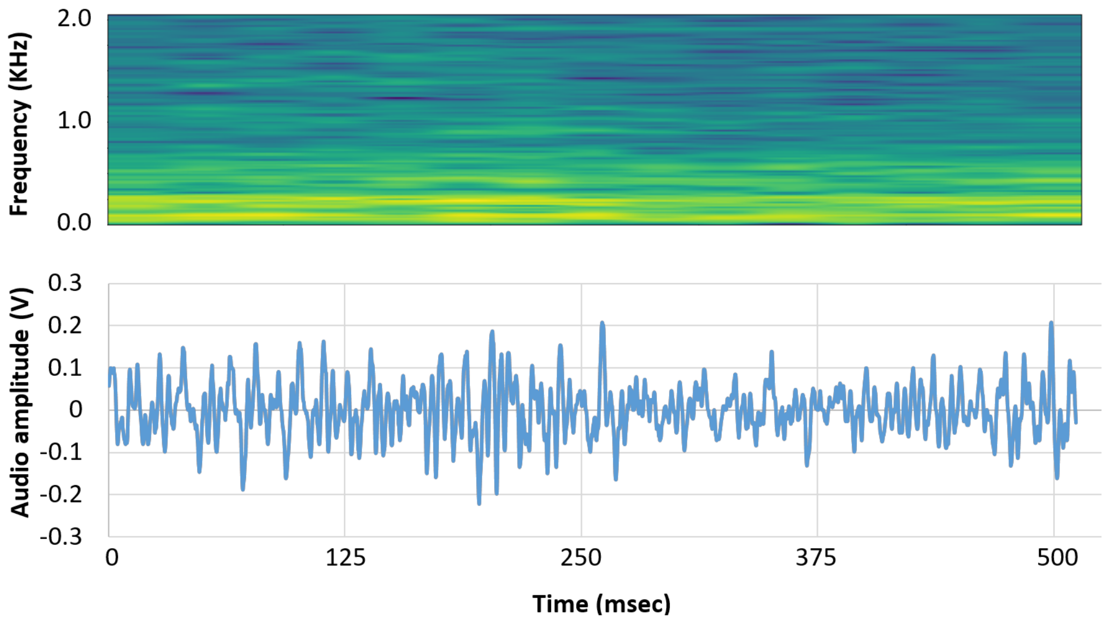

Figure 3.4: Example of a clean bee-buzz, sampled at 4 kHz with a 5 kHz LPF, using REF3433 as voltage reference for the ADC. The use of LPF and stable voltage reference has a significant impact on the audio quality.

result, an alternative voltage reference (REF3433) by Texas Instruments was used which produced a stable 3.3-V, low-drift voltage reference. A clean bee buzz is vital for reliable feature extraction [112], and a significant improvement in the audio quality was observed by switching to the external voltage reference. Using a second order, analogue RC Low Pass Filter (LPF), with a 5 kHz cut-off further improved the audio quality by filtering the noise, as seen in Figure 3.4.

Operation: The ADC of the MCU can sample up to 48 ksps (kilo samples per second), making it possible to record frequencies up to 24 kHz (via Nyquist criterion). However the limited memory available on the MCU restricts the number of samples that can be collected in a single burst of recording. Our studies found that it was important to record the audio signal for a minimum of 500 msec of bee buzz to extract key features. Bee buzz audio was investigated up to frequencies of 2 kHz (sample rate at 4 ksps), recording 2048 samples of audio every 10 minutes.

Processing: The bandwidth required to transmit audio data is much smaller than video data, but is still significant. The MCU is used to extract useful features from audio data, and transmit only the extracted features using a 2048 point Fast Fourier Transform (FFT). These features include the peak amplitude of buzz, highest frequency component for the recorded buzz, and the Power Spectral Density (PSD) of buzz with a 100 Hz resolution. Each audio recording requires 4096 bytes of data (2 bytes per sample), but the feature extraction process reduces this data volume down to 24 bytes, representing the audio features collected within a minimum of 500 msec sample duration.

Placement: While placement of multiple microphones on frames in a hive would improve the data collection, it would create problems for the beekeepers during hive inspection. The decision to use a single microphone was driven by our need to facilitate beekeepers, while using ultrasensitive MEMS microphones with 12-bit ADC to help detect bee buzz even from weak hives.

3.3.5 Gas Sensor(s)

CCS811 gas sensor from Sciosense is used to evaluate the gas emissions in the hive. This 10-bit, low-power digital metal oxide (MOX) sensor is capable of detecting low levels of equivalent Carbon dioxide (eCO_2) and equivalent Total Volatile Organic Compound (eTVOC) in ranges typically found indoors. The term *equivalent* is used for describing different greenhouse gases in a common unit, so these values do not correspond to pure content but to everything similar. The eCO_2 output range for the CCS811 is from 400 up to 29206 parts per million (ppm) and the eTVOC output range is from 0 up to 32768 parts per billion (ppb). Because of their wide usage and large scale manufacturing, gas sensors have seen a significant reduction in size and drop in cost. However, most of these sensors are designed for indoor deployment.

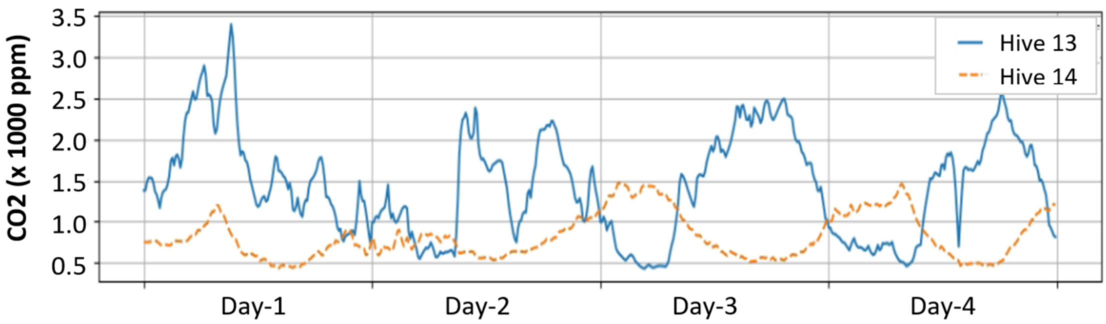

Figure 3.5: Sample of eCO_2 data collected form two hives with similar bee colony strengths, located at Jurien Bay, WA during late June 2021. Data shows contrasting patterns of eCO_2 fluctuations between the two hives.

Operation: The characteristics of MOX based sensors vary from sensor to sensor, and each sensor has a different baseline. CCS811 firmware is programmed to adjust its baseline at the system startup. It assumes that the sensor is in clean air at startup (not always the case) and uses the initial readings of first 20 min to stabilize and adjust the baseline. Within remote sensing systems that power down sensors to conserve power, this startup stabilization delay and baseline shift produce unacceptable variation in the collected data. The firmware is also programmed to re-calibrate itself on detecting high CO_2 levels, based on the assumption that high reading must be a result of sensor drift. To counter this, baseline records were obtained for each sensor in the lab environment, and the CCS811 was programmatically forced back to the baseline reference at regular intervals once deployed.

The CCS811 can be configured to operate in different modes to optimize the power usage. It can sample data every second, every 10 seconds or every 60 seconds. The greater the sampling interval, the lower the power consumption but this also lowers the sensor accuracy. This sensor achieves best accuracy when sampling every second, and hence some accuracy is compromised in a monitoring system designed to sample data much less frequently. Even though this sensor is able to detect eCO_2 and eTVOC, both of these readings were found to be highly correlated, carrying similar information.

Challenges: The cost of gas sensors is still high enough to limit their wide scale use in commercial monitoring systems. These sensors often have a much higher current rating compared to other sensors, which makes it difficult to use them in remote systems (see Table 3.5). These sensors also have a high drift and require frequent calibrations to adjust the baseline. It is difficult to regularly pull monitoring systems out of the hive, put them in a clean, stable environment for re-calibration, and then redeploy in hives. Frequent opening of a hive disturbs the bees and the environment inside the hive i.e., the temperature, humidity and composition of gases is lost every time a hive is opened. This makes the long-term deployment of these gas sensors difficult.

Reliability: Two major factors that affect the functionality of MOX based gas sensors are humidity and temperature [113], which vary from hive to hive. Figure 3.5 shows eCO_2

readings from 2 different hives, located at the same site. Even though the strength of bee colonies in these hives was observed to be very similar, the eCO_2 levels exhibit very different gradients, and don't follow a similar pattern. Despite considering this a good sensor based on the manufacturer's specifications, our study indicates that this particular gas sensor (and possibly this family of sensors based on MOX) may not be a reliable choice for hive monitoring.

3.3.6 Accelerometer

Accelerometer is included in BeeDAS to pick up the vibrations generated by the bee waggle dance. The MMA8452Q triple axis 12-bit accelerometer from NXP Semiconductors was used within deployed hives. This capacitive accelerometer is capable of detecting gravitational force in the range of $\pm$2g, $\pm$4g or $\pm$8g, depending upon the mode, and supports output data rates from 1.56 Hz to 800 Hz. The $\pm$2g scale was used in BeeDAS, as hives do not experience high g-forces. Datasets of 512 samples at 500 samples per second in each recording were obtained, which provided nominally 1 second of vibration data on all three axes every 10 minutes. Using the onboard MCU, the highest frequency component for each axis was extracted using Fast Fourier Transform (FFT), allowing vibrations in the range of 13 Hz and 250 Hz to be detected, with a resolution of approximately 1 Hz.

Placement Challenges: Honeybees have evolved to communicate using a waggle dance on hives made up entirely of wax. During this dance, the bees use a 15 Hz abdomen waggle and the 250 Hz thorax vibration [55, 58]. The composition of dance floor plays a significant role in determining the effectiveness of communication [114], and in commercial beehives made up of custom material, the wax is available only in the middle part of wooden/plastic frames. These frames are not ideal for the propagation of bee vibrations. In addition, the mounting screws for BeeDAS added further attenuation to the propagation of vibrations from the bees (through the wax and frame) to the accelerometer. If frames are allowed to remain at the same place for extended periods, the frames become glued together by the propolis produced by the bees, which improves the propagation of these vibrations. However, in commercial hives, the frames are regularly pulled out for honey extraction and bee inspection. Issues with the propagation of acoustic waves result in just a fraction of all vibrations inside the hive being picked up by the accelerometer. Further, accelerometers were configured to sample data at regular intervals to conserve power. The combination of periodic sampling and signal attenuation (given the sensor's resolution) result in majority of waggle dance events not being recorded.

3.3.7 Weight

Beehive weight scales are required to be rigid enough to support hives up to 120 kg in weight, sensitive enough to pick up changes of a few grams, and reliable enough to function outdoor in all weather conditions. For each designed weighing scale of BeeDAS, a total of 4 load-cells were used in a Wheatstone bridge configuration, with a maximum 50 kg load capacity per sensor. As shown in Figure 3.6, these load-cells are placed in indents designed as per size specifications to lock sensor position on the base aluminium frame. These scales were specifically designed for hives, where an aluminium lid on top of each scale ensures that the weight of the hive is transferred only through the 4 load-cells, and also protects the sensors from rain and direct exposure to sunlight. The size of the scale allows for easy placement of Langstroth hives (see Figure 2.2). The scale can measure weights upto 200 kgs, which more than satisfies the maximum weight requirements.

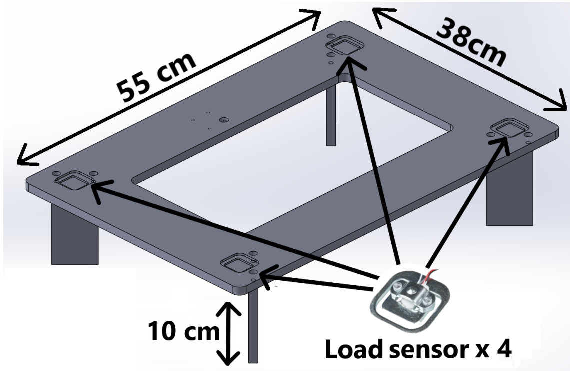

Figure 3.6: The 3-D model of the weighing scale designed for the beehives, with dimensions and load sensor placements.

HX711 from Avia Semiconductor is used as a load-cell amplifier and 24-bit ADC, which provides a resolution of 50 mg, with an operating range of -40 to 85°C. An average worker honeybee weighs about 100 mg [115], and theoretically this resolution should allow the measurement of the net number of bees entering or leaving the hive. Unfortunately, it is not practically possible to continuously monitor the weight of a hive in remote deployment, and weight is recorded every 10 minutes. During this interval, the net hive weight gain/loss is a result of foragers leaving the hive, foragers entering the hive, the amount of pollen/nectar brought in by the foragers, the evaporation of moisture from nectar, and the consumption of food by bees/brood inside the hive. On top of that, the variations in temperature and humidity, the accuracy of ADC, the exposure of hive to sun/shade also impact the weight readings. This makes it impossible to translate the weight change into exact number of bees participating in the foraging activity. Other methods such as camera monitoring at the hive entrance are much more suited to measuring the number of forager bees, however the power consumption made their use in this remote environment unfeasible.

Operation: The HX711 load-cell amplifier is powered down when not in use, and the readings of initial few seconds of ADC are discarded after it is powered up to allow for the warm-up time. To minimize the conversion error, several readings from ADC are recorded in multiple small bursts, from which the average of each burst is calculated, and the median value of these averages (ADC_avg) is used to calculate the weight using:

$$Weight = (Sensitivity \times ADC\_avg) - Offset \tag{3.1}$$

where *Sensitivity* is the coefficient to convert ADC readings to weight in grams, and

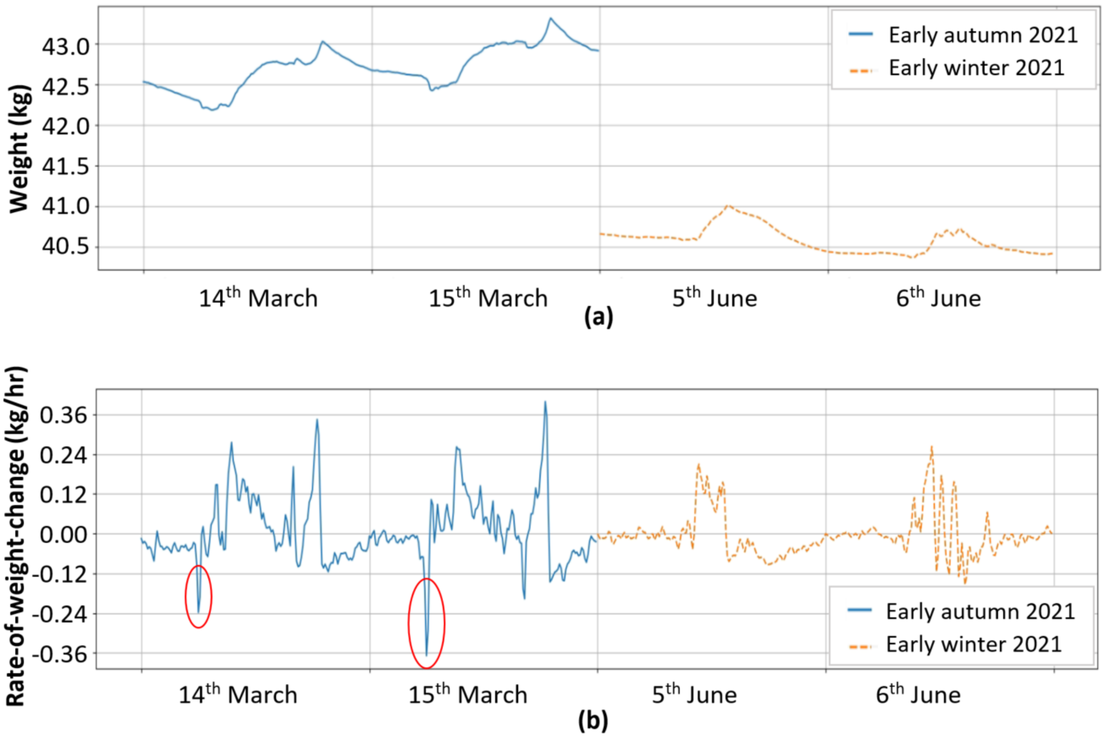

Figure 3.7: (a) Daily weight variations from a hive during two day period in March 2021 (early autumn) and two day period in June 2021 (early winter), located at The University of Western Australia (UWA), Crawley campus. (b) The difference in the rates of weight change for the same hive during two different seasons is an indicator of different levels of foraging activity. Red ellipses highlight sharp changes in weight due to many foragers simultaneously leaving the hive in the morning.

the *Offset* is the error in grams with no weight on the scale. The use of resistive load sensors with different manufacturing bias means that every weighing scale needs individual calibration. The *Offset* values not only vary a great deal between scales, but also drift over time for each scale, requiring re-calibration nominally every 6 months. The *Sensitivity* values for the different scales remain stable over time. However these values are not close enough to allow a single value to be used across all scales, requiring a separate *Sensitivity* value for each scale.

Observations: As the foragers of a hive start to return with pollen and nectar, the weight of the hive starts to increase. This increase is directly proportional to floral resources available around the hive. Once the sun sets, or the environmental conditions do not favour active foraging, the hive weight starts to decrease. This decrease is a net result of both nectar losing weight because of evaporation, and from the bees consuming the stored food to keep themselves warm.

Figure 3.7 (a) shows samples of weight data from a single hive during two different seasons. On 14th and 15th of March 2021 (early autumn), the hive gains close to 0.50 kg of weight over a two day period. The sharp dips in the weight in early hours of morning are the result of many foragers simultaneously leaving the hive. These are also reflected by negative rate of weight change of -0.24 kg/hr and -0.36 kg/hr, marked with red in Figure 3.7 (b). There are significant portions of positive rate of change during the day indicating productive foraging activity. However weight data from the 5th and 6th of June, 2021 (early winter) tells a different story. The hive loses approximately 0.25 kg of weight over two days, as little nectar and pollen is available for foraging. The foragers leaving the hive early in the morning are significantly less in number, which is also evident with the absence of negative rate of weight change for these days. The positive rates in the two day period in winter are also of lesser magnitude and duration when compared to the two day period in autumn 2021. Thus the rate of weight change of a hive acts like a *heartbeat* of the hive, providing vital information about hive health and bee activity.

Challenges: In Figure 3.7 (a), one can observe ripples in the hive weight on the 6th of June 2021, which are a result of rain on that day. Most of the beehives have horizontal top lids which are not perfectly flat, and a concave curved lid can accumulate rain water impacting the weight. Any wire connections or circuit boards of the scale exposed to rain can give incorrect readings. This makes it hard to get a precise weight reading of hive during/after rain. Other environmental factors such as temperature, exposure to sunlight/shade also impact the operation of scales. Since these scales are meant to be deployed in large numbers, they also should be cost effective and well calibrated. These issues make the design of these weighing scales very challenging, yet weight is perhaps the most important parameter to know when monitoring bees.

3.4 Sensor Evaluation Using Random Forests

Proper evaluation of sensors in a design is a fundamental step for system optimization. Researchers have used machine learning not only for evaluation and calibration of sensors [116], but also for aiding the sensor design [117]. *Ballard et al.* in their work [117] discuss how the design of hyperspectral image sensors, diagnostic sensor for Lyme disease, and other distributed sensing platforms can be improved by leveraging machine learning techniques and inverse design. They suggest that hardware for data acquisition can be redesigned for optimal sensing of data with respect to a parameter defined by user. This is especially useful for systems designed for specific tasks, which can be used as a benchmark for evaluation. In our case, the daily weight change of a hive is used as a benchmark to evaluate the in-hive sensors of BeeDAS on the basis of their contribution towards predicting/estimating this change. At this stage, the aim is not to achieve best possible weight estimates, but to evaluate the sensors and features in the design against each other. Later chapters will explore advance deep learning techniques not only to validate these findings but also generate best possible weight estimates.

One of the machine learning algorithms explored for the evaluation of sensors early in this study was Convolutional Neural Network (CNN). CNNs are good at identifying different patterns in the data, and can be used for both regression and classification tasks. However the fully connected layers and convolutional filters of CNN make it difficult to identify the contribution of individual inputs for a given task. After some experimentation, it was decided to not pursue CNNs for sensor evaluation. Also, classification of daily change of hive weight into low, medium or high does not provide beekeepers with a resolution they are interested in. Beekeepers are more interested in net change of the hive weight, thus this total change was later used as a benchmark to evaluate the contribution of different sensors. Therefore, the preliminary results of using CNN for hive weight classification are not included in this thesis.

As discussed above, the net change in the hive weight during a day is treated as a regression problem for sensor evaluation. Multiple machine learning techniques can be employed for regression, but random forest [106] are most widely used for the assessment of importance of the inputs for the regression task. They are also resistant to over-fitting and the ensemble of decision trees usually provides good accuracy. This makes random forest a good candidate for evaluating sensor feature importance in BeeDAS for the task of estimating the daily weight change.

As discussed in the previous sections, the changes in the weight of a beehive are one of the most obvious indicators of bee activity. If a sensor system is accurately measuring the honeybee activity throughout the day, it should also be able to estimate the daily weight change of the hive. The sensor features collected through BeeDAS are used to estimate the daily hive weight change using random forest [106]. This hive weight change varies between -1.0 kg to 2.5 kg per day, based on environmental conditions, season and strength of the bee colony. The effective regression learning of random forest allows the estimation of weight change for this entire range. Any sensor or feature that contributes well towards the weight estimation is a useful one. However one cannot discard a sensor from beehive monitoring system just on the basis of poor contribution towards weight estimation, as each sensor may contribute differently towards diagnosis of bee diseases. The scope of this study is however limited to evaluating sensor feature importance towards daily hive weight change, which in itself is a very challenging task. Given the difficult nature of measuring the hive weight using expensive scales, reasonable estimation of weight using affordable sensors using machine learning can be very valuable for the beekeepers.

The dataset used for this task has a total of 1,250 days of sensor data, collected from 3 different sites located at; Capel, UWA Crawley campus and Lesueur National Park. This data was collected using a total of 8 units of BeeDAS, deployed in hives at these sites over different time periods. Capel is approximately 170 km south of UWA Crawley campus, whereas Lesueur is 200 km north of UWA campus. Hives used in Capel were made of polystyrene, whereas hives at the other two sites were constructed using wood. The system performed adequately in both types of hives. Initially two units of BeeDAS were deployed to collect data from November 2020, whereas all eight units were deployed from March 2021. Only the data collected till the end of September 2021 is used in this analysis of feature importance. The deployments of BeeDAS between this period (Nov 2020 to Sept 2021) allowed collection of data during a variety of weather conditions, with reasonable spatial diversity. However, the deployment of BeeDAS was not continuous because of hardware and software problems in early stages of field deployment. Units of BeeDAS were

Table 3.1: Details of different sensors, extracted features and the number of features extracted per sensor/environment variable. Average of Mean Absolute Errors (MAE) computed for all 5-Folds for estimating the daily hive weight change using random forest is reported in the last column. The lower the error, the greater the sensor importance towards estimating the change in the hive weight.

Sensor/Parameter	Feature details	No. of features	MAE (kg)
All	-	34	0.200
Audio	Peak amplitude, critical freq components	17	0.220
Temperature	Temperature, temperature gradient	2	0.246
Pressure	mbar	1	0.280
Humidity	%	1	0.302
CO_2	Parts per million	1	0.311
Accelerometer	Dominant freq on x,y and z axis	3	0.339
Humidity (weather)	%	1	0.214
Temperature (weather)	Temperature, temperature gradient	2	0.216
Temp feel (weather)	°C	1	0.220
Wind speed (weather)	km/h	1	0.234
Rain (weather)	Total rain in mm	1	0.305
Season	Week number of the year	2	0.273
Hive size	Number of frames in the hive	1	0.382

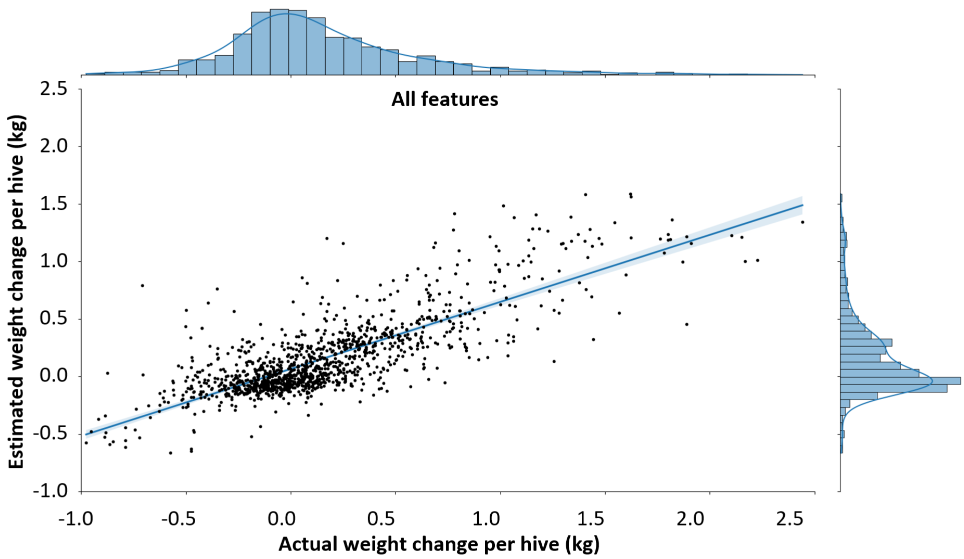

Figure 3.8: The scatter plot of actual weight change vs the estimated weight change per day for 1,250 days in the dataset, using all of the sensor features and the weather data.

repeatedly pulled out to address these issues. The sensor data where heavy rains impacted the weight readings was also discarded from the dataset.

BeeDAS is configured to collect data at 10 minute intervals, generating a total of 144 data points per day for each sensor/feature. For this study, this high resolution data is reduced to 8 data points per day, by computing the mean over three hour non-overlapping windows. This reduction enables the use of random forest for regression. The implemented random forest uses a total of 100 estimators, and depth of trees is set to the total number of features involved in each test. Features such as week of the year (season), daily weather information such as temperature, temperature feel, humidity, wind speed and rain are also part of the dataset. The weight change for each day is computed by taking the difference of hive weight between two consecutive mid-nights (12:00 AM to 12:00 AM).

To form a baseline for comparison, initially all of the sensors/features in the dataset are used to train and test the random forest. We estimate the daily weight change using 5-fold cross validation with a random shuffle of the dataset. This provides a 80-20 split for training and testing, and the average Mean Absolute Error (MAE) on test sets across all 5 folds is reported. Then, only the data/features from individual sensors are used to train and estimate the hive weight change. The results are reported in Table 3.1. From internal hive sensors, audio data has the highest contribution towards estimating the weight change as it generates the lowest MAE of 0.22 kg among all other internal sensors. From environmental parameters, humidity and temperature outside the hive play a key role towards determining the daily weight change.

Some of the qualitative results for the weight change estimations are provided as scatter plots. Figure 3.8 shows the actual weight change vs the estimated weight change using all of the sensor features, along with the seasonal and weather information. Results from testing

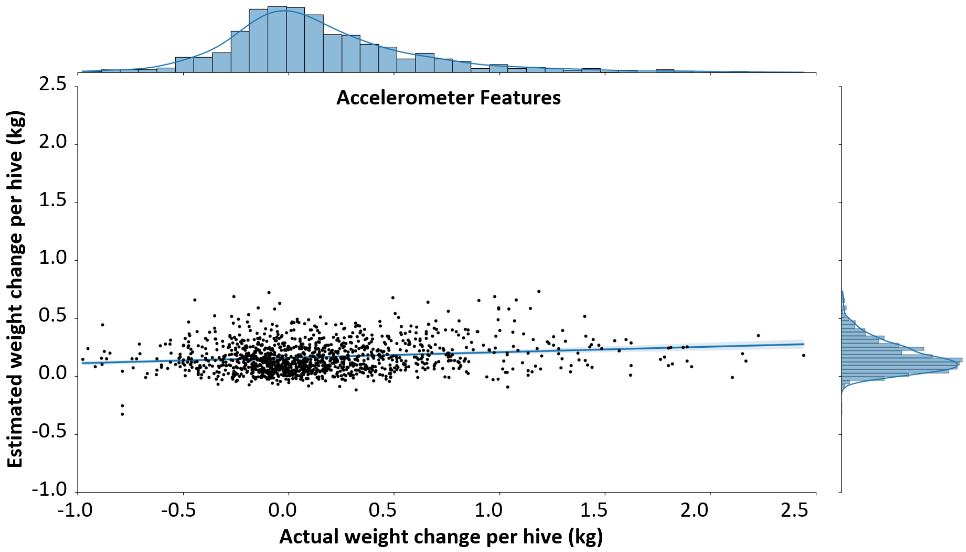

Figure 3.9: The scatter plot of actual weight change vs the estimated weight change of hives for each day using just the features from accelerometer. The narrow distribution of estimated weight change shows that the vibration features extracted using the accelerometer are not of adequate quality, as discussed in the placement challenges section for the accelerometer.

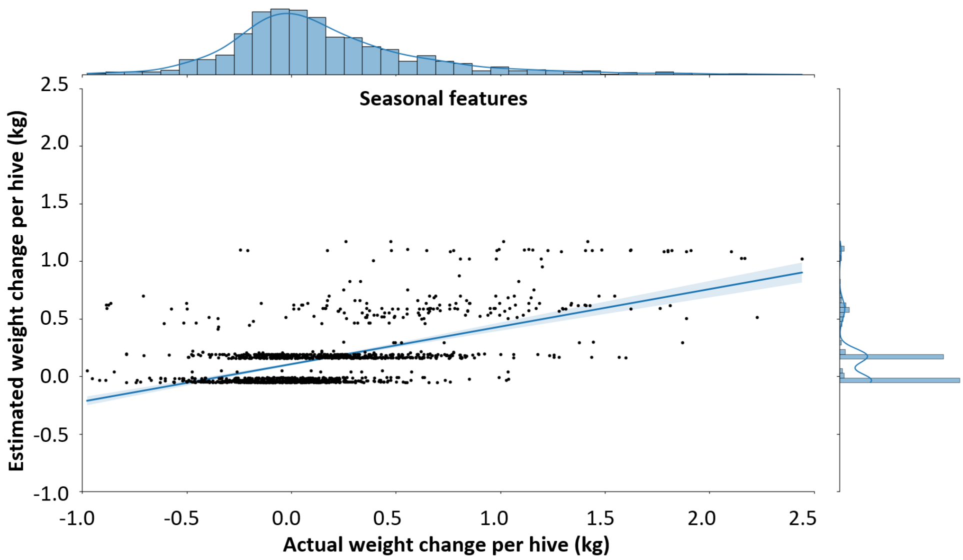

Figure 3.10: The scatter plot of actual weight change vs the estimated weight change using just the seasonal information. Each major season forms a horizontal cluster, which represents the most likely change in the daily weight of the hive for that season.

all 5 folds are pooled together to present a complete picture. As reported in Table 3.1, the best MAE of 0.2 kg is achieved using all of the features. The estimations using all the features have a fit for purpose accuracy when the change in hive weight is between -0.5 and 0.5 kg, which accounts for the majority of the days in the dataset. The weight estimations are inaccurate when the change in weight is significant, mostly due to an unbalanced dataset. Coming chapters will explore other machine learning models which are more robust on unbalanced datasets.

Figure 3.9 shows the weight estimations using just the accelerometer data. The accelerometer data consists of most dominant vibration frequencies extracted from X, Y and Z axes, from 1 second duration recordings captured every 10 minutes. These estimations are least accurate among all other estimations because of challenges associated with capturing vibrations generated by bees in a hive. These challenges are discussed in *Accelerometer* section. Figure 3.10 shows very interesting clustering, where random forest exploits the seasonal information to estimate the daily weight change. The seasonal information in the dataset is available as week number of the year (1 - 52), which provides seasonal details with very fine resolution. The wide standard deviation of estimations in this case is because of the huge impact seasons have on nectar/honey collection. The honey collected (weight gain) during the weeks of spring is significantly greater than any other season, and is represented by the top most cluster with approximately 1.2 kg of weight gain per day.

3.5 Data Communication from Remote Sites

Communication is one of the most crucial aspects of remote beehive monitoring systems. Most commonly used long-range communication modes are cellular and satellite communication. Some systems also employ short range communications such as WiFi, Bluetooth and ZigBee. The use of cellular communication such as 4G, relies on cellular infrastructure availability within 15 km radius [118]. Countries with vast landscape and relatively small population, such as Australia, have large areas with no or poor cellular coverage. Valuable foraging resources for honeybees in remote forests often have little to no cellular coverage. Some of the farmlands, where pollination can improve the production also face the same network coverage problem. The alternative is satellite communication, which is expensive and much more power hungry. However, with the increase in lower orbit satellite deployment for internet coverage, such as Starlink [119], both the cost and power consumption for satellite communication are decreasing rapidly.

BeeDAS has been designed to communicate using:

1- Long-Range Wide Area Network (LoRaWAN)

2- Narrow-Band Internet of Things (NB-IoT)

LoRaWAN is the most cost effective solution available for beehive communication. However, NB-IoT with its 35+ km coverage offers a range advantage over LoRaWAN, which has a typical coverage radius of 10+ km [120]. NB-IoT is relatively new and has not been tested by other researchers in beehive monitoring systems. This section will discuss both LoRaWAN and NB-IoT communication aspect of BeeDAS and their feasibility for remote deployment.

3.5.1 LoRaWAN

LoRaWAN is a communication protocol built upon physical layer of LoRa [121], designed to connect battery operated 'things' to the internet. It is low-power, and long-range with 10+ kms of coverage in perfect conditions [122]. Each LoRaWAN device needs a LoRa gateway to connect to, and uses license free radio frequency bands. The coverage area is significantly larger when compared to WiFi or Bluetooth. In LoRaWAN, the only operational costs are a) the internet connection for the LoRa gateway and b) power. Most of LoRa devices can operate for months with a small battery, and use of a solar panel with chargeable batteries makes them a very good candidate for beehive monitoring systems. The LoRa chips have a small footprint, and can work well with RF Ceramic chip antennas on PCBs. However, to maximize the communication range, an external antenna is preferable.

Challenges: The major problem with the use of LoRaWAN in data acquisition systems is the limited availability of LoRa gateways. LoRaWAN based devices and gateways are increasing quite rapidly, but are mostly concentrated in urban areas. For systems used by hobbyists, or deployed close to a LoRa gateway, LoRaWAN is the perfect solution with no operational communication costs. Also, this protocol is designed to carry small payloads of sensor data. The Things Network (TTN) is used in our application, and all the systems and data loggers connect via TTN to communicate with each other. Fair usage policy of TTN allows upto 30 seconds of uplink air time per day per node/sensor [123] to maximize the number of devices which can use this license free band. For hives deployed in regions with no LoRa gateways, a dedicated LoRa gateway can be setup. Unlike LoRa devices, LoRa gateways consume more power as they are continuously scanning for incoming LoRa messages. Along with a dedicated power supply, they also need an internet connection which can be hard to arrange in remote regions.

The system has been designed to work within the above mentioned constraints using LoRa transceiver from HopeRF [124]. System transmit 91 bytes of sensor data for each iteration, with a spread factor of 7, using Australian 915 MHz band at 500 kHz bandwidth. This calculates to 44.9 msec of air time for each transmission [125]. With a total of 144 transmissions per day, each system uses 6.5 sec of uplink air time per day, which is well within the 30 sec limit.

3.5.2 NB-IoT

NB-IoT is a Low Power Wide Area Network (LPWAN) standard, designed for cellular devices and services [127]. Due to the narrow-band allocation, NB-IoT does not support high data rates like 4G, but has an extended coverage. In good conditions, a 4G base-station can cover an area of upto 15 kms radius, whereas NB-IoT has a coverage radius of upto 35 kms. Figure 3.11 shows the Telstra coverage map of Australia for 4G and NB-IoT, both using the same cellular infrastructure.

Table 3.2: NB-IoT data packets transmitted from Capel, along with the percentage of packets that needed re-transmission because of timeouts. The nearest cellular tower was 15.10 km away from the hive site.

System ID	Packets transmitted	Packets with connection timeout	Packets with Ack. timeout
14	14,400	5.67 %	0.61 %
15	15,120	4.31 %	0.81 %
Total	29,520	4.98 %	0.71 %

An NB-IoT protocol has been designed for communication between BeeDAS units in the field, and a Raspberry Pi host at The University of Western Australia (UWA) Crawley campus, using Message Queuing Telemetry Transport (MQTT) [89]. The small sized packets of BeeDAS sensor data complement well with publish/subscribe messaging supported by MQTT. The NB-IoT first attempts to establish a connection with Telstra network, and then attempts to connect with the host system using MQTT, with timeouts allocated for both. Once successful, the date/time request and the data packets are published by the BeeDAS, and the host in return publishes the appropriate response and acknowledgements. Data communication is only considered complete when BeeDAS receives a final acknowledgement published by the host. In case such an acknowledgement is not received, BeeDAS saves the data on the SD card for re-transmission attempt(s) in next cycle(s). This guarantees that sensor data is never lost during transmission, unless the SD card fails simultaneously. Every time BeeDAS is unable to connect to the network, or fails to receive acknowledgement from the host, error flags are saved with the respective sensor data. Eventually BeeDAS transmits all the pending data, and the host can identify the reasons for delayed transmission using the error flags, if applicable.

In initial test deployments during 2019 and 2020, BeeDAS successfully communicated sensor data using NB-IoT from areas without 4G coverage such as Tincurrin WA and Mundaring State Forest WA [128]. For beehive sensor data collection using NB-IoT, a total

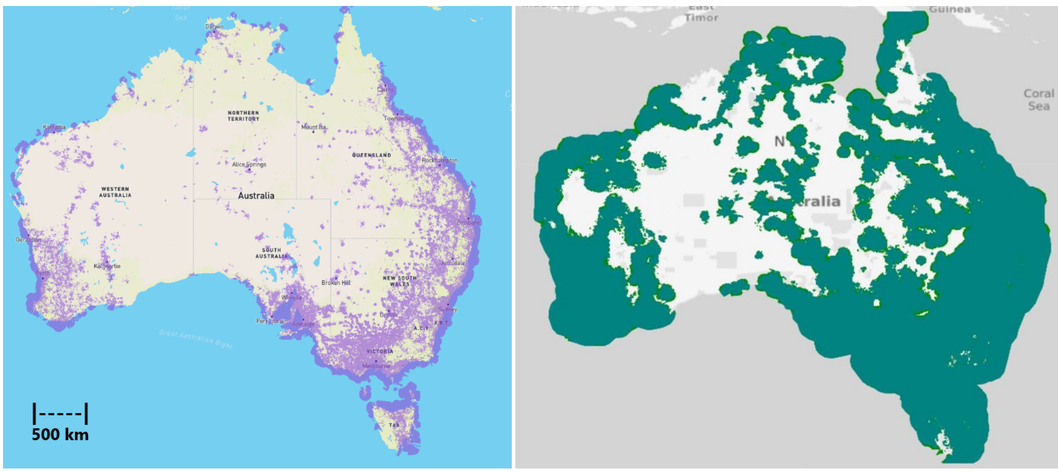

Figure 3.11: Telstra coverage for **4G** (left) and **NB-IoT** (right) in Australia [126], with approximately 700 km^2 and 3800 km^2 coverage per base-station respectively.

of 2 BeeDAS units were deployed from late 2020 onwards in hives located at Clover fields at Capel WA. Later, up to 8 BeeDAS units were deployed in UWA Crawley campus and Lesueur National Park WA with NB-IoT connectivity. Table 3.2 and Table 3.3 show systems deployed at Capel and UWA respectively, and the details of NB-IoT packet transmission from these hive sites. These details include the total NB-IoT data packets transmitted, and the percentage of packets re-transmitted from hive site because of network connection timeout or host acknowledgment timeout. The results from hives located at Capel are interesting because of significantly lower percentage of timeouts, despite being a good distance away from nearest base-station. These hives were used for pollination and placed on a movable platform to allow for easy transportation between fields, which gave these hives (and internal NB-IoT antennas) a 120 cm elevation from ground. This elevation advantage resulted in better connectivity to base-station and timely communication. For the other two locations, the hives were placed just 10 cm above the ground. The majority of re-transmissions occur because NB-IoT fails to connect to Telstra network within the allocated time. The duration of allocated timeout can be increased to reduce the percentage of re-transmitted packets, but that will come at the cost of increased duty cycle, leading to higher power consumption of the system.

During another deployment at a Tedera field near Yathroo WA, communication was significantly impacted as the vegetation in the field was quite dense, and more than 60 cm in height. Since the sensor frame (and the antenna) located in the brood chamber was surrounded by this dense vegetation, NB-IoT was unable to connect to the network. Based on this experience, the design of 4 BeeDAS units was modified and external antennas were mounted with 40 cm of elevation from ground. With external antenna for systems with ID 16, 17, 18 and 19, NB-IoT demonstrated significantly better connectivity. Table 3.4 shows transmission results from hive site at Lesueur after the modifications, where systems with external antennas have a significantly smaller percentage of connection timeouts compared to other systems. However external antennas are vulnerable, and the antenna of system 17 was damaged during hive transportation. The sensor data for this system was manually

Table 3.3: NB-IoT data packets transmitted from UWA, along with the percentage of packets that needed re-transmission because of timeouts. The nearest cellular tower was 0.10 km away from the hive site.

System ID	Packets transmitted	Packets with connection timeout	Packets with Ack. timeout
11	16,848	11.47 %	1.82 %
13	13,680	12.91 %	2.16 %
14	9,504	14.28 %	1.94 %
15	6,336	11.95 %	2.70 %
16	17,712	15.26 %	1.84 %
17	12,384	13.76 %	2.08 %
18	11,088	15.83 %	1.78 %
19	10,512	15.02 %	1.93 %
Total	98,064	13.82 %	1.98 %

Table 3.4: NB-IoT data packets transmitted from Lesueur, along with the percentage of packets that needed re-transmission because of timeouts. The nearest cellular tower was 21.40 km away from the hive site.

System ID	Packets transmitted	Packets with connection timeout	Packets with Ack. timeout
11	5,616	30.59 %	2.37 %
13	5,904	30.25 %	3.32 %
14	5,328	37.58 %	4.07 %
16	6,192	16.12 %	3.28 %
18	6,336	13.79 %	2.32 %
19	6,480	14.98 %	3.04 %
Total	35,856	23.28 %	3.06 %

retrieved from SD card at the end of deployment, and is not included in the communication table for Lesueur.

Challenges: MKR NB 1500 from Arduino was used in BeeDAS, and even though this NB-IoT board has been in production for a few years, it still suffers from stability issues. Initial sensor deployments at UWA used old firmware of the NB-IoT modem. This caused frequent timeouts in communication even though the hive site was just 100 meters away from a cellular tower (see Table 3.3), before we updated the modem firmware to the latest version. The primary issue with the modem on this board and similar communication platforms is the need to maintain and update firmware. Communications hardware is moving faster than software teams can patch issues and the short product life cycle of many communications platforms can compromise long-term remote sensor systems. In addition, one can see from Table 3.5 that the NB-IoT board is the most power hungry component of BeeDAS, thus requires careful management of sleep cycle during remote deployment.

3.5.3 LoRa – NB-IoT Hybrid

All the BeeDAS units deployed at a hive site can communicate with each other using raw LoRa. This gives us the flexibility of creating a star topology LoRa network of sensors, where one of the sensor system can act as a central connection point or hub. This hub can use the onboard NB-IoT for the long distance communication on behalf of all other systems at a particular site. This can save costs by reducing the number of NB-IoT devices, as well as conserve power for the systems using LoRa. Only the hub will need a dedicated NB-IoT device. Since this hub will listen/wait for other devices continuously, and communicate much more frequently, the power consumption for this system will be significantly higher. However pooling of sensor data can be used to optimize the data transmission. This hybrid configuration has not been tested because of time constraints of this PhD project, but we plan to explore this in the future.

3.6 Time Stamping and Sleep Intervals

Time series collection of data requires accurate timestamps associated with each set of data. Before transmitting sensor data, each BeeDAS unit communicates with the Raspberry Pi host at the UWA Crawley campus, and requests date and time information. The host responds with the current date and time, along with the time interval to be used for next data collection. BeeDAS saves these timestamps along with the raw sensor data on the SD card, and transmits the extracted features. BeeDAS then calculates the time it should sleep, based on the provided time interval by the host, and adjusting for the time it has spent in collecting and transmitting the sensor data. These sleep intervals are accurate to one hundredth of a second, and the time drift in data collection is negligible.

In earlier deployments, the sleep interval was altered in real-time based on the battery level of each BeeDAS unit, to avoid completely drained battery. This resulted in a dataset with variable intervals between collected samples, which can be problematic when used for training a machine learning model. To maintain consistency of data collection interval, BeeDAS units are now configured to use an adaptive transmission protocol and temporarily disable communications when the battery level falls below 10%, and continue the data collection at default interval of 10 minutes. This results in a reduced duty cycle. However, the host can override these settings and control the sleep interval of each BeeDAS unit individually. This interval can be decreased down to one minute in case of specific events (such as swarming), where the state of hive changes rapidly. The sleep intervals can also be increased to conserve power during nights, when bees are mostly inactive. However for the collection of data, a fixed interval of 10 minutes is used throughout these deployments, and adaptive transmission protocol is used to conserve power when needed.

3.7 Power Consumption

For remote deployment, power efficient design is crucial. To conserve power, sensors are either put into sleep mode when possible, or to a low-power mode by adjusting their sampling period. The minimum, maximum and typical current ratings of the sensors and components used in BeeDAS are shown in Table 3.5. The analog microphone cannot be switched off, but its data is only sampled when needed. The gas sensor can take up to 20 minutes to stabilize after being powered on, so switching it off is not feasible. However, the gas sensor can be configured to collect data every 60 seconds to conserve power, but that greatly impacts the accuracy. To provide a balance between power consumption and accuracy, the gas sensor has been configured to collect data every 10 seconds. In Figure 3.12, the current consumption of BeeDAS can be observed while collecting and storing the sensor data, extracting the features, and transmission using LoRaWAN. With 7 sec of awake time for every 600 sec cycle, BeeDAS has a duty cycle of around 1.2% using LoRaWAN. With

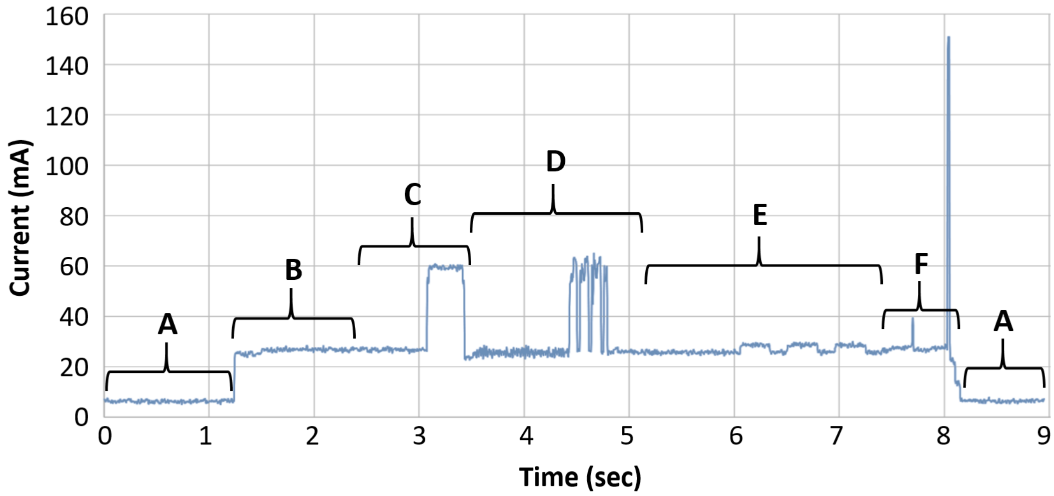

Figure 3.12 Current consumption of BeeDAS for an entire cycle of data collection, storage, feature extraction and transmission.
A- MCU in sleep mode with current consumption at 7 mA.
B- MCU awake and collecting temperature, humidity, pressure and CO_2 data with current consumption around 25 mA.
C- Audio sampling, storing of audio data on SD card and feature extraction using FFT. The storing and feature extraction process increases the current consumption to 60 mA.
D- Accelerometer data acquisition, storing and frequency feature extraction for X, Y and Z axis.
E- Weight data being sampled multiple times for filtering.
F- Data transmission using LoRaWAN with peak current consumption at 150 mA.

NB-IoT transmission, this duty cycle increases to 2% because of connection delays. With MCU and most of the sensors in sleep mode, the base-line current consumption of the system stands at 7 mA, where the gas sensor is the biggest contributor to this constant drain.

The time duration for which each sensor collects data contributes to the duty cycle of the system, and impacts the overall power consumption. At times, multiple readings are made and averaged out to minimize the noise. An array of data has to be sampled for the microphone and the accelerometer for frequency analysis. The size of such an array is restricted by the memory and processing resources available on the MCU, and is fixed to a size N. Careful considerations need to be given while selecting the sampling rate for such sensors. Data in the array of size N can be sampled in a shorter duration using higher sampling rates, thus reducing the duty cycle and minimizing the power consumption. However, these shorter duration recordings make it more likely to miss out on infrequent

Table 3.5: Current ratings of sensors and communication boards used in BeeDAS.

Device/Sensor	Min/Sleep	Typ.	Max
CCS811 – Gas Sensor	19 μA	30 mA	54 mA
BME280 – Temp, Humidity, Pressure	0.1 μA	3.6 μA	-
MMA8452Q – Accelerometer	6 μA	-	165 μA
HX711 – Load Cell Amplifier	0.3 μA	-	1.4 mA
ADMP401 - Microphone	-	210 μA	260 μA
LoRa Transceiver Module 915MHz	1.5 μA	12.1 mA	120 mA
MKR NB 1500 – NB-IoT Board	1 mA	60 mA	140 mA

events. For example, recording audio data at around 4 ksps, with 2048 samples in each of 144 daily recordings, amounts to a total of 72 seconds of audio per day. But recording at 20 ksps with the same 2048 samples at same 10 minutes interval, equates to 14.4 seconds of audio per day. Sampling at a rate higher by a factor of η enables us to observe frequencies η times higher, but for a window duration which is η times smaller. Since the sampling rate in a system with fixed array of size N determines the duration of sampling window, the bin width (resolution) of Fast Fourier Transform (FFT) also increases with the increase in sampling rate. If more than one important frequency component falls within the frequency resolution, the higher sampling rates can lead to loss of spectral information.

The sensor data is recorded every 10 minutes which provides a good resolution dataset, but at the cost of high power consumption and increased data transmission. A 3.7 V rechargeable LiPo battery with 6000 mAh charging capacity is used in BeeDAS. An external 10W-5V solar panel with IP65 rating is used for charging the battery. The high capacity battery ensures recordings of several days even when the solar panel is unable to charge. To prevent power hungry NB-IoT transmitters from completely draining the battery, the transmission is disabled when the battery level falls below 10%. This allows BeeDAS to collect data for another 3 to 4 days on the battery reserve. Once the solar panel charges the battery to more than 10%, the pending data is transmitted to the host system. This is not an ideal scenario if real-time communication is of high priority, rather a trade-off to conserve power to allow reliable collection of sensor data.

3.8 Data Storage

BeeDAS uses an on-board 32 GB SD card to record all the raw sensor data, as well as the extracted features. The raw sensor data from the SD card can later be analysed to efficiently evaluate the features, and to easily debug the feature extraction process. On-board storage also enables storing of the calibration data for each sensor, instead of hard coding the sensitivities and offsets. It also allows a cushion against failed communication by saving the sensor data, which can be later re-transmitted. However, one problem with SD cards is their reduced life inside the beehive due to the high humidity. The average life of standard SD cards during deployment was observed to be between 8-10 months, after which the faulty SD cards were not readable when tested on different platforms. This was a frequent problem which was overcome by transitioning to industrial grade SD cards. These high endurance SD cards have not yet failed during deployment.

3.9 Summary

This chapter discussed BeeDAS, a data acquisition system for beehives. The objective of designing this system was to evaluate different sensors for remote monitoring of hives, and to collect a high quality beehive sensor dataset. The system has been designed with a focus on low-power consumption and long-range communication. The sensors used in BeeDAS were carefully selected, and data processing within the hive was used to generate 91 bytes of data each cycle. This small data enabled the use of latest long-range communication technologies such as LoRaWAN and NB-IoT for monitoring of hives. The power efficient design of BeeDAS results in less than 2% duty cycle, achieved through low transmission capacity but which still permits reliable monitoring of hives with a 10 minute interval. This allowed the deployment of these systems in hives located at different geographical sites, and in varying environmental conditions to enable collection of a diverse dataset. The collected beehive sensor dataset and weather recordings of 1,250 days were also tested by estimating the daily weight change of a hive using random forest. Weight change of a hive is the biggest indicator of bee activity. Using all the features in the dataset for regression achieved a Mean Absolute Error of 200 grams per day per hive, on a dataset of 1,250 days of beehive sensor and weather recordings. These weight estimation results show that data collected by BeeDAS is rich with information about the bee activity. The issues related to system power optimisation, and data storage were also discussed in detail. This chapter helped identify the sensors, communication systems, and the design of a beehive monitoring system suitable for remote deployments.

Chapter 4

Weight Estimation of Beehives Using LSTM-Based Deep Neural Network

The design, development and deployment of data acquisition systems was a difficult task. As discussed in the previous chapter, the impact of harsh environment and weather, especially rain on the measurement of hive weight, and the communication issues using NB-IoT presented frequent challenges. It took us a little over two years and multiple design and development cycles before we were able to deploy all eight sensor systems in hives for reliable collection of data. This chapter explains the collected dataset and its significance, and also investigates soft sensing using machine learning to eliminate expensive and/or difficult to use sensors in beehive monitoring systems.

There are multiple parameters that are of paramount interest to beekeepers. One of them is the monitoring of day to day weight changes of a hive, as these weight variations are a good indicator of bee colony strength and honeybee activity [32]. But commercial beehive weighing scales are expensive and difficult to deploy. This chapter explores the use of deep learning to estimate the weight variations of a beehive using internal hive sensors, which cost much less and are not exposed to extreme weather conditions outside the hive. A model capable of estimating the correct trend of weight change is fit for purpose for most of the beekeepers. This chapter uses a combination of different internal hive sensors to gauge the complex activity of honeybees, along with external weather, seasonal, time and size information of the hives. The beehive sensor data used to train the model was collected using the systems described in previous chapter. To the best of our knowledge, this is the first work on estimating the daily weight variations of a beehive using machine learning. Figure 4.1 illustrates the proposed Weight Estimator for Beehives (*WE-Bee*).

The highlights of the proposed model are:

- Hybrid model for soft sensing and time series forecasting
- Cumulative beehive weight estimation over multiple weeks
- A fit for purpose design for cost sensitive market

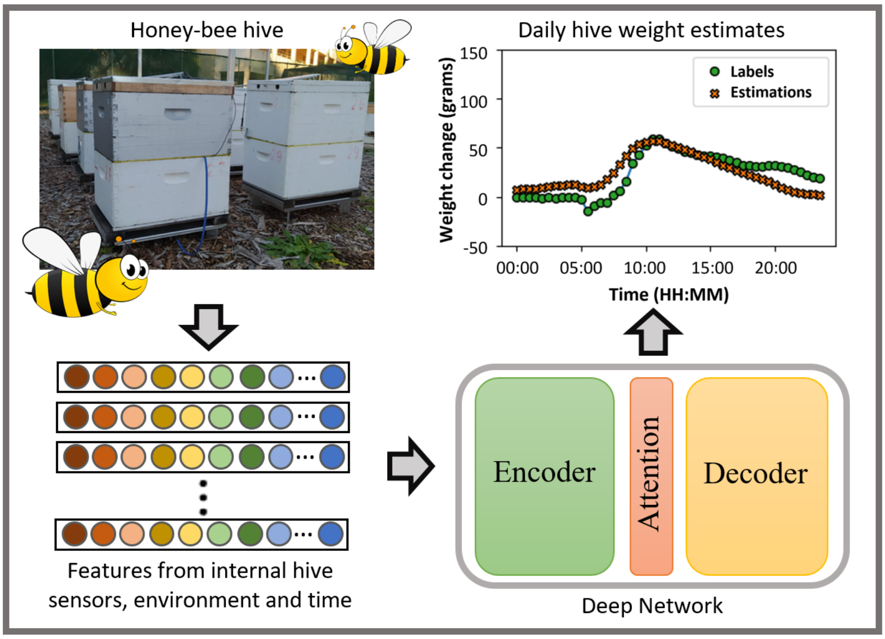

Figure 4.1: *WE-Bee* uses internal hive sensors, environmental features, season information, time and hive size to estimate the daily weight variations of a beehive per hive frame.

4.1 Significance

As discussed in chapter 1, honeybees play a critical role in pollination. Commercial beekeepers frequently move their hives between fields to provide pollination services. This hive movement is very stressful for bees and can adversely impact the colony strength. Excessive use of pesticides on agricultural land can also be harmful for the bees. Monitoring the health and strength of bee colony is a genuine concern for the beekeepers, as well as for the farmers who pay for the pollination service. Strong colonies contribute to pollination much better compared to weak ones. A strong bee colony can bring up to 3 kg pollen and nectar to the hive on a productive day. Thus, monitoring the weight of a hive provides a very good assessment of bee colony strength/activity, and the contribution of hive towards pollination [32].

Monitoring the weight of a hive however comes at a significant cost, with a single beehive weighing scale often exceeding 500 AUD in price. Commercial beehive monitoring systems use electronic sensors to collect data from inside the hive, and an external weighing scale to monitor the weight of hive. The weighing scale is mostly sold as an optional add-on to the monitoring system because of its cost, which is usually more than all of the internal sensors put together [19, 18, 129]. Many factors contribute to the high price of beehive weighing scales. Commercial beehives during peak honey flow in summer/spring can weigh upto 120 kg. The design of these scales should be rigid enough to support this weight, and the electronic load sensors should be sensitive enough to pick up small variations of a few grams. These scales are also designed to work in harsh weather conditions, to be able to withstand extreme heat, cold, and rain, which adds to the cost. But despite high costs, their performance in the field is often below expectations. Furthermore, these scales are often bulky, and have to be setup every time a hive is moved, adding to the setup time and the physical effort required by the beekeepers. Repeated deployments from one field to another also increase the wear and tear of these scales. Thus, an expensive and bulky scale under each hive is not feasible for a majority of commercial beekeepers.

Authors in [96] conduct a very thorough survey of Precision Beekeeping. It is a branch of Precision Agriculture focused on management of apiaries based on the monitoring of individual bee colonies in order to minimize resource consumption, and maximize bee productivity. The authors in their bibilometric analysis discover that from a total of 73 cited papers which use data for training machine models, only 6 use weight of the hives. From our experience, this is because collecting weight data from hives even for research purposes is expensive and very challenging. The authors in their conclusion state that "Looking at beekeepers' needs, it is clear that more affordable commercial solutions are to be developed." Given the high costs of weighing scales in beehive monitoring systems and the difficulties associated with their deployment, any progress towards cost-effective alternatives for weight measurements of hive will contribute significantly towards wide scale deployment of monitoring systems.

4.2 Challenges

The use of deep learning is explored to sense daily weight variations of a hive, using time series data from inexpensive sensors. Previous research has identified factors that contribute to weight variations of a hive [34, 130], including (but not limited to): number of forager bees, availability of floral resources and their distance from the hive, food consumption rate of the bees and the larvae, environment (temperature, rain, wind), and the evaporation rate of nectar. With so many variables involved, estimating the weight variations of a hive is a difficult, but an important problem to solve.

The strength of a honeybee colony is one of the biggest factors that contributes to the weight gain of a hive. The number of foragers (bees that go out in search of food) in a hive is directly proportional to total number of bees in the hive. A strong colony deploys more foragers to find pollen and nectar, and the hive gains weight at a faster rate. The most direct way to count foragers is to use cameras at the hive entrance. This however is not a very cost effective solution, and requires a lot of power and data bandwidth, a luxury remote beehive monitoring systems cannot afford. An indirect way is to estimate the strength of a bee colony, by monitoring the thermoregulation of the colony. A strong colony maintains the appropriate temperature and humidity levels inside the hive [62, 39, 28]. Hence the variations in temperature and humidity inside the hive compared to the variations outside the hive, provide a very good indication of the strength of the bee colony.

The work of *Hambleton et. al.* [33] is one of the earliest studies on the effect of weather and environment on beehive weight variations. Honeybees are inactive at night, thus the hive cannot gain weight. However the hive usually loses weight at night because of nectar evaporation and the food consumed by honeybees and larvae. The rate of food consumption depends on the number of bees in the colony, and the temperature. Bees consume more food in lower temperatures and increase their metabolism to keep the hive warm, which increases CO_2 concentration inside the hive. The rate of evaporation from the hive depends on the difference between the humidity levels inside and outside the hive.

In early morning, if the temperature outside the hive is appropriate, the forager bees leave the hive in large numbers to check for pollen and nectar availability. This results in a steep drop in the weight of a hive, which is referred to as 'Breakfast Canyon' [130]. The duration of Breakfast Canyon depends on the time foragers take to return to the hive. If there is plenty of pollen and nectar available closer to the hive, foragers return quickly, otherwise it takes longer for them to return. This availability of foraging resources, and their distance from the hive is a very difficult factor to estimate as it depends on the location of hives, season, weather and types of flora available. The magnitude and the frequency of bee buzz [43, 46], and bee waggle dance vibrations [55] are however good indicators of the level of foraging activity.

Based on temperature suitability and other environmental conditions, honeybees collect pollen and nectar throughout the day, resulting in the increase of hive weight [131]. In hot summer days, bees stop foraging activity when the temperature outside the hive increases in the middle of the day. Some flowers produce nectar only during early hours of the morning or late in the evening, dictating the pattern of hive weight variation. High wind speeds and rain disrupt bee activity and foraging. However rain can result in an increase in the weight of a hive because rain water can accumulate on the top of flat hive surface, and the wooden structure of the hive can absorb moisture resulting in weight gain. The effect of rain depends on the absence/presence/quality of paint on the outer-surface of wooden hives. Similarly, exposure to the sun or hot and dry weather can also lead to beehive structure losing moisture and weight. On the other hand, hives made of plastic or polystyrene cannot absorb any moisture, thus contribute very little to the hive weight variations.

The design and structure of a beehive weighing scale itself is a contributor to weight variations. The load sensors, Analog to Digital Converters (ADC), and the frame of weighing scale, are exposed to variations in temperature, humidity and other environmental factors, which impact their performance. Research is continuing on better designs to improve the performance of beehive weighing scales [34, 132, 133, 134]. Many commercial beehive monitoring systems are also competing with each other to provide affordable weighing scales. However given the durability and accuracy requirements of design, the cost of commercial beehive weighing scales is still high for a majority of beekeepers, preventing their large-scale deployment.

4.3 Deep Learning for Data-Driven Soft Sensing

Deep learning has shown a lot of promise in forecasting time series data, and in soft sensing for industrial processes. Soft sensing has been widely used in industrial processes to predict difficult to measure variables [135]. However to the best of our knowledge, no work has been done on estimating beehive weight or its variations. It is primarily because the weight of a hive depends on multiple factors, and a wide range of sensor data is required from the hive and its surroundings for reasonable estimations. Such datasets are not publicly available. Designing and developing sensor systems suitable for beehives is challenging, and the data collection from beehives is a time consuming process. One interesting work of soft sensing on humans is [136], where the authors used wearable strain sensors to measure angle of multiple joints in a human body to estimate the human gait. This work uses the same principle, where easy to use sensors are utilized to sense a difficult to measure quantity. However, the weight of a hive at any point of time is dependent upon the conditions at that time, as well as those in the past. This requires the deep network to pay attention to sensor data of an entire day to generate accurate estimates.

Long Short-Term Memory (LSTM) networks have been used successfully for time series forecasting. Authors in this work [137] demonstrate the ability of bidirectional LSTMs and Temporal Attention to learn long-term dependencies and correlation features which are hidden. Authors in [47] compared LSTM networks with other machine learning approaches to identify Queenlessness in hive using hive audio data, and found that LSTM networks provided promising results. The authors in [138] compared different LSTM based networks on time series data. Six different models were adapted for the task of forecasting hourly rainfall, using the historical weather data of two decades from five different cities in UK. The results showed that Bidirectional-LSTM Network had a comparable performance with the Stacked-LSTM Network with two hidden layers. Whereas the Stacked-LSTM Network with multiple hidden layers showed the worst performance. The authors also highlighted the problem of LSTM models not generalizing well on unseen data, and over-fitting the training data. This is a problem faced by many deep models and the training and testing of models needs extra care to minimize the bias in results.

WE-Bee is designed as a hybrid model to soft-sense/estimate the time series data of daily beehive weight variations using Bidirectional-LSTM. The cost effectiveness of *WE-Bee*, and fit for purpose weight estimation in the field makes it a very useful tool for the beekeepers in the cost-sensitive market.

4.4 Data Collection

To efficiently train any machine learning model, the quality and quantity of training data plays an important role. The quality of sensor data largely depends upon the sensor system itself. A total of eight sensor systems were designed, developed and deployed at three different sites to collect data from beehives. The microcontroller of the sensor system is used to extract sensor features, and NB-IoT is used to transmit these features from the remote site over a low bandwidth channel. The carefully designed feature extraction process significantly reduces the volume of data, e.g. each audio recording of 2048 samples is reduced to 17 features containing audio amplitude and frequency information of multiple 100 Hz bands in the bee buzz. Temperature and its gradient, both inside the hive and outside the hive are used as features along with the temperature feel. Table 4.1 lists all the parameters and features used by *WE-Bee*. The details of data collected using each sensor system are given in Table 4.2.

Sensor systems with ID 14 and 15 were deployed in multiple hives, over different periods of time. Whereas rest of the systems were deployed each in a single hive, and stayed in their respective hives when the hives were moved to different sites. The deployment of sensor systems in hives however was not continuous for several reasons. Hardware and software problems in early stages of field deployment often forced us to pull systems from the hives to address the issues. Sensors for CO_2 and hive weighing scales require frequent re-calibration, for which they were repeatedly pulled out. The water proofing of weighing scales has been a continuous concern, and heavy rains often cause malfunctioning of scales. Sensor data where weighing scale(s) showed unrealistic variations due to rain was also discarded from the dataset.

WE-Bee is designed to estimate the weight variation pattern for an entire day, hence problems with the data for even a few hours on a given day make the data for the entire day unreliable. Attempts to use interpolation to fix the outliers or missing sensor data did not

Table 4.1: The composition of parameters and the number of features contributed by each parameter.

Measured Parameter	**Number of Features**
Temperature inside the hive	2
Humidity inside the hive	1
Atmospheric pressure inside the hive	1
CO_2 inside the hive	1
Vibrations inside the hive	3
Bee buzz (audio) inside the hive	17
Temperature outside the hive	2
Temperature feel	1
Humidity outside the hive	1
Wind speed	1
Rainfall (in millimeter)	1
Time of the day	2
Week of the year (season information)	2
Number of frames in the hive	1
Total	36

Table 4.2: Break-down of data collected (days) for training and testing of *WE-Bee*. A total of 1200 days of sensor data has been used in this chapter, collected from 3 different sites using 8 sensor systems. Site-B is approximately 170 km north of Site-A, whereas Site-C is further 200 km north of Site-B. System 14 and 15 were deployed to collect data from November 2020, whereas rest of the systems were deployed from March 2021.

System ID	**Site-A (days)**	**Site-B (days)**	**Site-C (days)**	**Total days**
11	-	117	48	165
13	-	95	43	138
14	97	66	37	200
15	105	27	-	132
16	-	123	67	190
17	-	86	-	86
18	-	77	68	145
19	-	73	71	144
Total	202	664	334	1200

provide adequate results because of the complex nature of data from bee colonies. Also, the beehives need regular inspections to ensure the health of bees. The hives used in this study were inspected every fortnight for most of the year, and once a month during winter to make sure that the bees are healthy, and the queen is laying eggs. During these inspections, hive frames were pulled out one by one, with the hive open for up to 30 minutes during each inspection. Occasionally frames were added/removed/swapped during these inspections, which led to a change in the weight of the hive. For these reasons, data from the days of hive inspections is also not included in the dataset.

The variation in total days of data collected from each system in Table 4.2 is a result of different days of deployment, as well as a different number of data days discarded from the dataset for each system. The data is collected with an interval of 10 minutes, resulting in 144 data points per hive per day. *WE-Bee* is designed using 48 data points per day, with an interval of 30 minutes between consecutive samples, which is adequate to capture important variations in the hive weight and other sensor data. The 144 data points collected each day are used to increase the quantity of the training data, by extracting 3 sets of 48 data points from each day in the training set.

The activity of bees has a high correlation with the position of the sun. The position of the sun not only depends on the time of the day but also on the season. To a large extent, the seasons around the globe are determined by the latitude [139]. This information is important for the neural networks to generate reasonable daily estimates of weight. The information about the time of data acquisition and the week number of the year are included in the dataset after cyclical encoding using sine/cosine transformation. The cyclic encoding allows the start and end of the day/year to have a similar representation, and avoid big discontinuities that can result with linear transformation. For example, last week of December (week-52) and first week of January (week-1) have very similar times for sunrise/sunset, and the seasonal conditions for a specific geographic location. Hence their encoding also need to be reflective of this similarity. Same is applicable for the time around midnight, where both 23:59 and 00:01 are from the same part of the night despite belonging to two different days. The use of both sine and cosine overcomes the repetitive nature of curves, and time/week are encoded as a unique set of two points (features) each. For data collected from sites with huge geographic difference along north and south, the latitude of hive sites can be added to the dataset. This will allow the network to learn the seasonal variations that occur with respect to latitude. Given little variation in the seasons of sites used for the data collection in this study, latitude is not part of the dataset.

Weather has a huge impact on the honeybee activity, and is a significant factor in determining the hive weight variations. The data regarding external temperature, temperature feel, humidity, rain and wind speed was collected using the online reports generated by the Bureau of Meteorology (BOM) [110]. These reports are generated every 15 minutes for BOM weather stations which are available throughout Australia. We chose the closest weather station to the beehive site for our dataset. A significant lack of accuracy was observed in the weather data at Site-C (Lesueur near Jurien Bay, Western Australia), which is located approximately 48 km away from the nearest weather station. Rain was often reported when there was none at the site of hives, which was evident by solar panels charging the batteries. At times the beehives experienced rain, which resulted in noisy data on the weighing scales, but was not reported by the weather station site. Site-A (Capel, Western Australia) and Site-B (UWA Crawley Campus, Western Australia) were 8.4 km and 5.3 km away from the nearest BOM stations respectively, and their weather data was accurate for the hive positions. For future work, use of dedicated weather stations for hive sites located more than 20 km away from BOM stations would help improve the quality of weight estimates.

Size of a beehive also determines the capacity of the hive, which is one of the factors impacting the weight variations. Beehives come in different shapes and sizes, and the most common ones consist of multiple chambers stacked on top of each other. Each chamber contains multiple frames, which are used by bees to make a wax comb to raise the brood, or to store pollen and nectar. There is no standard design of a beehive, and most of the beekeepers have their own preferences. Some hives have 5 frames per chamber whereas some have up to 10 frames. Even the size (dimensions) of frames can vary from hive to hive. Beekeepers also change the number of chambers in a hive from time to time, depending upon the availability of nectar and the strength of the bee colony.

The data collected for this study is from hives of different sizes, however the size of the frames used in all these hives is the same. A hive consisting of N chambers with M frames per chamber, will be referred to as a hive of size $N \times M$ frames. This allows the number of frames to be used as a standard measurement for hive size, and the product of N

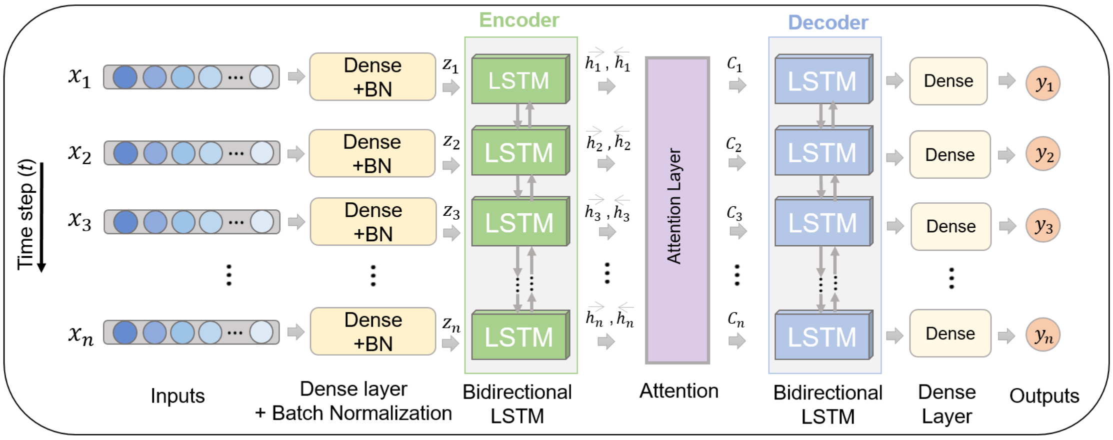

Figure 4.2: The Network architecture of *WE-Bee*. The input features are represented by x_t, whereas y_t is the output of estimated weight variation for a specific time step t. With a total of 48 samples of data per day, n is 48 in this particular case.

and M (number of total frames in the hive) can be used as the total capacity of the hive. The total number of frames in each hive is used to calculate the net weight variation of the entire hive, based on variations estimated per frame using *WE-Bee*. The baseline for hive weight was obtained by measuring several hives with empty frames and no honeybees. The average weight of empty hive structure (not part of the dataset) is 1.06 kg per frame, whereas the average weight of hives including the structure, pollen/nectar and bees in this dataset is 2.39 kg per frame.

4.5 Network and Experimental Setup

The deep learning architecture of *WE-Bee* is inspired by [137], where the authors use multivariate time series forecasting using attention-based encoder–decoder framework. The authors use their network to *predict* the values of a time series data in future, however we estimate the values of an unknown sensor (weight) for the same time. Previous sections explain how weight of a beehive is dependent on many different factors. *WE-Bee* exploits these dependencies to *estimate/predict* a series of weight values based on time series data collected from internal hive sensors and relevant information. Input to the network is a set of data collected from internal hive sensors such as temperature, humidity, atmospheric pressure, CO_2, acoustics and vibrations. Information about the weather, week of the year (seasonal information), time of the day, and the size of hive is also part of input. All the inputs are processed to create a feature vector x_t of size 36 (see Table 4.1) for each time step t, with a total of 48 time steps per day. Figure 4.2 shows the network architecture of *WE-Bee*, with hyper-parameter settings given in Table 4.3.

Daily weight variation estimation of a beehive is a many-to-many sequence-based problem, with both input and output having multiple time-steps. The change in hive weight at any time step, with midnight weight as a reference, is dependent upon all the bee activity and environmental conditions till that time step. The weight itself varies in a pattern and each estimation should properly fit between its neighboring estimations. The use of bidirectional LSTMs (Bi-LSTMs) as our encoder and decoder leverages both past and future contexts within a day. This allows the network to be robust against occasional noisy samples in the input features, and helps with accurate weight estimations. As obvious from the name, LSTMs have both long and short term memory. The short term memory is represented by h_t, or the hidden state. In bi-directional LSTMs, these hidden states are shared in both forward and backward direction across different time steps. Attention mechanism of a machine learning model helps the model to focus on data which is more important. The attention layer learns the importance of different inputs during the training process. The hidden states of the encoder are attended by the decoder via an attention layer, to utilise the most important information for transforming (decoding) input features to weight estimation y_t for each time step t.

Let x_t and y_t be the input feature vector and output weight estimate respectively for every time step t, where $t = [1 : n]$ for each day. Our network first projects x_t onto a sequence of 250 dimensional embeddings z_t. These embeddings are encoded by a Bi-LSTM into a context matrix, which is a concatenation of its hidden states h_t (forward $\overrightarrow{h_t}$ and backward $\overleftarrow{h_t}$).

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{4.1}$$

The decoder estimates the weight using the context vectors C_t. The dot-product attention mechanism is used to compute the context vectors, which are generated as a weighted sum of the hidden states of the encoder Bi-LSTM. The attention mechanism passes on the most useful encoder hidden representations to the decoder. The context vector C_t can be formalised as:

$$e_t = [W \times h_t] + b \tag{4.2}$$

Table 4.3: Hyper-parameter settings for *WE-Bee*.

Parameter	**Value**
Units in input dense layer	250
Activation function of dense layer	Leaky ReLU
Units in encoder	250
Units in decoder	500
Bi-LSTM merge mode	concat
Activation function of attention layer	softmax
Dropout (dense, encoder, decoder)	0.7
Units in output dense layer	1
Activation function of output layer	linear
Max training epochs	1000
Batch size	128
Loss function	MSE
Optimizer	Adam

Table 4.4: Results from 5-Fold cross-validation with a random shuffle of entire dataset. The Mean Square Error (MSE) is reported for 11,520 data-points (240 days $\times$ 48 data-points per day) per fold. Mean Absolute Error (MAE) is reported as a percentage of error, calculated using the average weight of 2.39 kg per frame in the dataset. The scatter plot for weight labels and weight estimations at the end of day is shown in Figure 4.4.

Fold	MSE (grams/frame)	MAE %	Variance of % errors
1	14.8	0.58	0.74
2	13.9	0.54	0.67
3	13.2	0.54	0.60
4	13.3	0.56	0.67
5	12.7	0.53	0.61
Avg	13.6	0.55	0.66
Std Dev	0.8	0.02	0.06

where W and b are attention weight and attention bias respectively.

$$a_t = Softmax(e_t) \tag{4.3}$$

$$C_t = a_t \times h_t \tag{4.4}$$

We used Keras, which is a high-level API of TensorFlow 2 to implement our model. The system used for training has an Intel® Core™ i7-10700K CPU @ 3.80GHz with 16 cores, 32 GB of RAM and a single NVIDIA GeForce RTX 2080 SUPER GPU with 8GB of memory. The network has approximately 5 million trainable parameters. 5-fold cross-validation was used to test the performance of *WE-Bee* and the MSE of test set was monitored during training with an early stopping (patience of 100) to avoid over-fitting.

4.6 Results and Analysis

The quantitative results for 5-fold cross-validation for all folds of dataset are shown in Table 4.4. The test scores of Mean Square Error (MSE) for each fold are reported in grams per frame. The average error for all folds is 13.58 grams, with a standard deviation of 0.8 grams per frame. To make more sense of what these errors represent, the label variations as well as the estimated variations are added with an offset of 2.39 kg, which is the average weight per frame in our dataset. The percentage error between estimated weight and label weight for the frame is then computed for each point in each day in the fold. The percentage Mean Absolute Error (MAE) and the variance of percentage error for each fold are reported in last two columns of Table 4.4 respectively.

Some examples of estimated weight variations per frame by *WE-Bee*, as well as the label weight variations per frame from the test set are shown in Figure 4.3. A total of 48 estimations are generated for each day, with the weight at midnight (00:00) as starting

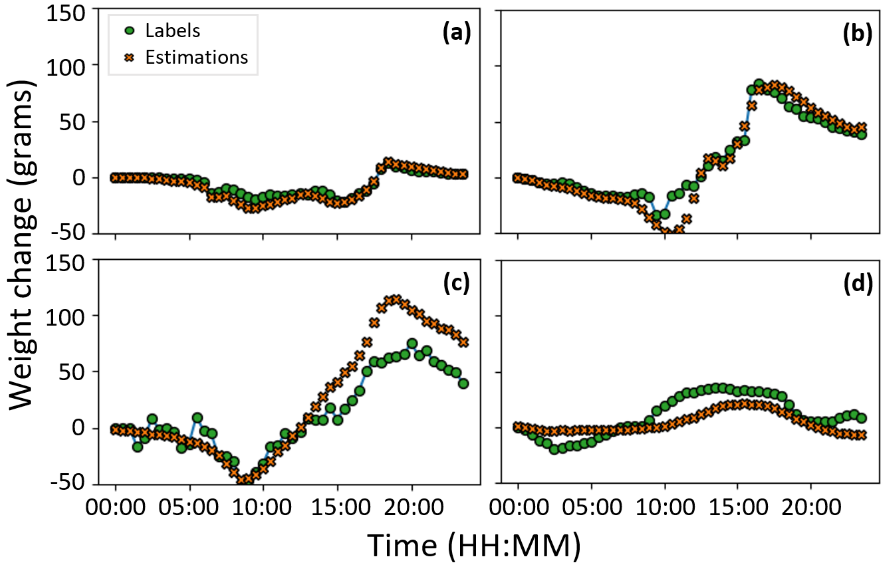

Figure 4.3: Test-set examples of daily weight variation labels per hive frame, and the estimations for the same using *WE-Bee*. First weight reading for each day at 00:00 is the reference for variations throughout the day. Daily weight variation estimations leading to a negligible error at the end of the day are shown in (a) and (b). An over-estimate of the daily weight on a day with occasional rain can be observed in (c). Example of an under-estimate of the daily weight is shown in (d).

reference for each day. The daily estimations can be divided into two categories. One where errors for all estimations within a day add to a negligible error by the end of the day, as shown in examples of Figure 4.3 (a) and (b). The second category is where the accumulated error for all 48 estimations within a day lead to either an over-estimate or under-estimate of weight variation, as shown in examples of Figure 4.3 (c) and (d) respectively.

For cumulative estimation of hive weight over multiple days, estimation for each day starts from where the estimation of previous day had ended. So the error in weight estimation at the end of the day (see Figure 4.3 (c) and (d)), propagates to the weight estimations for next day(s). A biased network with minor but consistent over/under-estimates will lead to a huge error over cumulative estimations. However a network with the Gaussian distribution of errors, will have a smaller accumulated error. Figure 4.4 shows the scatter plot of actual against estimated weight per frame per day. The network shows slight bias towards over-estimating the weight when the hive loses weight for a given day (cases where actual weight is less than 2.39 kg per frame). However when the hive gains weight by the end of day (cases where actual weight is more than 2.39 kg per frame), the network is slightly biased towards under-estimating the weight. Compared to the bias observed with random forests in Section 3.4, this bias in estimations is significantly less. The mean absolute error is around 0.5% at 2.35 kg, and around 1% at 2.50 kg with respect to actual weight of hive per frame.

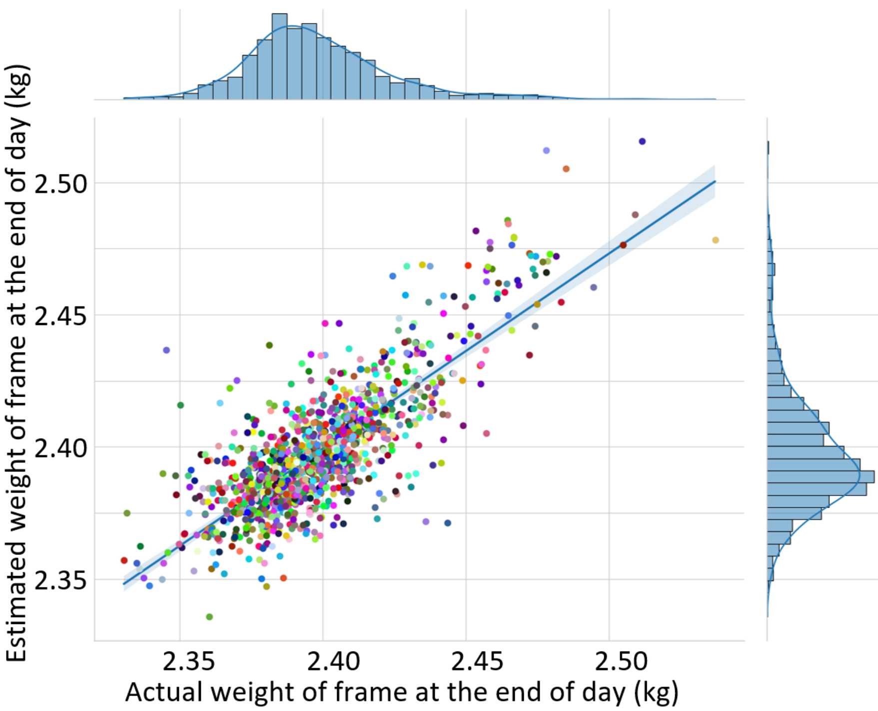

Figure 4.4: The scatter plot of actual weight of frame at the end of day, against the estimated weight for the same. These results for 1200 days in the dataset are obtained after merging the test results of all 5 folds (Table 4.4). The Pearson correlation between the two is r = 0.794, with p less than 0.001.

4.6.1 Performance on Unseen Sensors/Hives

The performance of deep networks is put to real test when they encounter unseen data. An issues with electronic sensor data is that sensors tend to add a specific bias to collected data, and this bias varies even between the sensors of the same type. In long term deployment of sensors, miss-calibration is a common problem and the biggest contributor to the sensor bias. This different bias of different senors for each system acts as a signature for the system, which deep networks can exploit and overfit for the systems in the training set. In a random shuffle, the data collected by each sensor system is available in both the training set and the test set. So instead of using a randomly shuffled dataset for k-fold cross-validation for training and testing, a system-fold (with 8 systems) cross validation is used where each fold contains the data collected from one particular sensor system.

System-fold cross validation allows thorough testing of the performance of the deep learning model on unseen data. For this, the models were trained using data collected by all sensor systems except for one, and the performance was tested on the data from the one system which was not used for training. This was repeated 8 times to test data from each sensor system. As sensor systems are allocated to specific hives, this allowed the evaluation of *WE-Bee* on hives which were not part of the training set. The training sets were shuffled whereas test sets were not. The non-shuffled test sets allow the graphical visualisation of the long term weight tracking capabilities of *WE-Bee*.

Table 4.5: Results from training on multiple hives and testing on an unseen hive. The details of data collected from each system are reported in Table 4.2.

System ID	MSE (grams/frame)	MAE %	Variance of % errors
11	20.7	0.79	1.24
13	16.8	0.73	0.98
14	12.2	0.58	0.66
15	10.3	0.54	0.51
16	16.8	0.72	1.04
17	7.6	0.38	0.27
18	16.9	0.65	0.90
19	21.7	0.69	1.23
Avg	15.4	0.64	0.85
Std. Dev.	4.9	0.13	0.35

The sensor system specific results are reported in Table 4.5, where the first column indicates the system ID which was used for testing, but not for the training. A more realistic deviation in these results can be observed, with relatively high MSE and percentage MAE compared to those reported in Table 4.4. Errors computed for each sensor system were pooled together and Figure 4.5 shows the histogram of errors in grams per frame between labels and estimations for all the sensor systems combined. However this histogram is only for the error at the end of each day, representing the propagation errors for 1200 days. The Gaussian distribution of these errors indicates that there is no major bias in weight estimations. The average weight of hives per frame in this dataset is 2,390 grams. Majority of daily errors in the histogram are between ± 20 grams per frame per day, which is less than 1% of average weight per frame.

4.6.2 Performance on Cumulative Estimation

The cumulative weight estimation capability of *WE-Bee* for unseen data was tested on data collected using sensor system 14. This system was deployed in hives for a total of 200 days, most for any system at the time of this study. The test set in this case consists of data collected via system 14 only, and has not been shuffled to preserve the order of days. As a first step, all the frame weight variation labels for sensor system 14 were stitched together. This was followed by converting the frame weight variations into hive weight variations, by multiplying them with the actual number of frames in hive on each day. This process was repeated for the daily frame weight variation estimations generated by *WE-Bee* as well. The starting offset for both the labels and estimations was set using the actual weight of the hive measured on day-0. This generated two sequences, one for hive weight labels and other for hive weight estimations as shown in Figure 4.6.

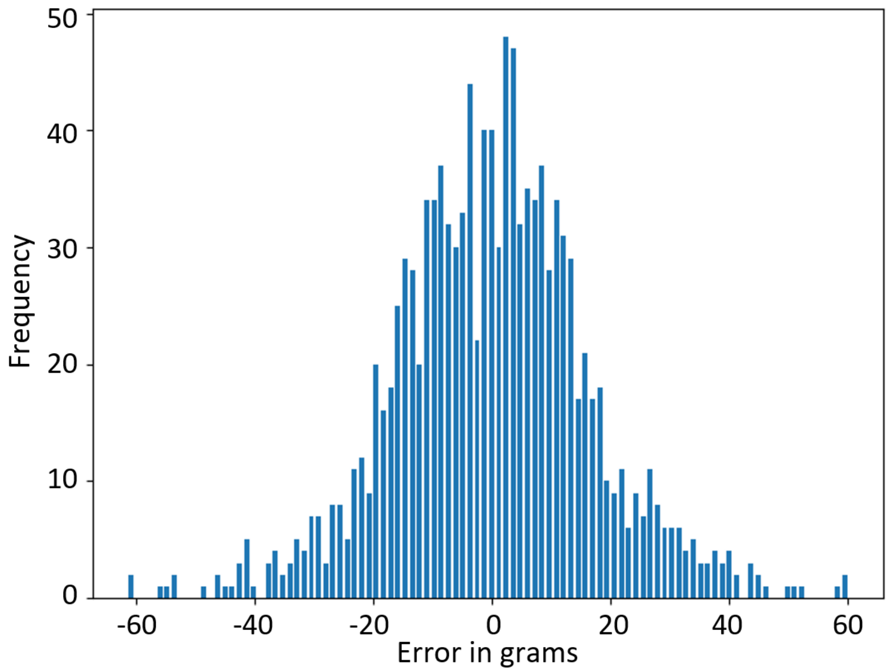

Figure 4.5: Histogram of error (in grams per frame) between final label and final estimation for each day (propagation error), for the 1,200 days in the dataset. Results obtained after individually testing each system (Table 4.5) are pooled together to obtain the error histogram.

The sharp increase in the weight around day-5, as shown in Figure 4.6, is a result of a beekeeper merging two hives together to make a stronger colony, which changed the size of the hive from 2×10 frames to 3×10 frames. The weight estimates are reasonable till around day-30, after which the model over-estimates the daily weights till day-97. The major reason for the over-estimations in this period is the limited training data. System 14 along with system 15 were the only two systems deployed at Site-A. During this period, other systems were not deployed in any hives. With system 14 being used for testing, the network only has data from system 15 available for training for this time of the year. The variations in weight are very season specific, and the lack of training data for this season makes the network under perform. After day-97, the system 14 was deployed in new hive of size 2×8 frames at Site-B, where the other seven systems were also deployed to collect data form beehives of varying strengths. With diverse data available for training, the performance of *WE-Bee* improves drastically. The trends of labels and estimations between day-98 and day-200 in Figure 4.6 are very similar.

WE-Bee is designed to estimate the weight of hive based on activity of bees picked up by the sensors. Sharp changes in the weight as a result of external agents, such as hive manipulations by beekeeper cannot be estimated using these sensors. There are two methods to manage such weight changes. First is to update the model with the total number of frames inside the modified structure, and let the model estimate the new net weight based on its old weight estimates per frame. This however is based on the assumption that the composition of honey, pollen, and bees in the new structure has not changed. This composition usually changes after major hive manipulations as honey filled frames in the hive are often replaced with empty or partially filled frames. This can lead to an error in net weight estimation. Also, if daily estimations until hive manipulation have an accumulated error, that error will also propagate to the estimates of new weight of modified hive structure. Second approach is to manually update the model with new net weight of the hive, and the model can then continue the estimations from that new baseline weight. This will limit the duration and the magnitude of the propagation error between consecutive inspections. However the new weight of the hive either has to be measured using a weighing scale, the instrument this work aims to eliminate, or the beekeeper has to take a calculated guess about the hive weight. In our experience, beekeepers can take a reasonably accurate guess about the weight of a hive once they have seen the status of frames and bees during the hive inspection, and know the extent of their manipulation. They can update the net weight of the hive via a user interface, which can be used as the new baseline weight for the hive.

The weight of hive at day-98, when sensor system 14 was moved to a new hive, is calculated using the first approach discussed in the paragraph above. The model uses the last estimated weight from the previous hive with a total of 30 frames (day-97), and the size of the new hive with a total of 16 frames to calculate the new weight baseline. This

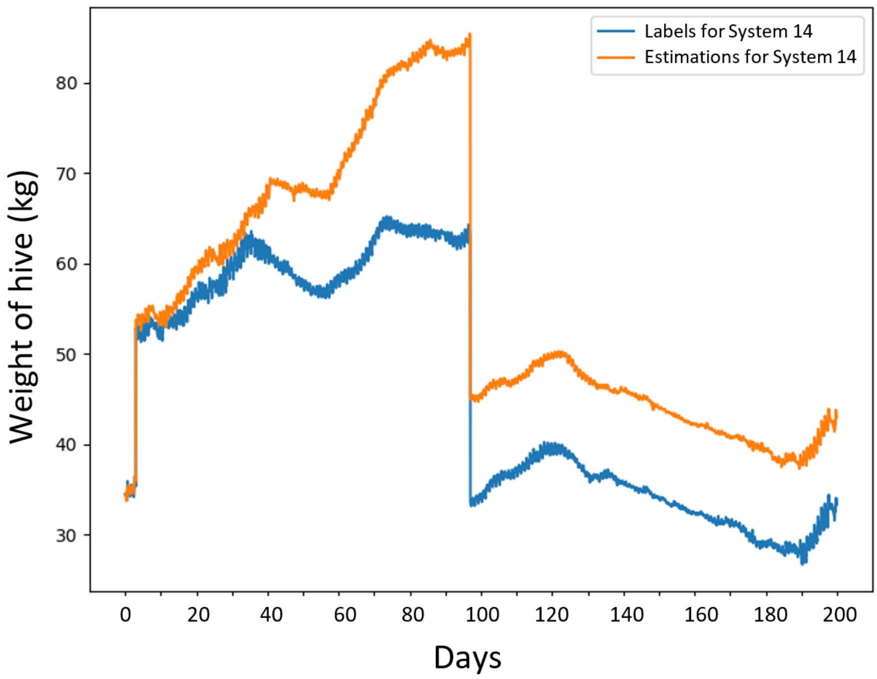

Figure 4.6: The weight of hive(s) with sensor system 14 estimated for 200 days. Between day-30 and day-97, *WE-Bee* over-estimates the daily weight change, and the error accumulates over time resulting in diverging patterns. System 14 is then deployed at a different hive on day-98, and the size of hive changes from 3×10 frames to 2×8 frames, resulting in a sharp drop in the weight. From day-98, the weight variation estimations are quite accurate till day-200, and only the previously accumulated error can be seen propagating during this period.

results in an error of net weight estimation on day-98. Despite this inherent error (mostly propagation), *WE-Bee* correctly estimates the trend of change in weight for more than 14 weeks after day-98. This trend is often an adequate piece of information for the beekeepers.

4.7 Summary

This chapter investigated a deep learning model for the task of estimating daily weigh variations of a beehive. The dataset used for training the machine learning model was explained, along with the designed model. The proposed model WE-Bee used LSTM encoders and decoders and was trained on 6 months of collected data. The features for training were selected after an in-depth study of bee behaviour, and the impact of environment on bee foraging activity. The performance of model was thoroughly tested using standard k-fold cross validation, and using system-fold cross validation to assess the performance on data from unseen hives/sensors. The model showed an average mean absolute error of 0.64% for estimating a total of 57,600 weight points for 1,200 days in the dataset. The results of daily weight estimations using WE-Bee showed decent accuracy on a model that is easy to train using a single GPU with 8 GB of memory. The cumulative estimation also showed promising results over multiple weeks for beehive weight estimation.

Chapter 5

A Deep Learning Model to Optimize Beehive Monitoring System for the Task of Daily Weight Estimation

Previous chapter discussed the use of LSTM encoders and decoders for the task of beehive weight estimation. The results demonstrate that deep learning can be used for estimating daily beehive weight variations with a good accuracy. In this chapter, we ask whether the same results can be achieved on a bigger dataset, but on an optimised system with fewer sensors. Fewer sensors reduce the cost and complexity of the monitoring systems. To investigate this, self-attention encoders are used to analyse the role of individual sensors in daily beehive weight variation estimations. The information about contribution of different sensors and features for a specific task can help design more efficient and cost effective beehive monitoring systems, by removing the non-contributing sensors and features from the design. This knowledge about contribution of different sensors will also improve the explainability of the AI, which can help build the confidence of beekeepers in the use of machine learning for beehive weight estimation.

5.1 Background

Deep learning models have shown promising results on performing different tasks on complex time series data [140]. Authors in [141] proposed the use of transformers, a deep network consisting of encoders and decoders with multi-head attention for predicting the next correct word in machine translation. Transformers have since been used for prediction in machine vision [142], bio-informatics [143] and many other areas.

Authors in [144] use multiple streams of encoders in parallel, where each encoder pays attention to only one resolution of input in each speech frame for the task of speech recognition. For beehive sensor data for an entire day, there are two aspects that need attention. The relationship between data/features from different sensors at a particular

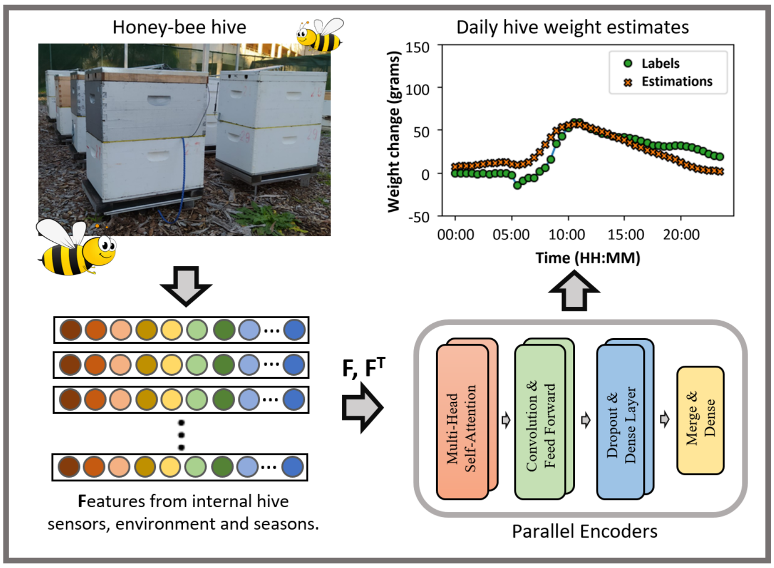

Figure 5.1: *Apis-Prime* uses internal hive sensors, environmental features, season information, and hive size to estimate the daily weight variations of a beehive per hive frame using two self-attention encoders in parallel.

time instance, and the relationship between data/features from a particular sensor for different times of the day. This chapter proposes a deep neural network *Apis-Prime*, which comprises two parallel self-attention encoders (without any decoders) to exploit these two independent relationships, for the task of beehive weight estimation. In chapter 4, attention layer was used between the encoder and decoder to help decoder focus on more important encodings generated by the encoder. In this chapter, self-attention layer at the input looks for dependencies within (self) the input data to identify what parts should be paid more attention to. This helps make most of the input data which is rich with information.

To optimize the system for the task of daily weight estimation, this chapter uses the attention weights of trained encoders of *Apis-Prime* to evaluate the sensors used by the monitoring systems. This evaluation is used to identify and remove the unnecessary sensor data/features from the dataset, reducing the number of sensor features from 36 to 23, hence providing a significant optimization. We provide a performance analysis of beehive weight estimations by *Apis-Prime* using the complete, as well as the optimized dataset on 2,170 days of beehive sensor recordings. Equally good results of daily weight estimations using the optimized feature set demonstrate the efficacy of proposed model for optimization of beehive monitoring systems for the task of weight estimation.

5.2 Network and Experimental Setup

The data used to train and test *Apis-Prime* was collected over a 12 month period using all 8 sensor systems (March 2021 - March 2022). The average weight of empty hive structure (not part of the dataset) is the same as reported in previous chapter i.e. 1.06 kg per frame, whereas the average weight of hives with structure, pollen/nectar and bees in this dataset has increased to 2.69 kg per frame. Same set of features is used to train *Apis-Prime* as discussed in previous chapter (see Table 4.1). The details of data collected using each sensor system for this dataset are given in Table 5.1.

The architecture of *Apis-Prime* is based on the encoder of transformer [141]. Transformers composed of encoders and decoders along with multi-head attention mechanism were first introduced for the task of machine language translation. *Apis-Prime* uses only the encoder part of transformer with some modifications to estimate the weight data, using the data from other sensors. Previous sections explain how weight of a beehive is dependent on many different factors. *Apis-Prime* exploits these dependencies to estimate/predict a series of weight values based on time series data collected from internal hive sensors and other relevant information. Input to the network is a set of daily data collected from internal hive sensors such as temperature, humidity, atmospheric pressure, CO_2, acoustics and vibrations. Information about the weather, week of the year (seasonal information), time of the day, and the size of hive is also part of input. Self-attention layer looks for dependencies within (self) the input data to identify what parts should be paid more attention to. This helps make most of the input data which is rich with information. The network is trained to generate the daily weight variations (per frame) of a hive as an output, the labels of which are obtained using custom built weighing scales deployed with each sensor system. All the inputs are normalized to create a feature matrix $\mathbf{F}$ of size 48×36, where 48 is the number of data samples per day and 36 is the number of features (see Table 4.1) extracted from different sensors within each data sample. Figure 5.2 shows the network architecture of *Apis-Prime*, with parameter settings provided in Table 5.2.

We use the weight of a hive at midnight as a reference, and any change in hive weight

Table 5.1: Break-down of data collected (days) for training and testing of *Apis-Prime*. A total of 2,170 days of sensor data has been collected from 3 different sites, using 8 sensor systems. System 14 and 15 were deployed to collect data from November 2020, whereas rest of the systems were deployed from March 2021. The data used in this chapter was collected till the end of March 2022.

System ID	Site-A (days)	Site-B (days)	Site-C (days)	Total days
11	-	232	59	291
13	-	182	70	252
14	97	155	37	289
15	105	124	-	229
16	-	214	96	310
17	-	201	50	251
18	-	191	86	277
19	-	182	89	271
Total	202	1481	487	2170

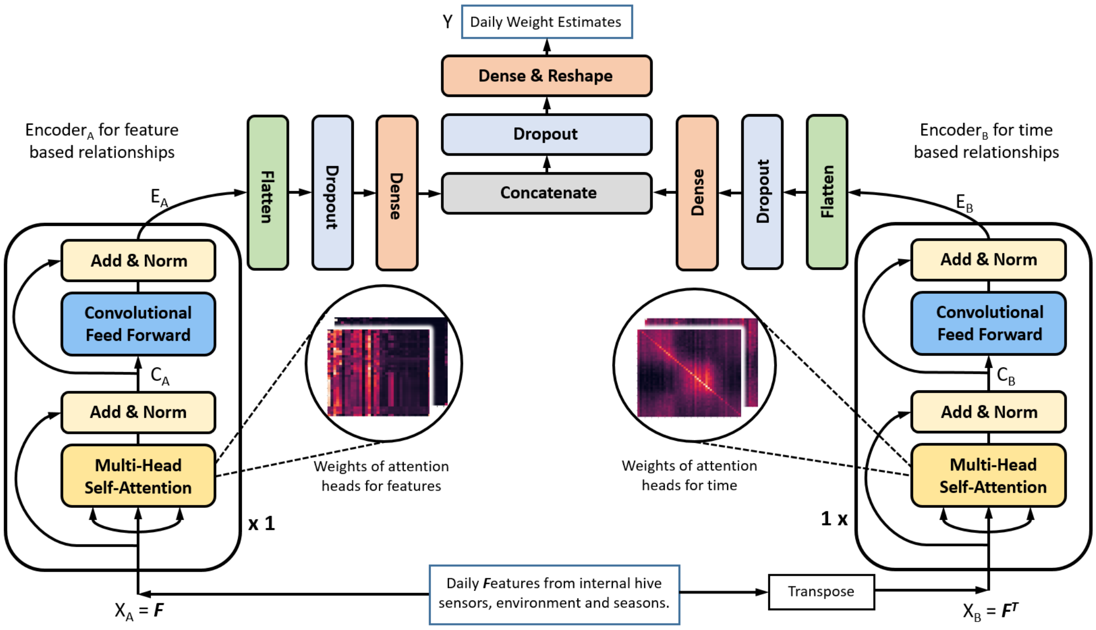

Figure 5.2: The network architecture of *Apis-Prime*, where sensor features of entire day are fed as inputs. The multi-head self-attention encoder on left exploits the feature based relationships, whereas the encoder at the right exploits the time based relationships in the daily sensor data. The outputs of these encoders are passed through flatten, dropout and dense layers before they are concatenated. Another dropout layer followed by dense and reshape layers generate the weight variation estimates for the entire day.

between midnight and a time instance in next 24 Hr window is dependent upon all the bee activity and environmental conditions till that time instance. The time-series data/features from all the sensors for an entire day (midnight to midnight) are shaped into a 2-dimensional matrix, with features on one axis and time on the other. Self-attention encoders use this 2-dimensional matrix as input, and exploit the relationships and dependencies within this

Table 5.2: Design and training parameters for *Apis-Prime*.

Parameter	**Value**
Number of encoders	2
Attention heads per encoder	2
Head size	512
Convolutional filters per encoder	512
Filter/kernel size	7
Convolution activation function	tanh
Dropout	0.8
Activation function of dense layers	linear
Max training epochs	2000
Early stopping patience	250
Batch size	16
Loss function	MSE
Learning rate scheduler	Decaying sinusoidal
Learning rate optimizer	Adam

data. Self-attention encoders and decoders were originally proposed for the task of language translation [141]. The dependencies and relationships within a sentence exist only in a single dimension, as words in a sentence appear in a sequence one after the other. Hence, self-attention encoders by design pay attention to data in one dimension. However in case of multi-sensory data from beehives, the relationships exists in two dimensions; one across different sensors and the other across time. To exploit the relationships in both these dimensions, *Apis-Prime* uses two self-attention encoders in parallel, but with different inputs. $Encoder_A$ in *Apis-Prime* uses feature matrix (**F**) of entire day as input, and exploits the relationships between different features across sensor data. Whereas $Encoder_B$ uses the transpose of feature matrix ($\mathbf{F}^T$) as input, and exploits the relationship between features across different times of the day. This simple manipulation of transposing the 2-D input matrix enables the two encoders to pay attention to relationships in different dimensions. This also enables the evaluation of the importance or usefulness of different sensors in the system, and different time periods for data collection during the day. The use of attention weights of the two trained encoders for evaluation and optimization of system is discussed in the last section.

Let X_A and X_B be the inputs of $Encoder_A$ and $Encoder_B$ respectively, and F be the matrix of size 48×36, consisting of daily features from hive, environment and weather. The inputs to the encoders are defined as

$$X_A = F \tag{5.1}$$

$$X_B = F^T \tag{5.2}$$

For $Encoder_j$ with j = [A, B] and
$Head_{ij}$ with i = [0, 1] for the two heads of each encoder

$$Head_{ij} = SelfAttention(Q, K, V) \tag{5.3}$$

where

$$Q = X_j W_{ij}^Q \tag{5.4}$$

$$K = X_j W_{ij}^K \tag{5.5}$$

$$V = X_j W_{ij}^V \tag{5.6}$$

where W_{ij}^Q, W_{ij}^K, and W_{ij}^V are independently learned projection matrices corresponding to Query, Key and Value respectively for each head for each encoder.

$$SelfAttention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5.7}$$

where d_k is the dimension of Q, K and V, which is set to 512 for this network.

For $Encoder_A$

$$MH_A = MultiHead_A(Q, K, V) = Concat(Head_{1A}, Head_{2A})W^A \tag{5.8}$$

$$C_A = Normalize(X_A + MH_A) \tag{5.9}$$

$$E_A = Normalize(Conv1d(C_A) + C_A) \tag{5.10}$$

For $Encoder_B$

$$MH_B = MultiHead_B(Q, K, V) = Concat(Head_{1B}, Head_{2B})W^B \tag{5.11}$$

$$C_B = Normalize(X_B + MH_B) \tag{5.12}$$

$$E_B = Normalize(Conv1d(C_B) + C_B) \tag{5.13}$$

where W^A and W^B are the learned projection matrices and E_A and E_B are the outputs of two encoders. These outputs are passed through layers to obtain D_A and D_B, where

$$D_A = Dense(Dropout(Flatten(E_A))) \tag{5.14}$$

$$D_B = Dense(Dropout(Flatten(E_B))) \tag{5.15}$$

and the final output Y, which is a vector of size 48×1 containing the daily weight variation estimates per hive frame is obtained as

$$Y = Dense(Dropout(Concat(D_A, D_B))) \tag{5.16}$$

We used Keras to implement Apis-Prime and used the same system described in previous chapter to train the model. The network of Apis-Prime has approximately 5.2 million trainable parameters. This network was also trained using the data collected from 7 sensor systems, and tested on the data collected from the 8^{th} sensor system. This process was repeated 8 times to test the performance of *Apis-Prime* on all 8 sensor systems. The MSE of test set was monitored during the training with an early stopping in place to avoid over-fitting.

5.3 Experimental Results

The same approach of using the system-fold cross validation described in Section 4.6.1 of previous chapter is used to evaluate *Apis-Prime* as well as *WE-Bee* on the bigger dataset for comparison. Use of system-fold cross validation allows thorough testing of the performance of the models on unseen data. For this, the models were trained using data collected by all sensor systems except for one, and the performance was tested on the data from the one

Table 5.3: Results from system-fold training on multiple hives and testing on unseen hives using *Apis-Prime*, and *WE-Bee*. The average results represent the scores for all 2,170 days with 48 points per day, making a total of 104,160 weight points in the dataset. Bold font indicates the better score between the two models. The details of data collected from each system are reported in Table 5.1.

Test System	**Test error (grams)**		**Absolute percentage error (Mean)**		**Absolute percentage error (Variance)**	
	Apis-Prime	WE-Bee	Apis-Prime	WE-Bee	Apis-Prime	WE-Bee
11	**24.1**	26.2	**0.60**	0.62	0.75	**0.86**
13	**28.5**	30.2	**0.65**	0.67	**0.80**	0.91
14	**13.7**	15.4	**0.47**	0.54	**0.43**	0.54
15	12.1	**11.3**	0.62	**0.56**	0.91	**0.68**
16	**16.8**	18.7	**0.51**	0.54	**0.50**	0.63
17	**17.0**	17.2	**0.48**	0.49	**0.42**	0.46
18	**26.9**	27.5	0.60	**0.59**	0.70	**0.69**
19	**18.8**	21.9	**0.47**	0.51	**0.46**	0.54
Average	**19.7**	21.05	**0.55**	0.56	**0.62**	0.66
Std. Dev.	**6.29**	6.55	0.08	**0.06**	0.20	**0.16**

system which was not used for training. This was repeated 8 times to test data from each sensor system. The training sets were shuffled whereas test sets were not. The non-shuffled test sets allow the graphical visualisation of the long term weight tracking capabilities of *Apis-Prime*.

Table 5.3 reports the sensor system specific results for both models. The first column is for the ID of system used for testing. The reported test errors are the Mean Square Errors (MSE) on test set, in grams per frame. The absolute percentage errors are also calculated for the entire test set, using the actual and estimated net daily weights of the frame. The mean of these absolute percentage errors for all the hives is consistently less than 1 percent, which is very promising. In most cases, *Apis-Prime* shows better weight estimation capabilities than *WE-Bee*. Only for the data collected using system with ID 15, *WE-Bee* shows a slightly better performance.

For long term tracking of the hive weight, the error at the end of the day is particularly important, as it propagates to the weight estimations for next day(s). The daily errors computed for each sensor system were pooled together, and Figure 5.3 shows the histogram of errors between labels and estimations for all the sensor systems at the end of each day, representing the accumulated propagation error for each day in a dataset of 2,170 days. The Gaussian distribution of errors indicate that there is no major bias in the estimations, and 83.5% errors at the end of the day are within 25 grams per frame. With an average hive frame weight of 2,687 grams in the dataset, this equates to less than 1% in daily error for 83.5% days.

Some examples of estimated weight variations per frame by *Apis-Prime*, as well as the label weight variations per frame from the test set are shown in Figure 5.4. A total of

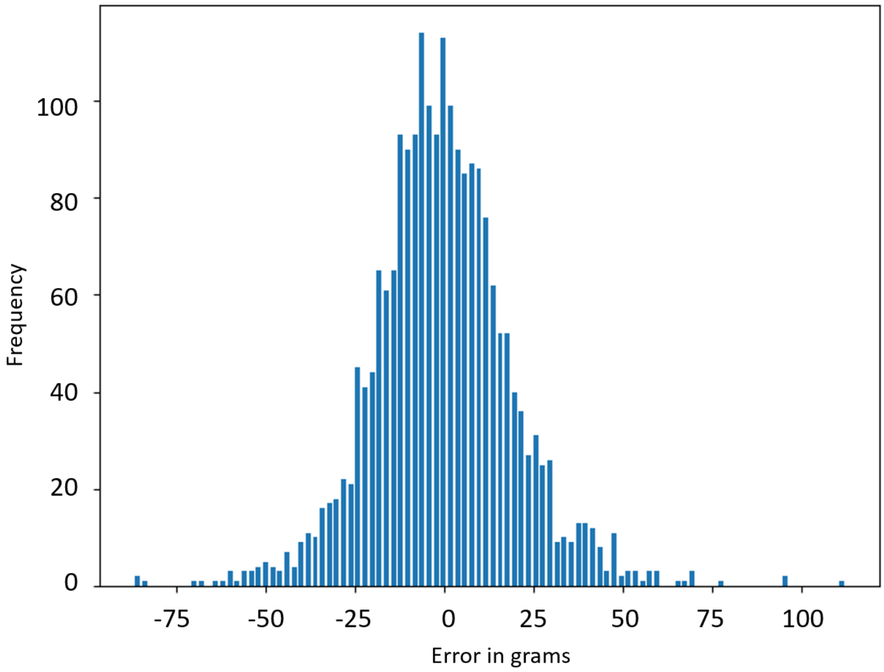

Figure 5.3: Histogram of error (in grams per frame) for *Apis-Prime* between final label and final estimation for each day (propagation error), for the entire dataset of 2,170 days. Results obtained after individually testing each system (Table 5.3) are pooled together to obtain this error histogram. A total of 357 out of 2,170 days have a daily error of more than 25 grams per frame.

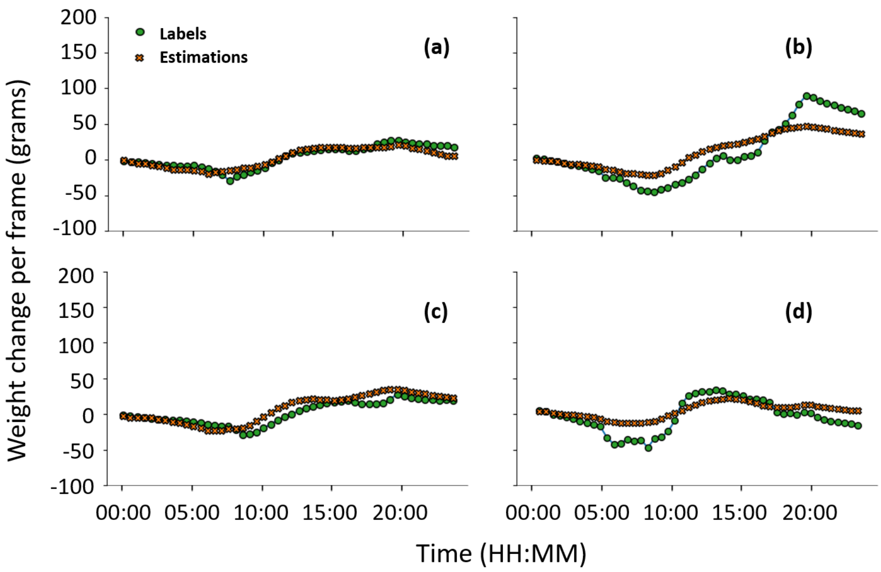

Figure 5.4: Test-set examples of daily weight variation labels per hive frame, and the estimations for the same using *Apis-Prime*. First weight reading for each day at 00:00 is the reference for variations throughout the day. Daily weight variation estimations leading to a negligible error at the end of the day are shown in (a) and (c). An under-estimate of the daily weight can be observed in (b). Example of an over-estimate of the daily weight on a day with noisy weight data is shown in (d).

48 estimations are generated for each day, with the weight at midnight (00:00) as starting reference for each day. The daily estimations can be divided into two categories. One where errors for all estimations within a day add to a negligible error by the end of the day, as shown in examples of Figure 5.4 (a) and (c). The second category is where the accumulated error for all 48 estimations within a day leads to an under-estimate or over-estimate of weight variation, as shown in examples of Figure 5.4 (b) and (d) respectively. As discussed above, the greater the over/under-estimate at the end of the day, greater the error that propagates to the estimations for following day(s) in long term or cumulative tracking.

5.3.1 Performance on Cumulative Estimation

The error in the weight estimation at the end of a day propagates to the weight estimations for next days as discussed in Section 4.6. The scatter plot of actual against estimated weight change per frame per day is shown in Figure 5.5. These estimations show bias when the actual weight change is less than -25 grams per frame, and when the actual weight change is greater than 75 grams per frame. Given the average weight of a hive frame at 2,687 grams

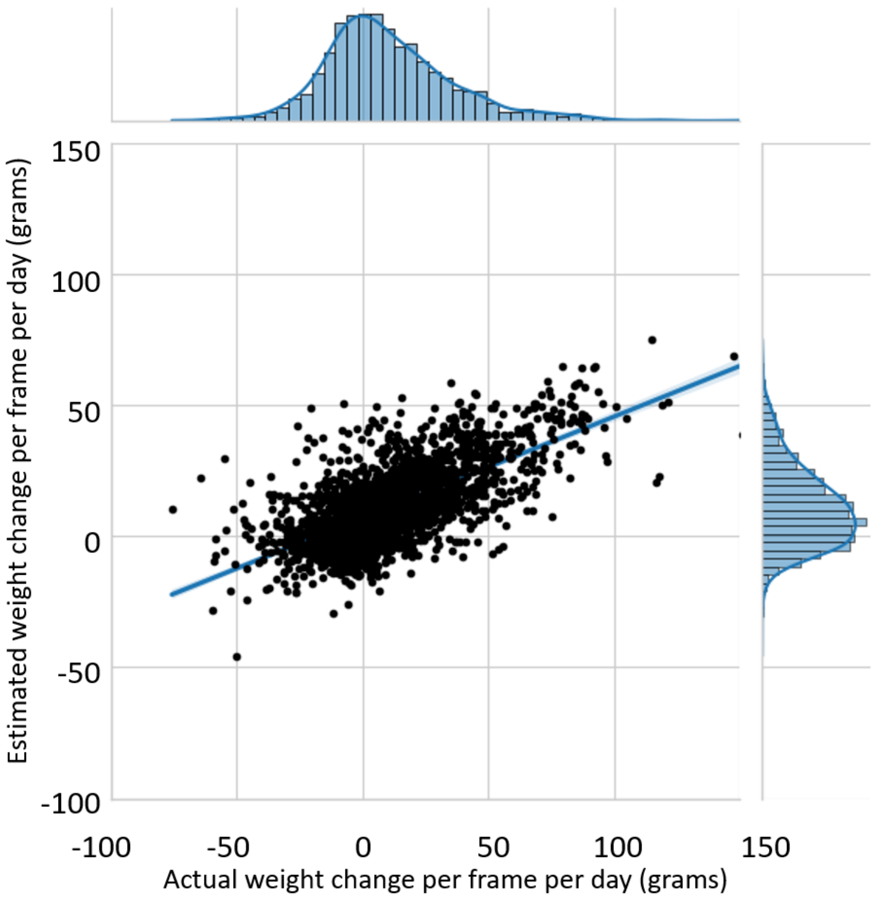

Figure 5.5: The scatter plot of actual weight change of a hive frame at the end of day, against the estimated weight change for the same. These results for 2,170 days in the dataset are obtained after merging the test results of all 8 systems tested separately (Table 5.3).

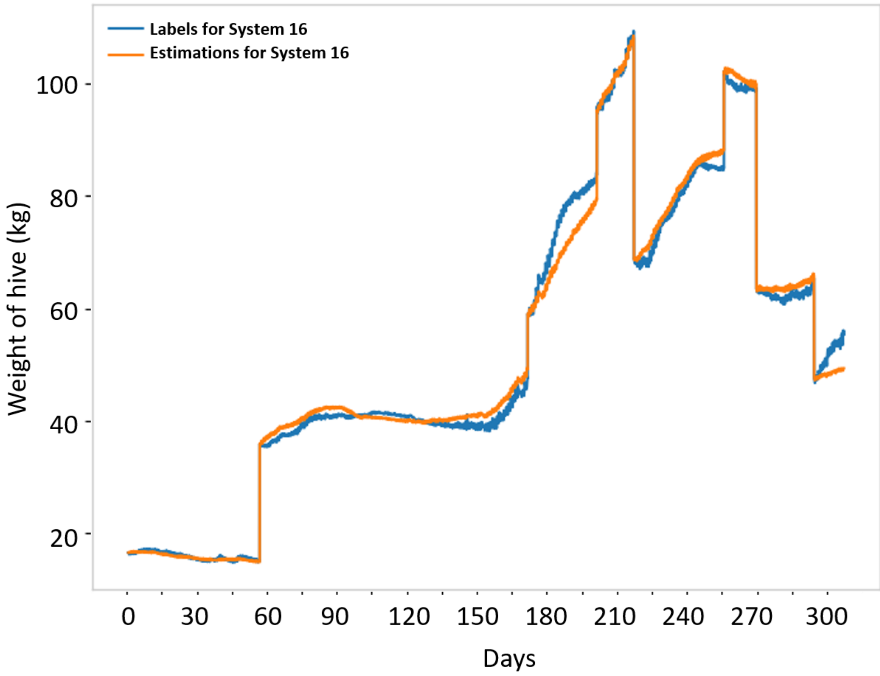

Figure 5.6: The weight of a hive with sensor system 16 estimated for 310 days. The sharp changes in the weight represent the structural changes in the hive, such as addition/removal of honey chambers or swapping of frames with other hives. Upon such change, the weight estimation model is provided with a new reference (true weight of the hive). *Apis-Prime* uses this new baseline/reference, and daily weight variation estimations to continue estimating the total weight. From day-60 to day-180, the model continuously tracks the weight of hive for 120 days with very good accuracy, and shows reasonable accuracy for other smaller segments between consecutive hive manipulations.

in the dataset, the bias error of around 50 grams and -75 grams at the extremes equate to an error of around 2% and -3% respectively, which is very reasonable for an estimate.

The cumulative weight estimation capability of *Apis-Prime* for unseen data is shared on data collected using sensor system 16. This system was deployed in hives for a total of 310 days, more than any other sensor system used in the dataset. The test set in this case consists of data collected via system 16 only, and has not been shuffled to preserve the order of days. The same methodology explained in Section 4.6.2 was used to test the cumulative estimation capability of *Apis-Prime*. This generated two sequences as shown in Figure 5.6, one for hive weight labels and other for hive weight estimations.

The sharp increase in the weight around day-60, as shown in Figure 5.6, is a result of a beekeeper adding a partially filled honey chamber to the hive, which changed the size of the hive from 1×8 frames to 2×8 frames. Around day-210, the beekeeper took two full honey chambers off from the hive, and replaced them with chambers containing empty frames to make more space for the honey. All other sharp changes in the weight are also a result of manipulations carried out by the beekeeper. These manipulations are a routine practice where a beekeeper changes the size of a hive based on his/her assessment of bee colony and the seasonal needs.

As discussed in Section 4.6.2, there are two methods to manage the sharp weight changes which originate from hive manipulations. The example of cumulative weight estimation in chapter 4 used the first method, whereas this case uses the second method. We update the base weight of the estimations using the actual weight of the hive after these manipulations, in order to test the long term tracking performance of *Apis-Prime*. If the actual weight of a hive is not available, the weight guessed by the beekeeper can also be used to update the base weight. A total of 8 instances of hive manipulation (including day-0), and update of base weight can be observed for sensor system 16 in Figure 5.6. The tracking of weights is very reasonable for a majority of 310 days. The only days where *Apis-Prime* significantly underestimates the weight gain is during the period around day-195 and from day-290 onwards. But even for these days, the network estimates the correct trend of weight change i.e. net increase in the weight of the hive. In most cases, the beekeepers are interested in finding out the correct trend of weight gain or loss, rather than the exact weight of the hive.

5.4 System Optimization Using Self-Attention Encoders

As discussed in the previous section, *Apis-Prime* uses two self-attention encoders to exploit the time based, and the feature based dependencies within the daily sensor data to estimate the daily weight variations. The high quality of these estimates show that *Apis-Prime* is capable of paying attention to the most relevant features in the dataset, and this attention can be used to evaluate the contribution of different sensors and features. The weights of self-attention layers of the encoders are updated during the training process, as the model gradually learns to pay more attention to important parts of daily sensor input and vice versa. Once the training of each system-fold is complete, we collect the self-attention weights (for a single head) from both encoders using 100 examples from the test set. This is repeated for all 8 folds for the systems, and then an average of these 800 attention weight matrices per encoder is computed.

5.4.1 Time-Based Evaluation

The map of time based attention weight average from $Encoder_B$ is shown in Figure 5.7. With a total of 48 samples per day, this 48×48 map provides insight into important parts of the day with a 30 min resolution. In Western Australia, hives gain most of their weight during summer/spring season, and the bees are very active in early parts of the day between 05:00 and 06:30. The foragers leave hive early in large numbers to scan the area for pollen and nectar availability, which results in sharp variation (decrease) in the weight of the hive. In figure 5.7, one can observe that the columns representing the sensor data collected during the early morning hours carry significantly higher weights when compared to other columns. The activity during early morning has a good correlation with other periods of

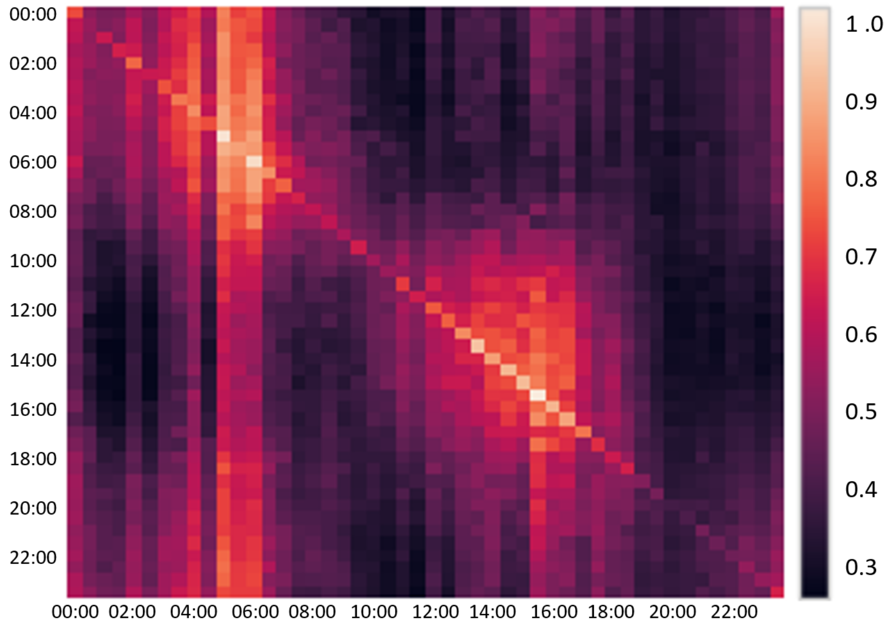

Figure 5.7: Daily attention weights map (with max values scaled to 1) for the time based self-attention Encoder$_B$ of *Apis-Prime*, with 48 data points per day. The higher values of $48\times$ 48 map during morning hours leading to 06:00 indicate that data collected at this time carries significantly more useful information for weight estimation.

the day, as this activity sets the tone for weight variations for the entire day. One can also observe that the foraging activity and resulting weight changes pick up in the afternoon and are at their peak around 15:30. On the contrary, the data collected around 20:00 is given minimum attention by the encoder. This is the time when bees have settled down in the hive at the end of the day, and are not bringing any more pollen and nectar. The temperatures at this time are also not that cold for bees to consume large amount of food to keep the hive warm. The weight of hives does not show much variation during this time period, hence the encoder pays less attention to data collected at this time. Based on these self-attention weights, it is possible to devise a variable sampling period for the sensor system, where the sensor data is collected at a faster rate during important periods of the day, and less frequently otherwise. This will optimize the power consumption of the system and enable collection of data with higher information content.

5.4.2 Feature-Based Evaluation

The weights for feature based self-attention in Figure 5.8 show that $Encoder_A$ focuses on exploiting very different dependencies when compared to those of $Encoder_B$. The top left quadrant of attention weights has significantly higher values when compared to other quadrants. This quadrant represents features such as time of the day, information about seasons as week number of the year, temperature features inside and outside the hive, weather data of hive site, CO_2, atmospheric pressure and some audio features of the bee

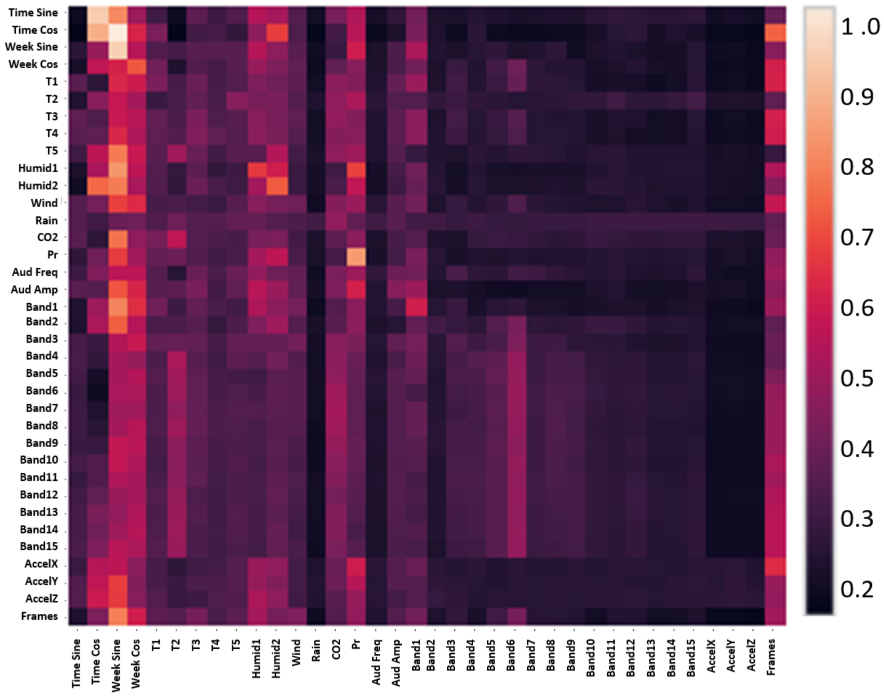

Figure 5.8: Daily attention weights map (with max values scaled to 1) for the feature based self-attention Encoder$_A$ of *Apis-Prime*, with 36 data features extracted at each time interval. The lower values of 36 $\times$ 36 map for some of the audio bands and accelerometer data indicate that these features carry significantly less useful information for weight estimation. The details of different features are provided in Section 4.4.

buzz from inside the hive. It is interesting that the two feature encodings of week number (sine and cosine), despite being constant throughout the day, have different attention weights assigned to them. Also note that not all of the audio features carry high weights. The column representing the weights for rain shows smaller values, and this is primarily because the rain in the dataset does not appear frequently. When averaged out over a longer period, the weights for infrequent events decrease significantly. The very last column, which is the number of frames in the hive, also carries high weights as the size of a managed hive is usually a good indicator of bee colony strength which greatly impacts the daily weight variations. The vibration data from the three axes of accelerometer carries the least weights. Detecting the vibrations caused by specific bee movements is very difficult because of the structure of commercial beehives, and has been discussed in detail in Section 3.3.6.

5.4.3 System Optimization for the Task of Weight Estimation

For the task of beehive weight estimation, the impact of data from accelerometer and some of the audio features needs further investigation. The attention maps indicate that these features carry little to no information contributing towards hive weight estimates. To validate this, another set of experiments was performed to estimate the daily weight variations where the features with low attention weights, as observed in Figure 5.8, were dropped from the dataset. This included 3 features of accelerometer, and 10 features of audio bands from Band6 to Band15. This resulted in a 36.11% reduction in the number of features, which reduced from 36 to 23 in the optimized dataset. Other parameters of experiment were kept exactly the same as the experiment with full set of features, as explained in Section 5.3. After training *Apis-Prime* with optimized dataset, new attention maps were generated. Figure 5.9 shows the 23 × 23 feature based self-attention map of trained $Encoder_A$ with optimized features. One can observe a significant increase in the overall values, and much fewer darker regions in the map. The question however remains if *Apis-Prime* trained using optimized features can perform at par with model trained with complete set of features.

To evaluate the performance of *Apis-Prime* trained using optimized feature set, weight estimates were generated for all the systems using same set of parameters as in earlier experiment. The comparison between the weight estimation results using full set of features, and optimized features is provided in Table 5.4. From the percentage errors computed using both sets, it can be observed that despite a reduction in the number of features from 36 to 23, the overall performance of the model in estimating the daily weight variations is very similar. In a number of cases, optimized feature set even out performs the full feature set. This implies that the attention weights of trained encoders can be used to fine tune the design of the sensor system. Completely dropping the accelerometer data did not impact the quality of the weight estimations. This means that the system cost can be reduced by not including the accelerometer in the design. The power consumed in the acquisition and

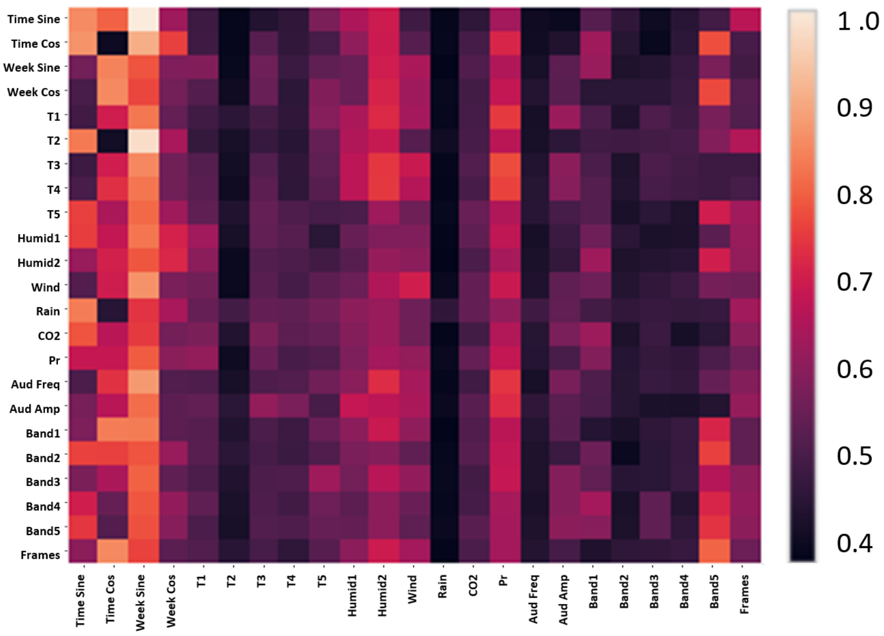

Figure 5.9: The attention weights map (with max values scaled to 1) of Encoder_A after reduction of features for *Apis-Prime*. With a total of 23 important features, the 23× 23 map shows much higher weight values for dependencies between different sensor features in the dataset. The quantitative results using reduced features are provided in Table 5.4.

Table 5.4: Comparative results between *Apis-Prime* using all of the sensor features, and the reduced features. Bold font indicates the better score between two feature sets.

Test System	Test error (grams)		Absolute percentage error (Mean)		Absolute percentage error (Variance)	
	All Features	Red. Features	All Features	Red. Features	All Features	Red. Features
11	**24.1**	24.4	**0.60**	0.61	0.75	**0.76**
13	**28.5**	28.7	0.65	**0.64**	0.80	**0.79**
14	13.7	**13.0**	0.47	**0.45**	0.43	**0.39**
15	**12.1**	13.0	**0.62**	0.63	0.91	0.91
16	16.8	**16.7**	0.51	**0.49**	0.50	**0.49**
17	**17.0**	17.8	**0.48**	0.49	**0.42**	0.47
18	26.9	**26.6**	0.60	**0.59**	0.70	**0.69**
19	**18.8**	19.4	0.47	0.47	**0.46**	0.47
Average	**19.7**	19.9	0.55	0.55	0.62	0.62
Std. Dev.	6.29	**6.19**	0.08	0.08	0.20	**0.19**

processing of the accelerometer data can also be saved. By dropping multiple features from audio data, the data transmission bandwidth and costs can be saved, as well as the processing power consumed in extraction of these features can be reduced. With a reduction of 13 features of size 2 bytes each, the data to be transmitted in each cycle will also reduce from 91 bytes to 65 bytes.

For remote deployment of beehive monitoring system, the reduction in system cost, power consumption, and data transmission bandwidth are considerable gains. This is only made possible with the use of self-attention encoders. The optimization carried out in this study is very specific to the task of beehive weight estimation. Similar optimizations can be performed for other tasks, where another set of sensors and features may come out as the best choice. For a beehive monitoring system designed for multiple tasks, self-attention encoders can help identify a superset of sensors contributing towards different tasks. Whereas for a monitoring system designed for a very specific task, self-attention encoders can be used to identify the minimal sensors, which will minimize the cost and maximize the performance of beehive monitoring systems.

One of the problems identified using attention maps is the low contribution of accelerometers in hive weight estimation. Theoretically, the vibrations inside the hive should contain the information related to honeybee foraging activity, which is an important parameter for hive weight change. However such information is not available in collected accelerometer data either due to poor placement and connection of accelerometer, or because of its low resolution. Problems with detecting these vibrations have been discussed in detail in Section 3.3.6. The low contribution from accelerometers requires further investigation. In the final stages of this study, an improved sensor system was designed and developed to overcome the problems identified during the system deployments for data collection. The new compact dimensions of the sensor board allows for better, more stable clamping onto the hive frame, and increase the chances of detecting the weak vibrations generated by honeybee waggle dance using an accelerometer.

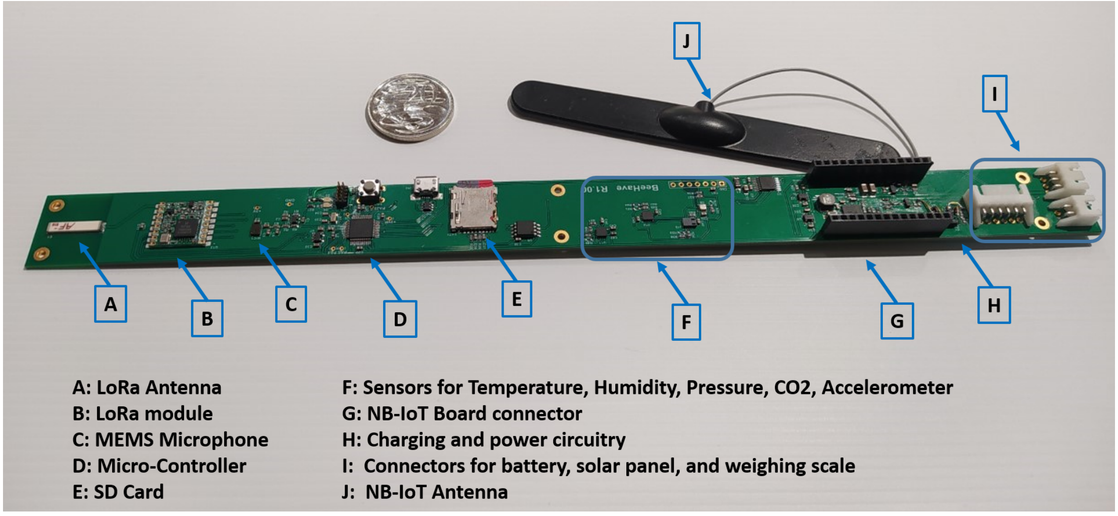

Figure 5.10: The new sensor board developed for beehive data collection, where sensors are better positioned to collect more hive information. All the components are soldered directly onto a single PCB, except for the NB-IoT board.

The firmware of the new system has also been upgraded for improved feature extraction from audio data. As discussed in Section 5.4.3, the 100 Hz sub-bands of bee buzz from 500 Hz to 1500 Hz contain significantly less information when compared to the sub-bands below 500 Hz. Based on this evaluation, the new firmware uses smaller 50 Hz sub-bands of audio to monitor bee buzz frequencies between 100 Hz and 850 Hz. This generates the same amount of audio features for transmission, but focuses on a narrower audio band with better resolution.

In the new sensor board (shown in Figure 5.10), all the sensors are better placed on the Printed Circuit Board (PCB) to avoid the impact of charging current, and to shield the sensors from other noises. All the components and sensors in this design are soldered directly on a single PCB, rather than smaller sensor boards soldered to a mother PCB. The only exception is the NB-IoT board which can easily be plugged/unplugged to the system. However due to the time limitations in this PhD project, the improved sensor board has not yet been tested in the beehives.

5.5 Summary

This chapter proposed Apis-Prime, a hybrid deep learning model for soft sensing and time series forecasting to estimate the daily weight variations of beehives. The results show an average error of 19.7 grams/frame for estimating a total of 104,160 weight points for 2,170 days in the dataset, compared to 21.05 grams/frame of earlier proposed model on same dataset. The cumulative estimation for extended periods also shows promising results, where the network demonstrated good tracking of actual hive weight for 120 consecutive days. The data was collected from hive sites at a significant distance from each other for geographical diversity. This diversity in the training data played a significant role in the quality of estimations, with an average mean absolute percentage error of 0.55% on the test set. From daily estimates, 83.5% days have errors of less than 25 grams per frame at the end of day. Furthermore, this work uses deep learning for the first time to optimize the design of beehive monitoring systems. The attention weights of self-attention encoders were used to gain insight into important parts of day for data collection, as well as the features which are less useful for weight estimation. Removing the features with minimal contribution from the dataset reduced the total number of features from 36 to 23, the size of each data transmission from 91 bytes to 65 bytes, while providing equally good results for weight estimation. This validates the additional usefulness of self-attention encoders for feature selection. These results demonstrate that fit for purpose, robust weight estimations for beehives can be achieved in real-world setting using low cost sensor systems and deep learning models. This study shows the potential of deep learning for improving the design of any hardware system using inverse design. Systems designed for specific tasks can be redesigned/optimized after proper evaluation using deep learning to identify the strengths and weaknesses.

Chapter 6

Conclusion

This thesis aimed to answer three questions related to the design of beehive monitoring systems. The first question was about the *selection of appropriate sensors for the design of a low power and long range beehive monitoring system.* This work examined the sensors used in beehives and analysed the monitoring systems proposed during the last ten years. Different sensors were evaluated for their relevance, hive deployment feasibility and the most appropriate sensors for remote hive monitoring were identified. This included sensors for temperature, humidity, atmospheric pressure, acoustics, vibration, carbon dioxide and weighing scales. The analysis of monitoring systems in literature facilitated the design and development of a Beehive Data Acquisition System (BeeDAS). A total of eight systems were deployed in beehives to thoroughly evaluate the data communication, power consumption and beehive feasibility of the designed system. Test deployments using NB-IoT for long-range communication were successful in the areas with no 4G coverage, and the solar power was adequate to keep the system operational for months. These traits of the system validate the selection of sensors, feature extraction processes, and the overall design.

Another contribution of this work is the collection of a diverse, high resolution beehive sensor dataset through deployment of BeeDAS systems in hives located at different geographical sites, and in varying environmental conditions. The collected dataset was first tested for its effectiveness using random forests to estimate the daily weight change of a hive. Daily weight change of a hive is a result of multiple complex factors ranging from bee activity inside the hive, to weather conditions outside the hive. The results of weight estimations demonstrate that the sensors used in BeeDAS, and the collected data features adequately capture the bee activity inside a hive. Audio data from inside the hive along with the external humidity and temperature were found to be specifically important in estimating the hive weight change using random forests.

A novel contribution of this work is the *use of soft sensor predictions for expensive and difficult to use components of a monitoring system*. This work explored the use of deep learning for estimating the weight variations of a beehive throughout a day. A hybrid model for soft sensing and time-series forecasting was designed using bi-directional LSTM encoders-decoders with temporal attention. The results on a dataset collected over 6 month period showed good accuracy for daily weight estimations, with a Gaussian profile of daily error histogram. This enabled cumulative estimation of hive weight over multiple weeks. This shows that soft sensing using machine learning can significantly reduce the cost of beehive monitoring systems by eliminating the costly beehive weighing scales from the design.

This work also explored the use of self-attention encoders for system optimisation for the task of beehive weight estimation using a bigger dataset collected over a 12 month period. The results demonstrate that fit for purpose, robust weight estimations for beehives can be achieved for all seasonal conditions using soft sensors in real-world settings. The self-attention encoders of this model helped investigate the *use of machine learning for selection of sensors, and improving the design of monitoring systems.* The attention weights of encoders provide a detailed insight into important parts of the day for data collection. The varying levels of attention by encoder on different times of a day help identify the parts of the day where data carries more information. Collecting data more frequently during early hours of the morning and in the afternoon can provide more useful information about bee activity, while less frequent collection at other times can conserve power and data bandwidth. These attention weights can also be used to identify sensors and features which are less useful for a specific beehive task. A decrease in the design cost of system can be achieved by excluding the sensors with an inadequate contribution, such as accelerometers. This will also reduce the power consumption and the required data bandwidth, as total size of data generated for each iteration will also reduce. For a system designed for remote deployment, these are noteworthy improvements in the design, and made possible with the help of machine learning. The equally good results of beehive weight estimation using reduced features validate the usefulness of machine learning for optimizing the design of beehive monitoring systems.

6.1 Limitations and Future Work

Performance of deep learning models largely depends upon the quality and quantity of training data. For beehive monitoring, the pace of data collection is restricted by a number of factors, and foremost is the natural bee activity. This natural activity cannot be sped up to collect more data in less time. With all the seasonal variations and associated behavioural changes of the bee colony, at least one year of data collection is required to properly train and test deep learning models. This work used data which was collected for little over a year. Conditions change from year to year, and data collected over more time would have helped train and test models for different conditions. Eight monitoring systems were deployed in parallel for the data collection phase. Development of more systems required more financial resources, and deployment needed more time allocation and logistic support. Taking all this into consideration, it was decided to not increase the number of systems for data collection.

Initial plans for the deployment of monitoring system were in a significantly larger area. However, the travel restrictions because of COVID-19 forced a change to those plans. A significant portion of data used in this study was collected from hives at UWA campus. This somewhat restricted the diversity in collected data, but enabled hive access to troubleshoot the sensor systems when needed, and the timely inspections/assessments of hives. The hive assessment is a fairly time consuming process which was made possible with the help of beekeepers involved in this project. One of the aims to collect these assessments was to document varying health states of the hives, and use deep learning to classify these states. We approached the beekeepers best known for their good management of hives to allow us to use their hives for sensor deployments. Unfortunately for us, the extremely well managed hives did not experience any noticeable decline in the bee health state during sensor deployment. Thus, the data collected about hive states was heavily biased towards healthy hives. Most of these hives were used for commercial purposes, and asking the commercial beekeepers to deliberately force some of their bee colonies to become weak was not an option. Weak colonies are susceptible to bee diseases, which can spread to other colonies and put the whole business at risk.

For work in the near future, our preference would be to deploy multiple units of BeeDAS in different states of Australia to collect a much bigger and diverse beehive sensor dataset. Given the huge size, seasonal variations, weather conditions and the floral diversity within Australia, this dataset will be a very good test of the capabilities of deep learning models that have been developed for beehive weight estimation. Based on the results, the deep learning models can be modified to generate even better estimations. A single model capable of generating accurate estimates for beehive weight, regardless of the location of hive will be an ideal outcome. An alternate approach would be to train different models on beehive data collected from different geographic locations/conditions. The models trained so far have used data from beehives which did not experience certain weather conditions, such as snowfall or sub zero temperatures. Collecting data from hives experiencing such conditions and using it for training and testing the models will be very interesting. Also, many different species of honeybees are used in commercial beekeeping across the world. This study collected data from hives of European Honeybees, which are the most commonly used species globally. However, data from hives of other honeybee species such as Asian Honeybees and African Honeybees will also be used to train and test weight estimation models.

Beehive health monitoring is a multi-disciplinary area with huge potential for further research and innovation. But progress in this domain requires knowledge of electronics, communication, data science, honeybees and familiarity with standard working practices of beekeepers. We have made headway towards acquiring the required skills, knowledge of relevant areas, the correct tools for advancing this research. Further research will aim to use machine learning to detect bee diseases. Based on the findings of this research, improvements have be made in the design of the sensor system. If funding is available to carry on this work, deployment of these improved systems will be carried out in carefully managed sick and healthy hives to collect a balanced beehive assessment dataset, and multiple beekeepers throughout Australia will be engaged to collect the beehive health assessments in a systematic manner. The sensor data and health assessments will be used to train deep learning models. These models once deployed will not only inform the beekeeper about the current health status of bees, but also predict the future vulnerability of the beehives towards specific diseases. This will change the role of beehive health monitoring systems from a tool which minimises the loss of beekeepers to one which prevents it.

The design of beehive monitoring systems will see rapid improvements in the future. Integrated Circuits (ICs) with multiple sensors, microprocessor and data transmitter on a single chip will result in tiny and cost effective solutions. These ICs will not only be integral part of each hive, but will be present in each and every frame of a hive to assist with a wide range of problems. The applications of the future systems will not just be limited to beehive health monitoring, but also for monitoring the type and quality of nectar and pollen, extent of pollination, honey extraction process, and traceability of honey from the source to the end consumer.

In a time when plans are in place to colonize Mars and to develop self-sustaining human settlements on the Red Planet, beehive monitoring systems have another crucial role to play. Progress cannot be made on the front of cultivating land on Mars for agriculture without suitable pollinators. Honeybees being the biggest pollinator on Earth, are also a top contender for this job on another planet. Selective breeding and genetic modifications will be required to create a bee specie that can survive on Mars. Such a landmark can only be achieved with thorough testing, and that is where beehive monitoring systems have a very important role to play. Monitoring systems can help evaluate the performance of honeybees in Mars like artificial environments created on Earth. They can be used to find the very extremes of conditions where different modified species of bees can survive. Monitoring systems will also evolve during this process, and a time will come when both the bees and the monitoring systems will be able to serve the basic physiological needs of humans on another planet.

Bibliography

Research Articles/Books

[1] Abraham H Maslow. "The instinctoid nature of basic needs." In: *Journal of personality* (1954). DOI: 10.1111/j.1467-6494.1954.tb01136.x.

[2] Marcelo A. Aizen et al. "How much does agriculture depend on pollinators? Lessons from long-term trends in crop production". In: *Annals of Botany* 103.9 (2009), pp. 1579–1588. ISSN: 10958290. DOI: 10.1093/aob/mcp076.

[3] Lucas A Garibaldi et al. "Stability of pollination services decreases with isolation from natural areas despite honey bee visits". In: *Ecology letters* 14.10 (2011), pp. 1062–1072. DOI: 10.1111/j.1461-0248.2011.01669.x.

[7] Thomas D Seeley. *Honeybee ecology: a study of adaptation in social life*. Vol. 36. Princeton University Press, 2014.

[8] Mark L Winston. *The biology of the honey bee*. Harvard University Press, 1991.

[9] Robin Moritz and Edward E Southwick. *Bees as superorganisms: an evolutionary reality*. Springer Science & Business Media, 2012.

[10] Simon G Potts et al. "Global pollinator declines: trends, impacts and drivers". In: *Trends in ecology & evolution* 25.6 (2010), pp. 345–353. DOI: 10.1016/j.tree.2010.01.007.

[11] Dave Goulson et al. "Bee declines driven by combined stress from parasites, pesticides, and lack of flowers". In: *Science* 347.6229 (2015), p. 1255957. DOI: 10.1126/science.1255957.

[12] Stephen Russell, Andrew B Barron, and David Harris. "Dynamic modelling of honey bee (Apis mellifera) colony growth and failure". In: *Ecological Modelling* 265 (2013), pp. 158–169. DOI: 10.1016/j.ecolmodel.2013.06.005.

[13] Dennis VanEngelsdorp et al. "Colony collapse disorder: a descriptive study". In: *PloS one* 4.8 (2009), e6481. DOI: 10.1371/journal.pone.0006481.

[22] Jonathan de Carvalho Silva et al. "LoRaWAN—A low power WAN protocol for Internet of Things: A review and opportunities". In: *2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*. IEEE. 2017, pp. 1–6.

[24] Aleksejs Zacepins et al. "Challenges in the development of Precision Beekeeping". In: *Biosystems Engineering* 130 (2015), pp. 60–71. DOI: 10.1016/j.biosystemseng.2014.12.001.

[25] GA Burdock. "Review of the biological properties and toxicity of bee propolis (propolis)". In: *Food and Chemical toxicology* 36.4 (1998), pp. 347–363. DOI: 10.1016/s0278-6915(97)00145-2.

[26] Fiona Edwards Murphy et al. "Development of an heterogeneous wireless sensor network for instrumentation and analysis of beehives". In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. IEEE. 2015, pp. 346–351. DOI: 10.1109/i2mtc.2015.7151292.

[27] WG Meikle and Niels Holst. "Application of continuous monitoring of honeybee colonies". In: *Apidologie* 46.1 (2015), pp. 10–22. DOI: 10.1007/s13592-014-0298-x.

[28] Zeaiter Zeaiter. "The Effect of Thermoregulatory Stress on Honey Bee Hives". PhD thesis. University of Sydney, 2019.

[29] Robert M Burrill and Alfred Dietz. "The response of honey bees to variations in solar radiation and temperature". In: *Apidologie* 12.4 (1981), pp. 319–328. DOI: 10.1051/apido:19810402.

[30] Hossam F Abou-Shaara, Ahmad A Al-Ghamdi, and Abdelsalam A Mohamed. "Tolerance of two honey bee races to various temperature and relative humidity gradients". In: *Environmental and experimental Biology* 10.4 (2012), pp. 133–138.

[32] William G Meikle et al. "Within-day variation in continuous hive weight data as a measure of honey bee colony activity". In: *Apidologie* 39.6 (2008), pp. 694–707. DOI: 10.1051/apido:2008055.

[33] James I Hambleton. *The Quantitative and Qualitative Effect of Weather Upon Colony Weight Changes*. Tech. rep. 1925. DOI: 10.1093/jee/18.3.447a.

[34] Darren W Fitzgerald et al. "Design and development of a smart weighing scale for beehive monitoring". In: *2015 26th Irish Signals and Systems Conference (ISSC)*. IEEE. 2015, pp. 1–6. DOI: 10.1109/issc.2015.7163763.

[35] Chiu Chen et al. "An imaging system for monitoring the in-and-out activity of honey bees". In: *Computers and electronics in agriculture* 89 (2012), pp. 100–109. DOI: 10.1016/j.compag.2012.08.006.

[37] Donald Howard, Olga Duran, and Gordon Hunter. "A low-cost multi-modal sensor network for the monitoring of honeybee colonies/hives". In: *Intelligent Environments, IOS Press* (2018).

[38] Keith M Doull. "The effects of different humidities on the hatching of the eggs of honeybees". In: *Apidologie* 7.1 (1976), pp. 61–66. DOI: 10.1051/apido:19760104.

[39] Ahmad A Al-Ghamdi, Hossam F Abou-Shaara, and Abdelsalam A Mohamed. "Hatching rates and some characteristics of Yemeni and Carniolan honey bee eggs". In: *Journal of Entomology and Zoology Studies* 2.1 (2014), pp. 6–10.

[40] HF Abou-Shaara et al. "A review of impacts of temperature and relative humidity on various activities of honey bees". In: *Insectes Sociaux* 64.4 (2017), pp. 455–463. DOI: 10.1007/s00040-017-0573-8.

[41] B Kraus and HHW Velthuis. "High humidity in the honey bee (Apis mellifera L.) brood nest limits reproduction of the parasitic mite Varroa jacobsoni Oud." In: *Naturwissenschaften* 84.5 (1997), pp. 217–218. DOI: 10.1007/s001140050382.

[42] Hannelie Human, Sue W Nicolson, and Vincent Dietemann. "Do honeybees, Apis mellifera scutellata, regulate humidity in their nest?" In: *Naturwissenschaften* 93.8 (2006), pp. 397–401. DOI: 10.1007/s00114-006-0117-y.

[43] Axel Michelsen, Wolfgang H Kirchner, and Martin Lindauer. "Sound and vibrational signals in the dance language of the honeybee, Apis mellifera". In: *Behavioral ecology and sociobiology* 18.3 (1986), pp. 207–212. DOI: 10.1007/bf00290824.

[44] Sara Ferrari et al. "Monitoring of swarming sounds in bee hives for early detection of the swarming period". In: *Computers and electronics in agriculture* 64.1 (2008), pp. 72–77. DOI: 10.1016/j.compag.2008.05.010.

[45] Nicolás Pérez et al. "Continuous monitoring of beehives' sound for environmental pollution control". In: *Ecological engineering* 90 (2016), pp. 326–330. DOI: 10.1016/j.ecoleng.2016.01.082.

[46] Alessandro Terenzi, Stefania Cecchi, and Susanna Spinsante. "On the importance of the sound emitted by honey bee hives". In: *Veterinary Sciences* 7.4 (2020), p. 168. DOI: 10.3390/vetsci7040168.

[47] Stenford Ruvinga et al. "Use of LSTM Networks to Identify "Queenlessness" in Honeybee Hives from Audio Signals". In: *2021 17th International Conference on Intelligent Environments (IE)*. IEEE. 2021, pp. 1–4. DOI: 10.1109/ie51775.2021.9486575.

[48] Amro Qandour et al. "Remote beehive monitoring using acoustic signals". In: *Edith Cowan University, Institutional Repository* (2014).

[49] Jason Campbell, Lily Mummert, and Rahul Sukthankar. "Video monitoring of honey bee colonies at the hive entrance". In: *Visual observation & analysis of animal & insect behavior, ICPR* 8 (2008), pp. 1–4.

[50] SN Fry et al. "Tracking of flying insects using pan-tilt cameras". In: *Journal of Neuroscience Methods* 101.1 (2000), pp. 59–67. DOI: 10.1016/s0165-0270(00)00253-3.

[51] Thi Nha Ngo et al. "A real-time imaging system for multiple honey bee tracking and activity monitoring". In: *Computers and Electronics in Agriculture* (2019), p. 104841. DOI: 10.1016/j.compag.2019.05.050.

[53] Kohei Shimasaki et al. "HFR-Video-Based Honeybee Activity Sensing Using Pixel-Level Short-Time Fourier Transform". In: *2018 IEEE SENSORS*. IEEE. 2018, pp. 1–4. DOI: 10.1109/icsens.2018.8589744.

[54] Martin Bencsik et al. "Identification of the honey bee swarming process by analysing the time course of hive vibrations". In: *Computers and Electronics in Agriculture* 76.1 (2011), pp. 44–50. DOI: 10.1016/j.compag.2011.01.004.

[55] Christoph Grüter and Walter M Farina. "The honeybee waggle dance: can we follow the steps?" In: *Trends in Ecology & Evolution* 24.5 (2009), pp. 242–247. DOI: 10.1016/j.tree.2008.12.007.

[56] James C Nieh et al. "The communication of food location by a primitive stingless bee". In: *Zoology* 102.1999 (2000), pp. 238–246.

[57] De Sandeman, J Tautz, and M Lindauer. "Transmission of vibration across honeycombs and its detection by bee leg receptors". In: *Journal of Experimental Biology* 199.12 (1996), pp. 2585–2594. DOI: 10.1242/jeb.199.12.2585.

[58] JAMES C Nieh and JÜRGEN Tautz. "Behaviour-locked signal analysis reveals weak 200–300 Hz comb vibrations during the honeybee waggle dance". In: *Journal of Experimental Biology* 203.10 (2000), pp. 1573–1579. DOI: 10.1242/jeb.203.10.1573.

[59] Fiona Edwards-Murphy et al. "b+ WSN: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring". In: *Computers and Electronics in Agriculture* 124 (2016), pp. 211–219. DOI: 10.1016/j.compag.2016.04.008.

[60] Andrzej Szczurek et al. "Semiconductor gas sensor as a detector of Varroa destructor infestation of honey bee colonies–Statistical evaluation". In: *Computers and Electronics in Agriculture* 162 (2019), pp. 405–411. DOI: 10.1016/j.compag.2019.04.033.

[61] Elke Genersch. "American Foulbrood in honeybees and its causative agent, Paenibacillus larvae". In: *Journal of invertebrate pathology* 103 (2010), S10–S19. DOI: 10.1016/j.jip.2009.06.015.

[62] Rangarajan Sudarsan et al. "Flow currents and ventilation in Langstroth beehives due to brood thermoregulation efforts of honeybees". In: *Journal of theoretical biology* 295 (2012), pp. 168–193. DOI: 10.1016/j.jtbi.2011.11.007.

[63] Edward E Southwick and Robin FA Moritz. "Social control of air ventilation in colonies of honey bees, Apis mellifera". In: *Journal of insect physiology* 33.9 (1987), pp. 623–626. DOI: 10.1016/0022-1910(87)90130-2.

[64] Thomas D Seeley. "Atmospheric carbon dioxide regulation in honey-bee (Apis mellifera) colonies". In: *Journal of insect physiology* 20.11 (1974), pp. 2301–2305. DOI: 10.1016/0022-1910(74)90052-3.

[65] Karel Van Nerum and Herman Buelens. "Hypoxia-controlled winter metabolism in honeybees (Apis mellifera)". In: *Comparative Biochemistry and Physiology Part A: Physiology* 117.4 (1997), pp. 445–455. DOI: 10.1016/s0300-9629(96)00082-5.

[66] EE Southwick, DW Roubik, JM Williams, et al. "Comparative energy balance in groups of Africanized and European honey bees: ecological implications." In: *Comparative Biochemistry and Physiology. A, Comparative Physiology* 97.1 (1990), pp. 1–7. DOI: 10.1016/0300-9629(90)90713-3.

[67] Hayward G Spangler. "Photoelectrical counting of outgoing and incoming honey bees". In: *Journal of Economic Entomology* 62.5 (1969), pp. 1183–1184. DOI: 10.1093/jee/62.5.1183.

[68] MH Struye et al. "Microprocessor-controlled monitoring of honeybee flight activity at the hive entrance". In: *Apidologie* 25.4 (1994), pp. 384–395. DOI: 10.1051/apido:19940405.

[69] Jerry Bromenshenk et al. "Honey bee monitoring system for monitoring bee colonies in a hive". In: (June 2005). US Patent 6,910,941.

[70] Joseph A Shaw et al. "Long-wave infrared imaging for non-invasive beehive population assessment". In: *Optics express* 19.1 (2011), pp. 399–408. DOI: 10.1364/oe.19.000399.

[71] Fiona Edwards Murphy et al. "Big brother for bees (3B)—Energy neutral platform for remote monitoring of beehive imagery and sound". In: *2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*. IEEE. 2015, pp. 106–111. DOI: 10.1109/iwasi.2015.7184943.

[73] SC Jay. "Spatial management of honey bees on crops". In: *Annual Review of Entomology* 31.1 (1986), pp. 49–65. DOI: 10.1080/0005772X.1986.11098882.

[74] Jonathan Paul Ruppert et al. "Portable RF ID tag and barcode reader". In: (June 1997). US Patent 5,640,002.

[75] P Nunes-Silva et al. "Applications of RFID technology on the study of bees". In: *Insectes sociaux* 66.1 (2019), pp. 15–24.

[76] Axel Decourtye et al. "Honeybee tracking with microchips: a new methodology to measure the effects of pesticides". In: *Ecotoxicology* 20.2 (2011), pp. 429–437. DOI: 10.1007/s10646-011-0594-4.

[77] Xujiang He et al. "Assessment of flight activity and homing ability in Asian and European honey bee species, Apis cerana and Apis mellifera, measured with radio frequency tags". In: *Apidologie* 44.1 (2013), pp. 38–51. DOI: 10.1007/s13592-012-0156-7.

[78] Dominic Clarke and Daniel Robert. "Predictive modelling of honey bee foraging activity using local weather conditions". In: *Apidologie* 49.3 (2018), pp. 386–396. DOI: 10.1007/s13592-018-0565-3.

[79] James L Gould, JL Kirschvink, and KS Deffeyes. "Bees have magnetic remanence". In: *Science* 201.4360 (1978), pp. 1026–1028. DOI: 10.1126/science.201.4360.1026.

[80] Chin-Yuan Hsu et al. "Magnetoreception system in honeybees (Apis mellifera)". In: *PloS one* 2.4 (2007), e395. DOI: 10.1371/journal.pone.0000395.

[81] Veronika Lambinet et al. "Linking magnetite in the abdomen of honey bees to a magnetoreceptive function". In: *Proceedings of the Royal Society B: Biological Sciences* 284.1851 (2017), p. 20162873. DOI: 10.1098/rspb.2016.2873.

[82] Rahman Tashakkori, Abdelbaset S Hamza, and Michael B Crawford. "Beemon: An IoT-based beehive monitoring system". In: *Computers and Electronics in Agriculture* 190 (2021), p. 106427. DOI: 10.1016/j.compag.2021.106427.

[83] Sergio Gil-Lebrero et al. "Honey bee colonies remote monitoring system". In: *Sensors* 17.1 (2017), p. 55. DOI: 10.3390/s17010055.

[84] Noor Hafizah Khairul Anuar et al. "Embedded Wireless Stingless Beehive Monitoring And Data Management System". In: *2021 IEEE International Conference in Power Engineering Application (ICPEA)*. IEEE. 2021, pp. 149–154. DOI: 10.1109/icpea51500.2021.9417758.

[85] Nirmal Anand et al. "Swarm Detection and Beehive Monitoring System using Auditory and Microclimatic Analysis". In: *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*. IEEE. 2018, pp. 1–4. DOI: 10.1109/cimca.2018.8739710.

[86] Andreas König. “IndusBee 4.0–Integrated Intelligent Sensory Systems for Advanced Bee Hive Instrumentation and Hive Keepers’ Assistance Systems”. In: *Sensors & Transducers* 237.9/10 (2019), pp. 109–121.

[87] Vladimir Kulyukin. “Audio, Image, Video, and Weather Datasets for Continuous Electronic Beehive Monitoring”. In: *Applied Sciences* 11.10 (2021), p. 4632. DOI: 10.3390/app11104632.

[88] Douglas S Kridi, Carlos Giovanni N de Carvalho, and Danielo G Gomes. “Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection”. In: *Computers and Electronics in Agriculture* 127 (2016), pp. 221–235. DOI: 10.1016/j.compag.2016.05.013.

[89] Nitin Naik. “Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP”. In: *2017 IEEE international systems engineering symposium (ISSE)*. IEEE. 2017, pp. 1–7. DOI: 10.1109/syseng.2017.8088251.

[91] Marcin Smereka and Ignacy Dulęba. “Circular object detection using a modified Hough transform.” In: *International Journal of Applied Mathematics & Computer Science* 18.1 (2008). DOI: 10.2478/v10006-008-0008-9.

[92] Beth Logan. “Mel frequency cepstral coefficients for music modeling”. In: *International Symposium on Music Information Retrieval*. Citeseer. 2000.

[93] Padraig Davidson. et al. “Anomaly Detection in Beehives using Deep Recurrent Autoencoders”. In: *Proceedings of the 9th International Conference on Sensor Networks - SENSORNETS,* INSTICC. SciTePress, 2020, pp. 142–149. ISBN: 978-989-758-403-9. DOI: 10.5220/0009161201420149.

[96] Hugo Hadjur, Doreid Ammar, and Laurent Lefèvre. “Toward an intelligent and efficient beehive: A survey of precision beekeeping systems and services”. In: *Computers and Electronics in Agriculture* 192 (2022), p. 106604. DOI: 10.1016/j.compag.2021.106604.

[101] Jun Zhang et al. “Evolutionary computation meets machine learning: A survey”. In: *IEEE Computational Intelligence Magazine* 6.4 (2011), pp. 68–75. DOI: 10.1109/mci.2011.942584.

[102] Laizhong Cui et al. “A survey on application of machine learning for Internet of Things”. In: *International Journal of Machine Learning and Cybernetics* 9.8 (2018), pp. 1399–1417. DOI: 10.1007/s13042-018-0834-5.

[103] Adnan Qayyum et al. “Secure and robust machine learning for healthcare: A survey”. In: *IEEE Reviews in Biomedical Engineering* 14 (2020), pp. 156–180. DOI: 10.1109/rbme.2020.3013489.

[104] Guillermo Barrenetxea et al. “The hitchhiker’s guide to successful wireless sensor network deployments”. In: *Proceedings of the 6th ACM conference on Embedded network sensor systems*. 2008, pp. 43–56. DOI: 10.1145/1460412.1460418.

[105] R Mark Goodwin, Joanne H Perry, and Anton Ten Houten. “The effect of drifting honey bees on the spread of American foulbrood infections”. In: *Journal of Apicultural Research* 33.4 (1994), pp. 209–212. DOI: 10.1080/00218839.1994.11100873.

[106] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32. DOI: 10.1023/A:1010933404324.

[107] J. Ross Quinlan. “Induction of decision trees”. In: *Machine learning* 1.1 (1986), pp. 81–106. DOI: 10.1007/bf00116251.

[108] Michael Simone-Finstrom et al. “Propolis counteracts some threats to honey bee health”. In: *Insects* 8.2 (2017), p. 46. DOI: 10.3390/insects8020046.

[109] James Simpson. “Nest climate regulation in honey bee colonies”. In: *Science* 133.3461 (1961), pp. 1327–1333. DOI: 10.1126/science.133.3461.1327.

[112] Sumit Biswas. “Intelligent beehive status monitoring in noisy environment”. MA thesis. 2020.

[113] Ashley M Collier-Oxandale et al. “Understanding the ability of low-cost MOx sensors to quantify ambient VOCs”. In: *Atmospheric Measurement Techniques* 12.3 (2019), pp. 1441–1460. DOI: 10.5194/amt-12-1441-2019.

[114] Juergen Tautz. “Honeybee waggle dance: recruitment success depends on the dance floor”. In: *The Journal of experimental biology* 199.6 (1996), pp. 1375–1381. DOI: 10.1242/jeb.199.6.1375.

[115] Warwick E Kerr and Nilton J Hebling. “Influence of the weight of worker bees on division of labor”. In: *Evolution* (1964), pp. 267–270. DOI: 10.2307/2406400.

[116] Minxing Si et al. “Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods”. In: *Atmospheric Measurement Techniques* 13.4 (2020), pp. 1693–1707.

[117] Zachary Ballard et al. “Machine learning and computation-enabled intelligent sensor design”. In: *Nature Machine Intelligence* 3.7 (2021), pp. 556–565.

[118] Ayman Elnashar, Mohamed A El-saidny, and Mahmoud Sherif. “Coverage and capacity planning of 4G networks”. In: (2014). DOI: 10.1002/9781118703434.ch6.

[120] Rashmi Sharan Sinha, Yiqiao Wei, and Seung-Hoon Hwang. "A survey on LPWA technology: LoRa and NB-IoT". In: *Ict Express* 3.1 (2017), pp. 14–21. DOI: 10.1016/j.icte.2017.03.004.

[122] Juha Petajajarvi et al. "On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology". In: *2015 14th international conference on its telecommunications (itst)*. IEEE. 2015, pp. 55–59. DOI: 10.1109/itst.2015.7377400.

[127] Rapeepat Ratasuk et al. "NB-IoT system for M2M communication". In: *2016 IEEE wireless communications and networking conference*. IEEE. 2016, pp. 1–5. DOI: 10.1109/wcnc.2016.7564708.

[130] Niels Holst and William G Meikle. "Breakfast canyon discovered in honeybee hive weight curves". In: *Insects* 9.4 (2018), p. 176. DOI: 10.3390/insects9040176.

[131] AR McLellan. "Honeybee colony weight as an index of honey production and nectar flow: a critical evaluation". In: *Journal of Applied Ecology* (1977), pp. 401–408. DOI: 10.2307/2402553.

[132] A Zacepins et al. "Solution for automated bee colony weight monitoring". In: *Agron Res* 15.2 (2017), pp. 585–593.

[133] A Terenzi et al. "Real-time System Implementation for Bee Hives Weight Measurement". In: *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. IEEE. 2019, pp. 231–236. DOI: 10.1109/metroagrifor.2019.8909252.

[134] Piotr Bratek and Piotr Dziurdzia. "Energy-Efficient Wireless Weight Sensor for Remote Beehive Monitoring". In: *Sensors* 21.18 (2021), p. 6032. DOI: 10.3390/s21186032.

[135] Qingqiang Sun and Zhiqiang Ge. "A Survey on Deep Learning for Data-Driven Soft Sensors". In: *IEEE Transactions on Industrial Informatics* 17.9 (2021), pp. 5853–5866. DOI: 10.1109/TII.2021.3053128.

[136] Yiğit Mengüç et al. "Wearable soft sensing suit for human gait measurement". In: *The International Journal of Robotics Research* 33.14 (2014), pp. 1748–1764. DOI: 10.1177/0278364914543793.

[137] Shengdong Du et al. "Multivariate time series forecasting via attention-based encoder–decoder framework". In: *Neurocomputing* 388 (2020), pp. 269–279. DOI: 10.1016/j.neucom.2019.12.118.

[138] Ari Yair Barrera-Animas et al. "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting". In: *Machine Learning with Applications* 7 (2022), p. 100204. DOI: 10.1016/j.mlwa.2021.100204.

[139] Kevin E Trenberth. “What are the seasons?” In: *Bulletin of the American Meteorological Society* 64.11 (1983), pp. 1276–1282.

[140] Hui Liu and Zhihao Long. “An improved deep learning model for predicting stock market price time series”. In: *Digital Signal Processing* 102 (2020), p. 102741. DOI: 10.1016/j.dsp.2020.102741.

[141] Ashish Vaswani et al. “Attention is All You Need”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6000–6010. ISBN: 9781510860964.

[142] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. “Vision transformers for dense prediction”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 12179–12188. DOI: 10.1109/iccv48922.2021.01196.

[143] Phasit Charoenkwan et al. “BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides”. In: *Bioinformatics* 37.17 (2021), pp. 2556–2562. DOI: 10.1093/bioinformatics/btab133.

[144] Kyu J Han, Ramon Prieto, and Tao Ma. “State-of-the-art speech recognition using multi-stream self-attention with dilated 1d convolutions”. In: *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE. 2019, pp. 54–61. DOI: 10.1109/asru46091.2019.9003730.

[145] Burton Noble Gates. *The temperature of the bee colony*. 96. US Dept. of Agriculture, 1914. DOI: 10.5962/bhl.title.64474.

[146] Piotr Medrzycki et al. “Influence of brood rearing temperature on honey bee development and susceptibility to poisoning by pesticides”. In: *Journal of Apicultural Research* 49.1 (2010), pp. 52–59. DOI: 10.3896/ibra.1.49.1.07.

[147] Keith S Delaplane, Jozef van der Steen, and Ernesto Guzman-Novoa. “Standard methods for estimating strength parameters of Apis mellifera colonies”. In: *Journal of Apicultural Research* 52.1 (2013), pp. 1–12. DOI: 10.3896/ibra/1.52.1.03.

[148] WG Meikle et al. “Using balances linked to dataloggers to monitor honey bee colonies”. In: *Journal of apicultural research* 45.1 (2006), pp. 39–41. DOI: 10.3896/ibra.1.45.1.09.

[149] TI Szabo and LP Lefkovitch. “Effects of honey removal and supering on honey bee colony gain”. In: *American bee journal (USA)* (1991).

[150] PA Sopade et al. “Application of the Williams–Landel–Ferry model to the viscosity–temperature relationship of Australian honeys”. In: *Journal of food engineering* 56.1 (2003), pp. 67–75. DOI: 10.1016/s0260-8774(02)00149-8.

[151] S Yanniotis, S Skaltsi, and S Karaburnioti. “Effect of moisture content on the viscosity of honey at different temperatures”. In: *Journal of Food Engineering* 72.4 (2006), pp. 372–377. DOI: 10.1016/j.jfoodeng.2004.12.017.

[152] RA Gleiter, H Horn, and H-D Isengard. “Influence of type and state of crystallisation on the water activity of honey”. In: *Food chemistry* 96.3 (2006), pp. 441–445. DOI: 10.1016/j.foodchem.2005.03.051.

[153] EF Woods. “Electronic prediction of swarming in bees”. In: *Nature* 184.4690 (1959), p. 842. DOI: 10.1038/184842a0.

[154] David G Dietlein. “A method for remote monitoring of activity of honeybee colonies by sound analysis”. In: *Journal of Apicultural Research* 24.3 (1985), pp. 176–183. DOI: 10.1080/00218839.1985.11100668.

[155] David Atauri Mezquida, J Llorente Martıénez, et al. “Platform for bee-hives monitoring based on sound analysis. A perpetual warehouse for swarm’s daily activity”. In: (2009). DOI: 10.5424/sjar/2009074-1109.

[156] EK Eskov and VA Toboev. “Analysis of statistically homogeneous fragments of acoustic noises generated by insect colonies”. In: *Biophysics* 55.1 (2010), pp. 92–103. DOI: 10.1134/s0006350910010161.

[157] Robert Denniston Milner and George S Demuth. *Heat production of honeybees in winter*. 988. US Department of Agriculture, 1921. DOI: 10.5962/bhl.title.108785.

[158] Stanley S Schneider, Judy A Stamps, and Norman E Gary. “The vibration dance of the honey bee. II. The effects of foraging success on daily patterns of vibration activity”. In: *Animal behaviour* 34.2 (1986), pp. 386–391. DOI: 10.1016/s0003-3472(86)80106-3.

[159] Christof W Schneider et al. “RFID tracking of sublethal effects of two neonicotinoid insecticides on the foraging behavior of Apis mellifera”. In: *PloS one* 7.1 (2012), e30023. DOI: 10.1371/journal.pone.0030023.

[160] Helge Schlüns et al. “Multiple nuptial flights, sperm transfer and the evolution of extreme polyandry in honeybee queens”. In: *Animal Behaviour* 70.1 (2005), pp. 125–131. DOI: 10.1016/j.anbehav.2004.11.005.

[161] Ina Heidinger et al. “Observation of the mating behavior of honey bee (Apis mellifera L.) queens using radio-frequency identification (RFID): factors influencing the duration and frequency of nuptial flights”. In: *Insects* 5.3 (2014), pp. 513–527. DOI: 10.3390/insects5030513.

URL References

[4] N. Mortensen Ashley, R. Schmehl Daniel, and Ellis Jamie. *Information sheet on Apis Mellifera*. https://entnemdept.ufl.edu/creatures/misc/BEES/euro_honey_bee.htm. Accessed: 2022-06-28. 2017.

[5] *Australain Honey Bee Industry Council, Pollination Factsheet*. https://honeybee.org.au/pdf/PollinationAwareFactSheet.pdf. Accessed: 2022-06-28.

[6] Agrifutures Australia. *Size and scope of the Australian honeybee and pollination industry – a snapshot*. https://agrifutures.com.au/wp-content/uploads/2021/02/20-136.pdf. Accessed: 2022-06-28.

[14] BeeAware. *Inspecting Hives*. http://beeaware.org.au/biosecurity/keeping-honey-bees-healthy/inspecting-hives/. Accessed: 2022-06-28.

[15] BeeInformed. *Relocating Hives*. https://beeinformed.org/2013/11/12/the-basics-of-moving-hives/. Accessed: 2022-06-28.

[16] *Beehive Monitoring 4G GSM Gateway*. https://www.beehivemonitoring.com/en/home/83-4g-gsm-gateway-classic.html. Accessed: 2022-06-28.

[17] *Libelium Beehive Monitoring*. http://www.libelium.com/temperature-humidity-and-gases-monitoring-in-beehives/. Accessed: 2022-06-28.

[18] *HiveMind Remote beehive monitoring*. https://hivemind.co.nz/. Accessed: 2022-06-28.

[19] Pollenity. *BeeBot Monitoring System*. https://pollenity.com/product/beebot/. Accessed: 2022-06-28.

[20] *BroodMinder Monitoring System*. https://broodminder.com/products/broodminder-citizen-science-kit. Accessed: 2022-06-28.

[21] *Cellular coverage comparison in Australia*. https://www.whistleout.com.au/MobilePhones/Guides/who-has-the-best-mobile-coverage. Accessed: 2022-06-28.

[23] *Narrow Band Internet of Things*. https://www.gsma.com/iot/narrow-band-internet-of-things-nb-iot/. Accessed: 2022-06-28.

[31] *Langstroth Beehive*. http://www.talkingwithbees.com/beekeeping/beehives/beehive-diagram. Accessed: 2022-06-28.

[36] *How to inspect a hive*. https://www.thespruce.com/inspect-your-honey-bee-hive-3016536. Accessed: 2022-06-28.

[52] *EyesOnHive video monitoring system*. https://www.keltronixinc.com/. Accessed: 2022-06-28.

[72] *Thermal Image of Beehives*. https://www.foxhoundbeecompany.com/beekeepingblog. Accessed: 2022-06-28.

[90] *ApisProtect Monitoring System*. https://www.apisprotect.com/home. Accessed: 2022-06-28.

[94] Hivemind. *Hivemind Weighing Scale User Manual*. https://hivemind.nz/static/manuals/hive-weight-scales.pdf. Accessed: 2022-06-28.

[95] *BeeMate Monitoring System*. https://beemate.buzz/. Accessed: 2022-06-28.

[97] Beeaware. *Biosecurity Code of Practice*. https://beeaware.org.au/wp-content/uploads/2017/09/Australian-Honey-Bee-Industry-Biosecurity-Code-of-Practice.pdf. Accessed: 2022-06-28. 2017.

[98] Agric Australia. *Prohibited Organism List*. https://www.agric.wa.gov.au/organisms?search_string=Paenibacillus+larvae. Accessed: 2022-06-28. 2022.

[99] Legislation. *Biosecurity and Agriculture Management 2007*. https://www.legislation.wa.gov.au/legislation/statutes.nsf/law_a146629.html. Accessed: 2022-06-28.

[100] Omar Anwar. *BeeDAS Github repository*. https://github.com/omaranwar85/BeeHave. Accessed: 2022-06-28. 2022.

[110] Australia. *Bureau of Meteorology (BOM)*. http://www.bom.gov.au/catalogue/data-feeds.shtml. Accessed: 2022-06-28. 2021.

[111] *Definition of Relative Humidity*. https://www.nfsa.gov.au/preservation/preservation-glossary/relative-humidity-rh. Accessed: 2022-06-28.

[119] *Starlink webpage, High-speed, low-latency broadband internet*. https://www.starlink.com/. Accessed: 2022-06-28.

[121] *LoRa-Alliance*. https://lora-alliance.org/. Accessed: 2022-06-28.

[123] *LoRa fair usage policy*. https://www.thethingsnetwork.org/forum/t/fair-use-policy-explained/1300. Accessed: 2022-06-28.

[124] *LoRa Module*. https://docs.rs-online.com/ff06/0900766b81544027.pdf. Accessed: 2022-06-28.

[125] *Airtime calculator for LoRaWAN*. https://www.thethingsnetwork.org/airtime-calculator. Accessed: 2022-06-28.

[126] *Telstra NB-IoT coverage*. https://www.telstra.com.au/business-enterprise/about-enterprise/our-network/iot-coverage-map. Accessed: 2022-06-28.

[128] *Mundaring State Forest on google maps*. https://goo.gl/maps/pwAzFaYhUSRnDBYE
Accessed: 2021-09-28.

[129] Equipment Arnia. *Arnia*. https://www.arnia.co/product-page/arnia-pro-scale. Accessed: 2022-06-28.

[162] *IoT Products Waspmote*. https://www.libelium.com/iot-products/waspmote/. Accessed: 2022-06-28.

[163] *Baxtran Weight tare indicator*. https://www.baxtran.com/en/indicators/br80.html. Accessed: 2022-06-28.

Appendix

This appendix contains some additional details about the bee hives, bee activity, and the setups that have been used to measure the relevant parameters. All the contents of this appendix link to chapter 2 of this thesis, which provides a background on beehive sensors and monitoring systems.

A Hive Thermoregulation

Initial studies of temperature in a beehive were conducted in early 20^{th} century [145]. The suitable temperature range in which honeybees can breed successfully is significantly smaller than their survivable temperature range. It is only during the summer when the queen bee lays eggs in the brood cells in large numbers. These eggs become capped brood after eight days. Capped brood is vulnerable to changes in hive temperature because it requires temperatures between 34°C and 36°C to mature properly into adult bees. A variation of more than a few degrees can kill or debilitate the brood [146]. If the temperature is below 34°C, hive bees form clusters over capped brood, consume honey and metabolise rapidly to generate heat with their flight muscles [28]. If the temperature inside hive exceeds 36°C, the honeybees use the fanning activity to bring the hive temperature down to an appropriate level [109] and foragers bring extra water in the hive to be passed around and reduce the temperature.

For adequate thermoregulation, studies suggest that healthy colonies require one hive bee for every two capped brood cells. A hive which can regulate its temperature effectively indicates a healthy and proportionate bee population [147]. On the other hand, weak hives with poor thermoregulation show decreased cognitive ability, weaker flight muscles and vulnerability to parasites [28]. This shows the importance of measuring the temperature of hive in assessing not only the current health, but also the future health of the bees. Back to temperature in Section 2.1.1.

B Hive Weight

The weight of a healthy hive goes through a cycle of increase and decrease on a typical foraging day. The total variation in the hive weight depends on the size of colony, number of forager bees, amount of floral resources available in the area and relative humidity. In suitable environmental conditions, foragers leave the hive early morning in large numbers in search of pollen and nectar. This results in a sharp decrease in the weight of the hive. The number of foragers in a hive depends on many factors, but usually account for one third of total bees in the hive. Thus, the early morning decrease in the weight can be used to roughly estimate the total number of bees in the hive. When the early foragers start to return to the hive with pollen and nectar, the weight of the hive increases [148, 149]. If forager bees continue to find good flowering resources, the net weight of hive continues to increase [32].

There are many factors which contribute towards the loss of hive weight. Nectar stored in the hives loses its water content because of evaporation, which results in reduction of the weight of the honey in the hive. The bees and brood inside the hive consume food, convert it into energy, further reducing the weight of the hive. Factors such as weak bee colonies, winter season, non-availability of resources to forage upon also result in the hive losing weight. A swarming event, when a portion of bees in the colony leave the hive with the (new) queen to find a new location for the hive, results in a sharp decrease in the weight of the hive [32]. This shows that the patterns of weight change carry very important information about the health and activity of the bees and status of the hive. A measuring resolution of a few grams can provide information about minor changes in the weight of the hive. From a honey production point of view, beekeepers use the weight related information to check the honey flow and to decide if they should move the hives to a new location with more flowering resources or keep the hives at their current location. Back to weight in Section 2.1.2.

C Hive Humidity

Humidity levels rise in the hive when bees bring in fresh nectar. Nectar is passed from bee to bee to reduce its high water content to around 17 - 21% [150, 151]. This nectar is then stored in a cell and because of evaporation, the water content is further reduced before it is capped by bees. Good quality honey has water contents of around 8% as honey with excessive water cannot be preserved for longer periods of time [152]. High humidity in a hive slows down the process of evaporation, and increases the risk of mould during winter. Honeybees use fanning [62] at the hive entrance to ventilate the air in order to keep the humidity and temperature at a reasonable level. Healthy colonies are good at regulating the micro-climate inside the hive, so humidity levels are also expected to be an indicator of hive health. However bees are able to regulate temperature much more effectively compared to humidity. The humidity levels at different locations inside the hive also vary, depending upon how effectively the bees are managing the ventilation of the hive. Honeybees maintain different humidity levels in the brood area and around honey frames [42]. The former requires higher humidity levels whereas the later needs lower levels of relative humidity to facilitate evaporation from nectar. Back to humidity in Section 2.1.4.

D Hive Acoustics

One of the earliest sound analyser for bees was 'Apidictor' [153], whose manufacturers claimed that it could detect swarming, acceptance and failure of the queen, and health of the bee colonies. However these claims were not verified by the wider community. The electronics involved in audio processing is much more sophisticated than that of temperature and weight. With the advancement of electronics, the work in this area gained momentum. Another study on audio signal acquisition and processing of beehive sounds was published by *Dietlein et al.* [154]. This study focused on the amplitude and duration of audio signals during different seasons and on identifying the major frequencies which are generated by honeybees. Bioacoustics has gained a lot of popularity recently and a focus of it has been on honeybees. Major work in this area [44, 155, 156] is primarily focused on using the acoustics for early detection of honeybee swarming. Swarming is a loss for the beekeeper, so prediction and early detection of swarming is a key priority and acoustics has shown a lot of promise in this area. Back to acoustics in Section 2.1.5.

E Hive Gas Composition

An initial study to understand the composition of gases inside the hive was conducted in 1921 [157]. In this study, the air from the brood chamber was extracted and passed through external detectors to find the O_2 and CO_2 levels inside the hive. Similar experiments were conducted by other scientists [64, 63, 65] in later years and all of them used the same method of extracting air from the beehive and then using external detectors to study the composition. This methodology was driven by two major factors. First, the size of equipment was too large to fit inside a hive, so gases were extracted after regular intervals and tested using the equipment outside. Second, there is a requirement of smooth flow of gases through/across the gas sensors for them to function properly. Such flow is not available inside the hive as air circulation in the hive is largely controlled by bees. Back to gas content in Section 2.1.8.

F Honeybee Waggle Dance

Vibrations generated by the waggle dance, the 15 Hz abdomen waggle and the 250 Hz thorax vibration are known for their relation with foraging activity [158, 57]. Using this dance, the forager bees inform the other foragers about relative position of flowering resources in the area with respect to position of the sun. This saves the forager bees a lot of time and effort as typical forager can travel up to 3 km from hive in search of pollen and nectar. This communication allows the foragers to make the most of the closest available foraging resources or the most favourite ones. Visits to a specific type of plant by foragers in a colony results in the collection of mono-floral honey. Other than these waggle dance vibrations, bees generate vibrations up to 1 kHz, but it is not known if this entire band carries useful information. Back to accelerometer in Section 2.1.7.

G Limitations of Thermal Imaging

Using low resolution thermal imaging, it is neither possible to estimate the number of bees, nor can anything be inferred about the presence of brood, stocks of food or presence of any disease in the colony. The amount of thermal radiation coming out from each hive is also dependent upon the thickness of hive wall, material used such as wood or plastic, reflectiveness of paint which varies with type and age, and the amount of moisture in the wood in case of a wooden hive. With so many variables adding to the noise of thermal images, it is very difficult to use thermal imaging as a standard way to estimate the strength of bee colony. Two colonies with similar strengths can have two very different thermal radiation patterns outside the hive depending upon the construction, composition and condition of the hive itself. Back to thermal imaging in Section 2.1.10.

H Use of RFID in Honeybee Research

A major use of this technology has been in identifying the impact of pesticides and insecticides on honeybees. Researchers use the RFID tags on bees so that the bees can be easily identified, and then feed with controlled dosages of sublethal substances, subsequently monitoring their flight activity over a period of time. These experiments allow researchers to determine the acute effects of sublethal doses of insecticides/pesticides [159]. Mating behaviour of honeybee queen has also been studied in detail using RFID. A virgin queen goes for mating flights, also known as Nuptial Flights only during a span of several days in her early life, mostly a few weeks after hatching from the cell [160]. During these flights, the queen mates with several drones and stores the sperms in her body. Once these nuptial flights are complete, the queen stays in her hive for rest of her life and lays eggs for upto 5 years. RFID has also been used in studies related to these complex nuptial flights [161]. Back to radio frequency identification in Section 2.1.12.

I Details of Beehive Monitoring Systems in Literature

This appendix will provide some details about the short listed beehive monitoring systems from literature.

I.1 Parameters/Sensors Used

Howard et al. [37] have performed the most comprehensive study of relevant parameters among the monitoring systems shortlisted for this review. They have used two very high relevance parameters i.e. temperature and weight, along with three high relevance parameters including humidity, acoustics and external imaging. Authors also measure gas contents in their experimental setup which is a parameter with medium relevance. *Tashakkori et al.* [82] propose a complete end-to-end beehive monitoring system 'Beemon'. This IoT based system comprises seven components including hardware, software, research tool for processing video and image data, research tool for audio data, web tool for streaming the live video from hive, web tool for data analytics/visualisation, and dashboards to be used as interface for beekeepers. However, the authors only mention the weighing scales in their work and do not provide any details of scale design, construction, usage or results.

Murphy et al. in both of their studies [26, 59] have used the same setup of sensors inside the hive. These studies however aimed to achieve different goals. The objective of *Murphy et al.* [26] was the testing of a wireless sensor network for beehive health monitoring with a focus on communication and power consumption of different types of sensors. Whereas the focus of *Murphy et al.* [59] was to use multiple deployments of their designed nodes to collect beehive data and induce a decision tree for hive classification. The authors in another study *Murphy et al.* [71] use a different set of sensors to investigate a low power and self-sustainable design with the aid of energy harvesting via solar panels.

Gil-Lebrero et al. [83] made use of three parameters in their proposed design. The focus of this design was on overall usability and scalability. The major difference between this study and others part of this review is the deployment of three temperature and humidity sensors in three different locations/frames of each hive.

In their work *Ferrari et al.* [44] and *Anand et al.* [85] used differernt set of parameters to monitor the swarming activity of bees, to allow for early detection of swarming. Swarming is a phenomenon when a part of the bee colony leaves the hive, which is a loss for the beekeepers. *Anuar et al.* [84] designed an embedded electronic beehive monitoring system along with its android application. However, the parameters were only monitored for the honey chamber in this study.

Konig et al. [86] in his work 'IndusBee 4.0' designed a monitoring system where the weight for each chamber of the hive was individually monitored. Acoustic monitoring of the hive was performed using a MEMS microphone. Further, a Volatile Organic Compounds (VOC) gas sensor was also added to the design to explore the in-hive detection of infestation and illness.

Kulyukin et al. [87] developed a custom monitoring system (BeePi), capable of recording audio, image and video data from the hives. The weather data was also recorded, and analyzed for correlation with bee activity. Later chapters of this thesis will also evaluate the importance of using weather data in beehive monitoring. This is one of the few works where the collected data has been made public.

The remaining two studies employed a single parameter each for beehive monitoring. *Kridi et al.* [88] used only the temperature sensor with focus on a light weight system. *Chen et al.* [35] used infrared imaging in their monitoring system. This work employed image processing to monitor the activity of forager bees, and is quite different from other works included in this review.

I.2 Design and Methodology

Howard et al. [37] used standard beehives in an apiary for their experimentation, with the aim of causing minimum interference to the normal bee activity. They used a total of 4 hives for their experiments, all placed within 30 meters of each other. The authors used commercially available Arnia monitoring systems, alongside their own designed system to read temperature, humidity and collect additional data of acoustics, video, gas concentrations, internal and external light levels, and the weight of hive. Arnia system does not facilitate the collection of video or audio data, and the Arnia weighing scales with a price tag of 699 GBP are quite expensive. However the authors used a single Arnia weight scale for one of the hives in the experiment. By reading some of the parameters twice, from Arnia and from their own design, they were able to validate the functioning and accuracy of their own system. Apart from acoustics, each of the deployed sensors read a value every minute which included 12 seconds of video during favourable light conditions. The microphone however was used to record 12 minutes of continuous audio every 20 minutes. The authors were able to collect the sensor data from hives effectively for several months. However the processing of audio and video data was not part of this study.

Tashakkori et al. [82] developed their own system with an aim of achieving cost-effectiveness, efficiency and reliability. The system used a Raspberry Pi 4 with a Raspberry camera interface and an SD card that was utilized for local storage. The camera was used to capture still images at a resolution of 3280 × 2464 pixels, and the video at 1080p. A USB microphone was used to record the audio data. An AM2302 sensor capable of recording humidity and temperature was connected to Raspberry Pi using the GPIOs. The Raspberry Pi ran an application that recorded the audio and video data, temporarily stored the data until it was uploaded to the server for long term storage and analysis. The plug and play system allowed for easy replacement of components. A client could easily relay commands from a terminal connected via a Secure Shell (SSH) connection to the server allowing control over the system.

Murphy et al. [26] targeted the deployment of a WSN for bee health monitoring, with the aim of being non invasive, easy for beekeeper, robust in hive, low energy and remote deployable. The Authors used Libelium Waspmote [162] as their basic platform for development of nodes. Two types of nodes were developed, one for generic hive measurements and other for gas detection. Both nodes were deployed on the top lid of a single hive, and analysed the air/gases flowing out of hive. Sensors were not exposed to bees directly, and a mesh excluder was used between the sensors and the bee compartment. Time stamped samples were collected every 4 hours, making a total of 6 samples per day. A third node was used as base station to collect the data from hive nodes. Authors have presented data collected during first two weeks of July. The collected data shows high temperature and humidity fluctuations in early July, despite this being a brood rearing period in the region of experimentation. Authors did not provide any explanation for these fluctuations.

Edwards et al. [59] used the same setup discussed in *Murphy et al.* [26] with some modifications in gas data collection and changed the sampling time to 8 hours (3 samples per 24 hours) instead of 4 hours to make battery last longer. Gas sensors have relatively high power consumption and reducing the sampling period of this node provided significant gains. Data was collected for 14 days with this setup and results again showed a lot of variations in the readings of temperature and humidity, which is not a norm for a healthy hive. The measurements from the lid of a hive were a non-linear average of parameters from inside, and outside the hive such as temperature, humidity. However, data gathered using this setup was used by the authors to classify ten different states of the hive. These ten states were divided into two groups, one which required a response from beekeeper and one which did not. Parameters used by authors in the decision tree classification were humidity, temperature, carbon dioxide and rain. This required the classification system to have access to meteorological data of hive location for accurate classification. Results from the experiments on single hive showed an accuracy of around 95%. However for a generic solution, a much bigger dataset from multiple hives is required.

*Murphy et al.*in their study [71] used a different set of parameters. Waspmote [162] was used to interface an accelerometer, microphone, infrared and thermal imaging cameras. Infrared Light Emitting Diodes (IR LEDs) were used with IR camera to illuminate the hive. As the IR spectrum is not visible to bees, authors claim that this setup is not invasive. However, the question still remains about the heat emitted by these IR LEDs, which can impact the thermoregulation of the hive if used in large numbers. A long term presence of IR camera inside the hive with bees may result in bees covering it with propolis, which will block the vision. The IR camera in this setup was only tested in an empty hive, and the practicality of proposed design can only be evaluated when tested with bees. Thermal images of 80×60 pixel resolution were also acquired using an external setup, where the authors did not provide any information on how the system will process this thermal data. The images, both thermal and IR, were acquired 5 times per day and the waspmote was used to turn off the Pis after each capture. Similarly, waspmote was used to boot the Pis before the next image capture, which reduced they duty cycle and maximized the energy performance. For acoustics, low sampling rates of 100 Hz were used by authors to minimize data rate and processing. Only the peak values of acquired audio samples of bee noise were used to alert the beekeeper if they crossed a certain threshold. This low sampling rate minimised the power consumption, but made it impossible to extract any frequency information from bee buzz (100 - 300 Hz). In addition to acoustics and imaging, authors used accelerometer to alert the beekeeper using a text message in case a significant movement was detected by the accelerometer.

Gil-Lebrero et al. [83] also used a Libelium Waspmote [162] with SHT15 sensors capable of measuring temperature and humidity, and a weighing scale. For each hive, authors deployed three SHT15 sensors in different parts of hive and the sensors were protected from propolisation by enclosing them in perforated queen expedition cages. Data collected from each sensor node was directly communicated to an embedded industrial computer, with XBee nodes connected to it using a star topology. They used a single 150 kg load cell for the weighing scale, with resolution set at 100g along with BR80 weight tare indicator [163]. Data from sensors and weighing scale was recorded every 5 minutes for 32 consecutive days, and then again during the honey flow. During the honey flow, the weight scale did not pick up minor daily variations because of used resolution, but the gradual increase in the weight during this period was evident. Results also showed very stable temperature and humidity in the brood chamber with minor fluctuations for sensors placed away from brood area. Environmental temperature and humidity variations had very little impact on readings from the brood area, which signifies the importance of proper placement of sensors in the hive. This designed monitoring system is being used in 20 bee colonies, which is significantly more than the number of hives used by other systems in this study.

Ferrari et al. [44] used temperature, humidity and acoustics to monitor the swarming activity of bees. This is one of the most significant works carried out for early detection of swarming. Three hives were monitored for 270 hours continuously, without removing the newborn queens from the hives. This allowed for multiple events of swarming and collection of event related sensor data. Audio data was sampled at 2 kHz with 16 bit resolution. The temperature and relative humidity sensors were placed in between the hive frames using HOBO dataloggers which recorded values every 2.5 minutes, and microphones were placed above the honey frames, continuously recording the sounds from inside the hive. All sensors were covered by a special net to protect against the propolisation from the bees. All the signal processing was done using Matlab including filtering to get rid of the noise. The authors labeled the audio data manually with the aid of spectrogram and used the timestamps to map audio data with humidity and temperature readings. A total of 9 swarming events occurred during this study and they were detected from sound as well as from the visual analysis of the observers. All the swarming events were recorded during day times and in the hottest hours. The total duration of a swarm, from the start of bees excitement till they leave the hive, ranged between 13 and 56 minutes. Bee noise generated during the swarming had a much higher Power Spectral Density (PS) when compared with the standard bee buzz, specially in the 500Hz to 1000Hz band. A drop in internal hive temperature was also observed just before the swarming events. Sound and temperature data were found to have more correlation with the swarming activity compared to humidity.

Anuar et al. [84] designed a system based on NodeMCU, and used a built-in low-power WiFi module for communication. A single hive was monitored for 36 hours and a total of 12,200 data points were collected with a 7 second interval. To estimate the number of forager bees going in and out of the hive, two pairs of infrared transmitters and receivers were used in a funnel to provide a single channel for forager movement. The funnel was fabricated using a 3-D printer and was installed at the main hive entrance/exit. However, the authors only tested the system at an unconventional hive, where brood chamber was cylindrical in shape.

Anand et al. [85] designed a beehive monitoring system to function in the practical environment and to be user friendly and compact. A major focus of this work was on using the audio from inside the hive to detect pre-swarming of bees. The Analog to Digital Converter (ADC) of Arduino UNO was used to sample the audio data at 38 kHz. A 256 point Fast Hartley Transform (FHT) was implemented on the sensor node to extract audio frequency components. However, the authors only shared the frequency analysis after processing in Matlab. The designed system was tested on a single hive.

Konig et al. [86] discussed the importance of beehive monitoring system and its evolution over time, with focus on Varroa mite infestation and hive state classification using acoustic data. Primary goal of this work was to achieve a simple, affordable and reliable hive keepers assistance system for varroa monitoring in the beehives with a sufficient screening coverage for event-driven treatment decisions. The designed system comprised of Raspberry Pi Zero, two temperature and moisture sensors (DHT22/11) at the top and the bottom of hive chamber, HX711 board for hive weight scale reading and an I2S MEMS microphone. The audio processing was performed offline on a remote computer, requiring all audio data to be captured and stored. Also, these tests were performed on a very limited data set acquired from a single mini bee colony. Further, a Volatile Organic Compounds (VOC) gas sensor was added to the design to explore the in-hive detection of infestation and illness. The very basic version of this system without scales and gas sensor cost less than US$ 30 for components.

Kulyukin et al. [87] shared the dataset they collected using a custom monitoring system. This data was collected from 2014 onwards, and authors have made an excellent effort in explaining the experimental setup, conditions, design limitations, power issues and data labelling. The designed system is based on raspberry Pi, and different models of Pi were used over different periods of time. The other setup however remained the same and consisted of a pi T-Cobbler, a breadboard, a waterproof temperature sensor, a pi camera, a ChronoDot clock, and a Neewer 3.5 mm mini lapel microphone.

Kridi et al. [88] developed a prototype for data collection based on the Arduino platform. The temperature sensors used were LM-35 which are very cost effective, and each sensor was calibrated by authors for consistent readings. They deployed two hives and recorded the temperatures inside for 40 consecutive days.

Chen et al. [35] in their study used an infrared CCD camera and infrared light source to monitor the passageway of bees going in and out of the hive. The setup forced the bees to use the confined passageways under camera observation. In order to help individually identify each forager bee, circular character tags were attached to the dorsum of the bees, where each tag had a black positioning dot to identify the orientation of the tag. The experiment was designed to record the timings of entry and exit for each bee, which provided information about the foraging habits of bees. The experiment lasted 15 days and by the end, the number of tagged honeybees going in and out had gone to zero. This validates the lifespan of foraging honeybee to be about two weeks. Authors observed that honeybees started the trips outside the hive around 05:00 AM and the activity reached its peak around 10:00 AM. After that the foraging activity gradually decreased, reaching a minimum level at around 07:00 PM. These findings however are very specific to the region and depend upon the sunrise and sunset times as well as the season. Almost half of the honeybees in this study spent less than 3 minutes outside the hive per trip, which indicates that flowering resources were located near by. Authors also noticed that for multiple trips in a single day, honeybees spent a minimum total of around 1 hour and a maximum total of 7 hours outside the hive.

I.3 Data Storage/Communication

Howard et al. [37] designed their system with a WiFi hotspot using a raspberry Pi for communication. All other Pi based sensor nodes were connected to this hotspot. However authors did not use this communication link for sensor data communication for unspecified reasons. Instead, they used onboard 32 GB SD cards to store small amounts of sensor data and an external 60 GB hard disk drive for large audio and video data. This data had to be manually copied from external HDD every 3-4 days.

Tashakkori et al. [82] used a class 10 SD card on raspberry Pi to temporarily store the high bandwidth audio and video data. The sensor data from low bandwidth sensors was transmitted using Message Queuing Telemetry Transport (MQTT) [89] protocol to the dashboard using a wired connection. The system used File Transfer Protocol (FTP) to transmit the high bandwidth audio and video data over same wired connection. This allowed a near real-time collection of data from the hives.

Murphy et al. in their works [26, 59] also used SD card for the storage of sensor data. They utilized low power XBee Series 2 radio for communication between two local WSN nodes and the third base station node which acted as 'Xbee to 3G radio bridge' for long distance communication. The base station node combined the data from all sensor nodes into a single 'csv' file and uploaded it to server via File Transfer Protocol (FTP).

In another study, *Murphy et al.* [71] designed the system with a GSM/GPRS module for data communication. This module had an advantage of ultra-low power operation and could be used for FTP upload/download. GSM/GPRS networking was selected in this study to allow the remote deployments of beehives using cellular network. Data collected using sensors was stored on SD card, and the proposed mechanism of data transfer using FTP was not tested by authors. Since the imaging equipment collected only 5 images per day, this generated small amounts of data, suitable for storage on SD cards.

Gil-Lebrero et al. [83] used Libelium's 'XBee USB-Serial gateway' module as the coordinator of the network, which used IEEE 802.15.4 standard to communicate with the deployed nodes. The SCADA system was used to broadcast a data request, and then collected data from all the nodes in a synchronized manner. The SCADA on gateway server used SQL database to store the sensor data. Local computer was connected to internet using ethernet connectivity and a periodic backup of the local database on each node was also carried out in the database server. This enabled the beekeepers to access data from anywhere.

Work of *Ferrari et al.* [44] was focused on detecting the swarming activity instead of coming up with a complete monitoring system. They used wired communication to collect the analog acoustic data and used a sound card to digitize and store it on a computer hard disk drive. Authors used in-hive data loggers for recording the humidity and temperature sensor data along with its time stamps and retrieved it after the experiment to correlate with audio recordings. *Chen et al.* [35] also used a simple wired connection to transfer CCD camera data to the computer where it was processed to extract the information about foraging activity.

The temperature monitoring system designed by *Kridi et al.* [88] sent the initial 4 hourly readings of temperature inside the hive to base station using a XBee module operating at 900 MHz. Authors used a high gain antenna to establish a wireless link between nodes in apiary and the base station at a distance of 210 meters. The specifications of XBee module claimed much bigger coverage area but authors found that not to be true. Besides sending the temperature readings to base, the designed system also used these readings to identify the closest pattern of acceptable microclimate. Once such pattern was identified, the monitoring system transmitted the ID of selected pattern and afterwards did not transmit any more hourly temperature readings as long as the hive temperature readings were within 1°C of selected pattern. If the hive temperature was found to be out of desirable range, the temperature was sent to the base station to indicate that the hive temperature has deviated from the selected pattern. The monitoring system then also attempted to fit a new pattern to the new temperature readings and if found, same process described above was repeated. If the monitoring system did not find any suitable pattern, it continued to send the temperatures to the base station, and after three or more consecutive failed attempts of finding a pattern, the system started generated an overheating alert as well. As only critical information was transmitted to the base station, this allowed for reduced communication from hive and made nodes more power efficient.

Anuar et al. [84] did not store the data locally on their system. All data were recorded and transferred to a Google Firebase real-time database using the onboard WiFi module. The authors also designed an Andriod application which fetches all the information from the database and represents it as graphs. *Anand et al.* [85] did not share any details about local storage capabilities of their system. However they analyzed the collected audio data from beehive using Matlab. It is not clear if the WiFi module of the system was used to transmit raw audio data to the cloud server or just the audio frequency components extracted through FHT. The cloud server used was BLYNK, which is an open source IOT cloud server. *Konig et al.* [86] also did not share any details about the communication or data storage capabilities of their system. However raspberry Pi Zero has support for both WiFi and an SD card.

Ferrari et al. [44] and *Chen et al.* [35] used a wired connection to log the data on a computer at the hive site. Their setups were not designed to test the communication or storage of hive data. *Kulyukin et al.* [87] did not design their system to handle data communication. The purpose of this design was to collect ample amounts of data for analysis. The total data collected by these systems over the years exceeds 1 TB (Tera Bytes), and USB storage devices were used with Raspberry Pis to store and collect this data.

I.4 Remote Deployability

Howard et al. [37] used SD cards and external Hard Disk Drives for data collection nodes based on raspberry Pi. This was not a user-friendly approach and was also invasive in nature. It required the beekeeper/researchers to open the hive every 2-3 days to collect data, which is significantly less than the recommended duration of two weeks. Also, the WiFi network between nodes was not used effectively to transmit the data. The backbone of connectivity for this study was a university network which cannot be used for remote deployments. Authors also used external power sources for their nodes. The hardware setup with frequent sampling proved so power hungry that use of batteries was found inadequate. This deployment strategy can work for data collection for research purposes for a limited period, but would fail to work for remote monitoring.

Tashakkori et al. [82] did not design their system for remote deployability. Since the system was entirely dependent on Ethernet (wired) connectivity for communication, the system was only deployed and tested on hives in urban areas where wired internet connections were available. Also, the systems based on raspberry Pi consume a lot of power. The collection of video data, and its transmission is also a power hungry process. The system was powered using a power line from a nearby building.

Murphy et al. [26] attempted to make adequate arrangements for remote deployability. Authors used wireless communication to relay the sensor data to the user, and included solar panels in their design for energy harvesting. The sensor nodes used 6.5V at 205mA solar panels with a 6600 mAh battery whereas the base station featured a sightly bigger solar panel of 7V at 500mA rating. The base station used XBee for inter node communication and 3G radio communication to upload node's data to server using FTP. The bigger solar panel of base station proved adequate for this operation, however the capacity of solar panel used for gas sensor nodes was not enough. This battery drained to around 20% within two weeks. Also, the base node and sensor nodes were deployed very close to each other, and the impact of increasing this distance on power usage by XBee radio was not investigated. In their next study, *Edwards et al.* [59] used the same setup but improved the deployability by reducing the number of samples of gas node from 6 to 3 per day. This provided a significant improvement in battery performance as the installed solar panel was able to replenish this smaller battery drain. This made their setup remotely deployable, subject to the availability of 3G coverage in the area.

Murphy et al. in another work [71] had the major objective of achieving system autonomy by being energy neutral. They implemented energy harvesting through solar panels and stored energy in a 1000 mAh battery, which would last several days even without sunshine. These nodes compared to their previous setup were much more energy efficient and the total energy requirement by each node in the worst case scenario could be met using the solar panel harvesting. However, this setup generated a lot of image data, and authors did not test the working of 3G network to transmit this data to server. SD cards were used to store this data, which required manual collection, not a desirable aspect for remote deployability.

Gil-Lebrero et al. [83] used batteries and external power supply for their nodes. The internode communication employed using XBee is a reasonable solution for remote deployment but the local server used in this setup was connected to internet using Ethernet connectivity. This system was not tested with wireless connectivity between local server and the cloud server which is essential for remote deployment.

Kridi et al. [88] developed a prototype which used SD card to record all the readings, and only the essential readings were communicated wirelessly to the base station. This significantly improved the battery life. The data from sensors was accessible using internet. These features add to the remote deployability of this system.

Anuar et al. [84] also based their system on two 1300 mAh lithium batteries and a charger module connected to AC power line. WiFi was used for communication of data for this system. The power and WiFi requirements limit the remote deployability options of this system. *Anand et al.* [85] did not provide any details about how the system was powered. And the use of WiFi significantly impacts remote deployability. [44, 35, 86] did not design their system for remote deployment.

I.5 Cost

Anand et al. [85], *Anuar et al.* [84], *Ferrari et al.* [44], and *Chen et al.* [35] did not provide any information about the cost of equipments used. The focus of their studies was on the development of algorithms and techniques rather than a complete monitoring system. *Kulyukin et al.* [87] designed their system to be low cost by using off the self sensors and components, but did not provide an estimated cost for their system.

I.6 Data Processing

Konig et al. [86] used the vibration and sounds inside the hive for classification of different hive states, which are:

- Okay/Calm or Normal State
- Agitated/Disturbed
- Knocking/Pecking at hive
- Scratching at hive
- Swarm mood
- Missing Queen
- Looting (robbing of honey by bees from other hives)

Other studies included in this review [37, 83, 44, 84, 85] did not focus on decision making aspects of the monitoring systems. However authors expressed their desire to use the experimental data for development of machine learning algorithms in the future.