

Toward an intelligent and efficient beehive: A survey of precision beekeeping systems and services.

Hugo Hadjur^{*1,2}, Doreid Ammar^{1,2}, and Laurent Lefèvre²

¹aivancity School for Technology, Business & Society, 57 Avenue du Président Wilson, 94230 Cachan, France

²Inria, Univ Lyon, EnsL, UCBL, CNRS, LIP, 46 Allée d'Italie, 69007 Lyon, France
{hadjur, ammar}@aivancity.ai, laurent.lefevre@ens-lyon.fr

December 2021

Abstract

Since the end of the 20th century, bees are suffering from increasing stress factors, leading domesticated colonies to die or at least be less productive. Precision beekeeping (PB) is an emerging field of agriculture that aims at protecting bees, supporting beekeepers, and optimizing apiary production thanks to digital infrastructures. The digitalization of apiculture first involves systems from the field of the Internet of Things (IoT), with the development of sensors to collect and transfer bee-related data. Then, data analysis comes into play, providing models that connect the data with the biological states of beehives, sometimes thanks to artificial intelligence (AI).

In this paper, we describe the recent advances in precision beekeeping as systems and as services. Different types of sensors, networks, and power sources in PB are covered. The collection and use of data are described, and the performances of PB services are assessed. We also estimate the sustainability of the proposed solutions, taking into account their scalability, efficiency, and economic cost, because beekeepers need deployable research results.

Keywords: Honey bee, Smart Agriculture, Internet of Things, Artificial Intelligence, Precision Beekeeping

*Corresponding author

Contents

1	Introduction	3
1.1	Bees & Beekeeping	3
1.2	Threats to bees	3
1.3	Precision Beekeeping	4
2	System Overview of a Connected Beehive	5
2.1	Architecture of a Connected Beehive	5
2.2	Types of sensors	6
2.3	Network Architecture: Transmission of Data	8
2.4	Power Supply	9
3	Costs of a Connected Beehive	9
3.1	Price	10
3.2	Energy	10
3.3	Open-Source access	11
3.4	Research Projects involved in Precision Beekeeping	11
4	Data Collection as a Service in a Smart Beehive	12
4.1	Challenges	12
4.2	Apiary level	12
4.3	Colony level	13
4.3.1	Inside the beehive	13
4.3.2	Outside the beehive	14
4.3.3	Presence detector	15
4.4	Bibilometric analysis: data types and their associated services	15
5	Data Analytics as a Service in a Smart Beehive	17
5.1	Audio	17
5.1.1	Biological knowledge about bees' sound signals	17
5.1.2	First step of sound data analysis: filtering	17
5.1.3	Second step: features extraction	18
5.1.4	Training models, accuracy & Conclusions	18
5.2	Image	20
5.2.1	Image pre-processing	20
5.2.2	Models & Accuracy	22
5.3	Other types of data	23
5.4	Discussion	23
6	Conclusion & discussion	24
6.1	The state of precision beekeeping research	24
6.2	The needs for precision beekeeping research	24
6.3	The future of precision beekeeping research	25

1 Introduction

1.1 Bees & Beekeeping

Honey bees are one of the main vectors of pollination: they transfer pollen between flowers to fertilize them. For some plant species, cross-pollination (pollination thanks to external organisms) is necessary for their reproduction. This type of pollination applies to both spontaneous and cultivated species. Agricultural production that depends on animal pollination quadrupled in the past 50 years ([FAO, 2016]).

The earliest human ancestors had started to hunt bee nests for their honey. However, bee has been domesticated since 2400 BC ([Crane, 1999]). Bees are extremely intelligent animals: they can learn tasks involving colors and shapes differentiation ([Giurfa et al., 2001]). Bees also cooperate within a colony. For instance, foragers perform a waggle dance to their nestmates to indicate the direction, distance, and quality of potential food resources ([von Frisch et al., 1967]). Beyond pollination, bees produce honey, royal jelly, propolis, pollen, and beeswax. Each of these substances has nutritional or health benefits. In temperate regions, bees hibernate during winter, start production in spring and continue throughout the summer, stop production and revert into the hibernation state in autumn. In hot and tropical regions, pollen and nectar sources can be available on a year-round basis, so bees do not hibernate.

Most of the cited articles in our survey focus on the *apis mellifera* species (western honey bee). This species is the most popular among beekeepers worldwide due to several selected advantages such as the ability to survive during dearth periods or "when honey flow is poor, the resistance to disease, the maximum amount of honey storage, the tendency to sting, and the ease of pacification by smoke" ([Weber, 2013]).

Honey bees are social insects. They live in colonies, with different kinds of members ([Crane, 1999]):

- The queen: there is usually one per healthy colony and it is the only female capable of reproduction. She is, therefore, the mother of all the bees of the colony. The queen is also the colony's barometer: if she is healthy, the colony is most often doing well. During their beehive control routines, beekeepers are looking for the queen to assess a beehive state.
- Female workers: they are the most numerous individuals of the colony and perform tasks like cleaning the honeycombs, producing royal jelly to feed larvae and the queen, building the honeycombs, storing pollen and nectar, protecting the beehive, and foraging (listed chronologically in the life of a worker).
- Males (drones): apart from mating with the queen, males honey bees do not perform productive work for the colony. They alternate between eating, resting, and mating during the high season (spring to fall).

As opposed to wild nests, beehives are referred to as the human-made adaptation of bees' habitat. Modern beehives consist of several honeycomb frames, where bees store brood, pollen, and honey. There are different kinds of models: some hang frames vertically, others horizontally. Despite the variety of models, they all have in common an orifice for the entrance, a central part containing the frames, and surrounding walls to protect bees from external threats and weather. Nowadays, the most popular type of beehive in beekeeping is the Langstroth hive ([Langstroth, 2004]), consisting of a box where frames are stored vertically. Smaller boxes with extra frames called honey supers can be added on top of the hive to boost honey production when a colony is expanding.

Nowadays, beekeeping is an activity that can go from a hobby to a full-time professional occupation. Even if professional beekeepers only represent small percentages among all beekeepers, their production is substantial: in France, in 2019, the top 4.1% of beekeepers (professionals owning more than 150 hives) produced 75% of the national volume of honey ([FranceAgriMer, 2020]). This subset also owns 57% of the total amount of beehives in France: 902 000 out of 1 584 000 ([ADA-France, 2019]).

1.2 Threats to bees

Colony Collapse Disorder (CCD) is an unusual phenomenon that appeared in the early 2000s ([vanEngelsdorp et al., 2009]) when honey bees suddenly started to vanish. Several factors are associated with bee mortality: pesticides, Varroa destructor mite, genetic strains, habitat degradation, the Asian hornet, and viruses. However, it is still difficult to prove causality and to predict future consequences. Beekeepers should refine their diagnosis of honey bee colonies' health and invest more time in prevention to reduce the mortality rate of bee colonies that is currently in rapid growth.

There are little to no affordable health prevention tools on the market. There are however a few promising treatments against the varroa mite which, even if not perfectly effective, do not seem to affect bees at short and

middle term. During winter, beekeepers can spray oxalic acid directly on frames. An alternative is formic acid during fall.

The main goal of parasite monitoring research is to develop tools that automatically diagnose parasite threats and prevent those threats as early as possible.

1.3 Precision Beekeeping

Precision Beekeeping (PB) is a branch of precision agriculture, first mentioned in [Zacepins et al., 2012] and described as “an apiary management strategy based on the monitoring of individual bee colonies to minimize resource consumption and maximize the productivity of bees”. PB tries to tackle challenges thanks to bee data gathered over time and throughout two main biological levels: apiary-level and colony-level.

- The apiary-level consists of several beehives consisting of one colony, each. The colonies share the same perimeter within a geographical location. It is also worth noting that the environmental conditions of the geographical location impact bees’ behavior.
- The colony-level focuses on a single beehive. More precisely, on the colony’s organization, bees’ life, and behaviors.

This paper uses the term **connected beehive** to designate a beehive that can collect bee-related data and possibly transfer data as part of a PB installation. A **smart beehive** is a connected beehive with some intelligence, for example, a beehive capable of diagnosing health issues.

Precision Beekeeping’s main objective is to allow beekeepers to lean on automated tools to decrease their workload while strengthening the production coming from hives. Apart from targeting each seasonal crop alone, another goal of PB is to protect bees over time. Technology is needed to help beekeepers save time: thanks to wireless installations, data about hives can be accessed directly on a smartphone in real-time. Therefore, traveling on-site is optimized.

The authors of [Lettmann and Chauzat, 2018] addressed a form to French beekeepers in 2018 which aimed at understanding the dynamics of this profession. Professional beekeepers represent 13% of the respondents, but this ratio increases to 44% when omitting those who do not use connected devices in their beehives. In the same direction, 52% of beekeepers who use digital technology have been performing beekeeping for more than ten years. Therefore, it is clear professional beekeepers fuel the demand for smart solutions.

As of today, professional beekeepers who choose to use technology rely on two main categories of metrics: on the one hand, the weight of the beehive, which is a good indicator of honey flow evolution and flowering. It facilitates optimizing the selection of flora and the timing of transhumance. On the other hand, internal and external temperature and humidity values can help track a colony’s health and development ([Cousin et al., 2019]).

Location of research papers Our PB survey cites a total of 86 publications, including 7 other surveys. Figure 1 shows the distribution of the 79 result-oriented cited publications. Our priority is to describe recent state-of-the-art techniques in a global context. Although centered around Europe, where historical biological bee research had been performed, PB research initiatives come from all over the world and seem to gain attention in developing countries, especially in the last few years. In particular, the authors of [Fiedler et al., 2020] are part of a 3-continent collaboration. They presented a PB system deployment in Ethiopia and Indonesia, involving researchers from the latter two countries and Germany and Latvia.

It is however hard to deduce any geographical tendency as many research papers are often the only PB-related contribution of the corresponding author.

Existing literature surveys In the past few years, a subset of PB surveys has been published in the literature. Each survey tackles PB with a different angle of analysis. The authors of [Zacepins and Karasha, 2013] describe the progress of the research in temperature monitoring of bee colonies, from simple thermometers and data collected manually to infrared imaging and the start of embedded solutions. The survey paper presented in [Meikle and Holst, 2015] reflects on the advancement of PB data analysis by listing comprehensively the types of data that a beehive can offer. The research described in [Zacepins et al., 2015] is also close to our survey. The authors of [Marchal et al., 2020] examine the range of data collected in an apiary and the expansion to smart technologies directly on flowers, bees, and nests. The author of [Bumanis, 2020] states the different trajectories that PB will follow in the future. One trajectory is to combine different sources of data to make them more sophisticated.

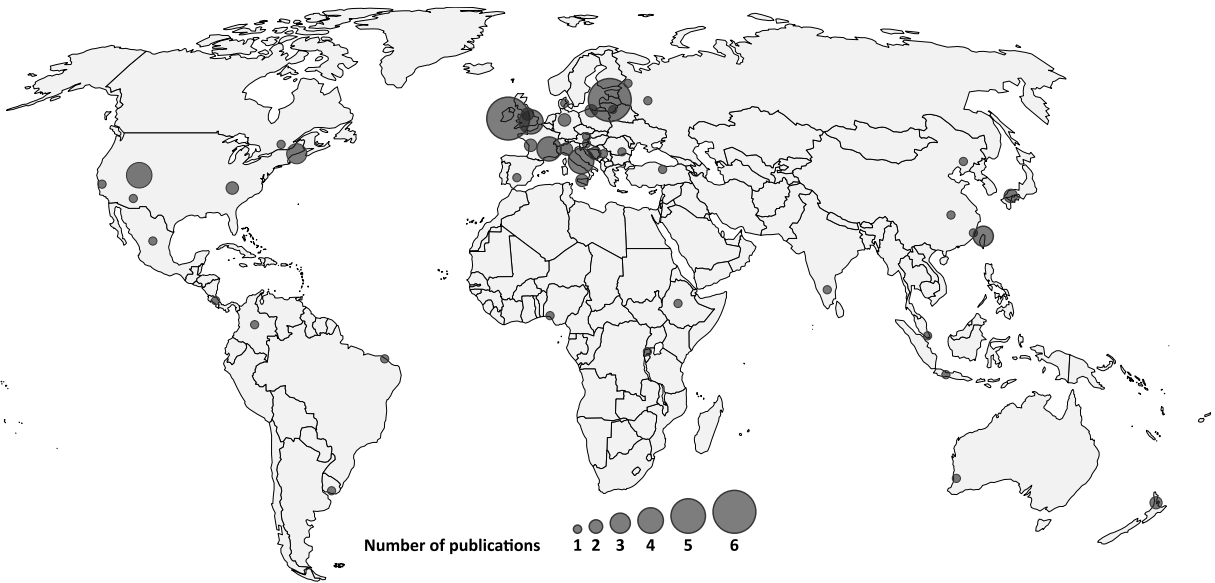


Figure 1. World distribution of PB research papers cited in this survey based on the location of the first author (surveys are omitted from this map)

In [Terenzi et al., 2020], the authors describe state of the art in PB sound analysis from early works to 2020. For each literature piece, they explain the methods and study their efficiency and drawbacks. The chronological order of the listing creates an interesting historical development.

Compared to other surveys, ours integrates the whole data pipeline from the design of systems and the collection of data to the underlying machine learning models. At every step, we weigh pros and cons of the solutions introduced in the literature. The main objective of this survey is to gain a global understanding of every layer of smart beehives. By criticizing different approaches, we also aim to raise questions for the future of PB, so that emerging research initiatives can gain a global view of the pivotal decisions to be made.

The paper is organized as follows: Section 2 focuses on the connected beehive system (mostly hardware). Section 3 highlights reproducible and sustainable researches in terms of costs. Section 4 and Section 5 go from how the data is collected to the AI models used to assist beekeepers. Specifically, Section 4 describes the types of data collected and the way it is gathered, and Section 5 goes through the different approaches used to develop PB learning models thanks to collected data. Finally, Section 6 concludes the paper and shares a number of lessons learned and directions for future work.

2 System Overview of a Connected Beehive

This section summarizes the different existing scientific approaches and outlines the architecture (i.e., the system). The services are then detailed in sections 4 and 5.

2.1 Architecture of a Connected Beehive

In the following sections, we use the term Internet of Things (IoT) as defined by [ITU, 2012] as "a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies".

In all PB IoT-related papers, a connected beehive consists of a standard beehive with sensors connected to a microprocessor, powered by a battery and usually connected to a network to send the collected data to a remote server. The design of a connected hive must ideally fulfill the needs of beekeepers, be technically feasible and be suitable for bees' life.

Every smart beehive has a common structure:

- The brain of it is its microcontroller. At regular intervals, it gathers and centralizes the data before transferring it to a distant server. There are different families of microcontroller boards, the main two being Arduino

and Raspberry Pi. Each type of board has strengths and weaknesses. For instance, Arduino-based boards consume very little power and are specialized, compared to Raspberry Pi models that consume more, but also run an entire Linux-based OS, bringing more possibilities like programming with different languages. Larger embedded GPU-based computing devices are used for systems requiring real-time results with complex processing (e.g., Jetson TX2 in [Ngo et al., 2021]).

- The organs of a smart beehive are its sensors. They are either directly plugged into the microcontroller or communicating the gathered data through wireless protocols. Section 2.2 lists the different types of sensors.
- Once the data is collected thanks to sensors, it can be transmitted to a remote server. The information is either formatted and displayed to beekeepers on an application that groups the data by hive ([Ammar et al., 2019]), or analyzed by researchers ([Giammarini et al., 2015]).
- Finally, an energy source is essential to power the system. The classical autonomous setup consists of a solar panel charging a battery. In cases when beehives are located near a power outlet, microcontrollers can directly be plugged into a reliable source of current.

In most PB system research papers, researchers place sensors at various locations around and inside the beehive. As for the microcontroller, they position it outside the beehive. For instance, inside the roof of the beehive ([Anuar et al., 2019]) or in a dedicated box ([Pérez et al., 2016]), sometimes 3-D printed ([Tashakkori et al., 2021]). Isolating the microcontroller has the advantage of being close to the bees while being protected from their defensive reactions. [Kulyukin and Reka, 2016] describe a PB architecture containing hardware placed in an empty super. There is no evidence of temperature imbalance in this case. An extension of the entrance of the hive as a container for the electronic system, serving as a take-off and landing platform for bees is proposed in [Chen et al., 2012], [Chen et al., 2015] and [Bjerger et al., 2019].

The design of a PB service needs to be well-thought and anticipate the barriers that come with deployment:

- The size of the beehive restrains the type of hardware. Training computationally heavy AI models on-site is an example of a challenging task, given this hypothesis.
- Gathering well-labeled data comes with a cost of time. Such autonomous systems must last a significant amount of time.
- Limiting the cost of the PB system is crucial in order to offer an affordable solution to beekeepers. For every PB system, there must be a trade-off between the cost of the installation for a single colony, the number of colonies equipped in an apiary and the quality of data that the system records.

2.2 Types of sensors

In the IoT field, the choice of appropriate sensors is essential to produce a solid base before data analysis. In PB, sensors can capture apiary-related or colony-related data. In the following paragraphs, we only focus on the latter since data that describe an apiary in its entirety can be collected without any connected beehive.

Weight The weight of a beehive indicates the activity level of a colony. During the foraging season, it reflects pollen and nectar collection as well as the number of bees inside their hive. During winter, there are fewer bees and fewer resources than during summer, honey stock, which is the food consumed by bees, can be estimated thanks to this metric. Abnormal weight variations can also be correlated to incoming swarming events ([Zacepins et al., 2015]).

For the majority of PB research, the measurement of the weight of a beehive is carried out in the following way: four load cells, which rely on mechanics and resistive theory, placed at each corner of the hive, integrated with amplifier modules that are linked to the microcontroller ([Anuar et al., 2019], for example). HaBEEtat ([Sakanovic and Kevric, 2020]) presents an original weight-measuring system, with a connected frame holder inserted between the roof and the base of a beehive, at the very top of the frames. This built-in system enables sensors inside the beehive, without being intrusive, to measure the weight of every single frame.

The system described in [Fitzgerald et al., 2015] sets up a single point impact cell placed under a supportive beehive platform. The results show a sensitivity of 10g which is ideal for beehive weight tracking.

Temperature A bee colony can regulate the temperature inside of its beehive. The internal hive temperature is an indicator of health for a colony. If the temperature is not regulated, the colony is in danger, especially during winter, when bees form a heating cluster to maintain living temperature conditions between 32°C and 37°C ([Markovic et al., 2016]). During summer, bees usually ventilate to allow air exchanges. An increase of the internal temperature can be correlated with swarming ([Zhu et al., 2019]).

To measure temperature, [Markovic et al., 2016] uses digital thermometers that have the advantage of communicating with a 1-Wire bus, thus enabling the use of several thermometers with a unique microprocessor. The research presented in [Catania and Vallone, 2019] describes thermometers placed inside the beehive, at the center, where the queen bee is located. In contrast, thermometers are placed in the upper part of the beehive close to the roof in [Meikle et al., 2016]. Thermometers mentioned in [Kviesis et al., 2020] are placed above the hive body and outside the beehive, under a cover to avoid sun exposure, to collect both internal and external temperatures. In [Gil-Lebrero et al., 2017], temperature and humidity sensors are installed in the middle of the brood area, on the honey and pollen area, and on honeycombs from non-brood frames. The authors of [Giammarini et al., 2015] choose to place sensors (temperature, humidity, microphone, accelerometer & carbon dioxide) on the inner side of the beehive base, rather than directly on the frames. This has the advantage of being non-intrusive toward bees. However, the quality of data might be lower than the on-frame alternative. In a very experimental setup, 36 probes per "space between frames" are placed in a 4-frame beehive to collect the evolution of a 3-D heat-map ([Zhu et al., 2019]).

Bees count The number of ingoing and outgoing bees is a crucial parameter correlated with the number of foraging bees, thus providing an indicator of the amount of pollen and nectar to be brought back inside the hive.

Optical, ultrasound and mechanical counting systems are tested in [Reyes et al., 2012] to measure the traffic and assess beehive activity thanks to a modified beehive entrance. In the electronic entrance shown in [Chen et al., 2015], infrared sensors are pointing at the holes that bees use to go in and out, in order to detect each passage.

In [Zhang et al., 2021], the temperature, humidity, and audio data are collected from different beehives, which all have a varying number of frames (from 1 to 23 for bee frames and 0 to 11 for brood frames). By doing so, the trained models have the potential to assess the proportion of bee frames and brood frames inside a beehive.

Sound Sounds emitted by honey bees reflect the state of a colony as a unit. For instance, the audio footprint can be linked with swarm preparation, queen identification, and pest infestation (Section 5.1). Knowing that bees emit sound ranging from 0 to a few thousand Hertz, the sound sensor's frequency range needs to include that window.

To collect audio samples, the choice of the microphone is essential. There are several types: electret microphones are used in [Qandour et al., 2014] and [Anand et al., 2018], MEMS microphones appear in [Cecchi et al., 2018], and the setups described in [Ferrari et al., 2008] and [Kulyukin et al., 2018] involve clip-on microphones. All existing audio installations are placed inside the hive.

The collection of sound samples is described in [Cecchi et al., 2018]. The main unit containing the microcontroller, sensors, and microphone seems to be placed in a protecting box on the backside of a hive. It is unclear whether interior sensors are linked through drilled holes or passing through the entrance. Also, propolisation is mentioned but it remains unknown whether authors met challenges with internal microphones. The system described in [Ferrari et al., 2008] shows that microphones (and other sensors) are protected from propolisation thanks to a particular net. Sound recording nodes are placed on top of the frames under the roof. The recording node from [Robles-Guerrero et al., 2017] is also protected, in this case, thanks to a metallic mesh. In [Ramsey et al., 2017] and [Ramsey et al., 2020], the two accelerometers which record vibrational data are directly placed on a frame. A cavity is made at the center of the frame to place the sensor, and some molten wax is then used to prevent contact between bee and metal components. Even after two seasons of use, it is reported that electronics are not affected by the colony.

In general, although most of the audio-based PB papers mention the challenges that internal microphones bring to bees' life-cycle, the design and position of the recording nodes are not always sufficient to let future research reproduce their setup. Beehive audio monitoring is not converging towards a common framework, but it also guarantees unbiased and exploratory research advancements.

Image Pictures or videos of bees are an indicator of the activity and health of a colony. Even if the place of a camera is restricted to the outside of a beehive (the inside is dense and dark), pointing at the entrance is the agreement that all computer vision PB researchers reach. Indeed, the most active outside location is the entrance, which regroups an important number of actions performed by bees (foraging, guarding and fanning).

In [Edwards-Murphy et al., 2015a] an infrared camera is used alongside infrared LEDs to monitor the inside of a beehive during times when the beehive is inaccessible (winter, night or bad weather). For now, it seems like IR images are only used by beekeepers, not computer vision models. In the same paper, a thermal camera is placed outside and estimates the hot area’s size inside the hive.

In [Yang and Collins, 2019], a camera is placed 30 cm above the entrance, and the landing platform is painted in green so that bees can be distinguished better. Image samples are also collected at the entrance of beehives in [Sledevic, 2018], with a camera placed 40 cm above the beehives’ entrances. Images are then cropped so that they each only contain one bee. Data from several beehives are collected, allowing different backgrounds and more consistency for the training part. The system described in [Ngo et al., 2021] includes an off-the-shelf web camera, and the entrance is enclosed in a black observation box with a lighting control system using a red LED. Compared with the two previous systems, this one allows real-time and long-term monitoring. Those three approaches aim at classifying the same thing: whether or not bees are carrying pollen sacs, and the last one allows bees counting and tracking too. Nevertheless, the first and third one focus on the final accuracy, whereas the second anticipates future deployment on the field. Both works described in [Tashakkori et al., 2015] (bees counting) and [Magnier et al., 2018] (bees tracking) have a similar approach as in [Yang and Collins, 2019] using a white background. This setup also brings the constraint a lower limit for the resolution: on the output images, bees need to appear clearly. For detection purposes, [Chiron et al., 2013a] opted for a limit of 6 pixels per bee. Other computer vision tasks require more pixels per bee (for example, varroa detection).

The authors of [Kulyukin and Mukherjee, 2019] (bees counting) work with cameras placed on top of supers. They focus on cases where one or two supers are added to the hive, making the distance between the camera and the landing platform either 35 cm or 60 cm, giving a field of view of either 50×80 cm or 100×160 cm. In each case, the region of interest retains sufficient pixels per bee to apply computer vision algorithms. The authors of [Kulyukin and Reka, 2016] (bees counting) put the camera above the entrance, in a compartment placed above supers. For research purpose, the experiment can be planned so that the setup is fixed, but supers are added and removed in a real beekeeping context. It changes the camera’s distance from bees over time. In this paper, the authors choose to crop images to only retain a region of interest containing the landing pad. More specialized cameras are used in [Shimasaki et al., 2018] and [Shimasaki et al., 2020] (bees tracking), where 1024×1024 images at 2000 and 500 frames per second are captured, respectively. These cameras are placed in front of the beehive, capturing a $4 \text{ m} \times 4 \text{ m}$ scene. The captured scene aims at tracking wing-flapping movements which happen at a lower frequency than these frame rates.

2.3 Network Architecture: Transmission of Data

The type of the chosen network in PB is determined thanks to the goal of the research itself.

In applied research cases, data can be directly sent to beekeepers. For example, the system presented in [Seritan et al., 2018] sends data through SMS.

When the goal is to train AI models, the focus is not to deploy a network or send the data: it can be stored on hard drives or locally transferred through Bluetooth ([Catania and Vallone, 2019]).

When transmitting data through a network, the network itself shapes the design of the system. Using wireless personal area networks (WPAN) like Zigbee or IEEE 802.15.4 brings the constraints of needing a central station to gather the data and share it through another protocol to the cloud. The range of WPANs is also limited to a few dozens of meters. Low-power wide-area networks (LPWAN) allow sensors to be further from the base station. The range can increase to a few kilometers, which can be ideal when beekeepers choose to move their apiary to different locations in a medium area. Those two groups of networks both have the advantage of consuming small amounts of energy, enabling autonomous systems and low maintenance cost. On the other hand, beehives equipped with cellular networks (GSM, EDGE) need more energy to transfer data but have the flexibility to share larger quantities of information. In the end, the choice of network is up to the objectives: working with stationary data points in time (temperature and gas data) is much lighter than working with heavy file formats like audio, image, or video samples.

The data can be transmitted to a server from each beehive or a central platform at an apiary. In [Edwards-Murphy et al., 2015c] a star network, also known as master-slave architecture is deployed. Each beehive (slave) has a Zigbee gateway capable of communicating with the base (master) Zigbee 3G/GSM station, located at the apiary, which sends the data (temperature, humidity and gas levels) to the cloud storage server. The data is transmitted six times a day from beehives to the base station, and once a day from the base station to the server. In [Kviesis et al., 2015], data is also transmitted from each hive to the central unit, using the 433 MHz radio frequency band. The systems’

structure in [Gil-Lebrero et al., 2017] and [Hong et al., 2020] are also star networks using IEEE 802.15.4 and Wi-Fi protocols, respectively. Each beehive can establish a connection with a local database server to send the data, and each server transfers the information to a global cloud database. The architecture in [Cejrowski et al., 2020] also corresponds to a master-slave schema, with the central device equipped with a GSM module connected to the Internet, whereas slaves communicate with the master through radio modules. In this case, particular attention is given to energy optimization. Each sensor is independent and can enter sleep mode to minimize power consumption. The authors of [Zacepins et al., 2018] chose a LoRaWAN network solution to connect three colonies and monitor their temperature. Since it is single-point values, temperature is a suitable type of data to use LoRa, which is a long-range, low-energy, and low-speed (for data transmission) network. There are two main alternatives when using LoRaWAN: commercial solutions with telecommunication companies or amateur installations for the gateway. In the cited paper, the first option is chosen, allowing connected beehives to include emitters and letting the Latvian operator taking care of the gateway part. In an urban setup, beehives can communicate data to a gateway placed 2.4 km away.

The design of a wireless sensor network (WSN) applied to precision beekeeping must consider the needs of beekeepers, the environment, and the specificities of the collected data. Each research topic has its combination of constraints so that there is no unique solution that fits all cases. In previously cited papers, most of the data is sent to the center of the network. One alternative to the star network is the peer-to-peer topology, where the network nodes are connected, enabling resource sharing for task parallelism. So far, it looks like none of the network-oriented PB papers opted for this kind of design.

In [Kridi et al., 2016], one of the focuses is only to share abnormal data that is worth informing the beekeeper. The objective is to monitor thermal dynamics inside beehives, so normal states are not sent, reducing packets and saving energy. The authors of [Tashakkori et al., 2021] mention the possibility of performing edge computing, i.e., performing the computationally heavy tasks close to the data source, directly near the beehive. By doing so, only the results need to be sent, and LPWAN becomes viable.

2.4 Power Supply

A self-sufficient approach for a power supply system is to use a solar panel to charge a battery. In [Seritan et al., 2018], [Pérez et al.], and [Ammar et al., 2019]’s examples, solar panels are used. Their energy is converted thanks to DC/DC step-down converters into a standardized 5 V that fuels batteries.

In [Edwards-Murphy et al., 2015c], the battery charge of sensors nodes is analyzed and optimized. There are different optimization levels: the rate of sampling (fewer data reads implies less energy consumed), the orientation of solar panels, and the rework of the order of sensors readings.

Other researchers focus on other tasks, simply design their experiments so that the hive is close to a power supply ([Chen et al., 2015] and [Howard et al., 2018]), or chose to power their system only thanks to batteries that are swapped and replaced when empty ([Catania and Vallone, 2020]).

3 Costs of a Connected Beehive

In this paper, we want to encourage using an approach that focuses not only on the quality of results but also on the inclusiveness of PB solutions. IoT and AI are fields which are exponentially developing since ten years ago. Still, the objectives that drive an IoT or AI research paper are often based solely on achieving the best performance. Ideally, including the economic, environmental, social costs, and reproducibility are vital. These factors would facilitate the access of IoT systems and AI models for all, especially beginners with simple tools. The authors of [Schwartz et al., 2020] state that scientific quality criteria should include the efficiency and related AI solutions measures more commonly. Designing efficient models also enables building applications on computationally limited platforms ([Howard et al., 2017]), which perfectly fits the design of PB systems and services.

Empirical analysis shows that PB research papers cited in our survey describing either an IoT system or an AI model do not mention energy nor price costs 70% of the time (51 papers out of 73 in that category). Among the same subset, there are 11 (15%) articles that mention only energy consumption and 7 (10%) which mention only the price of the hardware. 4 articles (5%) present a reflection on both of these challenges.

3.1 Price

The type of collected data directly sets a base for the price of sensors. The goal is then to optimize the return on investment in the quality of information that data reflect. On one side, single point measures like temperature, humidity and weight, temperature and weight seem to be the most obvious choices for a relatively cheap and basic PB installation that tracks the most crucial colony states. On the other, more complex information contained in sound and image data comes at a price, but has the potential, if used correctly, to bring advanced analyses.

The authors of [Zacepins et al., 2016] estimate their Raspberry Pi-based system to monitor 20 colonies at US\$ 140. The complementary network part is US\$ 475. Another Raspberry Pi-based system with audio and video monitoring costs US\$ 106 in [Tashakkori et al., 2021]. In their case, the system is placed in an urban area where the Internet and power are accessible through wires.

In [Anuar et al., 2019], the weight, temperature and humidity data collecting system is estimated at around US\$ 35 per beehive, given that there is a Wi-Fi access. In general, we can note that supposing that Wi-Fi availability considerably simplifies the costs, as many microcontrollers today include this functionality (ESP8266-based microcontrollers, Raspberry Pis). However, real deployment situations are almost always in areas where there is no Wi-Fi within range.

The addition of audio and video recording nodes adds around US\$ 135 to the existing US\$ 135 of the temperature, humidity, natural light level and air quality monitoring system in [Hunter et al., 2019].

Overall, sophisticated cameras aside, the investment in sensors and microcontrollers are usually more affordable than network solutions, although the architecture of a PB system often only requires one network component which centralizes the data transfer. In contrast, the number of each sensor must be equal to the number of monitored colonies. Concerning vision-based systems, simple cameras can be relatively cheap: an official Raspberry Pi camera is worth less than US\$ 35. It provides 3280×2464 pixels images, allowing a potential long-term integration in a PB system. There are, however, research results that are difficult to integrate and distribute to beekeepers. For example, the camera used in [Shimasaki et al., 2020] which produces high-frame-rate videos, is not in the same price range as a Raspberry Pi camera. Still, such research allows the development of computer vision methods applied to apiculture and biological knowledge about bees.

3.2 Energy

As mentioned in 2.4, the power system of a PB system can be based on different strategies. In every case, if the final goal is to produce a beekeeper-friendly solution, the system must be autonomous. Whether it includes solar panels or only batteries, the power consumption must be optimized to avoid any kind of breakdown.

A simple power consumption benchmark is performed for two hours in [Anuar et al., 2019]. The system collecting temperature, humidity, and weight can live up to 12 days with the 2600 mAh batteries used. In this case, this confirms that one-dimensional data collected and pushed to a cloud server is not an energy-heavy task.

A power analysis is performed in [Fitzgerald et al., 2015]. The weighing scale’s consumption is monitored: 200 mW is required to record weight, whereas 726 mW is necessary to send the data through radio (Zigbee). Even if it is unclear how much time each component needs to operate, we can speculate that weight recording is not an energy-heavy task. The authors of [Gil-Lebrero et al., 2017] also measure the energy consumption of the data recording node and the data transmission node. The results confirm the previous tendency: data transmission requires more than twice the current of the collection. In [Zacepins et al., 2020], the weight and temperature monitoring system is analyzed energetically. The main cost, as expected, is not the measurements themselves but the Wi-Fi transfer. The system can perform measurements every 2 minutes and last 18 days with four 1900 mAh batteries.

In [Edwards-Murphy et al., 2015a], the system relies on a 1000 mAh battery. It consists of an infrared camera, a thermal camera, a microphone and an accelerometer that consumes less energy than the 20 W solar panel production.

In [Kulyukin and Mukherjee, 2019], the computational cost is mentioned in terms of GPU hours. Deep learning methods require a much longer time to train, test and validate than the standard machine learning alternatives: 15 to 30 days for the selected convolutional neural networks, versus 15 hours for the random forests on the same task and dataset.

The battery charge level is tracked in [Edwards-Murphy et al., 2015c]. The sampling rate, sampling power, and solar panel orientation are all factors to be taken care of when tuning the parameters of the PB system. For some cases (maximum sampling rate and no solar panel adjustment), it looks like the battery charge tends towards zero. After some modifications, the charge level seems to go back to its initial maximum after the computational part. The system in [Edwards-Murphy et al., 2015d] has an audio recording software separated from the temperature and humidity node. This separation has the advantage of shutting down the audio part between sampling events

to save energy. In this case, the software is also designed so that the energy consumption fits the budget of the solar panel, taking into account its efficiency. To do so, the authors performed an energy analysis of several data recording cycles. In the end, a daily energy expenditure is calculated (1272 Joules) and compared to the solar panel budget (4220 Joules).

The energy footprint of an existing PB system is broken down in [Hadjur et al., 2020]. Boot, data collection, formatting, and transmission (Wi-Fi) are benchmarked. This analysis takes in account the current consumption and the duration of each task to retrieve the total costs, thus identifying the pressure points of the system. The consumption increases by 10% in cold (between 3°C and 5°C) conditions, than similar tasks in temperate conditions. A Raspberry Pi requires more energy (+0.5 W) to function if a USB peripheral is plugged in.

In general, energy-heavy operations are the collection of complex data such as video and image, the treatment of those data (computer vision) and the transmission of any type of data.

The power analysis papers cited in the subsection are useful prerequisites for potential future models which could select scenarios given energy budget and evolving constraints such as weather, battery charge and the needs of a smart beehive. For example, the choice between edge computing (performing the calculation thanks to on-hive microcontrollers) and cloud computing (performing the calculation thanks to cloud infrastructures, after the data is transmitted) could depend on several constraints, especially the energy budget.

3.3 Open-Source access

A trend for some research teams, publishing the code and the data from a paper is always a positive contribution: it allows other researchers to confirm the models on their building own data or test the data with other models. In some cases where the published work contains some weaknesses, it helps to correct it. In this part, we show a non-exhaustive list of diverse open-source initiatives in order to display examples of contributions that can serve as a base of new projects.

Over the years, the authors of [Kulyukin et al., 2018] kept publishing the code, dataset and detailed methodology for their system, alongside their results. They made available two audio datasets containing "bee", "cricket" and "noise" labeled sound samples¹, the code for the machine learning and deep learning experiments² and the rest of the BeePi project code³.

The authors of [Ammar et al., 2019] published the plans of their connected roof and beehive base⁴ as well as the code running inside their PB system⁵.

The dataset from [Nolasco and Benetos, 2018a] is constituted thanks to the combination of various sound data from individual beekeepers worldwide, who collected their data and added the state of each colony, forming a rich dataset⁶.

The authors of [Nolasco and Benetos, 2018b] published the code for their sound classification models⁷.

3.4 Research Projects involved in Precision Beekeeping

Throughout the literature review, we note that apart from a few exceptions, all the cited papers mention a link between their research and a broader project. Whether funded by one university or by a consortium of academics and governments, this brings money to fuel PB research. Each project addresses some specific issue.

For example, SAMS - Smart Apiculture Management Services⁸ is a project from the European Union's Horizon 2020 research and innovation programme focusing on PB in Europe and its links with countries such as Ethiopia and Indonesia. This shows that even if Figure 1 gives a global overview of the locations of PB papers, the underlying projects are sometimes international. On the opposite side of the spectrum, Hiveopolis⁹ aims at adapting PB to nowadays' constraints, especially in urban conditions, thanks to environmental data. In the same direction,

¹Audio Datasets from [Kulyukin et al., 2018]: <https://usu.app.box.com/v/BeePiAudioData>

²Python source code for the deep learning and machine learning experiments in [Kulyukin et al., 2018]: https://github.com/sarba-jit/EBM_Audio_Classification

³Source code used in BeePi project ([Kulyukin et al., 2018]): <https://github.com/VKEDCO/PYPL/tree/master/beeipi>

⁴Fabrication plans from [Ammar et al., 2019]: <https://github.com/emlyon/makers-beehives-fabrication> and <https://github.com/emlyon/makers-beehive>

⁵Source code used in [Ammar et al., 2019]: <https://github.com/emlyon/makers-beehives-hardware>

⁶To bee or not to bee: An annotated dataset for beehive sound recognition: <https://zenodo.org/record/1321278>

⁷Code from [Nolasco and Benetos, 2018b]: <https://github.com/inesnolas/Audio-based-identification-beehive-states>

⁸SAMS: https://sams-project.eu/wp-content/uploads/2020/06/GA_780755_D.5.3.-Evaluation-report-on-Bee-Mgmt-Health.pdf

⁹Hiveopolis: <https://www.hiveopolis.eu/>

Apicampus¹⁰ is a university project aiming at studying the impact of stress factors on bees' health, including pollutants, pesticides, pests, or other environmental threats.

Bpractices¹¹ is another European project. It aims at diagnosing potential honeybee diseases all the while designing connected beehives which are respectful towards bees.

ITApic (Application of information technologies in Precision Apiculture)¹² is a European project which objective is more system-oriented than others. It focuses on developing a beehive monitoring system, on the wireless sensor networks and their implementation in Europe.

Swarmonitor¹³ is a European consortium which objective is to develop monitoring tools based on beehive's vibrations. The NU-Hive project¹⁴ also focuses on a specific type of data: sound. Every contribution coming from this project explores ways to monitor the beehive condition thanks to audio samples.

In some cases, PB start-ups also participate in research projects, such as Pollenity¹⁵ – a company providing tools to monitor temperature, humidity, weight, and sound of a colony – that is part of Hiveopolis. The authors of [Howard et al., 2018] use Arnia¹⁶ commercial sensors. Moreover, they also add their system to complement and compare their solution to commercial sensors'.

The European Space Agency (ESA) also announced a project in PB: PASST - Precision Apiculture Supported by Space Technologies¹⁷. It gives access to the satellite data (Earth images) in order to perform crop analysis, thus selecting the best location for an apiary.

In general, research papers which come from North America are funded thanks university grants. Outside Europe most research papers are the results of small-scale projects. European Union's consortiums have been boosting the research in PB within the past few years as shown in Table 1.

4 Data Collection as a Service in a Smart Beehive

In this section, we go through every type of data collected in the PB literature and describe the different methods of collection for each category.

4.1 Challenges

The main challenge of data collection with the objective of building statistical models in apiculture is bias. Most of the time, the duration of the collected data ranges from a few minutes to a few days, whatever the type of the collected data ([Giammarini et al., 2015]). Moreover, each colony is unique, so analyzing just a few beehives for a small proportion of the life cycle of bees is not enough to understand all the subtleties of an apiary. Bias is even more present with models trained thanks to data from one beehive ([Sledevic, 2018]). Research initiatives that aim at understanding the life of bees need to include at least data from an entire year, since bees have drastically different lifestyles between summer and winter, and their behavior at the beginning and at the end of the high season differ too. The authors of [Davidson et al., 2020] tackle this challenge by combining different sources of data, coming from different apiaries, and different research initiatives, forming a robust dataset where a model can be trained on one part and tested on another one, independent of the first.

There are also different levels when it comes to collecting bee-related data. In [Zacepins et al., 2015], there are three categories: apiary-level, colony-level, and individual bee level. In our paper, we consider the latter as part of the colony-level category since data from counting systems or images of the entrance (considered "individual bee level" in the cited paper) are always analyzed colony-wise to assess the health of the entirety of a beehive.

4.2 Apiary level

The wind is measured and compared to honey production in the work described in [Catania and Vallone, 2020]: wind variations are correlated with beehive development over a period of a month. When long windy periods

¹⁰<https://www.irit.fr/neocampus/fr/on-en-parle/post/95/projet-apicampus-ruches-connectees>

¹¹Bpractices: <https://www.izslt.it/bpractices/en/the-project/>

¹²ITApic: <https://www.era-learn.eu/network-information/networks/ict-agri/ict-and-automation-for-a-greener-agriculture/application-of-information-technologies-in-precision-apiculture>

¹³Swarmonitor: <http://www.swarmonitor.com/>

¹⁴NU-Hive: <https://www.researchgate.net/project/NU-HIVE-New-technologies-for-bee-hive-monitoring>

¹⁵Pollenity: <https://pollenity.com/>

¹⁶<https://www.arnia.co.uk/>

¹⁷PASST: <https://business.esa.int/projects/passt>

occurred (several days), bees refrain from going out to collect pollen and the honey production stagnate.

The threats to bees can vary when considering different countries. The flame sensor deployed in Rwanda ([Ntawuzumunsi and Kumaran, 2019]) constantly monitors smoke levels and emits an alarm if a threshold is exceeded, indicating near forest fire which can cause apiary destruction.

The data gathered in [Edwards-Murphy et al., 2015b] include meteorological and environmental metrics. Outside temperature, wind speed, atmospheric pressure, and rainfall are collected thanks to a nearby weather station. Sunlight data is obtained thanks to an airport located 40 km from the apiary. Wind speed is sometimes collected thanks to a nearby weather station, such as [Ngo et al., 2021]’s case where the collaboration of a national weather bureau is mentioned.

Choosing the right place to install an apiary is often decided based on beekeepers’ experience. When realized correctly, it allows bees to focus their meals on unique flower types, enabling honey production with distinct flavors. This practice can sometimes be inaccurate, and the authors [Komasilova et al., 2020] propose a data-driven approach rather than experience-driven. Thanks to manually annotated aerial and satellite images of agricultural fields, the optimal position of an apiary is computed.

4.3 Colony level

4.3.1 Inside the beehive

Putting a sensor inside a beehive to gather data from a colony is an advantage over being outside. The concentration of life and the organization inside a hive are much more significant than monitoring individual bees flying outside.

There are, however, several challenges when it comes to measuring colony-level metrics inside a hive. First, it is essential to avoid affecting bees’ life with sensors that could disrupt the structure of frames. Bees also produce propolis which they usually use to seal unwanted holes and isolate them from foreign objects like sensors. As mentioned in Section 2.2, cases of propolised inner-hive sensors are not always mentioned in the literature, and even regardless of the type of collected data. Finally, modifying the structure of a beehive to fit IoT systems would create empty spaces that bees fill with combs.

Temperature, Weight & Gas levels collection The analysis of the temperature of the cluster that bees form during winter to save heat is performed in [Markovic et al., 2016]. Thermometers are placed vertically on a frame. The collected data in one dimension is correlated to the 2-dimension distribution of the actual temperature on the frame and can help monitor the beehive during winter. In this case, although there are up to six sensors and their wires directly on a frame, the level of intrusiveness and the reaction of bees toward the sensor network is unclear. In [Hunter et al., 2019], internal and external temperatures and internal gas levels are recorded for several weeks during spring. The daily periodic component of the data is highlighted. Inside a beehive, the temperature reaches its maximum in the afternoon, decreases during the night and becomes cool in the morning. Gas level fluctuations have daily and weekly periods explained by the road traffic near the apiary.

The authors of [Edwards-Murphy et al., 2015b] and [Murphy et al., 2016] work with temperature, relative humidity, and gas (air contaminants, CO₂, O₂ & NO₂) data collected over 16 days on a single hive at rates of 3 times per day for gas levels and 6 times for the rest.

Their goal is to develop models that can detect any abnormal colony state. The authors of [Stalidzans and Berzonis, 2013] use temperature data collected on 14 colonies every 15 minutes during a year. Their goal is to model temperature dynamics with regard to the different long-term phases in a colony. Five periods are identified and each is linked to a linear evolution of the temperature: winter brood rearing, spring brood rearing, summer brood rearing, autumn brood rearing and autumn broodless period. They mention that direct contact between sensors and bees did not impact measurements. In [Anuar et al., 2019], the weight, temperature and humidity of the inside of a beehive are recorded every 7 seconds for 36 hours. The variation of a beehive’s weight is correlated with the moment of the day when bee workers leave and come back inside their hive. It is concluded when there are fewer bees inside the beehive, it is suitable for a beekeeper to perform the checking routine. To detect theft, an alert is sent to the beekeepers when an abnormal variation of weight is detected ([Ntawuzumunsi and Kumaran, 2019] & [Anuar et al., 2019]).

Until now, intelligible data collection has been reviewed. In a PB system final solution, such information can directly be sent to beekeepers who can apply their expertise to make the right diagnostic.

Sound collection In the following paragraphs, data is less straightforward than the previously explained cases. It is up to researchers and beekeepers to ally themselves in order to define objectives and take advantage of artificial intelligence to produce intelligent services.

With the goal of publishing an open-source dataset, the contributors of [Nolasco and Benetos, 2018a] and [Cecchi et al., 2018] regrouped sound recordings collected from individual beekeepers together with corresponding beehives' states. The resulting dataset is diverse and rich compared to more classical research initiatives. The types of microphones and their place inside beehives vary and bees are different from an apiary to another.

There is still a discussion within the scientific community about the invasiveness of microphones when collecting audio data directly inside the beehive. Although researchers in [Nolasco et al., 2019] and [Cecchi et al., 2019] state that sound analysis of acoustic signals is not invasive, the authors of [Aumann et al., 2017], who opted for an outside alternative, argue that regular microphones systems are still a disruption to the colony.

Another inside vibrational type of data can be collected thanks to accelerometers placed directly on frames ([Ramsey et al., 2020]). Their setup is one of the most complete of the apiary literature: 25 beehives are recorded in different apiaries and over two years, resulting in roughly 2000 hours of vibrational recordings, making a prediction (not detection) of swarming events possible.

A posteriori observation of a key event in the life of the colony, such as swarming, is an important dimension to be considered before building learning models. However, beekeepers also gather a certain amount of information during their inspection visits (on average, once per week). For many papers, this data is simplified (prediction of only one trait, such as queen presence, swarming, or pest infestation). The protocol described in [Zhang et al., 2021] involves frame inspection (the percentage of bee or brood coverage is collected) and disease diagnosis in addition to potential rarer and more severe events. This is an original approach for building a model which allows thinner data labeling.

4.3.2 Outside the beehive

Image collection There are different factors when comparing ways of collecting images of bees: the place and orientation of the camera around the beehive, and whether the beehive is modified to allow better conditions (lighting or forcing bees to a specific path).

During peak hours, dozens of bees can be spotted around the entrance of their hives, providing much information but also making the tracking and counting challenging. On paper, one individual bee's trajectory could be misinterpreted, leading to misidentification, especially since bee motion is fast (8 m/s) and unpredictable. Another challenge arises from the outside constraint: environment. Light and background vary throughout a day or a year. Any computer vision system needs to take into account this external factor.

An artificial setup is established in [Schurischuster et al., 2018] to get around the aforementioned constraints. The entrance of the hive is modified so that a tunnel made of several lanes becomes the new path for bees to commute. Bees are forced into one of those lanes so that they navigate through this tunnel, allowing the perfect angle for video shooting. The modified entrance is also artificially lighted up and the sides are covered with dark adhesive foil to prevent reflection. This static setup also allows to pre-tune the focus and position of the camera. The authors of [Chen et al., 2012] proposed a modified version of the entrance too, where bees are confined when they go in and out. In their system, bees are tagged and individually tracked thanks to image recognition. This system inspired the one shown in [Bjerger et al., 2019], where bees also go through a particular darkened narrow passageway. Here, a mirror placed under the entranceway reflects bees into the camera and allows capturing bees images from their ventral side, where the varroa mite can be spotted. The mirror also allows illuminating bees with specific wavelengths since the mite is often well disguised. In this case, on top of the passive way of capturing images (waiting for bees to travel naturally), researchers opted for a more aggressive method which consists of shaking bees off from a frame to force them to go back in.

The authors of [Chiron et al., 2013a] compared "3D" cameras (stereo camera and time of flight camera) pointing at the entrance from above, with the goal of building an acquisition system without any artificial setup which changes the beehive. Stereo cameras' depth maps are more accurate and have a better resolution than the time of flight alternative. After segmentation and tracking are applied, results show that 3D bee tracking outperforms 2D tracking for numerous bees on simulated data. This technique was promising in 2013, but recent research about it lacks since then. Today, we can speculate that such cameras are still too expensive to be integrated into a cheap PB system. With a Raspberry Pi and its official camera, a 2D 3280×2464 pixels image can be captured, providing a good base for computer vision algorithms.

Vibrations One kind of non-invasive technique of PB that removes interactions between bees and wires, processing units, and sensors is the use of a Doppler radar. Placed on the outside of the entrance, this type of radar is described by the authors of [Aumann and Emanetoglu, 2016], who built an original non-intrusive setup gathering

acoustic data retrieving the typical sounds from a colony, like queen tooting and quacking. In another publication ([Aumann et al., 2017]), sounds and vibrations of a hive can be measured in the frequency spectrum thanks to the same type of radar microphone, and looking at the sign of frequency shifts, it is possible to spot individual bees flying out (negative) or back (positive) to their hive. Although only from the entrance of a beehive, the data collected by an outdoor radar still reflects the general health of a colony, as it is known that many events happen at the front of a hive, like foraging, guarding and fanning. The disruption of this activity is directly reflected by a weakness of a colony. Their last published work, [Cumha et al., 2020], retraces the analysis of data collected by their radar system over two years. The correlation between colony health and activity at the entrance is identified thanks to data collected from a dozen colonies and three apiaries. It is worth noting that such radar placed outside the beehive is designed so that its autonomy and its small size are suitable for professional deployment.

4.3.3 Presence detector

Counting the bees going in and out of the beehive can be performed thanks to a modified version of the entrance of the beehive ([Chen et al., 2015]). Thanks to an infrared transmitter-receiver module, the body of a bee as an obstacle to the infrared beam changes the value obtained at the receiver end. This modification is first registered as an analog signal, then transformed into a digital one thanks to a field-programmable gate array in order to perform computation.

4.4 Bibilometric analysis: data types and their associated services

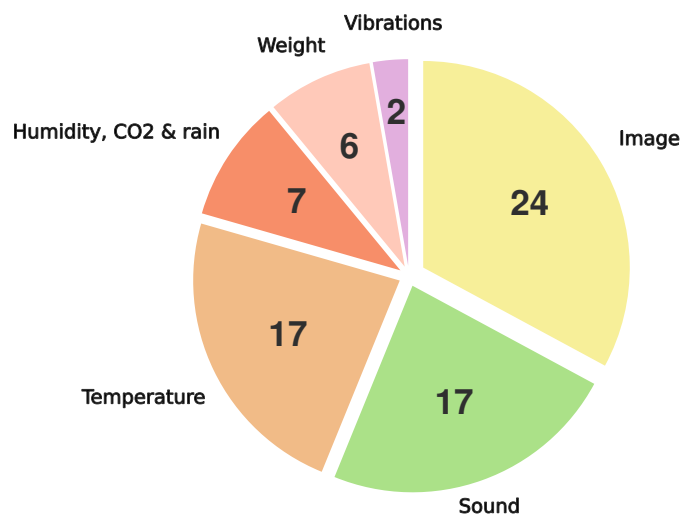
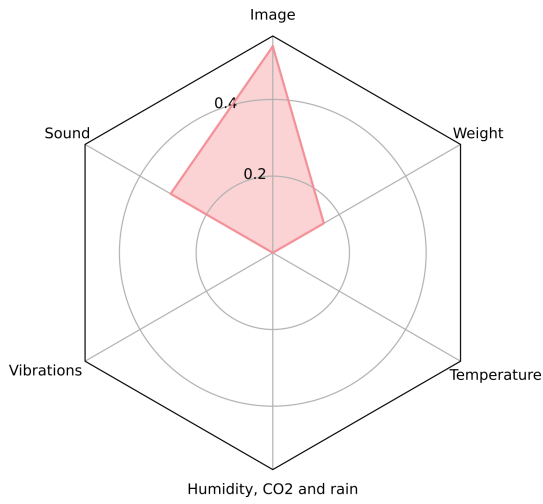


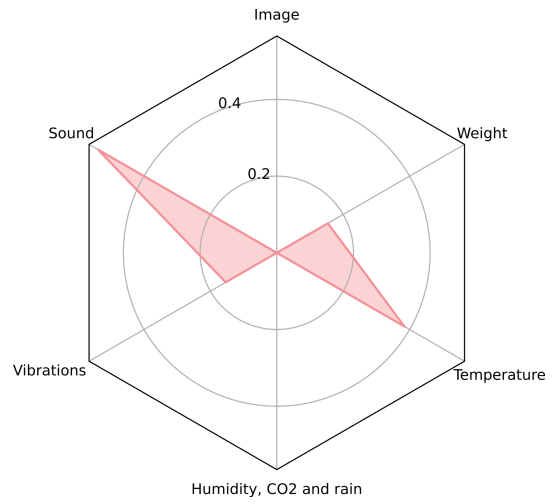
Figure 2. Number of cited papers using data in a learning model, grouped by data types

So far, Section 4 has introduced the different types of collected data. Figure 2 helps to gain an understanding of the proportions and number of articles related to each type of data (image, sound, temperature, humidity, CO2, rain, weight, and vibrations). In this figure, every article using bee-related data in a learning model is listed. Most articles focus on one type of data, but some can be counted several times if more than one of the types are used, like [Ferrari et al., 2008] with sound and temperature. The information gained thanks to this figure is that even though existing installations for beekeepers rely on external sensors like scales, PB scientific research focuses on more complex data like images and sounds thanks to sensors placed closer to bees. Systems used for research purposes are more intrusive than existing commercial solutions. Furthermore, they are more challenging to scale up in terms of the number of equipped beehives but enable a higher quality of data.

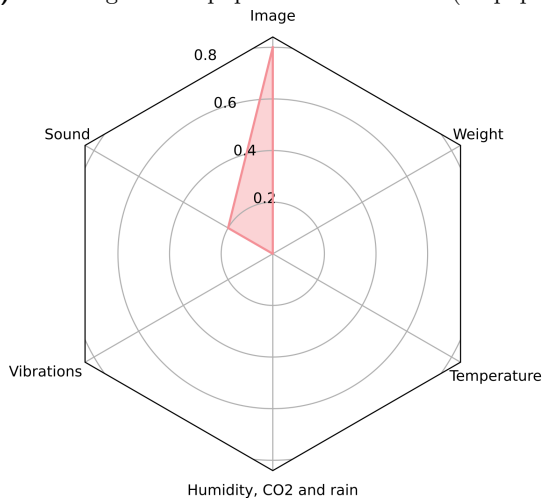
Figure 3 groups papers from Figure 2 by services. It illustrates the types of data required to perform specific tasks. For example, Figure 3a shows that 54% of articles counting bees or estimating the population of a colony use images for their models. Within the same subset, 31% of papers use sound, and 15% use weight. Swarm is detected or predicted mainly thanks to the sound. In addition to the sound, vibrations, weight, and temperature can also be used (Figure 3b). Pest detection (Figure 3c) is mainly performed thanks to images and occasionally sound. Finally, honey production (Figure 3d) is estimated thanks to weight, temperature, and the mix of humidity, CO2 level, and rain level.



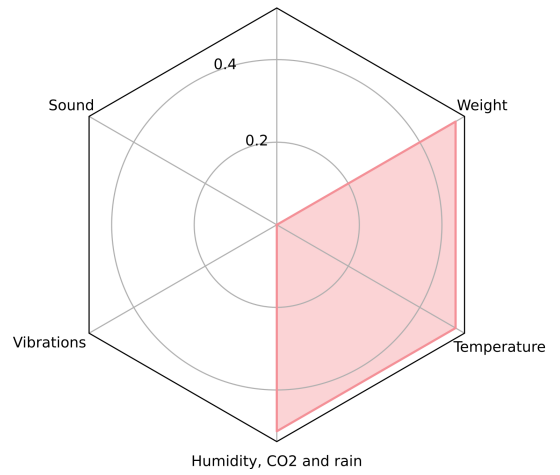
(a) Counting bees & population estimation (13 papers)



(b) Swarm detection & prediction (13 papers)



(c) Pest detection (5 papers)



(d) Honey production (4 papers)

Figure 3. Groups of services in PB and the data types needed to perform them

Overall, we focus on eight categories. Figure 3 shows four categories of services. The remaining ones are not shown in Figure 3 since they use fewer data types, thus less significant to represent graphically.

- **Counting bee & population estimation:** 13 papers (Figure 3a).
- **Swarm detection & prediction:** 13 papers (Figure 3b).
- **Pest detection:** 5 papers (Figure 3c).
- **Honey production:** 4 papers (Figure 3d).
- **Classification of health states** which are defined differently in each paper: 10 papers, all using a mix of temperature, humidity, CO2 level and rain level data.
- **Pollen detection:** 4 papers, all using image data.
- **Tracking:** 9 papers, all using image data.
- **Queen detection:** 5 papers, all using audio data.

5 Data Analytics as a Service in a Smart Beehive

In this section, different models used to handle bee-related data are exposed. We describe the selection process of the models, the tasks that they perform, and their accuracy.

5.1 Audio

5.1.1 Biological knowledge about bees' sound signals

The response from workers to induced acoustic vibrations has been analyzed since the middle of the 20th century ([Hansson, 1945] & [Frings and Little, 1957]). With vibrations between 100 and 2000 Hertz, bees stop their activities.

Audio and vibration measurements of beehives have been performed since the second half of the 20th century in order to understand interactions within a hive. Sound has a crucial role in the life of bees. They convey such signals thanks to the vibration of their thorax and the muscles of their wings, and the movement of their wings. Sound is transmitted through the air, and vibrations propagate through the wax comb or by contact between two individuals ([Kirchner, 1993]). In [Wenner, 1964], one of the first spectral analysis of bee's sound is performed: a microphone and a spectrograph are placed inside a beehive and the sound emitted during the waggle dance and soon after the birth of a new queen are examined.

Thanks to acoustic communication, bees share information with their colony about the distance, direction, and profitability of potential food resources. Bees communicate about food and potential dangers and their relationship together, especially for queens. Table 1 shows the associations between frequencies and their significance, which are vital to know for a beekeeper and scientists who wish to analyze beehive states. Because frequencies overlap, beehive sound data analysis is necessary to assess a colony's health and preventing dangerous situations. In all cases from the literature that include learning models, sound analysis involves classification models (rather than regression). The collection of a beehive's sound aims at capturing the general state of the colony. Inside a beehive, bees act as a super-organism. If an individual bee covers the sound of the whole colony, the global pattern cannot be picked up. This pattern changes based on events like swarming ([Anand et al., 2018]) and even over the length of a day, as bees do not emit in the same frequencies in the morning compared to the afternoon ([Pérez et al., 2016]).

5.1.2 First step of sound data analysis: filtering

Sound filtering is an optional first step when classifying audio samples. In [Qandour et al., 2014], a low-pass Butterworth filter is applied to restrain frequencies from 0 to 4 kHz, since the targeted sounds in this study usually reside below 1 kHz.

The work described in [Cejrowski et al., 2020] focuses on the whole colony's sound, so outlier sound samples where a single bee passes too close to the microphone are removed. Samples are only kept when the frequency of their most significant harmonic component is lower than 600 Hz, or when their Root Mean Square level does not deviate from more than 80% of the dataset average.

Action	Frequency range	Sender	Significance
Flying	250 Hz	Colony	
Fanning	225-285 Hz	Worker	Regulate hive temperature & Evaporate water from nectar
Hissing	0 to several thousand Hertz*	Colony	Response to danger
Tooting**	300-500 Hz	Queen	Prevent hatching (other queen cups) & Inform workers about other queens
Quacking**	<500 Hz	Queen larvae	Answer to tooting from confined queens to alert about their presence
Waggle	200-350 Hz	Forager	Share location and quality of food

*Although most of the literature pieces state that bees do not communicate at frequencies over a few thousand Hertz, [Papachristoforou et al., 2008] recorded the reaction of Cyprian honey bees to induced hornet attacks, showing a dominant frequency of 6 kHz and harmonics reaching 16 kHz.

** Tooting and quacking are some subtypes of queen piping sounds.

Table 1. Types of vibration and sound emitted by bees ([Qandour et al., 2014], [Zgank, 2018] & [Terenzi et al., 2020])

5.1.3 Second step: features extraction

Machine learning classification models take numbers, tables, or matrices as input. The first mandatory step in sound classification is to extract some numerical features from audio samples.

Mel Frequency Cepstral Coefficient (MFCC), Mel Spectra, and Hilbert-Huang transform (HHT), which are signal decomposition methods, are tested in [Nolasco et al., 2019]. The authors of [Zgank, 2018] and [Cejrowski et al., 2020] also work with MFCC, as well as [Robles-Guerrero et al., 2017], in which statistical descriptors (mean, standard deviation, kurtosis, skewness, median, etc.) are used to decrease dimensionality. The Fourier transform is used in [Qandour et al., 2014] to extract four features from the frequency domain: spectral centroid, peak frequency, root variance frequency, and bandwidth. After the training part complete, the weights of the model show that the bandwidth is the least important parameter, whereas the spectral centroid and the peak frequency are deciding features for pest detection. The S-transform ([Stockwell et al., 1996]) is used in [Howard et al., 2013] to extract Fourier-like features, with the advantage of using a variable window length. Due to the high number of dimensions, the self-organizing map method is used to simplify the features set. The Fast Hartley Transform is used in [Anand et al., 2018] to perform the same task as FFT. However, this variation is less computationally heavy and uses less power, enabling real-time results on limited installations. Wavelet transforms (Discrete Wavelet Transform DWT & Continuous Wavelet Transform CWT) are used in [Rybin et al., 2017] and are studied alongside HHT in [Cecchi et al., 2019], where frequency properties during swarming are highlighted. To extract features from audio samples, the authors of [Cejrowski et al., 2018] choose Linear Predictive Coding (LPC). With 14 coefficients of dimension 3000, the sound is compressed but keeps enough information to describe the signal well. Those vectors are then reduced to a size of 16 thanks to the t-Distributed Stochastic Neighbor Embedding (t-SNE) method. Such low dimension objects aim at assessing the possibility of future queen presence classification: with a 2-D visualization of the two principal components, clusters can be observed (See Table 2).

5.1.4 Training models, accuracy & Conclusions

The quality of the training data is essential to build a solid classification or regression model. Manually annotating the data is a necessary condition to build a robust dataset.

Singular Value Decomposition is applied to MFCC statistical descriptors in [Robles-Guerrero et al., 2017], and then a logistic regression is performed. With the data from two beehives (one healthy with the queen and one unhealthy without its queen) during one afternoon, which is a tiny and biased dataset, the presence of the queen is still detected perfectly.

The author of [Zgank, 2018] compared the swarm detection performances of different parameter tunings for Gaussian mixtures Hidden Markov Models (HMM). 81% of accuracy is obtained with 32 Gaussian mixtures, whereas 4 Gaussian mixtures perform at 68% of swarm/no-swarm classification accuracy. Although not precisely specified, the computational time and energy consumption of the latter alternative is the most compatible with an autonomous PB system.

In [Ramsey et al., 2020], swarming events are successfully predicted up to a month in advance thanks to vibrational data: one of the two methods used involves two cross-correlation products between the averaged spectra and specific discriminant functions. Based on the results, an alarm threshold is defined. This first method only performs well during swarm seasons. The second method follows the same cross-correlation product idea. However, this time the entry is a 3D Fourier Transform obtained thanks to the information contained in the result of the combination of the Fast Fourier Transform (FFT) from several days prior to the current data point and discrimination spectra

Reference	Nb of hives recorded & Duration	Method	Learning Task
[Ferrari et al., 2008]	3 during 270h	Filtering and spectrogram comparisons	Swarm detection
[Eskov and Tobeov, 2011]	12 over 1 spring and summer	POLS with Gaussian kernel	
[Howard et al., 2013]	4 over 1 week	FFT, STFT, S-transform & SOM	Queenless state identification
[Qandour et al., 2014]	*	FFT, SFTF, PCA, LDA & SVM	Varioa mites detection
[Kulyukin and Reka, 2016]	4 over 2 days	FFT & [Tolstov, 2012]	None
[Pérez et al., 2016]	3 over 2 months	Spectrogram comparisons	None
[Ramsey et al., 2017]	1 in UK & 1 in France over a season	Spectrogram comparisons, PCA	None ("whooping" detection)
[Aumann et al., 2017]	1 over 4 days	Doppler radar data and its features	Beehive activity
[Robles-Guerrero et al., 2017]	2 over 45 days	MFCC, SVD & Logistic Regression	Queenless state identification
[Cecchi et al., 2018]	3 over a season	Harmonics comparisons	Characteristics of different moments of a day Characteristics of an orphaned colony Characteristics of the reaction to smoke
[Zgank, 2018]	Dataset: [Nolasco and Benetos, 2018a]	MFCC & HMM	Swarm detection
[Cejrowski et al., 2018]	1 over 7 months	LPC, t-SNE & SVM	Queenless state identification
[Anand et al., 2018]	*	FHT	Swarm detection
[Nolasco and Benetos, 2018b]	Dataset: [Nolasco and Benetos, 2018a]	MFCC, Mel spectra, SVM & CNN	Presence of bees in recording
[Nolasco et al., 2019]	NU-Hive data: [Cecchi et al., 2018]†	MFCC; Mel spectrograms, HHT, SVM & CNN	Queenless state identification
[Kulyukin et al., 2018]	6 over 3 to 6 months	CNN; Logistic Regression, k-NN, SVM; Random forests	Classify "bee", "cricket" and "noise" audio samples
[Hunter et al., 2019]	4 over 18 months	Spectrogram comparisons	None
[Cecchi et al., 2019]	NU-Hive data: [Cecchi et al., 2018]†	STFT, MFCC, HHT, CWT & DWT	Swarm detection
[Cunha et al., 2020]	Dozens over 2 years	Doppler radar data	Beehive activity and health
[Ramsey et al., 2020]	25 over 2 seasons	Average vibrational spectra, 3DFT, PCA, DFA	Swarm prediction
[Cejrowski et al., 2020]	2 over 1 month	RMS, ACF, MFCC & SVM	Classify bees day and night
[Zhang et al., 2021]	26 over 79 days	FHT, MLP & GPN	Count bees frames, brood frames and assess disease severity

POLS: Procedure of the Optimal Linear Smoothing; FFT: Fast Fourier Transform; STFT: Short-Time Fourier Transform; SOM: Self-Organizing Maps; PCA: Principal Component Analysis; LDA: Linear Discriminant Analysis; SVM: Support Vector Machine; MFCC: Mel-Frequency Cepstral Coefficients; HMM: Hidden Markov Models; LPC: Linear Predictive Coding; t-SNE: t-Distributed Stochastic Neighbor Embedding; FHT: Fast Hartley Transform; CNN: Convolutional Neural Network; HHT: Hilbert Huang transform; k-NN: k-Nearest Neighbors; CWT: Continuous Wavelet Transform; DWT: Discrete Wavelet Transform; 3DFT: three Dimensional Fourier analysis; DFA: Discriminant Function Analysis; RMS: Root Mean Square; ACF: Autocorrelation Function; MLP: Multilayer Perceptron; GPN: Generative-Prediction Network.

* : unavailable

† : 3 hives over a season

Blue: Features extraction methods

Green: Learning methods

Table 2. Selection of PB papers involving sound and vibration analysis

functions. In the end, temporality aside, the first strategy outperforms the second in terms of false negatives (no alarm when there should be). This is a considerable finding, as it is known that other swarm-related studies can only detect swarming or predict it right before it happens ([Meikle et al., 2006]).

Support Vector Machine (SVM) The learning methods used in [Qandour et al., 2014] to classify pest infestations are SVM and LDA. The results show that only a few audio descriptors are required to obtain a good performance (spectral centroid is giving a 70% accuracy alone). With the four descriptors selected, LDA is able to detect varroa presence at 81%. This study does not specify how many colonies are involved in the sound recording process, but we can speculate that a unique beehive is the only source of the dataset. SVM is also used in [Cejrowski et al., 2018] to detect queen presence and make the difference between two queens introduced to a swarm one after the other. For a specific colony (focusing on one queen), SVM is between 97% and 98% correct at classifying the presence of the queen. SVM is also efficient when the task is to make the classification between the swarm living with the first queen and the swarm living with the second queen. However, the classification model for queen presence, which is trained thanks to the data from the first queen’s state, is not accurate at detecting queen presence in the second queen’s state. In a following paper, [Cejrowski et al., 2020], SVM is used, this time to differentiate day and night time for a colony. Several night lengths are considered and the model’s accuracy is shown for different times when bees start their night. At best, SVM predicts the phase of the day at an 81% accuracy using the main dataset. It is concluded that bees’ deep night is between 11 p.m. and 4 a.m.. The experiments described in [Nolasco et al., 2019] compare CNN with SVM in a sound classification task to determine the presence of the queen. In all setups, SVM performs better than the deep learning alternative. However, the common weakness of those two techniques is the inability to generalize results to new colonies. This also confirms the trend described in the previously cited article.

Deep Learning The authors of [Kulyukin et al., 2018] decide to split training and testing samples by beehive and location in order to validate the ability of models to extrapolate their knowledge. Testing models with data from the same apiary or even colony would introduce bias, and results could have been good thanks to the footprint of a colony rather than the general properties of bees. The latter bring more knowledge as models which are able to generalize are always important contributions. In this research, a second dataset is tested and separated by beehive, location, time, and bee race to try an even more challenging setup. The comparison between standard machine learning and deep learning models in [Kulyukin et al., 2018] shows that distinguishing between "bee", "cricket" and "noise" labeled data can be performed evenly if the data is separated by beehive and location thanks to a custom Convolution Neural Network (CNN), logistic regression, SVM, k-NN or random forests. In their second dataset, where the training and testing data is split by beehive, location, time, and bee race, deep learning outperforms standard machine learning approaches. CNNs also perform better when trained thanks to raw audio samples (rather than spectrogram images). A key question answered in [Kulyukin et al., 2018] is: is it worth embedding a CNN in a PB system? There is no doubt about the accuracy of deep learning, especially with no necessary feature extraction needed and in a more challenging setup where generalization is required. However, the actual cost of such a model resides in the training part, which is longer and more energy-heavy. Even if the experiment cited involves sound classification between "bee", "cricket", and "noise" samples, which is not useful for the beekeeper, the conclusions about the model’s efficiency may be transferred to more practical issues.

In general, advanced beehive audio analyses involving machine learning, which are growing at a fast rate, still rely on the knowledge from experimental studies from the past century. One challenge remains: the length and diversity of data always pose the question of the biological sense of such studies. Of course, models are validated because accurate and efficient in specific situations, but general results about the life of honey bee colonies are too far from the present isolated results.

5.2 Image

5.2.1 Image pre-processing

Pre-processing for image classification in PB involves removing backgrounds and isolating individual bees to count them, track them, and find pollen or parasites on their bodies. In general, the camera setup gives simple enough images (bees in front of a plain background) to prefer color-based segmentation methods rather than deep learning alternatives. Under those conditions, the main difficulty becomes the shadows of the bees in the background, which can lead models to count each bee twice, or to count a bee even if it is not in the camera’s angle. To solve this issue, the dominant blue color of natural light is taken into account. It is the reason why shadows have high saturation

Reference	Res. (FPS if video)	Method	Task
[Ramirez et al., 2012]	708 × 480 (30fps)	MAA detection	Varroa mites detection (without bees)
[Chiron et al., 2013a]	752 × 480 (50fps), stereo camera	GNN, Kalman filters	Bee tracking
[Chiron et al., 2013b]	752 × 480 (47fps), stereo camera	HIDS, GNN, MHT, Kalman filters	Bee tracking
[Tashakkori et al., 2015]	*	CC & OF	Bee tracking
[Sparavigna, 2016]	*	Image Segmentation	Honeycomb analysis
[Chazette et al., 2016]	320 × 240 per bee	RL/CI & CNN	Varroa mites detection
[Kulyukin and Reka, 2016]	720 × 480	OpenCV algorithms	Bee counting (omnidirectional)
[Schurischuster et al., 2018]	1920 × 1080 (30fps)	GMMs, Color Histograms, Dense SURF, Naive Bayes, SVM & RF	Varroa mites detection
[Magnier et al., 2018]	360p, 540p, 720p & 1080p (60 & 30fps)	Edge detection, fit ellipses, concordance	Bee tracking & counting
[Stojnic et al., 2018]	1280 × 720	SIFT, VLAD & SVM	Pollen sacs detection
[Stedevic, 2018]	1920 × 1024 (3fps)	CNN	Pollen sacs detection
[Shimasaki et al., 2018]	1024 × 1024 (2000fps)	STFT	Bee tracking & wing-flapping freq. measures
[Yang and Collins, 2018]	1920 × 1080 (50fps)	Hough transform, Kalman filter, Hungarian algorithm	Bee tracking (directional)
[Yang and Collins, 2019]	100 × 100 to 200 × 200 per bee	MOG & Faster RCNN	Pollen sacs detection
[Kulyukin and Mukherjee, 2019]	360 × 240 & 1920 × 1080	MOG, k-NN, SVM, RF & CNN	Bee counting (omnidirectional)
[Bjerge et al., 2019]	1296 × 966 (31fps)	SIFT, SURF, ORB, LDA & CNN	Varroa mites detection
[Shimasaki et al., 2020]	1024 × 1024 (500fps)	STFT	Bee tracking & wing-flapping freq. measures
[Mukherjee and Kulyukin, 2020]	160 × 120, 320 × 240 & 640 × 480 (25fps)	DPIV	Bee tracking (directional)
[Ngo et al., 2021]	640 × 480 (25fps)	Kalman filter, Hungarian algorithm & CNN	Bee detection, tracking, recognition of pollen and non-pollen bearing bees and counting

MAA: Movement Active Area; GNN: Global Nearest Neighbor; HIDS: Hybrid Intensity Depth Segmentation; MHT: Multiple Hypothesis Tracking; CC: Cascade Classification; OF: Optical Flow; RL/CI: Region Labeling/Color Identification; CNN: Convolutional Neural Network; GMMs: Gaussian Mixture Models; SVM: Support Vector Machine; RF: Random Forests; SURF: Speeded Up Robust Features; SIFT: Scale Invariant Feature Transform; VLAD: Vector of Locally Aggregated Descriptors; MOG: Mixture of Gaussians; STFT: Short-Time Fourier Transform; RCNN: Region-based Convolutional Neural Network; k-NN: k-Nearest Neighbors; LDA: Linear Discriminant Analysis; ORB: Oriented Fast and Rotated BRIEF; DPIV: Digital Particle Image Velocimetry.

* : unavailable

Blue: Image transformation without Machine Learning

Green: Classical Machine Learning

Purple: Deep Learning

Table 3. Selection of computer vision PB papers

in blue channels. Since bees are more yellow and brown than blue, filtering their shadows remains an easy task ([Magnier et al., 2018]). In [Stojnic et al., 2018], the approach starts with considering images in the CIE LAB space and considering the two non-illumination channels, so that differences in light levels are not seen. This segmentation method is compared with k-means. In [Schurischuster et al., 2018], from an original picture of bees to the training features, a foreground detection technique (median image and Gaussian mixture models are compared) and color thresholding (simple region threshold in CIELab and HSV color spaces are compared) are applied. Then, once the background is removed, parts of the image where bees are located are extracted (patch extraction, using connected component analysis). At the end of the processing part, images that are delimited by the shape of individual bees are generated. In [Sledevic, 2018], the same training is performed with RGB and HSV color spaces, in order to compare their effectiveness.

To create a background to be subtracted from pictures where bees appear, synthetic images without bees can be created thanks to the average of many other images where bees can appear ([Tashakkori et al., 2015]). In [Yang and Collins, 2019], Mixture of Gaussians (MOG) models are used to subtract the background. HSV color space is also explored here, with the potential of removing shadows.

5.2.2 Models & Accuracy

Classical Machine Learning Naive Bayes, SVM and random forests (RFs) are tested in [Schurischuster et al., 2018] to classify a balanced 200-image dataset with "mite" and "no mite" labels. As input, image features are obtained thanks to color histogram, color histograms moments, local binary patterns, and dense SURF. This study compares the accuracy and F1-score of different combinations of color models, image features, and classification methods and obtains a maximum accuracy of over 80%. In [Stojnic et al., 2018], pollen-bearing bees and non-pollen bearing bees are classified thanks to SVM mixed with PCA using SIFT and VALD to compute image descriptors. A maximum of 0.915 AUC score is obtained (See Table 3).

Deep Learning In [Sledevic, 2018], a CNN is applied to individual bee images in order to detect pollen-bearing bees. The size of the filter in the convolutional layers has to be larger than the pollen-bearing targets (5 pixels) and smaller than 15 pixels because of computational cost in embedded systems. The number of layers is also crucial as models with 3 layers outperform 1-layer and 2-layer models overall, giving a maximum of 94% classification rate with RGB images and 92% with HSV images from different hives. In [Yang and Collins, 2019], another CNN is tested with the same pollen detection task, using 13 convolutional layers and 3 fully connected layers. This model (7% error rate) is outperforming the statistical two lines model (33% error rate) on the same dataset. One common challenge of the two above-mentioned studies is the distribution of pollen-bearing bee samples (positive). There are 1000 positives and 1000 negative samples in the first analysis, and 10% of them are selected for validation. In contrast, the validation set for the second analysis contains roughly 20 times fewer pollen-bearing bees than non-pollen-bearing bees. In this case, we might question the information given by the error rate, because the model seems to be effective at not finding any pollen sacs but struggles when there is some.

The authors of [Kulyukin and Mukherjee, 2019] opt for a method where a motion detection model suggests areas of interest where potential bees are. Then, CNNs, SVM, and RFs are applied to classify images with and without bees. Results from hand-designed and auto-designed (designed thanks to automation) CNNs are compared, and hand-designed alternatives marginally outperform auto-designed CNNs. RFs perform better than other standard machine learning approaches like SVM. They also represent a middle ground between the maximum accuracy of DL techniques added to the computational resources, and the standard energy-efficient machine learning alternatives, which often fail to produce comparable results.

The main drawback of deep learning applied to PB is the reactivity of models. Even after the training is completed, results often cannot be obtained instantly, especially with CPU-based mini-computers. Although results given a few minutes after the images are collected can still be viable, Ngo et al. present in [Ngo et al., 2021] a real-time pollen sacks detection model using a GPU-based hardware and the YOLOv3-tiny architecture for its deep neural network, known for its fast detection speed. In this paper, the F1-score is 0.94 for pollen and non-pollen bees recognition, and the mean absolute percentage error (MAPE) is $8.45 \pm 2.72\%$ for the pollen-bearing honey count and $10.55 \pm 2.10\%$ for the total incoming bees. After pollen sacks recognition, a Kalman filter and Hungarian algorithm are used to perform tracking tasks.

Other Statistical Methods The authors of [Shimasaki et al., 2018] used Short-time Fourier Transform (STFT) to analyze video data at high frames per second. Doing so, the flight path of bees can be seen as the undulation of

the brightness of pixels and tracked with precision. A 180 Hz frequency component in the spectrum corresponds to the wing-flapping frequency. In [Shimasaki et al., 2020], the evolution of the flapping frequency is added. It ranges from 150 to 250 Hz.

Using stereo cameras (2D image & depth maps), the bee tracking techniques described in [Chiron et al., 2013b] rely on a novel hybrid intensity depth segmentation, Kalman filter, and Global Nearest Neighbors (GNN). In this case, each target is linked to a filter used to estimate the trajectory, and GNN used available information to help with uncertainties. With these tracking models, false positives mainly arise when the flight boards are crowded.

In [Magnier et al., 2018], the movement of bees at regular video frame rates (30 and 60 fps) was analyzed. In their setup, the challenge is to map an individual image per image when two or more bees appear in the video. To do so, the two previously known images are used to predict the next one. For each frame, each bee is represented as an ellipse representing a position and an orientation. Once trajectories are found, entering, departing and passing bees can be counted based on the starting point and finishing point of paths. In this study, four background removal methods are compared based on the counting accuracy.

The use of color distribution in an image (histogram analysis) can be used to detect anomalies such as varroa mites. However, such methods need to add the position of the mites in order to eliminate them. [Chazette et al., 2016] chooses Region Labeling to detect the position of mites and describes a real-world scenario with CNNs and lasers that locate and eliminate varroa mites.

In [Mukherjee and Kulyukin, 2020], a method that computes directional vectors rather than entire flight paths is performed (DPIV: Digital Particle Image Velocimetry) to measure traffic thanks to entrance images. On top of directions estimation, it performs as well as state-of-the-art techniques to measure omnidirectional traffic ([Kulyukin and Mukherjee, 2019]). In comparison to CNNs, DPIV also does not need a long dedicated training part.

5.3 Other types of data

In [Catania and Vallone, 2019], the collected temperature is compared with the data from [Meikle et al., 2016], where thermometers are in the upper part of the beehive, close to the roof, in a less dense area of the hive. The first case finds a weaker correlation between inside and outside temperatures, compared to the second.

The authors of [Kviesis et al., 2020] have a fuzzy logic approach for state detection. In their Fuzzy Inference System, they take into account the inside temperature, the ambient temperature, the difference between two, the evolution of the inside temperature, and the season to detect normal, abnormal and death states. The model obtains a F1-score of 98%.

In [Davidson et al., 2020], swarms are detected thanks to a rule-based algorithm and autoencoders, which consist of a pair of neural networks (encode and decoder) which are, in this case, two long-short-term memory networks (LSTMs). The results show that all swarms are detected thanks to the autoencoders method, and it is even triggered by other anomalies (mainly from beekeeper’s intervention: varroa treatment, control visit).

The authors of [Ngo et al., 2021] show the correlation between environmental data and the activity of a colony (in this case, pollen foraging). The results emphasize that temperature, relative humidity, wind speed, rain level, and light intensity influence pollen foraging. For example, heavy rain or gentle breeze negatively affect the harvest.

In [Edwards-Murphy et al., 2015b], a threshold-based algorithm is developed to assess the general health of a colony. For example, the authors state that the variation of temperature over 24 hours should not be greater than 20 degrees. All thresholds and intervals have been validated thanks to accurate alarms triggering. Advancement of threshold-based algorithm is proposed in [Murphy et al., 2016], where a decision tree is used to evaluate a beehive’s state thanks to heterogeneous data. One of the advantages of such a light-weight program is that it fits on a microcontroller with low capacity and low energy consumption (the ATmega1281 for example). To design the algorithm, a biological analysis is first performed to bring 10 important beehive states out. Some describe normal activities while others, when detected, must trigger an alarm sent to the beekeeper. Then, the training of an Iterative Dichotomiser 3 (ID3) decision tree algorithm is performed thanks to labeled data as input. In the end, over 95% of states are accurately classified. In comparison, the previously cited threshold-based method reached at most 89% of accuracy with the same dataset.

5.4 Discussion

Sections 4 and 5 show that every type of data has the potential to bring intelligence to a connected beehive. The differentiating factor between models that confirms results from the past and a significant contribution is the data

quality. Whether a study uses a single data source or heterogeneous data, the collection over several full seasons over different apiaries and precise labeling are necessary to produce unbiased and general biological results.

To collect data, only few articles document some of their challenges when installing in place their sensors. For better traceability, PB research papers need to specify why and how sensors should be placed, but maybe more importantly, explain the constraints that led to their choices and the reasons why other options are not selected, if possible.

Is it worth to embedding an AI model inside a PB system? The vital aspect to be taken care of is energy consumption, so "where to put the tradeoff between intelligence and energy consumption?" becomes the issue that beekeepers and researchers must agree on. From a research perspective, the final product and the preliminary data collection tools must be distinguished. The first one must fit in a PB system as described in Section 2: non-intrusiveness, efficient and connected. For the latter, non-intrusivity remains essential to respect the bees. Efficiency and connectivity is not a determining factor: training data can be manually downloaded and the data collection process can be monitored closely as opposed to the final product. From a beekeeper's point of view, knowing the needs depends on the size of the apiary. It also depends on the number of apiaries, whether transhumance is performed, and on the money that can be invested. The simple and cheap solutions regroup systems that collect and transfer raw data directly to beekeepers, who then extrapolate to make their opinion thanks to their apiary experience, which is possible. On the other hand, more advanced services provide results that are not manually reproducible but require extra energy resources to compute AI models on-site.

6 Conclusion & discussion

6.1 The state of precision beekeeping research

PB is still a new and emerging research topic: there are many publications in recent years, and those come from various scientific backgrounds, from computer science to agriculture.

It is not easy to generalize results found in the literature, first, because the life-cycle of bees continues to evolve after the data is collected. Then, the experimental conditions are unique for each project since they are conducted at different times, places, and with variations in the design of the beehive, or even the subtype of bee species.

However, all the sensors, network types, energy nodes, data and AI models cited show that there is some coherence and standard bases between PB literature pieces. Thanks to research initiatives and their accessibility, experiments that would have been instantaneously moved on from in the past, can be the source of new ideas and advancements in this field.

Tracking the research in PB remains tricky because the systems, services, and data are directly linked to business considerations. Several working groups appear to be intensely productive during a short time window and then disappear from the scientific community, most often because they stop publishing their findings, protect them, and start a business (ApisProtect from [Murphy and Whelan, 2017]).

6.2 The needs for precision beekeeping research

Looking at beekeepers' needs, it is clear that more affordable commercial solutions are to be developed. There is, of course, a duality between developing a unique solution that could regroup data from all beekeepers and build AI models, and the economic competition. The latter can still push PB in the right direction if many actors strive for cheaper and more adapted solutions. This survey helps to better understand this challenge thanks to Section 3 which help to understand the expenses when adopting precision beekeeping services, and the two following sections (Sections 4 and 5) which give an explanation about the potential benefits in terms of services in surveyed research papers. It is however difficult to imagine the return on investment for beekeepers when adopting the solutions described in the latter two sections and scaling them up in terms of quantity, since every service described uses DIY installations that are hard to mass-produce.

Bees are exposed to many stress factors, some better understood than others, and PB is considerably improving experimental conditions, particularly with the increasing data volumes. The state of PB research today shows solid advances with data like weight and temperature. During the past few years, sound and image-based methods came into play and opened new perspectives. The question is, how far can it lead us? So far, most of the projects are not reproducible. The next step is to integrate this expanding theoretical knowledge into beekeepers' everyday life. However, is it worth equipping all beehives with sensors? Probably not, first because the return on investment is not worth it, and secondly because different colonies still share the same environment.

Electromagnetic fields are also correlated with honey bees' health. Although there is less research on it than other stress factors, [Shepherd et al., 2019] showed an increase in aggressiveness and a decrease in the ability to learn when bees are exposed to low-frequency electromagnetic fields, at rates that are comparable to those found around power lines. [Henry et al., 2019] compared the internal temperature and relative humidity between wireless connected beehives (2.4 GHz Wi-Fi) and their wired twins for 30 days. No effect on bees was observed during that time span. It remains unknown whether PB research results are biased because of this reason.

6.3 The future of precision beekeeping research

This survey highlights the fact that PB research as a system and PB research as a service are often two different things: some papers focus on the IoT technologies and how to efficiently deploy a PB system, whereas others focus on machine learning tasks and classification performances. Those artificial intelligence techniques are often energy-consuming, thus making their deployment difficult. The future of PB lies in the integration of efficient, operational, and deployed AI models. To do so, the energy footprint of end-to-end systems needs to be analyzed to enable autonomy for a complete solution.

From a biological perspective, the 20th century has seen manual observations emerged, proving fundamental apiary knowledge. From the beginning of IoT systems around 2010 to today, digital technologies helped to exponentially improve the quality and the quantity of data collected, as well as the growth of apiary research all over the world. It confirmed previous biological conclusions. From now on, we can hope that more and more connected beehives will arise and help to produce large-scale studies searching for new apiary knowledge.

References

- [ITU, 2012] (2012). Overview of the internet of things. <https://handle.itu.int/11.1002/1000/11559>. Accessed: 2021-11-19.
- [FAO, 2016] (2016). Pollinators vital to our food supply under threat. <http://www.fao.org/news/story/en/item/384726/icode/>. Accessed: 2021-11-19.
- [ADA-France, 2019] ADA-France (2019). L'apiculture professionnelle en chiffres. <https://www.adafrance.org>. Accessed: 2021-11-19.
- [Ammar et al., 2019] Ammar, D., Savinien, J., and Radisson, L. (2019). The makers' beehives: Smart beehives for monitoring honey-bees' activities. In *Proceedings of the 9th International Conference on the Internet of Things, IoT 2019, Bilbao, Spain, October 22-25, 2019*, pages 16:1–16:4.
- [Anand et al., 2018] Anand, N., Raj, V. B., Ullas, M. S., and Srivastava, A. (2018). Swarm detection and beehive monitoring system using auditory and microclimatic analysis. In *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, pages 1–4.
- [Anuar et al., 2019] Anuar, N., Md Yunus, M. A., Baharuddin, M., Sahlan, S., Abid, A., Ramli, M., Amin, M., and Lotpi, Z. (2019). Iot platform for precision stingless bee farming. In *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pages 225–229.
- [Aumann et al., 2017] Aumann, H., Payal, B., Emanetoglu, N., and Drummond, F. (2017). An index for assessing the foraging activities of honeybees with a doppler sensor. In *2017 IEEE Sensors Applications Symposium (SAS)*, pages 1–5.
- [Aumann and Emanetoglu, 2016] Aumann, H. M. and Emanetoglu, N. W. (2016). The radar microphone: A new way of monitoring honey bee sounds. In *2016 IEEE SENSORS*, pages 1–2.
- [Bjerger et al., 2019] Bjerger, K., Frigaard, C. E., Mikkelsen, P. H., Nielsen, T. H., Misbih, M., and Kryger, P. (2019). A computer vision system to monitor the infestation level of varroa destructor in a honeybee colony. *Computers and Electronics in Agriculture*, 164:104898.
- [Bumanis, 2020] Bumanis, N. (2020). Data fusion challenges in precision beekeeping: a review. *Research for Rural Development 2020*.

- [Catania and Vallone, 2019] Catania, P. and Vallone, M. (2019). Design of an innovative system for precision beekeeping. In *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pages 323–327.
- [Catania and Vallone, 2020] Catania, P. and Vallone, M. (2020). Application of a precision apiculture system to monitor honey daily production. *Sensors*, 20:2012.
- [Cecchi et al., 2019] Cecchi, S., Terenzi, A., Orcioni, S., and Piazza, F. (2019). Analysis of the sound emitted by honey bees in a beehive. *Journal of The Audio Engineering Society*.
- [Cecchi et al., 2018] Cecchi, S., Terenzi, A., Orcioni, S., Riolo, P., Ruschioni, S., and Isidoro, N. (2018). A preliminary study of sounds emitted by honey bees in a beehive. In *Audio Engineering Society Convention 144*.
- [Cejrowski et al., 2020] Cejrowski, T., Szymański, J., and Logofătu, D. (2020). Buzz-based recognition of the honeybee colony circadian rhythm. *Computers and Electronics in Agriculture*, 175:105586.
- [Cejrowski et al., 2018] Cejrowski, T., Szymański, J., Mora, H., and Gil, D. (2018). Detection of the bee queen presence using sound analysis. In Nguyen, N. T., Hoang, D. H., Hong, T.-P., Pham, H., and Trawiński, B., editors, *Intelligent Information and Database Systems*, pages 297–306, Cham. Springer International Publishing.
- [Chazette et al., 2016] Chazette, L., Becker, M., and Szczerbicka, H. (2016). Basic algorithms for bee hive monitoring and laser-based mite control. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8.
- [Chen et al., 2012] Chen, C., Yang, E.-C., Jiang, J.-A., and Lin, T.-T. (2012). An imaging system for monitoring the in-and-out activity of honey bees. *Computers and Electronics in Agriculture*, 89:100–109.
- [Chen et al., 2015] Chen, W., Wang, C., Jiang, J., and Yang, E. (2015). Development of a monitoring system for honeybee activities. In *2015 9th International Conference on Sensing Technology (ICST)*, pages 745–750.
- [Chiron et al., 2013a] Chiron, G., Gomez-Krämer, P., and Ménard, M. (2013a). Outdoor 3D Acquisition System for Small and Fast Targets. Application to honeybee monitoring at the beehive entrance. In *GEODIFF 2013*, pages 10–19, Barcelona, France.
- [Chiron et al., 2013b] Chiron, G., Gomez-Krämer, P., and Michel, M. (2013b). Detecting and tracking honeybees in 3D at the beehive entrance using stereo vision. *EURASIP Journal on Image and Video Processing*, 2013(1):59.
- [Cousin et al., 2019] Cousin, P., Căuia, E., Siceanu, A., and de Cledat, J. (2019). The development of an efficient system to monitor the honeybee colonies depopulations. In *2019 Global IoT Summit (GIoTS)*, pages 1–5.
- [Crane, 1999] Crane, E. (1999). *The World History of Beekeeping and Honey Hunting*. Routledge.
- [Cunha et al., 2020] Cunha, A., Rose, J., Prior, J., Aumann, H., Emanetoglu, N., and Drummond, F. (2020). A novel non-invasive radar to monitor honey bee colony health. *Computers and Electronics in Agriculture*, 170:105241.
- [Davidson et al., 2020] Davidson, P., Steininger, M., Lautenschlager, F., Kobs, K., Krause, A., and Hotho, A. (2020). Anomaly detection in beehives using deep recurrent autoencoders. *SENSORNETS 2020*.
- [Edwards-Murphy et al., 2015a] Edwards-Murphy, F., Magno, M., O’Leary, L., Troy, K., Whelan, P., and Popovici, E. (2015a). 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)*, pages 106–111.
- [Edwards-Murphy et al., 2015b] Edwards-Murphy, F., Magno, M., Whelan, P., and Vici, E. (2015b). B+wsn: Smart beehive for agriculture, environmental, and honey bee health monitoring - preliminary results and analysis. *SAS 2015 - 2015 IEEE Sensors Applications Symposium, Proceedings*.
- [Edwards-Murphy et al., 2015c] Edwards-Murphy, F., Popovici, E., Whelan, P., and Magno, M. (2015c). Development of an heterogeneous wireless sensor network for instrumentation and analysis of beehives. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2015:346–351.

- [Edwards-Murphy et al., 2015d] Edwards-Murphy, F., Srbinovski, B., Magno, M., Popovici, E., and Whelan, P. (2015d). An automatic, wireless audio recording node for analysis of beehives. In *2015 26th Irish Signals and Systems Conference (ISSC)*.
- [Eskov and Toboev, 2011] Eskov, E. and Toboev, V. (2011). Changes in the structure of sounds generated by bee colonies during sociotomy. *Entomological Review*, 91:347–353.
- [Ferrari et al., 2008] Ferrari, S., Silva, M., Guarino, M., and Berckmans, D. (2008). Monitoring of swarming sounds in bee hives for early detection of the swarming period. *Computers and Electronics in Agriculture*, 64:72–77.
- [Fiedler et al., 2020] Fiedler, S., Zacepins, A., Kviesis, A., Komasilovs, V., Wakjira, K., Nawawi, M., Hensel, O., and Purnomo, D. (2020). Implementation of the precision beekeeping system for bee colony monitoring in indonesia and ethiopia. In *2020 21th International Carpathian Control Conference (ICCC)*, pages 1–6.
- [Fitzgerald et al., 2015] Fitzgerald, D., Edwards-Murphy, F., Wright, W., Whelan, P., and Popovici, E. (2015). Design and development of a smart weighing scale for beehive monitoring. In *2015 26th Irish Signals and Systems Conference (ISSC)*, pages 1–6.
- [FranceAgriMer, 2020] FranceAgriMer (2020). Observatoire de la production de miel et de gelée royale. franceagrimer.fr. Accessed: 2021-11-19.
- [Frings and Little, 1957] Frings, H. and Little, F. (1957). Reactions of honey bees in the hive to simple sounds. *Science*, 125(3238):122–122.
- [Giammarini et al., 2015] Giammarini, M., Concettoni, E., Zazzarini, C., Orlandini, N., Albanesi, M., and Cristalli, C. (2015). Beehive lab project - sensorized hive for bee colonies life study. In *2015 12th International Workshop on Intelligent Solutions in Embedded Systems (WISES)*.
- [Gil-Lebrero et al., 2017] Gil-Lebrero, S., Quiles-Latorre, F., Ortiz-López, M., Sánchez-Ruiz, V., Gàmez-López, V., and Luna-Rodríguez, J. (2017). Honey bee colonies remote monitoring system. *Sensors*, 17(1).
- [Giurfa et al., 2001] Giurfa, M., Zhang, S., Jenett, A., Menzel, R., and Srinivasan, M. (2001). The concepts of ‘sameness’ and ‘difference’ in an insect. *Nature*, 410:930–933.
- [Hadjur et al., 2020] Hadjur, H., Ammar, D., and Lefèvre, L. (2020). Analysis of energy consumption in a precision beekeeping system. In *Proceedings of the 10th International Conference on the Internet of Things, IoT ’20*, New York, NY, USA. Association for Computing Machinery.
- [Hansson, 1945] Hansson, Å. (1945). *Lauterzeugung und Lautauffassungsvermögen der Bienen*, volume 6. Entomologiska sällskapet i Lund.
- [Henry et al., 2019] Henry, E., Adamchuk, V., Stanhope, T., Buddle, C., and Rindlaub, N. (2019). Precision apiculture: Development of a wireless sensor network for honeybee hives. *Computers and Electronics in Agriculture*, 156:138 – 144.
- [Hong et al., 2020] Hong, W., Xu, B., Chi, X., Cui, X., Yan, Y., and Li, T. (2020). Long-term and extensive monitoring for bee colonies based on internet of things. *IEEE Internet of Things Journal*, 7(8):7148–7155.
- [Howard et al., 2017] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861.
- [Howard et al., 2018] Howard, D., Duran, O., and Hunter, G. (2018). A low-cost multi-modal sensor network for the monitoring of honeybee colonies/hives. *Intelligent Environments*, pages 69–78.
- [Howard et al., 2013] Howard, D., Duran, O., Hunter, G., and Stebel, K. (2013). Signal processing the acoustics of honeybees (*apis mellifera*) to identify the “queenless” state in hives. *Proceedings of the Institute of Acoustics*, 35:290–297.
- [Hunter et al., 2019] Hunter, G., Howard, D., Gauvreau, S., Duran, O., and Busquets, R. (2019). Processing of multi-modal environmental signals recorded from a “smart” beehive. *Proceedings of the Institute of Acoustics*, 41:337–348.

- [Kirchner, 1993] Kirchner, W. (1993). Acoustical communication in honeybees. *Apidologie*, 24:297–307.
- [Komasilova et al., 2020] Komasilova, O., Komasilovs, V., Kvišis, A., Bumanis, N., Mellmann, H., and Zacepins, A. (2020). Model for the bee apiary location evaluation. *Agronomy Research*, 18:1350–1358.
- [Kridi et al., 2016] Kridi, D., Carvalho, C., and Gomes, D. (2016). Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection. *Computers and Electronics in Agriculture*, 127:221–235.
- [Kulyukin and Mukherjee, 2019] Kulyukin, V. and Mukherjee, S. (2019). On video analysis of omnidirectional bee traffic: Counting bee motions with motion detection and image classification. *Applied Sciences*, 9(18).
- [Kulyukin et al., 2018] Kulyukin, V., Mukherjee, S., and Amlathe, P. (2018). Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples. *Applied Sciences*, 8:1573.
- [Kulyukin and Reka, 2016] Kulyukin, V. and Reka, S. (2016). Toward sustainable electronic beehive monitoring: Algorithms for omnidirectional bee counting from images and harmonic analysis of buzzing signals. *Engineering Letters*, 24:317–327.
- [Kvišis et al., 2020] Kvišis, A., Komasilovs, V., Komasilova, O., and Zacepins, A. (2020). Application of fuzzy logic for honey bee colony state detection based on temperature data. *Biosystems Engineering*, 193:90–100.
- [Kvišis et al., 2015] Kvišis, A., Zacepins, A., Durgun, M., and Tekin, S. (2015). Application of wireless sensor networks in precision apiculture. In *14th International Scientific Conference Engineering for Rural Development*.
- [Langstroth, 2004] Langstroth, L. (2004). *Langstroth’s Hive and the Honey-Bee: The Classic Beekeeper’s Manual*. Dover Publications, Incorporated.
- [Lettmann and Chauzat, 2018] Lettmann, M. and Chauzat, M.-P. (2018). Les outils connectés en apiculture : Evaluation de leurs application auprès des apiculteurs français. https://be.anses.fr/sites/default/files/0-028_2018-12-28_Outils-abeilles_Lettmann_VF.pdf. Accessed: 2021-11-19.
- [Magnier et al., 2018] Magnier, B., Ekszterowicz, G., Laurent, J., Rival, M., and Pfister, F. (2018). Bee hive traffic monitoring by tracking bee flight paths. In *13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, January 27-29, 2018, in Funchal, Madeira, Portugal*, pages 563–571.
- [Marchal et al., 2020] Marchal, P., Buatois, A., and Kraus, S. e. a. (2020). Automated monitoring of bee behaviour using connected hives: Towards a computational apidology. *Apidologie*.
- [Markovic et al., 2016] Markovic, D., Pesović, U., Djurasevic, S., and Randic, S. (2016). Decision support system for temperature monitoring in beehives. *Acta agriculturae Serbica*, 21:135–144.
- [Meikle and Holst, 2015] Meikle, W. G. and Holst, N. (2015). Application of continuous monitoring of honeybee colonies. *Apidologie*, 46(1):10–22.
- [Meikle et al., 2006] Meikle, W. G., Holst, N., Mercadier, G., Derouané, F., and James, R. R. (2006). Using balances linked to dataloggers to monitor honey bee colonies. *Journal of Apicultural Research*, 45(1):39–41.
- [Meikle et al., 2016] Meikle, W. G., Weiss, M., and Stilwell, A. R. (2016). Monitoring colony phenology using within-day variability in continuous weight and temperature of honey bee hives. *Apidologie*, 47(1):1–14.
- [Mukherjee and Kulyukin, 2020] Mukherjee, S. and Kulyukin, V. (2020). Application of digital particle image velocimetry to insect motion: Measurement of incoming, outgoing, and lateral honeybee traffic. *Applied Sciences*, 10(6).
- [Murphy et al., 2016] Murphy, F. E., Magno, M., Whelan, P. M., O’Halloran, J., and Popovici, E. M. (2016). b+wsn: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring. *Computers and Electronics in Agriculture*, 124:211–219.
- [Murphy and Whelan, 2017] Murphy, F. E. and Whelan, P. M. (2017). Apisprotect. <https://apisprotect.com/>. Accessed: 2021-11-19.

- [Ngo et al., 2021] Ngo, T. N., Rustia, D. J. A., Yang, E.-C., and Lin, T.-T. (2021). Automated monitoring and analyses of honey bee pollen foraging behavior using a deep learning-based imaging system. *Computers and Electronics in Agriculture*, 187:106239.
- [Nolasco and Benetos, 2018a] Nolasco, I. and Benetos, E. (2018a). To bee or not to bee: An annotated dataset for beehive sound recognition. <https://doi.org/10.5281/zenodo.1321278>. Accessed: 2021-11-19.
- [Nolasco and Benetos, 2018b] Nolasco, I. and Benetos, E. (2018b). To bee or not to bee: Investigating machine learning approaches for beehive sound recognition. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018)*, pages 133–137.
- [Nolasco et al., 2019] Nolasco, I., Terenzi, A., Cecchi, S., Orcioni, S., Bear, H. L., and Benetos, E. (2019). Audio-based identification of beehive states. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8256–8260.
- [Ntawuzumunsi and Kumaran, 2019] Ntawuzumunsi, E. and Kumaran, S. (2019). Design and implementation of smart bees hiving & monitoring system. In *2019 IST-Africa Week Conference (IST-Africa)*, pages 1–9.
- [Papachristoforou et al., 2008] Papachristoforou, A., Sueur, J., Rortais, A., Angelopoulos, S., Thrasyvoulou, A., and Arnold, G. (2008). High frequency sounds produced by cyprian honeybees *apis mellifera cypria* when confronting their predator, the oriental hornet *vespa orientalis*. *Apidologie*, 39(4):468–474.
- [Pérez et al., 2016] Pérez, N., Jesús, F., Pérez, C., Niell, S., Draper, A., Obrusnik, N., Zinemanas, P., Spina, Y. M., Letelier, L. C., and Monzón, P. (2016). Continuous monitoring of beehives sound for environmental pollution control. *Ecological Engineering*, 90:326 – 330.
- [Qandour et al., 2014] Qandour, A., Ahmad, I., Habibi, D., and Leppard, M. (2014). Remote beehive monitoring using acoustic signals. *Acoustics Australia / Australian Acoustical Society*, 42:204–209.
- [Ramírez et al., 2012] Ramírez, M., Prendas, J. P., Travieso, C. M., Calderón, R., and Salas, O. (2012). Detection of the mite *varroa destructor* in honey bee cells by video sequence processing. In *2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES)*, pages 103–108.
- [Ramsey et al., 2017] Ramsey, M., Bencsik, M., and Newton, M. I. (2017). Long-term trends in the honeybee ”whooping signal” revealed by automated detection. *PLOS ONE*, 12(2):e0171162.
- [Ramsey et al., 2020] Ramsey, M.-T., Bencsik, M., Newton, M., Reyes, M., Pioz, M., Crauser, D., Simon-Delso, N., and Le Conte, Y. (2020). The prediction of swarming in honeybee colonies using vibrational spectra. *Scientific Reports*, 10.
- [Reyes et al., 2012] Reyes, O. A. M., Àvila, A. A. M., Sebastian Eslava, G., and Rozo, G. B. (2012). Beekeeping monitoring module. In *2012 IEEE 4th Colombian Workshop on Circuits and Systems (CWCAS)*, pages 1–6.
- [Robles-Guerrero et al., 2017] Robles-Guerrero, A., Saucedo-Anaya, T., González-Ramírez, E., and Galván-Tejada, C. E. (2017). Frequency analysis of honey bee buzz for automatic recognition of health status: A preliminary study. *Res. Comput. Sci.*, 142:89–98.
- [Rybin et al., 2017] Rybin, V., Butusov, D., Karimov, T., Belkin, D., and Kozak, M. (2017). Embedded data acquisition system for beehive monitoring. In *2017 IEEE II International Conference on Control in Technical Systems (CTS)*, pages 387–390.
- [Sakanovic and Kevric, 2020] Sakanovic, S. and Kevric, J. (2020). Habeetat: A novel monitoring platform for more efficient honey production. In Badnjevic, A., Škrbić, R., and Gurbeta Pokvić, L., editors, *CMBEBIH 2019*, pages 193–200, Cham. Springer International Publishing.
- [Schurischuster et al., 2018] Schurischuster, S., Remeseiro, B., Radeva, P., and Kampel, M. (2018). A preliminary study of image analysis for parasite detection on honey bees. In Campilho, A., Karray, F., and ter Haar Romeny, B., editors, *Image Analysis and Recognition*, pages 465–473, Cham. Springer International Publishing.
- [Schwartz et al., 2020] Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020). Green ai. *Commun. ACM*, 63(12):54–63.

- [Seritan et al., 2018] Seritan, G., Enache, B.-A., Florin, A., Adochiei, F., and Toader, S. (2018). Low cost platform for monitoring honey production and bees health. In *2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, pages 1–4.
- [Shepherd et al., 2019] Shepherd, S., Hollands, G., Godley, V. C., Sharkh, S. M., Jackson, C. W., and Newland, P. L. (2019). Increased aggression and reduced aversive learning in honey bees exposed to extremely low frequency electromagnetic fields. *PLOS ONE*, 14(10):1–13.
- [Shimasaki et al., 2018] Shimasaki, K., Jiang, M., Takaki, T., Ishii, I., and Yamamoto, K. (2018). Hfr-video-based honeybee activity sensing using pixel-level short-time fourier transform. *2018 IEEE SENSORS*, pages 1–4.
- [Shimasaki et al., 2020] Shimasaki, K., Jiang, M., Takaki, T., Ishii, I., and Yamamoto, K. (2020). Hfr-video-based honeybee activity sensing. *IEEE Sensors Journal*, 20(10):5575–5587.
- [Sledevic, 2018] Sledevic, T. (2018). The application of convolutional neural network for pollen bearing bee classification. *2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, pages 1–4.
- [Sparavigna, 2016] Sparavigna, A. C. (2016). Analysis of a natural honeycomb by means of an image segmentation. *Philica*, 2016(897).
- [Stalidzans and Berzonis, 2013] Stalidzans, E. and Berzonis, A. (2013). Temperature changes above the upper hive body reveal the annual development periods of honey bee colonies. *Computers and Electronics in Agriculture*, 90:1 – 6.
- [Stockwell et al., 1996] Stockwell, R. G., Mansinha, L., and Lowe, R. P. (1996). Localization of the complex spectrum: the s transform. *IEEE Transactions on Signal Processing*, 44(4):998–1001.
- [Stojnic et al., 2018] Stojnic, V., Risojević, V., and Pilipovic, R. (2018). Detection of pollen bearing honey bees in hive entrance images. *2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–4.
- [Tashakkori et al., 2021] Tashakkori, R., Hamza, A. S., and Crawford, M. B. (2021). Beemon: An iot-based beehive monitoring system. *Computers and Electronics in Agriculture*, 190:106427.
- [Tashakkori et al., 2015] Tashakkori, R., Kae, D., and Parry, R. (2015). Automated beehive surveillance using computer vision. *Conference Proceedings - IEEE SOUTHEASTCON*, 2015.
- [Terenzi et al., 2020] Terenzi, A., Cecchi, S., and Spinsante, S. (2020). On the importance of the sound emitted by honey bee hives. *Veterinary Sciences*, 7(4).
- [Tolstov, 2012] Tolstov, G. (2012). *Fourier Series*. Dover Books on Mathematics. Dover Publications.
- [vanEngelsdorp et al., 2009] vanEngelsdorp, D., Evans, J. D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B. K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpay, D. R., and Pettis, J. S. (2009). Colony collapse disorder: A descriptive study. *PLOS ONE*, 4(8):1–17.
- [von Frisch et al., 1967] von Frisch, K., Chadwick, L., and Seeley, T. (1967). *The Dance Language and Orientation of Bees*. Belknap Press of Harvard University Press.
- [Weber, 2013] Weber, E. (2013). *Apis mellifera: The domestication and spread of european honey bees for agriculture in north america*. *University of Michigan Undergraduate Research Journal*, 9:20–23.
- [Wenner, 1964] Wenner, A. M. (1964). Sound communication in honeybees. *Scientific American*, 210(4):116–125.
- [Yang and Collins, 2018] Yang, C. and Collins, J. (2018). Improvement of honey bee tracking on 2d video with hough transform and kalman filter. *Journal of Signal Processing Systems*, 90.
- [Yang and Collins, 2019] Yang, C. and Collins, J. (2019). Deep learning for pollen sac detection and measurement on honeybee monitoring video. In *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 1–6.
- [Zacepins et al., 2015] Zacepins, A., Brusbardis, V., Meitalovs, J., and Stalidzans, E. (2015). Challenges in the development of precision beekeeping. *Biosystems Engineering*, 130.

- [Zacepins et al., 2018] Zacepins, A., Jelinskis, J., Kvišis, A., Dzenis, M., Komasilovs, V., and Komasilova, O. (2018). Application of lorawan technology in precision beekeeping. *IX International Agricultural Symposium "Agrosym 2018"*.
- [Zacepins and Karasha, 2013] Zacepins, A. and Karasha, T. (2013). Application of temperature measurements for bee colony monitoring: A review. *Engineering for Rural Development*.
- [Zacepins et al., 2016] Zacepins, A., Kvišis, A., Ahrendt, P., Richter, U., Tekin, S., and Durgun, M. (2016). Beekeeping in the future — smart apiary management. In *2016 17th International Carpathian Control Conference (ICCC)*, pages 808–812.
- [Zacepins et al., 2020] Zacepins, A., Kvišis, A., Komasilovs, V., and Muhammad, F. (2020). Monitoring system for remote bee colony state detection. *Baltic Journal of Modern Computing*, 8.
- [Zacepins et al., 2012] Zacepins, A., Stalidzans, E., and Meitalovs, J. (2012). Application of information technologies in precision apiculture. In *Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012)*, Indianapolis, IN, USA.
- [Zgank, 2018] Zgank, A. (2018). Acoustic monitoring and classification of bee swarm activity using mfcc feature extraction and hmm acoustic modeling. In *2018 ELEKTRO*, pages 1–4.
- [Zhang et al., 2021] Zhang, T., Zmyslony, S., Nozdrenkov, S., Smith, M., and Hopkins, B. (2021). Semi-supervised audio representation learning for modeling beehive strengths. *CoRR*, abs/2105.10536.
- [Zhu et al., 2019] Zhu, X., Wen, X., Zhou, S., Xu, X., Zhou, L., and Zhou, B. (2019). The temperature increase at one position in the colony can predict honey bee swarming (*apis cerana*). *Journal of Apicultural Research*, 58(4):489–491.