WE-Bee: Weight Estimator for Beehives Using Deep Learning

Omar Anwar,[1] Adrian Keating,[2] Rachel Cardell-Oliver,[3] Amitava Datta,[3] Gino Putrino [1]

[1] Department of Electrical, Electronic and Computer Engineering,
[2] Department of Mechanical Engineering,
[3] Department of Computer Science and Software Engineering
[1,2,3] The University of Western Australia (UWA), Perth WA 6009, Australia
omar.anwar@research.uwa.edu.au, adrian.keating@uwa.edu.au

Abstract

Abstract: We present *WE-Bee*, a hybrid for soft sensing and time series forecasting, to estimate the daily weight variations of honeybee hives. Weight variations of a honeybee hive are the most important indicator of hive productivity, and the health and strength of a bee colony. Precise measurement of the weight of a hive requires an expensive weighing scale under each hive. On the other hand, sensors deployed inside the hive are cheaper than a weighing scale, and are shielded from the extreme weather variations outside the hive. In this work, honeybee activity is monitored using data from sensors inside the hive, along with monitoring the information related to the seasons, time of the day, external weather and the size of hive. *WE-Bee*'s deep learning algorithm is based on Bidirectional LSTM encoder-decoder and attention mechanism. The results from field deployments demonstrate that the system is capable of cumulative weight estimation over multiple weeks. The results also show that the use of appropriate features and the diversity of training data is essential for robust performance of deep learning in the wild for this application.

This work presents a deep learning approach to estimate the weight variations of a beehive. The weight variations of a hive are a very good indicator of colony strength (Meikle et al. 2008). A model capable of estimating the correct trend of weight change fits the purpose for a majority of beekeepers. For this study, monitoring systems capable of sensing temperature, humidity, atmospheric pressure, CO_2, acoustics, vibrations and weight were designed and developed. Eight sensor systems were deployed in hives at different sites, in varying environmental and weather conditions, to collect a diverse dataset to train and test *WE-Bee*. To the best of our knowledge, this is the first work on estimating the weight variations of a beehive using machine learning. This work uses a combination of different internal hive sensors to gauge the complex activity of honey bees, along with external weather, seasonal, time and size information of hives. This work provides a thorough analysis of the performance of *WE-Bee* in the wild. The highlights of this work are:

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

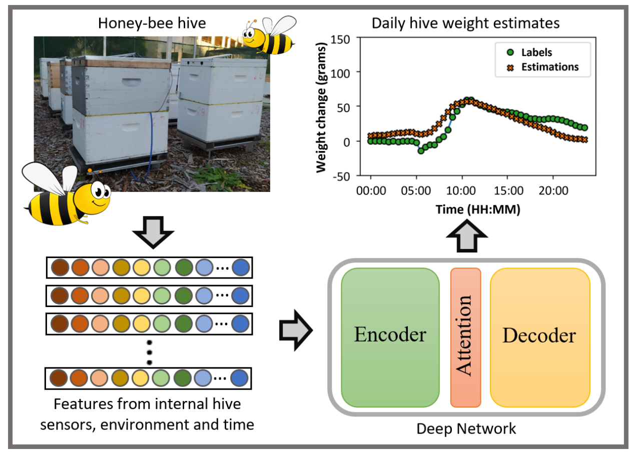

Figure 1: *WE-Bee* uses internal hive sensors, environmental features, season information, time and hive size to estimate the daily weight variations of a beehive per hive frame.

Honey bees play a critical role in pollination, which is vital for one third of global food production. Commercial beekeepers frequently move their hives between fields to provide pollination services. This hive movement is very stressful for bees and can adversely impact the colony strength (number of bees in the hive). Monitoring the colony strength is a genuine concern for the beekeepers, as well as for the farmers who pay for the pollination service. Strong colonies contributes to pollination much better compared to weak ones. A strong bee colony can bring up to 3 kg pollen and nectar to the hive in a single day. Thus, monitoring the weight of a hive provides a very good assessment of bee colony strength/activity, and the contribution of hive towards pollination (Meikle et al. 2008).

Monitoring the weight of a hive comes at a significant cost. Commercial beehive monitoring systems use electronic sensors to collect data from inside the hive, and an external weighing scale to monitor the weight of hive. Usually, the weighing scale is sold as an optional add-on to the monitoring system because of its cost, which is usually more than all the internal sensors put together (Pollenity 2021; Hivemind 2021; Arnia 2021). A majority of beekeepers purchase only the internal sensors and avoid expensive scales. Many factors contribute to the high price of beehive weighing scales. Commercial beehives during peak honey flow in spring can weigh upto 100 kg. The design of these scales should be rigid enough to support this weight, and the electronic sensors should be sensitive enough to pick up small variations of a few grams. These scales are also designed to work in harsh weather conditions, to be able to withstand extreme heat, cold, and rain, which adds to the cost. But despite high costs, their performance in the wild is often below expectations. Furthermore, these scales are often bulky, and have to be setup every time a hive is moved, adding to the setup time and effort required by the beekeepers. Repeated deployments from one field to another also increase the wear and tear of these scales.

Deep learning has shown a lot of promise in forecasting time series data, and in soft sensing for industrial processes. This work uses deep learning to sense daily weight variations of a hive, using time series data from inexpensive sensors. The weight variations of a hive are determined by many factors (but not limited to): number of forager bees, availability of floral resources and their distance from the hive, food consumption rate of the bees and the larvae, environment (temperature, rain, wind), and the evaporation rate of nectar. With so many variables involved, estimating the weight variations of a hive is a difficult, but an important problem to solve. The cost effectiveness of *WE-Bee*, and fit for purpose weight estimation in the wild makes it a very useful tool for the beekeepers in the cost-sensitive market.

Related Work

The weight of a hive is determined by many factors, some of which can be easily monitored using simple sensors, whereas some are very complex. The strength of a honey bee colony is one of the biggest factors which contribute to the weight gain of a hive. The number of foragers (bees which go out in search of food) in a hive is directly proportional to total number of bees in the hive. A strong colony deploys more foragers to find pollen and nectar, and the hive gains weight at a faster rate. The most direct way to count foragers is to use cameras at the hive entrance. This however is not a very cost effective solution, and requires a lot of power and data bandwidth, a luxury remote beehive monitoring systems cannot afford. An indirect way is to estimate the strength of a bee colony, by monitoring the thermoregulation of the colony. A strong colony maintains the appropriate temperature and humidity levels inside the hive (Sudarsan et al. 2012; Al-Ghamdi, Abou-Shaara, and Mohamed 2014; Zeaiter 2019). Hence the variations in temperature and humidity inside the hive compared to the variations outside the hive, provide a very good indication of strength of the bee colony.

The work of (Hambleton 1925) is one of the earliest studies on the effect of weather and environment on beehive weight variations. Honey bees are inactive at night, thus the hive cannot gain weight. However the hive usually loses weight at night because of nectar evaporation and the food consumed by honey bees and larvae. The rate of food consumption depends on the number of bees in the colony, and the temperature. Bees consume more food in lower temperatures and increase their metabolism to keep the hive warm, which increases CO_2 concentration inside the hive. The rate of evaporation from the hive depends on difference between the humidity levels inside and outside the hive.

In early morning, if the temperature outside the hive is appropriate, the forager bees leave the hive in large numbers to check for pollen and nectar availability. This results in a steep drop in the weight of hive, which is referred to as ‘Breakfast Canyon’ (Holst and Meikle 2018). The duration of Breakfast Canyon depends on the time foragers take to return to the hive. If there is plenty of pollen and nectar available closer to the hive, foragers return quickly, otherwise it takes longer for them to return. This availability of foraging resources, and their distance from the hive is a very difficult factor to estimate as it depends on the location of hives, season, weather and types of flora available. The magnitude and the frequency of bee buzz (Michelsen, Kirchner, and Lindauer 1986; Terenzi, Cecchi, and Spinsante 2020), and bee waggle dance vibrations (Grüter and Farina 2009) are however good indicators of the level of foraging activity.

Based on temperature suitability and other environmental conditions, honey bees collect pollen and nectar throughout the day, resulting in the increase of hive weight (McLellan 1977). In hot summer days, bees stop foraging activity when the temperature outside the hive increases in the middle of the day. Some flowers produce nectar only during early hours of the morning or late in the evening, dictating the pattern of hive weight variation. High wind speeds and rain disrupt bee activity and foraging. However rain can result in an increase of the weight of a hive because rain water can accumulate on the top of flat hive surface, and the wooden structure of the hive can absorb moisture resulting in weight gain. The effect of rain depends on the absence/presence/quality of paint on the outer-side of wooden hives. Similarly, exposure to the sun or hot and dry weather can also lead to beehive structure losing moisture and weight. On the other hand, hives made of plastic or polystyrene cannot absorb any moisture, thus contribute very little to the hive weight variations.

The design and structure of a beehive weighing scale itself is a contributor to weight variations. The load sensors, Analog to Digital Converters, and the frame of weighing scale, are exposed to variations in temperature, humidity and other environmental factors, which impact their performance. Research is continuing on better designs to improve the performance of beehive weighing scales (Fitzgerald et al. 2015; Zacepins et al. 2017; Terenzi et al. 2019; Bratek and Dziurdzia 2021). Many commercial beehive monitoring systems are also competing with each other to provide affordable weighing scales. However given the durability and accuracy requirements of design, the cost of commercial beehive weighing scales is still high for a majority of beekeepers, preventing their large-scale deployment.

Soft sensing has been widely used in industrial processes to predict difficult to measure variables, however no attempt has been made to estimate the beehive weight or its variations. One interesting work on humans is (Mengüç et al. 2014), where authors use wearable strain sensors to measure angle of multiple joints in a human body to estimate the human gait. Our work uses the same principle, where easy to use sensors are utilized to sense a difficult to measure quantity. However, the weight of a hive at any point of time is dependent upon the conditions at that time, as well as those in the past. Long Short-Term Memory networks (LSTM) have shown great promise with time series forecasting. Authors in this work (Du et al. 2020) demonstrate the ability of bidirectional LSTMs and Temporal Attention to learn long-term dependencies and correlation features which are hidden. *WE-Bee* is designed as a hybrid model to soft-sense/estimate the time series data of daily beehive weight variations.

Data Collection

To efficiently train any machine learning model, the quality and quantity of training data need to be adequate. The quality of sensor data largely depends upon the sensor system itself. A total of eight sensor systems were designed, developed and deployed at three different sites to collect data. Microcontroller of the sensor system is used to extract sensor features, and transmit them from the remote site over a low bandwidth channel. The carefully designed feature extraction process significantly reduces the size of data, e.g. each audio recording of 2048 samples is reduced to 17 features containing important amplitude and frequency information of bee buzz. Table 1 lists all the features used by *WE-Bee*. The details of data collected using each sensor system are given in Table 2.

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

Table 1: The composition of features used for *WE-Bee*.

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

WE-Bee is designed to estimate the weight variation pattern for an entire day, hence problems with the data for even a few hours on a given day make the data for entire day unreliable. Attempts to use interpolation did not provide adequate results because of complex nature of bee colonies. Also, the beehives need regular inspections to ensure the health of bees. The hives used in this study were inspected every fortnight to make sure that the bees are healthy, and the queen is laying eggs. During these inspections, hive frames were pulled out one by one, with the hive open for up to 30 minutes during each inspection. Occasionally frames were added/removed/swapped during these inspections, which led to a change in the weight of the hive. For these reasons, data from the days of hive inspections was also discarded.

The variation in total days of data collected from each system in Table 2 is a result of different days of deployment, as well as a different number of data days discarded for each system. The data is collected with an interval of 10 minutes, resulting in 144 data points per hive per day. *WE-Bee* is designed using 48 data points per day, with an interval of 30 minutes between consecutive samples, which is adequate to capture important variations in hive weight. We use the 144 data points collected each day to increase the quantity of the training data, by extracting 3 sets of 48 data points from each day in the training set.

Weather has a huge impact on the honey bee activity, and is a significant factor in determining the hive weight variations. The data regarding external temperature, humidity, rain and wind speed was collected using the online reports generated by the Bureau of Meteorology (BOM) (WeatherServices 2021). These reports are generated every 15 minutes for BOM weather stations which are available throughout Australia. We choose the closest weather station to the beehive site for our training data. A significant lack of accuracy was observed in the weather data at Site-C, which is located approximately 48 km away from nearest weather station. Rain was often reported when there was none at the site of hives, which was evident by solar panels charging

Table 2: Break-down of data collected (days) for training and testing of *WE-Bee*. A total of 1200 days of sensor data has been collected from 3 different sites, using 8 sensor systems. Site-B is approximately 170 km north of Site-A, whereas Site-C is further 200 km north of Site-B. System 14 and 15 were deployed to collect data from November 2020, whereas rest of the systems were deployed from March 2021.

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

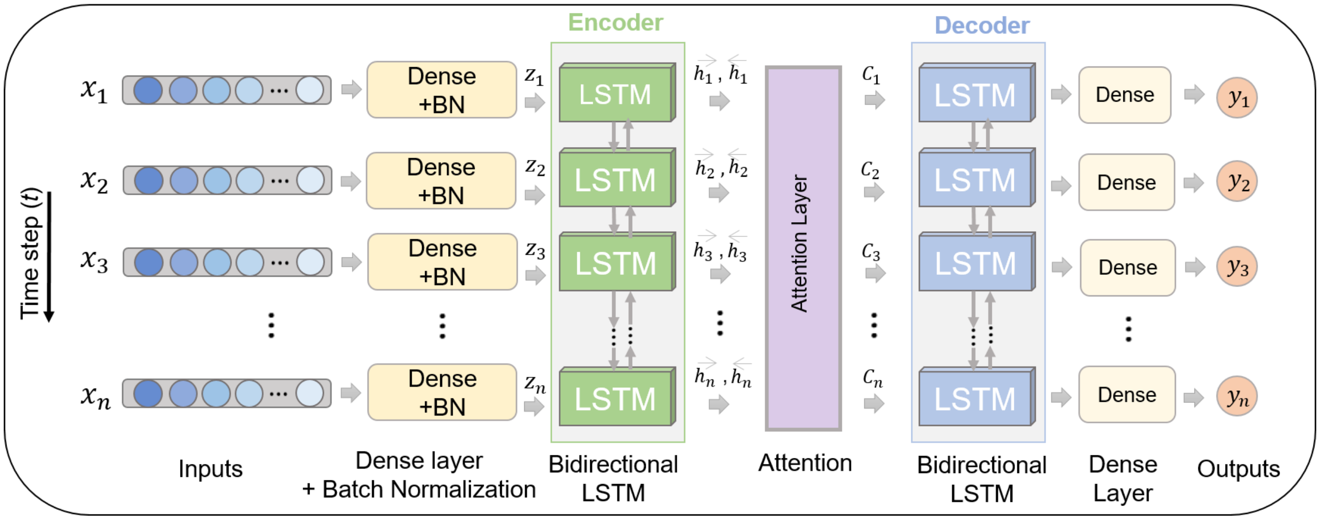

Figure 2: The Network architecture of *WE-Bee*. The input features are represented by x_t, whereas y_t is the output of estimated weight variation for a specific time step t. With a total of 48 samples of data per day, n is 48 in this particular case.

the batteries. At times the beehives experienced rain, which resulted in noisy data on the weighing scales, but was not reported by BOM. Inaccurate weather data, such as in this case makes it difficult to estimate the weight of a beehive.

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

The data collected for this study is from hives of different sizes, however the size of frame used in all these hives is the same. A hive consisting of N chambers with M frames per chamber, will be referred to as a hive of size $N \times M$ frames. This allows the number of frames to be used as a standard measurement for hive size, and the product of N and M (number of total frames in the hive) can be used as the total capacity of the hive. The number of frames in each hive is used to calculate the weight variation of an entire hive based on variations estimated per frame using *WE-Bee*. The baseline for hive weight was obtained by measuring several hives with empty frames and no honey bees. The average weight of empty hives (not part of the dataset) is 1.06 kg per frame, whereas the average weight of hives with pollen/nectar in our dataset is 2.39 kg per frame.

Network and Experimental Setup

The architecture of *WE-Bee* is inspired by (Du et al. 2020), where the authors use multivariate time series forecasting using attention-based encoder–decoder framework. The authors use their network to predict the values of a time series data in future, however we estimate the values of an unknown sensor (weight) for the same time. Previous sections explain how weight of a beehive is dependent on many different factors. *WE-Bee* exploits these dependencies to estimate/predict a series of weight values based on time series data collected from internal hive sensors and relevant information. Input to the network is a set of data collected from internal hive sensors such as temperature, humidity, atmospheric pressure, CO_2, acoustics and vibrations. Information about the weather, week of the year (seasonal information), time of the day, and the size of hive is also part of input. All the inputs are processed to create a feature vector x_t of size 36 (see Table 1) for each time step t, with a total of 48 time steps per day. Figure 2 shows the network architecture of *WE-Bee*, with hyper-parameter settings given in Table 3.

Table 3: Hyper-parameter settings for *WE-Bee*.

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

Daily weight variation estimation of a beehive is a many-to-many sequence-based problem, with both input and output having multiple time-steps. The change in hive weight at any time step, with midnight weight as a reference, is dependent upon all the bee activity and environmental conditions till that time step. The weight itself varies in a pattern and each estimation should properly fit between its neighbors. The use of bidirectional LSTMs (Bi-LSTMs) as our encoder and decoder leverages both past and future contexts within a day. This allows the network to be robust against occasional noisy samples in the input features, and helps with accurate weight estimations. The hidden states of our encoder are attended by the decoder (via an attention layer) to utilise the most important information for transforming (decoding) input features to weight estimation y_t for each time step t.

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

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{1}$$

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

$$e_t = [W \times h_t] + b \tag{2}$$

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

$$a_t = Softmax(e_t) \tag{3}$$

$$C_t = a_t \times h_t \tag{4}$$

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

Results and Analysis

The quantitative results for 5-fold cross-validation for all folds of dataset are shown in Table 4. The test scores of Mean Square Error (MSE) for each fold are reported in grams per frame. The average error for all folds is 13.58 grams, with a standard deviation of 0.8 grams per frame. To make more sense of what these errors represent, the label variations as well as the estimated variations are added with an offset of 2.39 kg, which is the average weight per frame in our dataset. The percentage error between estimated weight and label weight for the frame is then computed for each point in each day in the fold. The percentage Mean Absolute Error (MAE) and the variance of percentage error for each fold are reported in last two columns of Table 4 respectively.

Table 4: Results from 5-Fold cross-validation with a random shuffle of entire dataset. The Mean Square Error (MSE) is reported for 11,520 data-points (240 days × 48 data-points per day) per fold. Mean Absolute Error (MAE) is reported as a percentage of error for an average frame of 2.39 kg. The scatter plot for weight labels and weight estimations at the end of day is shown in Figure 4.

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

Some examples of estimated weight variations per frame by *WE-Bee*, as well as the label weight variations per frame from the test set are shown in Figure 3. A total of 48 estimations are generated for each day, with the weight at midnight (00:00) as starting reference for each day. The daily estimations can be divided into two categories. One where errors for all estimations within a day add to a negligible error by the end of the day, as shown in examples of Figure 3 (a) and (b). The second category is where the accumulated error for all 48 estimations within a day lead to either an over-estimate or under-estimate of weight variation, as shown in examples of Figure 3 (c) and (d) respectively.

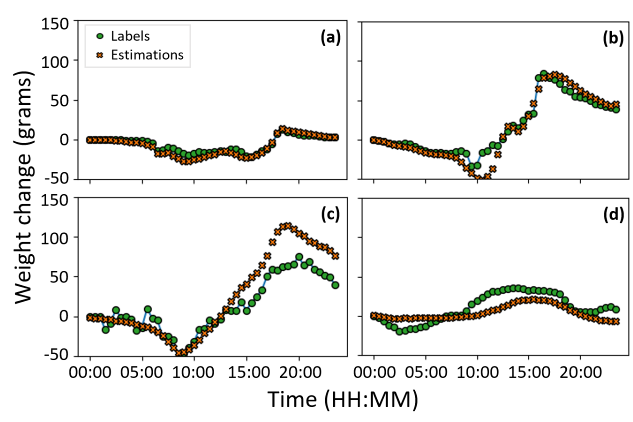

Figure 3: Test-set examples of daily weight variation labels per hive frame, and the estimations for the same. First weight reading for each day at 00:00 is the reference for variations throughout the day. Daily weight variation estimations leading to a negligible error at the end of the day are shown in (a) and (b). An over-estimate of the daily weight on a day with occasional rain can be observed in (c). Example of an under-estimate of the daily weight is shown in (d).

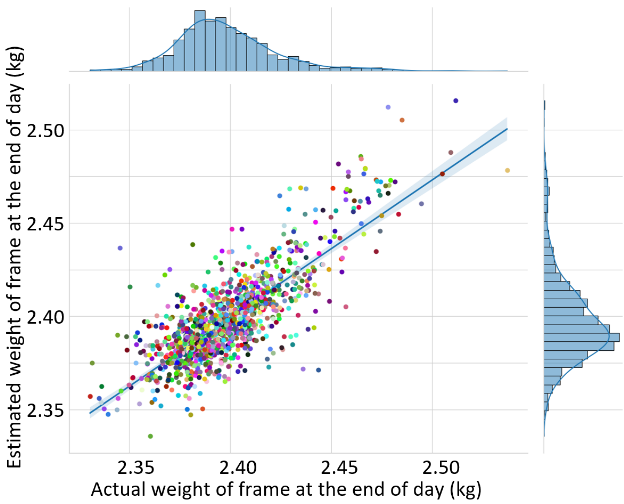

Figure 4: The scatter plot of actual weight of frame at the end of day, against the estimated weight for the same. These results for 1200 days in the dataset are obtained after merging the test results of all 5 folds (Table 4). The Pearson correlation between the two is r = 0.794, with p less than 0.001.

For cumulative estimation of hive weight over multiple days, estimation for each day starts from where the estimation of previous day had ended. So the error in weight estimation at the end of the day (see Figure 3 (c) and (d)), propagates to the weight estimations for next day(s). A biased network with minor but consistent over/under-estimates will lead to a huge error over cumulative estimations. However a network with the Gaussian distribution of errors, will have a smaller accumulated error. Figure 4 shows the scatter plot of actual against estimated weight per frame per day. The network shows slight bias towards over-estimating the weight when the hive loses weight for a given day (cases where actual weight is less than 2.39 kg per frame). However when the hive gains weight by the end of day (cases where actual weight is more than 2.39 kg per frame), the network is slightly biased towards under-estimating the weight. This is a classic example of an unbalanced dataset causing the network to be biased, with a mean absolute error of around 0.5% at 2.35 kg and around 1% at 2.50 kg of actual weight.

Performance on Unseen Sensors/Hives

A randomly shuffled dataset is used for the 5-fold cross-validation. As a result, data collected by each sensor system is available in both the training set and the test set. One of the issues with electronic sensor data is that sensors tend to add a specific bias to collected data, and this bias varies even between the sensors of same type. This bias acts as a signature which deep networks can exploit and overfit for sensor systems in the training set. The performance of deep networks is put to real test when they encounter data from unseen sensors, and hence cannot exploit the sensor bias. To thoroughly test the performance of *WE-Bee*, another set of evaluations was performed. This time the model was trained using data collected by all sensor systems except for one, and the performance was tested on the data from the system which was not used for training. This was repeated for every sensor system, and as sensor systems are allocated to specific hives, we were also able to evaluate the model on hives which were not part of training set.

Table 5: Results from training on multiple hives and testing on an unseen hive. The details of data collected from each system are reported in Table 2.

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

The sensor system specific results are reported in Table 5, where the first column indicates the system ID which was used for testing, but not for the training. A more realistic deviation in these results can be observed, with relatively high MSE and percentage MAE compared to those reported in Table 4. Percentage errors computed for each sensor system were pooled together and Figure 5 shows the histogram of percentage errors between labels and estimations for all the sensor systems combined. However this histogram is only for the errors at the end of each day, representing the propagation errors in percentage for 1200 days. The Gaussian distribution of percentage errors indicates that there is no major bias in estimations.

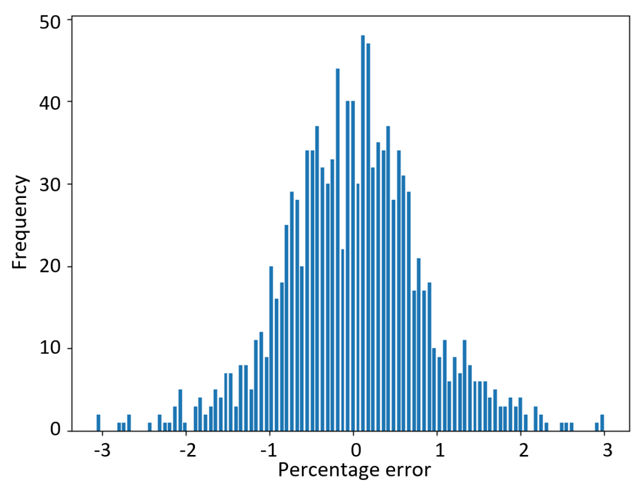

Figure 5: Histogram of percentage errors between final label and final estimation for each day (propagation error) in the entire dataset (1200 days). Results obtained after individually testing each system (Table 5) are pooled together to obtain the error histogram.

Performance on Cumulative Estimation

The cumulative weight estimation capability of *WE-Bee* for unseen data was tested on data collected using sensor system 14. This system was deployed in hives for a total of 200 days, more than any other sensor system used in this study. The test set in this case consists of data collected via system 14 only, and has not been shuffled to preserve the order of days. As a first step, all the frame weight variation labels for sensor system 14 were stitched together. This was followed by converting the frame weight variations into hive weight variations, by multiplying it with the actual number of frames in hive on that day. This process was repeated for the daily frame weight variation estimations generated by *WE-Bee* as well. The starting offset for both the labels and estimations was set using the actual weight of the hive measured on the first day. This provided us with two sequences, one for hive weight labels and other for hive weight estimations as shown in Figure 6.

The sharp increase in the weight around day-5, as shown in Figure 6, is a result of a beekeeper merging two hives together to make a stronger colony, which changed the size of the hive from 2×10 frames to 3×10 frames. The weight estimates are reasonable till around day-30, after which the model over-estimates the daily weights till day-97. The major reason for the over-estimations in this period is the limited training data. System 14 along with system 15 were the only two systems deployed at Site-A. During this period, other systems were not deployed in any hives. With system 14 being used for testing, the network only has data from system 15 available for training for this time of the year. The variations in weight are very season specific, and the lack of training data for this season makes the network under perform.

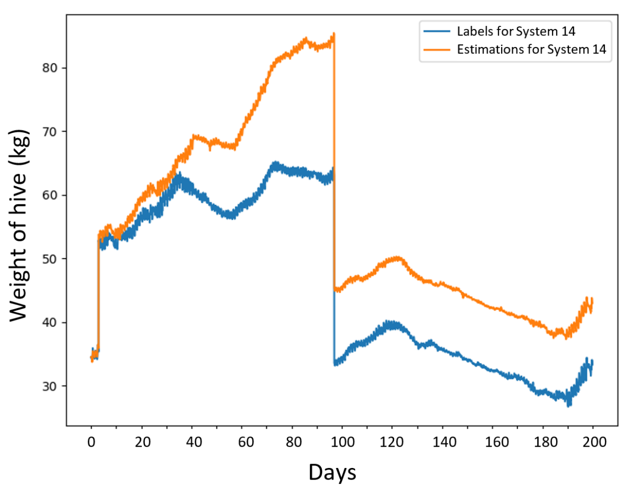

Figure 6: The weight of hive(s) with sensor system 14 estimated for 200 days. Between day-30 and day-97, *WE-Bee* over-estimates the daily weight change, and the error accumulates over time resulting in diverging patterns. System 14 is then deployed at a different hive on day-98, and the size of hive changes from 3×10 frames to 2×8 frames, resulting in a sharp drop in the weight. From day-97, the weight variation estimations are quite accurate till day-200, and only the previously accumulated error can be seen propagating in this period.

After day-97, the system 14 was deployed in new hive of size 2×8 frames at Site-B, where the other seven systems were also deployed to collect data form beehives of varying strengths. With diverse data available for training, the performance of *WE-Bee* improves drastically. The trends of labels and estimations between day-98 and day-200 in Figure 6 are very similar. The error during this period exists because we deliberately did not set a starting reference for estimations at day-98, when sensor system 14 was moved to a new hive. The model uses the last estimated weight from the previous hive with a total of 30 frames (day-97), and the size of the new hive with a total of 16 frames to calculate the new weight reference. In practice, beekeepers either weigh their hives using manual scales, or take a calculated and reasonably accurate guess about the weight of hive during hive inspections. This weight can be fed to the model, and used as a reference to estimate the weight and weight variations of the hive till the next inspection. This will significantly limit the duration and the magnitude of the propagation error. *WE-Bee* correctly estimates the trend of change in weight for more than 14 weeks after day-97, and this trend is often an adequate piece of information for the beekeepers.

Conclusion

This work proposes a hybrid for soft sensing and time series forecasting to estimate the daily weight variations of beehives. The results from Table 5 show an average mean absolute error of 0.64% for estimating a total of 57,600 weight points for 1,200 days in the dataset. These estimations are for the sensor systems and hives which were not part of the training set. This validates the good performance of *WE-Bee* for unseen data. With adequate training data, the cumulative estimation for extended periods also shows promising results. The hives used for the data collection were allowed their natural variations in colony strengths and forager activity, and were moved to sites at a significant distance from each other to collect geographically diverse data. The diversity of the training data played a significant role in the quality of estimations. The features for training were selected after an in-depth study of bee behaviour, and the impact of environment on bee foraging activity. Instead of trying to remove the impact of environmental noise (wind, rain and other external agents) from sensor data, the weather information was used as additional training features to increase the robustness of model against this noise. Rather than using raw sensor data such as audio, we used carefully selected features to achieve robust performance for beehive weight estimation. In future, we plan to explore the contribution of each input feature in estimating the weight variations of hive, while training and testing *WE-Bee* on a bigger dataset. We will also investigate the impact of using more than 48 samples per day on the performance of *WE-Bee*.

Acknowledgments

The authors would like to thank The Cooperative Research Centre for Honey Bee Products and The Department of Jobs, Tourism, Science and Innovation, Govt. of WA for supporting and funding this project. We are also thankful to the beekeepers, especially Tiffane Bates, Graham Brooks, and Peter Slater, who have assisted with our experiments by helping to test the monitoring systems, collect hive sensor data and perform hive health assessments.

References

Al-Ghamdi, A. A.; Abou-Shaara, H. F.; and Mohamed, A. A. 2014. Hatching rates and some characteristics of Yemeni and Carniolan honey bee eggs. *Journal of Entomology and Zoology Studies*, 2(1): 6–10.

Arnia, E. 2021. Arnia. https://www.arnia.co/product-page/arnia-pro-scale. Accessed: 2021-10-21.

Bratek, P.; and Dziurdzia, P. 2021. Energy-Efficient Wireless Weight Sensor for Remote Beehive Monitoring. *Sensors*, 21(18): 6032.

Du, S.; Li, T.; Yang, Y.; and Horng, S.-J. 2020. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing*, 388: 269–279.

Fitzgerald, D. W.; Murphy, F. E.; Wright, W. M.; Whelan, P. M.; and Popovici, E. M. 2015. Design and development of a smart weighing scale for beehive monitoring. In *2015 26th Irish Signals and Systems Conference (ISSC)*, 1–6. IEEE.

Grüter, C.; and Farina, W. M. 2009. The honeybee waggle dance: can we follow the steps? *Trends in Ecology & Evolution*, 24(5): 242–247.

Hambleton, J. I. 1925. *The effect of weather upon the change in weight of a colony of bees during the honey flow*. 1339. US Department of Agriculture.

Hivemind, E. 2021. Hivemind. https://modernbeekeeping.com.au/shop/hivemind-scale/. Accessed: 2021-10-21.

Holst, N.; and Meikle, W. G. 2018. Breakfast canyon discovered in honeybee hive weight curves. *Insects*, 9(4): 176.

McLellan, A. 1977. Honeybee colony weight as an index of honey production and nectar flow: a critical evaluation. *Journal of Applied Ecology*, 401–408.

Meikle, W. G.; Rector, B. G.; Mercadier, G.; and Holst, N. 2008. Within-day variation in continuous hive weight data as a measure of honey bee colony activity. *Apidologie*, 39(6): 694–707.

Mengüç, Y.; Park, Y.-L.; Pei, H.; Vogt, D.; Aubin, P. M.; Winchell, E.; Fluke, L.; Stirling, L.; Wood, R. J.; and Walsh, C. J. 2014. Wearable soft sensing suit for human gait measurement. *The International Journal of Robotics Research*, 33(14): 1748–1764.

Michelsen, A.; Kirchner, W. H.; and Lindauer, M. 1986. Sound and vibrational signals in the dance language of the honeybee, Apis mellifera. *Behavioral ecology and sociobiology*, 18(3): 207–212.

Pollenity, E. 2021. Pollenity. https://pollenity.com/shop/. Accessed: 2021-10-21.

Sudarsan, R.; Thompson, C.; Kevan, P. G.; and Eberl, H. J. 2012. Flow currents and ventilation in Langstroth beehives due to brood thermoregulation efforts of honeybees. *Journal of theoretical biology*, 295: 168–193.

Terenzi, A.; Cecchi, S.; and Spinsante, S. 2020. On the importance of the sound emitted by honey bee hives. *Veterinary Sciences*, 7(4): 168.

Terenzi, A.; Cecchi, S.; Spinsante, S.; Orcioni, S.; and Piazza, F. 2019. Real-time System Implementation for Bee Hives Weight Measurement. In *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, 231–236. IEEE.

WeatherServices. 2021. BOM. http://www.bom.gov.au/catalogue/data-feeds.shtml. Accessed: 2021-10-21.

Zacepins, A.; Pecka, A.; Osadcuks, V.; Kviesis, A.; Engel, S.; et al. 2017. Solution for automated bee colony weight monitoring. *Agron Res*, 15(2): 585–593.

Zeaiter, Z. 2019. *The Effect of Thermoregulatory Stress on Honey Bee Hives*. Master's thesis, University of Sydney.