A Machine Learning Integrated 5.8 GHz Continuous Wave Radar for Honeybee Monitoring and Behavior Classification

Nawaf Aldabashi, *Student Member, IEEE* Sam M. Williams, Amira Eltokhy, Edward Palmer, Paul Cross, Cristiano Palego, *Senior Member, IEEE*

Abstract- A 5.8 GHz Continuous Wave (CW) radar was developed and integrated on a compact printed circuit board (PCB) for near-hive monitoring of honeybees. It supported noninvasive detection of free-flying honeybees at 2 m and micro-Doppler extraction of bee wingbeat signatures at a closer range using both a double balanced and an In-phase/Quadrature (IQ) mixer. An original technique combining full-wave simulations and Doppler-radar monitoring of pendulum motion was used for: radar calibration through spherical targets of different material and size; precise extraction of bee radar cross section (RCS) at 5.8 GHz; estimation of detection range enhancement through partial silver coating of targets. Finally, the CW radar was integrated with machine learning (ML) to allow automated classification of incoming, outgoing and hovering honeybees. Different ML approaches were tested, where the highest accuracy of 93.37% was found in ternary classification via Support Vector Machine acting on Line Spectral Frequencies.

Index Terms— Continuous wave radar, doppler radar, machine learning, micro-doppler

I. INTRODUCTION

HE decrease of insect biomass has impacted global industries and agriculture [1], [2] as insects are relied upon for pollination, nutrients cycling and detrivory functions [3]–[6]. The significant annual contribution of insects to ecosystem services is estimated to range between \$235 to \$557 billion in value [7]. Among such insects are bees (*Apidae*) – known as the most important pollinating insects – which include the European honeybee (*Apis mellifera*) and the bufftailed bumblebee (*Bombus terrestris*) [8]. Approximately 80% of global pollination services are attributed to honeybees, making them a key component of the ecosystem [9]–[11].

Recent findings show that the combinational exposure of stressors inflict detrimental effects on colony health, hive performance and their population level [12]. For example, honeybee absconding, a significant cause of colony loss in pollinators, has been linked to intra-colony food pattern changes and foraging stresses [13]. The development of monitoring technologies for real-time assessment of insect

This paper is an expanded version from the IMS, Atlanta, GA, June, 2021. This paper was submitted for review on the 26/02/22. This work was supported by the Knowledge Economy Skills Scholarships (KESS 2, Ref: BUK298). N. Aldabashi, S. M. Williams, P. Cross and C. Palego are with Bangor University, Bangor, LL57 1UT UK (e-mail: c.palego@bangor.uk).

A.Eltokhy, was with Bangor University, Bangor, LL571UT UK. She is now with Rapid Bio-Labs, Talinn, 11314 Estonia (e-mail: a.eltokhy@rapidbiolabs.com).

activity is a key element in studies to further increase our understanding [14]–[16] and manage populations of Apidae.

Multiple sensors can be used for real-time probing of hive internal or in situ parameters, such as temperature, humidity, weight, and activity [17]. Radio telemetry, harmonic radars, and RFID tags have also enabled insect tracking and potentially support statistical analysis of their behavior over the entire forage range [16], [18], [19]. Nevertheless, there has been concern that tagging of Apidae for tracking purposes hinders natural movement and considerably affect insect response.

A generally adopted guideline prescribes that the tag's weight should not exceed the maximum nectar and pollen load which typically achieves 59% of a honeybee's body weight [20], but can reach 80% [20]. Thus, even placing tags as light as 30-50% of bee bodyweight can potentially alter their take-off ability, foraging inclination and overall performance [18], [21], [22].

Moreover, bee catching, optimum tag positioning and attachment are inherently stressful and time consuming processes in tag-based tacking approaches.

Tag-free CW radar systems have been designed ranging from 5.8 GHz to 24 GHz [23]–[27] and basing on Doppler shift correlation with bee's speed and acceleration. However, such systems were relatively invasive in that the radar was placed at the hive's entrance.

A primary aim for this work has been to develop a comparatively unobtrusive 5.8 GHz CW radar to monitor free-flying honeybees *at a 2 m range from the hive* and facing its entrance/exit. The 5.8 GHz frequency band was chosen as a compromise between signal quality, monitoring range, and the wide availability of off-shelf components. This allowed expanding on our original work [27] by integrating the CW radar into the single portable PCB in Fig. 1(b). The final aim of this work has been to create an autonomous machine learning based radar for long term data collection supporting continuous hive health monitoring.

Edward Palmer is with S&A Produce UK Ltd, Hereford, HR1 3ET, UK (e-mail: edward.palmer@sagroup.co.uk).



(c)

Fig. 1. (a) Block diagram of the 5.8 GHz CW Radar and (b) PCB implementation. (c) Setup for the hive monitoring experiment.

II. 5.8 GHZ CW RADAR DESIGN

The setup of the 5.8 GHz CW radar is shown in Fig.1 (c). An important requirement that dictates the design of the CW radar is the RCS of the target. The RCS of honeybees has been experimentally investigated at different frequency bands between 5.8 GHz and 24 GHz [23]–[27]. At 9.4 GHz, the RCS of honeybees ranged from -40 to -45 dBsm [24], whereas at 24 GHz, free-flying honeybees RCS averaged at -50 dBsm for 11 bees [25]. Wolf *et al.* mentioned honeybee RCS at 8-12 GHz to average between -65.9 dBsm and -61.5 dBsm for horizontal and vertical polarization, respectively [26]. Finally, at 5.8 GHz, the RCS of honeybees averaged between -55 to -60 dBsm [27].

The 5.8 GHz CW radar was designed with such parameters in mind, while factoring in a 2-3 m range from the hive. It was based upon inexpensive commercial Radio Frequency (RF) modules powered by a LiPo battery with Switch Mode Power Supplies (SMPS) for voltage regulation. This increased monitoring flexibility and allowed identification of a wider range of behaviors (e.g. hovering, flight patterns, and missing the entrance) than mere entry/exit.

The IF signal of the double balanced mixer (DBM) was amplified by a 60 dB custom-built Variable Gain Amplifier (VGA) with 100 dB CMRR and integrated a Low Pass Filter covering our frequency range of interest up to 408 Hz. The VGA's output was recorded in 16-bit wave formats, fed to a laptop using an external USB sound card sampled at 44.1 KHz. The recorded wave files were processed on a laptop equipped with MATLAB [28]. The 5.8 GHz CW radar parameters are summarized in Table 1.

Circuit optimization enabled integrating the functional system in [27] into the 45 mm \times 40 mm 4-layer PCB in Fig. 1(b) with an approximate cost of \$75 and same specifications as the original system.

While millimeter-wave modules afford superior radar resolution, the sensitivity-cost tradeoff achievable at 5.8 GHz is critical in greenhouse scenarios requiring multiple receiver deployment.

TABLE 1
5.8 GHz RADAR PARAMETERS

Parameter	Parameter Value
English	5 8 CH2

Frequency	5.8 GHz	
Peak transmitted power	28 dBm	
Radar range	2 meter	
Transmitter gain	12 dBi	
Transmitter beamwidth	42° H and V	
Receiver gain	17 dBi	
Receiver beamwidth	32° H and 76° V	
RF Receiver gain	38.40 dB	
Receiver noise figure	2.22 dB	

III. PREDICTED VS MEASURED RADAR SIGNATURES

A. Doppler Signatures: DBM Receiver

Although bee flight may involve composite/heterogeneous motion patterns, identification of uniform motion segments was a valuable abstraction to guide further analysis. Simplified Doppler shift models where first analytically derived by considering: ideal point scatterers; elementary motion segments with either uniform speed, uniform acceleration, or sinusoidally varying acceleration (pendulum motion); representative bee range, speed and acceleration as extracted from previous experiments [27].

MATLAB suites [28], [29] were then used for accurate spectrogram calculation accounting for: target range and angular deviation; spherical targets that matched the honeybee's extracted RCS at 5.8 GHz (see section IV); power spectral density variation from path loss and TX/RX antenna directivity.

The availability of measurement data for either real honeybees or spherical calibration targets, resulted in the spectrogram pairs visible in Fig. 2(a)-(b) through (e)-(f) for each elementary motion segment.

Notice that every modeled motion modulates the backscattered signal phase as a function of the motion basic parameters and transmitted signal wavelength λ . Since the models presented in (1)-(4) express Doppler frequency shift as a proxy for target radial speed $v_R(t)$, they effectively describe demodulation of frequency modulated waveforms with $\beta=2/\lambda$ modulation index and $v_R(t)$ modulating signal.



Fig. 2. Doppler signatures for target motion across elementary segments (a), (c), (e) as predicted from CW radar theory + MATLAB models and for measured honeybee (b), (d) or (f) metal sphere target. (g) Measured near-hive bee signatures reproducing some of the modeled elementary motions.

Thus, from CW radar theory, the expected Doppler shift signature for a target moving at uniform radial velocity v_R and a DBM receiver architecture is of the type:

$$f_{d1}(t) = \frac{2v_R}{\lambda} = \frac{2f_0 v_R}{c} \tag{1}$$

where λ is dictated by the transmitter signal frequency f_0 and the speed of light *c*, resulting in the predicted spectrogram in Fig. 2(a). The predicted Doppler signature for a uniform accelerated motion with radial acceleration a_R is:

$$f_{d2}(t) = \frac{2a_R t}{\lambda} \tag{2}$$

leading to theoretical and measured signatures in 2(c) and (d), respectively. The spectrograms in Fig. 2(b) (2(d)) represent rare examples of bee free flights at constant speed (acceleration) in front of the radar over appreciable time spans. The predicted Doppler shift for a simple pendulum is, neglecting friction over small angle swings:

$$f_{d3}(t) = \frac{2l\dot{\theta}}{\lambda}\cos\theta \tag{3}$$

where *l* is the pendulum length and θ is the swing angle coordinate. While not directly applied in bee flight analysis, such motion type was key to radar calibration and RCS extraction (section IV). It also supported the analysis of more articulate periodical motion patterns which were involved in bee hovering behavior. One such pattern is circular motion which was experimentally characterized by forcing pendulum circular swings at some distance from the observer. For a target rotating at angular speed ω_0 over a circumference of radius *r* centered at distance R_0 and in the same plane as the radar, the predicted Doppler signal is:

$$f_{d4}(t) = 2\omega_0 r \frac{\sin\left(\theta + \tan^{-1}\frac{\sin\theta}{R_0 - \cos\theta}\right)}{\lambda}$$
(4)

The resulting simulated and measured signatures also appear in Fig. 2(e) and (f), respectively, which outlines the resemblance with a simple pendulum motion.

Field experiments demonstrated the radar's capability to detect free-flying honeybees from a 2 m range and for motion segments comparable to the idealized models. Representative instances of flight detection are shown in Fig. 2(g). A leaving flight is shown to result in a red shift of 254 Hz corresponding to a speed of 6.46 m/s (23.25 km/h), which slightly deviated from the linear trend described by (1) due to saturation of the leaving speed. Similarly, a returning flight signature outlined blue shift from the initial 130 Hz value, corresponding to a speed of 3.35m/s (12.09 km/h), to the final 60 Hz value, corresponding to 1.55 m/s (5.52 km/h). The returning bee signal also featured a higher power spectral density indicating closer proximity to the hive.

Notice that the initial linear deceleration trend was followed by bee speed oscillation before landing. The stationary, winding, or periodical aspect in similar segments was visually confirmed through the camera data and generally branded as "hovering". An even more prominent instance of hovering appears in the Fig. 2(g) inset, evoking a spiraling down circular motion signature. Hovering was frequently observed both during takeoff and landing events, possibly in response to a busy hive entrance. Bees were also observed briefly flying out to instantly return to the hive's entrance, resulting in net zero in/out logging. This spurred the hovering classification effort described in section V.B.



Fig. 3. (a) Predicted micro-Doppler signal for bumblebee approaching the radar at 0.05 m/s constant speed and using IQ receivers. (b) micro-Doppler signal recorded though IQ mixer for mostly fluttering bumble bee motion in a transparent container.

B. Micro-Doppler Signatures: IQ Receiver

Micro-Doppler analysis was additionally afforded to disentangle minor motion signatures, such as bee thorax vibration and wingbeat, from gross body translation. This enabled identification of different insect species (e.g. honeybee vs bumblebee), while potentially supporting disambiguation of more complex patterns with similar Doppler signals. In order to gain approaching and receding micro-motion discrimination, a micro-Doppler module was also developed to include a quadrature receiver.

Micro-Doppler signatures of uniform speed flights were well approximated through linear scatterer combination of: constant speed body translation as ruled by (1) and harmonic (spring) motion to additionally model wing flapping:

$$f_{d5}(t) = \frac{2}{\lambda} [v_R + A_w \omega_w \cos \omega_w t]$$
(5)

where A_w and ω_w are the wingbeat amplitude (~1 mm) and angular speed (~2 π ·200 Hz), respectively. The ratio of honeybee wings to main body RCS used for Micro-Doppler signature prediction was of the order of 1:5 and both motions were radially directed in the approximate model. More accurate spectrogram calculation was obtained by modeling bee body as a 15 mm long 3 mm wide ellipsoid and wing length of 9.7 mm while also allowing both positive and negative Doppler shifts for IQ mixing response.

Calculations were specifically carried out for a bumblebee, flapping wings at lower frequency than typical honeybee (~175 vs 220 Hz [30], [31]), and approaching the IQ receiver at low body speed of v_R =0.05 m/s. Overlapping to uniform speed translation in (1) of the harmonic motion in (5) was more conveniently analyzed through phase modulation (PM) formalism. Hence, a quadrature-carrier description of the received micro-Doppler signal had the form:

$$x_r(t) = A[\cos(\beta \cos \omega_w t) - \sin(\cos(\beta \cos \omega_w t))] \quad (6)$$

where *A* is an RCS dependent amplitude term and $\beta = A_w \omega_w \sim 2$ for $A_w \sim 1$ cm is the modulation index, thus indicating departure from a narrowband tone modulation (β <1) scenario. Therefore, standard PM analysis [32] yielded:

$$\begin{aligned} x_{r}(t) &= A \sum_{\substack{B = -\infty \\ E ven}}^{\infty} J_{n}(\beta) \left(-1\right)^{\frac{n}{2}} \cos\left(2\pi (f_{B} + nf_{w})\right) + \\ &-A \sum_{\substack{B = -\infty \\ Odd}}^{\infty} J_{|n|}(\beta) \left(-1\right)^{\frac{n-1}{2}} \sin\left(2\pi (f_{B} + nf_{w})\right). \end{aligned}$$
(7)

where f_B , is the frequency shift for the main body translation, and $J_n(\beta)$ are the coefficient for the Bessel function of the 1st kind. (7) only slightly deviate from standard PM formulas, since (5) forces a $\beta \cos \omega_w t$, rather than the conventional $\beta \sin \omega_w t$ dependence, and are seen to result in spectral lines at $nf_w \pm f_B$ in Fig 3(a). Slow body speed also forces $f_B << f_w$ and $f_w \pm f_B - f_w$, unlike $f_B >> f_w$ in narrowband tone modulation, which underpin the following experimental scenario and allow direct wingbeat frequency extraction.

For experimental validation a bumblebee was placed in a transparent container allowing restrained flight range along with simultaneous video recording of both the insect motion and the radar output. Video segments were selected for comparison with the radar readouts where the insect was either dashing and rubbing its limbs against the box walls without flapping its wings, or fluttering/flying through wing flapping. The corresponding radar recorded files were processed as an STFT to extract micro-Doppler signatures with a window length of 256 samples out of the 44.1K samples and an overlap of 250. Fig. 3(b) represents the recorded micro-Doppler signature and clearly outlines the expected features from theoretical calculations.

The spectral power density mainly concentrated around the expected 2 Hz component corresponding to body fluttering at 0.05 m/s. Although this was somewhat cluttered by coexistence of additional motion types (e.g. rubbing of limbs against box walls) fluttering was captured in the video footage and precisely timed by audio recording of "buzzing bursts".

Weaker but clear horizontal bands at $\pm(nf_w \pm v_R)$ for n=1,2 confirmed coexistence and mixing of the translation and harmonic motions components, as from the model in (5)-(7). Furthermore, the extracted wingbeat frequency $f_w=175$ Hz matched the typical bumblebee range [30].

This confirmed the approach applicability to classify insect species, or even micro-Doppler profiling of individuals within a species. However, due to the late adoption of the IQ mixing approach, the ML analysis in section V could be only supported through double balanced mixing.

C. Classification of More Complex Free-Flight Patterns

While the elementary motion segments described so far assist identification and logging of near-hive activities, freeflying bee patterns are markedly different because: uniform motion patterns represent a minority of all recorded motions; direction and motion type changes occur frequently and at unpredictable rate; individual bee flights and their radar image are heavily influenced by other insect flights within the detection cone; non-flying non-targets such as bees crawling (and still flapping wings, or buzzing) on the hive walls in the detection cone result in significant clutter; in outdoor experiments wind is also a source of noise forcing flight path alteration, hive and radar setup shaking and EM background (e.g. grass, bushes, branches) fluctuation.

The specificity of honeybee behavior unpredictability and interference from collective dynamics add to the general challenge in radar detection, for example from warping of the received signal due to the antenna radiation pattern. As such, extraction of microwave features, from a traditional signal processing standpoint, is a challenging task in most outdoor bee detection scenarios. This explains why the ML algorithms described in section V focused on behavioral classifications more than, or independently of, direct microwave features extraction.

Building a ML model to classify simple and separable away/towards, and circular motion segments as in section II.A results in accuracy in excess of 95%. The availability of analytical models to tune upon enables extraction of Doppler radar features such as effective speed, and periodical features with accuracy between 85% (when circular motion is included) and 97% (without circular motion). However, the key point is that ML is largely unnecessary for such "well-behaved" flight instances where direct spectrogram readout allows efficient feature extraction.

Pursuing too strict a comparison between traditional signal processing and ML-extracted features might be misleading and overlook potential opportunities in new approaches. ML has the capacity to enable correct behavioral classification even when microwave features extraction, whether through traditional processing or ML, is impossible. This is largely due to the fact that confirmation from video footage and expert beekeeper insight is used for algorithm training even for instances where distinct motion types produce virtually identical spectrograms (e.g. radial vs circular oscillations in section II.A).

Assessing where ML-feature extraction breaks down due to overlapping of multiple signals or sheer clutter is a compelling research question. The topic appears however conveniently explorable in a more forgiving scenario than outdoor free-flying bee detection.

IV. RCS AND RANGE INCREASE USING SILVER COATING

Additional investigations were directed at increasing the honeybee's RCS as a means to enhance detection range without drastic hardware changes. While higher RCS could be achieved at 10.5 GHz, or 24 GHz [25], frequency increase was ruled out to maintain the low-cost and commercial availability of the present components, along with coherence with earlier pollinators telemetry systems [15], [16], [27].

The RCS for spherical targets of different material were first simulated as a function of the target diameter using CST Microwave studio. The simulated RCSs are shown in Fig. 4(a) for steel spheres and wooden spheres. The simulated RCS for a full-wave model of a honeybee is reported in Fig. 4(a). The honeybee's model length was varied between 10 and 15 mm. Interestingly, the bee model matched both the amplitude and slope of the 4mm steel sphere RCS vs size curve. The 4 mm water sphere dispersive model (DC ε '=78, σ =1.59 S/m) in Fig. 4(a) was also extrapolated to match the RCS calculation at 10.5 and 24 GHz in [25].

Although the simulated RCS of the 4mm wood sphere was -83.9 dBsm, which is below the radar detection threshold (-80 dBsm), coating the wood sphere with a ~100 μ m silver layer (σ =6.30×10⁷ S/m) was predicted to increase the RCS by an average of 11.7 dBsm and to a detectable -72.06 dBsm.

The $\sim 100 \ \mu m$ coating was obtained through a small brush stroke and validated by profilometer measurements to guide later application of silver nanoparticle layers.

The simulation results in Fig. 4(a) also show that while the coated wood sphere achieved higher RCS compared to the uncoated wood sphere, it featured a decisively lower RCS than the 4 mm steel or water sphere.

Experimental validation of the RCS increase through silver coating, was achieved through IQ radar detection of target oscillations in a pendulum setup [27]. Targets were suspended through a nonmetallic support and a thin 15 cm thread placed 0.5 m from the radar. Oscillations of 30° were allowed perpendicularly to the radar antenna's main beam axis and alternatively suspending: no target; a 4 mm wood sphere; a 4 mm coated wood sphere; a 4 mm steel sphere. The results shown in Fig. 4(b) indicate that the uncoated wood sphere was poorly detected as its RCS was below the minimum detectable signal, resulting in a barely distinct signature from the unloaded wire case.

By contrast, a clear oscillating behavior was detected for the coated wood sphere. Finally, the steel sphere showed an even stronger signature in accordance with the higher RCS, confirming the simulated trend shown in Fig. 4(a).

Silver coating was expected to enhance bee detection range in accordance with the modeled RCS increase, also through no hardware modifications. The predicted radar detection range for uncoated and coated targets is reported in Fig.5 (a). The present study informed a theoretical and experimental cost-benefit analysis of using coating to increase bee radar detection range. Notice the uncoated bee detection is within approximation of the 4 mm steel sphere, which agrees with [25], [27].



Fig. 4. (a) Simulated RCS values of steel, water, wooded spheres and a bee. (b) Signal amplitude increasing as pendulum target increases in RCS. The four wave files were combined to demonstrate the increase in amplitude.

A coated bee model was developed as suggested in Fig. 4(a) and used to plot the curves visible in Fig. 4(a) and 5(a). Such model aimed at maximizing the bee's coated surface while not impairing its flight capacity nor obstructing the breathing ability, by avoiding coating of bee wings and spiracles.

Feasibility of applying silver coating through a single small brush stroke was experimentally tested using (dead) specimens. The Scanning Electron Microscope (SEM) picture shown in Fig. 5(b) outlines a honeybee's layer of coated hair (*seta*) with length ranging from 300 to 900 μ m. The bee's surface available for coating favors distribution of silver coating in a closely packed arrangement that could approach the performance of the coating layers used for the RCS study. As shown in Fig. 4(a) and 5(a) the RCS and detection range improvement for a partially coated honeybee could achieve 18 dBsm and ~4 m, respectively. Such enhancement seems potentially interesting to dynamic tracking applications, drastically reducing tag load along with the challenges in higher frequency modules.

Nevertheless, for near-hive monitoring via stationary transceivers, the ~ 2 m range achieved while avoiding the invasiveness/inconvenience of the coating process might be preferable.



Fig. 5. (a) Increased detection range of both coated wooden sphere and coated bee (displayed in the top right corner). (b) SEM image of honeybee thorax displaying the coated honeybee seta.

V. MACHINE LEARNING

The volume of data collected using the DBM were suitable for ML. Many data quantification techniques are available for audio signals derived from the RF Doppler signature, such as Line Spectral Frequencies (LSFs), Linear Predictive Codes (LPCs), and Mel-Frequency Cepstral Coefficients (MFCCs) [33]–[35]. These were chosen to specifically address the lack of directional information in the DBM data.

Simple temporal features of the signal were additional candidates for ML including mean amplitude, RMS, zerocrossing rate, short-time energy, spectral centroid, kurtosis, skew, standard deviation, mean, variance, and energy. These were combined as the temporal data.

A. Binary Classification

Initial concern was the classification of the signals into two separate classes of bees entering the hive and bees leaving. This was not as simple as looking at the direction (toward/away from the radar) of flight as insects could move in a completely free environment irrespective of the radar. Our previous work looked at using spectrograms of the signatures as a means of classifying the radar signals, building from the MobileNet V2 architecture [36]. This achieved 88.7% accuracy but was a computationally

expensive approach. In addition, it was shown that such Neural Networks (NN) maladapted to this data due to their training on real-world objects rather than radar signatures. Spectrograms are a two-dimensional representation of a onedimensional signal, whereas image-processing NN are designed primarily to classify three dimensional objects represented in two dimensions.

To retrain such a network requires a large quantity of data on par with MobileNet's original work. Audio (rather than spectrogram) processing was a feasible alternative as all the 5.8 GHz output radar signals contained relevant information at sub 1 KHz frequencies. Audio files were subdivided into 0.4 s segments, with each being vetted for a minimum signal.

The window of 0.4 s was deemed a suitable tradeoff in view of the system application as a real-time monitor. This window matched our initial observation of the smallest complete event in the dataset. Subdividing this event further would risk failing to supply complete information to the ML. Larger windows did not offer sufficient improvements in accuracy to combat the additional cost in terms of samples. For rapid in-and-out signals the window of 0.4 s emerged as a threshold between clear separation and overlapped, unusable samples. Some windows without sufficient information, such as tail ends of a signal (representing less than 5% of the length of the sample), were discarded. This created approximately 700 bee-in to 600 bee-out signals.

To generate more data, augmentation approaches were investigated. These included adding artificial noise, timeshifting, and pitch shifting. The sensitive, RF nature of our data meant that noise and pitch alteration could not be enacted without compromising the ability for predictions, thus time-shifting was used.

As each signal was 0.4 s long, shifting could be affected at 0.1, 0.2, and 0.3 s, allowing for a fourfold volume increase, in addition to balancing the dataset. Support Vector Machine (SVM), Random Forest and NN were learning algorithms to generate the augmented data set [37]–[39]. A standard traintest split ratio of 4:1 was used, with the NN taking an additional 10% of the training data as a validation set.

Data quantized by the above methodologies were much smaller than spectrogram images and therefore did not need the raw predictive power associated with MobileNet's capabilities. A smaller sequential model was chosen, using two densely connected layers of 32 neurons each, activated by a SELU function. Although RELU was first used as an activation function, we noted that SELU performed slightly better across all datasets. This difference was small (<2% accuracy) but significant enough to warrant the change. The final layer had a dropout rate of 0.5 to minimize risk of overfitting from the smaller than ideal number of samples.

The model itself was optimized using the Adam algorithm due to the number of parameters (~1400 in the case of the MFCC/BFCC approach) [40]. Larger networks were also tried, both increasing the layer and neuron count, but it was found that the network quickly reached a point of overfitting even when increasing the dropout rate significantly. For comparison, the random forest and SVM used runtime hyperparameter tuning. For each data quantization method, the hyperparameters were chosen via Bayesian optimization, which was done on four-fold cross-validation with each cross-validator running five times. The random forest used values from the following possible hyperparameters:

- Estimator count between 100 and 1000, in increments of 100.
- Split criterion of either Gini impurities or entropy for information gain.
- The number of features to consider for a split being either the square root or log₂ of the total number of features.

Additionally, the SVM hyperparameters were from the following possibilities:

- A regularization parameter between 1e-6 and 100.
- Kernel coefficient of 1e-6 to 100.
- Polynomial kernel function degree between 1 and 5.
- A kernel choice of either linear, polynomial, radial basis function, or sigmoid.

Initial experiments showed that the MFCC and LSF approaches were the strongest predictors with both achieving approximately 85% accuracy. Mel-Frequency in MFCC refers to the melodic scale used to attune audio to match how it is processed by the human ear. The simplest form of this is expressed in (1) where F_m is the Mel-Frequency of natural frequency F. The constant (C) is 2595 and the denominator (D) is 700 in the original equation.

$$F_m = C * \log_{10}(1 + \frac{F}{D})$$
(8)

Though this gave relatively reliable results, it has little to do with MFCCs original use of approximating human ear perceived frequency. Our data was from the radar and, while stored as audio, did not represent physical sound. The two primary parameters of the equation (the constant and the denominator) allowed for tuning and subsequent impact assessment on the final algorithm precision.

The parameters were changed to 1250 for C and 175 for D as a test point. These were chosen to maintain the new curve close to the original MFCC curve but emphasized frequencies below 500 Hz, where most activity was observed. The resulting estimate is demonstrated in Fig. 6(a)and resulted in a 2% accuracy improvement for the Random Forest model. At each point, the random forest was retrained and tested on the data five times, with each iteration having a bootstrapped randomized copy of the data. This ensured that only a mean accuracy was used, and the results would be replicable. The generated grid gave a clear indication of the optimal parameters to be used as demonstrated in Fig. 6(b), though it is noted that this figure has been smoothed via a gaussian filter to allow improved parsing. The results suggest that the best parameters in our case were a constant (C) of 2325 and a denominator of 260 (D).

When these parameters were returned into the equation for testing it showed an increased accuracy of 91.1%, an improvement even on the more computationally expensive spectrogram approach. The hyperparameters found to achieve these results were 900 estimators, entropy as a split criterion, and the square root of total feature count being used as a baseline when looking for the best split. A full breakdown of results, including all algorithms used alongside each quantization method are present in Table 2. The strength of this Bee-Frequency Cepstral Coefficient (BFCC) technique is that the original experiment can now serve as a testbed for when the dataset size is increased. Sample points can be taken from the original, computationally expensive, modeling and compared to later results to test for deviation, which would indicate that the constant and denominator need further refinement. This can be achieved without remodeling the parameters entirely.

B. Ternary Classification

Adding hovering to entry and exit movement was deemed a critical improvement to the system. Hovering is defined as all behavior where the bee might fly close to the entrance of the hive but make no attempt to enter or leave the area.

The bee might move closer, or further away, from the radar. It might also move in and out of the detection cone rapidly or stay in view for prolonged periods. In essence, the bee moves freely causing signals that resemble those of bees entering and leaving as in Fig. 7(a). Classifying this behavior is valuable in both a commercial and research setting, firstly by removing the potential for these hovering flights to be falsely classified as entry and exit. In addition, it may prove that standalone hovering flights, or hovering before leaving, can be attributed to bee orientation flights, which can be a good indicator of growth, measured by the rate of young bees first leaving the hive [41].

Following the previous procedure of windowing, 200 samples were recovered of strictly hovering behavior. It should be emphasized that we also revisited the original data and split any samples that contained hovering and either of the other types. To balance the dataset, samples of hovering were given additional time shifts of 0.15 and 0.25s creating an approximate equilibrium between all three classes.

With log loss being unsuitable for ternary predictions, and to find comparable loss values between ML models, hinge loss was chosen as a suitable replacement [42]. It was observed from the results that both MFCC and BFCC algorithm lost almost all predictive power, in contrast to the binary results. At best, these techniques were closely matched at 64% accuracy.

Full accuracy and loss values are provided in Table 2 for each learning algorithm alongside each quantization method.

Line spectral frequencies, however, improved significantly over binary results. An improvement was seen in SVM prediction, with an achieved accuracy of 93.4%. Not only is this the best result in ternary results but it is also a significant improvement in prediction across all models trained.

		Binary Classification		Ternary Classification	
Algorithm	Approach	Accuracy	Loss	Accuracy	Loss
Neural Network	LPC	69.02%	0.4787	66.54%	0.5896
	LSF	85.48%	0.3647	89.22%	0.2933
	MFCC	79.01%	0.4718	58.36%	0.6636
	BFCC	79.77%	0.4528	58.18%	0.6498
	Temporal	72.05%	0.5602	52.23%	0.6293
Random Forest	LPC	73.42%	0.5979	65.43%	0.9348
	LSF	84.66%	0.4634	88.85%	0.7117
	MFCC	85.23%	0.4255	63.20%	0.9091
	BFCC	91.13%	0.3693	63.75%	0.8943
	Temporal	70.68%	0.5750	65.80%	0.8645
Support Vector Machine	LPC	55.61%	0.6590	65.80%	0.9149
	LSF	67.67%	0.5066	93.37%	0.2667
	MFCC	63.56%	0.5566	58.36%	0.9126
	BFCC	71.23%	0.5145	59.29%	0.8930
	Temporal	64.38%	0.5806	52.42%	0.9927

The hyperparameters found for this accuracy were a regularization parameter of 13.9, a polynomial degree of 2 and a kernel coefficient of 25.2, using a polynomial kernel.

These results appear as an outlier for SVM, both from the binary (best result 71.2%) and ternary (second-best result 65.8%) classifications. The SVM was trained in 15 separate instances with separate permutations of test and train data. This 15-fold cross-validation returned an average accuracy of 91.02% with a loss of 0.3046. Even when moving 20% additional random data from the training set into the testing set, for a 3:2 ratio of train and test, the accuracy only dropped to 89.5% with a loss of 0.3774. The results are also supported by the NN that achieved 89.2% accuracy and 0.2933 loss, only trailing the SVM results slightly.

The random forest also achieved 88.9% accuracy, though with significantly higher loss. The final step for ML was to compare with the potential of the previous spectrogram approach. The spectrogram system was expanded to incorporate three classes with no other changes with respect to the previous work [27]. This three-class, NN system achieved an accuracy of 75.5%, much below its original two class success of 88.7%. A full visual breakdown of the results across both works is demonstrated in Fig. 7(b). The spectrograms accuracy reduction for three way classification is likely due to the limitations in image resolution and lack of visual difference between the three classes. In theory, the availability of both positive and negative frequency shifts enabled by IQ data would likely increase spectrogram diversification and classification accuracy. This would however come at the expense of increased computational costs associated with image processing neural networks. Finally, a higher resolution ADC and a higher frequency radar providing a higher RCS for honeybees are expected to overcome the limitations in image resolution.



Fig. 6. (a) Natural versus modified frequency values as produced by altering the standard Mel-frequency algorithm, showing the original difference, the first estimate, and the eventual best version. (b) Plot of the two parameters of the MFCC algorithm (constant and denominator) and the effect that their change has on the accuracy of results. The z-axis shows the absolute difference between the accuracy and loss of the trained Random Forest.

C. Machine Learning Summary

Results from both binary and ternary (audio-based) classification show that the dataset has potential to be used in ML applications. In particular, the processed Doppler data has the potential to make predictions more accurate enabling species identification and monitoring of traffic in a more diverse area, such as wild woodland.

A real-time implementation of either binary or ternary classification to monitor hive entrance activity over extended periods of time would make for an interesting research topic, especially correlating activity to several other metrics such as hive health, pollination success, and reproductive success. The present algorithms reliance on audio-frequency rather than image/spectrogram processing is a key enabler for edge-computing architectures. Similarly, the cost-capacity tradeoff achieved at 5.8 GHz benefits from the requirement for processing boards with lower data compression than in higher frequency systems.

Commercially, being able to monitor traffic would provide information on which hives are inactive and in need of replacement for maximum pollination efficacy.

The future availability of IQ data could potentially support a comparative study of our ML based ternary classification



Fig. 7. (a) Visualization of a bee hovering in front of a hive. The bee might come closer or move further away in addition to its other movements. (b) Visual plot of the progress from two to three classes across all included audio quantization methods, in addition to two benchmark lines for comparison to the NN, spectrogram method featured in the previous work.

approach to classical signal processing techniques such as Hidden Markov Model (HMM).

It should be noticed that application of HMM to the DBM data to detect entering, leaving and hovering motion resulted in very poor accuracy of 53%. By providing insight on bee's direction IQ data would allow for more robust classification algorithms ultimately ushering higher prediction accuracy.

VI. CONCLUSION

This paper demonstrates that the current 5.8 GHz CW radar can be deployed to monitor free-flying honeybee activity. This allows long term data collection that facilitates hive surveillance. The radar system was able to identify different near-hive behavior such as leaving, entering and hovering. In addition to the ability to record free-flying honeybees, the radar was also capable of detecting micro-Doppler signals associated to bee wing/limb motion, using both a DBM and an IQ mixer. The developed simulation model accurately predicted calibration target RCS and detection range increase when adding silver coating whose applicability was both theoretically and practically explored.

The volume and quality of the data collected by the radar using the DBM setup was suitable for ML analysis, which was investigated in depth. The BFCC algorithm resulted in

the highest classification accuracy of 91.13% and loss of 0.3693 when using the Random Forest algorithm in Binary classification. In Ternary classification, the LSF approach exhibited the highest accuracy rate of 93.37% and lowest loss of 0.2667 when using the SVM algorithm. It is believed that IQ data, higher ADC resolution and advanced signal processing techniques could further improve the already considerable accuracy rate.

Due to the potential of both real-time hive status monitoring and complex behavior classification, the system can support the extraction of behavioral data as a proxy for important hive health metrics. Real-time and automated monitoring could additionally provide beehive owners with data on inactive hives or potential need for human mitigation. This would not only assist bee farms in hive surveillance tasks; it would also empower soft fruit industry stakeholders with innovative means to monitor biological pollinator behavior and efficiency at relevant locations and/or the impact of specific flower patches and polytunnels settings to increase the overall yield.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the Knowledge Economy Skills Scholarships

(KESS 2, Ref: BUK298) and the Bangor University Innovation and Impact Award (BUIIA, Ref. BU02888).

REFERENCES

- C. A. Hallmann *et al.*, "More than 75 percent decline over 27 years in total flying insect biomass in protected areas," *PLoS One*, vol. 12, no. 10, 2017, doi: 10.1371/journal.pone.0185809.
- [2] F. Sánchez-Bayo and K. A. G. Wyckhuys, "Worldwide decline of the entomofauna: A review of its drivers," *Biol. Conserv.*, vol. 232, no. January, pp. 8–27, 2019, doi: 10.1016/j.biocon.2019.01.020.
- [3] D. H. Morse, "The Insectivorous Bird as an Adaptive Strategy," *Annu. Rev. Ecol. Syst.*, vol. 2, no. 1, pp. 177–200, 1971, doi: 10.1146/annurev.es.02.110171.001141.
- [4] E. Öckinger and H. G. Smith, "Semi-natural grasslands as population sources for pollinating insects in agricultural landscapes," *J. Appl. Ecol.*, vol. 44, no. 1, pp. 50–59, 2007, doi: 10.1111/j.1365-2664.2006.01250.x.
- [5] J. Ollerton, R. Winfree, and S. Tarrant, "How many flowering plants are pollinated by animals?," *Oikos*, vol. 120, no. 3, pp. 321–326, 2011, doi: 10.1111/j.1600-0706.2010.18644.x.
- [6] L. H. Yang and C. Gratton, "Insects as drivers of ecosystem processes," *Curr. Opin. Insect Sci.*, vol. 2, pp. 26–32, 2014, doi: 10.1016/j.cois.2014.06.004.
- [7] IPBES and FAO, "The regional Report for Africa on Pollinators and Pollination and food production," pp. 1–49, 2016.
- [8] A. M. Klein *et al.*, "Importance of pollinators in changing landscapes for world crops," *Proc. R. Soc. B Biol. Sci.*, vol. 274, no. 1608, pp. 303–313, 2007, doi: 10.1098/rspb.2006.3721.
- [9] N. Carreck and I. Williams, "The economic value of bees in the UK," Bee World, vol. 79, no. 3, pp. 115–123, 1998, doi: 10.1080/0005772X.1998.11099393.
- [10] S. G. Potts, J. C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W. E. Kunin, "Global pollinator declines: Trends, impacts and drivers," *Trends Ecol. Evol.*, vol. 25, no. 6, pp. 345–353, 2010, doi: 10.1016/j.tree.2010.01.007.
- [11] T. H. Ricketts *et al.*, "Landscape effects on crop pollination services: Are there general patterns?," *Ecol. Lett.*, vol. 11, no. 5, pp. 499–515, 2008, doi: 10.1111/j.1461-0248.2008.01157.x.
- [12] G. Kairo *et al.*, "Nosema ceranae, Fipronil and their combination compromise honey bee reproduction via changes in male physiology," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, 2017, doi:

10.1038/s41598-017-08380-5.

- [13] S. S. Schneider and L. C. McNally, "Factors influencing seasonal absconding in colonies of the African honey bee, Apis mellifera scutellata," *Insectes Soc.*, vol. 39, no. 4, pp. 403–423, 1992, doi: 10.1007/BF01240624.
- [14] J. L. Woodgate *et al.*, "Harmonic radar tracking reveals that honeybee drones navigate between multiple aerial leks," *iScience*, vol. 24, no. 6, p. 102499, 2021, doi: 10.1016/j.isci.2021.102499.
- [15] J. Shearwood *et al.*, "C-Band Telemetry of Insect Pollinators Using a Miniature Transmitter and a Self-Piloted Drone," *IEEE Trans. Microw. Theory Tech.*, vol. 69, no. 1, pp. 938–946, 2021, doi: 10.1109/TMTT.2020.3034323.
- [16] J. Shearwood, D. M. Y. Hung, P. Cross, S. Preston, and C. Palego, "Honey-Bee Localization Using an Energy Harvesting Device and Power Based Angle of Arrival Estimation," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2018-June, pp. 957–960, 2018, doi: 10.1109/MWSYM.2018.8439173.
- [17] A. Zacepins, V. Brusbardis, J. Meitalovs, and E. Stalidzans, "Challenges in the development of Precision Beekeeping," *Biosyst. Eng.*, vol. 130, pp. 60–71, 2015, doi: 10.1016/j.biosystemseng.2014.12.001.
- [18] M. Hagen, M. Wikelski, and W. D. Kissling, "Space use of bumblebees (Bombus spp.) revealed by radio-tracking," *PLoS One*, vol. 6, no. 5, 2011, doi: 10.1371/journal.pone.0019997.
- [19] J. L. Woodgate, J. C. Makinson, K. S. Lim, A. M. Reynolds, and L. Chittka, "Life-long radar tracking of bumblebees," *PLoS One*, vol. 11, no. 8, pp. 1–22, 2016, doi: 10.1371/journal.pone.0160333.
- [20] A. M. Reynolds, J. L. Swain, A. D. Smith, A. P. Martin, and J. L. Osborne, "Honeybees use a Lévy flight search strategy and odourmediated anemotaxis to relocate food sources," *Behav. Ecol. Sociobiol.*, vol. 64, no. 1, pp. 115–123, 2009, doi: 10.1007/s00265-009-0826-2.
- [21] E. Feuerbacher, J. H. Fewell, S. P. Roberts, E. F. Smith, and J. F. Harrison, "Effects of load type (pollen or nectar) and load mass on hovering metabolic rate and mechanical power output in the honey bee Apis mellifera," *J. Exp. Biol.*, vol. 206, no. 11, pp. 1855–1865, 2003, doi: 10.1242/jeb.00347.
- [22] J. Kim, M. Jung, H. G. Kim, and D. H. Lee, "Potential of harmonic radar system for use on five economically important insects: Radar tag attachment on insects and its impact on flight capacity," *J. Asia. Pac. Entomol.*, vol. 19, no. 2, pp. 371–375, 2016, doi: 10.1016/j.aspen.2016.03.013.
- [23] H. Aumann, B. Payal, N. W. Emanetoglu, and F. Drummond, "An index for assessing the foraging activities of honeybees with a Doppler sensor," SAS 2017 - 2017 IEEE Sensors Appl. Symp. Proc., pp. 7–11, 2017, doi: 10.1109/SAS.2017.7894090.
- [24] A. Hajovsky, R.; Deam, A.; LaGrone, "Radar reflections from insects in the lower atmosphere.," *IEEE Trans. Antennas Propag.*, vol. 14, no. (2), pp. 224–227, 1966, doi: doi:10.1109/tap.1966.1138665.
- [25] H. M. Aumann, "A technique for measuring the RCS of free-flying honeybees with a 24 GHz CW Doppler radar," *IET Conf. Publ.*, vol. 2018, no. CP741, 2018, doi: 10.1049/cp.2018.0540.
- [26] W. W. Wolf, C. R. Vaughn, R. Harris, and G. M. Loper, "Insect Radar Cross-sections for Aerial Density Measurements and Target Classification," *Am. Soc. Agric. Biol. Eng.*, vol. 36, no. June, pp. 949– 954, 1993, doi: doi: 10.13031/2013.28420.
- [27] N. Aldabashi, S. Williams, A. Eltokhy, E. Palmer, P. Cross, and C. Palego, "Integration of 5.8GHz Doppler Radar and Machine Learning for Automated Honeybee Hive Surveillance and Logging," *IEEE MTT-S Int. Microw. Symp.*, pp. 625–628, 2021, doi: 10.1109/ims19712.2021.9574826.
- [28] MATLAB, "R2018a." The MathWorks Inc., Natick, MA, USA, 2018.
- [29] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, "Micro-doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, 2006, doi: 10.1109/TAES.2006.1603402.
- [30] K. Marcus J King, Stephen L Buchmann, H. Spanglar, "Activity of asynchronous flight muscle from two bee families during sonication (buzzing)," J. Exp. Biol., vol. 10, no. 199, pp. 2317–2321, 1996.
- [31] M. A. Jankauski, "Measuring the frequency response of the honeybee thorax," *Bioinspiration and Biomimetics*, vol. 15, no. 4, 2020, doi: 10.1088/1748-3190/ab835b.
- [32] A. B. Carlson, *Communication systems*, 3rd ed. 2000.
- [33] F. Itakura, "Line spectrum representation of linear predictor

coefficients of speech signals," J. Acoust. Soc. Am., 1975, doi: 10.1121/1.1995189.

- [34] D. O. Shaughnessy, "Linear predictive coding," *IEEE Potentials*, 1988, doi: 10.1109/45.1890.
- [35] R. Vergin, D. O'Shaughnessy, and A. Farhat, "Generalized mel frequency cepstral coefficients for large-vocabulary speakerindependent continuous-speech recognition," *IEEE Trans. Speech Audio Process.*, 1999, doi: 10.1109/89.784104.
- [36] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018, doi: 10.1109/CVPR.2018.00474.
- [37] L. Breiman, "Random forests," Mach. Learn., 2001, doi: 10.1023/A:1010933404324.
- [38] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Selfnormalizing neural networks," 2017.
- [39] V. Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A., & Vapnik, "Support vector regression machines," *Adv. Neural Inf. Process. Syst.*, vol. 9, 1997.
- [40] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–15, 2015.
- [41] E. A. Capaldi *et al.*, "Ontogeny of orientation flight in the honeybee revealed by harmonic radar," *Nature*, 2000, doi: 10.1038/35000564.
- [42] K. B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," 2005, doi: 10.1007/11494683_28.



Nawaf Aldabashi (Student member, IEEE) received the BEng degree in Computer Systems Engineering and the MSc degree in Electronic Engineering from Bangor University, UK in 2013 and 2016 respectively, where he is currently pursuing his Ph.D. degree with the department of

Electronic Engineering. His research interests are the design, integration and application of continuous wave and frequency modulated continuous wave radar in addition to nanotechnology and microfabrication.



Samuel M. Williams received the BSc in Computer Science from Bangor University, UK, where he is now a PhD candidate, in 2019. His research interests include the development of machine learning algorithms, acting as augmentation systems for RF and telemetry systems, with the aim of tracking,

monitoring, and predicting insect species and their behaviors.



Amira Eltokhy received the M.S. degree in Wireless Mobile Communication systems, in 2015, and the Ph.D. degree in Microwave Engineering from the University of Greenwich, UK, in 2019. She has been a postdoctoral associate at Bangor University, UK in 2019-

2020. Her research interests include Microwave systems and Biomedical Engineering in addition to Artificial Intelligent integration in both fields.



Edward Palmer is Group Technical Director and Business Development Manager and also leads the R&D Team within S&A Produce. He has extensive experience in all areas of agriculture having driven market outperformance with leading global retailers. Currently he is pioneering and leading other innovative and

industry leading projects around robotics and clear sustainable plant based packaging solutions. Edward is a key advocate for accurate tracking technology which allows bees to behave as naturally as possible, while promoting investigation of human impact and cohabitation strategies.



Paul Cross is a Lecturer in the Environment at Bangor University. His research portfolio encompasses an Agri-tech in China: Newton Network + (ATCNN) project to monitor UAV tracking system performance for pollinators; the use of honey-bee morphometrics to

determine racial strains of ecotypes in the UK involving collaboration with the National Bee Centre Wales; the impact of neonicotinoids on solitary bees; determining racial purity in UK honey bees; Dispersal rates and space use in bees: designing landscapes to accommodate pollinators; and evaluating the potential of beekeeping to alleviate poverty in sub-Saharan Africa.



Cristiano Palego received the M.S. degree in Electrical Engineering from the University of Perugia, Italy, in 2003, and the Ph.D. degree in Microwave Engineering and Optoelectronics from the University of Limoges, France, in 2007. He was a postdoctoral associate then a research scientist at Lehigh University

(USA) (2007-2017). In 2013 he joined Bangor University, UK, where he is now an associate professor in smart sensors and instrumentation and a researcher in the Medical Microwave Systems group. His research interests include micro/nanotechnology, minimally invasive animal telemetry and biomedical technology along with energy harvesting and RF-MEMS for reconfigurable antenna arrays.