DAVIS EXPEDITION FUND

REPORT ON EXPEDITION / PROJECT

Expedition/Project Title: Using Model-based Recognition of Acoustic Signatures to Identify UK Bumblebees to the Species & Caste Levels

Travel Dates: March 1st, 2022 to September 1st, 2022

Location: Royal Botanic Garden Edinburgh

Group Members: Alixandra Prybyla (PhD Student), Ava Nelson (Undergraduate Intern), Perrin Grant (Undergraduate Intern), Jack Seymour (Undergraduate Intern), Isla Hepburn (Undergraduate Intern)

Aims: To collect representative samples from each of Scotland's five most common bumblebees for the training of a convolutional neural network (CNN), thus determining if foraging flight buzzes could be used to discriminate between the species, sexes, and castes (i.e., queen, worker, and male) of Scottish bumblebees.

Photography consent form attached:
(*please refer to your award letter)*

☒ Yes
☐ No

Outcome (a minimum of 500 words):-

DAVIS EXPEDITION FUND

Using Model-based Recognition of Acoustic Signatures to Identify UK Bumblebees to the Species & Caste Levels

Project Proposal: 21st of January, 2022

Prepared by: Alixandra Prybyla & Katrine Redman
Contact: Alixandra Prybyla, A.N.Prybyla@sms.ed.ac.uk, (+44) 7565229127

Cover images: (left) a late season *Bombus hypnorum* queen foraging on a sedum flower (*Hylotelephium sp.*) in Jock Taimson's Gairden (one of ten sampling sites selected by the team); (right) Alixandra preparing to sample the *B. hypnorum* queen once she takes flight. Recording is done with an omnidirectional parabolic microphone (32 bit/44.1 kHz sampling rate) with a flat frequency response within ± 2 dB (range is 100 Hz to 10,000 Hz).

Table of Contents

DAVIS EXPEDITION FUND

I. SUMMARY

Over the course of six months, my team and I (consisting of four undergraduate research assistants and one PhD student) recorded the foraging flight buzzes of approximately 1,150 bumblebees from five different species in the Royal Botanic Garden Edinburgh. In addition to recording their foraging flight buzzes, bees were caught in a net, cooled to torpor (a lethargic state) for safe handling, marked with a small coloured and numbered disc to avoid recapture, measured and weighed, and then released back onto the flower from which they were caught. This allowed researchers to ensure that we were not sampling the same individual's buzzes multiple times. This is especially important given the type of analysis we would conduct on our samples: we would divide our data into train, test, and validation groups, and then train what is known as a convolutional neural network (a CNN) using these groups to estimate the species, sex, and caste of a recorded buzz. After attending a course at the California Institute of Technology (CalTech), I annotated a subset of the data collected during the field season funded by the Davis Expedition Fund (approximately 130 individuals), and programmed a series of CNNs to determine if sound could be used to identify these desired demographic and taxonomic traits. Results varied depending on which trait we were testing for: when testing for species & caste, the balanced accuracy score was 60%. When testing for species identification only, the balanced accuracy score was 69%. And when testing for caste only, the balanced accuracy score was 92%. I believe that, when we use the full dataset, and with some changing to they hyperparameters, we will increase the accuracy score substantially.

II. CONTEXT

There is clear evidence of worldwide bee declines, with approximately 25% fewer species reported between 2006 and 2015 compared to the period before 1990 [1]. These declines result in the loss of vital pollination services, with knock-on negative effects on the environment (ecosystem health and stability is threatened) [2], the economy (crop quality and production is jeopardised), and human welfare (food security and access to insect-pollinated phytopharmaceuticals are threatened) [3, 4]. A key challenge in reversing global bee declines is to detect shifts in local populations quickly enough for successful conservation interventions to be put into place. This is best done via population monitoring. The use of bioacoustics in bee population monitoring is relatively new, but has been successfully applied to approximately 23 bumblebee species [5], thus making them an excellent model for future bioacoustics applications. Of particular interest is the use of bioacoustic monitoring in bumblebee identification, an integral part of population monitoring. Research suggests that bumblebees can be identified to their species and caste (i.e., queen, worker, or male) levels based on their flight buzz [6, 7, 8, 9]. These preliminary studies, however, compared the sounds of organisms we would expect to be acoustically very different (like a bee and a hornet) [7], or have failed to adequately sample likely causes of acoustic diversity within a species, particularly body size and age-related wing wear [8, 6]. Both body size and age-related wear vary substantially within individual bee species and are expected to affect the sounds a bee makes, hence the ability of human ears and computers to distinguish them.

Furthermore, the data used to establish the theory that flight acoustics could be used to categorise individuals is limited in its geographic and temporal scope. By this, we mean that these conclusions were drawn from samples taken from one or two sampling sites per study, over the course of several days to several weeks. As we have established previously, these geotemporal variables are exceptionally important when it comes to bumblebee morphology, and so must be adequately accounted for in data collection. Therefore, to know how well sound might be used to identify bumblebees to the species and caste levels, we surveyed across six months to account for within species acoustic variation as well as between species variation (see: traits outlined in "Introduction to British Bee Biology").

There is also the question of the number of specimens sampled in each of these foundational studies. While Gradisek et al. [8] sampled 197 specimens over 12 species, the dataset was also divided into castes (in this case, queens, and workers) which they consider to be 21 sampled groups. As an example, they sampled *Bombus pascorum* queens and workers and considered them acoustically distinct. Among these 21 groups, 13 groups had less than 10 samples, and 7 groups had less than 5 samples. It is unclear where these specimens were sampled. Miller-Struttman et al. [6] sampled 15 *B. balteatus* specimens and 13 *B. sylvicola* specimens of two castes (again, queens and workers), and 17 additional bumblebee species from an online database which were not listed. Kawakita and Ichikawa [13] do not state how many biological samples they took (i.e., number of specimens) but instead focus on the number of acoustic samples. These numbers are likely not sufficient in capturing the full diversity of a species, which is what we ameliorate with our extended sampling period.

III. DATA COLLECTION AND ANALYSIS

Data collection took place at the Royal Botanic Garden Edinburgh (RBGE) from March to September of 2022. All sampling occurred from 9AM to 6PM on either sunny or warm (at least 10°C) days with wind levels low enough (usually below 27 MPH) so that recordings were still usable. Sampling involved my pursuit of individual bees using an omnidirectional microphone attached to a parabolic reflector. The parabolic reflector serves to focus and allow sampling of low-frequency sounds (down to 60Hz) while also capturing high-quality background noises, such as bird calls, cars, human voices, and any other background noise. These background noises are important so that the CNN can be taught what is a bee and what is not a bee during its training and testing. Each recording had the same sampling rate (32 bit/44.1 kHz sampling rate) for continuity. Bees were initially indiscriminately selected, and then as the season continued, we endeavoured to fill parts of our dataset that were lacking in representation while also keeping in mind that we needed data for all bees across their active season.

Once a usable audio clip has bee procured (usable in this context meaning a clip that had a steady foraging buzz recording of at least one full second), the bee was netted and cooled to torpor in a custom cooling chamber kept between -2°C and 2°C. Temperature was controlled manually by adding or removing reusable ice packs throughout the day. Once cooled, the bee's weight, body size (via intertegular distance), and wing wear and shape were recorded. We conducted a pollen count (i.e., scored the amount of pollen in the bee's corbiculae), and counted the number of commensal mites on its body. We additionally tagged the bee with a coloured and numbered disc to avoid recaptures and resamples. Bees, once fully measured, were warmed in our hands before being replaced on the flower they were foraging on before they were captured. If this flower was shaded, the bee was placed on a flower of similar size and structure in the sun. In addition to collecting information about the bee, we additionally collected environmental information at the time of its capture. This includes the irradiance at the time the bee was captured (given that bees are ectothermic, their body temperature is influenced by light and temperature, this making irradiance an important measure).

Figure 1: The Royal Botanic Garden Edinburgh (RBGE) serves as the perfect location for pollinator research as their living collections bloom throughout the year, thus provisioning pollinators throughout much of their active season. The above image is a representation of RBGE in map form, exhibiting the rich greenery available to local fauna. While we had several sites we preferentially surveyed at as we discovered which species would occur where and on which plants, we collected bumblebees across RBGE. Each location is well-represented in our dataset. Image copyright of the Royal Botanic Garden Edinburgh, all rights included.

Our surveying resulted in 1,152 unique foraging buzzes, with accompanying data for nearly each bee. The breakdown of data gathered is as follows:

Species and castes captured in 2022 field season

Species	**Total count**
Anthophora plumipes (F)	25
Bombus pascuorum (Q)	86
B. pascuorum (W)	105
B. pascuorum (M)	6
White-tailed aggregate (Q)	61
White-tailed aggregate (W)	169
B. terrestris (M)	71
B. lucorum (M)	13
B. pratorum (Q)	42
B. pratorum (W)	51
B. pratorum (M)	56
B. hortorum (Q)	23
B. hortorum (W)	23
B. hortorum (M)	4
B. hypnorum (Q)	38
B. hypnorum (W)	42
B. hynorum (M)	12

Table 1. In this table, you will see the breakdown of species and castes we caught in 2022. You will notice a group called "white-tailed aggregate", which refers to *B. lucorum* and *B. terrestris*. These females (i.e., workers and queens) of this species are undifferentiable without genetic analysis, something we were not aware of when we first embarked on this fieldwork. We have since developed a DNA barcoding system that can easily and cheaply differentiate between the two via a leg clipping. However, since we cannot be sure which bees we are which species, we are aggregating them here. You will also notice a non-bumblebee species, *A. plumipes*. These bees were included in the analysis because they have less physical variation than bumblebees and serve as a way for us to gauge how our CNN is working (i.e., if it is struggling with *A. plumipes*, then we should reconsider aspects of our hyperparameters).

Once data had been collected, it was first annotated (i.e., sound samples were labelled so that they could be read as having a buzz present or a buzz not present) in the free software Audacity, and then split into training, test, and validation sets (15%, 15%, and 70% split). The training set was used to train my algorithm known as BeeNet, a CNN built on a ResNet18 architecture. The construction of the convolutional neural networks that were used to analyse this data did not occur until summer of 2023 at CalTech (i.e., more than a year after the data collection took place). Because of their complexity, and because they were constructed outside the scope of this grant, I will keep the description of their construction simple. Once BeeNet was trained, it was tested, refined, and then finalised for purpose.

IV. PRELIMINARY RESULTS

BeeNet was reconfigured three times: the first time was to test to see if the algorithm could differentiate a bumblebee to the species and caste levels. The second time was to determine if BeeNet had an easier time differentiating bees based on their species. The results, given that we were using only a small subset of our data, were very promising. We visualise our results with confusion matrices. Confusion matrices, as you will see below, are used in machine learning to evaluate the performance of a classification algorithm. The rows represent the actual classes, and the columns represent the predicted classes. We have true positives (where an instance belongs to a class and is predicted to belong to that class), true negatives (where an instance does not belong to a class and is correctly predicted not to belong to that class), false positives (where an instance does not belong to that class but is predicted to belong to that class), and false negatives (where an instance that does belong to a class but is incorrectly predicted by the model not to belong to that class). You will notice on the following confusion matrices that we have labelled some of our classes as *B. terrestris*, though they should be construed as the conglomerate group of *B. terrestris* and *B. lucorum*.

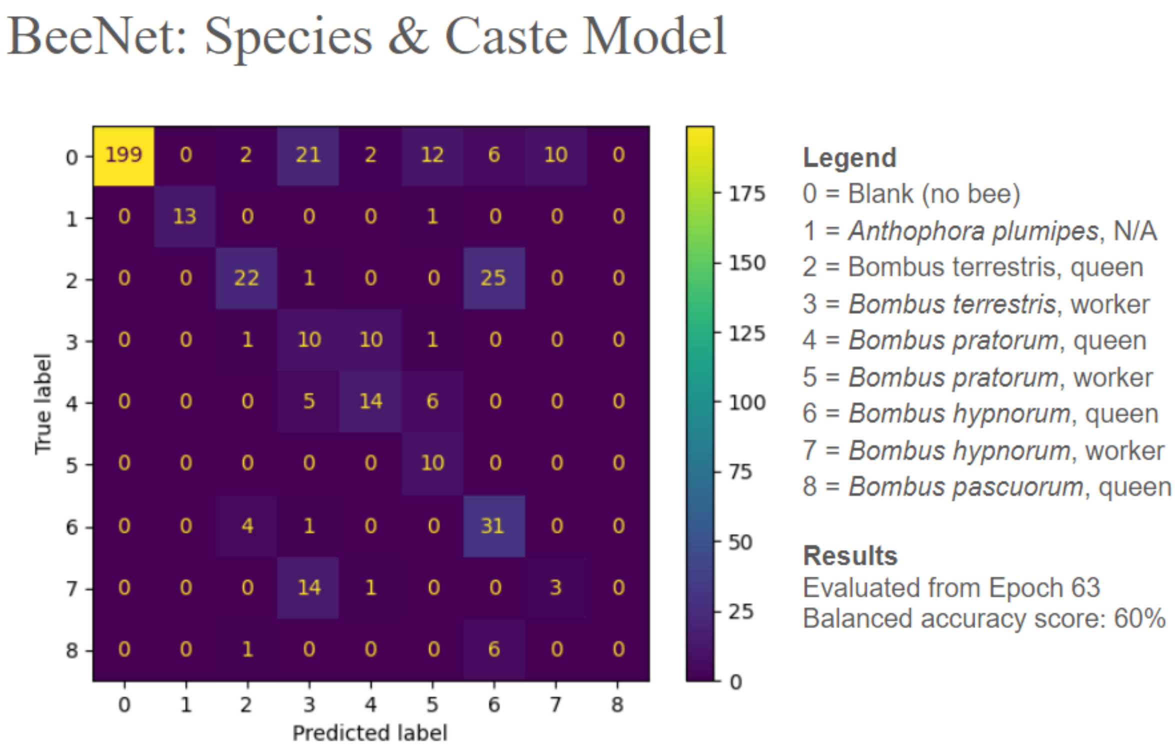

Figure 2. In this confusion matrix, which aims to identify a bee to its species and caste levels based only on its buzz, has a balanced accuracy score of 60%. Class 1, A. plumipes, is highly accurate with only one misidentification. This is likely because they don't exhibit high amounts of polymorphism, and so their sounds are all likely very similar to each other. Class 3, the white-tailed aggregate workers, are misidentified as 4, *Bombus pratorum* queens. I suspect this has something to do with size. White-tailed aggregate workers have a very wide

size range, while *B. pratorum* queens tend to remain quite small. There is likely morphological overlap here.

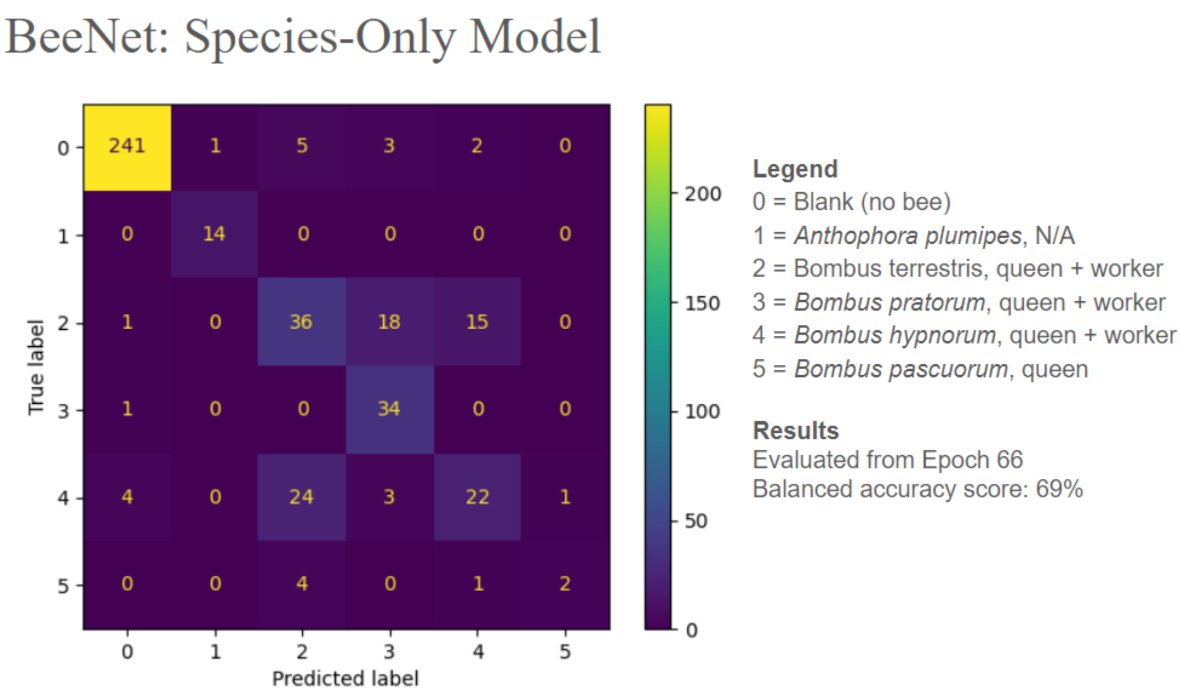

Figure 3. In this confusion matrix, which aims to identify a bee to its species level, we have a balanced accuracy score of 69%. This was surprising at first: the size and morphological differences that occur between queens and workers made me think that the algorithm would have a harder time differentiating them. This leads me to think that there might be other parts of a buzz that are species-specific, qualities that have yet to be identified.

V. DISCUSSION AND NEXT STEPS

In this ongoing study, the next steps involve refining BeeNet: I will move away from a binary classification model structure (i.e., is this one thing or is it not) and reprogram it into a more robust multi-class classification algorithm (where the algorithm can take many aspects of data into account at the same time, such as weather, irradiance, &c. and make a classification determination). My dataset, which is now fully annotated, will additionally play a crucial role in augmenting the algorithm's capabilities. Furthermore, the study aims to broaden the model's applicability by differentiating between the cryptic white tailed species, which we will be able to do with our DNA barcoding methods. This expansion will be facilitated by the utilization of newly annotated data.

To augment the temporal scope of the study, efforts will continue to gather acoustic data across several years. Building upon the data collected in the previous year, ongoing and forthcoming fieldwork will ensure a comprehensive dataset spanning a three-year timeframe. These collective efforts are poised to contribute significantly to the understanding of bumblebee bioacoustics and enhance the CNN's efficacy in species classification over a more extensive taxonomic range and temporal context.

VI. REFERENCES

[1] Zattara, Eduardo E., and Marcelo A. Aizen. "Worldwide occurrence records suggest a global decline in bee species richness." *One Earth* 4.1 (2021): 114-123.
[2] Ameixa, O. M. C. C., et al. "Ecosystem services provided by the little things that run the world." *Selected Studies in Biodiversity. Intechopen, Londin* (2018): 267-302.
[3] Gallai, N. (2008). Array for the economic valuation of the contribution of insect pollination to agriculture and impact on the welfare [WWW Document]. *FAOSTAT,. http://www. internationalpollinatorsinitiative. org/uploads/POLLINATION VALUE ARRAY. xls*.
[4] Van der Sluijs, Jeroen P., and Nora S. Vaage. "Pollinators and global food security: the need for holistic global stewardship." *Food Ethics* 1 (2016): 75-91.
[5] Abdollahi, Mahsa, Pierre Giovenazzo, and Tiago H. Falk. "Automated beehive acoustics monitoring: A comprehensive review of the literature and recommendations for future work." *Applied Sciences* 12.8 (2022): 3920.
[6] Miller-Struttmann, N. E., et al. (2017) Flight of the bumble bee: buzzes predict pollination services." *PLoS ONE* 12(6), e0179273
[7] Kawakita, S. & Ichikawa, K. (2019). Automated classification of bees and hornet using acoustic analysis of their flight sounds. *Apidologie* 50(1), 71-79.
[8] Gradišek, A. et al. (2017). Predicting species identity of bumblebees through analysis of flight buzzing sounds." *Bioacoustics* 26(1), 63-76.
[9] Prybyla, A. [Unpublished raw data]