File name	Sub-sheets	Variables	Description
`D1_sensor.csv` & `D2_sensor.csv`	-	Published_at	Time stamp (YYYY-MM-DD HH24:MI:SS)
		Hive ID	ID unique to each hive
		Apiary ID	ID of the apiary
		Temperature	Temperature in degree Celsius
		Relative humidity	Humidity in percentage
		Audio features	twenty audio feature values
`D1_ant.xlsx`	Visit	Yard	Name of the apiary
		Dates	Date of the visits
		Arrival time	Time of arrival (HH-MM)
		Departure time	Time of departure (HH-MM)
		Manipulations	Details of human evaluations
		Yard location	Coordinates of two apiaries
	Evaluation [NUMBER]	Dates	Date of the population evaluation
		Yard	Name of the apiary
		Hive ID	ID unique to each hive
		Number of boxes	Bottom chamber + honey supers
		NoF covered by bees	Population measurement
	Phenotypic measures	Brood surface	Capped, uncapped, and total brood cells
		Varroa infestation	Severity of *Varroa*
		Defensive behavior	Defensiveness measured by flag test
		Hygienic behavior	Cleaning capacity
		Honey weight	Total honey produced during summer
	READ ME	-	A quick introduction on file content of all sub-sheets
`D2_ant.xlsx`	-	Apiary	Name of the apiary
		Hive ID	ID unique to each hive
		Mortality cause	Causes of the failed hives
		Weight	Weight values (kg) before and after winterization
		Bees frames	Population before and after winterization
		Syrup consumption	Syrup consumption (kg) during winterization

Table 1: Structure of the multi-modal sensor data and phenotypic trait measurement files.

Data records

The MSPB dataset is made fully available at the Zenodo repository `https://doi.org/10.5281/zenodo.8371700`. The sensor data and phenotypic traits were stored separately in `.csv` format, each of which was further divided into two files based on the time range, resulting in a total of four `.csv` files. To distinguish summer and winter data, those collected between April, 2020 and October, 2020 received a 'D1' label in the file name, while the data between October, 2020 and April, 2021 were labelled as 'D2'. The detailed file composition is summarized in Table 1. The total size of the shared files is about 500 MB.

D1 and D2 sensor data are both paired with (1) the time stamp (date and time) of the data collection, (2) hive ID, which is a unique number to identify each hive, (3) apiary ID, which indicates the apiary location of the hive, (4) temperature values, (5) relative humidity values, and (6) twenty audio features. The D1 phenotypic traits file has three sub-sheets, which details (1) the visit date and time of the human evaluations, as well as the evaluation tasks, (2) the population size of the colonies measured at each visit, (3) other phenotypic trait measurements, such as *Varroa* infestation status, defensive and hygienic behavior, honey weight, etc. During the period of D2, hives were maintained in the winter chambers and only evaluated once in the Spring to check their winter survival rate. Hence, the D2 phenotypic traits file contains the survival status, as well as the mortality causes (if any) of all hives.