



# Using honey bee flight activity data and a deep learning model as a toxicovigilance tool



Ulises Olivares-Pinto <sup>a</sup>, Cédric Alaix <sup>b</sup>, Yves Le Conte <sup>b</sup>, Didier Crauser <sup>b</sup>, Alberto Prado <sup>a,\*</sup>

<sup>a</sup> Escuela Nacional de Estudios Superiores Unidad Juriquilla, Universidad Nacional Autónoma de México, Blvd. Juriquilla 3001, 76230, Juriquilla, Querétaro, Mexico

<sup>b</sup> INRAE, Abeilles et Environnement, Avignon, France

## ARTICLE INFO

### Keywords:

Honeybee flight activity  
Toxicovigilance  
Neurotoxic pesticides  
Recurrent neural network

## ABSTRACT

Automatic monitoring devices placed at the entrances of honey bee hives have facilitated the detection of various sublethal effects related to pesticide exposure, such as homing failure and reduced flight activity. These devices have further demonstrated that different neurotoxic pesticide molecules produce similar sublethal impacts on flight activity. The detection of these effects was conducted a posteriori, following the recording of flight activity data. This study introduces a method using an artificial intelligence model, specifically a recurrent neural network, to detect the sublethal effects of pesticides in real-time based on honey bee flight activity. This model was trained on a flight activity dataset comprising 42,092 flight records from 1107 control and 1689 pesticide-exposed bees. The model was able to classify honey bees as healthy or pesticide-exposed based on the number of flights and minutes spent foraging per day. The model was the least accurate (68.46%) when only five days of records per bee were used for training. However, the highest classification accuracy of 99%, a Cohen Kappa of 0.9766, a precision of 0.99, a recall of 0.99, and an F1-score of 0.99 was achieved with the model trained on 25 days of activity data, signifying near-perfect classification ability. These results underscore the highly predictive performance of AI models for toxicovigilance and highlight the potential of our approach for real-time and cost-effective monitoring of risks due to exposure to neurotoxic pesticide in honey bee populations.

## 1. Introduction

Honey bees, as major pollinators of crops (Garibaldi et al., 2013) and wild plants (Hung et al., 2018), are frequently exposed to multiple pesticides that are present in nectar, pollen, and water sources (Johnson, 2015; Poquet et al., 2016; Samson-Robert et al., 2014; Zioga et al., 2020). In fact, pesticide exposure has been identified as one of the major stressors on honey bee health (Goulson et al., 2015; Siviter et al., 2021). Hence, there is a strong interest in studying the sublethal effects of pesticides. Unwanted, sublethal effects of pesticides on honey bees have led to the regulation and banning of certain chemicals (Sgolastra et al., 2020). However, in order to improve pesticide risk assessment, one major challenge is to increase the ecological relevance of toxicity tests by assessing low-dose effects in free-flying bees (Barascou et al., 2021). This limitation has been partly circumvented by different remote sensing systems that have been used successfully to monitor the activity of foraging honey bees (Alaux et al., 2014; Barascou et al., 2022; Capela et al., 2022; Decourtye et al., 2011; Henry et al., 2012; Shi et al., 2020). These non-invasive remote sensing devices are based on detectors,

placed at the hive's entrance, that monitor the activity of hundreds of individually tagged bees, allowing the measurement of flight-related traits (Decourtye et al., 2011).

The use of radio frequency ID (RFID) and optical bee counters (OC) that monitor the activity of individually tagged bees have allowed for the identification of several undesirable sublethal effects of neurotoxic pesticides on honey bees, including homing failure (Capela et al., 2022; Decourtye et al., 2011; Henry et al., 2012) and reduced flight activity (Barascou et al., 2022; Colin et al., 2019; Coulon et al., 2020; Prado et al., 2019; Shi et al., 2020). Interestingly, similar sublethal effects on flight activity have been identified for different neurotoxic pesticide molecules. For example, Colin et al. (2019) found that bees exposed to imidacloprid, a neonicotinoid pesticide, performed 28% fewer foraging trips than control bees (Fig. 1A). Prado et al. (2019) found that bees exposed to an organophosphate or a pyrethroid mixed with fungicides performed ~1.4 fewer flights per day than control bees (Fig. 1B). Barascou et al. (2022) found that bees exposed to an insecticide from the sulfoximine group performed up to 33% fewer daily flights than controls (Fig. 1D). Coulon et al. (2020) found that honey bees orally exposed to

\* Corresponding author.

E-mail address: [aprado@unam.mx](mailto:aprado@unam.mx) (A. Prado).

thiamethoxam (neonicotinoid) exhibited a slower daily increase in flight duration, ~8 min less, compared to control bees (Fig. 1H). Thus, the reduction in honey bee daily activity could potentially be used as a general indicator of neurotoxins presence in the environment and, hence, as a toxicovigilance tool for real-time monitoring of colony health. Furthermore, the early identification of pesticide effects would benefit pesticide risk assessment, notably post-marketing authorization studies that monitor adverse effects linked to the use of pesticides.

Machine learning coupled with big data technology has permeated agriculture to inform crop, livestock, and water management (Liakos et al., 2018). For example, previous studies have successfully used artificial intelligence to analyze and classify animal behavior (Kleanthous et al., 2022), and evaluate animal well-being (McLennan et al., 2015). Furthermore, the use of neural networks or deep learning architectures has significantly increased for solving entomological problems (Marinho et al., 2023; Peng and Wang, 2022; Tuda and Luna-Maldonado, 2020). For example, Convolutional Neural Networks (CNN) have been applied to the classification of fruit fly pupae as parasitized or healthy (Marinho et al., 2023) and the recognition of insect pest species (Peng and Wang, 2022; Tuda and Luna-Maldonado, 2020). Regarding honey bees, recent studies have proposed deep learning-based image classification models for assessing individual bee health state (Berkaya et al., 2021) and CNN for evaluating colony health based on acoustical data (Truong et al., 2023). A more recent study has used OC at the entrance of beehives coupled with an artificial intelligence model that can classify returning bees as pollen foragers or non-pollen foragers to evaluate the effects of the neonicotinoid imidacloprid on bee foraging behavior (Wang et al., 2024). Here, we propose to use a similar approach based on OC data coupled with an artificial intelligence model, but our goal is to classify bee flight performances and determine whether they are healthy or pesticide-exposed. This is the first time an artificial intelligence model has been used to detect the sublethal effects of pesticides based on flight activity data. Our approach is novel as it first integrates the activity data individually in a time series for a single bee and then uses the recurrent neural network for classification (Fig. 2). This new technology could potentially provide real-time data to agricultural operations and environmental agencies for pesticide risk assessment.

## 2. Methods

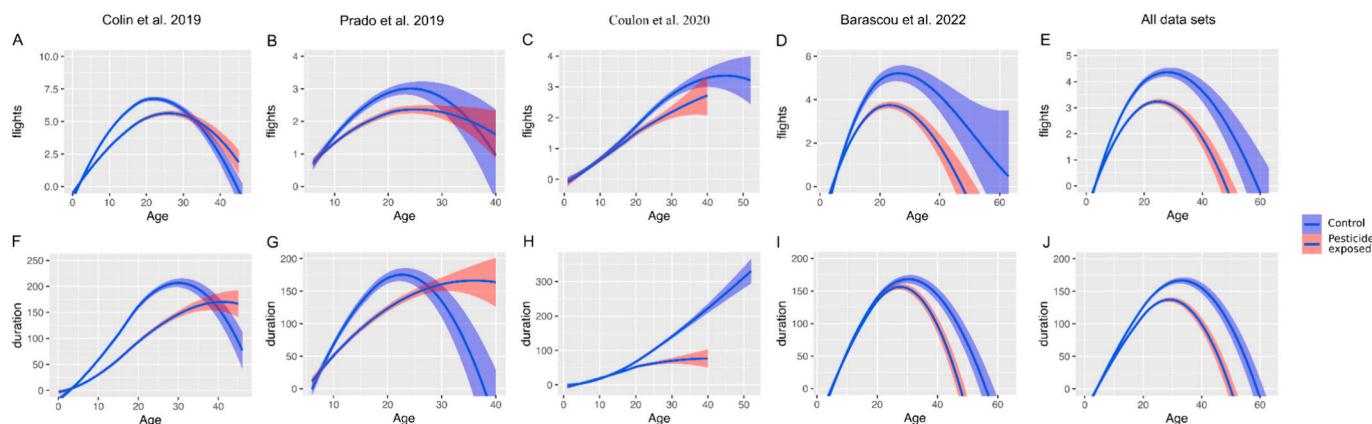
Fig. 2 summarizes a series of processes in designing a deep-learning architecture that acts as a toxicovigilance tool. Subsequent sections will provide a comprehensive description of each section.

### 2.1. Data collection

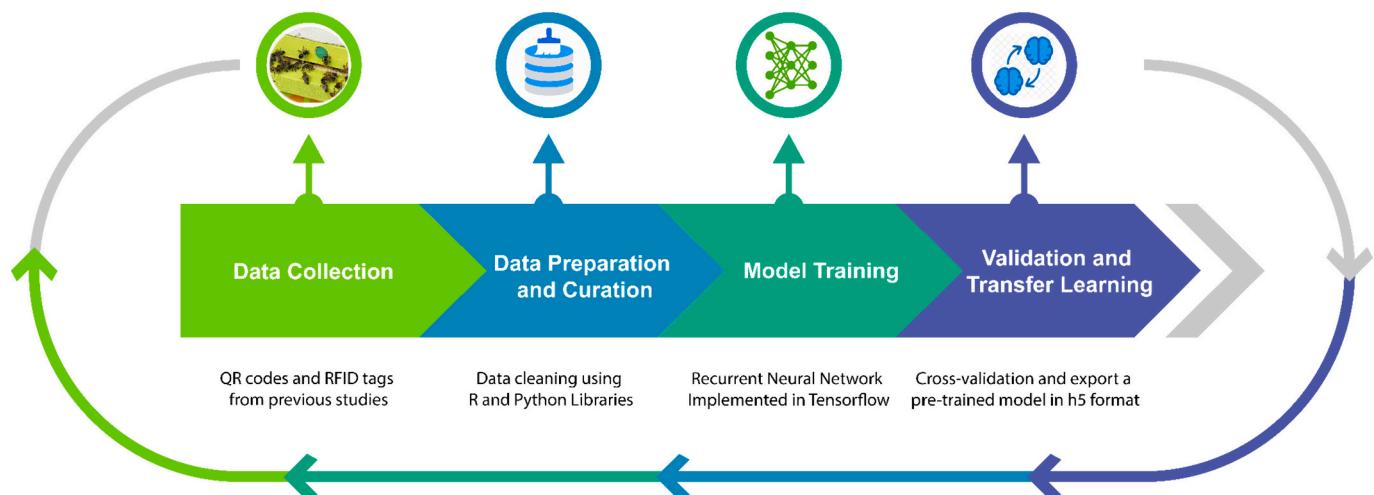
Four previously published data sets of honey bee flight activity, three using optical bee counters (Barascou et al., 2022; Coulon et al., 2020; Prado et al., 2019) and one using RFID (Colin et al., 2019), were used to train and test the recurrent neural network that is detailed in the sections below. All four data sets include the activity (exit /entrance) of individually-tracked bees in relation to their age in days. In the optical bee counter studies, bees were individually marked with a data-matrix barcode (3 mm diameter) printed on laminated paper and glued on the thorax with epoxy glue (Sader®). The activity of barcoded bees was recorded using optical bee counters at the hive entrance (Alaux et al., 2014). The optical bee counter consists of a camera that monitors the hive entrance and image analysis software that detects and registers the barcode. Prado et al. (2019) exposed bees to two common pollen-bound pesticide mixtures at field-realistic sublethal concentrations (209 control and 391 exposed). One mixture combined four fungicides: cyprodinil (540 ppb), fludioxonil (180 ppb), difenoconazole (38 ppb), dodine (34 ppb), and one neurotoxic pyrethroid insecticide, tau fluvalinate (310 ppb). The second mixture was comprised of three fungicides, cyprodinil (270 ppb), fludioxonil (250 ppb), and iprodione (65 ppb), and one organophosphate, chlorpyrifos (270 ppb). Coulon et al. (2020) exposed bees to two field-realistic concentrations of the neonicotinoid thiamethoxam (50 and 200 ppb; 224 control and 207 exposed), and Barascou et al. (2022) exposed bees to two field-realistic sublethal doses of sulfoxaflor, a sulfoximine neurotoxic insecticide (16 and 60 ng; 423 control and 796 exposed). In the RFID study, Colin et al. (2019) exposed bee larvae to a trace concentration of the neonicotinoid imidacloprid in sugar syrup (5 ppb; 251 control and 295 exposed). Once in the adult stage, Radio Frequency Identification tags (RFID; Invengo Technology) were glued to the bees' thoraxes. Hives were equipped with RFID antennae, which recorded bee activity.

### 2.2. Data preparation and curation

For the analysis of flight activity, exit-entrance sequences shorter than 1 min or longer than 240 min were excluded (not considered as true flights; Prado et al., 2019). For each bee, the daily activity was calculated with two parameters: the total number of minutes spent outside and the total number of flights. The `dropna` function from the `pandas` V 1.5.3 library was used to exclude missing values from the Colin et al., 2019 dataset. All other datasets contained complete records. Next, numerical variables were standardized to a specific range using the `MinMaxScaler` function of the `scikit-learn` V1.2.2 library. This normalization process ensures that each numerical variable contributes equally to the performance of the Deep Learning Model.



**Fig. 1.** Common sublethal effects of pesticides on honey bee flight activity. A-E) Number of flights performed per day as a function of bee age. F-J) Duration of daily flight activity in minutes (total time spent outside daily) as a function of bee age. Data from Colin et al., 2019 (A and F), Prado et al., 2019 (B and G), Coulon et al., 2020 (C and H), Barascou et al., 2022 (D and I). All datasets (E and J) were used to train a deep-learning model.



**Fig. 2.** Block diagram of deep-learning solution for the early detection of pesticide-exposed bees. The process consists of 1) Data collection, 2) Data preparation and curation, 3) Model training, and 4) Model validation and transfer learning.

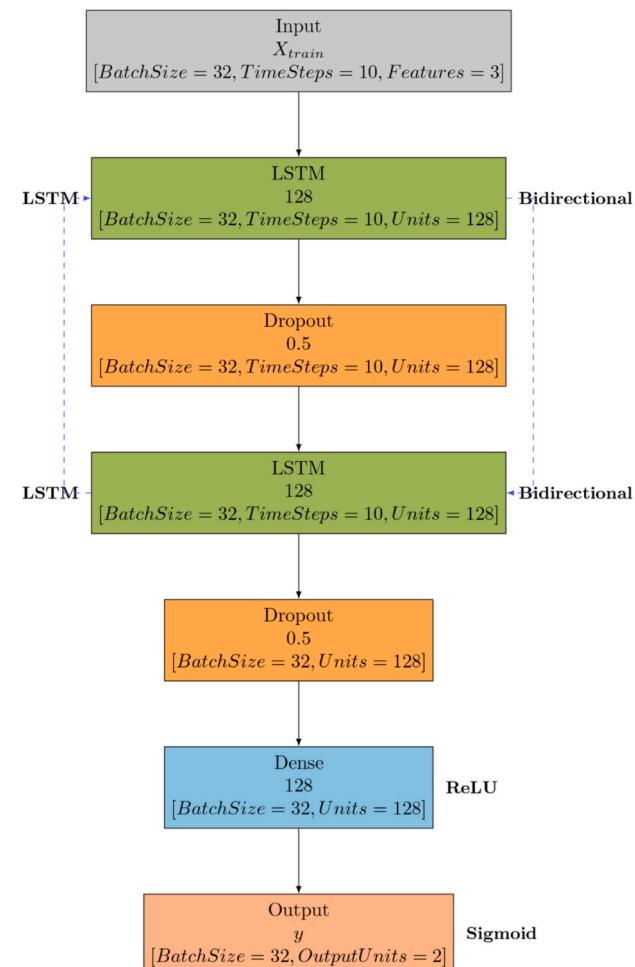
This study used the `train test split` function from the `scikit-learn` library (version 1.2.2), a function that facilitates randomized shuffling and division of datasets into distinct training and testing subsets. Specifically, 70% of the data was designated for training the classification model, while the remaining 30% was set aside for testing (Dobbin and Simon, 2011). Individuals from the control and pesticide-exposed groups were evenly distributed between these subsets, guaranteeing a balanced representation of each class. The recurrent neural network (RNN) model was designed to incorporate three primary input parameters: the bee's age, the daily number of flights, and the total time spent outside daily.

Following data preparation, labels for both training and testing datasets underwent a transformation known as one-hot encoding, facilitated by the `OneHotEncoder` function from the `scikit-learn` library (version 1.2.2). This function converts categorical variables into a binary matrix format. Each unique label in the original data is represented as a distinct binary column in the transformed output, where the presence of a particular category is marked by a "1" and its absence marked by a "0". In the context of the current study, a value of "1" was assigned to denote bees that have been exposed to pesticides, whereas a "0" value identified control bees (i.e., unexposed to such chemicals).

The model was trained using varying disjoint subsets of the dataset in order to assess the number of recordings necessary to yield satisfactory classification metrics. These subsets corresponded to different time steps, namely 5, 10, 15, 20, and 25 days of flight activity recordings. Subsequent experiments included training the model with various combinations of datasets: 1, 2, 3, or all four datasets combined. These raw and curated versions of the datasets are readily accessible in the Code and Data Availability section.

### 2.3. Deep neural network architecture

The bee activity dataset comprises daily entries from individual bees, organized into discrete time series containing 5, 10, 15, 20, or 25 days of registers, as depicted in the input layer of Fig. 3. Each record within the dataset captures critical details such as the bee's age, the total flights undertaken per day, and the cumulative duration spent outside the hive, collectively representing a sequence of bee flight performances. Given the structured nature of this data, this study urged a robust approach to handling time-series data effectively. Therefore, a recurrent neural network (RNN) with long short-term memory (LSTM) was selected for this study design (Malhotra et al., 2017; Rumelhart and Hinton, 1986). This decision leveraged the RNN with LSTM's proven capability to process sequential datasets and discern temporal patterns, which are crucial for the real-time analysis required in toxicovigilance (Tealab,



**Fig. 3.** Model architecture for predicting pesticide presence in bees using a Recurrent Neural Network. This framework integrates two Bidirectional Long Short-Term Memory (BiLSTM) layers. The model is designed to process three integrated features: the daily number of flights, the duration spent outside each day (measured in minutes), and the bee's age.

2018). Additionally, the incorporation of bidirectional LSTMs (BiLSTM) helps in decoding sequences from both historical and future records, enriching pattern discernment in a time series (Hochreiter and Schmidhuber, 1997; Karim et al., 2019; Khan et al., 2021).

In the proposed architecture, the model employs dual BiLSTM layers in its structure. Following each BiLSTM layer, dropout regularization is incorporated to mitigate the risk of overfitting. Specifically, this regularization method entails the random deactivation of 50% of the neural connections (Srivastava et al., 2014). After these first layers, the architecture introduces a densely connected layer comprised of 128 neurons. This layer utilizes a Rectified Linear Unit (ReLU) activation function (Nair and Hinton, 2010). A mechanism was proposed to facilitate the classification of the processed sequences, as depicted in Fig. 3. The model harnessed the capabilities of the NVIDIA CUDA Deep Neural Network (cuDNN) library to enhance computational efficiency and reduce training duration, as presented by Chetlur et al. (2014).

#### 2.4. Model training and hyperparameters

The model's weights were randomly initialized using a specific integer seed to guarantee consistent reproducibility of results. Furthermore, meticulous manual tuning of hyperparameters was conducted to achieve optimal model performance. The selected hyperparameters, which yielded the best performance metrics, are presented in Table 1.

#### 2.5. Model validation

Two distinct validation methodologies were implemented. The first method involved partitioning the dataset into training and validation sets, with 70% of the data designated for training and 30% for validation. This division was performed through random selection (Dobbin and Simon, 2011).

In contrast, the second method, cross-validation, was used as a robust statistical technique for assessing how the model can perform on an independent dataset. This approach divides the dataset into  $K$  distinct subsets (folds). During each cross-validation iteration, one of these folds was held out as the validation set, while the remaining  $K-1$  folds were combined to form the training set. The model was trained on the  $K-1$  folds and then validated on the remaining fold. This procedure was conducted  $K$  times, rotating the validation fold in each iteration and resulting in  $K$  unique pairs of training and validation datasets. A more accurate estimate of the model's ability to generalize new data was obtained by averaging the performance across all  $K$  iterations. Such a cross-validation approach effectively enhanced the model's performance. It reduced the likelihood of overfitting, ensuring that every data point has been included in the training and validation sets throughout the  $K$  iterations (Santos et al., 2018).

#### 2.6. Performance metrics

Conventional metrics were used to evaluate the model's performance, such as confusion matrices, accuracy, recall, precision, F1-score, Cohen Kappa, and the area under the ROC curve, some of which are defined below (Johnson and Khoshgoftaar, 2019).

**Table 1**

List of hyperparameters utilized for training the Recurrent Neural Network model.

Hyperparameter	Value
Optimizer	ADAM (Kingma and Ba, 2017)
Learning rate	0.003
Batch size	32
Epochs	100
Dropout rate	0.5
Loss function	Binary Cross Entropy

#### ● Confusion Matrix

A valuable and common performance metric within classification problems is the confusion matrix. This matrix compares predicted and true labels and contains four elements for binary classification problems. Those elements correspond to true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. These values are essential for identifying misclassifications and label-related mistakes predicted by the model.

#### ● Accuracy: Number of correct predictions over all predictions.

$$acc = \frac{TP + TN}{TP + TN + FP + FN}$$

#### ● Recall (sensitivity): Ratio between the true positives and the number of positives.

$$recall = \frac{TP}{TP + FN}$$

#### ● Precision: Ratio between true positives and the number of predicted positives.

$$precision = \frac{TP}{TP + FP}$$

#### ● F1-Score: Weighted average between precision and sensitivity.

$$f1 = \frac{2TP}{2TP + FP + FN} = \frac{precision * recall}{precision + recall}$$

#### ● Cohen Kappa: Measures the reliability and validity of the model predictions, typically used in imbalanced data sets.

$$k = \frac{2 * (TP * TN - FN * FP)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)}$$

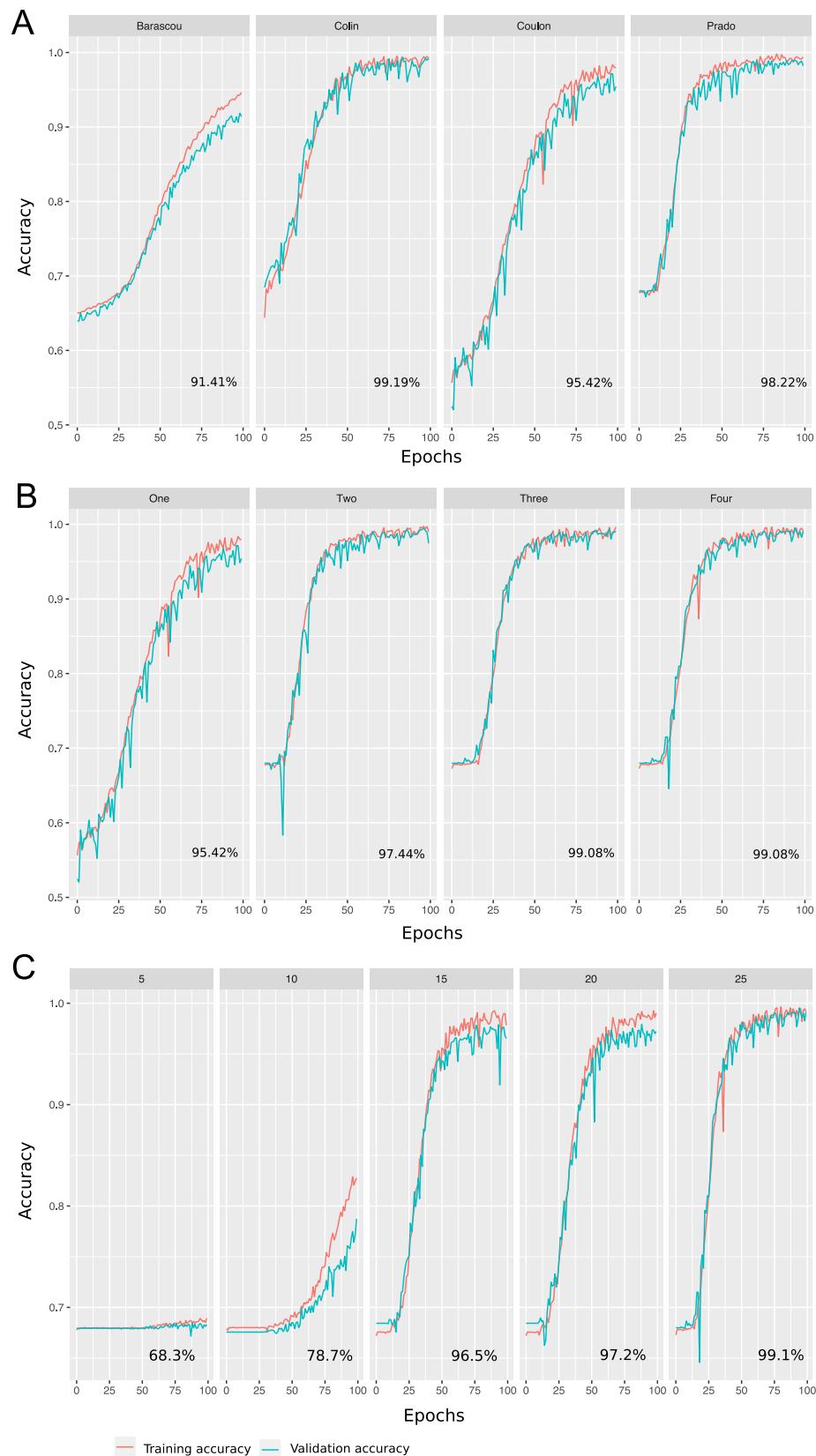
#### ● AUC (Area Under the Curve): The Area Under the Receiver Operating Characteristic (ROC) Curve compares the True Positive Rate (i.e., recall) and the False Positive Rate (Hand, 2001).

#### 2.7. Transfer learning

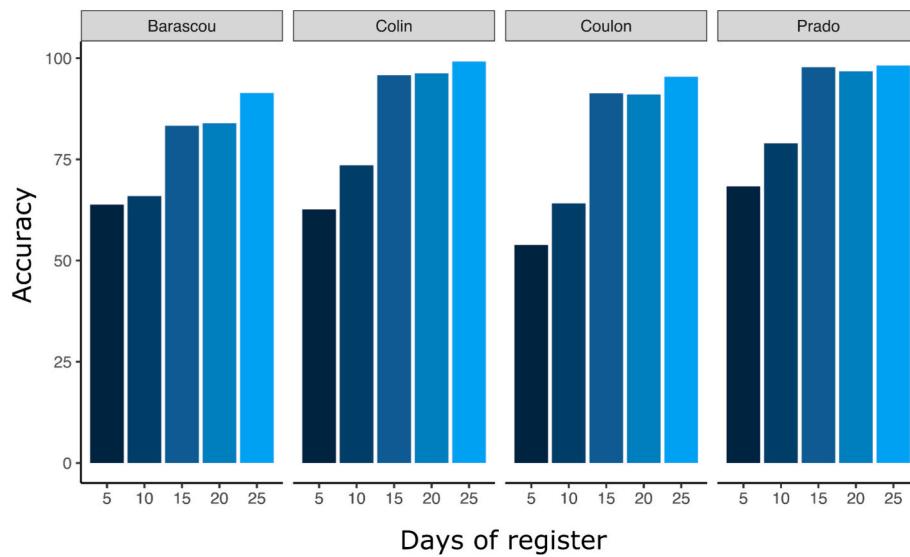
The model was designed with reproducibility and transferability in mind to guarantee consistent results and facilitate model deployment across various platforms, including embedded systems. A universally accepted pre-trained model, saved in the HDF5 file format, can be accessed from the GitHub repository listed in the Data and Source Code Availability section.

### 3. Results

Models trained with bees' life history traits (age, number of foraging trips and total amount of time spent outside per day) reached high accuracy values while classifying bees as control or pesticide-exposed (Fig. 4). When using each previously published dataset independently, accuracy values ranged from 91% to 99% after 100 epochs (Fig. 4A). Our results also show that combining two or more data sets increases accuracy (Fig. 4B). To determine the number of recording days the model needed to reach high accuracy values, a series of models were trained with a maximum of 5, 10, 15, 20, and 25 days of records per bee (Fig. 4C). According to our results, high accuracy values are reached starting at fifteen days of recording (Fig. 4C & 5).



**Fig. 4.** Accuracy of the training (red) and validation (blue) datasets. A) Accuracy of models for each previously published dataset. B) Accuracy of models trained with 1–4 data sets combined. C) Accuracy of models trained with a maximum of 5, 10, 15, 20, and 25 days of records per bee (all four datasets combined). Values at the bottom of the graphs indicate the final validation accuracy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** Accuracy of models trained with a maximum of 5, 10, 15, 20, and 25 records per bee.

**Fig. 6** shows the confusion matrices for the model predictions using all datasets combined, which demonstrates an increased performance of the model when larger subsets are used (25 daily records per subset). Correct predictions are shown on the main diagonal.

In addition, this model behavior can also be measured quantitatively using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, accuracy, precision, recall, and F-1 score (Sokolova and Lapalme, 2009). **Fig. 7** and **Table 2** illustrate how the model can improve performance by increasing the number of records in the disjoint sets, with AUC values ranging from 0.51 to 0.98, accuracy ranging from 0.68 to ~0.99, Cohen Kappa from 0.03 to 0.97, precision from 0.67 to 0.98, recall from 0.68 to 0.98, and F1-score from 0.58 to 0.99. The model is particularly good at identifying true positives (pesticide-exposed bees classified as such), and even with only five records (5 recording days), correctly classified 1394 out of 1413 pesticide-exposed bees (accuracy of 98.6%; **Fig. 6**).

**Fig. 8** underscores the predictive performance and scalability of the BiLSTM model as the number of time series increases. In **Fig. 8A**, a violin plot captures the variability and stability of the model's performance across the different datasets. This visualization highlights the model's consistent effectiveness, as evidenced by individual metric scores marked by red dots. **Fig. 8B** demonstrates that an increase in the time series leads to significant improvements in model precision and overall accuracy. Furthermore, **Fig. 8C** employs a Taylor diagram representing the model's accuracy and consistency. This diagram shows that as the number of time series increases, the performance points approach the reference line of a perfect model, highlighting the model's reliability and robustness. Collectively, these results confirm the BiLSTM model's aptitude for detecting pesticide-exposed bees and showcase its potential for practical application in real-time environments.

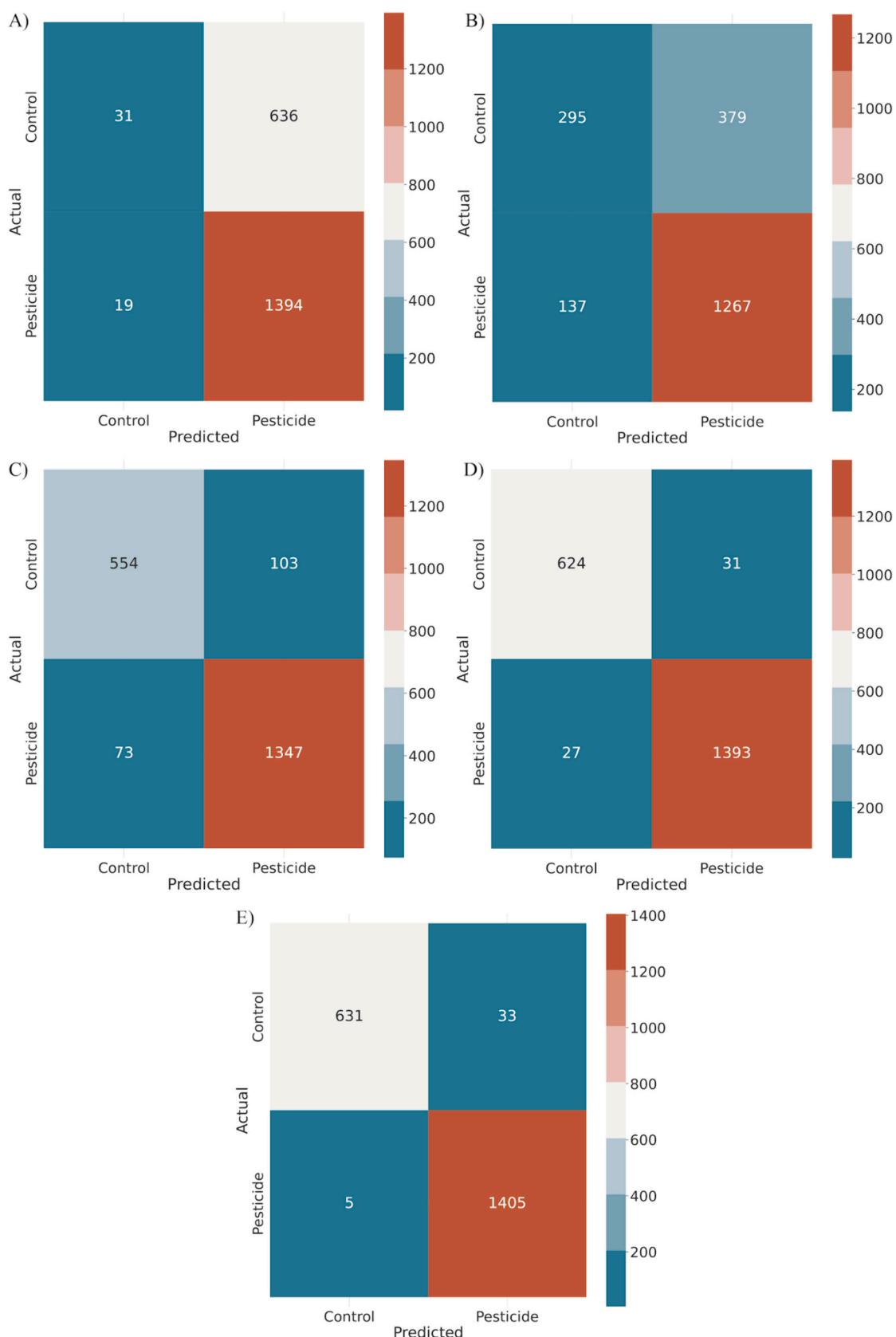
#### 4. Discussion

Within the current regulatory framework for pesticide risk assessment, pesticide registration applications must include data on the product's fate in the environment, as well as its toxicity for non-target organisms. Regulatory authorities can then assess the risk presented by the pesticide and determine to what extent it is safe for human and environmental health. As part of this risk assessment procedure, OECD (Organization for Economic Co-operation and Development) and EPPO (European and Mediterranean Plant Protection Organization) require toxicological data on honey bees (*Apis mellifera*), one of the most important pollinators worldwide (Hung et al., 2018; Klein et al., 2007).

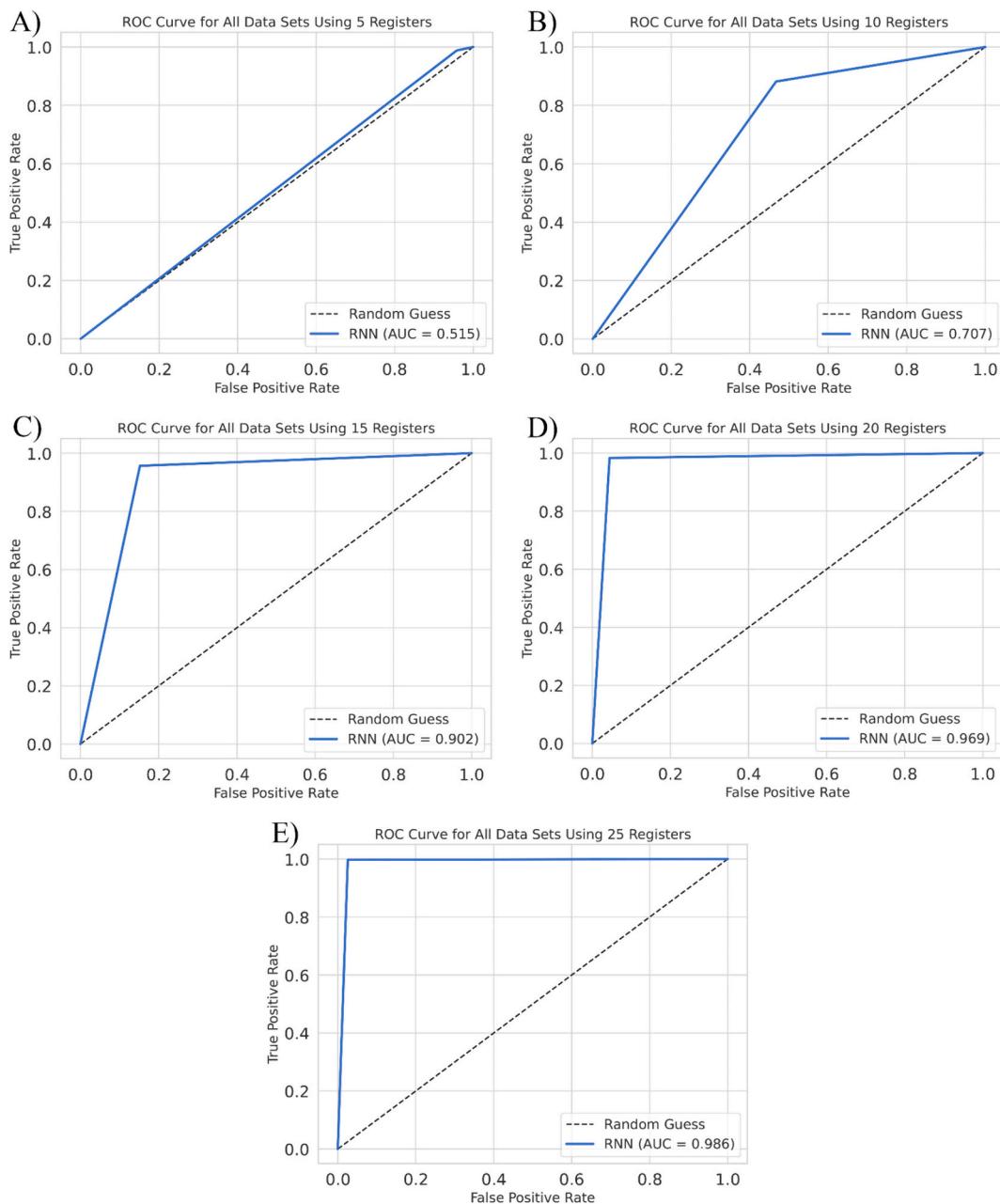
Semi-field and field surveys can notably be performed for the marketing authorization (MA), but also in post-MA studies monitoring adverse effects linked to the use of pesticides (i.e. phytopharmacovigilance). For that purpose, automatic monitoring devices at the entrance of beehives, such as RFID or OC, have proved to be valuable tools for assessing the ecotoxicological effects of pesticides on honey bees. For instance, different studies have found that exposure to low doses of neurotoxic pesticides can have lifelong effects on honey bee physiology and behavior (Decourtye et al., 2003; Kairo et al., 2017; Karahan et al., 2015; Zhang et al., 2020). While a variety of sublethal effects have been reported (i.e., hampered navigation, homing failure), reduced daily flight activity appeared to be a shared effect amongst different neurotoxic pesticide molecules (Barascou et al., 2022; Colin et al., 2019; Coulon et al., 2020; Karahan et al., 2015; Prado et al., 2019; Shi et al., 2020). In terms of the ecological services provided by honey bees, the reduction in the daily flight activity after pesticide exposure likely translates to fewer flowers being visited per day and, hence, a decrease in pollination.

Therefore, to protect honey bees and the pollination services they provide, having adequate tools for pesticide risk assessment and toxicovigilance in the field are paramount. Ideally, such tools should be able to inform us about the presence of harmful levels of pesticides in real time to assess pesticide risk as conditions change. In this paper, we proposed monitoring and analyzing in real-time, time-series data of honey bee flight activity to fill this gap. Neural Networks (NNs) have been extensively used to classify time-series data (Azad et al., 2018; Koh et al., 2021; Yao et al., 2019). Artificial intelligence approaches have spread into agricultural operations and are now used to classify animal behavior, identify unhealthy individuals, and guide livestock management (Kleanthous et al., 2022). CNNs have even been used to classify pesticides based on their chemical structure as potentially hazardous to honey bees (Wang et al., 2020). The use of artificial intelligence to monitor beehives and inform management practices is not novel but is still in its infancy (Cecchi et al., 2020; Troung et al., 2023). Sensors coupled with artificial intelligence models have been used to monitor hive weight, sounds emitted by the bees, temperature, humidity, and CO<sub>2</sub> inside the beehive, as well as weather conditions outside the hive (Cecchi et al., 2020; Troung et al., 2023). Here, we used a similar approach but within a context of toxicovigilance by coupling time-series data of individual bees from OC (flight activity) to an artificial intelligence model, in order to classify bees as healthy or pesticide-exposed.

In the domain of time-series analysis, BiLSTM networks have significantly enhanced the ability to decipher intricate patterns within



**Fig. 6.** Confusion matrices using all datasets for disjoint subsets of A) 5 days of records, B) 10 days of records, C) 15 days of records, D) 20 days of records, and E) 25 days of records per bee.



**Fig. 7.** Area Under the ROC curve of all datasets for disjoint subsets of A) 5 days of records, B) 10 days of records, C) 15 days of records, D) 20 days of records, and E) 25 days of records per bee. Increasing the number of recurring records increased the AUC metric.

**Table 2**

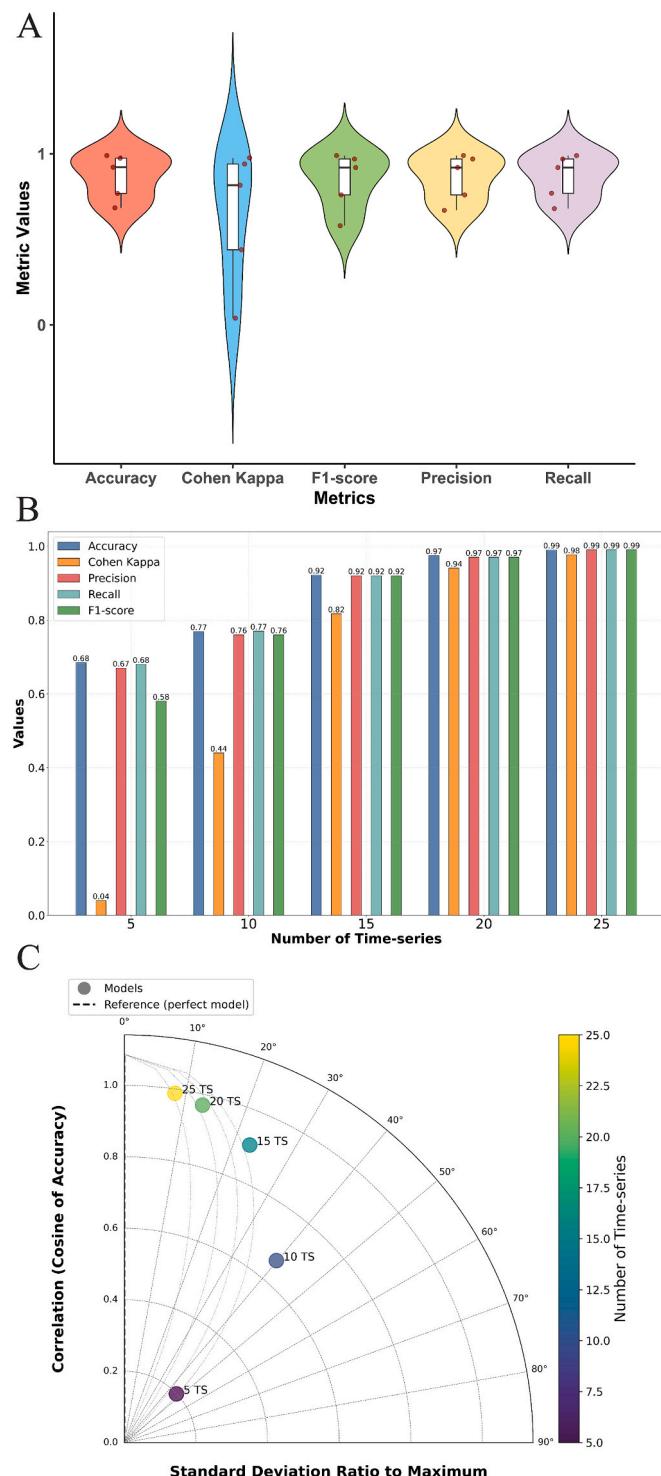
Performance metrics for the model were obtained by averaging the results from cross-validation tests across the four combined datasets.

Number of time-series	Accuracy	Cohen Kappa	Precision	Recall	F1-score
5	0.6846	0.0397	0.67	0.68	0.58
10	0.7685	0.4396	0.76	0.77	0.76
15	0.9220	0.8168	0.92	0.92	0.92
20	0.9744	0.9407	0.97	0.97	0.97
25	0.9898	0.9766	0.99	0.99	0.99

sequential data. Their bidirectional processing capability captures information from past and future data points, providing a comprehensive view of bee activity sequences. This detailed perspective is crucial for detecting subtle variations in bee behavior caused by pesticide exposure,

which traditional methods may overlook. The resilience of BiLSTM networks against the vanishing gradient problem allows them to retain long-term behavioral dependencies, a critical advantage over other neural architectures like Gated Recurrent Units (GRUs), which, while streamlining information flow, may miss crucial temporal variations. BiLSTM networks provided here a robust way to discern and predict the subtle effects of pesticides.

The model architecture in this study included two BiLSTM layers, complemented by a dropout regularization technique to mitigate overfitting and ensure a reliable generalization of unseen data. A classification layer of 128 neurons allowed the model to identify intricate patterns within the time-series data. This configuration proved to be effective in accurately predicting the effects of pesticides. Indeed, this model categorized bees with an accuracy exceeding 98%. It has been trained and validated across four distinct datasets, demonstrating high validation accuracies and its ability to generalize effectively the



**Fig. 8.** Comprehensive evaluation of predictive model performance across multiple metrics and time series. A) Violin plot for the distribution of the performance metrics: Accuracy, Cohen Kappa, F1-score, Precision, and Recall (the time-series data 5, 10, 15, 20 and 25 are represented by red dots). B) Bar plot across different performance metrics and numbers of time series data points (ranging from 5 to 25 days of records per bee) C) Taylor plot representing the standard deviation ratio to the maximum, cosine of accuracy correlation, and the number of time series data points (colour-coded). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictions. As depicted in Fig. 4-B, integrating multiple datasets further enhanced the model's stability and predictive reliability, achieving up to 99.08% validation accuracy. These findings reflect the model's robust learning dynamics, emphasizing its potential for real-time monitoring of pesticide effects. In particular, the number of time series significantly influenced the performance metrics of the BiLSTM model. As shown in Fig. 6, the confusion matrices revealed an improved ability of the model to differentiate between control and pesticide-exposed bees as the dataset size increased from 5 to 25 days of records. Notably, a significant reduction in false positives indicated a marked increase in the model's precision for correctly classifying unexposed bees. This improvement is also mirrored in Fig. 7, where the ROC curves display a progressive enhancement in the Area Under the Curve (AUC) metrics; starting at an AUC of 0.515 with just five records, the model's discrimination increased to 0.986 with 25 records.

Despite these advancements, a limitation remained in the two-week minimum of data recording required for the detection of pesticide effects. This lag between the exposure time and the symptom detection (reduced flight activity) is due to honey bee biology. Honey bees start their first orientation flights at around six days of age (Capaldi et al., 2000), and the average age at which bees start foraging is highly variable but has been estimated at  $\sim 20 \pm 7$  days (Prado et al., 2020). Accordingly, our model's results follow honey bee life history traits as its accuracy jumps suddenly at the two-week mark when bees switch to predominantly foraging each day (Fig. 1). Healthy bees exhibit more substantial increases in daily activity compared to neurotoxic pesticide-exposed bees, and the artificial intelligence model we have developed can detect differences in the sequences of the daily time spent outside and the daily number of flights for healthy or exposed bees. Employing a pre-trained model with transfer learning could however effectively address this lag-time issue in the detection of effects by significantly reducing the model's adaptation time to new data and enhancing its applicability in real-world scenarios (see Data and Source Code Availability section).

The proposed model is efficient and an excellent candidate for implementation into embedded systems because of its simplicity and size. Consequently, this technology can be quickly adopted into various contexts, such as MA and phytopharmacovigilance studies. Incorporating this classification model into a monitoring system (RFID or OC) would create a cutting-edge device that could analyze honey bee flight performances and assess the risk due to exposure to pesticides in real-time. The classification model could be embedded in a small computer coupled to the monitoring device and bee flight data could be curated and fed to the RNN model in real-time. The model could then make its predictions on hundreds of tagged bees and provide the user with an estimate of the percentage of bees exposed to neurotoxic pesticides. The utility of such a device should however be first validated using experimental set-ups, including RFID and OC technologies on a variety of neurotoxic pesticides.

The behavior of honey bees inside the hive is currently being studied using barcoded bees, automated monitoring devices, and artificial intelligence models to identify specific behaviors like trophallaxis (food exchange) (Gernat et al., 2023). In a recent study, artificial intelligence successfully distinguished flight events performed inside flight cages by either healthy or neurotoxic pesticide-exposed bumblebees (Chatzaki et al., 2023). However, to our knowledge, this study is the first to propose an artificial intelligence model to detect pesticide-exposed bees using their daily foraging activity in field conditions.

## 5. Conclusion

In response to the pressing need for practical tools to assess the ecological impacts of pesticides, our research introduces a pioneering approach using a BiLSTM-based artificial intelligence model. This model significantly improves our ability to detect the sublethal effects of neurotoxic pesticide exposure by monitoring honey bee flight activity in

real-time. Employing advanced neural network architectures to analyze complex time-series data has enabled precise classifications of bees as pesticide-exposed or healthy, based on their behaviors. Our findings highlight that the model's accuracy increases substantially with the volume of temporal data used for training, achieving a near-perfect accuracy with extensive datasets. Integrating this AI model into broader ecological monitoring initiatives could represent a significant advancement in environmental toxicology, offering a cost-effective, efficient method to detect neurotoxic pesticides in the environment and potentially helping pesticide management and conservation efforts. In conclusion, this model addresses urgent needs in honey bee protection by pioneering the use of AI tools in ecological and agricultural research.

### CRediT authorship contribution statement

**Ulises Olivares-Pinto:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Validation, Writing – original draft, Writing – review & editing. **Cédric Alaux:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Yves Le Conte:** Supervision. **Didier Crauser:** Data curation. **Alberto Prado:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

### Data availability

The Deep Learning architecture's activity data, source code, and pre-trained model are available in the following GitHub repository: <https://github.com/HpcDataLab/ToxicovigilanceTool>.

### Acknowledgments

The authors acknowledge the grants provided by the North American Pollinator Protection Campaign (NAPPC) Honey Bee Health Improvement Project and the CONAHCyT: 191982 (UOP).

### References

- Alaux, C., Crauser, D., Pioz, M., Saulnier, C., Le Conte, Y., 2014. Parasitic and immune modulation of flight activity in honey bees tracked with optical counters. *J. Exp. Biol.* 217 (19), 3416–3424.
- Azad, A.P., Meher, P.K., Rautaray, S.S., 2018. Time series classification using recurrent neural networks and learning vector quantization. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 598–603. <https://doi.org/10.1109/ICMLA.2018.00058>.
- Barascou, L., Brunet, J.L., Belzunges, L., Decourtey, A., Henry, M., Fourrier, J., Le Conte, Y., Alaux, C., 2021. Pesticide risk assessment in honeybees: using behavioral and reproductive performances as assessment endpoints. *Chemosphere* 276, 130134.
- Barascou, L., Requier, F., Sené, D., Crauser, D., Le Conte, Y., Alaux, C., 2022. Delayed effects of a single dose of a neurotoxic pesticide (Sulfoxaflor) on honeybee foraging activity. *Sci. Total Environ.* 805, 150351.
- Berkaya, S.K., Gunal, E.S., Gunal, S., 2021. Deep learning-based classification models for beehive monitoring. *Eco. Inform.* 64, 101353.
- Capela, N., Sarmento, A., Simões, S., Azevedo-Pereira, H.M., Sousa, J.P., 2022. Sub-lethal doses of sulfoxaflor impair honey bee homing ability. *Sci. Total Environ.* 837, 155710.
- Capaldi, E.A., Smith, A.D., Osborne, J.L., Fahrbach, S.E., Farris, S.M., Reynolds, D.R., Edwards, A.S., Martin, A., Robinson, G.E., Poppy, G.M., Riley, J.R., 2000. Ontogeny of orientation flight in the honeybee revealed by harmonic radar. *Nature* 403 (6769), 537–540.
- Cecchi, S., Spinsante, S., Terenzi, A., Orcioni, S., 2020. A smart sensor-based measurement system for advanced bee hive monitoring. *Sensors* 20 (9), 2726.
- Chatzaki, V., Montoro, M., El-Rashid, R., Jensen, A.B., Lecocq, A., 2023. A new approach for detecting sublethal effects of neonicotinoids on bumblebees using optical sensor technology. *Insects* 14 (8), 713.
- Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., Shelhamer, E., 2014. cuDNN: Efficient Primitives for Deep Learning:1–9. <http://arxiv.org/abs/1410.0759>.
- Colin, T., Meikle, W.G., Wu, X., Barron, A.B., 2019. Traces of a neonicotinoid induce precocious foraging and reduce foraging performance in honey bees. *Environ. Sci. Technol.* 53 (14), 8252–8261.
- Coulon, M., Dalmon, A., Di Prisco, G., Prado, A., Arban, F., Dubois, E., Ribière-Chabert, M., Alaux, C., Thiéry, R., Le Conte, Y., 2020. Interactions between thiamethoxam and deformed wing virus can drastically impair flight behaviour of honey bees. *Front. Microbiol.* 11, 766.
- Decourtey, A., Lacassie, E., Pham-Delègue, M.H., 2003. Learning performances of honeybees (*Apis mellifera* L) are differentially affected by imidacloprid according to the season. *Pest Manag. Sci. formerly Pest. Sci.* 59 (3), 269–278.
- Decourtey, A., Devillers, J., Aupinel, P., Brun, F., Bagnis, C., Fourrier, J., Gauthier, M., 2011. Honeybee tracking with microchips: a new methodology to measure the effects of pesticides. *Ecotoxicology* 20 (2), 429–437.
- Dobbin, K.K., Simon, R.M., 2011. Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med. Genet.* 4, 1–8.
- Garibaldi, L.A., Steffan-Dewenter, I., Winfree, R., Aizen, M.A., Bommarco, R., Cunningham, S.A., Kremen, C., Carvalheiro, L.G., Harder, L.D., Afik, O., Bartomeus, I., et al., 2013. Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* 339 (6127), 1608–1611.
- Gernat, T., Jagla, T., Jones, B.M., Middendorf, M., Robinson, G.E., 2023. Automated monitoring of honey bees with barcodes and artificial intelligence reveals two distinct social networks from a single affiliative behavior. *Sci. Rep.* 13 (1), 1541.
- Goulson, D., Nicholls, E., Botías, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347, 1255957.
- Hand, D.J., 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* 45, 071–186.
- Henry, M., Beguin, M., Requier, F., Rollin, O., Odoux, J.F., Aupinel, P., Aptel, J., Tchamitchian, S., Decourtey, A., 2012. A common pesticide decreases foraging success and survival in honey bees. *Science* 336, 348–350.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hung, K.L.J., Kingston, J.M., Albrecht, M., Holway, D.A., Kohn, J.R., 2018. The worldwide importance of honey bees as pollinators in natural habitats. *Proc. R. Soc. B Biol. Sci.* 285 (1870), 20172140.
- Johnson, R.M., 2015. Honey bee toxicology. *Annu. Rev. Entomol.* 60, 415–434.
- Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. *J. Big Data* 6 (1), 1–54. <https://doi.org/10.1186/s40537-019-0192-5>.
- Kairo, G., Poquet, Y., Hajji, H., Tchamitchian, S., Cousin, M., Bonnet, M., Pelissier, M., Kretzschmar, A., Belzunges, L.P., Brunet, J.-L., 2017. Assessment of the toxic effect of pesticides on honey bee drone fertility using laboratory and semifield approaches: a case study of fipronil. *Environ. Toxicol. Chem.* 36, 2345–2351. <https://doi.org/10.1002/etc.3773>.
- Karahan, A., Çakmak, I., Hranitz, J.M., Karaca, I., Wells, H., 2015. Sublethal imidacloprid effects on honey bee flower choices when foraging. *Ecotoxicology* 24, 2017–2025. <https://doi.org/10.1007/s10646-015-1537-2>.
- Karim, F., Majumdar, S., Darabi, H., Harford, S., 2019. Multivariate LSTM-FCNs for time series classification. *Neural Netw.* 116, 237–245. <https://doi.org/10.1016/j.neunet.2019.04.014>.
- Khan, M., Wang, H., Riaz, A., Elfatyan, A., Karim, S., 2021. Bidirectional LSTM-RNN-based hybrid deep learning frameworks for univariate time series classification. *J. Supercomput.* 77 (7), 7021–7045. <https://doi.org/10.1007/s11227-020-03560-z>.
- Kingma, D.P., Ba, J., 2017. *Adam: A Method for Stochastic Optimization* (arXiv: 1412.6980). arXiv. <http://arxiv.org/abs/1412.6980>.
- Kleanthous, N., Hussain, A.J., Khan, W., Sneddon, J., Al-Shamma'a, A., Liatsis, P., 2022. A survey of machine learning approaches in animal behaviour. *Neurocomputing* 491, 442–463.
- Klein, A.M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B* 274, 303–313.
- Koh, B.H.D., Lim, C.L.P., Rahimi, H., Woo, W.L., Gao, B., 2021. Deep temporal convolution network for time series classification. *Sensors* 21 (2). <https://doi.org/10.3390/s21020603>.
- Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. *Sensors* 18 (8), 2674.
- Malhotra, P., Vig, L., Agarwal, P., Shroff, G., 2017. TimeNet: Pre-Trained Deep Recurrent Neural Network for Time Series Classification. ArXiv.
- Marinho, R.S., Silva, A.A., Mastrangelo, C.B., Prestes, A.J., de Costa, L.Z.M., Toledo, C.F., Mastrangelo, T., 2023. Automatic parasitized fruit fly pupae classification from X-ray images by convolutional neural networks. *Eco. Inform.* 78, 102382.
- McLennan, K.M., Skillings, E.A., Rebelo, C.J., Corke, M.J., Moreira, M.A.P., Morton, A.J., Constantino-Casas, F., 2015. Validation of an automatic recording system to assess behavioural activity level in sheep (*Ovis aries*). *Small Rumin. Res.* 127, 92–96.
- Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814.
- Peng, Y., Wang, Y., 2022. CNN and transformer framework for insect pest classification. *Eco. Inform.* 72, 101846.
- Poquet, Y., Vidau, C., Alaux, C., 2016. Modulation of pesticide response in honeybees. *Apidologie* 47 (3), 412–426.
- Prado, A., Pioz, M., Vidau, C., Requier, F., Jury, M., Crauser, D., Brunet, J.L., Le Conte, Y., Alaux, C., 2019. Exposure to pollen-bound pesticide mixtures induces longer-lived but less efficient honey bees. *Sci. Total Environ.* 650, 1250–1260.
- Prado, A., Requier, F., Crauser, D., Le Conte, Y., Bretagnolle, V., Alaux, C., 2020. Honeybee lifespan: the critical role of pre-foraging stage. *R. Soc. Open Sci.* 7 (11), 200998.
- Rumelhart, D.E., Hinton, G.E., 1986. Learning representations by back-propagating errors. *Nature* 2, 3–6.
- Samson-Robert, O., Labrière, G., Chagnon, M., Fournier, V., 2014. Neonicotinoid-contaminated puddles of water represent a risk of intoxication for honey bees. *PLoS One* 9 (12), e108443. <https://doi.org/10.1371/journal.pone.0108443>.

- Santos, M.S., Soares, J.P., Abreu, P.H., Araujo, H., Santos, J., 2018. Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [research frontier]. *IEEE Comput. Intell. Mag.* 13 (4), 59–76. <https://doi.org/10.1109/MCI.2018.2866730>.
- Sgolastra, F., Medrzycki, P., Bortolotti, L., Maini, S., Porrini, C., Simon-Delso, N., Bosch, J., 2020. Bees and pesticide regulation: lessons from the neonicotinoid experience. *Biol. Conserv.* 241, 108356.
- Shi, J., Yang, H., Yu, L., Liao, C., Liu, Y., Jin, M., Yan, W., Wu, X.B., 2020. Sublethal acetamiprid doses negatively affect the lifespans and foraging behaviors of honey bee (*Apis mellifera* L.) workers. *Sci. Total Environ.* 738, 139924.
- Siviter, H., Bailes, E.J., Martin, C.D., Oliver, T.R., Koricheva, J., Leadbeater, E., Brown, M.J., 2021. Agrochemicals interact synergistically to increase bee mortality. *Nature* 596 (7872), 389–392.
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* 45 (4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15 (1), 1929–1958.
- Tealab, A., 2018. Time series forecasting using artificial neural networks methodologies : a systematic review. *Future Comput. Inform. J.* 3 (2), 334–340. <https://doi.org/10.1016/j.fcij.2018.10.003>.
- Troung, T.H., Nguyen, H.D., Mai, T.Q.A., Nguyen, H.L., Dang, T.N.M., Phan, T.T.H., 2023. A deep learning-based approach for bee sound identification. *Eco. Inform.* 78, 102274.
- Tuda, M., Luna-Maldonado, A.I., 2020. Image-based insect species and gender classification by trained supervised machine learning algorithms. *Eco. Inform.* 60, 101135.
- Wang, F., Yang, J.F., Wang, M.Y., Jia, C.Y., Shi, X.X., Hao, G.F., Yang, G.F., 2020. Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction. *Sci. Bull.* 65 (14), 1184–1191. <https://doi.org/10.1016/j.scib.2020.04.006>.
- Wang, M., Tausch, F., Schmidt, K., Diehl, M., Knaebe, S., Bargen, H., Materne, L., Groeneveld, J., Grimm, V., 2024. Honeybee pollen but not nectar foraging greatly reduced by neonicotinoids: insights from AI and simulation. *Comput. Electron. Agric.* 221, 108966.
- Yao, H., Cui, P., Wu, Y., Xu, W., Chen, S., Wang, X., Yan, J., 2019. Attention-based recurrent neural network models for time series classification. In: 2019 IEEE International Conference on Data Mining (ICDM), pp. 218–227. <https://doi.org/10.1109/ICDM.2019.00140>.
- Zhang, Z.Y., Li, Z., Huang, Q., Li, K., Yan, W.Y., Zhang, L.Z., Zeng, Z.J., 2020. Deltamethrin impairs honeybees (*Apis mellifera*) dancing communication. *Arch. Environ. Contam. Toxicol.* 78, 117–123. <https://doi.org/10.1007/s00244-019-00680-3>.
- Zioga, E., Kelly, R., White, B., Stout, J.C., 2020. Plant protection product residues in plant pollen and nectar: a review of current knowledge. *Environ. Res.* 189, 109873.