



# Machine Learning for Predicting Mycotoxin Occurrence in Maize

Marco Camardo Leggieri, Marco Mazzoni and Paola Battilani\*

Department of Sustainable Crop Production, Università Cattolica del Sacro Cuore, Piacenza, Italy

Meteorological conditions are the main driving variables for mycotoxin-producing fungi and the resulting contamination in maize grain, but the cropping system used can mitigate this weather impact considerably. Several researchers have investigated cropping operations' role in mycotoxin contamination, but these findings were inconclusive, precluding their use in predictive modeling. In this study a machine learning (ML) approach was considered, which included weather-based mechanistic model predictions for AFLA-maize and FER-maize [predicting aflatoxin B<sub>1</sub> (AFB<sub>1</sub>) and fumonisins (FBs), respectively], and cropping system factors as the input variables. The occurrence of AFB<sub>1</sub> and FBs in maize fields was recorded, and their corresponding cropping system data collected, over the years 2005–2018 in northern Italy. Two deep neural network (DNN) models were trained to predict, at harvest, which maize fields were contaminated beyond the legal limit with AFB<sub>1</sub> and FBs. Both models reached an accuracy >75% demonstrating the ML approach added value with respect to classical statistical approaches (i.e., simple or multiple linear regression models). The improved predictive performance compared with that obtained for AFLA-maize and FER-maize was clearly demonstrated. This coupled to the large data set used, comprising a 13-year time series, and the good results for the statistical scores applied, together confirmed the robustness of the models developed here.

**Keywords:** aflatoxins, *Aspergillus flavus*, cropping system, deep learning, *Fusarium verticillioides*, fumonisins, predictive models

## OPEN ACCESS

### Edited by:

Alicia Rodríguez,  
University of Extremadura, Spain

### Reviewed by:

Alejandro Hernández,  
University of Extremadura, Spain  
Naresh Magan,  
Cranfield University, United Kingdom

### \*Correspondence:

Paola Battilani  
paola.battilani@unicatt.it

### Specialty section:

This article was submitted to  
Food Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 30 January 2021

**Accepted:** 16 March 2021

**Published:** 09 April 2021

### Citation:

Camardo Leggieri M, Mazzoni M  
and Battilani P (2021) Machine  
Learning for Predicting Mycotoxin  
Occurrence in Maize.  
Front. Microbiol. 12:661132.  
doi: 10.3389/fmicb.2021.661132

## INTRODUCTION

Mycotoxin contamination of maize is a major concern worldwide (Eskola et al., 2020). The colonization of maize ears by *Aspergillus* section *Flavi* and *Fusarium* spp. can lead to ear rots whose impact on the amount of grain yield is minor or negligible yet their mycotoxin contamination levels are high; therefore, the main impact of mycotoxin producing fungi in maize regards grain safety and its compliance with the legal limits. Concerning those mycotoxins produced by *Aspergillus* section *Flavi*, among the aflatoxins (AFs), aflatoxin B<sub>1</sub> (AFB<sub>1</sub>) is classified by IARC (International Agency for Research on Cancer) as a class-1A, human carcinogen. Such AFs were first detected in Italy in the early 2000s (Piva et al., 2006; Battilani et al., 2008a), but since 2012 they have spread all over southeastern Europe, presumably aided by warmer and drier conditions during summer attributed to ongoing climate change (Dobolyi et al., 2013; Levic et al., 2013; Battilani et al., 2016), and their incidence and severity can vary markedly among years. *Fusarium* spp. can produce a wide range of mycotoxins, of which the fumonisins B<sub>1</sub>, B<sub>2</sub>, and B<sub>3</sub> (FBs)—predominantly produced by *F. verticillioides*—are the key ones reported in maize grain worldwide, thus posing a serious risk of possible human carcinogenicity (IARC, 1993). Other mycotoxins produced by *Fusarium* genus are the trichothecenes (TCTs) and zearalenone (ZEN), these being prevalent in temperate

and wet areas especially in rainy years, optimal conditions for their main producer, *F. graminearum* (Pietri et al., 2004).

These main mycotoxins threaten the maize supply chain worldwide, in all producing areas; nevertheless, the prevailing mycotoxin and level of contamination depends both on the growing area and year, intended as the meteorological conditions occurring during the crop growing season (Logrieco et al., 2021). Support for farmers coming from predictive modeling, using meteorological data as input variables (Battilani, 2016), has been pursued in Europe in the form of two mechanistic models for AFB<sub>1</sub> and FBs predictions: respectively, AFLA-maize (Battilani et al., 2013) and FER-maize (Battilani et al., 2003). They both aim to predict the risk of contamination above current legal limits in force in Europe (European Commission, 2006b, 2007), and they strongly support stakeholders in the maize chain management (Battilani and Camardo Leggieri, 2015; Battilani, 2016; Palumbo et al., 2020). However, mounting uncertainty of climate conditions and extreme events, often emphasized as issues in climate change, has recently increased the importance of deriving reliable predictions at the farm level (Camardo Leggieri et al., 2020a). Addressing the variability in mycotoxin occurrence among years and geographic areas, even those quite close to each other, in addition to the emerging issue of co-occurring mycotoxins (Camardo Leggieri et al., 2019; Giorni et al., 2019), will require making reliable predictions to support the maize chain management.

Weather variables are the leading factors contributing to mycotoxin occurrence, but the cropping system used is a powerful tool of farmers to mitigate grain contamination. Accordingly, several authors have studied the role of the cropping system and the rationale behind its impact on mycotoxin contamination (Munkvold, 2003; Battilani et al., 2008a; Blandino et al., 2009; Kos et al., 2013; Palumbo et al., 2020). A rationale crop rotation, leaving the land fallow (unsown with maize), is recommended to reduce mycotoxin contamination in maize fields, even if the impact on it cannot be readily demonstrated, especially in intensive maize-growing areas (Guo et al., 2005; Marocco et al., 2008; Munkvold, 2014). A significant impact of the season length of maize hybrids, frequently reported as FAO class (Food and Agriculture Organization classification), upon FBs and AFB<sub>1</sub> contamination was reported (Pietri and Bertuzzi, 2012). Scientist do not all agree on this statement (Battilani et al., 2008a; Mazzoni et al., 2011), but the number of days elapsed from sowing to harvest was positively related to mycotoxin contamination (Battilani et al., 2008a; Torelli et al., 2012). The sowing date has been confirmed to influence the likelihood and extent of mycotoxin contamination (Jones, 1981; Alma et al., 2005), with late sowing generally associated with a higher content of mycotoxins at harvest (Alma et al., 2005; Blandino et al., 2009; Mazzoni et al., 2011). Nonetheless, irrigation has a strong impact as well, particularly upon AFs occurrence (Palumbo et al., 2020). Both the severity of European corn borer (ECB) (Jones, 1981) and the use of insecticide treatments may also significantly affect contamination, especially from FBs (Alma et al., 2005; Saladini et al., 2008; Mazzoni et al., 2011). Harvest time, or rather the kernel moisture at harvest, can also be crucial, notably for AFs contamination of maize (Munkvold, 2003; Battilani et al., 2008a); in fact, AFs production increases significantly from maize

physiological maturity, when kernel moisture is lower than 28–30% (Payne et al., 1988; Giorni et al., 2016). Therefore, delays in maize harvest after that stage means giving the fungus time to efficiently increase the contamination. Then, keep kernel moisture below 14% is mandatory in the postharvest stages, from drying to the whole storage period (Danso et al., 2018).

The above research findings have contributed to developing guidelines for mitigating mycotoxin contamination, but a quantitative evaluation of cropping system's impact on mycotoxins remains an unresolved issue. The little work done so far to predict the effect of cropping system on mycotoxin contamination (Battilani et al., 2008b; Camardo Leggieri et al., 2015) is neither complete nor satisfactory; however, it does clarify that cropping factors cannot be considered in isolation and that applying conventional statistical methods is not suitable for the task (Battilani, 2016; Palumbo et al., 2020). Hence, alternative approaches should be explored and possibly used.

Machine learning's emergence, alongside big data technologies and high-performance computing, introduces new opportunities for data-intensive science in precision farming and sustainable agriculture (Liakos et al., 2018). ML is the scientific field in which machines are trained to learn without being strictly programmed (Samuel, 2000), which has three main categories: (1) supervised learning (SL), (2) unsupervised learning (UL), and (3) reinforcement learning (RL). The SL algorithms use a training data set of labeled data to infer a function that is used to map new data. The UL algorithms directly look at the data and learn patterns from them, without human supervision. Last is RL, the ML branch in which entities called “software agents” take action, in a specific context, to optimize a given function. These ML approaches are increasingly applied in different subject areas to solve complex problems, often those with many factors involved, to which agriculture is no exception. In fact, ML is used in a variety of contexts and all the three main categories are now applicable (Liakos et al., 2018; Elavarasan and Vincent, 2020). Recently, Liakos et al. (2018) reviewed the ML approach in agriculture, highlighting that ML models had been applied in the multi-disciplinary agri-technologies domain for crop management (61%), yield prediction (20%), and disease detection (22%), but never accounting specifically for mycotoxins' co-occurrence. In crop yield prediction, which depends on many different factors operating simultaneously, deep neural networks (DNN), a type of artificial neural network (ANN) for SL models, are the most used (Khaki and Wang, 2019; Khaki et al., 2019; Nevavuori et al., 2019; Niedbała, 2019). DNNs are also very useful for plant disease identification, which is done via convolutional neural networks (CNN), which is a specific DNN-architecture used for image recognition (Boulet et al., 2019). Mycotoxins are mainly detected via high-performance liquid chromatography (HPLC) and mass spectrometry; however, DNNs coupled with rapid analytical tools, such as the electronic nose or infrared attenuated total reflection spectroscopy, have been recently applied and found to improve the assessment reliability (Evans et al., 2000; Jia et al., 2019; Öner et al., 2019; Camardo Leggieri et al., 2020b).

Torelli et al. (2012), in the first example of ML applied to mycotoxins, performed a 2-year study (2007–2008) that included seven cropping system variables—FAO class, sowing and harvest

dates, crop duration, kernel moisture, ECB treatment, and irrigation—as input for an ANN to classify maize samples based on their contamination with FBs. A fair correlation between the predicted and observed contaminated samples was reported ( $R^2 = 0.67$  and  $R^2 = 0.57$ , respectively, for the training and validation data sets), so the approach seems promising.

Therefore, this study aimed to develop ML models by combining the AFLA-maize and FER-maize predictions, together with cropping system information, as input to improve the mycotoxin risk predictions for AFB<sub>1</sub> and FBs. For this, a 13-year data set was considered and two ML models, one for each type of mycotoxin, were trained and validated (5-fold cross-validation) and their respective performance discussed.

## MATERIALS AND METHODS

### Data Collection

The data used in this study came from several surveys managed in the Emilia Romagna region (northern Italy) during 2004–2018, partially published (Battilani et al., 2013; Camardo Leggjieri et al., 2015, 2020b). The protocol for data collection was the same both for the published and unpublished data.

Briefly, meteorological data were downloaded from the Emilia Romagna meteorological service, available on request for research applications. They were based on a grid of squares, each 5 km wide, that encompassed the Emilia Romagna territory; all sources (both meteorological stations and radar) are interpolated for each square, to deliver reliable data (Bottarelli and Zinoni, 2002). Hourly data on air temperature (T, °C), relative humidity (RH, %), and rain (R, mm), during the period of January through September, were downloaded. These squares were filtered, to locate those corresponding to the maize field site sampled.

Maize fields sampling was performed at harvest, managed between mid August and September, during the combine machine discharge, according to European Commission Regulation (EU) 401/2006 (European Commission, 2006a). Relevant cropping system data were collected in each georeferenced field, based on a questionnaire filled by farmers, supported by extension services. Empirical information of different site variables were collected: the type of soil (percentages of sand, clay, and silt), maize hybrid FAO class, preceding crop, type of tillage, sowing date, plants per m<sup>2</sup>, silk emission date, harvest date, damage caused by hail or wind and ECB, fertilization type and dose, the number of irrigation intervention with relative volumes of water used, pest and disease control practices, and kernel moisture at harvest. Mycotoxin analysis was performed for all the sampled fields according to Bertuzzi et al. (2012) for the AFs [limit of detection (LOD): 0.05 µg/kg and limit of quantification (LOQ): 0.15 µg/kg], and by following Pietri and Bertuzzi (2012) for the FBs (LOD: 10 µg/kg and LOQ: 30 µg/kg).

### Data Processing

AFB<sub>1</sub> and FBs content allowed in maize grain, according to legal limits, were used as a threshold to separate the field samples in two classes: (1) *contaminated*, consisting of those

samples equal or exceeding the respective legal limit; (0) *non-contaminated* comprising all samples below the legal limit. Thresholds were therefore set to 5 µg/Kg for AFB<sub>1</sub> and 4,000 µg/Kg for FBs (FB<sub>1</sub> + FB<sub>2</sub>), the legal limits currently set by the European Commission for unprocessed maize destined for human consumption (European Commission, 2006b, 2007).

Meteorological data were used as input for the two predictive models, AFLA-maize and FER-maize, and cumulative risk indexes were obtained as the output, AFI for AFB<sub>1</sub> and FK for FBs, for each station and year. DNN models were implemented within the frame of Scikit-Learn (v0.21.3) in the Python module library (Pedregosa et al., 2011).

### Input Features

After the exclusion of those variables with many missing data points, eight different variables were considered as input for the ML approach; sowing date and harvest date were grouped on a per week basis. Of these eight, five variables were categorical (maize hybrid FAO class, preceding crop, sowing week, harvest week, ECB damage; **Table 1**) and three were continuous variables (growing days, days from crop sowing to harvest, calculated variable, kernel moisture at harvest, and mycotoxin cumulative indices AFI and FK, the output of predictive models).

For each continuous variable, its average ( $\mu$ ) and standard deviation ( $\sigma$ ) were computed for the data standardization, using Eq. (1).

$$X = \frac{X_0 - \mu}{\sigma} \quad (1)$$

**TABLE 1** | Summary of categorical data used for the two pathosystems analyzed: *A. flavus*-maize and *F. verticilloides*-maize.

Variable	N. of categories	Categorical value	Integer encoding
Maize hybrid FAO class	4	200–300	1
		400	2
		500	3
		600–700	4
Preceding crop	3	arable crops	1
		small grain	2
		maize	3
Sowing week*	4	10–12	1
		13–14	2
		15–16	3
		17+	4
Harvest week*	4	32–35	1
		36–37	2
		38–39	3
		40+	4
Severity of ECB attack	3	No/Minor-damage	1
		Medium damage	2
		Severe damage	3

Variables (hybrid FAO classes, preceding crop, and sowing and harvest weeks of the year (Julian date) and European corn borer [ECB] damage) and numbers of utilized categories are reported.

\*n<sup>th</sup> week of the year.

Even if this procedure is entirely facultative in neural network training, it is nonetheless useful for reducing the variance, speeding up the computational process, and improving the model's accuracy (Jin et al., 2015).

To deal with categorical data, the integer encoding procedure was applied. This assigns to a specific category an integer value that ranges from 1 to N, where N is the last category.

## Deep Neural Network (DNN) Development

A typical ANN consists of a network of connected computational units called neurons; these units are organized in layers, with input data passing through the network, and an activation function used to produce an output. DNNs are a particular class of ANNs in which, between the input and the output layer, there is an arbitrary number of hidden layers. The fully connected architecture has been adopted, meaning that each neuron in each layer is connected to each neuron in the next adjacent layer.

The development of a DNN model is a two-step process: (i) training and (ii) validation. The final aim of the training is to minimize a given error function by using an optimization algorithm. The training phase ends when the error converges to a pre-determined value, or when it does not decrease for a specific number of cycles, both decided *a priori* by the user (Camardo Leggieri et al., 2020b). A "batch training" mode was applied in this work. Briefly, during the training phase, training data (i.e., the mycotoxin contamination data in this study), were split into subgroups called *batches*. Neural network weights are updated when every sample inside a batch passes through the network. The iteration ends when all batches have passed through the network. After each iteration, a penalization term, called the weight decay (L2 regularization term), is introduced into the model to avoid overfitting the model to the data.

The final output is the result of a linear or non-linear activation function. In our study, a non-linear activation function between the input layer and the hidden layers, called Rectified Linear Unit (ReLU, Eq. 2), was applied:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (2)$$

where,  $x$  represents the weighted sum in a given input to a neuron (Dahl et al., 2013; Zeiler et al., 2013; LeCun et al., 2015).

The activation function used between the last hidden layer and the output layer (classification function) took the logistic form (Eq. 3):

$$f(x) = \frac{1}{1 + e^{-\vartheta \sum_{i=1}^N x_i w_{ij}}} \quad (3)$$

where,  $j$  is relative to the  $j^{\text{th}}$  output neuron, and  $i$  represents the  $i^{\text{th}}$  input neuron. The numerical result is between 0 and 1, for which 0.5 served as a threshold to discriminate between the positive and negative classes.

Different DL models were tested following a grid search procedure, done as described in Camardo Leggieri et al. (2020b).

Briefly, each hyperparameter was tested with a combination of every other hyperparameter. The hyperparameters tested were weight decay, number of hidden layers, number of neurons per hidden layer, and the optimization algorithm. In this work, two algorithms were tested; the first approximates the Broyden-Fletcher-Goldfarb-Shanno algorithm and is called LBFGS (Byrd et al., 1995; Kingma and Ba, 2014), while the second is an optimization of the classical stochastic gradient descent, called Adam (Kingma and Ba, 2014). Matthew's correlation coefficient (MCC) and accuracy were used as metrics to select the best combination of hyperparameters, for both NN models relevant to the pathosystem *A. flavus*-maize (DNN-*A. flavus*-maize) and *F. verticilloides*-maize (DNN-*F. verticilloides*-maize).

## DNN Validation

As in Camardo Leggieri et al. (2020b), both AFB<sub>1</sub> and FBs original datasets were splitted into two "sub-dataset." These four "sub-datasets" (two for AFB<sub>1</sub> and two for FBs) were generated by random sampling, but keeping the proportion of contaminated vs. non-contaminated samples constant. The first "sub-data set" accounted for 75% of the original data set and was used to perform a 5-fold cross-validation (CV). The other, called the "blind set" in this work, accounted for the rest of the original data set (25% of the original data set), and was used for a further validation of the models. The goodness-of-fit of each DNN-*A. flavus*-maize and DNN-*F. verticilloides*-maize model was assessed by computing several statistical scores, which were also applied to the blind set:

- True positive and true negative rates (TPR, TNR; Kohavi and Provost, 1998);

$$TNR = \frac{TN}{TN + FP} \quad (4)$$

where, TP and TN denote the number of true positives and true negatives, respectively, and likewise FP and FN denote the number of false positives and false negatives.

- Positive predictive value (PPV): the PPV (Eq. 5) index represents the proportion of positives samples identified as true positives. It ranges from 0 to 1 (Kohavi and Provost, 1998)

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

- Receiver operator characteristic (ROC) curve and area under the curve (AUC): the ROC curve and its AUC measure the quality of a binary classifier: the higher the area under the curve, the better the model performs. The AUC value ranges from 0 to 1 (Bradley, 1997).
- The MCC (Eq. 6) is used to assess the quality of binary classification (Matthews, 1975). This index takes into consideration TPR, TPN, and both false discoveries (false positives and negatives). The MCC ranges between -1 (complete disagreement between predicted and observed values) and +1 (perfect agreement). The MCC is considered a balanced measure, and it can be used



**TABLE 2** | Descriptive statistics of aflatoxin B1 (AFB<sub>1</sub>) and fumonisins (FBs, intended as the sum of FB<sub>1</sub> + FB<sub>2</sub>) levels of contamination (μg/kg) in maize grain samples collected in Emilia Romagna, Italy, over the years 2005–2018 (with some exceptions both for AFB<sub>1</sub> and for FBs).

Data set	Year	N.*	%positives <sup>§</sup>	Mean	StDev	Minimum	Maximum
<b>AFB<sub>1</sub></b>							
	2005	70	41.4	13.8	29.65	<0.05	154.9
	2006	25	24.0	18.8	52.78	<0.05	258.3
	2007	29	27.6	8.34	18.50	<0.05	68.43
	2008	40	40.0	9.11	16.99	<0.05	93.79
	2009	31	29.0	23.3	88.51	<0.15	494.3
	2010	35	28.0	14.4	36.60	<0.05	173.3
	2011	31	12.9	14.9	61.03	<0.05	334.8
	2014	26	23.1	10.2	23.17	<0.15	93.77
	2015	15	20.0	11.2	32.34	<0.05	129.3
	2016	20	45.0	30.4	62.33	0.42	208.3
	2017	28	50.0	22.3	29.15	<0.05	116.2
	2018	28	25.0	5.76	13.09	<0.05	65.00
	Total	378	35.54	14.66	42.59	<0.05	494.3
<b>FB<sub>1</sub> + FB<sub>2</sub></b>							
	2009	31	16.1	2,721.93	2,423.335	139.3	8,829.7
	2010	36	40.0	3,975.79	3,221.642	142.7	12,637.0
	2011	30	6.67	2,344.87	3,689.235	74.0	21,007.0
	2014	45	84.4	17,586.88	18,300.650	1718.4	106,053.5
	2016	21	19.0	3,035.33	3,251.853	204.3	14,020.8
	2017	29	13.8	2,643.91	3,107.761	<10.0	14,767.4
	2018	33	21.2	3,846.22	6,298.140	51.8	29,632.7
	Total	225	38.22	6,029.34	10,566.241	<10.0	106,053.5

The percentage of positive samples above the legal limits (5 and 4,000 μg/kg for AFB<sub>1</sub> and FBs, respectively; European Commission, 2006a, 2007) is reported.

\*N. = number of samples collected in the year; § % of positive = % of fields with contamination above the legal limits.

even if the two classes differ in size (Boughorbel et al., 2017).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

All the scores were computed using a home-built Python (v3.6.9) script that implemented the equations reported above. The ROC curves and AUCs, the grid search procedure, the MCCs, and the DNN architecture were implemented in the framework of scikit-learn (v0.23.2; Pedregosa et al., 2011).

The output indexes of the two mechanistic models (AFI and FK) were used to classify the blind data set, as described in Battilani et al. (2003) for the FBs and in Battilani et al. (2013) for AFB<sub>1</sub>. Finally, the results were compared using the classification obtained by the two DNNs.

## RESULTS

A total of 378 and 225 samples were included in the *A. flavus*-maize and *F. verticilloides*-maize data sets, respectively (Table 2). No data were retrieved for FBs before 2009 or during 2012 and 2013. Concerning the AFB<sub>1</sub>, data for it were not retrieved for

the years 2012 and 2013. The sample sizes per year were slightly different for the two data sets because of this missing data.

Fields with AFB<sub>1</sub> contamination above 5.0 μg/kg were found, whose incidence and mean values differed across the considered years. The highest amount of AFB<sub>1</sub> was found in 2009, at 494.3 μg/kg. In all years AFB<sub>1</sub> was greater than LOQ, but lower than LOD, except in 2009, 2014, and 2016. Regarding the incidence of fields found positive for AFB<sub>1</sub>, this was highest at 50% in 2017 and the lowest (12.9%) in 2011 (Table 2). In the whole maize data set, for AFB<sub>1</sub>, the mean incidence of positive samples was 32.9%.

Fields with FBs' contamination above 4,000 μg/kg were found in all the years considered, but this incidence differed across all years. The only year when the FBs was below the LOD was in 2017. The highest amount of FBs (106,053.5 μg/Kg) and the highest incidence of positive samples (89.8%) were both detected in 2014. The lowest incidence of FBs was 6.7%, scored in 2011 (Table 2). Considering the FER-maize data set as a whole, the overall mean incidence of positive samples, above the legal limit, was 32.2%.

Means and standard deviations computed for AFI and FK, the kernel moisture, and the growing days, are reported in Table 3.

Categorical data were grouped into three or four categories (Table 1). The FAO class was represented by four categories: 200–300, 400, 500, and 600–700. The preceding crop was represented by three categories: arable crops, small grain, and maize. The sowing and harvest weeks both accounted for four categories. Considering the sowing week, category #1 was assigned to weeks 10–12 of the year, #2 to weeks 13–14, #3 to weeks 15–16 and #4 to week ≥ 17. For the harvest week, those weeks of the year from 32 to 35 were designated category #1, and likewise 36–37 to #2, 38–39 to #3, with all weeks > 40 assigned to #4. The damage caused by ECB was divided into three categories: no damage and small damage were grouped into category #1, medium damage was assigned to category #2, and severe damage was assigned to category #3 (Camardo Leggieri et al., 2015).

Standardized continuous data were joined to categorical data, to form the neural network's input vector; thus, the final input array was formed by five encoded and three continuous variables (Tables 1 and 3), respectively.

The generation of the "sub-datasets" included a total 283 (75%) and 95 (25%) samples in the CV and blind set respectively

**TABLE 3** | Basic statistics of the continuous data included as the input into the model for the two pathosystems, *A. flavus*-maize and *F. verticilloides*-maize.

Model	Variable	Mean	StDev	Maximum	Minimum
<i>A. flavus</i> -maize	AFI index	2,906.17	2,226.792	8,944.5	11.6
	Kernel moisture (%)	20.66	3.541	31.5	11.9
	Growing days	158.0	16.68	234	66
<i>F. verticilloides</i> -maize	FK index	246,407.4	626,994.91	537,8545.3	2,102.3
	Kernel moisture (%)	19.99	3.712	31.5	11.8
	Growing days	156.6	16.30	207	115

for AFB<sub>1</sub>, and 169 and 56 samples in the CV and blind set respectively for FBs.

## Deep Neural Network (DNN) Training

The two DNNs were trained to be able to predict the content of AFB<sub>1</sub> and FBs, respectively. The NN-*A. flavus*-maize model was implemented with one hidden layer consisting of 80 neurons, a ReLU activation function, and an L2 regularization term of 0.0001. The parameters of that model were updated using the Adam algorithm (Kingma and Ba, 2014). By contrast, the NN-*F. verticilloides*-maize model's sole hidden layer had 50 neurons, a ReLU activation function, and an L2 regularization term of 0.1. Its parameters were updated using the LBFGS algorithm (Byrd et al., 1995; Table 4). The two developed NN models were validated using both the 5-fold cross-validation and a blind data set.

## DNN Validation

Cross-validation ROC curves and their relative AUCs were computed for the two NN models, to assess the quality of the two classifiers (Figures 1A,B). The 5-fold-cross-validation for NN-*A. flavus*-maize achieved an accuracy of  $66.56 \pm 3.381$  (mean  $\pm$  SE), and even higher at 78.94 for the blind data set. An MCC of  $0.10 \pm 0.157$  and  $0.49$  were achieved by the cross-validation and blind data set, respectively. The AUC and the TPR averaged  $0.58 \pm 0.063$  and  $0.08 \pm 0.073$  during the model's cross-validation. The model scored an AUC of 0.64 and a TPR of 0.42 when tested against the blind data set (Table 5).

The NN-*F. verticilloides*-maize model's 5-fold cross-validation attained an accuracy of  $69.63 \pm 10.892$ ; the accuracy was 79.31% using the independent data set. A TPR of  $0.53 \pm 0.118$  and  $0.65$  were, respectively, achieved by cross-validation and blind data set, respectively. Moreover, the model had an AUC of

$0.72 \pm 0.103$  and an MCC of  $0.35 \pm 0.229$  during the cross-validation phase, with corresponding values of 0.75 and 0.56 when tested against the unseen data set (Table 4).

Finally, to check whether the new approach represented a major step forward in the prediction of mycotoxin contamination in maize, our two DNN-models were compared with two counterpart mechanistic models, both run with the whole available data set. The resulting confusion matrix (Table 6) shows the performances of the NN-*A. flavus*-maize and NN-*F. verticilloides*-maize models and those of the two mechanistic models vis-à-vis the blind data set. The NN-*A. flavus*-maize model correctly estimated about 78% of samples (14% true positives, 64% true negatives). The wrong classification accounted for 19% of them being underestimated and 3% overestimated. The NN-FER-maize model correctly classified approximately 75% of the data set (25% true positives, 50% true negatives), with underestimations and overestimations amounting to 15 and 11%, respectively. In stark contrast, the AFLA-maize model correctly predicted just ~53% of samples (11% true positives, 42% true negatives). Further, a wrong classification accounted for more underestimations (22%) and overestimations (25%). Similarly, the FER-maize model correctly classified only ~52% of samples (31% true positives, 20% true negatives), but its wrong classifications included fewer (7%) underestimated cases being more prone to overestimations (41%).

## DISCUSSION

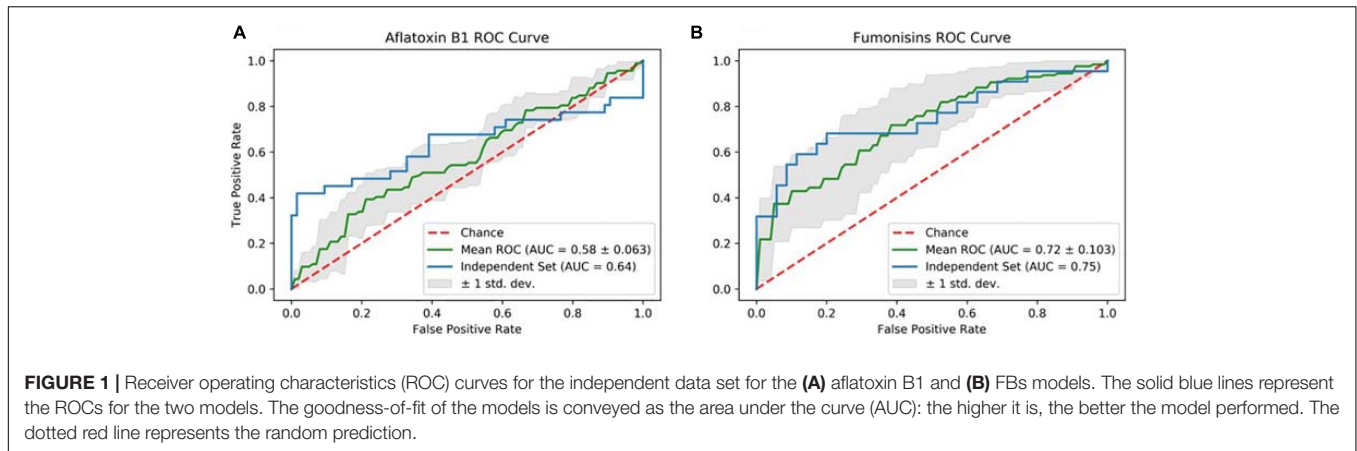
Maize is exposed to mycotoxins, which threaten human and animal health, and represent the major non-tariff trade barrier for agricultural products, negatively affecting the income of small-holder farmers and disrupting regional and international trade (Palumbo et al., 2020; Logrieco et al., 2021). Timely identification of contaminated lots is not a trivial challenge since mycotoxin contamination relies on several factors, including meteorology and how farmers manage the crop during the season and in the postharvest stages of storage and distribution (Munkvold, 2014; Logrieco et al., 2021). Different methodologies for the rapid detection of mycotoxin contamination are currently available (Öner et al., 2019; Camardo Leggieri et al., 2020b), but since they are applied at harvest or postharvest stages, they offer no support for taking preventive action and for optimizing lot use and management. On the contrary, farmers can benefit from model predictions of the risk of mycotoxin occurrence above the legal limit, when delivered before or during the cropping season, in the form of risk maps or risk indexes. Therefore, predictive modeling has garnered mounting interest over the last two decades (Battilani, 2016). Predictions refer to maize at harvest and it is assumed that the postharvest management guarantee a rapid grain drying to humidity  $\leq 14\%$ , kept stable during storage, to avoid fungal activity and further mycotoxin production.

Meteorological factors jointly determine whether fungi can grow and produce toxins, while the site's cropping system modulates the amount of contamination that ensues (Battilani and Camardo Leggieri, 2015). The former are the driving

**TABLE 4** | Hyperparameter values used to implement the neural networks *A. flavus*-maize (NN-*A. flavus*-maize) and *F. verticilloides*-maize (NN-*F. verticilloides*-maize) models.

Hyperparameter	NN- <i>A. flavus</i> -maize	NN- <i>F. verticilloides</i> -maize
Number of input neurons	7	7
Number of hidden layers	1	1
Number of neurons per hidden layer	80	50
Activation function input–hidden	ReLU (Eq. 1)	ReLU (Eq. 1)
Activation function hidden–output	Logistic (Eq. 2)	Logistic (Eq. 2)
L2 regularization term	0.0001	1.024
Parameters update algorithm	Adam	LBFGS

The number of input neurons is defined as the sum of the categorical and continuous data. The hyperparameter values were selected by following a grid searching procedure.



**FIGURE 1 |** Receiver operating characteristics (ROC) curves for the independent data set for the **(A)** aflatoxin B1 and **(B)** FBs models. The solid blue lines represent the ROCs for the two models. The goodness-of-fit of the models is conveyed as the area under the curve (AUC): the higher it is, the better the model performed. The dotted red line represents the random prediction.

**TABLE 5 |** Classification results summary for the prediction of aflatoxin B<sub>1</sub> (AFB<sub>1</sub>) and fumonisins (intended as the sum of FB<sub>1</sub> + FB<sub>2</sub>, FBs) in the maize samples.

	CV	Blind data set (DNN-models)	Blind data set (mechanistic models)
<b>AFB<sub>1</sub></b>			
ACC	66.56 ± 3.381	78.94	52.63
TPR	0.08 ± 0.073	0.42	0.32
TNR	0.94 ± 0.053	0.96	0.63
PPV	0.59 ± 0.424	0.90	0.29
MCC	0.10 ± 0.157	0.49	-0.051
<b>FBs</b>			
ACC	69.63 ± 10.892	79.31	52.63
TPR	0.53 ± 0.118	0.65	0.81
TNR	0.80 ± 0.122	0.88	0.34
PPV	0.63 ± 0.175	0.78	0.44
AUC	0.72 ± 0.103	0.75	n.c.
MCC	0.35 ± 0.229	0.56	0.17

The deep neural networks (DNNs) considered were evaluated using a 5-fold cross-validation and a blind data set. Cross-validation (CV) results were reported as an average value and related standard deviation (Avg ± Stdev). ACC, accuracy; MCC, Matthew's correlation coefficient; AUC, area under the curve; TPR, true positive rate; TNR, true negative rate; PPV, positive predictive value; n.c., not computed.

variables for predictive modeling, whereas the latter are rarely included, especially in mechanistic models. This omission is starting to gravely limit the reliability of predictions; in fact, during the last two decades, the typical cropping system has changed significantly due to the knowledge transfer from scientists to farmers; farmers are now following the guidelines to optimize crop management and mitigate mycotoxin contamination, with good results so far in term of a reduced mycotoxin occurrence. Both the meteorological data and the cropping system data have been used before as model inputs, to predict the content of mycotoxin in maize at the time of its harvest (Battilani et al., 2003, 2008a; Bertuzzi et al., 2014; Camardo Leggieri et al., 2015), yet they were used independently and only supported by basic statistical approaches. The aim of this study was to evaluate how

combining cropping system information with mechanistic predictive models could support the sought-after improvement in prediction performance.

Here an ML approach was developed using the AFLA-maize and FER-maize outputs (mycotoxin risk indexes) combined with cropping system information—this being known to significantly influence mycotoxin contamination in maize according to other studies (Palumbo et al., 2020; Logrieco et al., 2021)—as input variables. Other crop-related variables should have been included, like fertilization, irrigation, and pest control (Mazzoni et al., 2011; Munkvold, 2014); however, we excluded them because this data was largely unavailable to us. Moreover, the geolocation of maize fields was excluded as an input variable in our modeling; actually, even when the field location is known to be relevant (Torelli et al., 2012; Camardo Leggieri et al., 2015), the idea of this work was to obtain models applicable at a global level, without geographical constraints. When combining weather data and cropping system no information is lost, even when the maize fields' geolocation is excluded.

The predictions of NN-A. *flavus*-maize and NN-F. *verticilloides*-maize, the two neural network models developed in this study, were capable of an accuracy approaching ~78% for AFB<sub>1</sub> and ~79% for FBs, with a good correlation between

**TABLE 6 |** Confusion matrix computed from the blind data set results for the predicted and observed values of aflatoxin B<sub>1</sub> (AFB<sub>1</sub>) and fumonisins (intended as the sum of FB<sub>1</sub> + FB<sub>2</sub>, FBs). The predicted vs. observed results are reported as percentages.

DNNs	Predicted		
	Observed	Negative	Positive
AFB <sub>1</sub>	Negative	65	2
	Positive	19	14
FBs	Negative	53	7
	Positive	14	26
<b>Mechanistic models</b>			
AFLA-maize	Negative	42	25
	Positive	22	11
FER-maize	Negative	21	41
	Positive	7	31

predicted and observed data. This result is supported by the MCC results, which reached values of 0.49 and 0.56 when computed for the NN-A. *flavus*-maize and NN-F. *verticilloides*-maize, and by their corresponding AUC of 0.64 and 0.75 for NN-A. *flavus*-maize and NN-F. *verticilloides*-maize. Both AFLA-maize and FER-maize, the mechanistic models which served as the starting point of our investigation, achieved accuracies one-third lower, of about 50%, and their respective MCC was very close to 0; this indicates that their predictions were comparable to random one, when based on the same data set (i.e., the blind data set) used for NN model evaluation. It is, therefore, evident the proposed ML approach significantly improved the prediction of mycotoxins' content across the studied maize fields, making the successful use of this tool to detect maize grain not compliant with the current legal limit in Europe now more realistic and feasible to implement. Nevertheless, the NN-F. *verticilloides*-maize model performed better than NN-A. *flavus*-maize; apparently, it is easier to predict levels of FBs than AFB<sub>1</sub> as contaminants. The reason for this is not entirely clear, but the very low limit fixed by the legislation for AFB<sub>1</sub> could surely play a role. Further, irrigation is known to be very relevant for AFB<sub>1</sub> contamination, and grain humidity at harvest too, but they were excluded as input variables because of many missing data and this has surely a considerable impact on the predictive capacity of the model.

The NN-models were developed based on a large data set, one that included data collected from over 12 different years, with 378 samples in the AFB<sub>1</sub> data set and 225 samples in the FBs data set. The large data sets used, and the range of years considered, including the period of significant changes in the cropping system, strongly support the robustness of NN-A. *flavus*-maize and NN-F. *verticilloides*-maize models and their promising utility as a tool to support farmers in their decision-making. Future applications in other pathosystems is also foreseeable, as previously done for AFLA-maize (Kaminariis et al., 2020).

## CONCLUSION

To conclude, despite omitting some relevant cropping system variables, a substantial improvement at correctly predicting maize fields contaminated with mycotoxins above their legal limits was gained. Further improvements should be obtained by optimizing the data collection. Solving the missing data problem might be an easy task in the future, once scientists succeed in convincing farmers of the crucial role they can play in such data collection and also of the added value of predictive models in mycotoxin management. This will be a matter of building with the maize chain stakeholders a knowledge exchange approach and make them more involved compared to what actually happened in the past. Another aspect of improvement could be gleaned through our modeling approach, that of emerging issues related to mycotoxins. The main example of this concerns the recent report of co-occurring *A. flavus* and *F. verticilloides* in maize ears due to climate change effects, resulting in their complex interaction, with their dominance alternating during the growing season (Camardo Leggieri et al., 2019, 2020a; Giorni et al., 2019). Looking ahead, we anticipate

that elucidating the impact of these interactions between co-existing fungi upon mycotoxin production in maize will become crucial. Data collection to develop a joint model for the prediction of AFB<sub>1</sub> and FBs, including the impact of fungi co-occurrence, is ongoing and in the next future it is expected to contribute to a step up in mycotoxin prediction, possibly joining also NN-A. *flavus*-maize and NN-F. *verticilloides*-maize in a NN-mycotoxin predictive model.

The current work represents a notable step forward in modeling and predicting mycotoxins in crops. We retrieved evidence that ML can effectively combine cropping system data and meteorological data, thereby improving the accuracy and robustness of predictions. Big data is a relatively new concept for agriculture and plant disease research and management, but massive volumes of data with several components that interact within the pathosystem can also be captured in this context, and their elaboration can enhance the decision-making process (Wolfert et al., 2017). Applying machine learning to farm management systems is quickly evolving into a real artificial intelligence (AI) system, providing richer recommendations and insight for subsequent decisions and timely actions (Liakos et al., 2018). Further research that aims to integrate automated data recording, mycotoxin analysis, ML implementation, and decision support systems will provide practical tools in line with so-called "knowledge-based agriculture." This should move us closer toward sustainable agriculture and smart farming that also improves food safety and quality.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

PB, MC, and MM contributed to the conception and design of the study and wrote sections of the manuscript. MC organized the database. MM performed the statistical analysis. MC and MM wrote the first draft of the manuscript. All authors contributed to manuscript revision, read it, and approved the final submitted version.

## FUNDING

This study was partially supported by the regional government, Emilia Romagna region, "SERVICE – SistEmi infoRmatiVi rIschio miCotossinE" (*IT systems for mycotoxin risk*) No. 5149128.

## ACKNOWLEDGMENTS

We are grateful to the Emilia Romagna meteorological web service for providing meteorological data, to CRPV for the project coordination and to all the farmers and technicians who contributed to maize samples collection.



## REFERENCES

- Alma, A., Lessio, F., Reyneri, A., and Blandino, M. (2005). Relationships between *Ostrinia nubilalis* (Lepidoptera: crambidae) feeding activity, crop technique and mycotoxin contamination of corn kernel in northwestern Italy. *Int. J. Pest Manag.* 51, 165–173. doi: 10.1080/09670870500179698
- Battilani, P. (2016). Recent advances in modeling the risk of mycotoxin contamination in crops. *Curr. Opin. Food Sci.* 11, 10–15. doi: 10.1016/j.cofs.2016.08.009
- Battilani, P., Barbano, C., and Piva, G. (2008a). Aflatoxin B1 contamination in maize related to the aridity index in North Italy. *World Mycotoxin J.* 1, 449–456. doi: 10.3920/WMJ2008.x043
- Battilani, P., and Camardo Leggieri, M. (2015). Predictive modelling of aflatoxin contamination to support maize chain management. *World Mycotoxin J. (Special Issue Aflatoxins Maize Other Crops)* 8, 161–170. doi: 10.3920/Wmj2014.1740
- Battilani, P., Camardo Leggieri, M., Rossi, V., and Giorni, P. (2013). AFLA-maize, a mechanistic model for *Aspergillus flavus* infection and aflatoxin B1 contamination in maize. *Comput. Electron. Agric.* 94, 38–46. doi: 10.1016/j.compag.2013.03.005
- Battilani, P., Pietri, A., Barbano, C., Scandolaro, A., Bertuzzi, T., and Marocco, A. (2008b). Logistic regression modeling of cropping systems to predict fumonisin contamination in maize. *J. Agric. Food Chem.* 56, 10433–10438. doi: 10.1021/jf801809d
- Battilani, P., Rossi, V., and Pietri, A. (2003). Modelling *Fusarium verticillioides* infection and fumonisin synthesis in maize ears. *Aspects Appl. Biol.* 68, 91–100.
- Battilani, P., Toscano, P., Van der Fels-Klerx, H. J., Moretti, A., Camardo Leggieri, M., Brera, C., et al. (2016). Aflatoxin B<sub>1</sub> contamination in maize in Europe increases due to climate change. *Sci. Rep.* 6:24328. doi: 10.1038/srep24328
- Bertuzzi, T., Camardo Leggieri, M., Battilani, P., and Pietri, A. (2014). Co-occurrence of type A and B trichothecenes and zearalenone in wheat grown in northern Italy over the years 2009–2011. *Food Addit. Contam. Part B Surveill.* 7, 273–281. doi: 10.1080/19393210.2014.926397
- Bertuzzi, T., Rastelli, S., Mulazzi, A., and Pietri, A. (2012). Evaluation and improvement of extraction methods for the analysis of aflatoxins B<sub>1</sub>, B<sub>2</sub>, G<sub>1</sub> and G<sub>2</sub> from naturally contaminated maize. *Food Anal. Methods* 5, 512–519. doi: 10.1007/s12161-011-9274-5
- Blandino, M., Reyneri, A., Vanara, F., Pascale, M., Haidukowski, M., and Campagna, C. (2009). Management of fumonisin contamination in maize kernels through the timing of insecticide application against the European corn borer *Ostrinia nubilalis* Hübner. *Food Addit. Contam. Part A Chem. Anal. Control Exposure Risk Assess.* 26, 1501–1514. doi: 10.1080/02652030903207243
- Bottarelli, L., and Zinoni, F. (2002). La rete meteorologica regionale. *Il Divulgatore* 5, 13–17.
- Boughorbel, S., Jarray, F., and El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS One* 12:e0177678. doi: 10.1371/journal.pone.0177678
- Boulent, J., Foucher, S., Théau, J., and St-Charles, P. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10:941–941. doi: 10.3389/fpls.2019.00941
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* 30, 1145–1159. doi: 10.1016/s0031-3203(96)00142-2
- Byrd, H. R., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Stat. Comput.* 16, 1190–1208. doi: 10.1137/0916069
- Camardo Leggieri, M., Bertuzzi, T., Pietri, A., and Battilani, P. (2015). Mycotoxin occurrence in maize produced in Northern Italy over the years 2009–2011: focus on the role of crop related factors. *Phytopathol. Mediterr.* 54, 212–221.
- Camardo Leggieri, M., Giorni, P., Pietri, A., and Battilani, P. (2019). *Aspergillus flavus* and *Fusarium verticillioides* interaction: modeling the impact on mycotoxin production. *Front. Microbiol.* 10:2653. doi: 10.3389/fmicb.2019.02653
- Camardo Leggieri, M., Lanubile, A., Dall'Asta, C., Pietri, A., and Battilani, P. (2020a). The impact of seasonal weather variation on mycotoxins: maize crop in 2014 in northern Italy as a case study. *World Mycotoxin J.* 13, 25–36. doi: 10.3920/WMJ2019.2475
- Camardo Leggieri, M., Mazzoni, M., Fodil, S., Moschini, M., Bertuzzi, T., Prandini, A., et al. (2020b). An electronic nose supported by an artificial neural network for the rapid detection of aflatoxin B1 and fumonisins in maize. *Food Control* 123:107722. doi: 10.1016/j.foodcont.2020.107722
- Dahl, G., Sainath, T., and Hinton, G. E. (2013). “Improving deep neural networks for LVCSR using rectified linear units and dropout,” in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC.
- Danso, J. K., Osekre, E. A., Opit, G. P., Manu, N., Armstrong, P., Arthur, F. H., et al. (2018). Post-harvest insect infestation and mycotoxin levels in maize markets in the Middle Belt of Ghana. *J. Stored Products Res.* 77, 9–15. doi: 10.1016/j.jspr.2018.02.004
- Dobolyi, C., Sebők, F., Varga, J., Kocsubé, S., Baranyi, N., Szécsi, Á, et al. (2013). Occurrence of aflatoxin producing *Aspergillus* isolates in maize kernel in Hungary. *Acta Aliment.* 42, 451–459. doi: 10.1556/AAlim.42.2013.3.18
- Elavarasan, D., and Vincent, P. M. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE Access* 8, 86886–86901. doi: 10.1109/ACCESS.2020.2992480
- Eskola, M., Kos, G., Elliott, C. T., Hajšlová, J., Mayar, S., and Krska, R. (2020). Worldwide contamination of food-crops with mycotoxins: validity of the widely cited ‘FAO estimate’ of 25%. *Crit. Rev. Food Sci. Nutr.* 60, 2773–2789. doi: 10.1080/10408398.2019.1658570
- European Commission (2006a). Commission regulation (EC) No 401/2006 laying down the methods of sampling and analysis for the official control of the levels of mycotoxins in foodstuffs. *Official J. Eur. Union* 70:12.
- European Commission (2006b). Commission regulation (EC) No 1881/2006 setting maximum levels for certain contaminants in foodstuffs. *Official J. Eur. Union* 364:5.
- European Commission (2007). Commission regulation (EC) No 1126/2007 amending regulation (EC) No1881/2006 setting maximum levels for certain contaminants in foodstuffs as regards *Fusarium* toxins in maize and maize products. *Official J. Eur. Union* 255:14.
- Evans, P., Persaud, K., McNeish, A., Sneath, R. W., Hobson, N., and Magan, N. (2000). Evaluation of a radial base function neural network for the determination of wheat quality from electronic nose data. *Sens. Actuators B Chem.* 69, 348–358. doi: 10.1016/S0925-4005(00)00485-8
- Giorni, P., Bertuzzi, T., and Battilani, P. (2016). Aflatoxin in maize, a multifaceted answer of *aspergillus flavus* governed by weather, host-plant and competitor fungi. *J. Cereal Sci.* 70, 256–262. doi: 10.1016/j.jcs.2016.07.004
- Giorni, P., Bertuzzi, T., and Battilani, P. (2019). Impact of fungi co-occurrence on mycotoxin contamination in maize during the growing season. *Front. Microbiol.* 10:1265. doi: 10.3389/fmicb.2019.01265
- Guo, X. W., Fernando, D., and Entz, M. (2005). Effects of crop rotation and tillage on blackleg disease of canola. *Can. J. Plant Pathol. Revue Can. Phytopathol.* 27, 53–57. doi: 10.1080/07060660509507193
- IARC (1993). “IARC monographs on the evaluation of carcinogenic risks to humans,” in *Some Naturally Occurring Substances: Food Items and Constituents, Heterocyclic Aromatic Amines and Mycotoxins*, ed. World Health Organization (Lyon: IARC Press), 445–466.
- Jia, W., Liang, G., Tian, H., Sun, J., and Wan, C. (2019). Electronic nose-based technique for rapid detection and recognition of moldy pples. *Sensors* 19:1526. doi: 10.3390/s19071526
- Jin, J., Li, M., and Jin, L. (2015). Data normalization to accelerate training for linear neural net to predict tropical cyclone tracks. *Math. Probl. Eng.* 2015, 1–8. doi: 10.1155/2015/931629
- Jones, R. (1981). Effect of nitrogen fertilizer, planting date, and harvest date on aflatoxin production in corn inoculated with *Aspergillus flavus*. *Plant Dis.* 65:741. doi: 10.1094/PD-65-741
- Kaminari, M. D., Camardo Leggieri, M., Tsitsigiannis, D. I., and Battilani, P. (2020). Afla-pistachio: development of a mechanistic model to predict the aflatoxin contamination of pistachio nuts. *Toxins* 12:445. doi: 10.3390/toxins12070445
- Khaki, S., and Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant Sci.* 10:621. doi: 10.3389/fpls.2019.00621
- Khaki, S., Wang, L., and Archontoulis, S. (2019). A CNN-RNN framework for crop yield prediction. *arXiv [preprint] arXiv:1911.09045*

- Kingma, D., and Ba, J. (2014). "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA
- Kohavi, R., and Provost, F. (1998). Glossary of terms. *Mach. Learn.* 30, 271–274. doi: 10.1023/A:1017181826899
- Kos, J., Mastilović, J., Janić Hajnal, E., and Šarić, B. (2013). Natural occurrence of aflatoxins in maize harvested in Serbia during 2009–2012. *Food Control* 34, 31–34. doi: 10.1016/j.foodcont.2013.04.004
- LeCun, Y., Bengio, Y., and Hinton, G. E. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Levic, J., Gosic-Dondo, S., Ivanovic, D., Stanković, S., Krnjaja, V., Bočarov-Stančić, A., et al. (2013). An outbreak of *Aspergillus* species in response to environmental conditions in Serbia. *Pestic. Fitomed.* 28, 167–179. doi: 10.2298/PIF1303167L
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors* 18:2674.
- Logrieco, A. F., Battilani, P., Camardo Leggieri, M., Haesaert, G., Jing, X., Lanubile, A., et al. (2021). Perspectives on global mycotoxin issues and management from MycoKey Maize working group. *Plant Dis.* doi: 10.1094/PDIS-06-20-1322-FE [Epub ahead of print].
- Marocco, A., Gavazzi, C., Pietri, A., and Tabaglio, V. (2008). On fumonisin incidence in monoculture maize under no-till, conventional tillage and two nitrogen fertilisation levels. *J. Sci. Food Agric.* 88, 1217–1221. doi: 10.1002/jsfa.3205
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta (BBA) Protein Struct.* 405, 442–451. doi: 10.1016/0005-2795(75)90109-9
- Mazzoni, E., Scandolara, A., Giorni, P., Pietri, A., and Battilani, P. (2011). Field control of Fusarium ear rot, *Ostrinia nubilalis* (Hubner), and fumonisins in maize kernels. *Pest Manag. Sci.* 67, 458–465. doi: 10.1002/ps.2084
- Munkvold, G. (2003). Epidemiology of *Fusarium* diseases and their mycotoxins in maize ears. *Eur. J. Plant Pathol.* 109, 705–713. doi: 10.1023/A:1026078324268
- Munkvold, G. (2014). "Crop management practices to minimize the risk of mycotoxins contamination in temperate-zone maize," in *Mycotoxin Reduction in Grain Chains*, eds J. F. Leslie and A. F. Logrieco (New Delhi: Wiley Blackwell), 59–77. doi: 10.1002/9781118832790.ch5
- Nevavuori, P., Narra, N., and Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* 163:104859. doi: 10.1016/j.compag.2019.104859
- Niedbala, G. (2019). Application of artificial neural networks for multi-criteria yield prediction of winter rapeseed. *Sustainability* 11:533. doi: 10.3390/su11020533
- Öner, T., Thiam, P., Kos, G., Krska, R., Schwenker, F., and Mizaikoff, B. (2019). Machine learning algorithms for the automated classification of contaminated maize at regulatory limits via infrared attenuated total reflection spectroscopy. *World Mycotoxin J.* 12, 1–10. doi: 10.3920/WMJ2018.2333
- Palumbo, R., GonçAlves, A., Gkrillas, A., Logrieco, A., Dorne, J., Dall'Asta, C., et al. (2020). Mycotoxins in maize: mitigation actions, with a chain management approach. *Phytopathol. Mediterr.* 59, 5–28. doi: 10.14601/Phyto-11142
- Payne, G. A., Hagler, W. M., and Adkins, C. R. (1988). Aflatoxin accumulation in inoculated ears of field-grown maize. *Plant Dis.* 72, 422–424. doi: 10.1094/PD-72-0422
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pietri, A., and Bertuzzi, T. (2012). Simple phosphate buffer extraction for the determination of fumonisins in masa, maize, and derived products. *Food Anal. Methods* 5, 1088–1096. doi: 10.1007/s12161-011-9351-9
- Pietri, A., Bertuzzi, T., Pallaroni, L., and Piva, G. (2004). Occurrence of mycotoxins and ergosterol in maize harvested over 5 years in Northern Italy. *Food Addit. Contam.* 21, 479–487. doi: 10.1080/02652030410001662020
- Piva, G., Battilani, P., and Pietri, A. (2006). "Emerging issues in Southern Europe: aflatoxins in Italy," in *The Mycotoxin Factbook*, eds D. Barug and D. Bhatnagar (Wageningen: Wageningen Academic Publishers), 139–153.
- Saladini, M., Blandino, M., Reyneri, A., and Alma, A. (2008). Impact of insecticide treatments on *Ostrinia nubilalis* (Hübner) (Lepidoptera: Crambidae) and their influence on the mycotoxin contamination of maize kernels. *Pest Manag. Sci.* 64, 1170–1178. doi: 10.1002/ps.1613
- Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* 44, 206–226. doi: 10.1147/rd.441.0206
- Torelli, E., Firrao, G., Bianchi, G., Saccardo, F., and Locci, R. (2012). The influence of local factors on the prediction of fumonisin contamination in maize. *J. Sci. Food Agric.* 92, 1808–1814. doi: 10.1002/jsfa.5551
- Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M. J. (2017). Big data in smart farming – a review. *Agric. Syst.* 153, 69–80. doi: 10.1016/j.agry.2017.01.023
- Zeiler, M. D., Ranzato, M., Monga, R., Mao, M., Yang, K., Le, Q. V., et al. (2013). "On rectified linear units for speech processing," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 3517–3521

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Camardo Leggieri, Mazzoni and Battilani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.