



# Analyzing and assessing explainable AI models for smart agriculture environments

Andrea Cartolano<sup>1</sup> · Alfredo Cuzzocrea<sup>2,3</sup>  · Giovanni Pilato<sup>4</sup>

Received: 13 August 2022 / Revised: 6 December 2023 / Accepted: 21 December 2023 /

Published online: 15 January 2024

© The Author(s) 2024

## Abstract

We analyze a case study in the field of smart agriculture exploiting Explainable AI (XAI) approach, a field of study that aims to provide interpretations and explanations to the behaviour of AI systems. The study regards a multiclass classification problem on the Crop Recommendation dataset. The original task is the prediction of the most adequate crop, according to seven features. In addition to the predictions, two of the most well-known XAI approaches have been used in order to obtain explanations and interpretations of the behaviour of the models: SHAP (SHapley Additive ExPlanations), and LIME (Local Interpretable Model-Agnostic Explanations). Both packages provide easy-to-understand visualizations that allow common users to understand explanations of single predictions even without going into the mathematical details of the algorithms. Within the scientific community criticisms have been raised against these approaches, and recently some papers brought to light some weaknesses. However, the two algorithms are among the most popular in XAI and are still considered points of reference for this field of study.

**Keywords** Smart agriculture · XAI · SHAP · LIME

---

This research has been made in the context of the Excellence Chair in Big Data Management and Analytics at University of Paris City, Paris, France.

---

✉ Alfredo Cuzzocrea  
alfredo.cuzzocrea@unical.it

Andrea Cartolano  
andrea.cartolano@community.unipa.it

Giovanni Pilato  
giovanni.pilato@icar.cnr.it

<sup>1</sup> DSEAS Department, University of Palermo, Palermo, Italy

<sup>2</sup> iDEA Lab, University of Calabria, Rende, Italy

<sup>3</sup> Department of Computer Science, University of Paris City, Paris, France

<sup>4</sup> National Research Council, ICAR Institute, Palermo, Italy

## 1 Introduction

Smart agriculture is an ideal field of application for the key concepts of industry 4.0 [1–4]. Global warming and climate change have been threatening the agricultural and food production processes, impacting both contexts that are becoming more and more important to manage to avoid potentially huge losses for what concerns the cultivation of the crop. Lack of water put the chain of food production at risk, especially in some world regions. A new approach, known as Smart Agriculture, combining traditional agriculture with artificial intelligence (AI) and autonomous systems, is then needed to tackle the challenges represented by climate change [5].

In this paper, we tackle the problem of smart agriculture. We focus on Explainable AI (XAI), a field of study that aims to obtain interpretations and explanations to model outputs by making the logic and behaviour of machine learning models comprehensible for humans. After the first papers that have been published on the theme at the turn of the Eighties and the Nineties, in the following decades, the largest part of research focused on new high-performance but very complex algorithms. However, since the predictions of these machine learning models cannot be interpreted, it is also difficult to trust them, and this can lead to dangerous consequences, particularly in certain sectors such as healthcare and self-driving cars, in which wrong predictions could potentially cost human lives. As a result, new needs for interpretability arose, and nowadays, with the field of XAI that has been experiencing a revival, it is preferable to have an interpretable yet slightly less accurate model than a more accurate black box model, within a certain extent [6]. In this paper, we decided to propose an XAI approach for a smart agriculture task, since we consider it crucial for the reasons stated above.

Classification tasks are part of machine learning (ML) techniques, whose goal is to predict unordered, and discrete values. When proper labels are provided, we are facing a supervised learning problem, otherwise, if labels are not given, we have to train our models for an unsupervised learning task. Classification can be binary, and in this case we have only two classes, or multiclass, and it is the case in our study where we have more than two classes to predict [7].

Many classification tasks associated with smart agriculture have been proposed in the literature. However, among the different approaches, an interpretable methodology is needed since the audience is represented in the first place by farmers and agronomists. The kind of audience involved is crucial in XAI [8]: farmers and agronomists, generally speaking, are pretty skeptical in trusting AI and machine learning model predictions, especially in the presence of non-interpretable black-box models, as reported in [9]. Furthermore, the visualization of the results plays a crucial role in this field since the final users of this methodology are usually not experts in machine learning and visual output is therefore desirable (e.g., [10–12]).

XAI allows us to investigate the knowledge learned by machine learning models trained to recognize an adequate crop to cultivate. More in detail, we show that through some XAI charts, even non-machine learning experts can understand why a model predicts a particular crop for a specific observation and, overall, which are the best combinations of features that lead the models to the prediction of one class. In particular, we show how different ML models can obtain high accuracy scores, and how visualization XAI charts transform our black box models into interpretable, transparent models.

SHapley Additive ExPlanations (SHAP) [13] and Local Interpretable Model-Agnostic Explanations (LIME) [14] libraries have been used to get explanations and interpretations

of the model. We show how both XAI packages can be advantageous in showing the behavior of models about the recommendation of the best crop to select among 22 possible classes. Through the functions and the different kinds of plots that both libraries provide, we can collect interpretations on single predictions of the model and, particularly with SHAP `summary_plot`, even the patterns behind the recommendation of a single class.

The remainder of this paper is organized as follows. In Section 2, we investigate and report on related work for our research. Section 3 discusses general guidelines and main assumption of applying XAI to the emerging field of smart agriculture, and we also introduce the two main tools used in our research, i.e. SHAP and LIME. In Section 4, we describe how we trained the reference five classification models of our experiments, along with their accuracy analysis. Section 5 reports our experimental campaign where we show methods and related graphical tools of SHAP and LIME used to produce suitable XAI charts, along with a detailed discussion. In Section 6, we discuss limitations and strengths of SHAP and LIME. Finally, Section 7 contains conclusions and future work of our research.

This paper significantly extends the conference paper [15], where we introduced our general framework.

## 2 Related work

As already mentioned in the *Introduction*, the field of smart agriculture is an ideal application domain for the fundamental concepts of Industry 4.0. As a consequence, there have been several projects that have been studied and pursued in recent years that have included multiple objectives: *a)* managing greenhouse gas emissions through sensors; *b)* improving energy efficiency; *c)* observing phenological stages; *d)* detecting the presence of insects or diseases in crops [16].

In previous projects, in many cases, these plans have been implemented through decision support systems (Decision Support Tools). Still, this approach is not without its limitations. Although they have proved to be effective tools in decision-making and easier to use than in the past, these systems return data that are difficult to analyze and interpret. For these reasons, farmers and agronomists have shown several reluctance to use them [1].

To tackle these problems, a system named *Solarfertiligation* has been implemented, which deals with integrating and unifying both the decision-making process and the process related to the automation of irrigation and fertilization processes. Another peculiarity of the system just mentioned is that it manages the entire crop cultivation cycle. It is capable of changing the amounts and types of fertilizers, as well as the amount of water to be used in the irrigation process, based on the detection of meteorological data and others related to the characteristics of the cultivated soil. *Solarfertiligation* is also equipped with an independent weather station in addition to sensors that can extract data from weather stations located throughout the territory [1].

In the few projects that have combined precision agriculture with XAI, the use of models that can generate explanations for predictions has achieved numerous benefits, including incentivizing farmers to use artificial intelligence systems. For example, a Fuzzy Rule-Based System (FRBS) model, called *Vital*, has been developed in one such project. It can automatically manage sensors scattered over fields and decide the appropriate amount of water for irrigation through interpretable outputs. There are essentially three reasons why this approach was followed in that study: 1) expert knowledge is not, in most cases, absolute and well-defined, but it has degrees of fuzziness and approximation; a model that is based on fuzzy

logic can therefore more easily integrate it; 2) real-world data, even in the agricultural field, and even those collected by sensors, always carry noise and are therefore reported with some degree of approximation; a model capable of obtaining predictions based on these data can therefore return indications that are reasonably closer to real-world ones 3) such logic allows for outputs that are simpler to interpret than others, more complex models. In addition, FRBS models have been shown to be superior to similar crisp-type systems in different applications and for various tasks (classification, regression, big data analysis) [9, 17].

However, another study presented a Case-Based Reasoning (CBR) model. The approach aims, on the one hand, to determine the ideal growth rate for farming according to sustainability and affordability criteria; and, on the other hand, to make this model easily interpretable and understandable for agricultural workers. The objectives of this study are twofold: 1) the accuracy of the predictions, which translates into obtaining a sufficiently low mean square error (less than or equal to 10 kilograms per hectare per day) of the dry grass growth rate; 2) explanatory success, defined as the percentage of nearest adjacent cases, or within the same farm or county. Interpretations of the model are then obtained through customized post-hoc explanations with examples, with the aim of excluding outliers as much as possible, reducing noise, and thus providing clear guidelines for farmers [18].

Another interesting study comes from a UK Natural Environment Research Council (NERC) project. It aims, through joint techniques of probabilistic inference, machine learning and XAI, to develop a framework that can identify the key factors that have led to land-use changes in the two pilot regions selected for project development (i.e., Oxfordshire and Lincolnshire) and build a model that can predict changes that will occur in the coming years [19].

### 3 Applying explainable AI to smart agriculture

The selection of the adequate crop to cultivate in relation to the soil characteristics and climate conditions is extremely important in smart agriculture, because it allows implementing ML models that can classify and predict which agricultural products are more likely to grow in the presence of specific input data. The problem can be seen as a classification task. There are many classification approaches in the literature. However, we need an explainable methodology for many reasons. First of all, there are some fields of application in which is extremely dangerous to be confident in predictions without an explanation, for instance in medical science where it has been proven that not interpretable models could potentially cost human lives [20]. The accuracy score is not enough to gain trust in the algorithms because a model can learn pieces of knowledge not included in the training set, and we may have data leakage [21]. Also, we may have models that, when used on real-world data, could obtain worse performances than expected, resulting in negative economic consequences. In the second place, XAI can encourage farmers and agronomists to use ML models or AI systems, allowing them to investigate the knowledge learned by the models, on the one hand, it is also possible to compare human expertise with ML knowledge.

For the best crop to be predicted, accurate, and structured data must be obtained. It is vital to know the *nitrogen* (*N*), *phosphorus* (*P*) and *potassium* (*K*) concentration values contained in the fertilizer used for the crop. Every crop needs the right concentration of these three elements, which are responsible for indispensable steps in the plant's growing process: in particular, *nitrogen* (*N*) goes to affect leaf growth in the plant; *phosphorus* (*P*) focuses on

root, flower, and fruit development; and *potassium* (*K*) enables the plant to absorb water more easily and to resist frost or harmful actions by pests more effectively <sup>1</sup>.

The issue that has been tackled is a multi-classification problem: the models have to predict an adequate crop for the field condition based on seven numeric features, respectively *N*, *P*, and *K* (showing the concentration values of nitrogen, phosphorus, and potassium within the fertilizer), *temperature* (in Celsius), rate of *humidity* (percentage), *ph* (acidity of the soil), and *rainfall*. We have used the Crop Recommendation dataset <sup>2</sup> experimentally to investigate explanations and common patterns among different models.

In particular, the dataset consists of 2200 observations and 8 columns. There are no missing values or duplicates. The seven features are all numerical and are as follows: *N*, concentration of the nitrogen value used in the fertilizer; *P*, concentration of phosphorus value used in the fertilizer; *K*, concentration of potassium value used in the fertilizer; *ph*, value representing the measure of soil acidity; *rainfall*, rainfall expressed in mm; *temperature*, temperature expressed in degrees Celsius; *humidity*, relative humidity in percentage values.

The first three attributes contain integer numeric values, while the remaining four are of float type. Added to these attributes is the target categorical variable, *label*, of string type that contains the names of agricultural products grown. There are 22 classes, each of which has 100 instances. The agricultural products that make up the classes are: *apples*, *bananas*, *black Indian bean*, *chickpeas*, *coconut*, *coffee*, *cotton*, *grapes*, *jute*, *red beans*, *lentils*, *maize*, *mango*, *aconitifolia vine*, *green Indian bean*, *melon*, *orange*, *papaya*, *cayenne*, *pomegranate*, *rice*, and *watermelon*.

This dataset has been selected because of its completeness and simplicity. This characteristic is essential, since one of the targets that should be taken into consideration is the comparison between the knowledge learned by the algorithm and the knowledge of farmers and agronomists.

For every model, the dataset has been divided into a training set (80% of observations) to train the models and a test set (20%) to obtain the predictions. In the next subsections, the two interpretable algorithms will be described, by focusing on the main characteristics.

Furthermore, while our paper effectively delineates the utility of XAI in smart agricultural practices, it is imperative to scrutinize and acknowledge the inherent constraints that may impede its implementation and efficacy in practical settings.

- **Implementation Challenges in Real-World Agricultural Settings:** The integration of XAI techniques into actual agricultural operations faces practical challenges. The complexity of agricultural environments, variations in soil types, climate conditions, and crop varieties can pose hurdles in deploying XAI models effectively. Furthermore, the requirement for specialized sensors, infrastructure, and the need for high-quality data acquisition systems may be financially burdensome for smaller or resource-constrained agricultural setups. Addressing these challenges demands not only technological advancements but also infrastructural support and financial investments, which might not be feasible for all stakeholders.
- **Limitations of the Dataset and Generalizability of Results:** The quality, diversity, and representativeness of the dataset used in training XAI models significantly impact their performance and generalizability. Issues such as biased or incomplete data, limited data samples, or data collected from specific geographical regions or timeframes might restrict the applicability and generalizability of the developed models. Consequently, the predic-

<sup>1</sup> For a more thorough description of what *N*, *P* and *K* are actually responsible for, see: <https://www.gardenersworld.com/how-to/grow-plants/understanding-npk-fertiliser/>

<sup>2</sup> <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>

tive capability of these models may suffer when applied to different agricultural contexts or unforeseen scenarios.

- *Interpretability and Complexity*: While XAI models aim to provide interpretability, the complex nature of some machine learning algorithms might result in black-box scenarios where the decision-making process becomes challenging to comprehend. This lack of interpretability could hinder the acceptance and trustworthiness of these models among agricultural stakeholders who require transparent and understandable decision-making processes.
- *Regulatory and Ethical Concerns*: Implementing AI in agriculture also raises regulatory and ethical concerns. Privacy issues related to data collection from farms, ownership and sharing of data, as well as potential biases encoded in algorithms, need to be thoroughly addressed. Compliance with existing agricultural regulations and ethical considerations related to AI adoption in farming practices necessitate careful attention and adherence.

In conclusion, acknowledging these limitations is crucial for a holistic understanding of the practical applicability and challenges associated with employing XAI in smart agriculture. Addressing these limitations requires collaborative efforts among researchers, policymakers, technologists, and agricultural practitioners to devise innovative solutions and frameworks that mitigate these constraints and facilitate the successful integration of XAI into real-world agricultural settings.

### 3.1 SHAP

SHAP [13] is an additive feature attribution method that has its roots in Shapley values and Game Theory. Starting from the base value, the predicted value from the null model (i.e., the model without any features), SHAP calculates the average marginal contribution of each player, a portion, or a group of features. For each observation, the sum of SHAP values of each feature is equal to the difference between the model's predicted value and the base value.

Explanation models use simplified inputs  $x'$  rather than the original ones through the following mapping function:  $x = h_x(x')$ . The explanation function of such methods is a linear function of binary variables. SHAP calculates the contribution  $\phi_i$  to each feature, and by summing, it is able to approximate the prediction function of the original model, where  $z'_i$  can be equal to 0 or 1 and  $M$  is the number of simplified input features.

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i + z'_i \quad (1)$$

SHAP values have three different desirable properties. Local accuracy is the first one and prescribes that the explanation model must be able to approximate the output of the original model either when  $x = h_x(x')$  or  $\phi_0 = f(h_x(0))$ .

$$f(x) = g(z') = \phi_0 + \sum_{i=1}^M \phi_i x'_i \quad (2)$$

The second property, missingness, requires the missing features to have no impact on the model's output.

$$x'_i = 0 \implies \phi_i = 0 \quad (3)$$

The third property, consistency, states that if the contribution of a simplified input, regardless of the other ones, does not also decrease, the original input should do the same.

$$f'_x(z') - f'_x(z' \setminus i) \geq f_x(z') - f_x(z' \setminus i) \tag{4}$$

for all inputs  $z' \in \{0, 1\}^M$ , then  $\phi_i(f', x) \geq \phi_i(f, x)$ , where  $f_x(z') = f(h_x(z'))$  and  $z' \setminus i$  equates to set  $z_i = 0$ .

In order to compute Shapley values, the model has to be trained for each possible subset  $S$  of the entire set of features  $F$ . In this way, it is possible to attribute to each feature an importance value that corresponds to the contribution of each feature to the model prediction. To compute this value, a model  $f_{S \cup \{i\}}$  is both trained with a particular feature and without the same one  $f_S$ . By doing so, predictions from the two different models are compared:  $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ , where  $x_S$  represents the input features in the subset  $S$ . By replying to this procedure for each feature, it is possible to obtain a feature attribution  $\phi_i$  for each observation.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \tag{5}$$

While model-agnostic SHAP KernelExplainer calculates the average marginal contribution of each feature, the model-specific TreeExplainer calculates the contributions conditioned on the subset  $S$  of features [22], where  $S$  corresponds to the non-zero indexes of  $z'$  and  $N$  is the set of all input features. The expected value is  $E[f(x) | x_S]$ .

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!(M - S - 1)!}{|M|!} [f_x(S \cup \{i\}) - f_x(S)] \tag{6}$$

DeepExplainer, for neural networks, assumes feature independence and the linearity of deep model and is based on the Deep Learning Important FeaTures (DeepLIFT) [23]. It assigns each input  $x_i$  a  $C_{\Delta x_i \Delta x_j}$  value, correspondent to the effect of an input  $x_i$  set to a reference value in contrast to the original one. Through the mapping function  $x = h_x(x')$ , DeepLIFT converts original values into binary values, where 0 represents the input  $x_i$  taking the reference value, and 1 the original value. DeepExplainer combines small components of neural network in those of the entire model by recursively passing DeepLIFT's multipliers, defined as:

$$m_{\Delta x_i \Delta x_t} = \frac{C_{\Delta x \Delta t}}{\Delta t} \tag{7}$$

where  $\Delta x$  is the difference between the input value and the reference value,  $\Delta t$  describes the difference between the target neuron  $t$  and the reference value, and  $C_{\Delta x \Delta t}$  is the contribution of the two inputs.

### 3.2 LIME

LIME [14] aims to explain the prediction function of the original complex model through a simpler linear model. An explanation is a local linear approximation of the original model. LIME is model-agnostic, which means that, regardless of the complexity of the original model, this algorithm will interpret it as a non-transparent model. The authors of LIME tried to find solutions to two common problems in machine learning: the gain of trust in the single predictions, on the one hand, and the gain of trust in the behavior of the model as a whole,

on the other hand. Indeed, if users cannot understand why a model behaves as it does, they will tend not to use it. An explanation should also be locally faithful, which means that in the proximity area around the instance explained, the explanation model  $g$  should reply to the behavior of the original model  $f$ .

To explain a local prediction that is locally faithful, the algorithm will minimize the loss function  $L$ , which includes the original model  $f$ , the simpler linear model  $g$  and  $\pi_x(z)$ , that is, the proximity measure between the instance  $x$  and the instance  $z$ , and  $\Omega(g)$ , that represents the complexity of the explanation model (e.g., depth of the trees in a decision tree). The simpler the model, the better for the interpretability of the explanation.

$$\xi(x) = \operatorname{argmin}_{g \in G} L(f, g, \pi_x) + \Omega(x) \quad (8)$$

To gain trust in the behavior of the entire model, the authors of LIME developed another algorithm, the Submodular Pick (SP). SP aims to select instances characterized by a non-redundant coverage of the area of the model, where non-redundant means that it is made up of instances with different explanations. Within a set  $B$  of instances that a human being is willing to inspect, through the SP, it is possible to obtain a  $n \times d'$  explanation matrix, where  $n$  is the number of explanations selected by the SP and  $d'$  represents the interpretable features, while  $I$  is the total importance of interpretable features that are contained in at least one selected instance. Non-redundant coverage is obtained by the function  $c$  that for  $W$  and  $I$ , computes the total importance of features contained in the set  $V$  of explanations.

$$c(V, W, I) = \sum_{j=1}^{d'} \mathbb{1}_{[\exists i \in V: W_{ij} > 0]} I_j \quad (9)$$

The SP maximizes the mentioned coverage function by adding, for each iteration, the instances with the highest impact on coverage inside the set  $V$ .

$$\operatorname{Pick}(W, I) = \operatorname{argmax}_{V, |V| \leq B} c(V, W, I) \quad (10)$$

## 4 Trained models

Five different models have been implemented on top of a Python environment running on a machine characterized by: (i) Microsoft Windows 10 operating system; (ii) Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz processor; (iii) 16.0GB of main memory; (iv) 512 GB of SSD storage. Three of these algorithms have been used to get explanations: an Extreme Gradient Boosting (XGB), a Multi Layer Perceptron neural network (MLP), and three different Support Vector Machines (SVM), the first one with linear kernel, the second one with a polynomial kernel, and the third one with a radial basis function kernel: only the linear SVM has been selected since it obtains a higher accuracy score compared to the other two. The reason behind the choice of these particular models was due to the different kinds of explainers that the SHAP package provides: in addition to the model-agnostic `KernelExplainer` (used for the linear SVM), the `DeepExplainer` has been used for MLP, and the `TreeExplainer` is specific for tree-based algorithms, for XGB. As Fig. 1 shows, each model achieves a very high accuracy score.



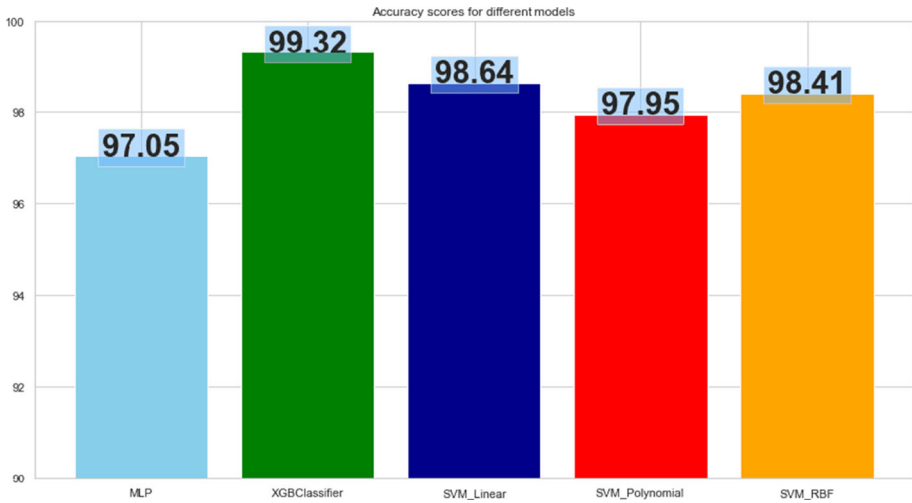


Fig. 1 Accuracy scores for each model

### 5 XAI charts

The functions of both SHAP and LIME packages allow us to investigate on single predictions. With the method `shap_values()` a list of 22 arrays is obtained, one for each class. In every model, we have noted the tendency to misclassify the class *rice* in favor of the class *jute*. The Fig. 2 shows the output of the XGB model for an observation in which *jute* has been predicted on behalf of *rice*.

The LIME class `TabularExplainer` has the `explain_instance()` method shown in Fig. 3, by which it is possible to obtain an easily interpretable HTML visualization for the same observation of Fig. 2, where prediction probabilities are displayed in addition to the coefficients of the features that seem to impact both positively and negatively on the prediction of one particular class.

SHAP `summary_plot` allows us to display both features with the mean biggest effect on model output and how they impact single classes. For instance, *humidity* and *P* are the most important features for the prediction of the class apples.

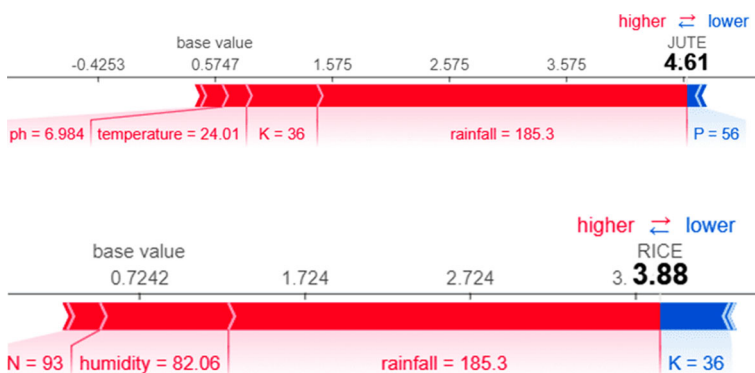


Fig. 2 XGB - force\_plot

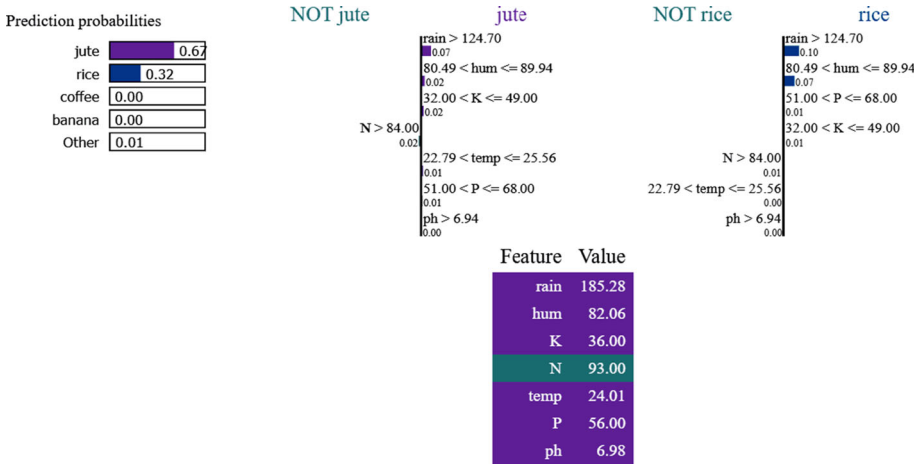


Fig. 3 XGB - explain\_instance

It is also possible to investigate a single class with the same plot. It is clearly visible what has been observed in the previous one: XGB tends to predict apples just with the contribution of *P* and *humidity*, and more precisely in the presence of high values for both of these features, as shown in Fig. 4. *Rainfall* seems to have a slight impact, whereas the other features have no impact at all on the cultivation of this crop, according to what has been learned by the XGB. Figure 5 shows the XGB outcome for the class *apple*.

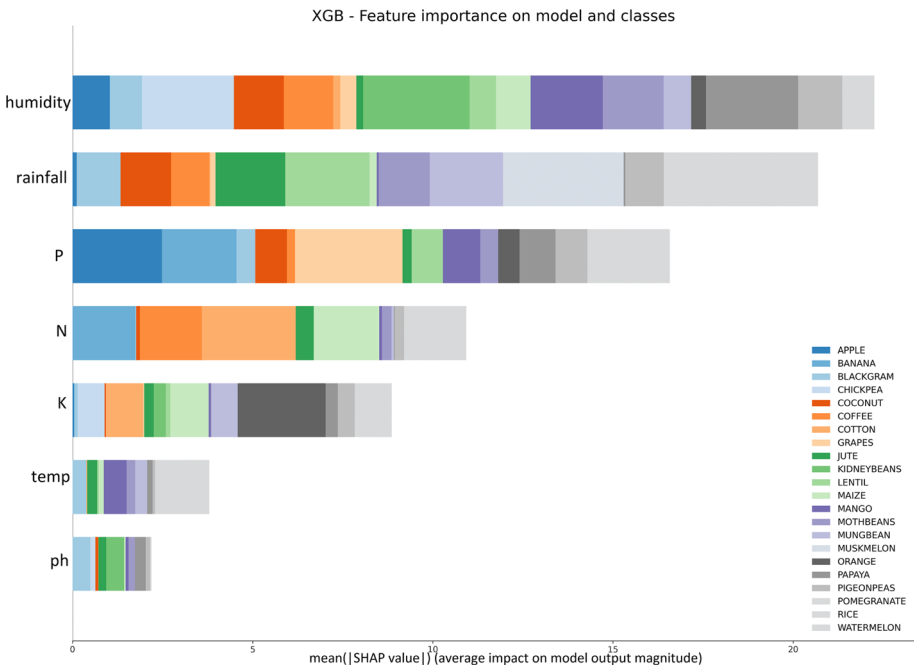


Fig. 4 XGB - summary\_plot

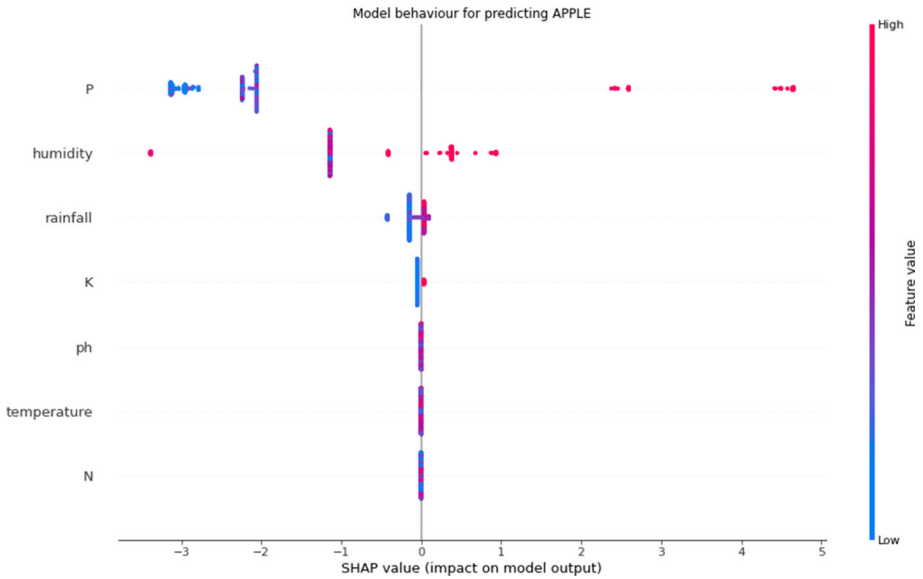


Fig. 5 XGB - summary\_plot for class *apple*

With LIME SubmodularPick, as it has been mentioned in the LIME section, it is possible to obtain a matrix  $W$  made up of  $n$  instances and  $d'$  interpretable features. In Fig. 6, it has been displayed the mean effect of the interpretable attributes selected by SP within a set  $B$  of 50 instances for which we wanted to obtain 10 explanations. We can note that the interpretable attribute with the biggest mean effect is *humidity* > 89.94, and that the attributes with poor or negative impact are related to *ph* and *temperature*, which is consistent with what SHAP summary\_plot has shown.

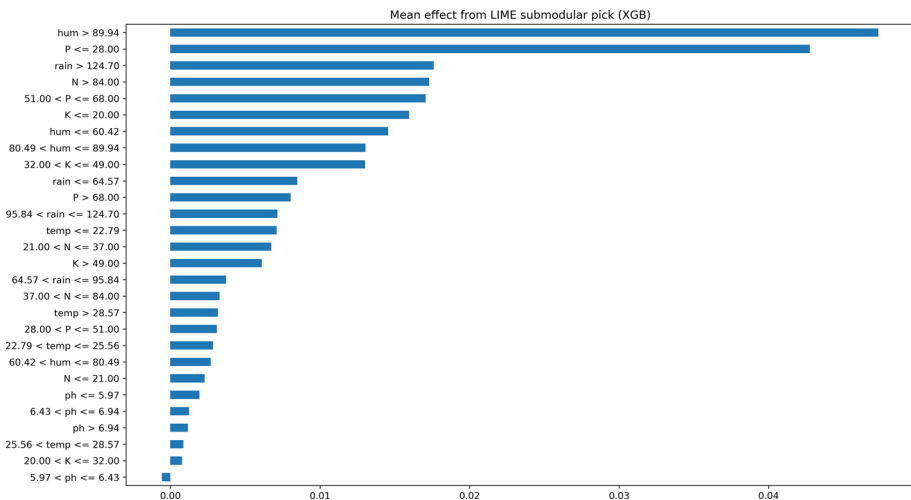


Fig. 6 XGB - LIME submodular\_pick

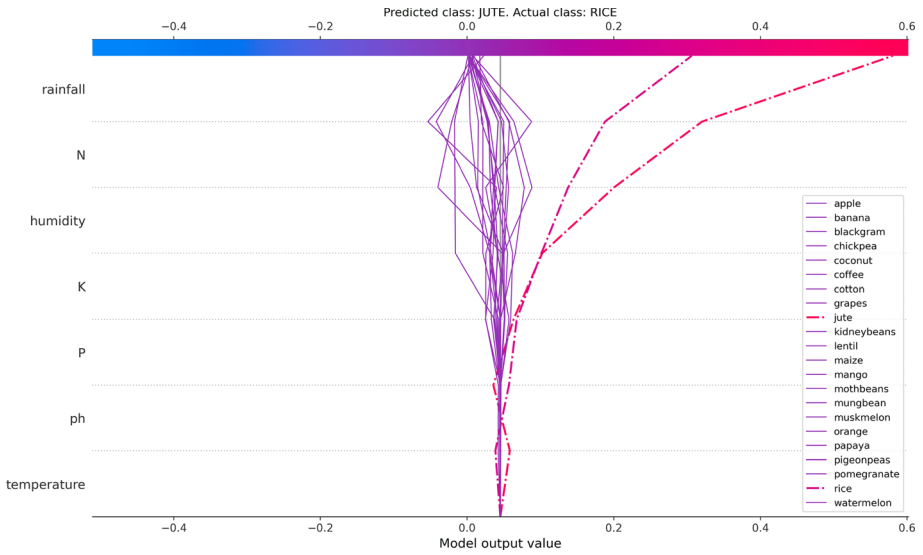


Fig. 7 SVM - multioutput\_decision\_plot

SVM also tends to misclassify rice in favor of jute. In Fig. 7, the SHAP multioutput\_decision\_plot shows the model output for the 22 classes on a single observation, but just two of them are highlighted with the dashed lines. *Rice* is the expected, class but despite its positive contributions, the predicted class from the model is *jute*. The two classes have very similar performances until the feature *humidity*, when a small gap starts to divide the two classes.

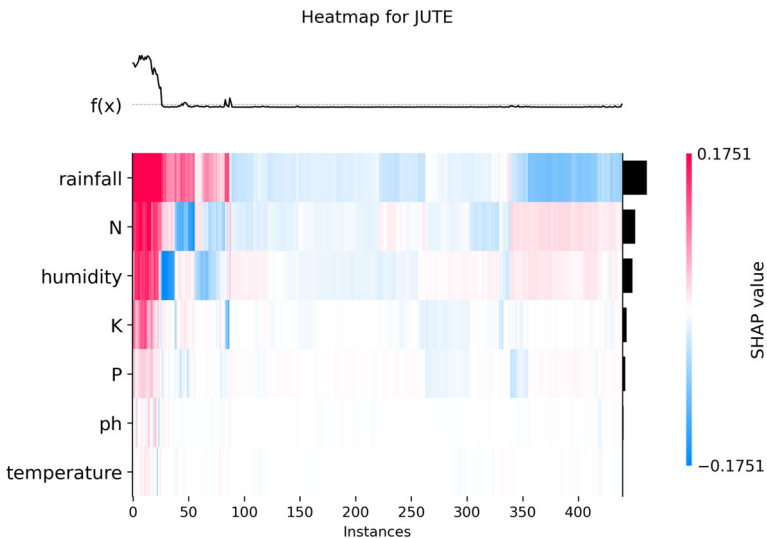


Fig. 8 SVM - heatmap\_plot for *jute*

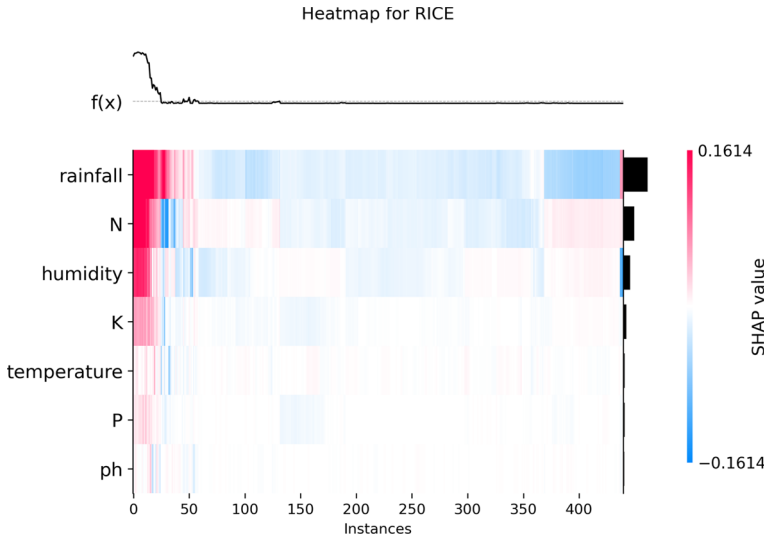


Fig. 9 SVM - heatmap\_plot for rice

Another interesting visualization is given by the SHAP heatmap\_plot where it is visible how much each feature impacts the instances of the test set, in addition to the global importance of each feature for the prediction of a single class. In Fig. 8, we reported the heatmap for class *jute*, while in Fig. 9 the same experimental pattern for class *rice*.

If we compare the previous heatmap with the class *rice* heatmap, we will also have an idea of the reason why the two classes are not always correctly classified by our model. Both *rice* and *jute* are classified with the contribution of the same features: for these classes, the behaviour of the model seems quite similar. However, if we look more closely, we can see

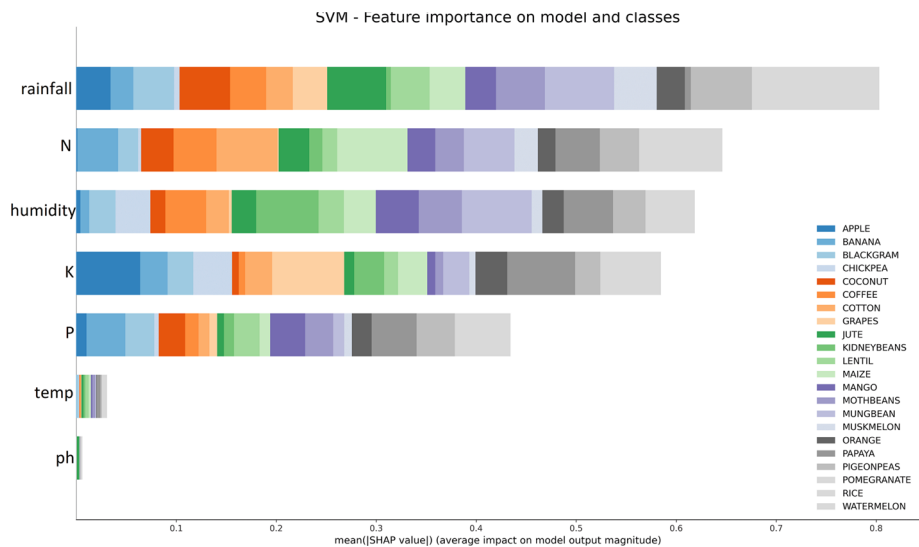


Fig. 10 SVM - summary\_plot

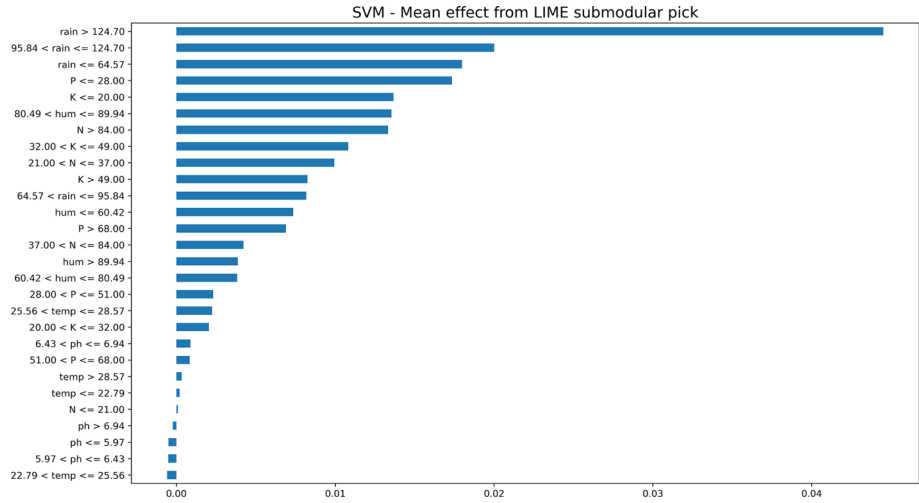


Fig. 11 SVM - submodular\_pick

that the *rainfall* attribute has a positive impact on more instances of *jute* than *rice*. Because of this, it is likely that *jute* is more easily classified than *rice* in presence of similar input data.

Unlike XGB, in SVM, the feature with the biggest impact is *rainfall*, followed by *N*; *temperature* and *ph* are still the least important attributes on model output. Also, the single classes are selected differently: *apple*, for instance, is selected mainly for *K*, *rainfall*, *P* and *humidity*, as shown in Fig. 10.

In Fig. 11, we report the mean values. As shown in Fig. 11, the visualization of the mean effect of the interpretable attributes selected by LIME SubmodularPick is somehow similar to what `summary_plot` shows. The interpretable attributes related to rainfall have the biggest mean effect, in particular, *rainfall* > 124.70, whereas those related to *temperature* and *ph* have a poor or negative effect.

Figure 12 shows the prediction output with MLP. Here, MLP also tends to misclassify rice in favor of jute, but if we look more closely at the misclassified instances, we will note that

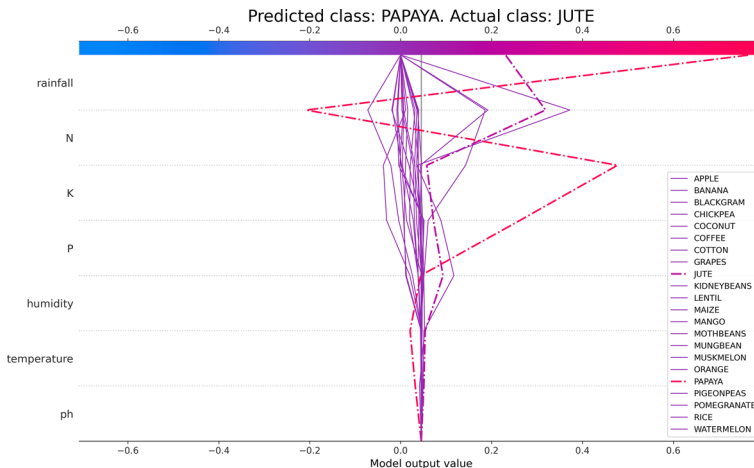


Fig. 12 MLP - multioutput\_decision\_plot

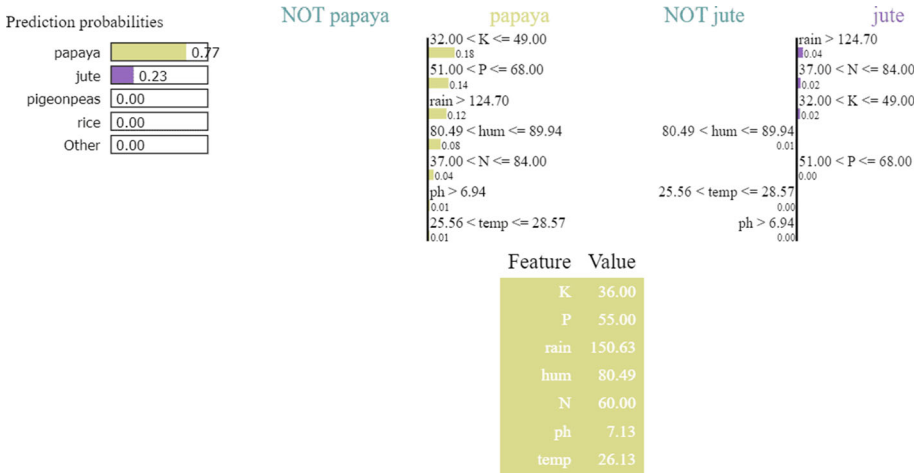


Fig. 13 MLP - explain\_instance

with this model, papaya is selected more times than jute. Indeed, papaya is a kind of crop that benefits from different values of rainfall, even lower than rice and jute, A clear representation of this pattern learned by MLP is given by the multioutput\_decision\_plot, in which for an observation that has  $rainfall = 150.6$ , the model output for *papaya* improves dramatically, unlike the expected class (*jute*).

By analyzing the same observation, we can see how the explanation linear model approximates the behaviour of the neural network. All the interpretable attributes impact positively on the prediction of *papaya*, but the attribute  $rainfall > 124.70$  is only the third most important, whereas  $32.00 < k \leq 49.00$  and  $51.00 < P \leq 68.00$  have the biggest impact on the prediction of the class. Figure 13 reports the visual explanation for this experiment.

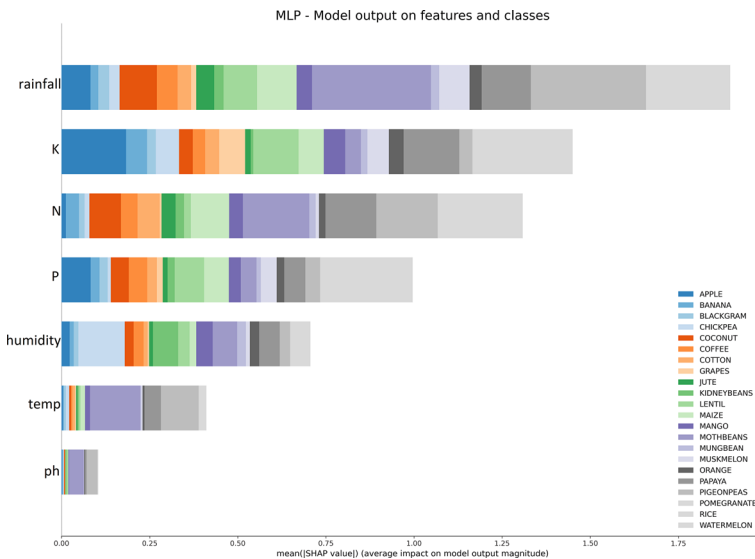


Fig. 14 MLP - summary\_plot

In Fig. 14, we again show the feature analysis for MLP. As shown in Fig. 14, *rainfall* is still the feature with the most important mean impact on the model, but, unlike SVM, *humidity* is only the fifth attribute in order of importance. The three chemical elements *N*, *P*, *K* are more important, with the latter that has the second biggest impact.

After that, in Fig. 15, we report the mean values for MLP. Here, LIME SubmodularPick seems to confirm the importance of the chemical elements: 7 out of the first 8 interpretable attributes are related to the values of *N*, *P* and *K*. The only one that is related to a different feature is *rainfall* > 124.70, an attribute that we had already found in SVM.

Finally, we explore the practical implications and challenges associated with our proposed approach in the context of agriculture as it is crucial for its successful implementation and adoption. We, address these aspects in order to elucidate the integration of the approach within existing agricultural frameworks, infrastructure requirements, and the training necessary for stakeholders to effectively utilize the system.

- *Integration with Existing Agricultural Practices:* An in-depth discussion regarding the seamless integration of the proposed approach within established agricultural practices is imperative. Highlighting how the proposed system complements or enhances existing methodologies, such as precision agriculture techniques, crop management practices, or decision-making frameworks, would elucidate its practical relevance. Emphasizing the compatibility and adaptability of the approach with diverse farming systems, crop varieties, and regional agricultural practices is essential to showcase its versatility and applicability across different contexts.
- *Infrastructure Requirements:* Detailing the necessary infrastructure for implementing the proposed approach is fundamental. This encompasses technological prerequisites such as sensors, data collection devices, computing resources, and communication networks. Additionally, elucidating the scalability and cost implications of the required infrastructure, especially for smallholder farmers or resource-limited settings, would provide insights into the feasibility and potential barriers to adoption.
- *Training for Farmers and Agronomists:* A comprehensive discussion on training requirements is essential to empower farmers and agronomists in utilizing the proposed system effectively. Describing the training modules, workshops, or educational programs necessary to familiarize stakeholders with the technology, its functionalities, and interpretation of results would facilitate its uptake. Additionally, addressing the need for user-friendly

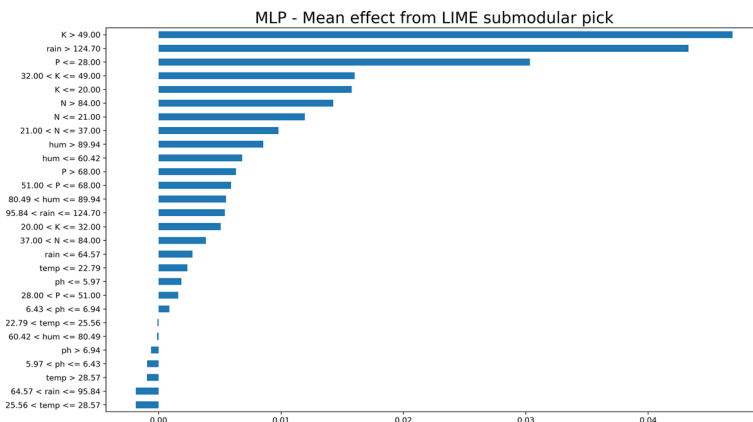


Fig. 15 MLP - submodular\_pick



interfaces, manuals, or support systems to aid stakeholders in navigating and understanding the system would enhance its usability and acceptance.

- *Adoption Challenges and Mitigation Strategies*: Recognizing the challenges associated with the adoption of new technologies in agriculture is crucial. Discussing potential barriers such as technological literacy, resistance to change, or financial constraints and proposing mitigation strategies, including capacity-building initiatives, demonstration projects, or collaborative partnerships, would pave the way for smoother implementation and uptake.

Finally, addressing these practical implications and challenges would facilitate a clearer roadmap for the successful implementation and adoption of the system within the dynamic landscape of agricultural practices.

## 6 Limitations and strengths of both packages

The two libraries have received some criticism, and some problems have to be solved to gain solidity and reliability within the scientific community.

As shown in [24], both packages can be deceived by adversarial attacks, leading SHAP and LIME to generate inconsistent explanations. In particular, the authors of the paper created a *scaffolding* that fundamentally hides the predictions of a biased classifier on input data, so that the explanations generated on the perturbed data points are unable to detect potentially discriminatory models behaviour. Another paper [25] demonstrated that it is also possible (either accidentally or intentionally) to create models that use one particular feature to obtain predictions, but when the explanations are generated the same feature could not be included, leading both XAI algorithms to generate misleading explanations.

More in detail, LIME lacks stability in its explanations, which basically means that if an explanation regarding the same observation is repeated multiple times, different explanations can be generated; furthermore, on one hand, a linear model may not be the best one to interpret a complex model, especially if we take into consideration a large part of the area of the model; on the other hand, the surrogate model must be as simple as possible in order to keep its interpretability and transparency [26]. Also, LIME seems to be unable to discriminate between relevant and non-relevant features when it comes to provide explanations for high-dimensional data sets [27].

With regard to SHAP, `KernelExplainer` has two big weaknesses: it ignores possible feature dependence by assuming that each one is independent from one another, and it is extremely slow in the calculation of SHAP values (LIME has the advantage of being much faster in the generation of the explanation). Suffice it to say that to calculate the SHAP values of our small test set, it took around 15 minutes. This could be a very big limitation, especially in the presence of many features. `TreeExplainer` is much faster than `KernelExplainer` and allows to have global model interpretations in reasonable time, but since it changes the value function, it could lead to obtain possible counter intuitive values [26]. From our experience, we also found several conflicts among `DeepExplainer` and different versions of PyTorch and TensorFlow; hopefully they will be fixed soon.

However, SHAP and LIME were undoubtedly the first packages that opened the way for XAI applications and despite their limitations, they still remain valid solutions, even considering that this discipline has been studied in particular since 2016, the year of LIME's

development; we must therefore bear in mind that XAI is a rather recent field of study, research is still in progress, so it would have been almost impossible to propose flawless algorithms at this stage. Both packages are also included in several libraries, such as the Microsoft open-source *InterpretML*<sup>3</sup> (which proposes the Explainable Boosting Machine, a new XAI algorithm [28]), the interactive *SHAPASH*<sup>4</sup> and *ExplainerDashboard*<sup>5</sup>, just to name a few. An interesting description of different XAI approaches can be found at [29].

Both packages work on tabular data, images, and texts; in particular, SHAP includes two specific charts for images and texts, the `image_plot` and the `text_plot`. LIME is able to identify and select just a handful of features that have a significant impact on the prediction and generate interpretable features that allow to obtain deeper insights into the final explanations; especially in the presence of high-dimensional datasets, this can be a great help [26]. SHAP, being derived from Shapley values and Games Theory, has a solid theoretical foundation and is also able (with the `KernelExplainer`) to connect LIME local explanations with Shapley values [26].

Furthermore, we discuss other limitations and potential issues associated with XAI packages SHAP and LIME for a better comprehensive understanding of their applicability in agricultural contexts.

- *Interpretability Issues*: While SHAP and LIME are renowned for their ability to provide interpretability to complex machine learning models, certain limitations exist regarding the interpretability of the explanations generated. The complexity of models or instances with high dimensionality might pose challenges in providing easily interpretable explanations. Additionally, these methods might generate explanations that are not intuitive or are difficult for non-technical stakeholders, such as farmers or agronomists, to comprehend. Addressing the potential shortcomings in delivering understandable explanations is crucial for ensuring the practical utility of these XAI packages in agricultural decision-making.
- *Computational Cost*: The computational cost associated with SHAP and LIME methods can be substantial, especially when applied to large-scale agricultural datasets or complex machine learning models. The calculation of Shapley values in SHAP or generating local approximations in LIME may demand significant computational resources and time, making real-time or on-field application challenging. Discussing strategies to optimize computational efficiency without compromising accuracy and reliability is essential for making these XAI packages more feasible for practical agricultural settings.
- *Potential Conflicts with Other Software or Libraries*: Integrating SHAP and LIME into existing software environments or utilizing them alongside other libraries may pose compatibility issues or conflicts. Incompatibilities with specific versions of programming languages, dependencies, or conflicts with other AI/ML frameworks could hinder seamless integration and usage. A detailed exploration of these potential conflicts and recommendations for mitigating such issues would be beneficial for practitioners intending to implement these XAI packages in agricultural applications. To prevent these conflicts, adherence to version compatibility is key. Ensuring consistent versions across dependencies, including machine learning frameworks and Python libraries, mitigates conflicts. Leveraging virtual environments or containerization aids in isolating environments, reducing compatibility issues. Regular updates and thorough documentation facilitate smoother integration, fostering reliable and robust XAI applications.

<sup>3</sup> <https://interpret.ml/>

<sup>4</sup> <https://github.com/MAIF/shapash>

<sup>5</sup> <https://explainerdashboard.readthedocs.io/en/latest/>

- **Robustness and Validation:** Another pertinent aspect is the robustness and validation of the explanations generated by SHAP and LIME. Highlighting potential scenarios where these XAI packages might provide misleading or inaccurate explanations is crucial for ensuring trustworthiness and reliability in agricultural decision-making.

Addressing these concerns can contribute to enhancing the practical utility and reliability of these XAI packages in facilitating transparent and interpretable decision-making processes in agricultural settings.

## 7 Conclusions and future work

In this paper, we have discussed how XAI can help people understand why a model selects a specific class and the logic that leads it to recommend a particular crop. Through the graphic options of both SHAP and LIME, we have been able to obtain explanations for single predictions and intuitive representations of how a specific model predicts a class. The visualizations that both packages provide are easy to understand and allow to have an immediate and intuitive comprehension of the explanation. More in detail, we can make different observations, as it follows. Common tendencies have been found among different models: for instance, rice being misclassified in favor of jute. This is also because heavy rainfall leads to the growth of both classes. It is possible to investigate single, misclassified observations. We can understand how the model behaves under the hood and why it selects a crop different from the expected one. Especially through the SHAP summary plot, we can have an intuitive idea of which class will be predicted according to the input data. This is possible even without knowing the mathematical rules included in the explanation model. We can investigate single classes and the knowledge learned by the model. Last but not least, even non-ML experts can partially understand explanations for single predictions and which features are the most important for that particular prediction. This last element is extremely important because it can open the way to discuss with agronomists and farmers on what the model has learnt. Both SHAP and LIME make the original models transparent and, regardless of their complexity, allow us to make comparisons between what a ML model learned and what farmers and agronomists know, which should always be our first concern if our target is a model that is not just accurate but also trustworthy. Future work will regard the improvement of the approach by exploiting different XAI approaches and visualization techniques, as well as using the XAI approaches in different multidisciplinary fields like computational creativity [30, 31], and emotion detection [32, 33]. Another relevant line of research consists in embedding *flexibility* in our proposed framework, for instance, by adopting a semi-structured data representation format (e.g., [34–37]), which may turn out useful to align data with AI explanations (e.g., [38–41]).

**Acknowledgements** This work was funded by the Next Generation EU - Italian NRRP, Mission 4, Component 2, Investment 1.5 (Directorial Decree n. 2021/3277) - project Tech4You n. ECS0000009.

**Author Contributions** A.C provided part of the conceptualization and writing of the manuscript. A.C provided the main conceptualization, models and writing of the manuscript. G.P provided the case studies and writing of the manuscript. All authors reviewed the manuscript.

**Funding** Open access funding provided by Università della Calabria within the CRUI-CARE Agreement. Not applicable.

**Availability of Supporting Data** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Competing Interests** The authors declare that they have no competing interest.

**Ethical Approval and Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

**Human and Animal Ethics** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Valecce G, Strazzella S, Radesca A, Grieco LA (2019) Solarfertigation: internet of things architecture for smart agriculture. In: 2019 IEEE International conference on communications workshops (ICC workshops), IEEE, pp 1–6
2. Qazi S, Khawaja BA, Farooq QU (2022) Iot-equipped and ai-enabled next generation smart agriculture: a critical review, current challenges and future trends. *IEEE Access* 10:21219–21235
3. Sharma A, Georgi M, Tregubenko M, Tselykh AA, Tselykh AN (2022) Enabling smart agriculture by implementing artificial intelligence and embedded sensing. *Comput Ind Eng* 165:107936
4. Shaikh FK, Memon MA, Mahoto NA, Zeadally S, Nebhen J (2022) Artificial intelligence best practices in smart agriculture. *IEEE Micro* 42(1):17–24
5. Ray PP (2017) Internet of things for smart agriculture: technologies, practices and future direction. *J Ambient Intell Smart Environ* 9(4):395–420
6. Linardatos P, Papastefanopoulos V, Kotsiantis S (2020) Explainable ai: a review of machine learning interpretability methods. *Entropy* 23(1):18
7. Kotsiantis SB, Zaharakis ID, Pintelas PE (2006) Machine learning: a review of classification and combining techniques. *Artif Intell Rev* 26(3):159–190
8. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R et al (2020) Explainable artificial intelligence (xai): concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf Fusion* 58:82–115
9. Tsakiridis NL, Diamantopoulos T, Symeonidis AL, Theocharis JB, Iossifides A, Chatzimisios P, Pratos G, Kouvas D (2020) Versatile internet of things for agriculture: an explainable ai approach. In: IFIP International conference on artificial intelligence applications and innovations, Springer, pp 180–191
10. Cuzzocrea A, Saccà D, Serafino P (2006) A hierarchy-driven compression technique for advanced OLAP visualization of multidimensional data cubes. In: Data warehousing and knowledge discovery, 8th international conference, DaWaK 2006, Krakow, Poland, September 4–8, 2006, Proceedings, pp 106–119
11. Barkwell KE, Cuzzocrea A, Leung CK, Ocran AA, Sanderson JM, Stewart JA, Wodi BH (2018) Big data visualisation and visual analytics for music data mining. In: 22nd International conference information visualisation, IV 2018, Fisciano, Italy, July 10–13, 2018, pp 235–240
12. Jin BT, Xu F, Ng RT, Hogg JC (2022) Mian: interactive web-based microbiome data table visualization and machine learning platform. *Bioinform.* 38(4):1176–1178
13. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30
14. Ribeiro MT, Singh S, Guestrin C (2016) “why should i trust you?” explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144

15. Cuzzocrea A, Pilato G, Cartolano A (2022) explainable AI for smart agriculture. In: The 28th international DMS conference on visualization and visual languages, DMSVIVA 2022, KSIR Virtual Conference Center, USA, June 29-30, 2022, pp 69–76
16. Barcelo-Ordinas JM, Chanet J-P, Hou K-M, García-Vidal J (2013) A survey of wireless sensor technologies applied to precision agriculture. In: Precision agriculture'13, pp 801–808
17. Langone R, Cuzzocrea A, Skantzos N (2020) Interpretable anomaly prediction: predicting anomalous behavior in industry 4.0 settings via regularized logistic regression tools. *Data Knowl Eng* 130:101850
18. Kenny EM, Ruelle E, Geoghegan A, Shalloo L, O'Leary M, O'Donovan M, Keane MT (2019) Predicting grass growth for sustainable dairy farming: a cbr system using bayesian case-exclusion and post-hoc, personalized explanation-by-example (xai). In: International conference on case-based reasoning, Springer, pp 172–187
19. Nerc Centre for Ecology and Hydrology, Explainable AI for UK agricultural land use decision-making, UKRI (United Kingdom Research and Innovation) (2022). <https://gtr.ukri.org/projects?ref=NE%2FT003952%2F1>
20. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B (2019) Interpretable machine learning: definitions, methods, and applications. [arXiv:1901.04592](https://arxiv.org/abs/1901.04592)
21. Kaufman S, Rosset S, Perlich C, Stitelman O (2012) Leakage in data mining: formulation, detection, and avoidance. *ACM Trans Knowl Discov Data (TKDD)* 6(4):1–21
22. Lundberg SM, Erion GG, Lee S-I (2018) Consistent individualized feature attribution for tree ensembles. [arXiv:1802.03888](https://arxiv.org/abs/1802.03888)
23. Shrikumar A, Greenside P, Kundaje A (2017) Learning important features through propagating activation differences. In: International conference on machine learning, PMLR, pp 3145–3153
24. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H (2020) Fooling lime and shap: adversarial attacks on post hoc explanation methods. In: Proceedings of the AAAI/ACM conference on AI, ethics, and society, pp 180–186
25. Lakkaraju H, Bastani O (2020) “how do i fool you?” manipulating user trust via misleading black box explanations. In: Proceedings of the AAAI/ACM conference on AI, ethics, and society, pp 79–85
26. Molnar C (2019) Interpretable Machine Learning. <https://christophm.github.io/interpretable-ml-book/>, <https://christophm.github.io/interpretable-ml-book/>
27. Visani G, Bagli E, Chesani F, Poluzzi A, Capuzzo D (2022) Statistical stability indices for lime: obtaining reliable explanations for machine learning models. *J Operat Res Soc* 73(1):91–101
28. Nori H, Jenkins S, Koch P, Caruana R (2019) Interpretml: a unified framework for machine learning interpretability. [arXiv:1909.09223](https://arxiv.org/abs/1909.09223)
29. Bhatnagar P (2021) Explainable AI (XAI) — A guide to 7 Packages in Python to Explain Your Models. <https://towardsdatascience.com/explainable-ai-xai-a-guide-to-7-packages-in-python-to-explain-your-models-932967f0634b>
30. Augello A, Infantino I, Pilato G, Rizzo R, Vella F (2013) Introducing a creative process on a cognitive architecture. *Biologically Inspired Cognit Architect* 6:131–139
31. Augello A, Infantino I, Pilato G, Rizzo R, Vella F (2015) Creativity evaluation in a cognitive architecture. *Biologically Inspired Cognit Architect* 11:29–37
32. Cuzzocrea A, Pilato G (2018) Taxonomy-based detection of user emotions for advanced artificial intelligent applications. In: International conference on hybrid artificial intelligence systems, Springer, pp 573–585
33. Pilato G, D'Avanzo E (2018) Data-driven social mood analysis through the conceptualization of emotional fingerprints. *Procedia Comput Sci* 123:360–365
34. Cannataro M, Cuzzocrea A, Mastroianni C, Ortale R, Pugliese A (2002) Modeling adaptive hypermedia with an object-oriented approach and XML. In: Proceedings of the second international workshop on web dynamics, WebDyn@WWW 2002, Honolulu, HI, USA, May 7, 2002, pp 35–44
35. Bonifati A, Cuzzocrea A (2006) Storing and retrieving xpath fragments in structured P2P networks. *Data Knowl Eng* 59(2):247–269
36. Cannataro M, Cuzzocrea A, Pugliese A (2002) XAHM: an adaptive hypermedia model based on XML. In: Proceedings of the 14th international conference on software engineering and knowledge engineering, SEKE 2002, Ischia, Italy, July 15-19, 2002, pp 627–634
37. Bonifati A, Cuzzocrea A (2007) Efficient fragmentation of large XML documents. In: Database and expert systems applications, 18th international conference, DEXA 2007, Regensburg, Germany, September 3-7, 2007, Proceedings, pp 539–550
38. Pölsterl S, Aigner C, Wachinger C (2021) Scalable, axiomatic explanations of deep alzheimer's diagnosis from heterogeneous data. In: Medical image computing and computer assisted intervention - MICCAI 2021 - 24th international conference, Strasbourg, France, September 27 - October 1, 2021, Proceedings, Part III. Lecture Notes in Computer Science, 12903:434–444

39. Cocarascu O, Cyrus K, Rago A, Toni F (2022) Mining property-driven graphical explanations for data-centric AI from argumentation frameworks. In: Human-Like machine intelligence, pp 93–113
40. Prasad N, Boughanem M, Dkaki T (2023) A hierarchical neural framework for classification and its explanation in large unstructured legal documents. [arXiv:2309.10563](https://arxiv.org/abs/2309.10563)
41. Zhao Z, Shen Z, Mao A, Wang H, Hu C (2023) Pandadb: An ai-native graph database for unified managing structured and unstructured data. In: Database systems for advanced applications - 28th international conference, DASFAA 2023, Tianjin, China, April 17-20, 2023, Proceedings, Part IV. Lecture Notes in Computer Science, 13946:669–673

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.