IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 March 2023, accepted 30 March 2023, date of publication 7 April 2023, date of current version 12 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265597

APPLIED RESEARCH

Classification and Forecasting of Water Stress in Tomato Plants Using Bioristor Data

MANUELE BETTELLI[1], FILIPPO VURRO[1], RICCARDO PECORI[1,2], MICHELA JANNI[1], NICOLA COPPEDÈ[1], ANDREA ZAPPETTINI[1], AND DANIELE TESSERA[3]

[1]Institute of Materials for Electronics and Magnetism (IMEM), National Research Council, 43124 Parma, Italy
[2]SMARTEST Research Centre, eCampus University, 22060 Novedrate, Italy
[3]Faculty of Mathematical, Physical and Natural Sciences, Catholic University of the Sacred Heart, 25133 Brescia, Italy

Corresponding author: Riccardo Pecori (riccardo.pecori@cnr.it)

This work was supported in part by the E-Crops—Technologies for Digital and Sustainable Agriculture Project funded by the Italian Ministry of University and Research (MUR) through the Programma Operativo Nazionale (PON) Agrifood Program under Contract ARS01_01136.

ABSTRACT Water stress and in particular drought are some of the most significant factors affecting plant growth, food production, and thus food security. Furthermore, the possibility to predict and shape irrigation on real plant demands is priceless. The objective of this study is to characterize, classify, and forecast water stress in tomato plants by means of in vivo real time data obtained through a novel sensor, named bioristor, and of different artificial intelligence models. First of all, we have applied classification models, namely Decision Trees and Random Forest, to try to distinguish four different stress statuses of tomato plants. Then, we have predicted, through the help of recurrent neural networks, the future status of a plant when considering both a binary (water stressed and not water stressed) and a four-status scenario. The obtained results are very good in terms of accuracy, precision, recall, F-measure, and of the resulting confusion matrices, and they suggest that the considered novel data and features coming from the bioristor, together with the used machine and deep learning models, can be successfully applied to real-world on-the-field smart irrigation scenarios in the future.

INDEX TERMS AI modeling and forecasting, bioristor, precision agriculture, recurrent neural network, tomato plants, tree-based classifiers, smart irrigation, water stress.

I. INTRODUCTION

Drought is one of the major drivers of water stress and yield losses in agro-ecosystems. The year 2022 has seen one of the most severe water shortage all over Europe since the beginning of the recorded historical series of data. In particular, Italy's severe drought has caused crop yields to fall by up to 45%.[1]

Water and heat stress have substantially reduced summer crop yields, with grain maize, soybeans, and sunflowers being the most affected crops. In this context, the rational use of the water resources in agriculture is mandatory to achieve a more sustainable and successful food production. Water stress negatively affects many physiological functions of plants, including photosynthesis, transpiration, and nutrient uptake, and reduces vegetative growth and crop yield, thus threatening food security [13], [14]. The severe drought affecting many regions of Europe further expanded and worsened as of early August 2022. Dry conditions were related to a wide and persistent lack of precipitations combined with a sequence of heatwaves from May 2022 onward [15]. In this scenario, the early detection of drought, and the water stress characterization and prediction via different approaches are critical for agritech decision support systems and more sustainable water management [16].

As a result, increasing efforts are made towards drought characterization and modeling thanks to machine learning (ML) approaches and artificial intelligence (AI) techniques. The possibility to predict the onset and termination of droughts, given the critical stage in mitigating the effects

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

[1]https://www.ansa.it/english/news/general_news/2022/07/25/drought-crop-yields-down-by-up-to-45-coldiretti_a926e415-5ecd-4bed-9f59-2f2197d8aad8.html

of drought, will be key to achieve the goal of agriculture sustainability. So far, a major challenge is the development of a method to accurately predict plant drought conditions for the upcoming short-/medium term period [17].

Recently, a novel in vivo sensor, named ''bioristor,'' was developed, patented, and applied in plants [18], [19], [20], [21]. Bioristor provided new insights in the dynamic changes of the chemical composition of the sap in the xylem, occurring in drought-stressed tomato and grapevine plants [18], [19]. The bioristor has been used as a smart sensor for precision phenotyping in greenhouse conditions to fine tune the regulation of Vapor Pressure Deficit (VPD) and achieve increased water use efficiency and yield [19]. Moreover, salinity stress was detected in a specific manner, allowing one to hypothesize the ability to define the use of sea water for tomato irrigation and improve fruit quality by constantly monitoring the plant health, boosting water reuse, and improving quality without compromising the yield itself [20].

The main contribution of this paper is twofold. The first one is the exploitation, for the first time, of the data acquired through the bioristor, during previously reported experiments in a controlled environment, in order to build a proper and innovative feature model to classify, by using traditional machine learning techniques, four possible water-stress statuses of a tomato plant. The second one regards taking advantage of the novel bio-electrical features, obtained from the bioristor, to predict, 24 hours ahead and by using LSTM-based neural networks, the status of a tomato plant, in order to be able to automate the irrigation process and allow the plant to be healthy and fruit-bearing as much as possible.

The rest of the paper is divided into the following sections. Section II summarizes some recent work about the application of AI techniques to precision agriculture and water stress detection. Section III accurately details the practical experiments on the plants and the data collection, dealing with how the tomato plants have been cultivated in a controlled environment as well as which novel bio-electrical data and features have been extracted from the bioristor. Section IV summarizes the main theoretical characteristics of the AI models we applied to the bioristor data and features. Section V reports the results we obtained, in terms of accuracy, precision, recall, F-measure, and confusion matrices, for both the classification and the prediction analyses. Section VI discusses the main achievements and observations that can be drawn from the results presented in Section V, while Section VII draws some conclusions and highlights possible future research directions on the prediction and classification of water stress in plants, by using the data of the bioristor.

II. RELATED WORK

The usage of artificial intelligence, and specifically machine learning techniques, in the agricultural field has been an increasing topic in the last ten years [22]. For example, in [1] the authors try to detect leaf miners in tomato plants by applying two types of deep neural networks to classify and segment plant images, while in [2] a convolutional neural network is used to predict some growth indices of lettuce plants, always using image data. In [3] the target crop regards apples and in particular the classification of their maturation degree in order to facilitate the work of picking robots. This is achieved by exploiting histograms of oriented gradients, SVM, and a fast identification technique for multiple targets, very suitable for a complex occlusion environment. The focus of the contribution in [4] is always the apple tree, but with the aim of classifying leaf diseases by using image data and deep neural networks with a robust speed up feature technique. In [5] the focus is the nitrogen status of wheat crops, predicted through artificial neural networks and genetic algorithms on image data.

In particular, water status estimation and smart irrigation via machine learning techniques and for different plantations have been thoroughly studied and reviewed [23], [24], also in the context of remote sensing and decision support systems [25]. For example, Sharma et al. [5] try to estimate soil moisture by exploiting artificial neural networks on meteorological data to improve rice crop yield, while in [8] machine learning techniques are applied to an IoT-based smart irrigation system to predict water needs in various crops when considering soil moisture, air temperature, relative humidity, and ultraviolet radiation as input data. Also in [7] an IoT-based smart irrigation system is proposed for various crops, namely spinach, beans, carrots, walnuts, corn, barley, and maize. The module considers soil evaporation and plant transpiration as input of a deep neural network predictive model.

Romero et al. [9] focused on vineyards by using multi-spectral images from an Unmanned Aerial Vehicle and artificial neural networks to estimate their stem water potentials. Revathy and Balamurali [10] try to predict, through traditional and deep neural networks, enhanced by the firefly optimization algorithm, the irrigation amount needed for sugarcane crops in the next 30 years. As a benchmark, they exploit future forecasts by the Community Climate System Model and they achieve an accuracy well beyond 90% on four different future irrigation strategies. Nagappan et al. [11] try to predict evapotranspiration for smart irrigation using optimized deep neural networks. Their model is based on a multivariate analysis of correlated variables, such as, for example, temperature, wind speed, sunshine hours, etc., whose number has been reduced to three via principal component analysis, while the data samples were collected from 1995 to 2016 in Veeranam, India. Farooque et al. [12] face the same problem, the daily forecasting of evapotranspiration, using a similar set of features, but applying different deep neural network models, namely, LSTM, 1D-CNN, and convolutional LSTM. Moreover, the data were collected across Prince Edward Island, Canada, from 2011 to 2017 and the dataset was split into four subsets, corresponding to the four seasons.

Table 1 summarizes the main characteristics of the reviewed researches, highlighting the used input data, the used artificial intelligence techniques, the target objective, and the performed task (classification or prediction).

TABLE 1. Summary of the reviewed related work.

Reference	Crop	Sensed data	Target	Machine/Deep Learning	Class./Pred.
[1]	Tomato	images	leaf miner	DL	class.
[2]	Lettuce	images	growth indices	DL	pred.
[3]	Apple	images	fruit maturation	ML	class.
[4]	Apple	images	leaf disease	DL	class.
[5]	Wheat	images	nitrogen status	ML	pred.
[6]	Rice	meteorological data	soil moisture	DL	pred.
[7]	various	soil evaporation, plant transpiration	water stress	DL	pred.
[8]	various	soil moisture, air temperature, relative humidity, ultraviolet radiation	water stress	ML	pred.
[9]	Vineyard	images	water stress	ML	pred.
[10]	Sugarcane	soil permanent wilting point and weather forecast	water stress	ML/DL	pred.
[11]	various	temperature, wind speed, sunshine hours, etc.	water stress	ML/DL	pred.
[12]	various	air temperature, solar radiation, relative humidity, and wind speed	water stress	DL	pred.

In contrast to the them, the main novelty of the study proposed in this paper, compared to the relevant literature, is the application, for the first time, of machine learning techniques, namely Decision Tree-based methods and recurrent neural networks, to the data of a novel and recently patented in vivo sensor, named "bioristor," by using a novel and proper feature set, for both the classification and prediction of water stress in tomato plants, and so foster a proper smart irrigation system thereof.

III. EXPERIMENTS AND DATA COLLECTION

A. TOMATO PLANTS GROWTH AND DROUGHT TREATMENT

The data analyzed in the current paper are those reported in [20], including the details of the plant growth and the drought imposition.

In summary, two experiments in controlled conditions have been performed.

In the first experiment, seven plants of the cultivar (cv.) Red Setter were grown up to the stage of 5^{th} fully expanded leaves. The plants were kept fully irrigated until their fifth true leaf had fully expanded, after which a bioristor was inserted in the stem of each plant. After three days, four of the plants were exposed to drought stress by withholding watering for 14 days; the plants were then irrigated over two days of recovery, and, finally, a 7-day stress episode was imposed by withholding water again. A set of three plants was kept fully watered as a control set.

On the basis of the results obtained in the pilot experiment, the same experiment was performed in the greenhouse of the ALSIA plant phenomics facility (Metaponto, Italy), on cv. Ikram. When plants reached the 5^{th} fully expanded leaf stage, the bioristors were integrated into the plants, and one day after the implantation of the bioristor, watering was withheld from four of the plants for 16 days to impose the drought stress and then they were restored for a recovery phase of seven days.

B. BIORISTOR PREPARATION

As concerns the inherent structure of the bioristor, it is a biosensor based on an Organic Electrochemical Transistor (OECT). It is made of two textile fibers, treated by soaking for 5 minutes in aqueous poly (3,4-ethylenedioxythiophene) doped with polystyrene sulfonate (Clevios PH1000, Starck GmbH, Munich, Germany) after which ethylene glycol (10%v/v) and dodecyl benzene sulfonic acid (2%v/v) were added. Fibers were then baked at 150°C for 45 minutes in three different steps. Finally, the whole process, from deposition to heat treatment, was repeated three times to complete the preparation. Before functionalization, each thread was cleaned by plasma–oxygen cleaner treatment (Femto, Diener electronic, Ebhausen/Germany) to increase its wettability and to facilitate the adhesion of the aqueous conductive polymer solution.

C. DATA WORKFLOW

After the novel bioristor was prepared, it was inserted into the plant stem (Figure 1), and connected to a computer as indicated in [26]. More in detail, data were acquired by means of a customized local control unit including a digital acquisition board, namely a NI USB-6343 multifunction I/O device (National Instruments, Austin, TX, USA), connected to the bioristor through electrical wires. Moreover, the board was endowed with a multi-channel analog-to-digital converter connected to a PC, where the readout data were processed by a custom software application and then saved into the Cloud. The sensor currents were converted to voltage, more easily readable by the NI board, through a resistor. Resolutions were 8 μA and less than 100 nA for drain-source currents and gate-source currents, respectively. The data from each connected bioristor were acquired every second, locally saved on the PC, and then sent to the Cloud wireless connections.

Indeed, this was only a possible first experiment performed in a controlled scenario. The final scenario, where we expect to apply the bioristor, is depicted in Figure 2, wherein the complete workflow of data is represented.

As a matter of fact, we expect to apply the bioristor to several plants in real crop fields, wire-connecting it to a local processing unit that applies bias and reads currents from all bioristors and sends, via 4G or LoRA-WAN [27] connection, the collected data to the Cloud, where the AI engine will analyze them in order to recognize the presence of hydric stress. If hydric/water stress is present or is going to start in the next hours, the AI engine will send via 4G or LoRA-WAN

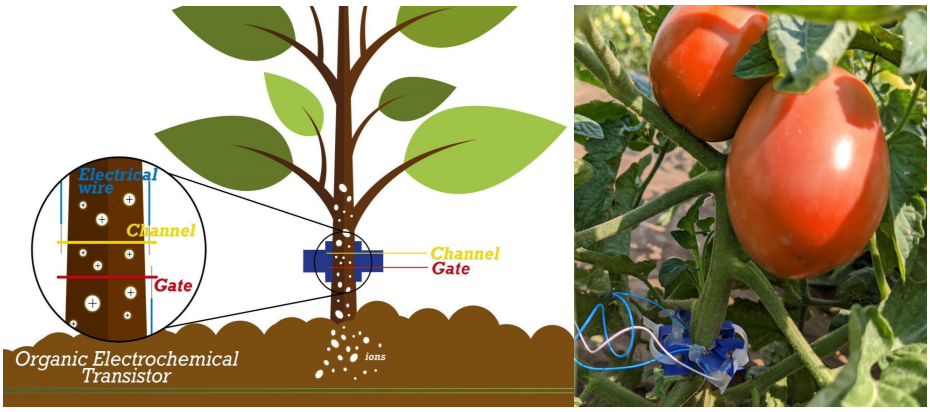

FIGURE 1. Scheme of the bioristor (left) and photo of the sensor installed on a tomato plant (right) in an open field.

a command to the local smart actuator, i.e., an automated irrigator.

D. DATA COLLECTION AND CONSIDERED FEATURES

The bioristor is utilized to measure the concentration and movement of ions in the sieve tubes, with a specific focus on xylem vessels, which are responsible for the unidirectional transport of water and mineral nutrients from roots to aerial tissues through the transpiration stream [28]. The bioristor features are the result of the relationship between the sensor and the changes occurring in the plant sap composition and ion concentration during growth, development, and under abiotic stress. When stomata are closed and the transpiration stream is blocked, there is a reduction in water loss that affects both ion concentration and the sensor wetness status [29]. The concentration of ions decreases as a result of the compartmentalization of ions to balance the water potential inside the plant xylem. Additionally, the water content in the plant decreases due to the blockage of transpiration and reduced water uptake from the soil [29].

The application of positive gate potential in the bioristor results in a decrease of the current flowing into the channel, caused by the entry of positive ionic charges into the polymer. The bioristor was operated by applying a constant voltage ($V_{ds} = -0.05V$) to drain and source terminals across the main transistor channel. When a negative V_{ds} is applied, the holes in the channel will flow from drain to source, generating an I_{ds} current in the channel; V_g is applied to the gate terminal and it is set to 0. Turning on V_g with a positive value, the cations of the electrolyte are pushed into the channel, where they interact with the organic semiconductor, causing an alteration of its doping level and a consequent decrease in I_{ds} [30]. The modulation of the current between on and off status (R_{ds}), proportional to the cations present in the plant xylem sap, is given by the following formula: $R_{ds} = |I_{ds} - I_{ds0}|/I_{ds0}$, where I_{ds} and I_{ds0} are the current across the channel when V_g is ON and when it is OFF, respectively. Upon application of a positive gate bias, the gate electrode and the device channel charge capacitively, then a current flowing also through the

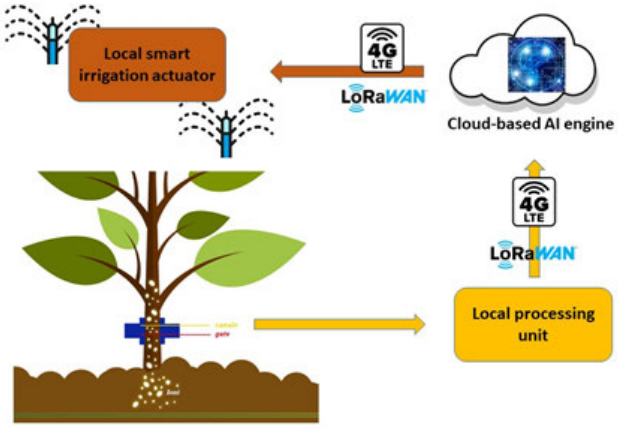

FIGURE 2. Workflow of the bioristor data processing pipeline, with data coming from the plants to the actuator, passing through the local processing collector unit and the cloud-based AI engine.

liquid from the gate to the main channel (I_{gs}) was monitored continuously. The difference $\Delta I_{gs} = I_{gs} - I_{gs0}$ was calculated to evaluate the sensor wetting status, where I_{gs0} represented the current across the solution when $V_g = 0$ [29], [31]. Moreover, two time constants have been calculated by fitting the non-linear drain and gate current curves (t_{ds} and t_{gs}), which are related to the inverse of both cation quantity and mass [21], [32]. Time constants are related to the time ions enter the polymer (t_{ds}), and to the diffusivity of ions in the solution (t_{gs}). The value of t_{ds} constant is derived from the formula: $A \times (1 - e^{-t/t_{ds}}) + B$, which represents how fast the curve reaches a plateau situation.

Therefore, the presence of different ionic species that may have overlapping trends is not taken into consideration. The change of t_{ds} as well as the composition of the effect of all ionic species will be assessed. With the same approach t_{gs} is calculated by fitting the curves of I_{gs} current as a function of time (Figure 3) resulting from the formula: $A \times (e^{-t/t_{gs}}) + B$. Under drought stress, R_{ds} and ΔI_{gs} drop significantly as a result of the diminished ion concentration and amount of

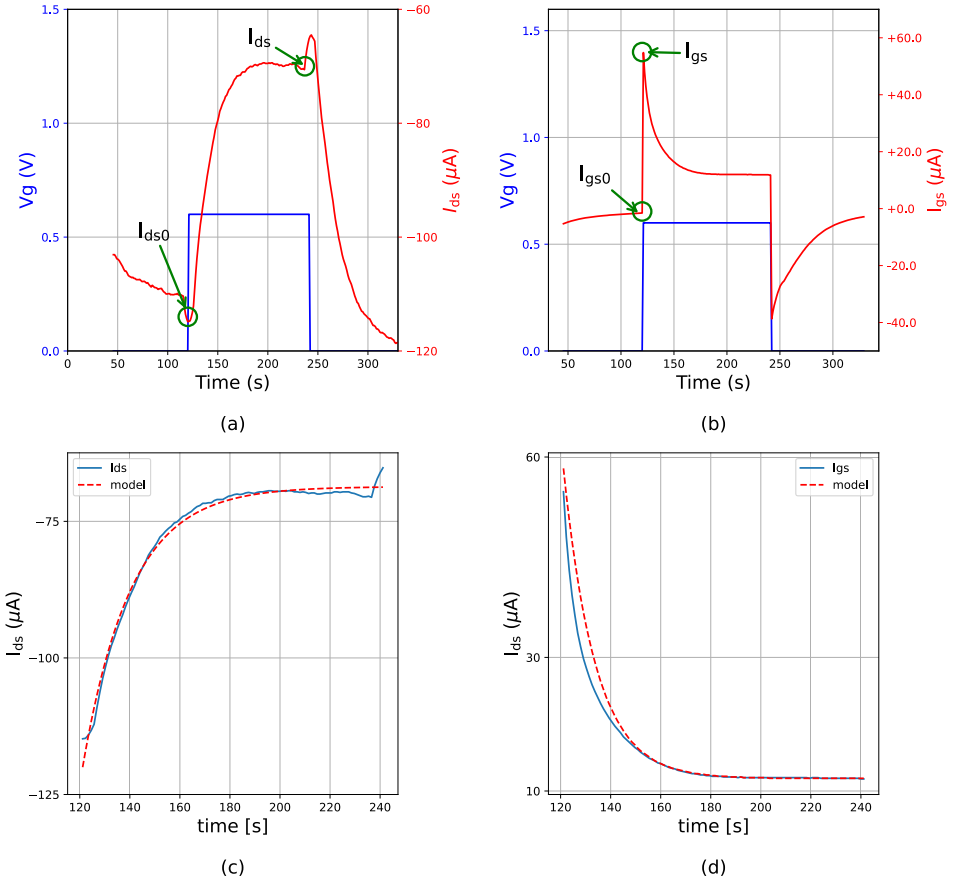

FIGURE 3. a) I_{ds} modulation when gate voltage is applied in OECT. Green circles indicate the drain-source current when $V_g = 0$ (I_{ds0}), and the drain source-current when $V_g > 0$ (I_{ds}); b) I_{gs} behavior when gate voltage is applied in OECT. Green circles indicate the gate-source current when $V_g = 0$ (I_{gs0}), and the gate-source current when $V_g > 0$ (I_{gs}); c) I_{ds} curve fitted with an exponential model; d) I_{gs} curve fitted with an exponential model.

water in the plant sap, as a consequence of the low plant transpiration and root water absorption [19], [29].

The considered dataset is composed of observations gathered by ALSIA and IMEM facilities related to 13 tomato plants. Each observation consists of four numerical parameters and one qualitative parameter (the status label or class). Measurements and classifications were sampled every 15 minutes for about 14 and 16 days, depending on the experimental facility. More precisely, the dataset is defined as: $D = (Rds_{i,t}, \Delta Igs_{i,t}, tds_{i,t}, tgs_{i,t}, status_{i,t})$ where i is the plant identifier (Experiment, Plant), t is the measurement sampling time, and $status_{i,t} = \{healthy, uncertain, stress, recovery\}$ is the target classification label provided on the basis of the acquired physiological and morphometric traits. Indeed, at 6 days after drought initiation, the plant triggered physiological responses such as the reduction of the stomatal conductance and significant reduction of the digital bio-volume [19].

Figure 4 shows the box-plots for the numerical features we have considered, whereas Figure 5 reports the frequency distribution of the four considered statuses, i.e., "healthy" representing tomato plants for sure in good conditions, "stress" representing tomato plants for sure in a stress condition (at least six days passed after the last irrigation), "recovery" which represents the status of tomato plants watered after a drought stress period, and "uncertain" which is the status of tomato plants in the six days between the end of the irrigation and a sure stress condition revealed. As can be seen from the figure, all classes are well-represented. The excess of instances of the healthy class, compared to the other classes, is explained by the experimental design that involved monitoring healthy plants and stressed plants periodically, throughout the duration of the experiments.

IV. MACHINE LEARNING MODELS

In the carried out analyses we have tested several machine and deep learning models, but in this paper we only present the best obtained results. Thus, in the following, we only describe the best performing models to both classify and predict the water stress status of a tomato plant by using bioristor data.

As for the classification, we employed Decision Tree-based models, namely standard Decision Tree, which is also an explainable model, and an ensemble technique, i.e., Random Forest, which is a black-box model. This choice comes

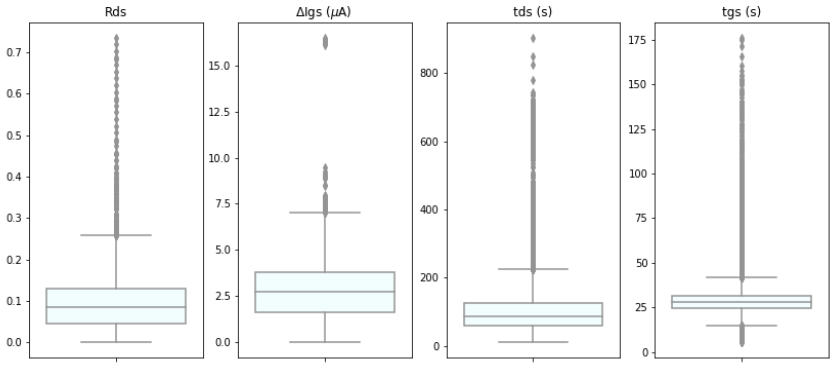

FIGURE 4. Box-plots showing the variability of the four considered features in the considered dataset.

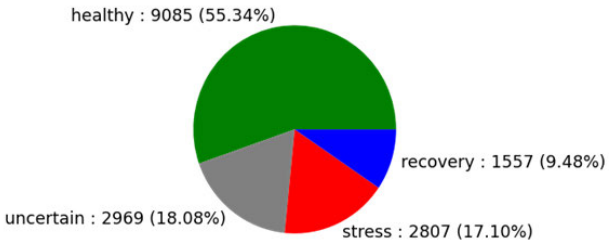

FIGURE 5. Pie chart showing the distribution of the four regarded health statuses in the considered dataset.

from the twofold aim to achieve both top performance and the most interpretable results.

Conversely, as regards the prediction part, we used a neural network-based model, i.e., a recurrent neural network (RNN) with long short-term memory (LSTM), focusing only on performance. All models have been implemented by using Python 3.9 and its SciKitLearn package.

A. DECISION TREE

A Decision Tree can be regarded as a directed acyclic graph, whose internal nodes iteratively execute a test on a certain feature, while edges determine the outcome of the test, and terminal nodes, i.e., the so-called ''leaves,'' can have instances belonging to one or more class labels. External nodes are: the first node to be created, named ''root'' of the tree, and the leaves. Generally, in each leaf node, each class k is associated with a weight w_k, which determines the strength of the class in the leaf node itself. During the training phase the Decision Tree model is built recursively exploiting the training set according to the following procedure [33]:

- one of the features is chosen as the root, maximizing a certain metric, usually either the Information Gain or the Gini Index.
- Each node, except for the leaves, has a set of outbound edges and corresponding inbound children nodes. A Decision Tree can be binary (only two outbound edges per node) or multi-path (multiple outbound edges per node). Each outbound edge corresponds to the fulfillment of a condition on the values of the tested feature.
- Recursively, for each internal node, a new feature is selected from the set of the features of the training set, considering only the instances satisfying the test associated with the edge itself. When no more features can be selected, the node is considered as a leaf.

Once the construction of the Decision Tree has terminated, the classification of a new, never seen, instance of the test set follows the following steps:

- the given unlabeled instance follows an activation path of nodes from the root to one leaf, passing different tests on the features. The path to follow is determined by the specific values of the features in the considered instance.
- each leaf is labeled with a unique class label. This determines that the instance to classify is assigned to the class corresponding to the reached leaf.

Moreover, the structure and representation of a Decision Tree model is inherently explainable, because the numerical tests on the different features can be considered as the antecedents of IF-THEN rules with the following structures:

$$\textbf{IF}\ \text{condition on}\ f_1\ \textbf{AND}\ \text{condition on}\ f_2\ \textbf{AND}\ \ldots.\ \textbf{THEN}\ \ i_k\ \text{is}\ \mathit{Class}_j \qquad (1)$$

where i_k is the k^{th} instance to classify and Class_j is the j^{th} class label in the set of all possible class labels.

B. RANDOM FOREST

Ensembles are machine learning techniques grouping a set of classifiers together in order to perform a combined voting classification [34]. Practically, the outcome of each classifier is taken into account as a vote for the final decision and all votes are combined, according to specific rules, to issue the final output.

These techniques create a sort of super classifier, much more accurate than any single composing classifier by itself, because the single algorithms may suffer from issues usually solved by minimizing the following two kinds of errors in the ensemble: variance in sensitivity and bias in the model

itself. Moreover, ensemble techniques usually exploit the same basic classifier repeated n times and the output of each classifier is different because of different allocations strategies of the training samples. Indeed, ensemble methods can be classified not only according to the voting strategy, but also according to their training allocation strategy [35].

Random Forest is an ensemble method, made up of various Decision Trees, that exploits bagging and features randomness in order to create an uncorrelated forest of Decision Trees themselves [36]. These allocation strategies create a random subset of features, ensuring low correlation among Decision Trees. As for the final voting strategy, Random Forest usually decides by considering the majority voting techniques, i.e., choosing the class label output the most among the considered trees, or the average predicted class label of the single Decision Trees making up the forest itself.

Indeed, Random Forest only selects a subset of the features, whereas Decision Trees regard all the possible feature splits, marking a key difference between both techniques. Another important difference is that Random Forest is a black-box model, failing to be interpretable and explainable.

C. RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNNs) [37] are used to analyze and process sequences of data with a certain chronological order by sharing parameters across different parts of the model itself; thus they are sequential networks allowing for the information to persist while the training goes on following the temporal sequence of data. This past information can be also used when processing a current input datum. However, the main disadvantage of an RNN is that it cannot remember long term dependencies because of the vanishing gradient issue.

LSTMs [38] are advanced RNNs, able to handle the vanishing gradient problem threatening RNNs, and thus explicitly designed to avoid long-term dependency problems. They are suitable to process time series data as well as lags of unknown duration between important events in a time series itself. The basic component unit of an LSTM network is made up of a cell, an input gate, an output gate, and a forget gate. Cells are used to remember values over arbitrary time intervals, while gates are used to regulate the information flow entering and exiting a cell.

V. RESULTS

In this section, we present the obtained results for both the classification and the prediction tasks we have carried out.

A. CLASSIFICATION

As regards classification, we have compared the results of an explainable model, i.e., Decision Tree (DT), and of a black-box model, i.e., Random Forest (RF).

The main hyper-parameters of both classifiers assume the default values as in their SciKit-Learn implementation. We have only varied the splitting criterion in the Decision Tree by considering both the Gini impurity index and the Information Gain, while the number of tree-classifiers in the Random Forest model is exactly the default one, i.e., equal to 100. In the Random Forest model we have also set the random state parameter, in order to provide randomness to the bootstrapping of the instances used to build the tree-classifiers and simultaneously to the feature sampling when considering the best splitting at each internal node.

TABLE 2. Performance metrics over a 5-fold cross validation for the considered classification models.

	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
DT (Gini)	92.24 ± 0.83	92.26 ± 0.75	92.28 ± 0.87	92.11 ± 0.87
DT (InfoGain)	92.30 ± 0.72	92.27 ± 0.70	92.38 ± 0.63	92.36 ± 0.54
RF	94.83 ± 0.42	94.80 ± 0.42	94.83 ± 0.42	94.80 ± 0.42

In Table 2, we report the accuracy and the weighted precision, recall, and F-measure (F1) obtained in a 5-fold cross validation classification test, when using the aforementioned models. The table shows the average of the values as well as the relative standard deviation. Moreover, in the case of Decision Trees, it presents the outcomes when using both the well-known Gini impurity and Information Gain [39] as a metric to determine the features to use for the splits in the trees.

From the outcomes of Table 2, we can infer that Random Forest improves only a little the overall performance and that, among the Decision Trees, the one reaching slightly better results is the one employing Information Gain as a metric to determine the splitting features. The good performance is also evident in the confusion matrices depicted in Figures 6, 7, and 8. Figures 6 and 7 report the confusion matrices for the Decision Tree models using Gini and Information Gain, respectively. Figure 8 shows the confusion matrix for the Random Forest model. Looking at the confusion matrices, when using a Decision Tree and the Gini index, the ''Healthy'' class is correctly recognized in the 96.21% of the cases, the ''Uncertain'' class in the 88.75% of the cases, the ''Stress'' class in the 88.28% of the cases, and the ''Recovery'' class in the 82.53% of the cases. When using a Decision Tree and the Information Gain, the ''Healthy'' class is correctly recognized in the 96.38% of the cases, the ''Uncertain'' class in the 88.68% of the cases, the ''Stress'' class in the 89.06% of the cases, and the ''Recovery'' class in the 82.72% of the cases. When using a Random Forest model, the ''Healthy'' class is correctly recognized in the 97.97% of the cases, the ''Uncertain'' class in the 92.25% of the cases, the ''Stress'' class in the 92.70% of the cases, and the ''Recovery'' class in the 85.23% of the cases. Table 3 summarizes the just mentioned results.

Indeed, we can state that Random Forest can help to better understand whether a tomato plant is water stressed, with an increase of 3.60% compared to the best performing Decision Tree model. The performance in detecting healthy plants is, instead, quite similar, with an increase of only 1.66% passing from the best performing Decision Tree model to Random Forest. However, being Random Forest a black box model, it lacks in interpretability, while from the Decision Tree model we can infer some rules or, however, an order of

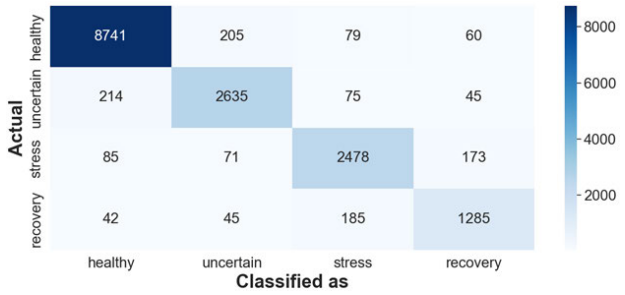

FIGURE 6. Confusion matrix when using a decision tree model and the Gini impurity index.

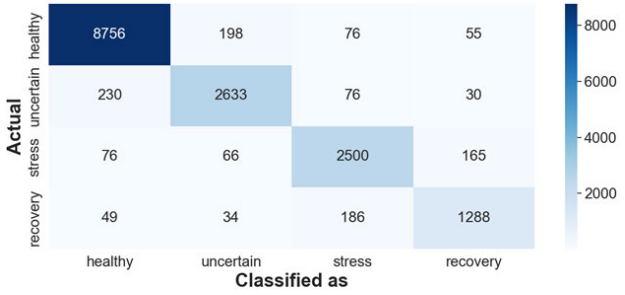

FIGURE 7. Confusion matrix when using a decision tree model and the Information Gain.

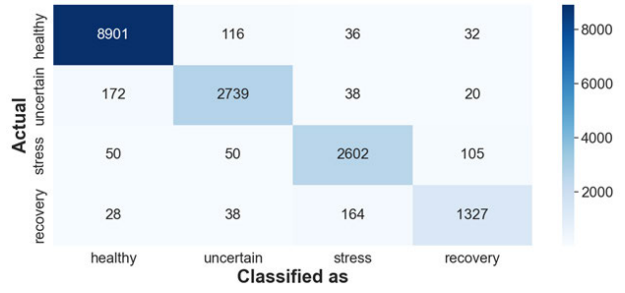

FIGURE 8. Confusion matrix when using a random forest model.

TABLE 3. Percentage of correct identification per class label.

	Healthy (%)	Uncertain (%)	Stress (%)	Recovery (%)
DT (Gini)	96.21	88.75	88.28	82.53
DT (InfoGain)	96.38	88.68	89.06	82.72
RF	97.97	92.25	92.70	85.23

importance of the analyzed features. For example, when considering the Decision Tree models using Information Gain, i.e., the best performing DT model, R_{ds} is always the root feature, i.e., the most discriminant one, followed by t_{gs} or t_{ds} at the second level, while the third level is usually occupied by R_{ds} or t_{ds} in case t_{gs} is the second most important feature, or t_{gs} in case t_{ds} is the second most important feature.

B. PREDICTION

In this section, we present the principal results obtained to predict the future status of a plant based on historical measurements about R_{ds}, ΔI_{gs}, t_{ds}, and t_{gs}. As a starting point, measurements collected every 15 minutes have been regularized by applying a four step rolling mean kernel (i.e., one hour smoothing) and sub-sampled to collect hourly measurements. In order not to clutter the presentation, in what follows, the time series index t refers to the hourly sampling time. The recent past behavior of each plant is then summarized by considering a few lagged historical values of its time series. The number of these lagged values and the time difference between value pairs is usually driven by domain knowledge, that is, the biological processes of tomatoes.

In details, we performed a grid search varying both the number of lagged measurements from 1 up to 12 hours, as well as the time difference between them. The result of this feature selection tuning was that using the current measurements and the last two lagged ones with a time difference of 3 hours accurately captured the dynamic behavior of the tomato plant and allowed for accurate forecasting. More in detail, for each plant i, for each time t, we created a 12 element sequence composed of the parameter values at time t as well as their values collected at $t-3$ and $t-6$ hours, namely: $Rds_{i,t}$, $Rds_{i,t-3}$, $Rds_{i,t-6}$, $\Delta Igs_{i,t}$, $\Delta Igs_{i,t-3}$, $\Delta Igs_{i,t-6}$, $tds_{i,t}$, $tds_{i,t-3}$, $tds_{i,t-6}$, $tgs_{i,t}$, $tgs_{i,t-3}$, $tgs_{i,t-6}$.

Moreover, we investigated various prediction time horizons, that is, for how long in the future we want to predict the plant status. For such a purpose we considered predictions one, two, three, six, twelve, twenty-four, and thirty-six hours ahead. In particular, we found the best trade-off between model complexity, accuracy, and prediction horizon times, was to consider one day ahead forecasting, i.e., 24 hours. We underline that using shorter time horizons would either improve the model accuracy or require smaller time lags for predictions.

In total, we obtained 3, 978 sequences that have been split into training and validation sets, accounting for 80% and 20% of sequences, respectively. The target of this modeling was to predict the status of the plant one day ahead (i.e., at time $t + 24$ hours) from its current time lagged measurements. As a pre-processing phase, we normalized the four features (i.e., R_{ds}, ΔI_{gs}, t_{gs}, and t_{ds}), by applying the z-score normalization. More precisely, the values of each feature x, that is, $\{x_i; 1 \leq i \leq N\}$, where N is the number of measurements, were recomputed as: $x'_i = \frac{x_i - \mu_x}{\sigma_x}$ where x'_i is the normalized value of x_i, while μ_x and σ_x are the mean and standard deviation of x, respectively. This normalization is required to cope with the heterogeneity of the considered features. Indeed, R_{ds} is a current modulation (expressed as a fractional ratio), ΔI_{gs} is the difference between drain and source currents (measured in μA), t_{ds} and t_{ds} are time constants derived to model the ion diffusion in the polymer and on the solution, respectively (measured in seconds).

Furthermore, we have derived two types of models, one full predictor that forecasts one day ahead the status (i.e., healthy, uncertain, stress, and recovery) and a simplified model that only predicts if the plant will be stressed or not stressed. Both models are based on recurrent neural networks for their capability to describe sequences of observations.

Figure 9 shows the architecture of the full predictor model, implemented using Keras[2] framework. Each input sequence

[2]https://keras.io/

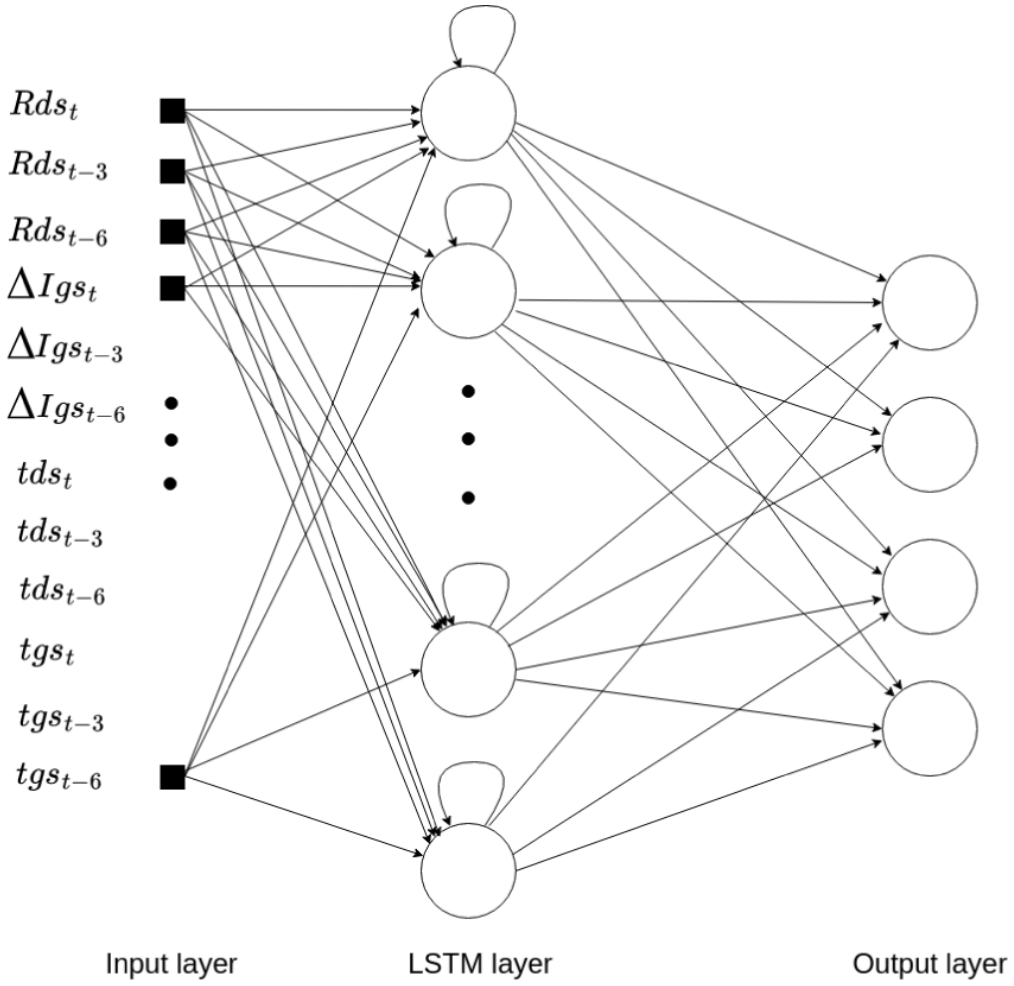

FIGURE 9. Architecture of the RNN based predictor.

is fully connected to a hidden layer composed of 30 standard LSTM cells, which, in turn, are fully connected to a four node output layer that provides the one-hot encoding of the prediction of tomato plant status. We used `ReLu` as activation function for LSTM cells, `sigmoid` as recurrent activation function, and `softmax` for the output layer. In summary, the proposed model has a total of 3,964 trainable parameters. We performed also a fine-tuning of the model hyper-parameters, varying the number of LSTM cells and also considering an additional hidden layer of neurons between the LSTM and the output layers. In particular, as a result of our testing, we found that adding another layer led to a decrease in forecast accuracy when using validation (i.e., not included in the training) sequences because of over-fitting. Similarly, increasing the number of LSTM cells did not result in any performance improvements. On the other hand, our testing has shown that for one hour ahead prediction, the optimal number of LSTM cells is 20.

The architecture of the simplified model differs from the one depicted in the figure in that the LSTM layer is composed of 15 cells and the output layer consists of only one neuron whose two states are associated with the `stressed` and `not stressed` status. These simplifications have reduced the number of trainable parameters to 1,036 only.

For the training we used, as loss function, the categorical cross entropy metric for the full predictor and the binary cross entropy for the simplified one. We have also applied custom weights, derived from the marginal frequencies of the plant status, to the loss function, in order to cope with the unbalanced occurrences of status labels themselves. Finally, for training both models we used the `Adam` [40] optimizer.

Table 4 summarizes the details of the full predictor model and of its validation.

In addition, to prevent over-fitting, the training has been stopped when there was no improvement on the loss function in the last ten epochs. Figure 10 depicts the behavior of the loss function (Fig. 10.a) and of the accuracy (Fig. 10.b) as a function of the number of processed epochs during the training of the full predictor.

In detail, the training has stopped after processing 168 out of the 300 epochs since the loss function did not get any reduction in the last ten epochs, achieving an accuracy of 95.5% in the training set and 91.1% in the validation phase. Similar results have been obtained also for the

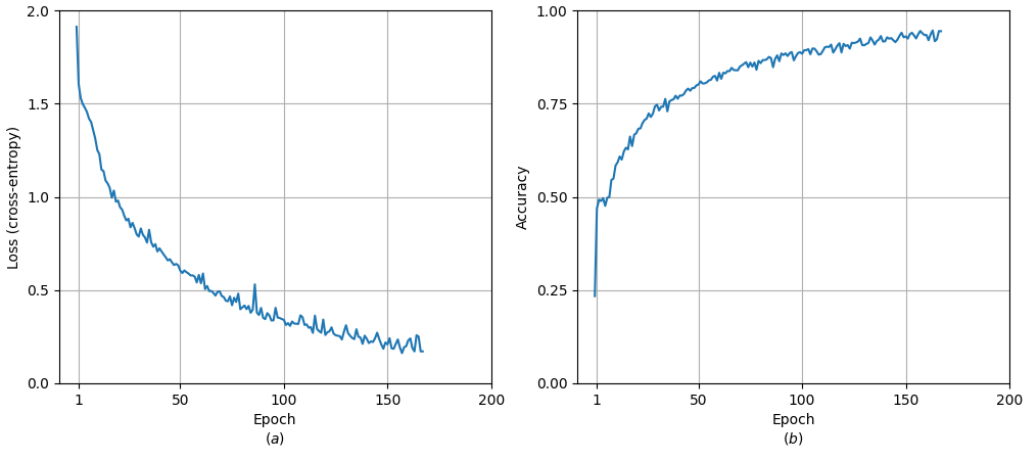

FIGURE 10. Loss function (a) and accuracy (b) as a function of the number of processed epochs for training the full predictor.

TABLE 4. Summary of the hyperparameters for the full predictor.

Hyper parameter	value
Number of input nods	12
Number of LSTM cells	30
Number of output nodes	4
LSTM activation function	`relu`
LSTM recurrent activation function	`sigmoid`
Output layer activation function	`softmax`
Optimizer	`Adam`
Loss function	categorical cross entropy
Maximum number of epochs	300
Number of KFolds	5

simplified model, where the training stopped after 209 out of 300 epochs.

In addition, to evaluate the diagnostic capabilities of our models, Fig. 11.a and Fig. 11.b plot the Receiving Operation Characteristic (ROC) curves, for the full and simplified model, respectively. Each curve plots, for the given plant status, the true positive rate as a function of the false positive rate and provides a measure of the sensitivity of the model as a function of the fall-out. For the full model, the ROC and area under the curve (AUC) are computed, on a per tomato plant status basis, as target status vs all other statuses. As can be seen from the figure both models achieve a very good accuracy.

The evaluation of the performance achieved by the proposed prediction models have been investigated by a cross validation approach. In particular, we applied a stratified KFold with 5 splits to ensure that each model fitting/validation trial had the same fraction of tomato statuses. As can be seen from the confusion matrix of the full model, shown in Figure 12, the model achieves a very good performance in predicting whether a given plant will be subject to hydric stress after 24 hours (97.40% of correct identification). A similar performance is also achieved by the simplified model, whose confusion matrix is reported in Figure 13. In this case the percentage of correct identification of the stress label is equal to 88.15%.

TABLE 5. Performance metrics for the prediction task over a stratified 5-fold cross validation for both 4- and 2-status models.

	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
4-status model	88.90 ± 1.13	89.10 ± 1.10	88.54 ± 1.12	85.85 ± 5.70
2-status model	93.35 ± 0.71	82.36 ± 2.51	81.72 ± 5.50	81.88 ± 2.46

Tables 5 summarizes the performance metrics for both models. As can be seen, the two prediction models achieve very good and comparable performance with respect to all metrics. In particular, the simplified 2-status model has an overall accuracy of about 93% in predicting the irrigation needs, while the full predictor of about 89%.

VI. DISCUSSION

As concerns the classification, the very good results we obtained (more than 90% for all the considered metrics) demonstrate, first of all, the successful applicability of the tree-based models to the considered dataset and feature set, even if it is slightly unbalanced in favor of the "Healthy" class. It also appears that there is no clear difference in using Gini impurity or Information Gain in the Decision Tree models, and that Random Forest, although being an ensemble of Decision Trees, only achieve a very little improvement in comparison with simple Decision Tree models. These models, moreover, also guarantee interpretability, thus they should be preferred even if they provide a little worse performance. In terms of explainability, R_{ds} appears to be the first feature used by the Decision Trees in dividing the dataset, i.e., R_{ds} is the root feature, and this is consistent with the ability of the bioristor in revealing changes in the concentration of the ions, already identified as the main players in the drought stress response [19].

Looking at the confusion matrices, the "Healthy" class always achieves very good performance, well beyond 90% of TPR (True Positive Rate), while the "Recovery" class is always the worst (less than 90% in terms of TPR even with Random Forest) when considering the correct identification. However, this is not a very critical issue, given that the recovery phase takes place when the plant has been

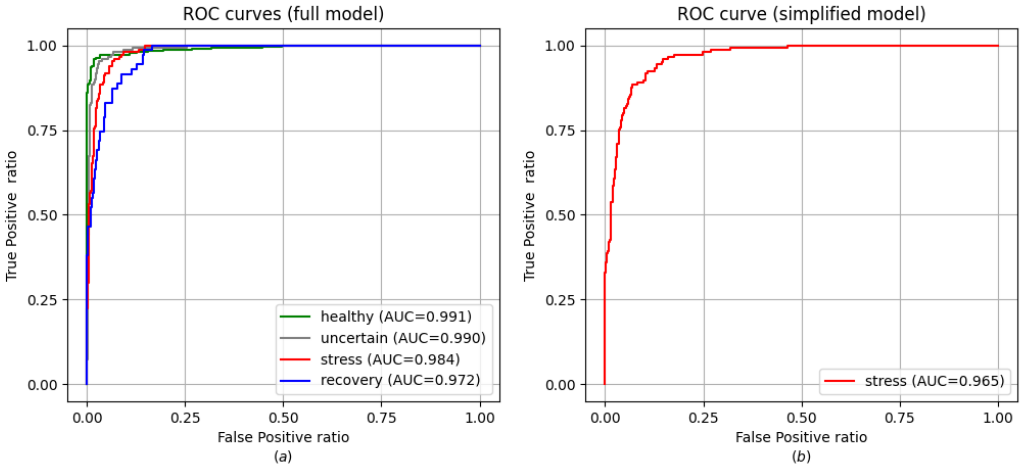

FIGURE 11. Receiving operating characteristic curves for the full model (4 statuses) (*a*) and the simplified model (2 statuses) (*b*).

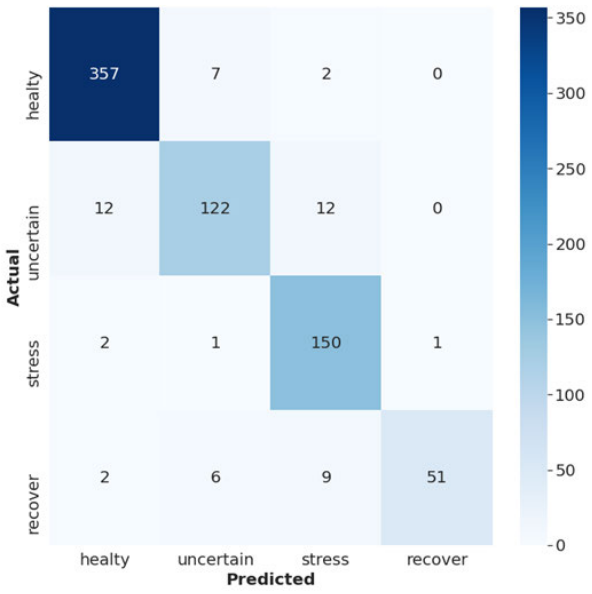

FIGURE 12. Confusion matrix for the 24h forecasts in the full model (4 statuses).

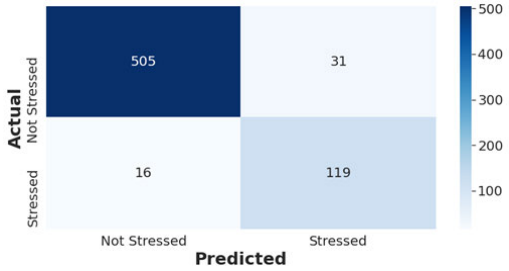

FIGURE 13. Confusion matrix for the 24h forecasts in the simplified model (2 statuses).

re-irrigated for sure (through rain or artificial irrigation) and so its perfect identification can be easily obtained in other ways.

On the other hand, the time series predictive analysis has demonstrated very high accuracy both in predicting, 24 hours in advance, the four different statuses (with a 97.40% of correct prediction of the stress status) as well as in the early identification of the stress/non-stress status (with a 88.15% of correct prediction of the stress status) with a simplified model. In addition, the precision for predicting the stress status, defined as the ratio between the number of correctly predicted stress statuses over the number of predicted stress statuses, for the two models is 86.7% and 79.3%, respectively. This will be very useful in order to know in advance the possible manifestation of a hydric stress in the plants and thus enacting proper countermeasures, e.g., the activation of a proper smart irrigation system, to keep the plants healthy and fruit-bearing as long as possible within their natural life cycle.

Finally, we remark that the considered models take into account only measurements collected from the bioristor during the previous six hours, at times t, $t-3$, and $t-6$. In particular, the models could be augmented by regarding external actions, such as irrigation times, that were not available in this study, and that can be used to determine better the plant recovery times.

VII. CONCLUSION

In this paper, we applied for the first time machine and deep learning models and techniques to the features and data coming from a bioristor positioned into tomato plants to monitor their water stress status. In particular, we tried to classify the water stress status in four classes, by using Decision Trees and Random Forest models, achieving very high results in terms of accuracy, precision, recall, and F-measure. Moreover, we have tried to forecast the water stress status after 24 hours from the current status of the plant, by considering both a binary and a four-status model and using recurrent neural networks, obtaining very good diagonal confusion matrices in both cases. The very good results demonstrate the feasibility of the application of machine and deep learning models to bioristor data to analyze and predict the water stress status of plants both from a static and a dynamic point of view.

Future developments will target the application of the used models on a real field scenario, and the identification of

classification and prediction models that integrate environmental measurements, such as, temperature, humidity, and irrigation quantities and times, in order to improve further the achieved performance. Another promising direction for future research is the identification of time series models for forecasting the behavior of R_{ds}, ΔI_{gs}, t_{ds}, and t_{gs} based on their historical data. These models will be integrated into classification systems to better explain the behavior of plants under various stimuli and environmental conditions and to better forecast their irrigation needs.

ACKNOWLEDGMENT

Riccardo Pecori is a member of the INdAM GNCS Research Group. The authors would like to thank Fondazione Cariparma (project Biomontans), RGVFAO VI DM 10271, and ALSIA Metaponto Agrobios for hosting some of the experiments and for funding a PhD doctoral fellowship. They also would like to thank the anonymous reviewers for their valuable comments and suggestions, which helped in improving the quality of the manuscript. This work was supported in part by the E-Crops—Technologies for Digital and Sustainable Agriculture Project funded by the Italian Ministry of University and Research (MUR) through the Programma Operativo Nazionale (PON) Agrifood Program under Contract ARS01_01136.

REFERENCES

[1] S. Jeong, S. Jeong, and J. Bong, "Detection of tomato leaf miner using deep neural network," *Sensors*, vol. 22, no. 24, p. 9959, Dec. 2022.

[2] M.-S. Gang, H.-J. Kim, and D.-W. Kim, "Estimation of greenhouse lettuce growth indices based on a two-stage CNN using RGB-D images," *Sensors*, vol. 22, no. 15, p. 5499, Jul. 2022.

[3] Q. Hao, X. Guo, and F. Yang, "Fast recognition method for multiple apple targets in complex occlusion environment based on improved YOLOv5," *J. Sensors*, vol. 2023, pp. 1–13, Feb. 2023.

[4] J. S. H. Al-bayati and B. B. Ustundag, "Artificial intelligence in smart agriculture: Modified evolutionary optimization approach for plant disease identification," in *Proc. 4th Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2020, pp. 1–6.

[5] A. Sharma, M. Georgi, M. Tregubenko, A. Tselykh, and A. Tselykh, "Enabling smart agriculture by implementing artificial intelligence and embedded sensing," *Comput. Ind. Eng.*, vol. 165, Mar. 2022, Art. no. 107936.

[6] C. Arif, M. Mizoguchi, B. I. Setiawan, and R. Doi, "Estimation of soil moisture in paddy field using artificial neural networks," 2013, *arXiv:1303.1868*.

[7] N. K. Nawandar and V. R. Satpute, "IoT based low cost and intelligent module for smart irrigation system," *Comput. Electron. Agricult.*, vol. 162, pp. 979–990, Jul. 2019.

[8] A. Goap, D. Sharma, A. K. Shukla, and C. R. Krishna, "An IoT based smart irrigation management system using Machine learning and open source technologies," *Comput. Electron. Agricult.*, vol. 155, pp. 41–49, Dec. 2018.

[9] M. Romero, Y. Luo, B. Su, and S. Fuentes, "Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management," *Comput. Electron. Agricult.*, vol. 147, pp. 109–117, Apr. 2018.

[10] R. Revathy and S. Balamurali, "Developing an efficient irrigation scheduling system using hybrid machine learning algorithm to enhance the sugarcane crop productivity," Res. Square, 2022, doi: 10.21203/rs.3.rs-1504824/v1.

[11] M. Nagappan, V. Gopalakrishnan, and M. Alagappan, "Prediction of reference evapotranspiration for irrigation scheduling using machine learning," *Hydrol. Sci. J.*, vol. 65, no. 16, pp. 2669–2677, Dec. 2020.

[12] A. A. Farooque, H. Afzaal, F. Abbas, M. Bos, J. Maqsood, X. Wang, and N. Hussain, "Forecasting daily evapotranspiration using artificial neural networks for sustainable irrigation scheduling," *Irrigation Sci.*, vol. 40, no. 1, pp. 55–69, Jan. 2022.

[13] S. S. Bashir, A. Hussain, S. J. Hussain, O. A. Wani, S. Zahid Nabi, N. A. Dar, F. S. Baloch, and S. Mansoor, "Plant drought stress tolerance: Understanding its physiological, biochemical and molecular mechanisms," *Biotechnol. Biotechnol. Equip.*, vol. 35, no. 1, pp. 1912–1925, Jan. 2021.

[14] V. Buffagni, F. Vurro, M. Janni, M. Gullì, A. A. Keller, and N. Marmiroli, "Shaping durum wheat for the future: Gene expression analyses and metabolites profiling support the contribution of BCAT genes to drought stress response," *Frontiers Plant Sci.*, vol. 11, p. 891, Jul. 2020.

[15] A. Toreti, D. Bavera, J. Acosta Navarro, C. Cammalleri, A. de Jager, C. Di Ciollo, A. Hrast Essenfelder, W. Maetens, D. Magni, D. Masante, M. Mazzeschi, S. Niemeyer, and J. Spinoni, "Drought in Europe: August 2022," Publications Office Eur. Union, Luxembourg, Tech. Rep. EUR 31192 EN, 2022.

[16] A. Gorlapalli, S. Kallakuri, P. D. Sreekanth, R. Patil, N. Bandumula, G. Ondrasek, M. Admala, C. Gireesh, M. S. Anantha, B. Parmar, B. K. Yadav, R. M. Sundaram, and S. Rathod, "Characterization and prediction of water stress using time series and artificial intelligence models," *Sustainability*, vol. 14, no. 11, p. 6690, May 2022.

[17] A. K. Rico-Chávez, J. A. Franco, A. A. Fernandez-Jaramillo, L. M. Contreras-Medina, R. G. Guevara-González, and Q. Hernandez-Escobedo, "Machine learning for plant stress modeling: A perspective towards hormesis management," *Plants*, vol. 11, no. 7, p. 970, Apr. 2022.

[18] A. Finco, D. Bentivoglio, G. Chiaraluce, M. Alberi, E. Chiarelli, A. Maino, F. Mantovani, M. Montuschi, K. G. C. Raptis, F. Semenza, V. Strati, F. Vurro, E. Marchetti, M. Bettelli, M. Janni, E. Anceschi, C. Sportolaro, and G. Bucci, "Combining precision viticulture technologies and economic indices to sustainable water use management," *Water*, vol. 14, no. 9, p. 1493, May 2022.

[19] M. Janni, N. Coppede, M. Bettelli, N. Briglia, A. Petrozza, S. Summerer, F. Vurro, D. Danzi, F. Cellini, N. Marmiroli, D. Pignone, S. Iannotta, and A. Zappettini, "In vivo phenotyping for the early detection of drought stress in tomato," *Plant Phenomics*, vol. 2019, pp. 1–10, Jan. 2019, Art. no. 6168209.

[20] J. Michela, C. Claudia, B. Federico, P. Sara, V. Filippo, C. Nicola, B. Manuele, C. Davide, F. Loreto, and A. Zappettini, "Real-time monitoring of arundo donax response to saline stress through the application of in vivo sensing technology," *Sci. Rep.*, vol. 11, no. 1, p. 18598, Sep. 2021.

[21] F. Vurro, M. Janni, N. Coppedè, F. Gentile, R. Manfredi, M. Bettelli, and A. Zappettini, "Development of an in vivo sensor to monitor the effects of vapour pressure deficit (VPD) changes to improve water productivity in agriculture," *Sensors*, vol. 19, no. 21, p. 4667, Oct. 2019.

[22] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artif. Intell. Agricult.*, vol. 2, pp. 1–12, Jun. 2019.

[23] E. A. Abioye, O. Hensel, T. J. Esau, O. Elijah, M. S. Z. Abidin, A. S. Ayobami, O. Yerima, and A. Nasirahmadi, "Precision irrigation management using machine learning and digital farming solutions," *AgriEngineering*, vol. 4, no. 1, pp. 70–103, Feb. 2022.

[24] Y. Ahansal, M. Bouziani, R. Yaagoubi, I. Sebari, K. Sebari, and L. Kenny, "Towards smart irrigation: A literature review on the use of geospatial technologies and machine learning in the management of water resources in arboriculture," *Agronomy*, vol. 12, no. 2, p. 297, Jan. 2022.

[25] M. K. Saggi and S. Jain, "A survey towards decision support system on smart irrigation scheduling using machine learning approaches," *Arch. Comput. Methods Eng.*, vol. 29, no. 6, pp. 4455–4478, Oct. 2022.

[26] N. Coppedè, M. Janni, M. Bettelli, C. L. Maida, F. Gentile, M. Villani, R. Ruotolo, S. Iannotta, N. Marmiroli, M. Marmiroli, and A. Zappettini, "An in vivo biosensing, biomimetic electrochemical transistor with applications in plant science and precision farming," *Sci. Rep.*, vol. 7, no. 1, p. 16195, Nov. 2017.

[27] M. A. M. Almuhaya, W. A. Jabbar, N. Sulaiman, and S. Abdulmalek, "A survey on LoRaWAN technology: Recent trends, opportunities, simulation tools and future directions," *Electronics*, vol. 11, no. 1, p. 164, Jan. 2022.

[28] P. Carella, D. C. Wilson, C. J. Kempthorne, and R. K. Cameron, "Vascular sap proteomics: Providing insight into long-distance signaling during stress," *Frontiers Plant Sci.*, vol. 7, p. 651, May 2016.

[29] F. Gentile, F. Vurro, M. Janni, R. Manfredi, F. Cellini, A. Petrozza, A. Zappettini, and N. Coppedè, "A biomimetic, biocompatible OECT sensor for the real-time measurement of concentration and saturation of ions in plant sap," *Adv. Electron. Mater.*, vol. 8, no. 10, 2022, Art. no. 2200092.
[30] G. Tarabella, M. Villani, D. Calestani, R. Mosca, S. Iannotta, A. Zappettini, and N. Coppedè, "A single cotton fiber organic electrochemical transistor for liquid electrolyte saline sensing," *J. Mater. Chem.*, vol. 22, no. 45, pp. 23830–23834, 2012.
[31] F. Gentile, F. Vurro, F. Picelli, M. Bettelli, A. Zappettini, and N. Coppedè, "A mathematical model of OECTs with variable internal geometry," *Sens. Actuators A, Phys.*, vol. 304, Apr. 2020, Art. no. 111894.
[32] N. Coppedè, M. Villani, and F. Gentile, "Diffusion driven selectivity in organic electrochemical transistors," *Sci. Rep.*, vol. 4, no. 1, p. 4297, Mar. 2014.
[33] L. Breiman, H. Jerome Friedman, A. Richard Olshen, and J. Charles Stone, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, Oct. 2017.
[34] G. Thomas Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*. Berlin, Germany: Springer, 2000, pp. 1–15.
[35] M. C. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
[36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org
[39] M. A. I. H. Hall Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

MANUELE BETTELLI received the Ph.D. degree in material science and technology from the University of Parma, in March 2018. He is currently a Postdoctoral Researcher with IMEM-CNR (PR), where he works on two main topics, such as I) X-ray and gamma-ray room-temperature detectors based on CZT, particularly, transport and spectra simulations, detector realization, and characterization; II) organic electrochemical transistors for the in-vivo monitoring of plants, the sensors are based on PEDOT:PSS. In particular, he deals with electronic readout systems and data analysis. He had designed and programmed embedded systems. He is an expert in instrumentation–PC interfacing. He has been and is scientifically responsible for IMEM-CNR of several activities and projects. He is currently involved in the foundation of the "BIORISTOR" spin-off (of CNR) which is the finalist of the national PNI 2021 competition. Since 2014, he has been coauthoring 41 papers in international journals, sent 33 contributions to international conferences (including an invited oral presentation), and participated in one international summer school.

FILIPPO VURRO received the bachelor's degree in biology and the master's degree in chemistry from the University of Parma, and the Ph.D. degree in material science and technology from IMEM-CNR. He is a Postdoctoral Researcher with IMEM-CNR. He is the author of 14 scientific publications. He has worked on metabolomics through NMR spectroscopy both in the biomedical and agronomic fields, as well as on the synthesis of ceramic nanoparticles for human tissue regeneration. His research is mainly focused on the development of in-vivo OECT-based electrochemical biosensors to monitor plant physiology in real time. In May 2019, in the framework of the Fascination of Plants Day 2019 in Matera, he won the first prize in the Smart Plants Contest. In December 2019, during the Nanoday19 in Milano, he received recognition as the Best Presentation in the life science session.

RICCARDO PECORI received the Ph.D. degree in information technologies from the University of Parma, Parma, Italy, in 2011. He was an Assistant Professor of computer engineering with eCampus University, Novedrate, Italy, from 2015 to 2019; and the University of Sannio, Benevento, Italy, from August 2019 to January 2022. He was a Senior Assistant Professor with the Catholic University of the Sacred Heart, Brescia, Italy, from February 2022 to April 2022. He has been an Associate Professor of computer engineering with Universitas Mercatorum, since May 2022, and eCampus University, since October 2022. Since June 2022, he has been a Researcher of computer science with IMEM-CNR. He has authored or coauthored more than 50 papers in refereed international journals and conferences. His current research interests include network and cybersecurity, the Internet of Things, machine- and deep-learning applications, analysis of complex systems, and educational big data mining. He was a leading organizer and the program chair of computer science-related special sessions and workshops at international conferences. He is currently on the editorial board of *SoftwareX*.

MICHELA JANNI received the degree in physics from the University of Tuscia, in 1995, and the Ph.D. degree in plant biotechnology, in 2006, focusing her activities on plant physiology and genetics. She is currently a permanent Researcher with IMEM-CNR. She is the author of 42 peer-reviewed articles. Her H-index is 17. Her main research interests include the application of in-vivo sensors in plant phenotyping and precision agriculture.

NICOLA COPPEDÈ received the degree in physics from the University of Pisa, in 2001, and the Ph.D. degree in physics from the University of Trento, in 2006. He is a Researcher with IMEM-CNR and the Head of the SIGNAL Research Group. He is the author of over 90 scientific publications in international journals (source: Web of Science), three international patents, and four Italian patents. His H-index is 21. His main research interests include the development of integrated sensors, from organic electrochemical transistors on textiles for wearable applications, to biocompatible fiber biosensors for plant monitoring, to soft pressure sensors in polymers and composites.

ANDREA ZAPPETTINI received the degree in physics and the master's degree in material science from the University of Parma. After working for Eni and Pirelli, he became a Researcher with the Institute of Materials for Electronics and Magnetism (IMEM), an Institute of the Italian National Research Council (CNR). Since 2019, he has been the Director of IMEM. He has published over 200 articles in international journals (H-index is 35). His research activity has been mainly devoted to the development of sensors based on novel and multifunctional materials.

DANIELE TESSERA received the Ph.D. degree in computer engineering from the University of Pavia, Italy. He is currently an Associate Professor of computer science with the Department of Mathematics and Physics, Catholic University of the Sacred Heart, Italy. He is the coauthor of more than 40 papers in international journals and conference proceedings. His research interests include performance debugging and benchmarking, workload characterization, cloud computing, and artificial intelligence applications.

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE-ITALY' within the CRUI CARE Agreement