Modeling Earth Systems and Environment (2025) 11:173
https://doi.org/10.1007/s40808-025-02343-w

ORIGINAL ARTICLE

Assessing salinity-induced impacts on plant transpiration through machine learning: from model development to deployment

Niguss Solomon Hailegnaw[4] · **Girma Worku Awoke**[1] · **Aline de Camargo Santos**[2] · **Bruce Schaffer**[1,5] · **Ana I. Vargas**[5] · **Edivan Rodrigues de Souza**[3] · **Haimanote K. Bayabil**[1]

Received: 6 January 2025 / Accepted: 21 February 2025 / Published online: 13 March 2025

Abstract

Soil salinity poses a global threat to crop production. Early understanding of plant physiological responses to salinity stress can be critical to implementing timely stress management strategies. One of the initial plant physiological responses to salinity is a reduction of transpiration. This study used papaya as a model crop to better understand the effect of salinity on whole-plant transpiration using a greenhouse experiment. Treatments consisted of four electrical conductivity (EC_IR) levels of irrigation water: ~ 0 (tap water), 2, 4, or 8 dS m^{-1} were considered for the experiment. An automated phenotyping platform measured whole plant transpiration from the time of papaya transplanting until they reached approximately 15 weeks of age. Five machine learning models: extreme gradient boosting (XGBt), categorical boosting (CATBt), light gradient boosting (LAGBt), random forest (RF), and decision tree (DT) were fitted to the transpiration data and machine learning algorithms were deployed on a new data set. The impact of salinity on transpiration started to become evident 16 days after initiation of salinity treatments, where only the 8 dS m^{-1} treatment induced a significant decline in transpiration. All machine learning models efficiently captured salinity-induced impacts on transpiration. The use of salinity as an input feature improved the performance of all machine learning models. Salinity contributed up to 32% to the predictive capability of the machine learning models by improving the R^2, root mean squared error (RMSE), and mean absolute error (MAE) of the machine learning models up to 19, 25, and 25%, respectively. Overall, the extreme gradient boosting model outperformed all the machine learning models. Furthermore, deployment of machine learning algorithms to a new data set effectively indicated a critical level of EC_IR of 6 dS m^{-1} on whole-plant transpiration above which transpiration significantly drops. Apart from transpiration the effect of salinity on biomass, the concentration of Na^+ and Cl^- contents in leaves and roots were evident. Overall, machine learning models can be useful tools for capturing salinity-induced impact on plant water use and its integration with crop stress management practices could be valuable.

Keywords Salinity stress · Automated phenotyping platform · Extreme gradient boosting · Categorical boosting · Light gradient boosting, decision tree, and random forest

---

✉ Niguss Solomon Hailegnaw
nhailegnaw@ufl.edu

Girma Worku Awoke
girma.awoke@ufl.edu

Aline de Camargo Santos
alinedcsantoss@gmail.com

Bruce Schaffer
bas56@ufl.edu

Ana I. Vargas
avargas22@ufl.edu

Edivan Rodrigues de Souza
edivan.rodrigues@ufrpe.br

Haimanote K. Bayabil
hbayabil@ufl.edu

[1] Department of Agricultural and Biological Engineering, Tropical Research and Education Center, IFAS, University of Florida, Homestead, FL 33031, USA

[2] School of Environment and Natural Resources, The Ohio State University, Columbus, OH 43204, USA

[3] Department of Agronomy, Federal Rural University of Pernambuco, Recife, Brazil

[4] Department of Agronomy, Everglades Research and Education Center, IFAS, University of Florida, Belle Glade, FL 33430, United States

[5] Horticultural Sciences Department, Tropical Research and Education Center, IFAS, University of Florida, Homestead, FL 33031, United States

Introduction

Soil salinity is among the main abiotic factors affecting agriculture worldwide causing major crop losses and significantly reducing agricultural productivity (Majeed and Muhammad 2019). Currently, 20% of the total cultivated land and 33% of irrigated agricultural lands are affected by high levels of soil salinity. Moreover, there is an estimated 10% annual increase in salinized agricultural areas worldwide (Chele et al. 2021; Shrivastava and Kumar 2015). Several reasons are contributing to such a high annual rate of increasing soil salinity such as rising sea levels and saltwater intrusion, overuse of chemical fertilizers, low precipitation, high evaporation, and irrigation with saline water (Chele et al. 2021; Nicolas et al. 2023; de Souza et al. 2024). Soils are classified as saline when the electrical conductivity (EC) of saturated extract from the root zone exceeds 4 dS m^{-1} and sodic when there is an exchangeable sodium percentage of 15% (Shrivastava and Kumar 2015). Evaporation of water from saline soils causes the precipitation of salts, predominantly Na^+ and Cl^- ions, and nutrient imbalances in the root zone, leading to salt uptake by the roots and consequent stunting of plant development at all stages (Chele et al. 2021). Multiple studies highlight the severe and multifaceted impacts of soil salinity on agricultural productivity. For instance, Nicolas et al. (2023) evaluated the effects six irrigation water salinity levels (0.5 to 5.5 dS m^{-1}) on the yield of alfalfa, almonds, grapes, and tomatoes in sandy loam soils of North America. Their results indicated the negative impact of salinity levels of 5.5 dS m^{-1}, decreasing the yield of alfalfa, almonds, grapes, and tomatoes by up to 10%, 45%, 18%, and 12%, respectively (Nicolas et al. 2023). Similarly, Stavridou et al. (2017) investigated the impacts of soil salinity on stomatal conductance and dry matter yield of the perennial grass. Their findings demonstrated a significant decline in stomatal conductance and dry matter yield by up to 56% for soil salinity levels over 5 dS m^{-1}. Likewise, de Souza et al. (2024) tested the influence of six soil salinity levels on plant growth, and physiological and nutritional responses of 'Red Lady' papaya plants in an inert calcined clay (Turface®) potting medium. Their results indicated a decline in stomatal conductance and transpiration by up to 63 and 52% respectively at soil salinity levels of 6 dS m^{-1}.

A decline in plant transpiration, is one of the initial physiological responses to salinity stress, which can be easily captured at the whole-plant level using tools such as high-throughput physiological phenotyping platforms that generate dynamic and intensive datasets of plant transpiration. Such intensive datasets are ideal for machine learning models that employ various learning algorithms to understand the complex and nonlinear relationship between the input and target variables, which can be a viable alternative to model salinity effects on crop physiology. For example, Fan et al. (2021) tested the applicability of four machine-learning models, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBt), Artificial Neural Network (ANN), and Deep Neural Network (DNN) models in simulating daily transpiration of maize. Their result suggested that the addition of soil water content and leaf area index significantly improved the performance of machine learning models for simulating daily transpiration. They highlighted the effectiveness of DNN models in simulating daily maize transpiration due to their competitive advantage in modeling the complex relationship between transpiration and its driving factors (Fan et al. 2021). Furthermore, several studies successfully simulated plant evapotranspiration using machine learning models (Du et al. 2024; Fan et al. 2021; Guo et al. 2024; Hailegnaw et al. 2024; He et al. 2024; Lee et al. 2024; Li et al. 2020; Wang et al. 2024).

Understanding salinity-induced plant water use and transpiration responses using continuously measured data from systems like the Plantarray (Plant-DiTech Ltd., Yavne, Israel; PDT) – a high-throughput, multi-sensor physiological gravimetric phenotyping platform – offers significant benefits. Integrating this data with machine learning analytics could give a better understanding and prediction of plant transpiration changes to gauge salinity stress. The primary goal of this study was to leverage the capabilities of the Plantarray system to gather large datasets, enabling a more in-depth analysis of plant responses to salinity. By employing machine learning algorithms, we aimed to gain a better understanding of the trends and patterns in the data, ultimately leading to improved insights into how plants respond and adapt to increased soil salinity.

Materials and methods

Experimental setup

A study was conducted in a greenhouse at the University of Florida, Tropical Research and Education Center in Homestead, Florida, USA (25.5°N longitude and 80.5°W latitude) from September to November 2022 to assess the impacts of salinity on whole-plant transpiration through machine learning. Papaya plants (*Carica papaya* L.) cultivar Red Lady grown in Turface® were used as a model system because a recent study (de Souza et al. 2024) determined the amount of salinity that negatively affected transpiration and caused salinity stress of this species in this potting medium. Papaya seeds were germinated in flats in Promix® potting medium (Premier Tech, Quebec, Canada) and ~6 weeks after germination seedlings were then transplanted to pots that were part of a high-throughput physiological phenotyping platform (Plantarray, Plant-DiTech Ltd., Yavne, Israel; PDT) for constant monitoring of plant water use and transpiration and to precisely control irrigation scheduling. After a week of transplanting, salinity stress treatments were started at 4 different irrigation water electrical conductivity levels (0, 2, 4, and 8 dS m^{-1}). In this manuscript the word salinity has been used interchangeably with electrical conductivity of irrigation water (EC_IR). A description of the PDT phenotyping platform is provided by Halperin et al. (2017) and the experimental setup is described in detail for a similar study by de Camargo Santos et al. (2024). The PDT system is composed of 3.9-L pots, each placed on top of highly accurate (± 10 g) and temperature-compensating lysimeter (Tadea-Huntleigh, model 1042 C4; Vishay Intertechnology, Malvern, PA, USA). Each pot had a sensor (model 5 TE, Meter Group, Pullman, WA, USA) for measuring soil water content and electrical conductivity. Each pot was irrigated with a drip irrigation system connected to a controller that controls the irrigation schedules for each plant individually. Data were downloaded every 3 min to a datalogger (model CR 1000, Campbell Scientific, Logan, UT, USA) and sent to a server in real-time (SPAC Analytics - Plant-Ditech Ltd., Yavne, Israel). The system was installed in a climate-controlled greenhouse with a weather station (WatchDog 2800 Weather Station, Spectrum Technologies, Inc., Aurora, IL, USA) that continuously monitored and recorded daily variations in photosynthetically active radiation (PAR) and vapor pressure deficit (VPD). Temperature in the greenhouse ranged from 24 to 32 °C and relative humidity ranged from 60 to 80% during the experiment.

A 1:1 (v/v) mixture of Turface MVP® and Turface Profile Greens Grade® (Profile Products LLC, Buffalo Grove, IL, USA), an inert calcined clay oven-dried for 48 h, was used as the growing medium. The gravimetric water content of the saturated substrate was 0.77 g g^{-1} or 77%, and 0.60 g g^{-1} or 60% after drainage (field capacity). A layer of 300 cm^3 of 6 mm plastic beads was placed on top of the substrate to prevent any water loss by evaporation, thus enabling accurate water mass balance calculations and estimates of whole-plant transpiration. The irrigation regime was programmed to occur during the night, between 23:00 and 02:00 h, in 2–3 sequential pulses scheduled every 30 min to avoid the confounding effects of water use during the day. Plants were irrigated with a fertigation solution consisting of a Hoagland's solution (Hoagland and Arnon 1950).

Experimental design

Papaya seedlings were divided into four treatments consisting of four EC_IR levels corresponding to ~0 (tap water with not salt added), 2, 4, or 8 dS m^{-1} achieved by adding different amounts of artificial sea salt (Instant Ocean®, Aquarium Systems, Blacksburg, VA, USA) (https://www.instantocean.com/products/sea-salt-mixes/sea-salt-mixture.aspx) to the irrigation water to obtain the desired EC_IR. There were eight single plant replicates per treatment arranged in a randomized complete block design.

Physiological assessment and biomass measurements

Whole-plant continuous physiological measurements were obtained from the PDT system. Daily whole-plant transpiration was determined for each plant by calculating the difference between the system's start and end-of-the-day reference points obtained for each individual lysimeter. These reference points were obtained by averaging the lysimeter's readings over a 30-min period, between 05:00 and 05:30 h for the start-of-the-day, and between 21:00 and 21:30 h for the end-of-the-day. Whole-plant transpiration (plant-Tr) and root water uptake were calculated from the lysimeter and soil sensor readings (according to the substrate volume) temporal series as described by Halperin et al. (2017). In short, those variables were determined by multiplying the derivative of the initial recorded measurement time series by −1. The average of measurements recorded from 11:00–13:00 h were used to calculate the midday transpiration rate and root water uptake. Then the transpiration efficiency ($g_{(dry\ weight)}$ L^{-1}) was calculated as the ratio of papaya biomass (g) and cumulative transportation (L) (Kemanian et al. 2005).

Plants were harvested 50 days after stress imposition (DAS) and tissues were separated into shoot and root portions, dried in an oven at 60 °C until they reached a constant mass. Dried tissue samples were ground to a fine powder in an electrical blender. Leaf and root Cl contents were determined by extraction in water and titration with silver nitrate (Malavolta et al. 1997). Leaf and root Na^+ concentrations were determined at the University of Florida, USA, Analytical Research Laboratory in Gainesville, Florida, where Na^+ content was determined by inductively coupled plasma emission spectrometry (Hanlon et al. 1994).

Machine learning models

The machine-learning models utilized in this study included five different algorithms: eXtreme Gradient Boosting Trees (XGBt), Decision Trees (DT), Random Forest (RF), Categorical Boosting Trees (CATBt), and Light Gradient Boosting Machine (LAGBt). Python version v3.14 wasused to develop all machine learning models. A total of 22,363 data points collected from plant array system were used for all machine learning by grouping 70% to training and 30% to the test set. Therefore, all models have been trained with

70% of the observed data and tested with the remaining 30% of the observed data. The RandomizedSearchCV hyperparameter optimization method was used to randomly search over a specified grid of hyperparameters and find the best combination for the model. It efficiently explores a wide range of hyperparameters without the exhaustive computation required by GridSearchCV technique. The evaluation criteria used for the models are R^2: Coefficient of determination, RMSE: Root Mean Squared Error and MAE: Mean Absolute Error.

Extreme gradient boosting (XGBt) is a powerful machine-learning model known for its scalability, flexibility, and efficiency. It employs advanced techniques such as row and column sampling to address common overfitting issues encountered in machine learning. XGBt takes advantage of first and second-order statistics to effectively optimize the loss function (Zhang et al. 2021). A categorical boosting (CATBt) is a supervised machine learning in the family of gradient-boosting decision tree models that is capable of handling both classification and regression problems (Hancock and Khoshgoftaar 2020). Light gradient boosting (LAGBt) combines the idea of boosting with a decision tree type of modeling. It is ideally different than the extreme gradient boosting in its way of handling the algorithm training to speed up the process (Ke et al. 2017). Random forest (RF) is a type of supervised machine learning that can learn trends in data sets and therefore able to estimate trends based on learning. It has good estimation accuracy as it integrates multiple decision trees. The algorithm constructs a tree by randomly selecting a subset of variables and then the estimation of trends will be generated by aggregating the prediction of all trees (Breiman 2001). The decision tree (DT) model is known to handle both categorical and regression types of problems. The algorithm is known for its tree-like structure allowing it to take advantage of obtaining the best weight for training and improving the predicting capability of the algorithm (Tahraoui et al. 2022).

Evaluation of developed machine learning algorithms

Simulated experiments with 9 treatments (0, 1, 2, 3, 4, 5, 6, 7, and 8 dS m^{-1}) were developed using a 15-minute interval of weather data comprising 36,801 observations from a weather station located in the same greenhouse, salinity levels randomly assigned in 3 replications to the data. These data included parameters used for the development of machine learning algorithms, such VPD and PAR. Afterward, the developed machine learning models were imported directly into Jupiter notebook and fitted with the newly prepared data set to predict whole-plant transpiration (Fig. 1).

Statistical analysis

Tukey's Honest Significant Difference (HSD) test was applied to identify pairwise significant differences between different levels of electrical conductivity. A p-value of <0.05 was considered statistically significant unless otherwise stated. The statistical analysis and figures were generated using Python 3.

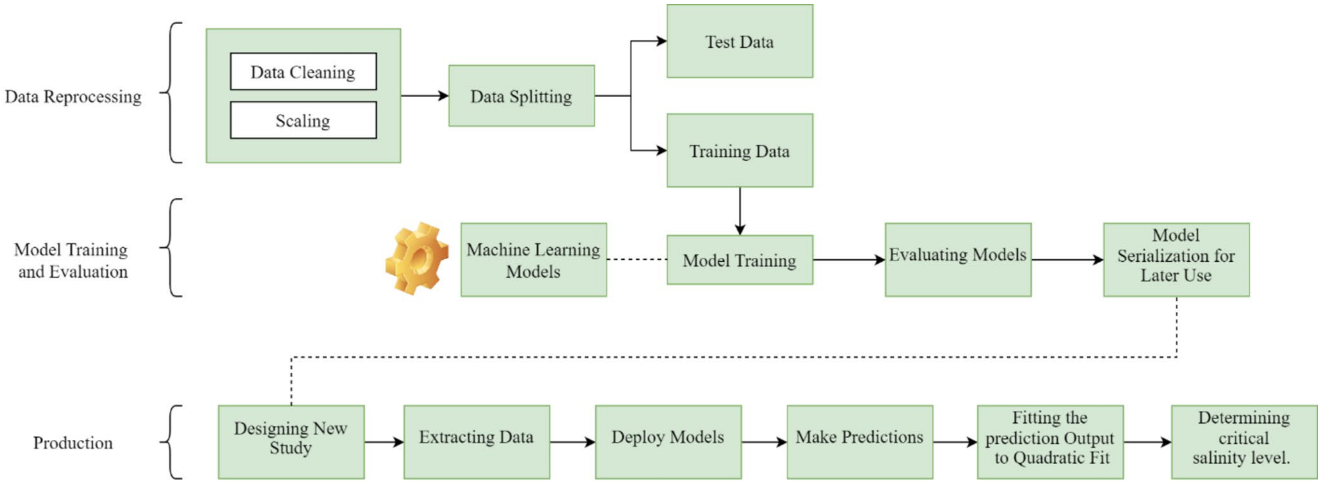


Fig. 1 Schematic diagram depicting the machine learning process including data preprocessing, training machine learning models, and production using algorithms

Results and discussion

Performance of machine learning models in capturing salinity-induced changes of papaya transpiration

The Plant-Ditech system allows for the continuous measurement of plant transpiration at a 3-minute interval, providing highly detailed plant transpiration data. The effect of salinity on transpiration began 16 days after salinity treatments were initiated (Fig. 2A). The EC_IR of 8 dS m^{-1} significantly decreased papaya cumulative transpiration by 53% compared to the control, whereas there was a 14% and 2% decline in cumulative transpiration in the 4 and 2 dS m^{-1} treatments, respectively (Fig. 2B). Cumulative transpiration was negatively correlated with Na^+ and Cl^- contents in leaves (r = -0.78 and −0.71) and roots (r = -0.82 and −0.66). Salinity stress can reduce transpiration by interfering with root water uptake and due to the toxicity of salts (de Souza et al. 2024; Lu and Fricke 2023; Munns and Tester 2008). A higher accumulation of salts in plant tissue triggers a reduction in transpiration in an attempt to conserve water and minimize further salt uptake (Peçanha et al. 2017).

Electrical conductivity of irrigation water negatively correlated with transpiration (r=−0.26), confirming the negative impact of salinity on transpiration (Fig. 3). As expected, the day after planting, VPD and PAR were positively correlated with transpiration.

To evaluate the ability of machine learning models in capturing salinity-induced impacts on transpiration, two models were developed with and without salinity as an input feature. Four factors were selected for the first set of models: EC_IR, solar radiation, days after planting, and VPD (XGBt_1, CATBt_1, LAGBt_1, RF_1, and DT_1). In contrast, EC_IR is excluded for XGBt_2, CATBt_2, LAGBt_2, RF_2, and DT_2. As evidenced by all three statistical evaluations (R^2, RMSE, and MAE), the inclusion of EC_IR as an input feature improved the efficiency of all the machine learning models (Fig. 3; Table 1).

As shown in Fig. 4, the density plot indicates that all machine learning models exhibit a slightly higher distribution in the transpiration range of 0.15 to 0.25 g water per plant per minute compared to the measured values and a slightly off distribution for transpiration above 0.4 g water per plant per minute. Additionally, within this range, models that do not account for salinity displayed a marginally higher distribution than their respective salinity-included counterparts. The inclusion of salinity in the models improved the performance of all machine learning models by increasing R^2 up to 25% and decreasing RMSE and MAE up to 21 and 23%, respectively. The substantial improvements in R^2, RMSE, and MAE indicate that salinity was indeed an important parameter in estimating plant transpiration. This was further confirmed by the feature importance analysis result of all models (Fig. 5). Salinity contributed to 29, 21, 24, 28, and 32% to the predictive power of XGBt, CATBt, LAGBt, RF, and DT, respectively.

The distribution of transpiration, from measured data or data estimated by machine learning models (Fig. 6), further indicated a better match between observed and predicted transpiration when salinity was considered (XGBt_1, CATBt_1, LAGBt_1, RF_1, and DT_1) compared to when it was not (XGBt_2, CATBt_2, LAGBt_2, RF_2, and DT_2). To our knowledge, there is no published scientific paper considering soil salinity in modeling plant transpiration while several papers have highlighted the efficiency of machine learning models in estimating plant evapotranspiration under optimal conditions (Du et al. 2024; Fan et

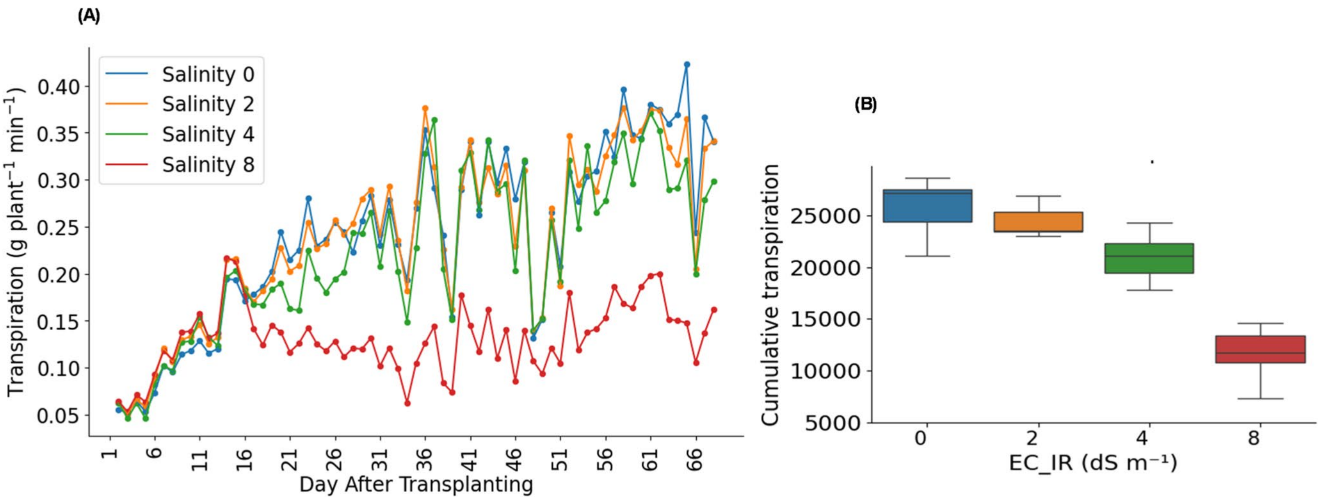


Fig. 2 (**A**) Papaya transpiration as affected by variation in electrical conductivity of irrigation water (EC_IR) of 0, 2, 4, or 8 dS m^{-1}), (**B**) The effect of salinity on the cumulative transpiration of papaya as affected by salinity stress

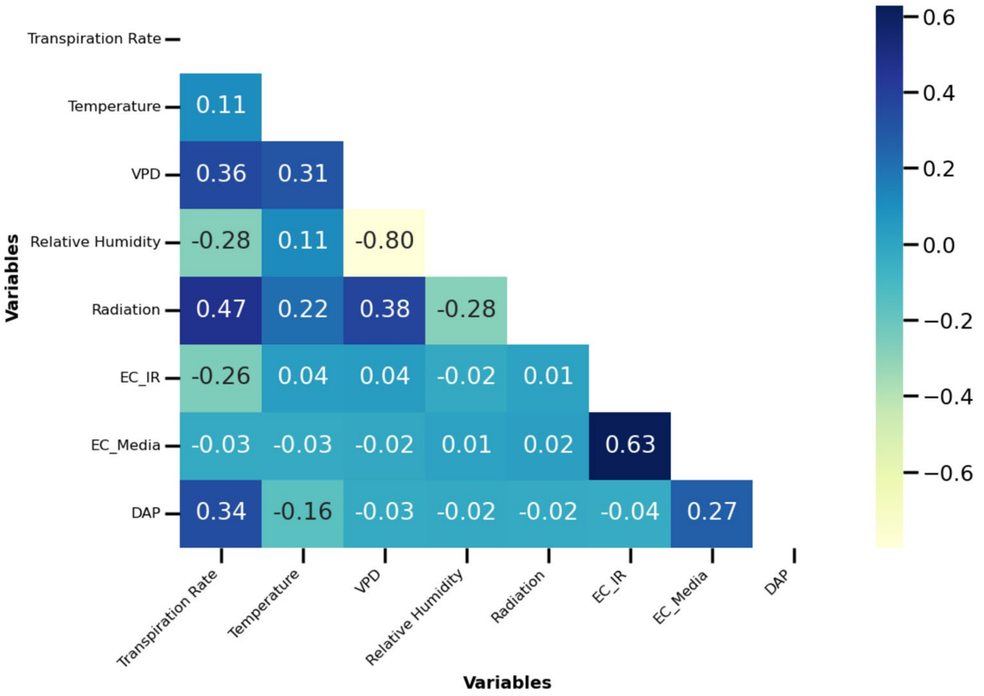


Fig. 3 Correlation matrix of input parameters. VPD: vapor pressure deficit, EC_IR: Electrical conductivity of irrigation water, EC_Media: Electrical conductivity of growing media, DAP: Day after planting

Table 1 Statistical evaluation of XGBt_1, CATBt_1, LAGBt_1, RF_1, DT_1, XGBt_2, CATBt_2, LAGBt_2, RF_2, and DT_2 on test data set

		R^2	RMSE (g/m)	MAE (g/m)	R2 rank	RMSE rank	MAE rank
XGBt _1	Rs, VPD, DAP, EC_IR	0.828	0.0651	0.0444	1	1	1
RF_1		0.8256	0.0656	0.0447	2	2	2
DT_1		0.8098	0.0685	0.0467	3	3	4
LAGBt_1		0.8076	0.0689	0.0462	4	4	3
CATBt_1		0.8074	0.0689	0.0490	5	5	6
XGBt_2	Rs, VPD, DAP	0.8071	0.0690	0.0473	6	6	5
RF_2		0.6934	0.0870	0.0599	7	7	7
LAGBt_2		0.6922	0.0871	0.0599	8	8	8
CATBt_2		0.6921	0.0871	0.0599	9	9	9
DT_2		0.6915	0.0872	0.0601	10	10	10

Rs: Solar radiation, VPD: Vapor pressure deficit, EC_IR: Electrical Conductivity of irrigation water, DAP: Day After Planting

al. 2021; Guo et al. 2024; Hailegnaw et al. 2024; He et al. 2024; Lee et al. 2024; Li et al. 2020; Shao et al. 2022; Wang et al. 2024). For instance, Fan et al. (2021) indicated the efficiency of machine learning models such as XGBt with R^2 value of 0.929.

The extreme gradient model outperformed all machine learning models by the overall consideration of R^2, RMSE, and MAE (Table 1). A superior accuracy of the random forest model in estimating plant evapotranspiration was evident (Ge et al. 2022; Zhou et al. 2021). A similar better performance of random forest over other machine learning models in simulating plant evapotranspiration is evident, which succeeded extreme gradient (Li et al. 2020). It is essential to accurately simulate plant transpiration to effectively develop crop stress management strategies. This helps combat the effects of salinity by ensuring that plants receive sufficient water to meet their evapotranspiration needs, reducing salt content, and flashing out salts from the root zone. Integrating plant transpiration data with an automated irrigation system that adjusts based on real-time or simulated plant transpiration rate will ultimately help mitigate the impact of salinity on plant growth and yield.

Model deployment

A precise identification of salinity levels that negatively affect plant physiology and growth is critical for effective salinity management. Understanding these levels allows for the determination of critical salinity thresholds that, if exceeded, could lead to adverse physiological responses in plants. Such information is vital for planning targeted interventions to protect crop health and optimize yields under

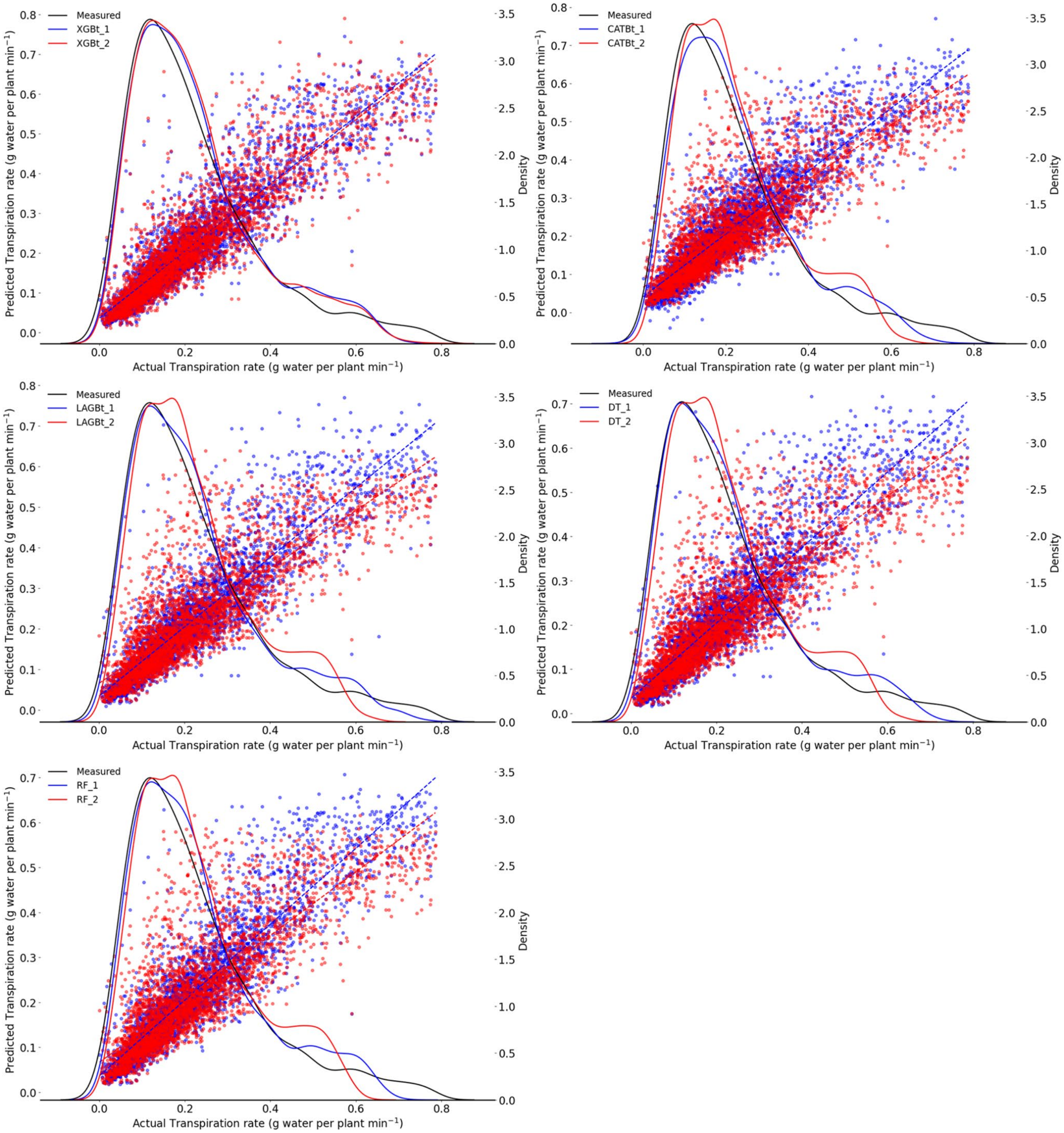


Fig. 4 Scatter and density plot of measured and estimated transpiration rate (g water per plant per min) using three input variables (Solar Radiation, Vapor Pressure Deficit, and Salinity Level) for XGBt_1, CATBt_1, LAGBt_1, RF_1, and DT_1, and Two Input Variables (Solar Radiation and Vapor Pressure Deficit) for XGBt_2, CATBt_2, LAGBt_2, RF_2, and DT_2

saline conditions. However, conducting extensive studies to assess the impact of salinity on plant physiological responses at a field or greenhouse scale can be challenging. These studies often require significant time, and resources, and are practically limited due to space, time, and equipment constraints. In this study, we used papaya in Turface® potting medium as the model crop system and included a control alongside three EC_IR levels 2, 4, and 8 dS m^{-1} as a baseline information. A previous study was available about the effects of these salinity levels on papaya in this medium from a previous study (de Souza et al. 2024) where a phenotyping system was not employed and transpiration was

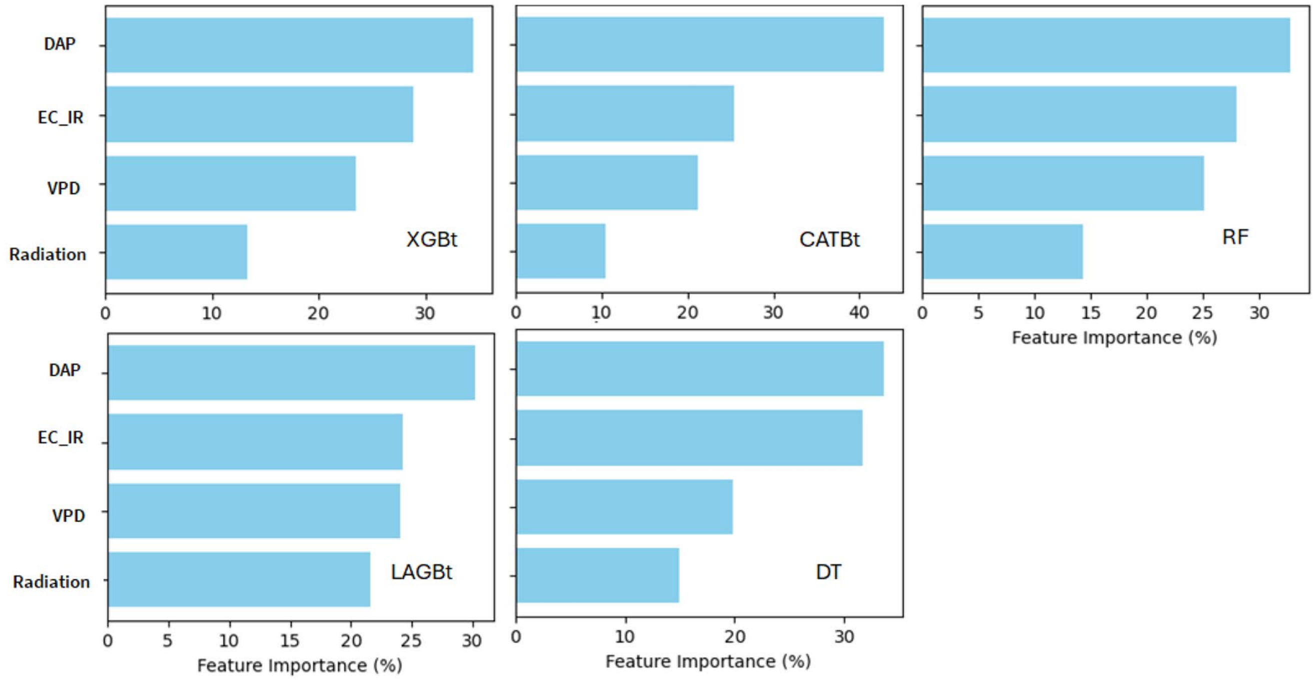


Fig. 5 Feature importance evaluation of Input Variables (DAP: Day after planting, Solar Radiation, Vapor Pressure Deficit, and EC_IR) in modeling transpiration using machine learning models (XGBt, CATBt, LAGBt, RF, and DT)

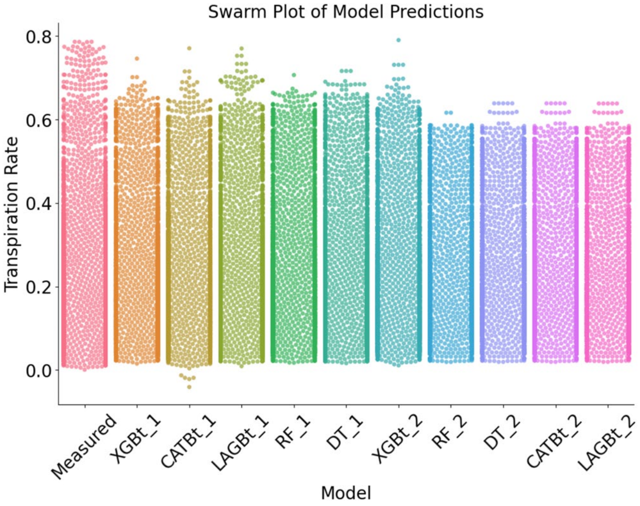


Fig. 6 Swarm plots of measured transpiration rate and estimated by XGBt_1, CATBt_1, LAGBt_1, RF_1, and DT_1 using solar radiation, vapor pressure deficit, days after planting and EC_IR and XGBt_2, CATBt_2, LAGBt_2, RF_2, and DT_2 using solar radiation, vapor pressure deficit and days after planting

measured only on a leaf area basis. Thus, we indicated the salinity levels that may elicit a response of the same plant in the same medium using the phenotyping platform. The results of the current study showed a significant negative impact of an EC_IR level of 8 dS m^{-1} on papaya transpiration. The effects of intermediate EC_IR levels such as 5, 6, and 7 dS m^{-1}, were not tested due to the limitations of the plant phenotyping platform and the number of replicated treatments that the phenotyping system could accommodate. To fill in the "gaps" in the data and better predict the response of papaya to varying EC_IR levels, we developed and deployed five machine-learning models capable of simulating plant transpiration under a broader range of EC_IR levels, and therefore extended our analysis to include nine distinct EC_IR levels with 1 dS m^{-1} increments (0, 1, 2, 3, 4, 5, 6, 7, and 8 dS m^{-1}). This finer gradual increment allowed us to capture more detailed insights into the salinity response, particularly at levels where empirical data was previously lacking. All algorithms successfully predicted transpiration as affected by EC_IR levels. Transpiration increased up to EC_IR of 2 dS m^{-1} and then decreased beyond in all machine learning models. However, a significant decline in daily average transpiration was induced at 6 dS m^{-1} in all machine learning models used in this study (Fig. 7).

This finding agrees with de Souza et al. (2024), who reported reduced transpiration (measured on a leaf area basis) of papaya at an EC_IR of 6 dS m^{-1}. The predicted cumulative transpiration values from the XGBt, CATBt, LAGBt, DT, and RF models demonstrated a strong fit to a quadratic regression model, each yielding R² value of 0.99 (Fig. 8). These enabling the identification of maximum EC_IR level for papaya plants up to the age of 15 weeks. By simulating the impacts of previously untested EC_IR levels, our study was able to develop a quadratic equation, which could be used to predict the possible effect of EC_IR on the

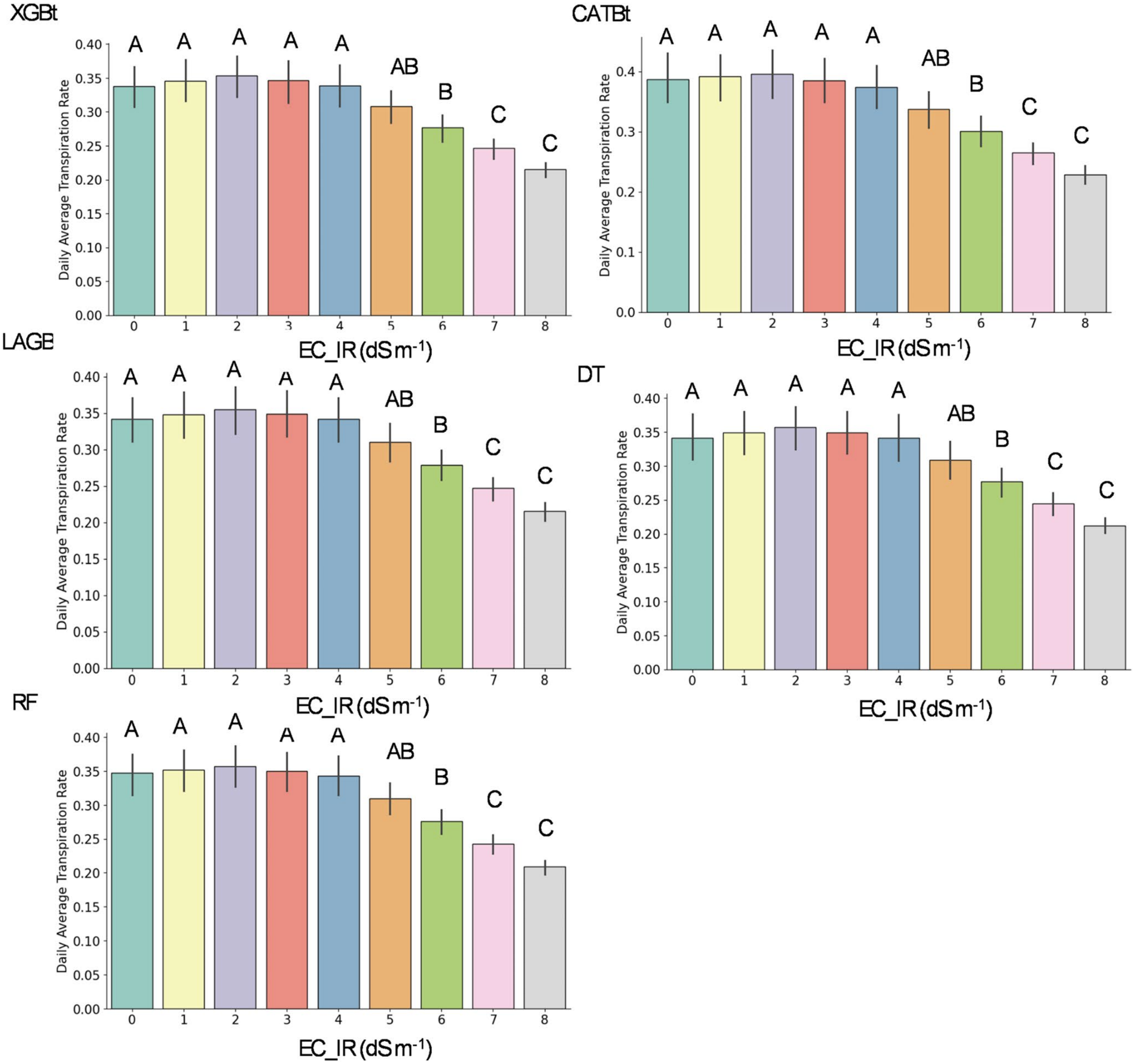


Fig. 7 The impact of salinity on the daily average whole-plant transpiration of papaya (g water per plant per minute) at EC_IR levels of 0, 1, 2, 3, 4, 5, 6, 7, and 8 dS m^{-1} (data predicted by machine learning models)

transpiration rate of papaya plants to the age of 15 weeks. This information is crucial for developing effective salinity mitigation strategies in papaya cultivation, potentially aiding in protecting yield loss and overall plant health in saline-prone regions. Hence, there is a need to confirm this finding with an observed data set collected from a real environment with a wide range of salinity.

Effect of salinity on papaya biomass

The impact of salinity on the biomass of papaya was evident, where 8 dS m^{-1} significantly decreased the total biomass, and dry weight of stem, leaf, and root weight by 37%, 31%, 36%, and 47%, respectively (Fig. 9). However, there was no significant biomass change at 2 and 4 dS m^{-1}. These results agree with de Souza et al. (2024) where a EC_IR level of 6 dS m^{-1} and below did not significantly impact papaya biomass. Generally, the most notable decline was observed in the plant's root system. Salinity stress causes a significant impact on the growth and development of plants by hindering transpiration (Munns and Tester 2008). The decline in papaya dry biomass with salinity has been reported multiple times (Aguilar-Bautista et al. 2022; Álvarez-Méndez et al. 2022; De Lima-Neto et al. 2016; Sá et al. 2016).

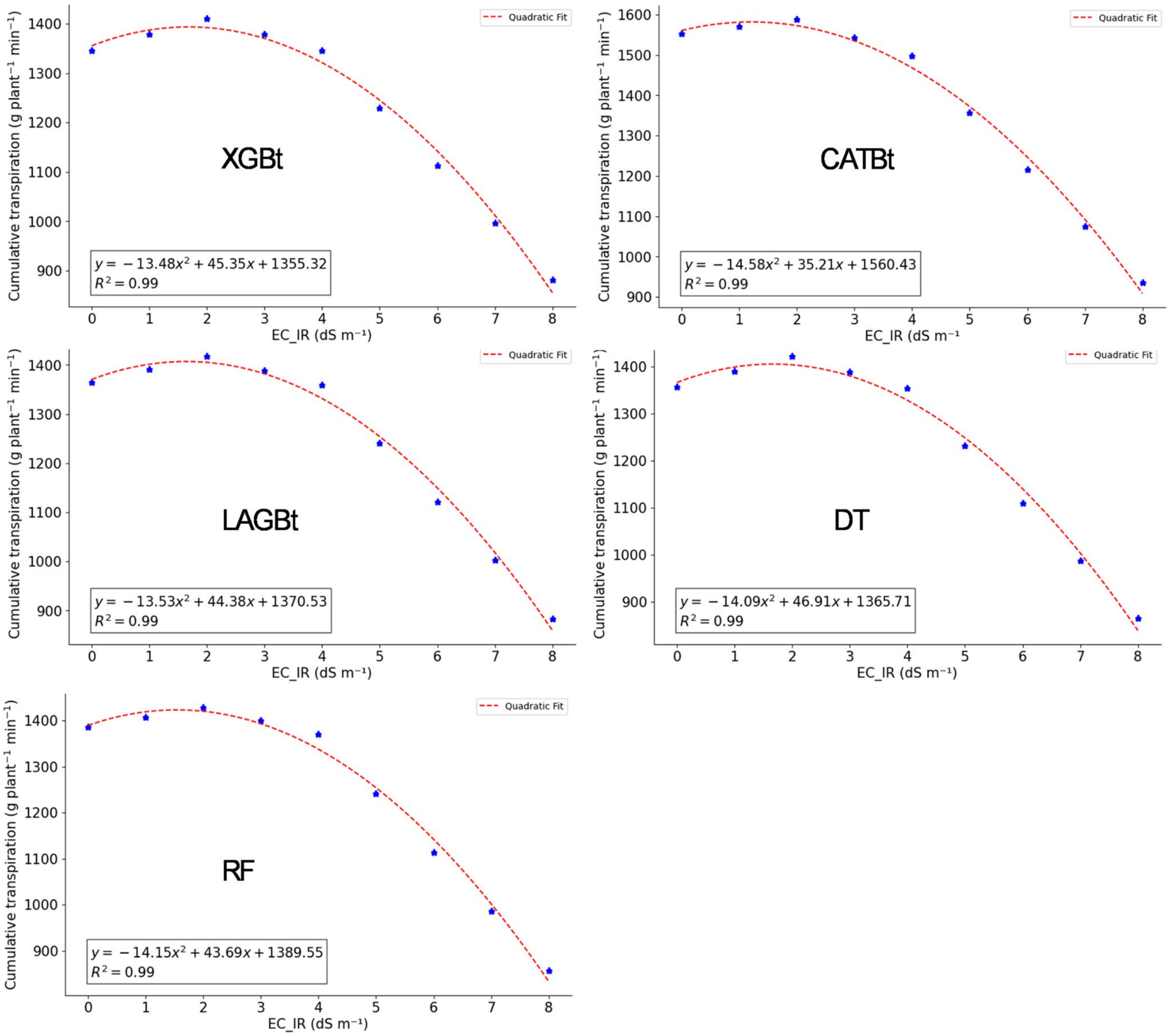


Fig. 8 Relation between EC_IR and machine learning models (XGBt, CATBt, LAGBt, RF, and DT) predicted transpiration based on new set of experimental data

The decline in papaya biomass could be induced due to Na^+ and Cl^- toxicity, thus competing with the uptake of nutrients such as Ca^{2+}, K^+, and NO_3^- (Huang et al. 2017; Mengel and Kirkby 2010). Figure 10 shows the negative correlation between the content of Na^+ in papaya leaves with the total biomass, and dry weight of stem, leaf, and root (*r* = -0.64, -0.48, -0.59, and −0.73−0.72, respectively). Additionally, the content of Na^+ in the root was negatively correlated with total biomass, stem, leaf, and root dry weights (*r* = -0.64, -0.55, -0.65, and −0.62, respectively). The negative correlation between Na^+ content in plants and their biomass and growth rate is well-documented in the literature (Munns and Tester 2008).

Plants employ various physiological mechanisms to combat salinity-induced stress. One such mechanism involves reducing leaf area to minimize water loss, which limits water and consequently salt uptake by the roots (Sá et al. 2016). Plants can also manage salinity by lowering the hydraulic conductivity of their roots, thus reducing the uptake and transport of salts to the upper tissues of the plant. This adaptive response significantly reduces plant transpiration, which is crucial in maintaining water balance and transporting nutrients. However, decreased hydraulic conductivity leads to salt accumulation in the root zone, causing osmotic stress and impeding water and nutrient absorption, ultimately decreasing plant biomass (Vysotskaya et al. 2010). Moreover, the decline in plant growth is directly

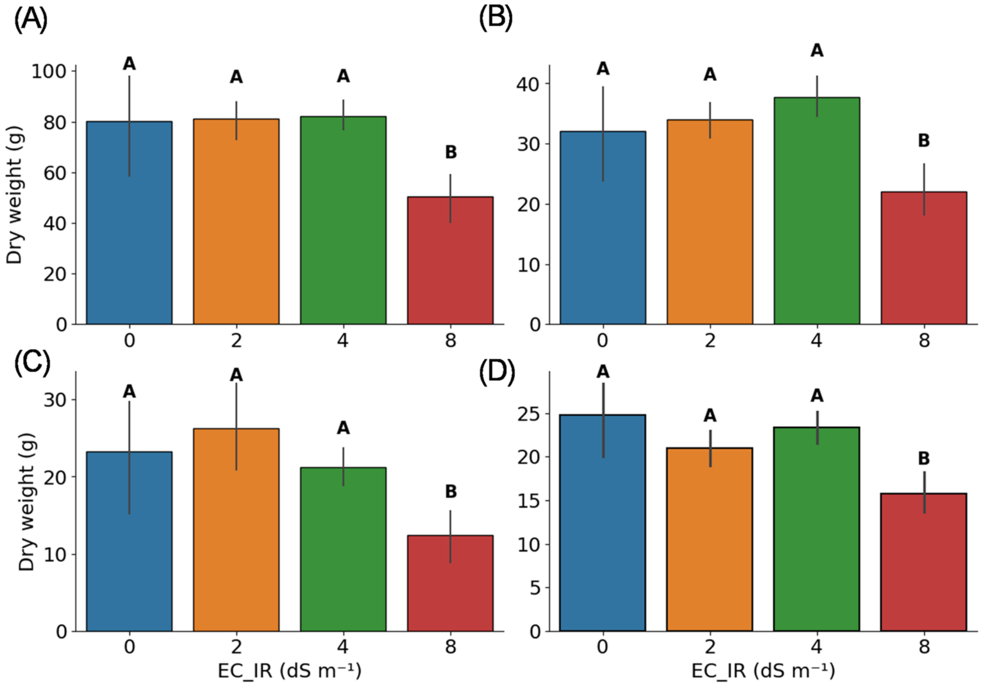
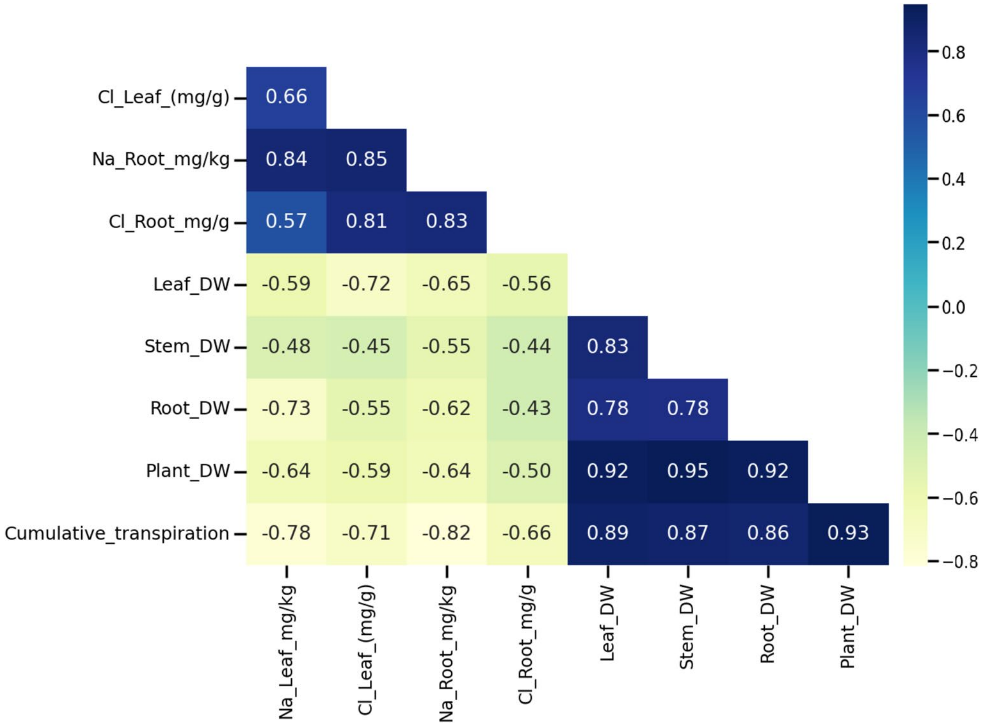


Fig. 9 The effect of salinity on the (**A**) total biomass, (**B**) stem, (**C**) root, and (**D**) leaf dry weight of papaya



	Na_Leaf_mg/kg	Cl_Leaf_(mg/g)	Na_Root_mg/kg	Cl_Root_mg/g	Leaf_DW	Stem_DW	Root_DW	Plant_DW
Cl_Leaf_(mg/g)	0.66							
Na_Root_mg/kg	0.84	0.85						
Cl_Root_mg/g	0.57	0.81	0.83					
Leaf_DW	-0.59	-0.72	-0.65	-0.56				
Stem_DW	-0.48	-0.45	-0.55	-0.44	0.83			
Root_DW	-0.73	-0.55	-0.62	-0.43	0.78	0.78		
Plant_DW	-0.64	-0.59	-0.64	-0.50	0.92	0.95	0.92	
Cumulative_transpiration	-0.78	-0.71	-0.82	-0.66	0.89	0.87	0.86	0.93

Fig. 10 Correlation between plant parameters and cumulative transpiration

associated to a decline in transpiration as papaya was experiencing a decline in transpiration as salinity increased beyond certain level.

Effect of salinity on the uptake of Na+ and Cl-

Salinity induced a significant increase of Na^+ content in the leaves and roots of papaya (Fig. 11), with a significant increase starting at 4 dS m^{-1} in leaves and 2 dS m^{-1} in the roots. The content of Na^+ increased by 362%, 260%, and 69% in the leaves, and by 320%, 219%, and 151% in the roots at EC_IR levels of 8, 4, and 2 dS m^{-1}, respectively. The accumulation of ions in plant-growing media generally leads to their preferential uptake and subsequent accumulation in plant tissue, which can potentially reach toxic levels. The toxic level of Na^+ in fruit plant tissue ranges between

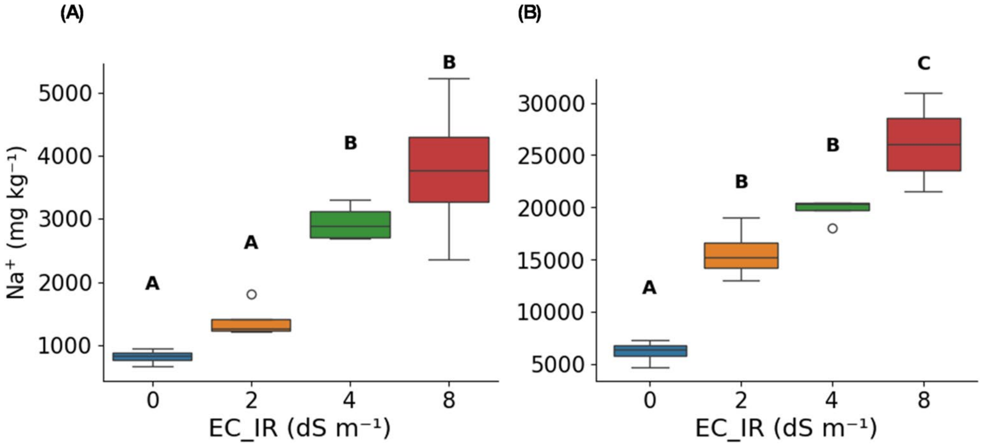


Fig. 11 The effect of salinity on Na^+ content of (A) Papaya leaves, and (B) Papaya roots

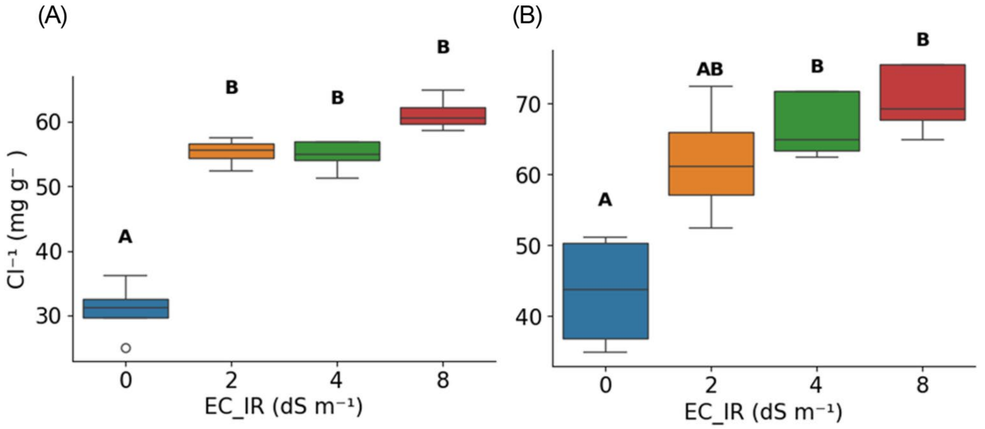


Fig. 12 The effect of salinity on Cl^- content of (**A**) Papaya leaves, and (**B**) Papaya roots

0.2 and 0.5%. (Ayers and Westcot 1985). Na^+ and its preferential uptake can also antagonistically affect the uptake of other cations such as K^+ and Ca^{2+}. For instance, based on the study of Aguilar-Bautista et al. (2022), there was a 25% decline in the uptake of K^+ and a 39% decline in the uptake of K and Ca, respectively. Therefore, salt stress can reduce the uptake of nutrients such as N, K^+ and Ca^{2+} (Uygur and Yetisir 2009).

The contents of Cl^- in both roots and leaves increased with an increment of EC_IR levels (Fig. 12). The Cl^- content showed a significant increase starting from the EC_IR level of 2 dS m^{-1} in leaves and 4 dS m^{-1} in roots. The increase in Cl^- content was substantial, reaching up to 98% in the leaves and 71% in the roots of the papaya plant at a EC_IR level of 8 dS m^{-1}. Chloride toxicity in tree crops such as papaya can be seen when Cl^- content in plant tissue reaches above 0.3% (Ayers and Westcot 1985).

The abundance of Cl^- in solution may interfere with the uptake of nutrients such as NO_3^- (Mengel and Kirkby 2010). Both Na^+ and Cl^- were more accumulated in the roots of papaya. This indicates further uptake and translocation of salts were hindered and accumulated at the root of the papaya plant. In saline soil, after a certain time, plants have to almost fully exclude Na^+ and Cl^- to maintain osmotic balance, and the process is very costly in terms of energy expenditure (Munns et al. 2020).

Conclusion

The use of whole-plant high-throughput phenotyping systems, such as the PlantArray utilized in this study, overcomes the challenges of measuring plant transpiration and generates extensive datasets on plant water use dynamics. The study used plant transpiration data collected from papaya (from 1 week to ~15 weeks after transplanting) grown in PlantArray system. All machine learning models (extreme gradient boosting, categorical boosting, light gradient boosting, random forest, and decision tree) effectively captured the salinity-induced change in transpiration. Machine learning models that included salinity as an input variable performed substantially better than the other models, highlighting the importance of considering salinity in predictive models for accurate irrigation scheduling. Salinity significantly contributed to the predictive capacity of all machine learning models by up to 32%. When salinity was included in the models, the accuracy of transpiration estimation increased by up to 19% for R^2, 25% for RMSE, and 25% for MAE. Furthermore, this study deployed the developed machine learning algorithms on a new data set and identified 6 dS m^{-1} as a critical salinity level (EC of irrigation water) for 6- to 15-week-old papaya plants grown in an inert clay potting medium. A decline in biomass and an increase in the concentration of Na^+ and Cl^- were also evident due to salinity. The observed decline in biomass can be attributed to Na^+ and Cl^- induced toxicity, which competes with the uptake of essential nutrients such as Ca^{2+}, K^+, and NO_3^-. By integrating real-time transpiration data with automated irrigation systems, farmers can gauge plant stress due to soil salinity and adjust water application based on the plant's actual needs, thus mitigating the adverse effects of salinity levels. This approach not only ensures sufficient water supply to meet evapotranspiration demands but also helps in flushing out salts from the root zone, thereby maintaining plant health and productivity.

Acknowledgements This publication is based on research funded by the United States Department of Agriculture's National Institute of Food and Agriculture under award number 2020-67019-31163. Any opinions, results, conclusions, or recommendations contained in this publication are solely those of the authors and do not necessarily reflect the views of the USDA.

Author contributions Niguss Solomon Hailegnaw and Haimanote Bayabil: Conceptualization of the paper idea, Bruce Schaffer: Research design and review, Niguss Solomon Hailegnaw: Writing the draft paper, data analysis, and modeling; Haimanote Bayabil: Supervising and reviewing, Girma Awoke: Writing and Reviewing, Aline Camargo: writing and reviewing; Edivan Rodrigues: Reviewing.

Data availability Data will be available up on request.

Declarations

Competing interests The authors declare no competing interests.

References

Aguilar-Bautista A, Alejo-Santiago G, Aburto-González CA, Juárez-Rosete CR, Bugarín-Montoya R, López-Guzmán GG, García-Herrera T, Aguilar-Bautista A, Alejo-Santiago G, Aburto-González CA, Juárez-Rosete CR, Bugarín-Montoya R, López-Guzmán GG, García-Herrera T (2022) Salinity Effect of Nutrient Solution on Maradol papaya. Cienc. Tecnol. Agropecu. 23. https://doi.org/10.21930/rcta.vol23_num3_art:2396

Álvarez-Méndez SJ, Urbano-Gálvez A, Mahouachi J, Álvarez-Méndez SJ, Urbano-Gálvez A, Mahouachi J (2022) Mitigation of salt stress damages in *Carica papaya* L. seedlings through exogenous pretreatments of gibberellic acid and proline. Chil J Agric Res 82:167–176. https://doi.org/10.4067/S0718-58392022000100167

Ayers, Robert S., and Dennis W. Westcot. Water quality for agriculture. Vol. 29. Rome: Food and agriculture organization of the United Nations, 1985.

Breiman L (2001) Random Forests Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Chele KH, Tinte MM, Piater LA, Dubery IA, Tugizimana F (2021) Soil salinity, a serious environmental issue and plant responses: A metabolomics perspective. Metabolites 11. https://doi.org/10.3390/metabo11110724

de Camargo Santos, A, Schaffer, B, Fotopoulos, V, Moon, P, Shahid, M, Rowland, DL, Tillman, B, Bremgartner, M, Iannou, AG, Bassil, E. (2024) Melatonin seed priming improves early establishment and water stress tolerance in peanut. Plant Physiology and Biochemistry 211:108664. https://doi.org/10.1016/j.plaphy.2024.108664

de Lima-Neto, AJ, Cavalcante, LF, Mesquita, F. de O, Souto, AG de L, dos Santos, GP, dos Santos, JZ, de Mesquita, EF, (2016) Papaya seedlings irrigation with saline water in soil with bovine biofertilizer. Chil. J Agric Res 76, 236–242. https://doi.org/10.4067/S0718-58392016000200014

de Souza ER, Schaffer B, Vargas AI, de Camargo Santos A, Rodriguez EAG (2024) Contrasting salinity effects of artificial seawater and sodium chloride on Carica papaya L. cultivar red lady physiology and growth. CABI Agric Biosci 5:15. https://doi.org/10.1186/s43170-024-00216-3

Du C, Jiang S, Chen C, Guo Q, He Q, Zhan C (2024) Machine Learning-Based Estimation of daily cropland evapotranspiration in diverse climate zones. Remote Sens 16:730

Fan J, Zheng J, Wu L, Zhang F (2021a) Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. Agric Water Manag 245:106547. https://doi.org/10.1016/j.agwat.2020.106547

Fan J, Zheng J, Wu L, Zhang F (2021b) Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. Agric Water Manag 245:106547. https://doi.org/10.1016/j.agwat.2020.106547

Ge J, Zhao L, Yu Z, Liu H, Zhang L, Gong X, Sun H (2022) Prediction of greenhouse tomato crop evapotranspiration using XGBoost machine learning model. Plants 11:1923. https://doi.org/10.3390/plants11151923

Guo N, Chen H, Han Q, Wang T (2024) Evaluating data-driven and hybrid modeling of terrestrial actual evapotranspiration based on an automatic machine learning approach. J Hydrol 628:130594

Hailegnaw NS, Bayabil HK, Berihun ML, Teshome FT, Shelia V, Getachew F (2024) Integrating machine learning and empirical evapotranspiration modeling with DSSAT: implications for agricultural water management. Sci Total Environ 912:169403. https://doi.org/10.1016/j.scitotenv.2023.169403

Halperin O, Gebremedhin A, Wallach R, Moshelion M (2017) High-throughput physiological phenotyping and screening system for the characterization of plant-environment interactions. Plant J 89(4):839–850. https://doi.org/10.1111/tpj.13425

Hancock JT, Khoshgoftaar TM (2020) CatBoost for big data: an interdisciplinary review. J Big Data 7:94. https://doi.org/10.1186/s40537-020-00369-8

Hanlon EA, Gonzalez JG, Bartos JM (1994) IFAS extension soil testing laboratory chemical procedures and training manual. Fla Coop Ext Serv. Circular 812

He X, Liu S, Bateni SM, Xu T, Jun C, Kim D, Li X, Song L, Zhao L, Xu Z (2024) Innovative approach for estimating evapotranspiration and gross primary productivity by integrating land data assimilation, machine learning, and multi-source observations. Agric Meteorol 355:110136

Hoagland DR, Arnon DI (1950) The water-culture method for growing plants without soil. Calif Agric Exp Stn 347:1

Huang L, Liu X, Wang Z, Liang Z, Wang M, Liu M, Suarez DL (2017) Interactive effects of pH, EC and nitrogen on yields and nutrient absorption of rice (*Oryza sativa* L.). Agric. Water Manag 194:48–57. https://doi.org/10.1016/j.agwat.2017.08.012

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: A highly efficient gradient boosting decision tree, in: advances in neural information processing systems. Curran Associates, Inc

Kemanian AR, Stöckle CO, Huggins DR (2005) Transpiration-use efficiency of barley. Agric Meteorol 130:1–11. https://doi.org/10.1016/j.agrformet.2005.01.003

Lee J, Bateni SM, Jun C, Heggy E, Jamei M, Kim D, Ghafouri HR, Deenik JL (2024) Hybrid machine learning system based on multivariate data decomposition and feature selection for improved multitemporal evapotranspiration forecasting. Eng Appl Artif Intell 135:108744

Li L, Chen S, Yang C, Meng F, Sigrimis N (2020) Prediction of plant transpiration from environmental parameters and relative leaf area index using the random forest regression algorithm. J Clean Prod 261:121136. https://doi.org/10.1016/j.jclepro.2020.121136

Lu Y, Fricke W (2023) Salt Stress—Regulation of root water uptake in a Whole-Plant and diurnal context. Int J Mol Sci 24:8070. https://doi.org/10.3390/ijms24098070

Majeed A, Muhammad Z (2019) Salinity: A major agricultural Problem—Causes, impacts on crop productivity and management strategies. In: Hasanuzzaman M, Hakeem KR, Nahar K, Alharby HF (eds) Plant abiotic stress tolerance: agronomic, molecular and biotechnological approaches. Springer International Publishing, Cham, pp 83–99. https://doi.org/10.1007/978-3-030-06118-0_3

Malavolta E, Vitti GC, Oliveira AS (1997) Avaliação do Estado Nutricional das Plantas: Princípios e Aplicações. Piracicaba: Potafos, p 201

Mengel K, Kirkby E. Principles of plant nutrition. Vol. 1. springer science & Business Media, 2001.

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59:651–681. https://doi.org/10.1146/annurev.arplant.59.032607.092911

Munns R, Passioura JB, Colmer TD, Byrt CS (2020) Osmotic adjustment and energy limitations to plant growth in saline soil. New Phytol 225:1091–1096. https://doi.org/10.1111/nph.15862

Nicolas F, Kamai T, Ben Gal A, Ochoa-Brito J, Daccache A, Ogunmokun F, Kisekka I (2023) Assessing salinity impacts on crop yield and economic returns in the central Valley. Agric Water Manag 287. https://doi.org/10.1016/j.agwat.2023.108463

Peçanha AL, da Silva JR, Rodrigues WP, Ferraz TM, Torres Netto A, Lima RSN, Lopes TS, Ribeiro MS, Deus BCdaSde, Couto TRdo, Schaffer B, Campostrini E (2017) Leaf gas exchange and growth of two Papaya (*Carica Papaya* L.) genotypes are affected by elevated electrical conductivity of the nutrient solution. Sci Hortic 218:230–239. https://doi.org/10.1016/j.scienta.2017.02.018

Sá, Francisco Vanies da Silva, Marcos Eric Barbosa Brito, Rômulo Carantino Lucena Moreira, Alberto Soares de Melo, Luderlândio de Andrade Silva, Hans Raj Gheyi, Lizaiane Cardoso de Figueiredo, Emanoela Pereira de Paiva (2016) Balance of salts and growth of papaya cultivars irrigated with saline water. Biosci J 32(4). https://doi.org/10.14393/BJ-v32n4a2016-32753

Shrivastava P, Kumar R (2015) Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. Saudi J Biol Sci 22:123–131. https://doi.org/10.1016/j.sjbs.2014.12.001

Stavridou E, Hastings A, Webster RJ, Robson PRH (2017) The impact of soil salinity on the yield, composition and physiology of the bioenergy grass *Miscanthus* × *giganteus*. GCB Bioenergy 9:92–104. https://doi.org/10.1111/gcbb.12351

Tahraoui H, Amrane A, Belhadj A-E, Zhang J (2022) Modeling the organic matter of water using the decision tree coupled with bootstrap aggregated and least-squares boosting. Environ Technol Innov 27:102419. https://doi.org/10.1016/j.eti.2022.102419

Uygur V, Yetisir H (2009) Effects of rootstocks on some growth parameters, phosphorous and nitrogen uptake watermelon under salt stress. J Plant Nutr 32:629–643. https://doi.org/10.1080/01904160802715448

Vysotskaya L, E Hedley P, Sharipova G, Veselov D, Kudoyarova G, Morris J, G Jones H (2010) Effect of salinity on water relations of wild barley plants differing in salt tolerance. AoB PLANTS 2010(plq006). https://doi.org/10.1093/aobpla/plq006

Wang J, Gao X, Zhou Y, Wu P, Zhao X, Cai Y (2024) Modeling the dynamics of evapotranspiration of wolfberry (*Lycium barbarum* L.) under different cultivation methods on the Tibetan plateau. J Hydrol 131537

Zhang W, Wu C, Zhong H, Li Y, Wang L (2021) Prediction of undrained shear strength using extreme gradient boosting and random forest based on bayesian optimization. Geosci Front 12:469–477. https://doi.org/10.1016/j.gsf.2020.03.007

Zhou Z, Zhao L, Lin A, Qin W, Lu Y, Li J, Zhong Y, He L (2021) Exploring the potential of deep factorization machine and various gradient boosting models in modeling daily reference evapotranspiration in China. Arab J Geosci 13:1287. https://doi.org/10.1007/s12517-020-06293-8