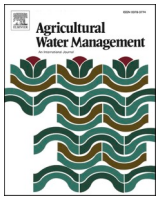
Agricultural Water Management 324 (2026) 110113

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat



Modelling water - use and yield of selected irrigated subtropical crops using machine learning and hybrid models in north - eastern South Africa

Prince Dangare [a,f,*], Paul J.R. Cronje [a,b], Zama E. Mashimbye [c], Joseph Masanganise [a,h], Zanele Ntshidi [e,g], Shaeden Gokool [i,j], Vivek Naiken [i,k], Tendai Sawunyama [d,i], Sebinasi Dzikiti [a]

[a] Department of Horticultural Science, Stellenbosch University, Stellenbosch, South Africa
[b] Citrus Research International, Mbombela, South Africa
[c] Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch, South Africa
[d] Inkomati – Usuthu Catchment Management Agency, South Africa
[e] Arid Lands Node, South African Environmental Observation Network (SAEON), Kimberley, South Africa
[f] Department of Electronics and Telecommunications, University of Zimbabwe, Harare, Zimbabwe
[g] Department of Physical and Earth Sciences, Sol Plaatje University, Kimberley, South Africa
[h] Department of Physics and Engineering, Bindura University of Science Education, Bindura, Zimbabwe
[i] Centre for Water Resources Research, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa
[j] South African Environmental Observation Network (SAEON), Grasslands-Forests-Wetlands Node, Pietermaritzburg 3201, South Africa
[k] Department of Agrometeorology, Schools of Agriculture and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

ARTICLE INFO

Handling Editor-Dr. Brent Clothier

Keywords:
Crop coefficients
Evapotranspiration
Extreme gradient boosting
Light gradient boosting machine learning
Random forest
Transpiration
Yield response factor

ABSTRACT

The Inkomati-Usuthu Water Management Area in Mpumalanga, South Africa, is a main producer of subtropical crops. These crops are mainly produced under irrigation, yet water resources in this catchment are nearly fully allocated. This calls for improved irrigation efficiency in the region to save water needed for supporting agricultural expansion. Accurate derivation of crop coefficients (K_c) and yield response factors (K_y) is vital for irrigation management and yield prediction. In this study, transpiration (T) for the crops was measured using the heat ratio method of monitoring sap flow while evapotranspiration (ET) was quantified using eddy covariance and surface renewal techniques. Leaf area index for the fields was derived from Landsat 8 imagery. Light gradient boosting machine (LightGBM), Random Forest (RF) and Extreme gradient boosting (XGBoost) machine learning models was investigated for predicting the crop ET and T of banana, grapefruit, litchi, mango and sugarcane. The best performing ET and T machine learning-based models were used for developing the crop coefficients (K_c) and a hybrid model for predicting K_y, respectively. The LightGBM achieved the highest accuracy in predicting banana, grapefruit, litchi and sugarcane ET. The XGBoost achieved the highest accuracy in predicting mango ET. The LightGBM achieved the highest accuracy in predicting the grapefruit, litchi and mango T. All the ET and T models produced coefficient of determination in the range 0.83–0.96, root mean square error ranging from 0.02 to 0.10 mm/h, mean absolute error ranging from 0.01 to 0.06 mm/h and Kling-Gupta efficiency in the range 0.88–0.97. The grapefruit, litchi and mango produced K_y values of 2.70, 2.50, and 2.90 respectively. The derived K_c and K_y information can assist irrigation managers optimize irrigation to promote productive water use in the water scarce regions.

1. Introduction

The agricultural sector in South Africa is under increasing threats from serious water shortages (Mabhaudhi et al., 2021). Aggravating these shortages is the ever-increasing pressure for the limited resource due to population growth, increased industrial activities, periodic droughts due to climate change impacts, among others (Dangare et al., 2018a; Mpakairi et al., 2025). The Inkomati-Usuthu Water Management

* Correspondence to: Department of Horticultural Science, Stellenbosch University, Lombardi Building, Corner of Victoria and Neethling streets Private Bag X1, Stellenbosch, 7599, South Africa.
E-mail addresses: pdangare@ceic.uz.ac.zw, 22284168@sun.ac.za (P. Dangare).

https://doi.org/10.1016/j.agwat.2025.110113
Received 24 September 2025; Received in revised form 20 November 2025; Accepted 25 December 2025
Available online 6 January 2026

Area in Mpumalanga, South Africa, is a main producer of subtropical crops (Feliciano et al., 2025). These crops are mainly produced under irrigation, and water resources in this catchment are nearly fully allocated (Inkomati-Usuthu CMA Business Case, 2012). Furthermore, climate change is anticipated to amplify evaporation, thereby reducing available freshwater resources and increasing the water scarcity (Bringhenti et al., 2025; Engelbrecht et al., 2024). Currently, little information exists on the water - use of these irrigated crops, which adversely affects the sustainability and growth of the agricultural industry. This calls for improved irrigation efficiency in the region, leading to a reduction in irrigation freshwater use and saving water needed to support agricultural expansion (Reinders et al., 2013).

Evapotranspiration comprises of both evaporation (*E*) and transpiration (*T*) (Taheri et al., 2025, 2022), which is a passive process controlled by available energy, vapour pressure deficit, and constrained by soil water availability (Sadras et al., 2020). Evapotranspiration information is very vital in irrigation scheduling. Additionally, formulating irrigation schedules that optimize accessible water requires information on crop yield responses to water, and this can be estimated using empirical water-yield relation models (Paredes et al., 2014). The crop yield models relate yield to either evapotranspiration (*ET*) or crop transpiration (*T*) (Doorenbos and Kassam, 1979; Hanks, 1974; Jensen, 1968; Paredes et al., 2018; Stewart et al., 1977). The crop transpiration (*T*) is preferred for yield modelling because it is directly related to the crop yield development (Paredes et al., 2014). Crop transpiration occurs at the same time with photosynthetic uptake of carbon dioxide through the plant leaves stomatal openings, thus it is linked to the plant productivity (Ding et al., 2013; Jones and Tardieu, 1998; Pieruschka et al., 2010). Whole plant transpiration can be measured using sap flow sensors (Capurro et al., 2024), and this can be scaled to the field level using results from several sap flow sensors installed in the field (Bethenod et al., 2000; Dzikiti et al., 2022). However, the use of multiple sap flow sensors in transpiration measurement is expensive and requires complex implementation (Ballester et al., 2013; Dangare et al., 2018b), thus limiting their application in irrigation management. Therefore, an accurate transpiration model that requires simple model inputs is needed to estimate crop transpiration for yield modelling in a wide range of cultivation environments (Bringhenti et al., 2025; Villalobos et al., 2013).

The *ET* data is fundamental to the discernment and management of water resources (Allen et al., 2011), and can be determined using two primary techniques namely direct and indirect methods (Taheri et al., 2025). Direct methods that measure *ET* include weighing lysimeters (Marek et al., 1988), eddy covariance (Baldocchi et al., 1988), scintillometry method (De Bruin et al., 1995), Bowen ratio (Bowen, 1926) and surface renewal (Paw U et al., 1995), among others. Direct methods for *ET* measurement are time-consuming, costly, and not readily available in several regions (Li et al., 2025a; Tang et al., 2018). Thus, these direct methods are however not suited for routine use in orchard water management. Indirect methods of determining *ET* apply mathematical models (physical models) to compute the latent heat flux (Abdullah and Malek, 2016). The *ET* can be determined using FAO-56 models (Allen and Pereira, 2009; Er-Raki et al., 2010; Qiu et al., 2015; Wei et al., 2013), however this has often led to incorrect water use estimates. This is because the division of crop growth period and related crop coefficients are intimately linked to local climate and crop conditions (Zhao et al., 2015). In recent years, technological advancements have introduced machine learning (*ML*) techniques for modelling *ET* and *T* (En-Nia et al., 2025; Li et al., 2025a). The *ML* techniques can model *ET* and *T* owing to their ability to accurately capture complex non-linear relationships between input and target variables (Hailegnaw et al., 2025). For instance, support vector machine (SVM), extreme gradient boosting (XGBoost), single-layer artificial neural network (ANN) and deep neural network (DNN) were tested on modelling daily maize transpiration with good accuracy using meteorological, soil water content and crop leaf area index input variables (Fan et al., 2021). In another study, four *ML* models, RF, SVM, XGBoost and backpropagation neural network (BP) were used to model daily crop *ET*, using 3, 6, 9-factor input combinations with substantial accuracy across several climatic zones (Du et al., 2024). There is potential for further improvement of the modelling methodology by coupling *ML* models with physical models, thereby creating a hybrid model (He et al., 2024; Hu et al., 2021). This is because *ML* models and physically based models are complementary rather than being treated as two independent scientific paradigms (Reichstein et al., 2019). Several studies have reported the successful application of hybrid models (Bennett and Nijssen, 2021; He et al., 2023, 2022; Koppa et al., 2022; Shang et al., 2023). For example, Shang et al. (2023) developed two hybrid models to estimate regional *ET* on the Tibetan Plateau. To the best of our knowledge, the comparison of the accuracy of gradient boosting frameworks in modelling *ET* and *T* have not yet been investigated on the economically important subtropical crops in South Africa. Furthermore, a hybrid model has not yet been developed to estimate yield response factors (K_y) of subtropical crops in South Africa. To improve knowledge on the subtropical crop water-use and yield characteristics, the objectives of the current study are: (1).To model ET for banana, grapefruit, litchi, mango and sugarcane crops using light gradient boosting machine (LightGBM), random forest (RF) and extreme gradient boosting (XGBoost) models, to gap fill the eddy covariance ET measurements which spanned only a few months per crop species per season.(2).To use the best performing model to derive highly accurate crop coefficients.(3).To model T for grapefruit, litchi and mango crops using light gradient boosting machine (LightGBM), random forest (RF), extreme gradient boosting (XGBoost) models and selecting the best performing T model for yield modelling. (4).To formulate a hybrid water – yield model for deriving the grapefruit, litchi, and mango yield response factors by combining the empirical water-yield relation with the best machine learning T model.

(1) To model *ET* for banana, grapefruit, litchi, mango and sugarcane crops using light gradient boosting machine (LightGBM), random forest (RF) and extreme gradient boosting (XGBoost) models, to gap fill the eddy covariance ET measurements which spanned only a few months per crop species per season.
(2) To use the best performing model to derive highly accurate crop coefficients.
(3) To model *T* for grapefruit, litchi and mango crops using light gradient boosting machine (LightGBM), random forest (RF), extreme gradient boosting (XGBoost) models and selecting the best performing *T* model for yield modelling.
(4) To formulate a hybrid water – yield model for deriving the grapefruit, litchi, and mango yield response factors by combining the empirical water-yield relation with the best machine learning *T* model.

The derived information will provide accurate information on crop water use and potential yield dynamics, hence allowing the farmers to plan irrigation schedules that increase crop water productivity.

2. Materials and methods

2.1. Study site and plant material

The studies were conducted at Riverside farm, Malelane (25.447924°S; 31.5547226°E and 144 m above sea level) and Welgelegen farm, Komatipoort (25.50909°S; 31.94411°E and 173 m above sea level) in the Mpumalanga province of South Africa (Fig. 1). The sites were approximately 50 km apart, with the Riverside farm bordering the Kruger National Park in Malelane and the Welgelegen farm located close to the South Africa-Mozambique border. The sites were situated within the summer rainfall region of South Africa, which receives most rain in the months of November to March (Mangani et al., 2025; Ubisi et al., 2020). Five crops were used in this study which comprised of four fruit

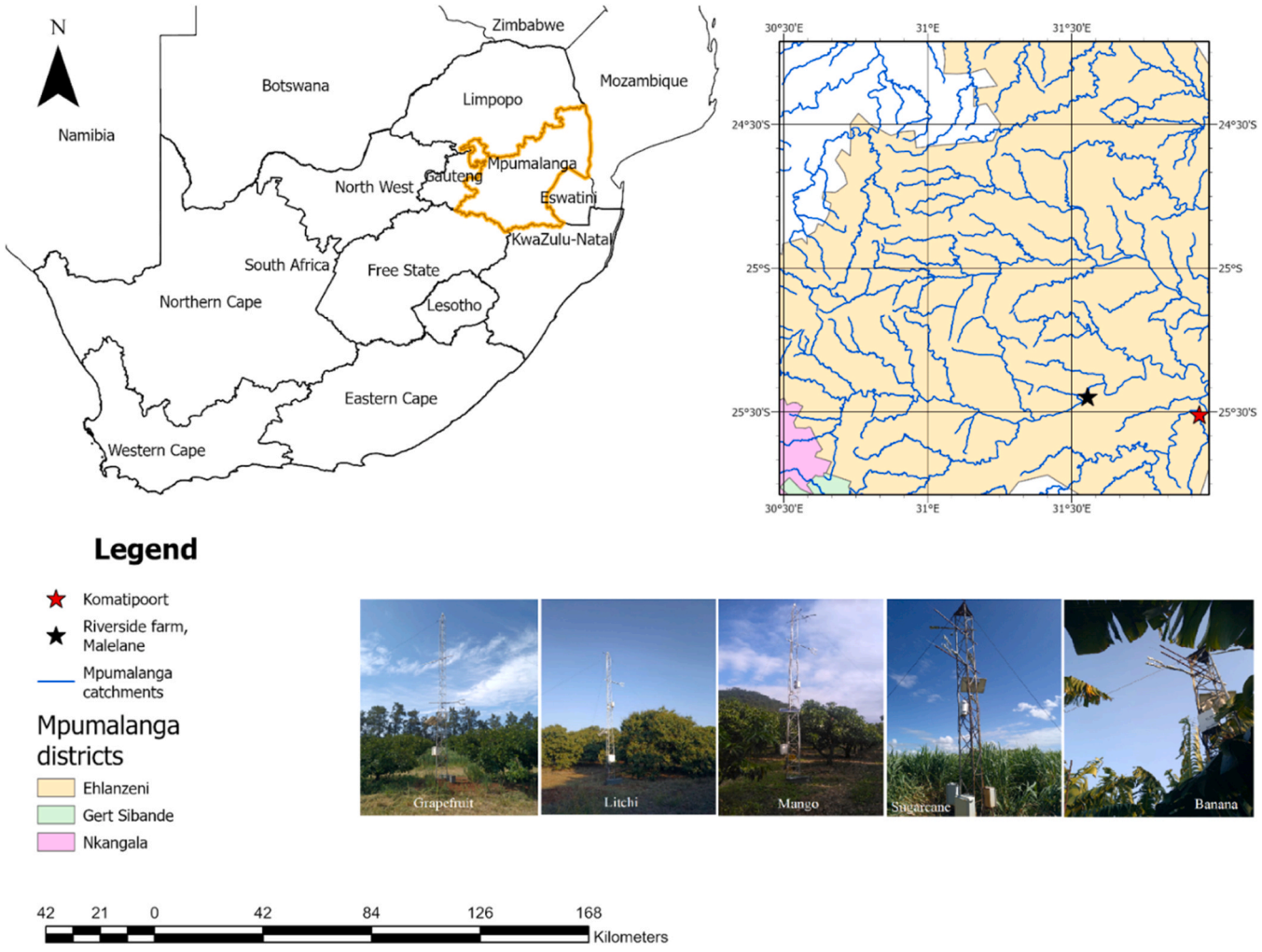


Fig. 1. Study sites in Malelane and Komatipoort, Mpumalanga, South Africa.

(banana, grapefruit, litchi and mango) and sugarcane. The grapefruit, litchi, mango and sugarcane study sites were located at Riverside farm, Malelane while the banana study site at Welgelegen farm, Komatipoort, Mpumalanga. A summary of the plant material and type of irrigation methods used in this study are shown in Table 1. Irrigation for the crops at both sites (Riverside and Welgelegen farms) were scheduled using the DFM (Dirk Friedhelm Mercker, Westen Cape, South Africa) soil moisture sensors. The study sites were selected because they are part of major subtropical fruit producers which contribute substantially to the South African agricultural export and local markets. Furthermore, the sites were situated in a semi-arid zone which requires irrigation water supplementation for growing the fruit crops (Camacho Suarez et al., 2015; Dangare et al., 2024; Feliciano et al., 2025; Nel et al., 2024).

2.2. Data

2.2.1. Microclimate

The microclimate data were measured at Riverside farm, Malelane and Welgelegen farm, Komatipoort using automatic weather stations which were mounted close to the orchards. The weather stations were installed on an open space covered with short grass and all meteorological data were measured at a height of 2 m above the ground. The monitored meteorological parameters included solar radiation, air temperature, relative humidity, rainfall as well as windspeed and direction. The solar radiation was measured using digital thermopile pyranometer (CS320, Campbell Scientific, USA), air temperature and relative humidity were measured using the digital temperature and relative humidity probe (CS215, Campbell Scientific, USA), rainfall was measured using an electronic tipping bucket rain gauge (TE25, Campbell Scientific, USA), windspeed and direction were measured using a digital sonic anemometer (ATMOS-22, METER Group, USA). The

Table 1
Plant material and irrigation method.

Crop type	Cultivar	Rootstock	Age (yrs)	Orchard size (ha)	Number of plants per hectare	Plant height (m)	Irrigation system
Banana	Williams	-	2	3.82	2222	4.0	Microsprinkler
Grapefruit	Star Ruby	C35/5 × 5B	14	6.25	476	2.9	Microsprinkler
Litchi	Mauritius	Mauritius	53	13.10	70	6.3	Microsprinkler
Mango	Tommy Atkins	-	39	9.50	303	4.7	Microsprinkler
Sugarcane	N53	-	3	8.54	87	2.8	Drip

meteorological sensors were connected to a CR1000 datalogger (Campbell Scientific, USA) which was programmed to collect measurements every 5 s and process hourly outputs. Historical hourly weather data at the Riverside farm, Malelane site from the year 2014–2021was provided by the Agricultural Research Council, South Africa (Moeletsi et al., 2022). Average weather data coverage at each study site was approximately 98 % with missing data caused by instruments failure.

2.2.2. Leaf area index (LAI)

The *LAI* for the banana, grapefruit, litchi, mango and sugarcane sites was determined using the fractional vegetation cover (F_c). The F_c was calculated using the Normalized Difference Vegetation Index (*NDVI*), which was derived using Landsat 8 Tier 1 surface reflectance data (U.S. Geological Survey). The Landsat 8 surface reflectance data were downloaded in Google Earth Engine (GEE) and images of more than 20 % cloud cover were filtered out. The 20 % cloud mask was applied to minimize the influence of cloud cover on the *NDVI* output. The *NDVI* was calculated using the red and near infrared bands of the Landsat 8 surface reflectance data using Eq. 1:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

where *NIR* is the near infrared band, and *R* is the red band of the Landsat 8 satellite (Tucker, 1979; Yilgan et al., 2025). The Landsat 8 satellite data were selected for computing *NDVI* in the current study because it provided satellite coverage since 2013, which was essential for our crop yield modelling.

The F_c was calculated using Eq. 2:

$$F_c = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \tag{2}$$

Where F_c is the factional vegetation cover, *NDVI* is the normalized difference vegetation index, $NDVI_{min}$ is the minimum value of *NDVI* and $NDVI_{max}$ is the maximum value of *NDVI* (Gao et al., 2020; Wu et al., 2025; Zhang et al., 2006). The values of $NDVI_{min}$ and $NDVI_{max}$ were determined from the Landsat 8 datasets for the five sites (Table 2).

The *LAI* (m^2/m^2) was calculated using Eq. 3:

$$LAI = \frac{-\ln(1 - F_c)}{k} \tag{3}$$

Where *k* is the extinction coefficient. A value of 0.5 was used for *k* in the current study by assuming leaf with random distribution (Li et al., 2010; Norman et al., 2003, 1995). We interpolated the Landsat 8 derived *LAI* data to hourly data using the cubic spline interpolation method (Liu et al., 2020).This was done to synchronize the meteorological and remote sensing data to a consistent temporal scale.

2.2.3. Transpiration measurements

Transpiration in the grapefruit, litchi and mango orchards was measured on four trees of different stem sizes per orchard using the heat ratio method (HRM) sap flow method (Burgess et al., 2001). A metal drilling template with three holes arranged in line and 5 mm apart was used to guide the drill bit to make holes in the tree stems. The drilling template was used to minimize errors due to HRM sap flow probes misalignment (Dzikiti et al., 2018; Nel et al., 2024). Each instrumented tree was implanted with four pairs of T-type thermocouples and four heaters, which were arranged in the north, south, east and west directions around the stem to measure the radial sap velocity of the stem. Pairs of T-type thermocouples were installed upstream and downstream of the central heater probes which were inserted inside brass sleeves to facilitate effective heat transfer from the heat source.

For each instrumented tree, the T-type thermocouple pairs were inserted at depths of 10, 20, 30, and 40 mm, respectively below the tree bark. The thermocouples and heaters were connected to the tree boxes which were controlled by CR1000 dataloggers (Campbell Scientific, USA) through the AM16/32B multiplexers (Campbell Scientific, USA). The CR1000 (Campbell Scientific, USA) data loggers were programmed to control hourly sap flow measurements. The measured sap velocity was corrected for tree wounding due to hole drilling and sap flow probes insertion using the method proposed by Swanson and Whitfield (1981). The conducting sapwood for grapefruit and mango was established by injecting a weak solution of methylene blue dye into the tree stems. The process was followed by collecting a stained wood sample above the dye injection points using a stem corer. The conducting sapwood for the litchi was determined by visual inspection of the changes in the colour of a slice of wood that was cut from the tree stem. The method was adopted because the colour of the litchi wood obscured a clear identification of several dyes staining. Sap flow volume (L/tree/h) for each instrumented tree was calculated as a weighted sum of the products of the corrected tree sap velocity at each thermocouple probe depth and the sapwood area represented by the respective probe (Dzikiti et al., 2018, 2017). The transpiration for the grapefruit, litchi, mango orchards in equivalent water depth units (mm/h) was calculated from the hourly sap flow volumes by using the number of trees per hectare in each stem diameter class divided by one hectare area (Nel et al., 2024; Ntshidi et al., 2021).

2.2.4. Evapotranspiration measurements

The evapotranspiration for grapefruit, litchi and mango orchards was measured on selected days (Table 3) using an eddy covariance (EC) flux tower. The flux tower was equipped with an open path eddy covariance system which comprised an IRGASON (Campbell Scientific, USA) which was connected to an EC100 (Campbell Scientific, USA) control system. The IRGASON comprised of a 3-dimensional sonic anemometer for measuring 3-dimensional components of windspeed and an infrared gas analyzer (IRGA) for measuring the concentration of water vapour and carbon dioxide in the atmosphere at a frequency of 10 Hz.

The IRGASON was mounted at a height of 1.5 m above the tree canopies facing in the direction of the prevailing wind, giving a flux footprint of approximately 150 m around the EC flux tower. The tower was equipped with additional sensors which included a net radiometer (NR-LITE2, Kipp & Zonen, Netherlands), two soil heat flux plates (HFP01, Hukseflux, Netherlands), two soil moisture sensors (CS616, Campbell Scientific, USA), temperature and relative humidity sensor (EE181, Campbell Scientific, USA), and averaging soil thermocouple probe (TCAV, Campbell Scientific, USA). The NR-LITE2 was mounted at a height of about 9.6 m to measure the orchard net radiation, HFP01 sensors were installed at a depth of about 8 cm to measure the soil heat flux, soil moisture content was measured above the HFP01 sensors using

Table 2
Normalized Difference Vegetation Index (*NDVI*) maximum, mean, minimum and standard deviation values for banana, grapefruit, litchi, mango and sugarcane fields.

	Banana	Grapefruit	Litchi	Mango	Sugarcane
Minimum *NDVI*	0.36	0.32	0.19	0.30	0.24
Maximum *NDVI*	0.91	0.64	0.82	0.80	0.60
Mean *NDVI*	0.83	0.53	0.59	0.63	0.50
NDVI Standard deviation	0.14	0.04	0.11	0.06	0.07

Table 3
Orchard evapotranspiration measurement campaigns.

Orchard	Evapotranspiration measurement campaign period
Banana	4 May 2022–30 September 2023
Grapefruit	6–9 December 2022; 7–21 March 2023
Litchi	16–31 July 2022; 20 September 2022–18 October 2022; 14 November 2022–1 December 2022
Mango	3–28 March 2022; 11 April 2022–30 May 2022; 26 June 2022–3 July 2022
Sugarcane	14 October 2023–1 February 2024

CS616 sensors, and TCAV was installed at depths of 2 and 6 cm for correcting the measured soil heat fluxes. All high and low frequency data were stored onto a 2 GB secure digital (SD) card which was connected to a CR3000 datalogger (Campbell Scientific, USA). The EC fluxes were collected and corrected for coordinate rotations, spikes, among others using the EasyFluxDL CR3000P software (Campbell Scientific, USA) which was installed on the CR3000 datalogger (Bambach et al., 2022; Dzikiti et al., 2018; Gush and Taylor, 2014; Mashabatu et al., 2025; Pagano et al., 2023). Average Eddy Covariance data coverage at each study site was approximately 94 %, with missing data caused by instruments failure. The flux data for all the sites were filtered using the quality flags as described by Foken et al. (2012). Data of good quality (flags 0–4) according to the Foken et al. (2012) quality flagging system were used for further analysis, and the remaining data were discarded (Maluleke et al., 2025). Linear interpolation was used to fill data gaps shorter than 2 h (Li et al., 2025b), and longer data gaps were filled using the REddyProc method (Wutzler et al., 2018). The eddy covariance data were corrected for energy imbalance using the Bowen ratio method (Twine et al., 2000).

Evapotranspiration for the banana and sugarcane fields were measured using the surface renewal (SR) method (Paw U et al., 1995; Snyder et al., 1996). The *ET* for banana field was measured from May 2022 to September 2023, while that of sugarcane was measured from October 2023 to February 2024 (Table 3). A 6 m tall SR flux tower was installed at the centre of the field, and this was equipped with two type-E fine-wire thermocouples (Campbell Scientific, USA) which were mounted on extension arms above the crop canopy. Additionally, the SR tower was equipped with a 4-component net radiometer (CNR4, Kipp & Zonen, Netherlands) which was extended above the crop to measure the field net radiation. In the banana field, two type-E fine-wire thermocouples were installed facing the direction of the prevailing wind at heights of 5.15 and 5.65 m, respectively above the ground. In the sugarcane field, the type-E fine-wire thermocouples were installed facing the direction of the prevailing wind at heights of 4 and 4.5 m respectively above the ground. The maximum crop height for banana and sugarcane is shown in Table 3. The height for the two crops was measured every month using a height gauge and the missing height values were interpolated using the cubic spline. The fine-wire thermocouples were installed to make high frequency air temperature measurements above the crop canopy at different heights and time lags, thereby facilitating the accurate determination of the best parameters for calculating field sensible heat flux. Measurements of the air temperature were conducted at a frequency of 10 Hz and time lags of 0.4 s and 0.8 s, respectively. The high and low frequency data measured at intervals of 100 ms and 30 min respectively, were stored onto a 2 SD card which was connected to a programmed CR1000 datalogger (Campbell Scientific, USA). An open path EC system which comprised of an IRGASON (Campbell Scientific, USA) was mounted on the SR tower at the height of the lower fine-wire thermocouple for three days from 30 January to 1 February 2024 owing to equipment shortages. This was done to obtain direct measurements of the field sensible heat flux which were used to calibrate the SR measured sensible heat flux. The calibration process involved plotting a regression graph between the EC sensible heat flux and SR sensible heat flux, then using the slope of the graph as a calibration factor (α) to correct the SR sensible heat flux measurements. The field *ET* was calculated as a residual of the energy balance equation, assuming that the soil heat flux was 10 % of the field net radiation (Masanganise et al., 2022; Meyers and Hollinger, 2004) as these data were not measured at this site.

2.3. Machine learning (*ML*) models

We applied three machine learning models namely light gradient boosting machine, random forest and extreme gradient boosting to model *ET* and *T* in the current study. These tree-based machine learning models were selected because they possess excellent computational efficiency and reliable accuracy (Zhao et al., 2024). Description of the *ML* models are as follows:

2.3.1. Light gradient boosting machine (LightGBM)

The LightGBM is a powerful gradient-boosting framework that utilizes tree-based learning algorithms in machine learning (Alsulamy, 2025; Daniel, 2024). The LightGBM has outstanding computational efficiency and reduces memory usage (Ke et al., 2017). Furthermore, the model integrates Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) (Alsulamy, 2025; Yaseen and Alhalimi, 2025). During the data sampling process, the GOSS technique retains samples with greater gradient values and randomly samples those that possess smaller gradient values. The approach reduces computational load, while ensuring the model training accuracy.

The EFB technique minimizes the complexity of the model training by reducing the dimensionality of the feature space through bundling of mutual exclusive features. Additionally, the LightGBM employs a histogram-based decision tree algorithm to obtain optimal splitting points, thereby allowing it to efficiently process large datasets (Ma et al., 2025). The LightGBM model optimizes the following objective function to update the model:

$$\mathscr{L} = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{N} \Omega(f_k) \tag{4}$$

Where $\mathscr{L}$ is the objective function, $\ell(y_i, \hat{y}_i)$ is the loss function, $\hat{y}_i$ is the prediction for instance i, $\Omega(f_k)$ is the regularization term for the k-th tree f_k, n is the number of instances, N is the number of trees, and y_i is the actual value for the i-th observation (Alsulamy, 2025; Ismail et al., 2024; Ma et al., 2025).

2.3.2. Random forest (RF)

Random Forest (RF) is a sophisticated ensemble learning algorithm that is based on decision trees (Breiman, 2001; Imani et al., 2025; Wang et al., 2020). The RF model applies the Bootstrap Aggregating (bagging) technique to combine multiple decision trees to enhance predictive performance of the model (Kouloumpris and Vlahavas, 2025; Yang et al., 2025). The RF possesses powerful interpretability, avoids overfitting problems, is less prone to data noise and outliers, among others (Yu et al., 2025). The final prediction of a RF model is given by:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{5}$$

Where $\bar{y}$ is the mean of the output of the individual trees used, $T_i(x)$ are the results from each tree, and N is the number of samples (Imani et al., 2025; Yaseen and Alhalimi, 2025).

2.3.3. Extreme gradient boosting (XGBoost)

The Extreme Gradient Boosting model is a tree-based and machine learning algorithm (Chen and Guestrin, 2016; EL Bilali et al., 2025). The XGBoost constructs an ensemble of weak learners (decision trees) sequentially, with each tree rectifying the errors of its predecessor. The process creates a strong predictive model for both regression and classification tasks (Aldosary et al., 2025; Lavaei et al., 2025). Additionally, the XGBoost prevents overfitting during model training by pruning trees and it controls model complexity via limiting the tree depth (Alsulamy, 2025). It also applies advanced techniques such as parallelization and handles missing data values, rendering it suitable for large and complex datasets (Fayzullo et al., 2025; Lavaei et al., 2025; Niazkar et al., 2024). The XGBoost algorithm is an iterative technique that aims to sequentially minimize the objective function given in Eq. 6:

$$\mathscr{L}(\Theta) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{j=1}^{N} \Omega\left(f_j\right) \tag{6}$$

Where $\mathscr{L}(\Theta)$ is the objective function, $\ell(y_i, \hat{y}_i)$ is the loss function, $\hat{y}_i$ is

the prediction for instance i, $\Omega\left(f_j\right)$ is the regularization term for the j-th tree f_j, n is the number of instances, N is the number of trees, and y_i is the actual value for the i-th observation (Amiri-Ramsheh et al., 2025; Nagro, 2025; Qin et al., 2025; Sadig et al., 2025; Tolun et al., 2025; Zhang et al., 2025).

2.4. Model development

2.4.1. Evapotranspiration, transpiration, crop coefficient and yield modelling

We proposed three *ML* models (LightGBM, Random Forest, and XGBoost) for estimating *ET* and *T* for banana, grapefruit, litchi, mango and sugarcane (Fig. 2). The best performing *ML* models for *ET* were used to derive the crop coefficients for the five crops (banana, grapefruit, litchi, mango and sugarcane). Hybrid models for modelling grapefruit, litchi and mango yield were created by combining the empirical water-yield equation (Giménez et al., 2016; Paredes et al., 2014), with the best performing *ML* models for *T* (Fig. 2).

The yield modelling for banana and sugarcane was excluded from the study because their yield data were not available. The *T* was used for yield modelling because it is directly connected to the crop yield development (Sinclair et al., 1984). The set of input variables for *ML* models (Fig. 2) were chosen because they defined crucial factors in the *ET* and *T* dynamics.

The hourly *ET* and *T* for the crops were modelled using Eqs. 7 and 8 respectively,

$$ET \sim ML\left(LAI,\ R_s,\ T_{\min},\ T_{avg},\ T_{\max},\ RH_a,\ U_{avg},\ d_r\right) \tag{7}$$

$$T \sim ML\left(R_s,\ LAI,\ T_{avg},\ RH_a,\ U_{avg}\right) \tag{8}$$

Where *ML* are the LightGBM, RF, XGBoost machine learning models, $R_s(\mathrm{W}/m^2)$ is the solar radiation, *LAI* is the mean monthly leaf area index, $T_{\min}$ (°C)is minimum air temperature, T_{avg} (°C) is the mean air temperature, $T_{\max}$ (°C) is maximum air temperature, RH_a (%) is the mean relative humidity, U_{avg} (m/s) is the mean windspeed at 2 m height, and d_r is relative distance between the earth and the sun. The yield response factor (K_y) for each crop was estimated using the hybrid model in Eq. 9:

$$\left(1-\frac{Y_a}{Y_m}\right)=K_y\left(1-\frac{T_a}{T_c}\right) \tag{9}$$

Where Y_a is the actual seasonal yield (tonne/ha), Y_m is the seasonal maximum (expected) yield (tonne/ha), T_c is the seasonal maximum transpiration (mm), and T_a is the seasonal transpiration (mm). The K_y for each crop was calculated as the slope of the linear plot of $\left(1-\frac{Y_a}{Y_m}\right)$ against $\left(1-\frac{T_a}{T_c}\right)$, forced through the origin. The T_a was calculated as the seasonal total of the hourly *T* which was predicted using the best performing *ML* in Eq. 8. The Y_m and T_c values are shown in Table 6, and these were obtained from the 10-year historic yield and modelled T_a data from different orchards at the farm. To gain further insights on the modelled K_y to the uncertainties in transpiration, we performed sensitivity tests on the water – yield model (Eq. 9). The sensitivity test involved varying the T_a for each crop by ± 5 % while keeping the T_c, Y_a, Y_m fixed and observing the changes in K_y (Table 6).

The yield for each crop was estimated using Eq. 10:

$$Y_a = Y_m - \frac{Y_m K_y (T_c - T_a)}{T_c} \tag{10}$$

The crop coefficient, K_c was calculated using the equation:

$$K_c = \frac{ET}{ET_o} \tag{11}$$

Where ET_o (mm) is the evapotranspiration for well-irrigated short grass uniformly covering the whole soil surface that is used as a reference (Allen et al., 1998). The monthly ET_o was taken as the the cumulative total of the hourly ET_o, calculated using the automatic weather station data. ET_c (mm) is the crop evapotranspiration which was modelled using

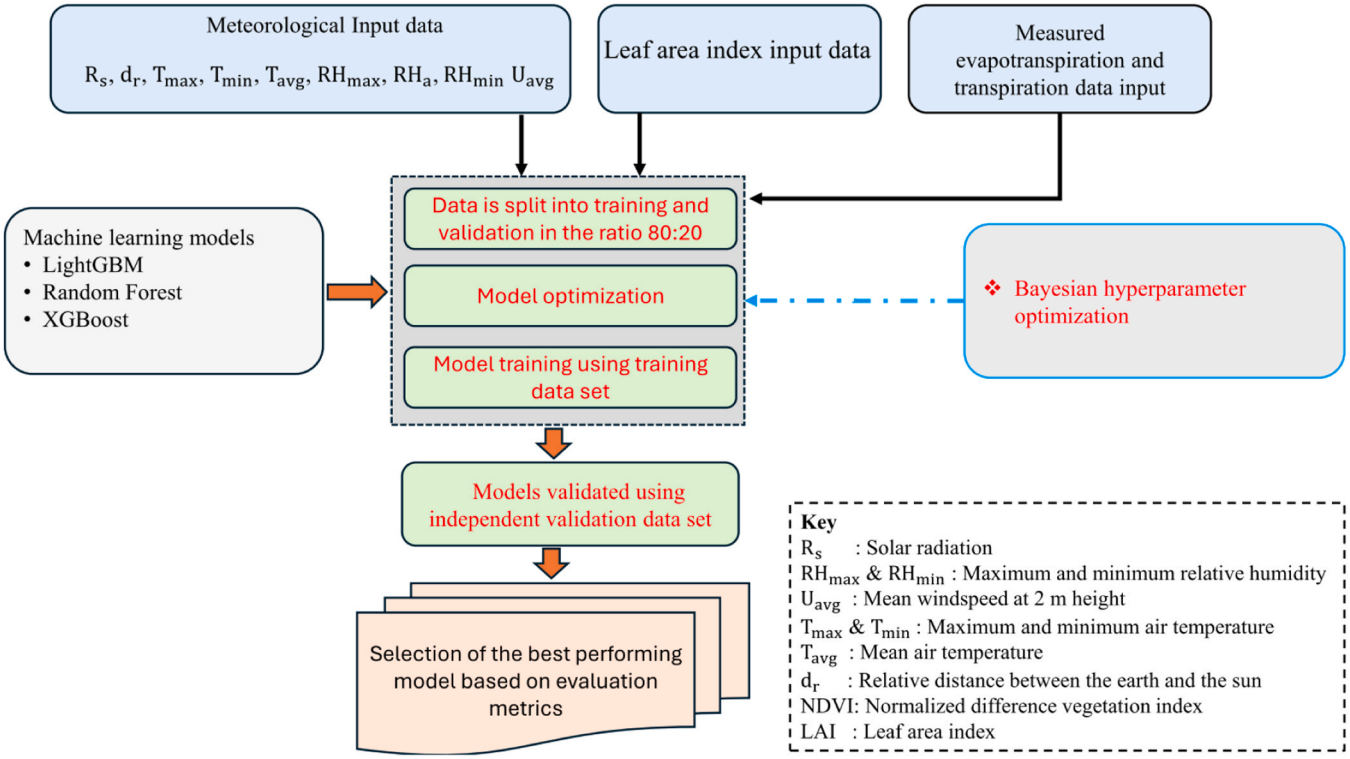


Fig. 2. Development framework for the orchard crop coefficients and crop yield response factors.

best performing *ML* model. Monthly ET_c was taken as the cumulative total of the modelled hourly ET_c.

We used LightGBM, RF and XGBoost machine learning models from the Python (v3.12.6) scikit-learn library to construct the *ET* and *T* models in PyCharm Community Edition (version 2025.1). The measured datasets for *ET* and *T* were divided into training and validation sets in the proportion 80:20 %, respectively. In the model development, 80 % of the data was assigned for model training to allow the models to learn reliable data patterns, and the remaining 20 % was assigned for estimating the model performance.

Additionally, a random state of 42 was applied to ensure that the model was dependable and reproducible (Yaseen and Alhalimi, 2025). To ensure reliability, all the *ML* models were optimized using the Bayesian hyperparameter optimization (Abbassi and Cheng, 2025; Chen et al., 2025). The Bayesian hyperparameter optimization was conducted on the *ML* models within the training datasets using the Optuna hyperparameter optimization framework (Akiba et al., 2019; Shen et al., 2025), in the scikit-learn machine learning library. We adopted the 10-fold cross validation (Pande et al., 2025), to randomly evaluate the performance of the *ML* models using the testing datasets (80 % of the measured data). An independent validation was performed on the developed *ML* models using the validation data (20 % of the measured data) to estimate the performance of the developed models on unseen data.

2.4.2. Model explanation: shapley additive exPlanations (SHAP) method

Ensemble *ML* models are complex models, which often operate as "black boxes" as it is difficult to understand how they reach their predictions (Pervez and Jamal, 2025). Various methods have been proposed to assist users interpret the complex models' predictions, and these include Shapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). The SHAP method is based on cooperative game theory, and it utilizes Shapley values to assess the significance of the *ML* model input features (Zhang et al., 2025). In the SHAP framework, the *ML* model input parameters are regarded as players, and the *ML* model outputs are regarded as the game. The Shapley value then determines the contributions of the players (*ML* model input parameters) to the game (*ML* model outputs) (Zheng et al., 2023). The SHAP value for model input feature *i* is calculated using the equation:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [k(S \cup \{i\} - k(S))] \quad (12)$$

Where ϕ_i is the SHAP value for feature *i*, *S* is a subset of features excluding *i*, *N* is the set of all the features, and $k(S)$ is the value function for the subset *S* (Nohara et al., 2022; Wang et al., 2025; Zheng et al., 2023). In the current study, we used the SHAP summary plots to investigate the importance ranking of the *ML* models input features and how they influence the predicted outputs.

2.4.3. Model evaluation metrics

The developed *ML* models were evaluated for predicting *ET* and *T* measurements, using the coefficient of determination, R^2 (Liao et al., 2025); root mean square error, *RMSE* (Karunasingha, 2022; Liu et al., 2025); mean absolute error, *MAE* (Ahmed et al., 2025), and Kling-Gupta efficiency, *KGE* (Gupta et al., 2009; Mohammadi et al., 2025). The R^2 has a range of 0–1, with values closer to 1 indicating a near perfect relationship and absence of statistical correlation between the measured and modelled data respectively (Barzegar et al., 2016). A *RMSE* value close to zero indicates a good model prediction (Alhakeem et al., 2022). The *MAE* has a range of 0 to infinity (∞), with a value of 0 showing a perfect relationship between the measured and modelled data (Duarte et al., 2019). The *KGE* has a range of $-\infty$ to 1, with a value of 1 showing a perfect relationship between measured and modelled data (Zhang et al., 2025). The R^2, *RMSE*, *MAE*, and *KGE* were calculated using Eqs. 13 - 16:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (13)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \quad (15)$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_p}{\sigma_o} - 1\right)^2 + \left(\frac{\mu_p}{\mu_o} - 1\right)^2} \quad (16)$$

Where y_i is the measured value, $\hat{y}_i$ is the modelled value, $\bar{y}$ is the mean of measured values, *n* is the number of observations, *r* is the Pearson correlation coefficient, σ_p is standard deviation of the predictions, μ_p is the mean value of the predictions, σ_o is standard deviation of the observations and μ_o is mean value of the observations. The overall model performance was determined by calculating the R^2, *RMSE*, *MAE* and *KGE* of each of the 10 test subsets. Finally, the reported evaluation metric was calculated as an average of the 10 subsets evaluation metrics to produce a good estimate of the model performance.

3. Results

3.1. Microclimate

The monthly summary of the weather conditions measured at the two study sites (Riverside farm, Malelane and Welgelegen farm, Komatipoort) between September 2021 and September 2023 is shown in Fig. 3. The monthly trend of the average solar radiation produced maximum and minimum values in the summer and winter months, respectively. The maximum monthly average solar radiation was recorded in February 2022 with a value of 21 MJ/m^2/d and a minimum value of 11.6 21 MJ/m^2/d occurring in May 2023. The monthly average maximum and minimum temperature patterns followed that of the solar radiation. The monthly average maximum temperature ranged between 33.9 and 24.6 °C, while minimum temperature ranged between 20.7 and 11.3 °C, respectively in the summer and winter seasons. The study sites received rain throughout the study period except in the month of August 2023, with maximum rainfall recorded in the summer seasons. A value of 461 mm was recorded in February 2023, and this was exceptionally higher than the value of 203 mm recorded in January 2022. The difference was mainly caused by a single storm that occurred in February 2023, causing widespread flooding.

3.2. Orchard leaf area index (*LAI*)

The Landsat 8 derived *NDVI* was used to calculate the monthly average orchard *LAI* from January 2015 to December 2024, and the *LAI* used for the *ML* models' development is shown in Fig. 4. A summary of the banana, grapefruit, litchi, mango and sugarcane field maximum *NDVI*, minimum *NDVI*, mean *NDVI* and *NDVI* standard deviation is shown in Table 2. Monthly average field *LAI* ranged between 0.09 and 6.6 m^2/m^2 for banana, 1.3 and 4.3 m^2/m^2 for grapefruit, 1.9 and 5.3 m^2/m^2 for litchi, 2.0 and 4.7 m^2/m^2 for mango, 0.2 and 4.3 m^2/m^2 for sugarcane. The monthly average *LAI* was used to model both the field *ET* and tree *T*. The minimum monthly average *LAI* value for banana in September – October 2021 coincided with the crop planting period. The minimum monthly average *LAI* values for sugarcane in June 2022 and July 2023 coincided with the field harvesting periods.

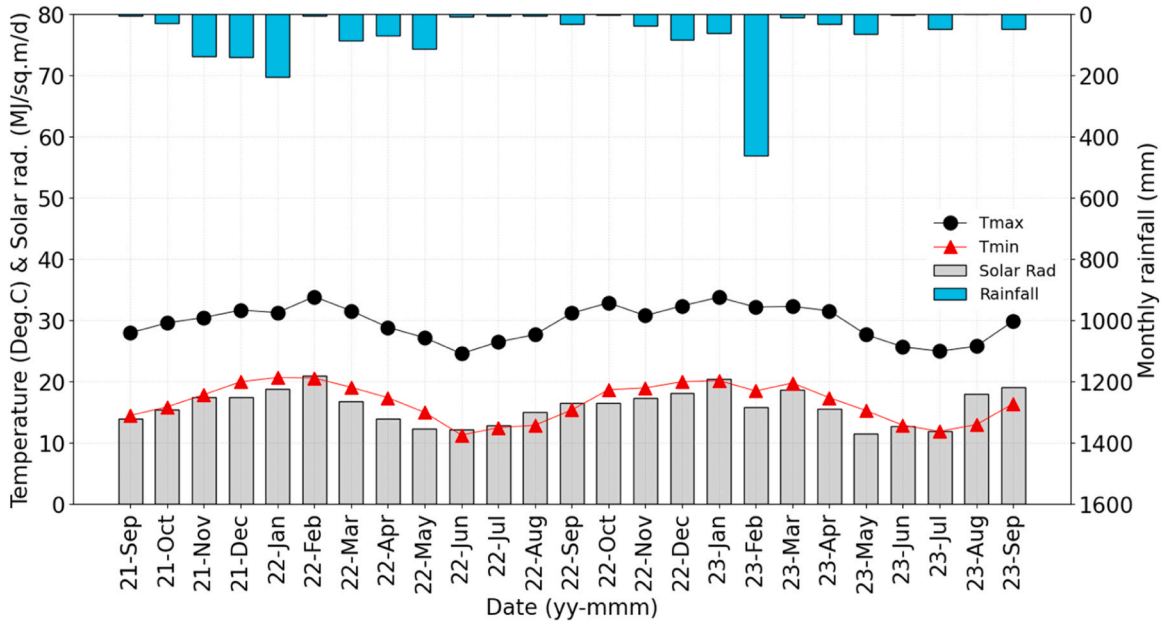


Fig. 3. Weather conditions at experimental sites from September 2021 to September 2023 showing the monthly average solar radiation, monthly average maximum and minimum air temperature, and monthly total rainfall.

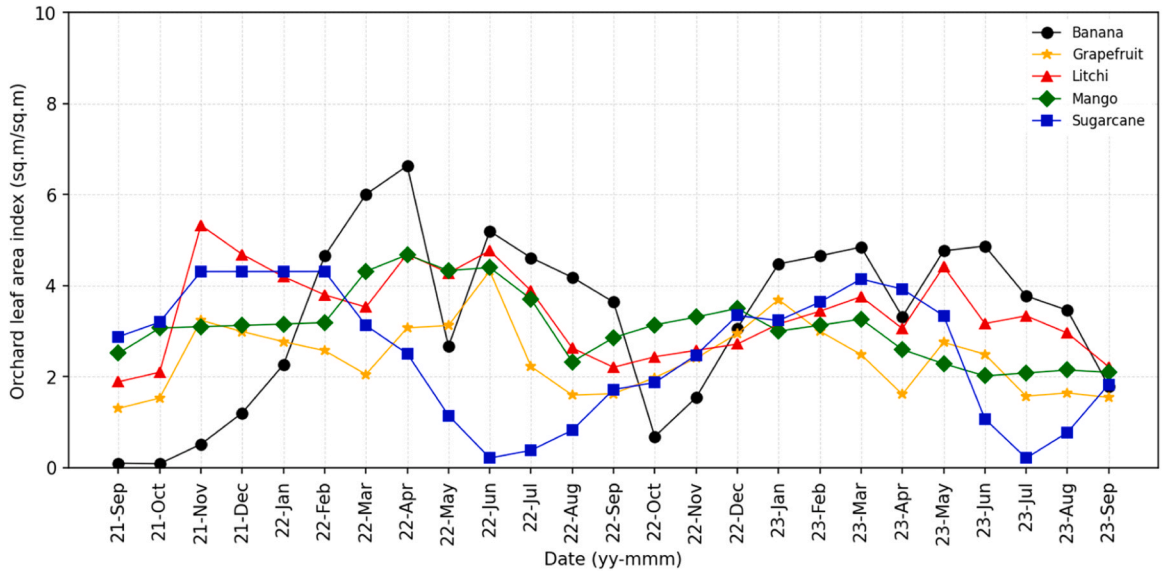


Fig. 4. Monthly average orchard leaf area index for banana, grapefruit, litchi, mango, sugarcane, calculated using Normalized Difference Vegetation Index (*NDVI*) derived from Landsat 8 data.

3.3. Comparison of the evapotranspiration and transpiration models

To select the best models for predicting *ET* and *T*, three *ML* models namely LightGBM, RF and XGBoost were investigated. The best performing models were selected based on statistical metrics between the measured and modelled data. Five LightGBM, RF, and XGBoost *ET* models respectively, were trained using the measured *ET* data for banana, grapefruit, litchi, mango, and sugarcane. Additionally, three LightGBM, RF, and XGBoost *T* models respectively, were trained using the measured *T* data for grapefruit, litchi and mango trees. Bayesian optimization was applied to each model to search for optimal parameters, and the resulting hyperparameters are shown in Table 7 – 12 (supplementary material). To evaluate the developed *ET* and *T* models' robustness, we conducted a 10-fold cross-validation on the training datasets for the crops. The 10-fold cross-validation involved dividing the training datasets into 10 subsets. At each moment the model was trained using 9 subsets and the remaining one was used for model testing. The independent validation was conducted on the developed models using the validation datasets to evaluate the performance of the models on unseen data. The calculated metrics for all the *ET* and *T* models are shown in Table 4 and Table 5, respectively. The scatter plots for the relationship between measured and modelled *ET* and *T* for the training and validation datasets are shown in Fig. 5, Fig. 6 and Fig. 11–14 (supplementary material).

In the model validation, the LightGBM and XGBoost banana evapotranspiration models produced similar model performance statistics, and they both outperformed the RF banana evapotranspiration model. The LightGBM and XGBoost banana evapotranspiration models produced an R^2 of 0.90, *RMSE* of 0.08 mm/h, *MAE* of 0.04 mm/h, and *KGE* of 0.93 when compared to the measured banana *ET* data (Fig. 5f and

Table 4
Evaluation metrics for the LightGBM, Random Forest and XGBoost evapotranspiration models.

Models	Cross validation				Independent validation			
	R^2	*RMSE*	*MAE*	*KGE*	R^2	*RMSE*	*MAE*	*KGE*
Banana LightGBM	0.91	0.07	0.04	0.94	0.90	0.08	0.04	0.93
Grapefruit LightGBM	0.93	0.03	0.02	0.93	0.93	0.03	0.02	0.93
Litchi LightGBM	0.92	0.03	0.02	0.93	0.89	0.03	0.02	0.93
Mango LightGBM	0.88	0.04	0.02	0.91	0.89	0.04	0.02	0.91
Sugarcane LightGBM	0.85	0.09	0.05	0.88	0.83	0.10	0.06	0.88
Banana Random Forest	0.91	0.07	0.03	0.93	0.90	0.08	0.04	0.92
Grapefruit Random Forest	0.93	0.04	0.02	0.92	0.92	0.04	0.02	0.91
Litchi Random Forest	0.89	0.03	0.02	0.86	0.87	0.03	0.02	0.87
Mango Random Forest	0.87	0.04	0.03	0.87	0.89	0.04	0.03	0.87
Sugarcane Random Forest	0.84	0.10	0.05	0.85	0.80	0.11	0.06	0.84
Banana XGBoost	0.91	0.07	0.04	0.94	0.90	0.08	0.04	0.93
Grapefruit XGBoost	0.90	0.04	0.03	0.90	0.91	0.04	0.03	0.89
Litchi XGBoost	0.92	0.03	0.02	0.93	0.89	0.03	0.02	0.93
Mango XGBoost	0.89	0.04	0.02	0.91	0.90	0.04	0.02	0.91
Sugarcane XGBoost	0.86	0.10	0.05	0.88	0.83	0.10	0.06	0.88

Table 5
Evaluation metrics for the LightGBM, Random Forest and XGBoost transpiration models.

Models	Cross validation				Independent validation			
	R^2	*RMSE*	*MAE*	*KGE*	R^2	*RMSE*	*MAE*	*KGE*
Grapefruit LightGBM	0.95	0.02	0.01	0.96	0.95	0.02	0.01	0.97
Litchi LightGBM	0.90	0.02	0.01	0.92	0.90	0.02	0.01	0.93
Mango LightGBM	0.96	0.02	0.01	0.97	0.96	0.02	0.01	0.97
Grapefruit Random Forest	0.95	0.02	0.01	0.95	0.95	0.02	0.01	0.95
Litchi Random Forest	0.89	0.02	0.01	0.91	0.89	0.02	0.01	0.91
Mango Random Forest	0.96	0.02	0.01	0.96	0.95	0.02	0.01	0.96
Grapefruit XGBoost	0.95	0.02	0.01	0.96	0.95	0.02	0.01	0.96
Litchi XGBoost	0.90	0.02	0.01	0.92	0.90	0.02	0.01	0.92
Mango XGBoost	0.96	0.02	0.01	0.97	0.96	0.02	0.01	0.97

Fig. 13(f8) in supplementary material).

The grapefruit LightGBM evapotranspiration model outperformed both the grapefruit RF and XGBoost evapotranspiration models, respectively producing an R^2 of 0.93, *RMSE* of 0.03 mm/h, *MAE* of 0.02 mm/h, and *KGE* of 0.93 when compared to the measured grapefruit *ET* data (Fig. 5g). The litchi LightGBM and XGBoost evapotranspiration models produced similar model performance statistics and outperformed the litchi RF evapotranspiration model. The LightGBM and XGBoost evapotranspiration models produced an R^2 of 0.89, *RMSE* of 0.03 mm/h, *MAE* of 0.02 mm/h, and *KGE* of 0.93 when compared to the measured litchi *ET* data (Fig. 5h and Fig. 13(h8) in supplementary material). For mango crop, the XGBoost evapotranspiration model outperformed both the mango LightGBM and RF evapotranspiration models, producing an R^2 of 0.90, *RMSE* of 0.04 mm/h, *MAE* of 0.02 mm/h, and *KGE* of 0.91 when compared to the measured mango *ET* data (Fig. 13(i8) in supplementary material). The sugarcane LightGBM and XGBoost models produced similar model performance statistics and outperformed the sugarcane RF evapotranspiration model. The sugarcane LightGBM and XGBoost models produced an R^2 of 0.83, *RMSE* of 0.10 mm/h, *MAE* of 0.06 mm/h, and *KGE* of 0.88 when compared to the measured sugarcane *ET* data (Fig. 5j and Fig. 13(j8) in supplementary material). Consequently, LightGBM was selected for modelling the banana, grapefruit, litchi and sugarcane crop coefficients because of good performance, computational speed and light memory usage. The XGBoost was used for modelling the mango crop coefficients because it outperformed the LightGBM and RF evapotranspiration models.

In the validation of the grapefruit *T* model, the LightGBM outperformed both the grapefruit RF and XGBoost transpiration models, producing an R^2 of 0.95, *RMSE* of 0.02 mm/h, *MAE* of 0.01 mm/h, and *KGE* of 0.97 when compared to the measured grapefruit *T* data (Fig. 6 (d1)). The litchi LightGBM transpiration model outperformed the litchi RF and XGBoost transpiration models, producing an R^2 of 0.90, *RMSE* of 0.02 mm/h, *MAE* of 0.01 mm/h, and *KGE* of 0.93 when compared to the measured litchi *T* data (Fig. 6(e1)). The mango LightGBM and XGBoost transpiration models produced similar model performance statistics and outperformed the mango RF transpiration model. The mango LightGBM and XGBoost transpiration models produced an R^2 of 0.96, *RMSE* of 0.02 mm/h, *MAE* of 0.01 mm/h, and *KGE* of 0.97 when compared to the measured mango *T* data (Fig. 6(f1) and Fig. 14(f9) in supplementary material). Therefore the LightGBM transpiration model was selected to model T_a in Eqs. 9 and 10 owing to its good performance, computational speed and low memory usage.

3.4. Machine learning model interpretability

In the current study, we investigated the contributions of the input features to the selected banana, grapefruit, litchi, mango, and sugarcane *ET* models output using SHAP heat maps (Fig. 7(a2) - (e2)). The horizontal axis of the SHAP heatmap indicate instance samples, vertical axis indicates the model inputs, and the SHAP values are encoded on a colour scale. The deeper and lighter colours denote stronger and weaker impacts of input features on model output, respectively. The output of the model is shown above the heatmap matrix, and the global importance of each model input is shown as a bar plot on the right-hand side of the heatmap plot. We used the layered violin summary plots to investigate the significance of model inputs and establish their effects on modelled *T* (Fig. 8(a3) - (c3)). The violin summary plot provides a concise representation of the distribution and variability of SHAP values for each model input. The magnitude of the SHAP value is displayed on the horizontal axis, with negative and positive values indicating a negative and positive influence on the modelled output, respectively.

The banana, grapefruit, litchi, mango, and sugarcane *ET* models heatmaps indicated that SHAP values for the input features produced alternating colours indicating variable contributions to the model

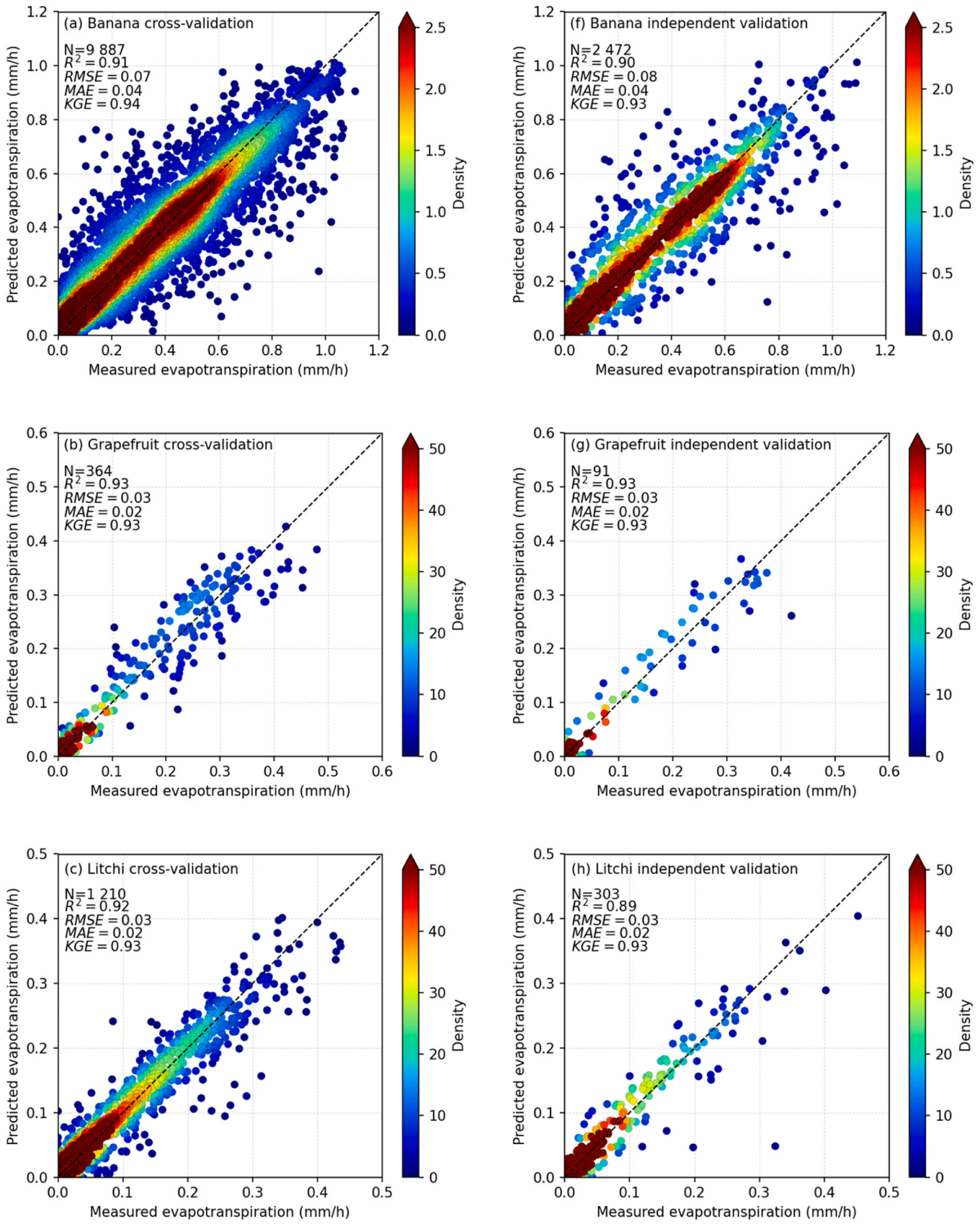


Fig. 5. Performance of the LightGBM learning model in predicting evapotranspiration using 10-fold cross-validation for (a) banana, (b) grapefruit, (c) litchi, (d) mango, (e) sugarcane, and the corresponding model independent validation for (f) banana, (g) grapefruit, (h) litchi, (i) mango, (j) sugarcane, respectively.

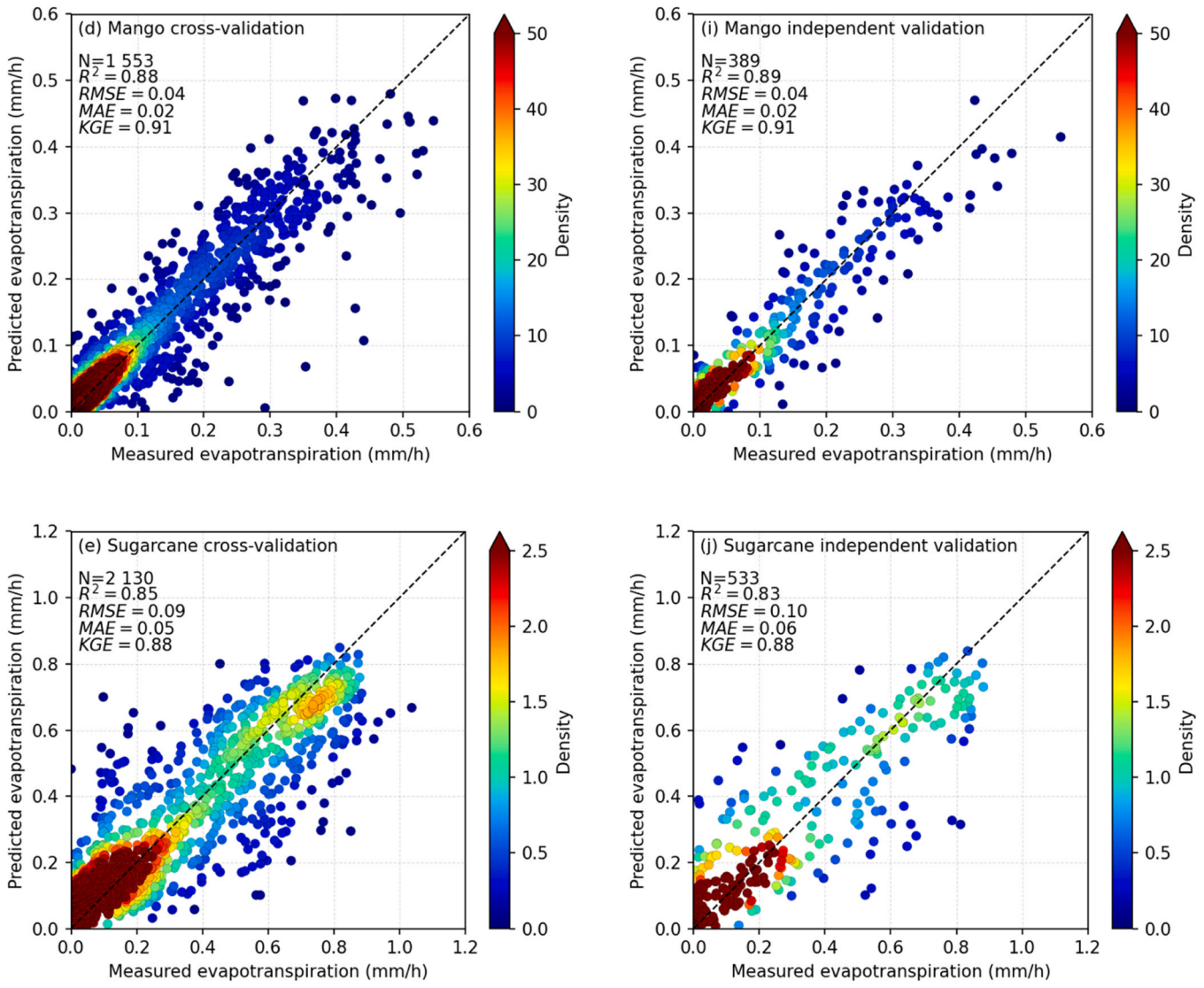


Fig. 5. (*continued*).

output. In all the five crop *ET* models, R_s produced the deepest red and blue colours which coincided with the high and low model outputs, respectively. The deep red and blue colours implied a strong positive and negative effect on the *ET* model output, respectively (Fig. 7(a2) - (e2)). The bar plots displaying the mean absolute SHAP values for each *ET* model input are shown in Fig. 7(f2) – (i2). These bar plots provided a ranking of the model input features, ordered by the size of the calculated mean absolute SHAP values. The input feature with the largest mean absolute SHAP value had the greatest contribution to the model prediction. The bar plots indicated that the R_s had the largest influence on the five *ET* model outputs as compared to the other model inputs. Additionally, the other model inputs () varied in importance to the *ET* model output between the crops (Fig. 7(f2) - (i2)). The violin summary and bar SHAP plots for grapefruit, litchi and mango *T* modelling are shown in Fig. 8(a3) - (f3). The R_s had the highest ranking on the list of *T* model inputs for the 3 crops (grapefruit, litchi, and mango), with high and low R_s values producing a positive and negative effect on the modelled output respectively (Fig. 8(a3) - (c3)). This implied that R_s had the largest influence on the modelled output as compared to other model inputs (LAI, T_{avg}, RH_a, U_{avg}). Additionally, the influence of the other predictors (LAI, T_{avg}, RH_a, U_{avg}) on the modelled *T* varied among the three crops (Fig. 8(d3) - (f3)).

3.5. Evapotranspiration and crop coefficients

The *ET* for banana, grapefruit, litchi, sugarcane were modelled using the LightGBM evapotranspiration models, and the mango *ET* was modelled using the XGBoost evapotranspiration model following the outcome of the performance of *ML* models in predicting crop *ET*. The modelled monthly *ET* for banana, grapefruit, litchi, mango, and sugarcane followed the reference evapotranspiration trend between September 2021 and September 2023 (Fig. 9(a4) - (e4)), implying that the crops had a positive response to the changes in the atmospheric evaporative demand. The mean annual *ET* for banana, grapefruit, litchi, mango and sugarcane was approximately 1524 mm, 824 mm, 639 mm, 960 mm, and 1137 mm, respectively. Among the five crops, banana had the largest cumulative *ET*, followed by sugarcane, mango, grapefruit, and then litchi. The derived monthly crop coefficients for banana, grapefruit, litchi, mango, and sugarcane are shown in Fig. 9(f4) - (j4). The crop coefficient (K_c) for each crop was calculated by dividing the modelled *ET* with reference evapotranspiration (ET_o). The banana K_c varied throughout the study period, producing a maximum and minimum value of approximately 1.3 and 0.92, respectively (Fig. 9(f4)). The grapefruit K_c varied throughout the period September 2021 and September 2023, with K_c values ranging between 0.53 and 0.75 (Fig. 9 (g4)). The litchi K_c values ranged between 0.40 and 0.57 for the period

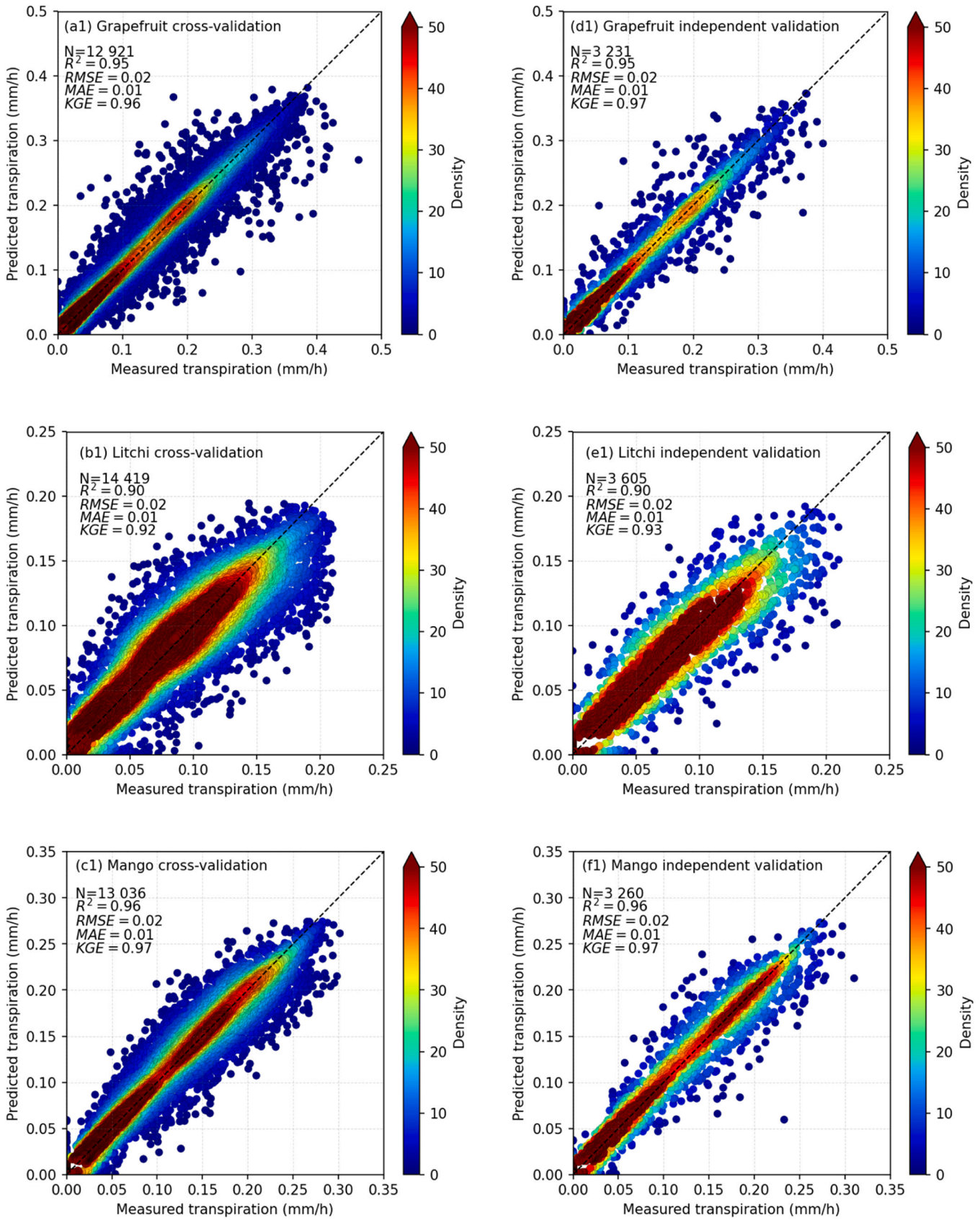


Fig. 6. Performance of the LightGBM learning model in predicting transpiration using 10-fold cross-validation for (a1) grapefruit, (b1) litchi, (c1) mango, and the corresponding model independent validation for (d1) grapefruit, (e1) litchi, (f1) mango, respectively.

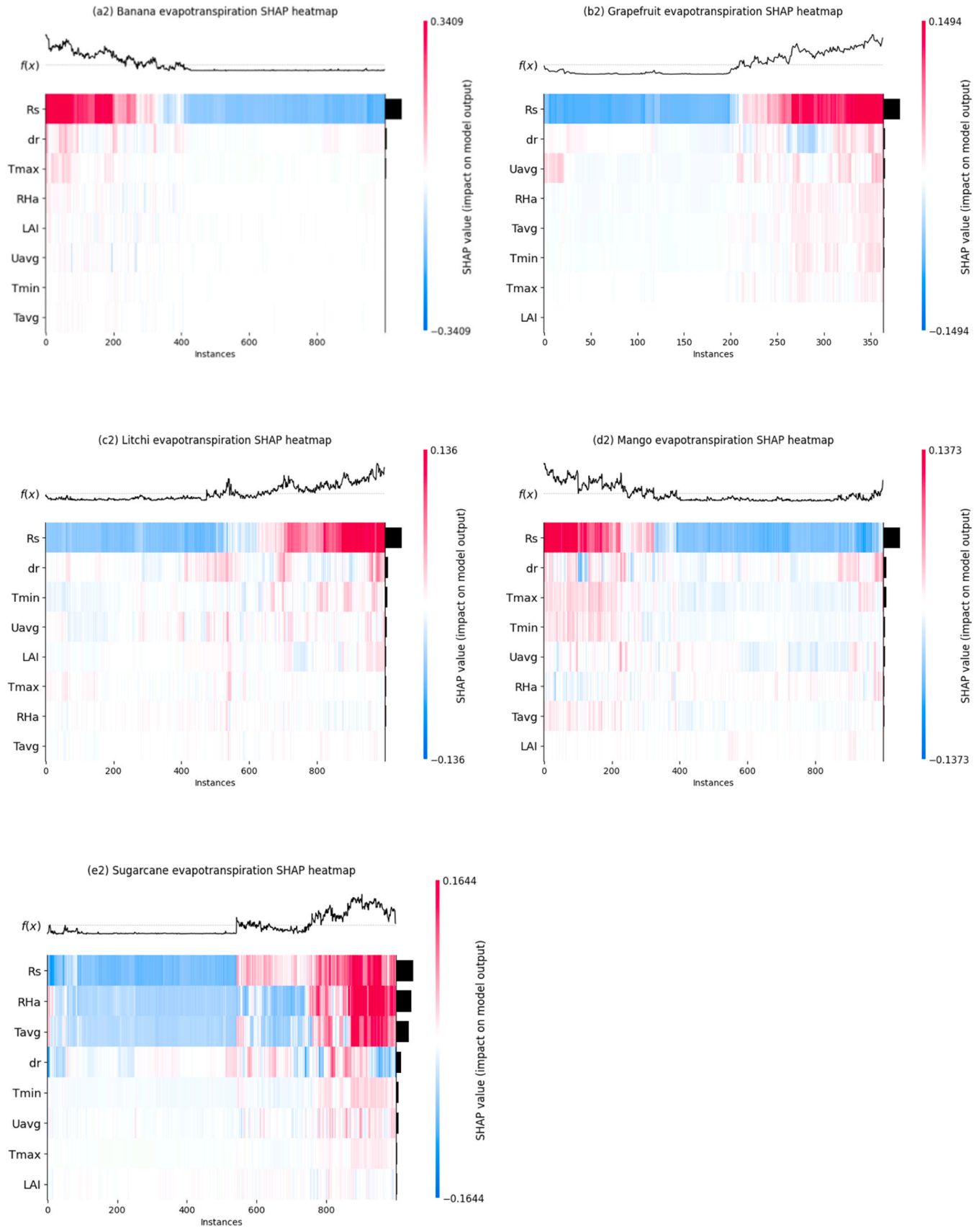


Fig. 7. SHAP heatmap and bar plot of the evapotranspiration prediction for banana, grapefruit, litchi, mango, and sugarcane.

September 2021 to September 2023 (Fig. 9(h4)). The mango K_c values ranged between 0.63 and 0.84 for the period September 2021 to

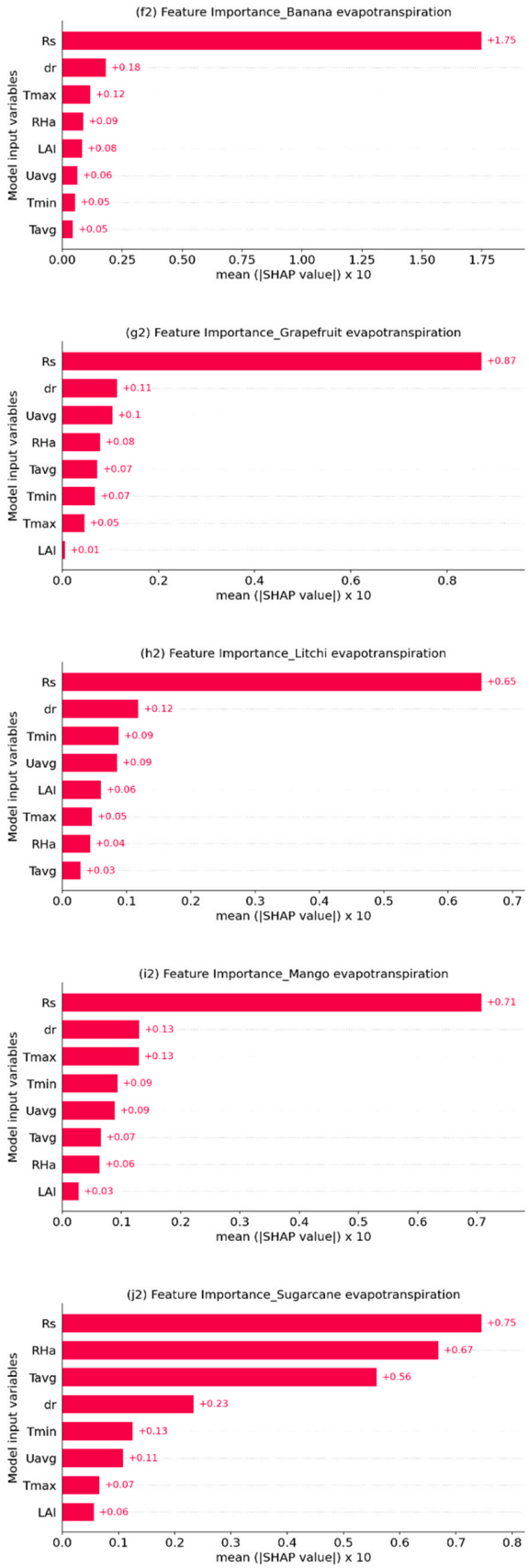


Fig. 7. (*continued*).

September 2023 (Fig. 9(i4)). The sugarcane K_c varied between 0.62 and 1.16 between September 2021 and September 2023 (Fig. 9(j4)).

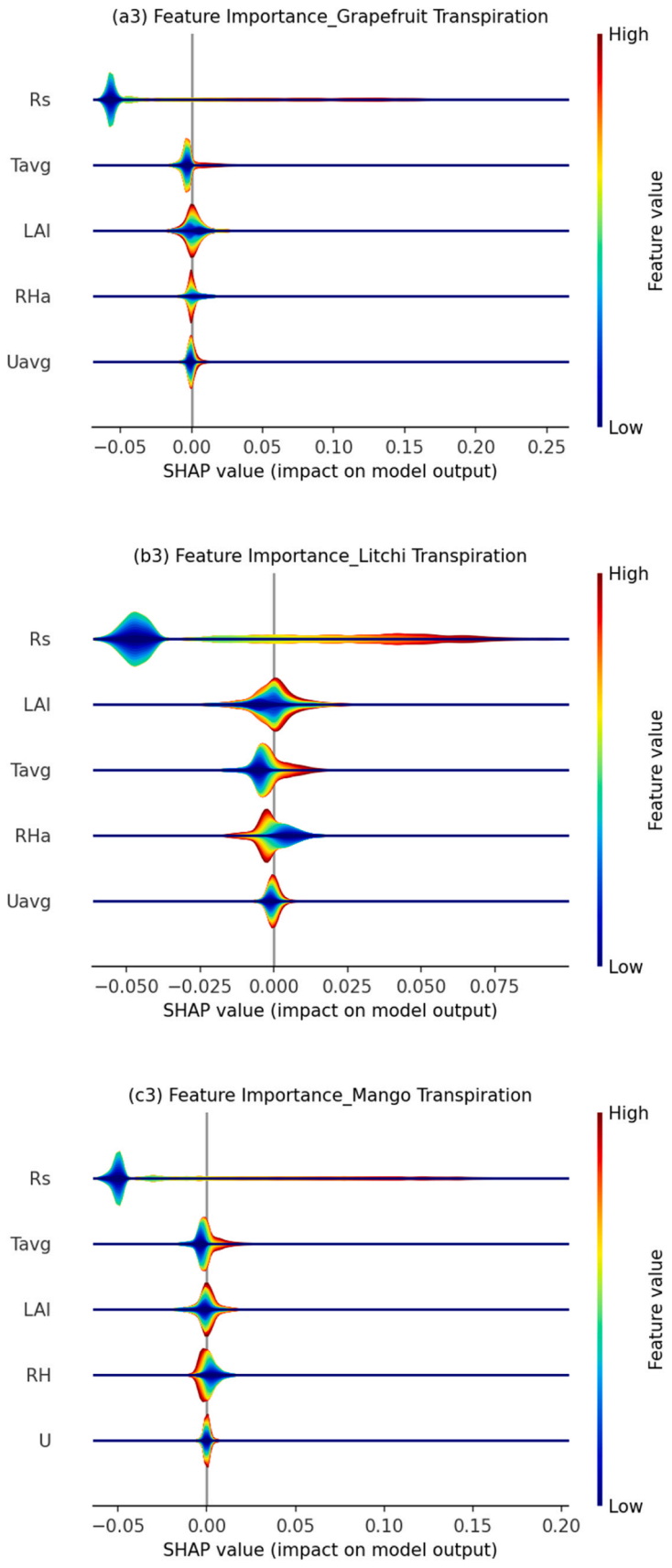
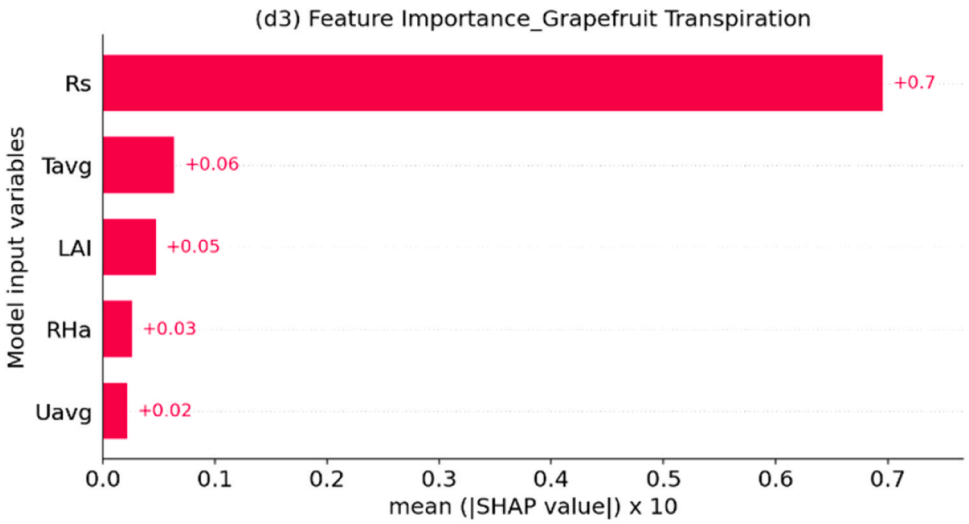
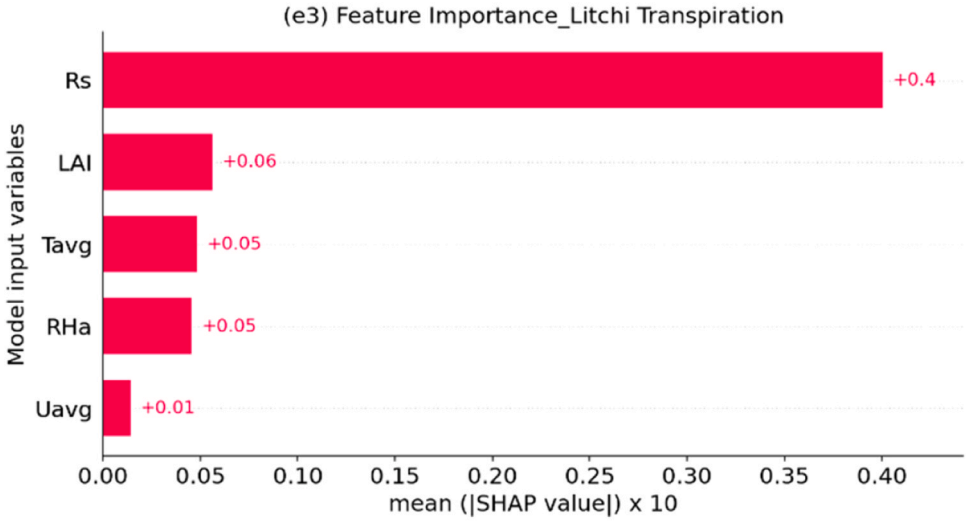


Fig. 8. SHAP summary and bar plot of the transpiration prediction for grapefruit, litchi, and mango.





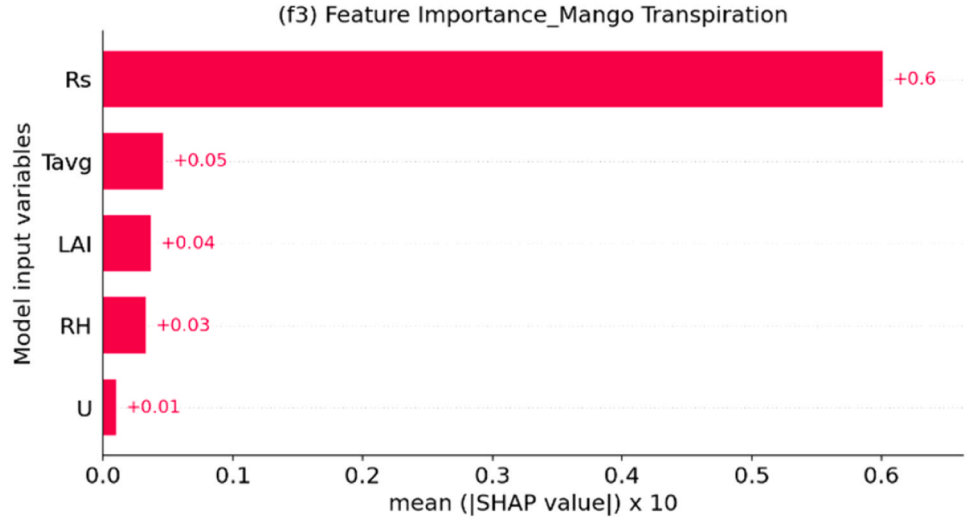


Fig. 8. (*continued*).

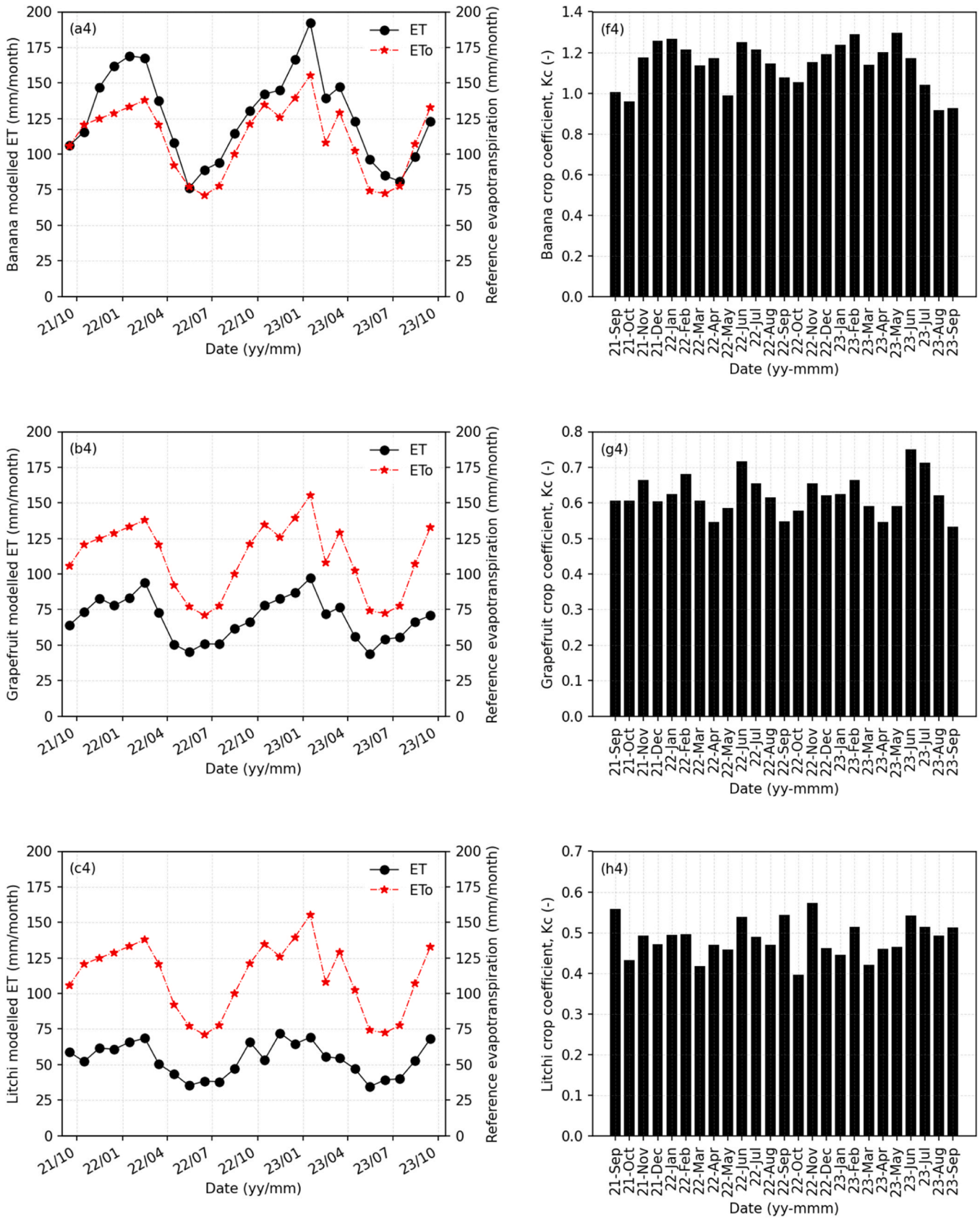


Fig. 9. The variation of the monthly total reference evapotranspiration and modelled evapotranspiration for (a4) banana, (b4) grapefruit, (c4) litchi, (d4) mango, (e4) sugarcane and the corresponding derived monthly crop coefficients for (f4) banana, (g4) grapefruit, (h4) litchi, (i4) mango, and (j4) sugarcane.

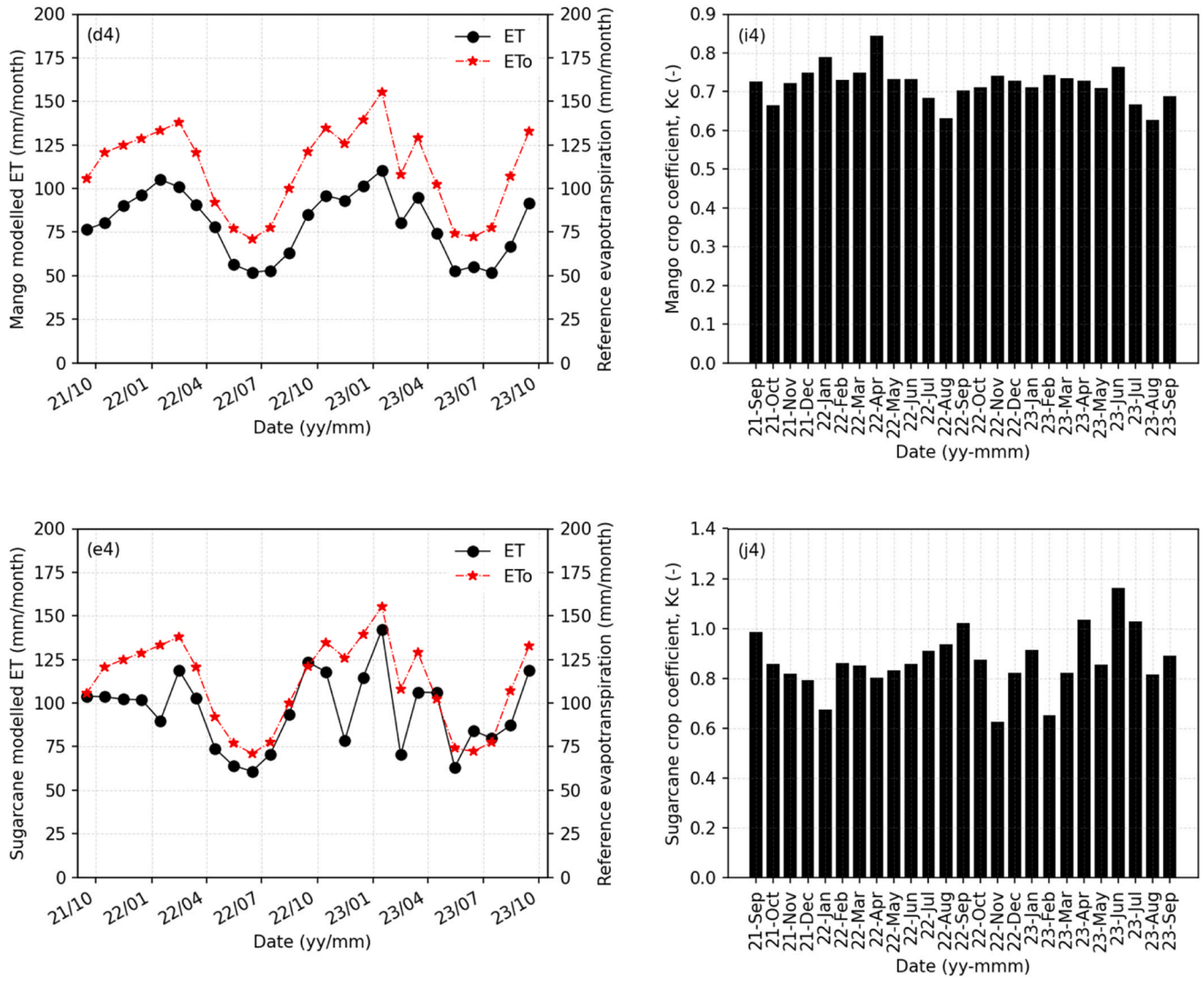


Fig. 9. (*continued*).

3.6. Crop yield response factor

The K_y was taken as the slope of the linear regression graph which was plotted using Eq. 9. The K_y was derived and validated using measured yield and modelled transpiration data from different orchards. The grapefruit, litchi and mango produced K_y values of 2.70, 2.50, and 2.90 respectively. The sensitivity test indicated that ±5 % change in grapefruit T_a, produced a 43 and 21 % change in grapefruit K_y respectively. A ± 5 % change in litchi T_a produced a 13 and 11 % change in litchi K_y, respectively. A ± 5 % change in mango T_a produced a 35 and 21 % change in mango K_y respectively. The derived K_y values are shown in Table 6. The developed K_y values excluding the sensitivity tests derived K_y values were used to model the crop yield (Eq. 10). The validation of yield modelled using grapefruit K_y produced an R^2 of 0.81, *RMSE* of 4.49 tonne/ha, *MAE* of 3.84 tonne/ha, and *KGE* of 0.74 when compared to the measured grapefruit yield data (Fig. 10(a5)). The validation of yield modelled using litchi K_y produced an R^2 of 0.87, *RMSE* of 0.78 tonne/ha, *MAE* of 0.59 tonne/ha, and *KGE* of 0.92 when compared to the measured litchi yield data (Fig. 10(b5)). The validation of yield modelled using mango K_y produced an R^2 of 0.83, *RMSE* of 2.79 tonne/ha, *MAE* of 2.23 tonne/ha, and *KGE* of 0.88 when compared to the measured mango yield data (Fig. 10(c5)). Thus, the derived K_y values for grapefruit, litchi and mango produced good results in modelling yield.

4. Discussion

Accurate and timely prediction of orchards *ET* is vital for productive

Table 6
Seasonal maximum yield (Y_m), seasonal maximum transpiration (T_c), seasonal transpiration (T_a) and the water – yield response factor (K_y) for the grapefruit, litchi and mango.

Crop	Y_m (tonnes/ha)	T_c (mm/yr)	K_y (*using* T_a)	K_y (*using* $1.05T_a$)	K_y (*using* $0.95T_a$)
Grapefruit	90	600	2.70	3.86	2.13
Litchi	30	500	2.50	2.87	2.26
Mango	55	550	2.90	3.91	2.30

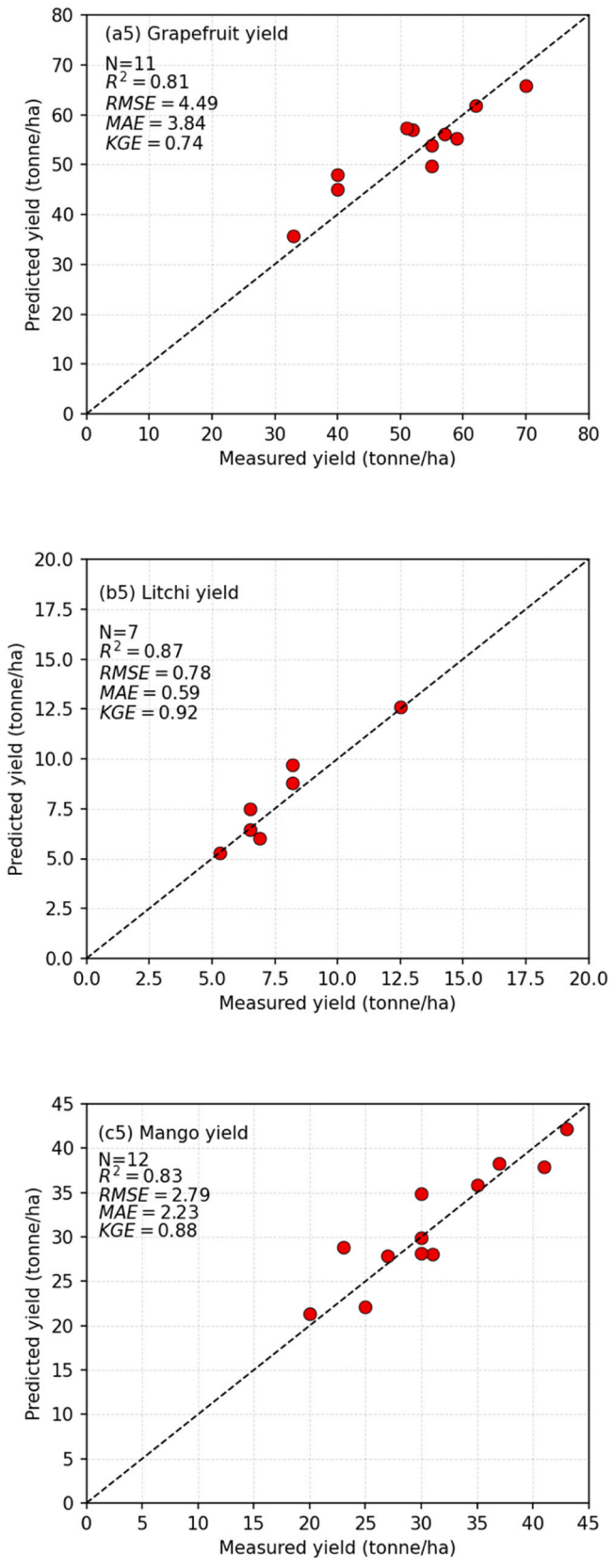


Fig. 10. Relationship between measured and modelled yield for (a5) grapefruit, (b5) litchi, (c5) mango.

irrigation management (Belarbi and El Younoussi, 2025; Rahman et al., 2025). Recent advancements in *ML* have provided improved *ET* predictive capabilities as compared to the traditional *ET* models, due to the *ML* capabilities in capturing intricate connections between various predictor variables and *ET* (Afridi et al., 2025; Dangare et al., 2025).

In the current study, the validation of the five *ET* models used for calculating the crop coefficients for banana, grapefruit, litchi, mango, and sugarcane produced an $R^2 \geq 0.83$, $RMSE \leq 0.10$ mm/h, $MAE \leq 0.06$ mm/h, and $KGE \geq 0.88$. Additionally, the LightGBM transpiration models used for yield modelling produced an an $R^2 \geq 0.90$, $RMSE = 0.02$ mm/h, $MAE = 0.01$ mm/h, and $KGE \geq 0.93$. The results obtained for *ET* and *T* machine learning modelling had the same order of accuracy as those observed in other studies. For instance, a study was conducted to predict *ET* in Bangladesh using four deep learning models namely, feed forward neural networks (FFNNs), convolutional neural networks (CNNs), gated recurrent units (GRUs), and long short-term memory networks (LSTMs). The modelled *ET* was compared to that *ET* derived from the Bangladesh Agricultural Research Council (BARC) data, producing an $R^2 \geq 0.80$ for both the model training and validation across the four deep learning models (Rahman et al., 2025). In the United States of America and Europe, six machine learning models namely artificial neural networks (ANN), extremely randomized trees (ExtraTrees), gradient boosting decision tree (GBDT), light gradient boosting machine (LightGBM), random forest (RF), and extreme gradient boosting (XGBoost) were used to model *ET* for 16 croplands. The modelled *ET* was compared against 20-year *ET* data that was measured using eddy covariance systems. The XGBoost evapotranspiration model outperformed the ANN, ExtraTrees, GBDT, LightGBM and RF models, producing correlation coefficient (*r*) of 0.88, *RMSE* of 6.87 W/m^2, *MAE* of 3.41 W/m^2, and Nash-Sutcliffe Efficiency (*NSE*) of 0.77 when compared to the measured *ET* (Zhang et al., 2025). In southern Italy, partial least squares (PLS), the generalized linear model (GLM), and three machine learning models namely, random forest (RF), elastic net (Enet) and support vector machine (SVM) were used to model actual *ET* of watermelon crops. The random forest model produced R^2 of 0.74, *RMSE* of 0.577 mm, and mean bias error (*MBE*) of 0.03 mm when compared to the *ET* measured by eddy covariance system (Garofalo et al., 2025). The CatBoost, RF, ExtraTrees, multi-layer perceptron neural network, and K-nearest neighbor (KNN) models were used to model daily actual *ET* of maize in Texas, United States of America. The models' output was compared to the *ET* measured using weighing lysimeters. The CatBoost model produced the best accuracy among the investigated models, giving R^2 of 0.96 and *RMSE* of 0.5594 mm/d when compared to the measured *ET* (Rahimi et al., 2025). In China, a study was conducted to analyse and compare the accuracy of estimating *ET* using the Long-Term Sequence Feature Space Method (LTSFSM) and six machine learning models. The machine learning models included the gradient boosting decision tree, random forest regression, partial least square regression, KNN, backpropagation neural network and support vector regression. The modelled *ET* was compared against the eddy covariance flux measurements and the random forest regression produced the highest accuracy among the investigated models with $R = 0.79$, $RMSE = 0.61$ mm/d, $MAE = 0.42$ mm/d, and $MBE = -0.02$ mm (Sun et al., 2025). In a study to estimate the seasonal crop water requirements for wheat in India, the support vector regression was compared against the crop *ET* estimated using the FAO-PM56 method. The support vector regression produced an accuracy of $R^2 = 0.9921$, $MAE = 0.11$ mm/d, $RMSE = 0.1439$ mm/d and mean squared error (*MSE*) = 0.02 mm/d (Jain and Gupta, 2025). In a study in Mount Gongga China, three machine learning models namely Random Forest, Multi-Layer Perceptron and Support Vector Regression were used to predict daily transpiration of a subalpine coniferous forest. The Support Vector Regression achieved the highest performance with $R^2 = 0.81$, $RMSE = 0.20$ mm/d, $MSE = 0.04$ mm/d and $MAE = 0.16$ mm/d (Xiang et al., 2025). In a study in Southern China, 8 machine learning models were used to

predict *ET* of evergreen forests. The eXtreme Gradient Boosting achieved the highest overall performance in modelling *ET* among the investigated models giving $R^2 = 0.84–0.98$ (Xing et al., 2025). Thus, the current study indicated that the use of *ML* models can successfully predict the *ET* and *T* for subtropical crops with acceptable accuracy. Furthermore, the relative simplicity and availability of input parameters used, strengthen the applicability of *ML* models in predicting *ET* and *T* for subtropical crops as compared to the traditional physical models.

Evapotranspiration and transpiration modelling involves various interacting factors that exhibit complex and non-linear relationships. The LightGBM is a gradient boosting framework that adapts well to capturing complex relationships in the input data and models with great accuracy (Li et al., 2025). On the other hand, each decision tree in RF predictions may excessively depend on certain input features, which leads to inconsistent predictions and lower accuracy (Yang et al., 2025). Thus, this may have led to LightGBM producing higher accuracy in modelling *ET* and *T* as compared to RF model. The XGBoost outperformed LightGBM and RF in modelling *ET*. This could be attributed to microclimate variability in the mango orchard as XGBoost is known to effectively reduce errors, learn complex data patterns and employs regularization techniques that improve model's generalizability (Chen and Guestrin, 2016).

The modelled *ET* for banana crop in the current study was 1524 mm and the calculated K_c values ranged between 0.92 and 1.30 (Fig. 9(f4)). The values obtained for both *ET* and K_c were higher than those obtained in the first crop of red banana grown in the semi-arid region of India. The study in India produced an *ET* of 1448 mm using soil water balance approach and a K_c ranging between 0.57 and 1.19 (Arunadevi et al., 2025). In the current study, the monthly litchi K_c varied between 0.47 and 0.57, lacking a seasonal trend Fig. 9(h4). The obtained K_c range were less than the range 0.4 – 1.2 obtained in the 10-year-old 'Mauritius' litchi orchard in South Africa. However, there was consistency in the lack of seasonal trend between the two studies (Menzel et al., 1995). The mean annual grapefruit *ET* in the current study was 824 mm, and this was higher than the values of 716.9 mm and 642 mm measured on a grapefruit orchard in Turkey using Bowen ratio-energy balance and eddy covariance measurements respectively (Ünlu et al., 2014). The differences in the grapefruit *ET* between the two studies were attributed to different orchard planting densities, where in Turkey the orchard had 156 trees/ha and the orchard in the current study had 476 trrees/ha. In the current study, the mean annual mango *ET* was 960 mm and the K_c values ranged between 0.63 and 0.84 (Fig. 9(i4)). In Brazil, a study on a 7-year-old mango orchard produced a maximum *ET* value of 555 mm over a period of 6 months (de Azevedo et al., 2003). Furthermore, the maximum K_c for mango obtained in the current study matched the mid-season K_c value of 0.84 that was reported for 'Tommy Atkins' mango in the study of Paredes et al. (2024). The mean annual *ET* for sugarcane in the current study was 1137 mm, and this was less than the *ET* obtained in a study in Ethiopia which obtained a mean annual *ET* range 1318 – 1682 mm. The study in Ethiopia used 4 surface balance models that utilized Landsat imagery and weather data to predict *ET*. The surface balance models included SEBAL (Surface Energy Balance Algorithm for Land), METRIC (Mapping Evapotranspiration with Internalized Calibration), SSEB (Simplified Surface Energy Balance), and SSEBop (Operational Simplified Surface Energy Balance) (Woldemariam et al., 2025).

The yield response factor (K_y) indicates a reduction in crop yield as a function of *ET* and *T*, and can be vital for yield prediction and irrigation management (Lovelli et al., 2007). The $K_y < 0.85$ indicates low sensitivity to water deficit, $0.85 < K_y < 1$ indicates a low to medium sensitivity, $1 < K_y < 1.15$ indicates a medium to high sensitivity and $K_y > 1.15$ indicates high sensitivity (Silva et al., 2019). The K_y obtained for grapefruit in the current study was 2.70 and this was higher than the K_y range 1.1–1.3 reported for citrus trees (Allen et al., 1998). Additionally, K_y obtained for litchi was 2.50 and that for mango, it was 2.90, showing a high sensitivity for the three crops to water deficits. The high K_y values obtained in the current study are consistent with findings in other studies. For example, in Brazil a study on 'Bragantina' black pepper reported K_y values ranging between 1.72 and 2.96 in response to four different irrigation treatments which were established using emitters with different flow rates (Carvalho et al., 2023). In Italy, a study reported a K_y of 1.37 for the eggplant (Lovelli et al., 2007). The sensitivity tests indicated that a small change in T_a has a significant impact on the developed K_y values. This implies that large uncertainties in the T_a predictions will cause large variations in the derived K_y values for grapefruit, litchi and mango crops.

The results indicate that *ET*, *T*, K_c, and K_y are variable between regions due to the differences in climate, soil nutrients, soil moisture status, crop cultivars, crop characteristics and irrigation management practices. Thus, the determination of representative crop *ET*, *T*, K_c, and K_y under local conditions is a necessity (Dzikiti et al., 2024; Gush and Taylor, 2014). The derived K_c and K_y in this study have a strong potential to improve irrigation planning and yield prediction of subtropical crops. This can be achieved by incorporating the derived coefficients into practical decision support systems for irrigation management, such as the SIMDualKc model (Rosa et al., 2012).

5. Conclusion

The main objectives of the current study were to investigate and select the best evapotranspiration and transpiration *ML* model, develop the crop coefficients for five crops using this model, and derive the yield response factors for three crops using a transpiration-yield hybrid model. The *T* for the crops were measured using the heat ratio method of monitoring sap flow and the *ET* was quantified using eddy covariance technique. Landsat 8 satellite was used to calculate the field leaf area index (*LAI*). Three evapotranspiration and transpiration *ML* models (LightGBM, Random Forest, XGBoost) were developed, trained and validated in Python (v3.12.6) using the scikit-learn machine learning library. All the models were optimized using Bayesian hyperparameter optimization, a 10-fold cross validation was conducted on the training data sets which comprised 80 % of the measured data to test the model robustness, and an independent validation was performed on the validation data sets which comprised 20 % of the measured data. The LightGBM achieved the highest accuracy in predicting banana, grapefruit, litchi and sugarcane *ET*, whereas XGBoost achieved this for mango. The LightGBM produced the best accuracy in predicting the grapefruit, litchi and mango *T*. The mean annual *ET* for banana, grapefruit, litchi, mango and sugarcane was approximately 1524 mm, 824 mm, 639 mm, 960 mm, and 1137 mm, respectively. The calculated banana K_c varied in the range 0.92–1.3, grapefruit K_c ranged between 0.53 and 0.75, litchi K_c ranged between 0.40 and 0.57, mango K_c ranged between 0.63 and 0.84, and sugarcane K_c varied between 0.62 and 1.16, respectively. The grapefruit, litchi and mango produced K_y values of 2.70, 2.50, and 2.90 respectively. The K_y values obtained in the current study indicated the grapefruit, litchi and mango have a high sensitivity to water deficits.

As a result of equipment shortages, limitations developed as the study was only done on two sites, in addition to only measuring crop *ET* partially during the season. This produced a small training data set for the *ET* machine learning models' and could lead in the developed models being site specific. Consequently, while the derived values are accurate and applicable to the current study, it is recommended that extensive validation be conducted using independent data before applying them in other regions.

CRediT authorship contribution statement

Cronje Paul J.R.: Writing – review & editing, Validation, Supervision, Resources. **Prince Dangare:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation,

Conceptualization. **Joseph Masanganise:** Writing – review & editing, Investigation. **Zama E. Mashimbye:** Writing – review & editing, Validation, Supervision, Resources, Methodology. **Shaeden Gokool:** Writing – review & editing, Investigation. **Zanele Ntshidi:** Writing – review & editing, Visualization, Investigation. **Tendai Sawunyama:** Funding acquisition. **Vivek Naiken:** Writing – review & editing, Investigation. **Sebinasi Dzikiti:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research reported here formed part of a bigger project entitled: "Developing a decision support system for estimating the water use and efficiency of irrigated crops in the Inkomati-Usuthu Water Management Area (WMA)"-Water Research Commission-Project WRC C2020/2023–00399. We acknowledge and express our gratitude to the Water Research Commission (WRC) and the Inkomati-Usuthu Catchment Management Agency in South Africa for funding and managing the project. We also thank Riverside Farm and Welgelegen Farm for allowing us to work in their productive fields. We gratefully acknowledge the farm managers Mr Dean van Heerden, Mr Jannie Small, Mr Jaco Henery, Mr Toekie Badenhorst, Mr Kyle Negrao, Mr Jannes van de Merwe, Mr Ferdie Thiele, Mr Gerbrand and sugarcane field supervisor Mr Bheki Lubisi for all the assistance.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.agwat.2025.110113.

Data availability

Data will be made available on request.

References

Abbassi, B., Cheng, L.-Z., 2025. Curvilinear lineament extraction: Bayesian optimization of principal component wavelet analysis and hysteresis thresholding. Comput. Geosci. 194, 105768. https://doi.org/10.1016/j.cageo.2024.105768.

Abdullah, S.S., Malek, M.A., 2016. Empirical penman-monteith equation and artificial intelligence techniques in predicting reference evapotranspiration: a review. Int. J. Water 10, 55. https://doi.org/10.1504/IJW.2016.073741.

Afridi, W.A.K., Mukhopadhyay, S.C., Mainali, B., 2025. Machine learning enabled smart sensor node for evapotranspiration prediction in diverse soil and environmental conditions. IEEE Sens J. 25, 22720–22728. https://doi.org/10.1109/JSEN.2025.3566935.

Ahmed, Y., Akter, T., Prima, M., Dutta, K.R., Mukut, S., Ahsan, M., Rahman, M.M., Mohammad Ziaul Hyder, M.K., 2025. Advanced ciprofloxacin quantification: a machine learning and metaheuristic approach using ultrasensitive chitosan-gold nanoparticle based electrochemical sensor. J. Environ. Chem. Eng. 13, 115094. https://doi.org/10.1016/j.jece.2024.115094.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, New York, NY, USA, pp. 2623–2631. https://doi.org/10.1145/3292500.3330701.

Aldosary, A.S., Al-Ramadan, B., Kafy, A., Al, Altuwaijri, H.A., Rahaman, Z.A., 2025. Forecasting climate risk and heat stress hazards in arid ecosystems: machine learning and ensemble models for specific humidity prediction in Dammam, Saudi Arabia. Nat. Hazards 121, 9281–9309. https://doi.org/10.1007/s11069-025-07140-3.

Alhakeem, Z.M., Jebur, Y.M., Henedy, S.N., Imran, H., Bernardo, L.F.A., Hussein, H.M., 2022. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques. Materials 15, 7432. https://doi.org/10.3390/ma15217432.

Allen, R.G., Pereira, L.S., 2009. Estimating crop coefficients from fraction of ground cover and height. Irrig. Sci. 28, 17–34. https://doi.org/10.1007/s00271-009-0182-z.

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: II. Recommended documentation. Agric. Water Manag. 98, 921–929. https://doi.org/10.1016/j.agwat.2010.12.016.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Guidelines for computing crop water requirements. In: FAO Irrigation and Drainage Paper No, 56. FAO, Rome, Italy.

Alsulamy, S., 2025. Predicting construction delay risks in Saudi Arabian projects: a comparative analysis of CatBoost, XGBoost, and LGBM. Expert Syst. Appl. 268, 126268. https://doi.org/10.1016/j.eswa.2024.126268.

Amiri-Ramsheh, B., Larestani, A., Atashrouz, S., Nasirzadeh, E., Essakhraoui, M., Abedi, A., Ostadhassan, M., Mohaddespour, A., Hemmati-Sarapardeh, A., 2025. Toward accurate prediction of carbon dioxide (CO2) compressibility factor using tree-based intelligent schemes (XGBoost and LightGBM) and equations of state. Results Eng. 25, 104035. https://doi.org/10.1016/j.rineng.2025.104035.

Arunadevi, K., Valliammai, A., Nagarajan, M., Sakthivel, N., 2025. Development of crop coefficient of drip irrigated banana for semi-arid region. Plant Sci. Today. https://doi.org/10.14719/pst.6704.

de Azevedo, P.V., da Silva, B.B., da Silva, V.P.R., 2003. Water requirements of irrigated mango orchards in northeast Brazil. Agric. Water Manag. 58, 241–254. https://doi.org/10.1016/S0378-3774(02)00083-5.

Baldocchi, D.D., Hincks, B.B., Meyers, T.P., 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. Ecology 69, 1331–1340. https://doi.org/10.2307/1941631.

Ballester, C., Castel, J., Testi, L., Intrigliolo, D.S., Castel, J.R., 2013. Can heat-pulse sap flow measurements be used as continuous water stress indicators of citrus trees? Irrig. Sci. 31, 1053–1063. https://doi.org/10.1007/s00271-012-0386-5.

Bambach, N., Kustas, W., Alfieri, J., Prueger, J., Hipps, L., McKee, L., Castro, S.J., Volk, J., Alsina, M.M., McElrone, A.J., 2022. Evapotranspiration uncertainty at micrometeorological scales: the impact of the eddy covariance energy imbalance and correction methods. Irrig. Sci. 40, 445–461. https://doi.org/10.1007/s00271-022-00783-1.

Barzegar, R., Adamowski, J., Moghaddam, A.A., 2016. Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. Stoch. Environ. Res. Risk Assess. 30, 1797–1819. https://doi.org/10.1007/s00477-016-1213-y.

Belarbi, Z., El Younoussi, Y., 2025. A review on optimizing water management in agriculture through smart irrigation systems and machine learning. E3S Web Conf. 601, 00078. https://doi.org/10.1051/e3sconf/202560100078.

Bennett, A., Nijssen, B., 2021. Deep learned process parameterizations provide better representations of turbulent heat fluxes in hydrologic models. Water Resour. Res. 57. https://doi.org/10.1029/2020WR029328.

Bethenod, O., Katerji, N., Goujet, R., Bertolini, J.M., Rana, G., 2000. Determination and validation of corn crop transpirationby sap flow measurement under field conditions. Theor. Appl. Clim. 67, 153–160. https://doi.org/10.1007/s007040070004.

Bowen, I.S., 1926. The ratio of heat losses by conduction and by evaporation from any water surface. Phys. Rev. 27, 779–787. https://doi.org/10.1103/PhysRev.27.779.

Breiman, L., 2001. Random forests. Mach. Learn 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Bringhenti, T., Moriondo, M., Abdulai, I., Joubert, E., Roetter, R.P., Taylor, P.J., Hoffmann, M.P., 2025. Adopting and evaluating a simple model for macadamia tree transpiration in periodically water-scarce subtropical regions. Sci. Hortic. 341, 113970. https://doi.org/10.1016/j.scienta.2025.113970.

Burgess, S.S.O., Adams, M.A., Turner, N.C., Beverly, C.R., Ong, C.K., Khan, A.A.H., Bleby, T.M., 2001. An improved heat pulse method to measure low and reverse rates of sap flow in woody plants. Tree Physiol. 21, 589–598. https://doi.org/10.1093/treephys/21.9.589.

Camacho Suarez, V.V., Saraiva Okello, A.M.L., Wenninger, J.W., Uhlenbrook, S., 2015. Understanding runoff processes in a semi-arid environment through isotope and hydrochemical hydrograph separations. Hydrol. Earth Syst. Sci. 19, 4183–4199. https://doi.org/10.5194/hess-19-4183-2015.

Capurro, M.C., Ham, J.M., Kluitenberg, G.J., Comas, L., Andales, A.A., 2024. A novel sap flow system to measure maize transpiration using a heat pulse method. Agric. Water Manag 301, 108963. https://doi.org/10.1016/j.agwat.2024.108963.

Carvalho, D.F. de, Teles, G.C., Cruz, E.S. da, Valença, D. da C., Medici, L.O., 2023. Yield response factor (Ky) and initial growth in black pepper in a tropical environment. Sci. Agric. 80. https://doi.org/10.1590/1678-992x-2022-0171.

Chen, H., Cheng, Y., Du, T., Wu, X., Cao, Y., Liu, Y., 2025. Enhancing the performance of recycled aggregate green concrete via a Bayesian optimization light gradient boosting machine and the nondominated sorting genetic algorithm-III. Constr. Build. Mater. 458, 139527. https://doi.org/10.1016/j.conbuildmat.2024.139527.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Dangare, P., Mashimbye, Z.E., Cronje, P.J.R., Masanganise, J.N., Gokool, S., Ntshidi, Z., Naiken, V., Sawunyama, T., Dzikiti, S., 2025. Evapotranspiration Partitioning in Selected Subtropical Fruit Tree Orchards Based on Sentinel 2 Data Using a Light Gradient-Boosting Machine (LightGBM) Learning Model in Malelane, South Africa. Hydrology 12, 189. https://doi.org/10.3390/hydrology12070189.

Dangare, P., Mhizha, T., Mashonjowa, E., 2018b. Design, fabrication and calibration of a low cost smart sap flow measuring system based on ATmega 328/P microcontroller. In: Proceedings of the EAI International Conference for Research, Innovation and Development for Africa. EAI, Victoria Falls, pp. 108–119. https://doi.org/10.4108/eai.20-6-2017.2275846.

Dangare, P., Mhizha, T., Mashonjowa, E., 2018a. Design, fabrication and testing of a low cost Trunk Diameter Variation (TDV) measurement system based on an ATmega

328/P microcontroller. Comput. Electron Agric. 148, 197–206. https://doi.org/10.1016/j.compag.2018.03.022.
Dangare, P., Nel, G.P., Sawunyama, T., Cronje, P.J.R., Dzikiti, S., 2024. Measurement and modelling of water use by litchi (*Litchi sinensis*) orchards under subtropical conditions. Acta Hortic. 167–174. https://doi.org/10.17660/ActaHortic.2024.1409.23.
Daniel, C., 2024. A robust LightGBM model for concrete tensile strength forecast to aid in resilience-based structure strategies. Heliyon 10, e39679. https://doi.org/10.1016/j.heliyon.2024.e39679.
De Bruin, H.A.R., Van Den Hurk, B.J.J.M., Kohsiek, W., 1995. The scintillation method tested over a dry vineyard area. Bound. Layer. Meteor. 76, 25–40. https://doi.org/10.1007/BF00710889.
Ding, R., Kang, S., Zhang, Y., Hao, X., Tong, L., Du, T., 2013. Partitioning evapotranspiration into soil evaporation and transpiration using a modified dual crop coefficient model in irrigated maize field with ground-mulching. Agric. Water Manag. 127, 85–96. https://doi.org/10.1016/j.agwat.2013.05.018.
Doorenbos, J., Kassam, A.H., 1979. Yield response to water. In: FAO Irrigation and Drainage Paper, 33. FAO, Rome, pp. 1–193.
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. https://doi.org/10.3390/rs16050730.
Duarte, F.S.L.G., Rios, R.A., Hruschka, E.R., de Mello, R.F., 2019. Decomposing time series into deterministic and stochastic influences: a survey. Digit Signal Process 95, 102582. https://doi.org/10.1016/j.dsp.2019.102582.
Dzikiti, S., Dangare, P., Nel, G.P., Masanganise, J.N., Kapangaziwiri, E., Kleinert, A., Cronje, P.J., Midgley, S.J.E., Raath, P., Mashimbye, E.Z., Ntshidi, Z., 2024. Developing A Decision Support System For Water Use And Water-Use Efficiency Of Irrigated Crops In The Inkomati-Usuthu Water Management Area. Pretoria.
Dzikiti, S., Lotter, D., Mpandeli, S., Nhamo, L., 2022. Assessing the energy and water balance dynamics of rain-fed rooibos tea crops (Aspalathus linearis) under changing Mediterranean climatic conditions. Agric. Water Manag. 274. https://doi.org/10.1016/j.agwat.2022.107944.
Dzikiti, S., Ntshidi, Z., Le Maitre, D.C., Bugan, R.D.H., Mazvimavi, D., Schachtschneider, K., Jovanovic, N.Z., Pienaar, H.H., 2017. Assessing water use by Prosopis invasions and Vachellia karroo trees: implications for groundwater recovery following alien plant removal in an arid catchment in South Africa. Ecol. Manag. 398, 153–163. https://doi.org/10.1016/j.foreco.2017.05.009.
Dzikiti, S., Volschenk, T., Midgley, S.J.E., Lötze, E., Taylor, N.J., Gush, M.B., Ntshidi, Z., Zirebwa, S.F., Doko, Q., Schmeisser, M., Jarmain, C., Steyn, W.J., Pienaar, H.H., 2018. Estimating the water requirements of high yielding and young apple orchards in the winter rainfall areas of South Africa using a dual source evapotranspiration model. Agric. Water Manag. 208, 152–162. https://doi.org/10.1016/j.agwat.2018.06.017.
EL Bilali, A., Hadri, A., Taleb, A., Tanarhte, M., EL Khalki, E.M., Kharrou, M.H., 2025. A novel hybrid modeling approach based on empirical methods, PSO, XGBoost, and multiple GCMs for forecasting long-term reference evapotranspiration in a data scarce-area. Comput. Electron Agric. 232, 110106. https://doi.org/10.1016/j.compag.2025.110106.
Engelbrecht, F.A., Steinkopf, J., Padavatan, J., Midgley, G.F., 2024. Projections of Future Climate Change in Southern Africa and the Potential for Regional Tipping Points. pp. 169–190. https://doi.org/10.1007/978-3-031-10948-5_7.
En-Nia, S., Liouaeddine, M., Mansouri, Z., 2025. Impact of climate change on agricultural GDP in Morocco using machine learning techniques. Model Earth Syst. Environ. 11, 285. https://doi.org/10.1007/s40808-025-02486-w.
Er-Raki, S., Chehbouni, A., Boulet, G., Williams, D.G., 2010. Using the dual approach of FAO-56 for partitioning ET into soil and plant components for olive orchards in a semi-arid region. Agric. Water Manag 97, 1769–1778. https://doi.org/10.1016/j.agwat.2010.06.009.
Fan, J., Zheng, J., Wu, L., Zhang, F., 2021. 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.
Fayzullo, N., Sariyev, S., Sherzodjon, Y., 2025. Analyzing the Effectiveness of Ensemble Methods in Solving Multi-Class Classification Problems. In: 2025 International Russian Smart Industry Conference (SmartIndustryCon). IEEE, pp. 788–793. https://doi.org/10.1109/SmartIndustryCon65166.2025.10986248.
Feliciano, D., Smith, P., Mabhaudhi, T., 2025. Climate change exacerbates inequalities between small-scale and large-scale farmers in South Africa's fruit export market. Reg. Environ. Change 25, 28. https://doi.org/10.1007/s10113-024-02354-w.
Foken, T., Leuning, R., Oncley, S.R., Mauder, M., Aubinet, M., 2012. Corrections and Data Quality Control. In: Eddy Covariance. Springer Netherlands, Dordrecht, pp. 85–131. https://doi.org/10.1007/978-94-007-2351-1_4.
Gao, L., Wang, X., Johnson, B.A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., Gu, X., 2020. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: a review. ISPRS J. Photogramm. Remote Sens. 159, 364–377. https://doi.org/10.1016/j.isprsjprs.2019.11.018.
Garofalo, S.Pietro, Ardito, F., Sanitate, N., De Carolis, G., Ruggieri, S., Giannico, V., Rana, G., Ferrara, R.M., 2025. Robustness of actual evapotranspiration predicted by random forest model integrating remote sensing and meteorological information: case of watermelon (Citrullus lanatus, (Thunb.) Matsum. & Nakai, 1916). Water 17, 323. https://doi.org/10.3390/w17030323.
Giménez, L., Petillo, M., Paredes, P., Pereira, L., 2016. Predicting maize transpiration, water use and productivity for developing improved supplemental irrigation schedules in western uruguay to cope with climate variability. Water 8, 309. https://doi.org/10.3390/w8070309.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.
Gush, M.B., Taylor, N.J., 2014. The water use of selected fruit tree orchards: Technical report on measurements and modelling. WRC report No. 1770/2/14 2, 1–285.
Hailegnaw, N.S., Awoke, G.W., Santos, A. de C., Schaffer, B., Vargas, A.I., Souza, E.R. de, Bayabil, H.K., 2025. Assessing salinity-induced impacts on plant transpiration through machine learning: from model development to deployment. Model Earth Syst. Environ. 11, 173. https://doi.org/10.1007/s40808-025-02343-w.
Hanks, R.J., 1974. Model for predicting plant yield as influenced by water use 1. Agron. J. 66, 660–665. https://doi.org/10.2134/agronj1974.00021962006600050017x.
He, X., Li, Y., Liu, S., Xu, T., Chen, F., Li, Z., Zhang, Z., Liu, R., Song, L., Xu, Z., Peng, Z., Zheng, C., 2023. Improving regional climate simulations based on a hybrid data assimilation and machine learning method. Hydrol. Earth Syst. Sci. 27, 1583–1606. https://doi.org/10.5194/hess-27-1583-2023.
He, X., Liu, S., Bateni, S.M., Xu, T., Jun, C., Kim, D., Li, X., Song, L., Zhao, L., Xu, Z., Wei, J., 2024. Innovative approach for estimating evapotranspiration and gross primary productivity by integrating land data assimilation, machine learning, and multi-source observations. Agric. Meteor. 355, 110136. https://doi.org/10.1016/j.agrformet.2024.110136.
He, X., Liu, S., Xu, T., Yu, K., Gentine, P., Zhang, Z., Xu, Z., Jiao, D., Wu, D., 2022. Improving predictions of evapotranspiration by integrating multi-source observations and land surface model. Agric. Water Manag. 272, 107827. https://doi.org/10.1016/j.agwat.2022.107827.
Hu, X., Shi, L., Lin, G., Lin, L., 2021. Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. J. Hydrol. 601, 126592. https://doi.org/10.1016/j.jhydrol.2021.126592.
Imani, M., Beikmohammadi, A., Arabnia, H.R., 2025. Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. Technologies 13, 88. https://doi.org/10.3390/technologies13030088.
Inkomati-Usuthu CMA Business Case, 2012. Bus. Case InkomatiUsuthu Catchment Manag. Agency.
Ismail, S., Nouman, M., Dawoud, D.W., Reza, H., 2024. Towards a lightweight security framework using blockchain and machine learning. Block. Res. Appl. 5, 100174. https://doi.org/10.1016/j.bcra.2023.100174.
Jain, S.K., Gupta, A.K., 2025. Estimation of seasonal crop water requirement using support vector regression in India's arid zone. Curr. World Environ. 20, 144–154. https://doi.org/10.12944/CWE.20.1.12.
Jensen, A.R., 1968. Social class, race, and genetics: implications for education. Am. Educ. Res J. 5, 1–42. https://doi.org/10.3102/00028312005001001.
Jones, H.G., Tardieu, F., 1998. Modelling water relations of horticultural crops: a review. Sci. Hortic. 74, 21–46. https://doi.org/10.1016/S0304-4238(98)00081-8.
Karunasingha, D.S.K., 2022. Root mean square error or mean absolute error? Use their ratio as well. Inf. Sci. (N. Y) 585, 609–629. https://doi.org/10.1016/j.ins.2021.11.036.
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: 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA.
Koppa, A., Rains, D., Hulsman, P., Poyatos, R., Miralles, D.G., 2022. A deep learning-based hybrid model of global terrestrial evaporation. Nat. Commun. 13, 1912. https://doi.org/10.1038/s41467-022-29543-7.
Kouloumpris, E., Vlahavas, I., 2025. Markowitz random forest: weighting classification and regression trees with modern portfolio theory. Neurocomputing 620, 129191. https://doi.org/10.1016/j.neucom.2024.129191.
Lavaei, H., Esmaeili, M., Mehraein, M., 2025. Enhanced prediction of scour dimensions: temporal variations induced by turbulent plane wall jets using FFNN, CatBoost, and XGBoost models. Ocean Eng. 333, 121539. https://doi.org/10.1016/j.oceaneng.2025.121539.
Li, Y., Liu, X., Zhang, X., Gu, X., Yu, L., Cai, H., Peng, X., 2025a. Using solar-induced chlorophyll fluorescence to predict winter wheat actual evapotranspiration through machine learning and deep learning methods. Agric. Water Manag. 309, 109322. https://doi.org/10.1016/j.agwat.2025.109322.
Li, S., Luo, L., Li, J., 2025. Advanced machine learning approaches for predicting energy and fossil fuel consumption for green growth. Unconv. Resour., 100262 https://doi.org/10.1016/j.uncres.2025.100262.
Li, Y., Peng, X., Liu, Z., Lu, X., Gu, X., Yu, L., Xu, J., Cai, H., 2025b. A machine learning-driven semi-mechanistic model for estimating actual evapotranspiration: integrating photosynthetic indicators with vapor pressure deficit. Agric. Water Manag. 315, 109563. https://doi.org/10.1016/j.agwat.2025.109563.
Li, X., Yang, P., Ren, S., Li, Y., Liu, H., Du, J., Li, P., Wang, C., Ren, L., 2010. Modeling cherry orchard evapotranspiration based on an improved dual-source model. Agric. Water Manag. 98, 12–18. https://doi.org/10.1016/j.agwat.2010.07.019.
Liao, W., Chen, Z., Li, P., Tan, J., Li, H., Zhao, C., 2025. Enhanced battery health monitoring in electric vehicles: a novel hybrid HBA-HGBR model. J. Energy Storage 110, 115316. https://doi.org/10.1016/j.est.2025.115316.
Liu, X., Tang, Z., Wei, J., 2025. Multi-layer perceptron model integrating multi-head attention and gating mechanism for global navigation satellite system positioning error estimation. Remote Sens. 17, 301. https://doi.org/10.3390/rs17020301.
Liu, H., Wang, Y., Chen, W., 2020. Three-step imputation of missing values in condition monitoring datasets. IET Gener. Transm. Distrib. 14, 3288–3300. https://doi.org/10.1049/iet-gtd.2019.1446.
Lovelli, S., Perniola, M., Ferrara, A., Di Tommaso, T., 2007. Yield response factor to water (Ky) and water use efficiency of Carthamus tinctorius L. and Solanum melongena L. Agric. Water Manag. 92, 73–80. https://doi.org/10.1016/j.agwat.2007.05.005.

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA, pp. 4766–4777.
Ma, W., Zhang, Xiao, Xie, J., Zuo, G., Luo, F., Zhang, Xu, Jin, T., Yang, X., 2025. Prediction of non-stationary daily streamflow series based on ensemble learning: a case study of the Wei River Basin, China. Stoch. Environ. Res. Risk Assess. 39, 509–529. https://doi.org/10.1007/s00477-024-02877-y.
Mabhaudhi, T., Nhamo, L., Mpandeli, S., 2021. Enhancing crop water productivity under increasing water scarcity in South Africa. Climate Change Science. Elsevier, pp. 1–18. https://doi.org/10.1016/B978-0-12-823767-0.00001-X.
Maluleke, A., Feig, G., Brümmer, C., Jaars, K., Hamilton, T., Midgley, G., 2025. Paired eddy covariance site reveals consistent net C sinks over three growing seasons in an African arid and grassy shrubland. Agric. Meteor. 372, 110705. https://doi.org/10.1016/j.agrformet.2025.110705.
Mangani, R., Mazarura, J., Matlou, S., Marquart, A., Archer, E., Creux, N., 2025. The impact of past and current district-level climatic shifts on maize production and the implications for South African farmers. Theor. Appl. Clim. 156, 109. https://doi.org/10.1007/s00704-024-05334-6.
Marek, T.H., Schneider, A.D., Howell, T.A., Ebeling, L.L., 1988. Design and construction of large weighing monolithic lysimeters. Trans. ASAE 31, 0477–0484. https://doi.org/10.13031/2013.30734.
Masanganise, J., Kunz, R., Clulow, A.D., Mabhaudhi, T., Savage, M.J., 2022. Evapotranspiration estimates of soybean using surface renewal: comparison with crop coefficient approach. Phys. Chem. Earth Parts A/B/C. 128, 103244. https://doi.org/10.1016/j.pce.2022.103244.
Mashabatu, M., Motsei, N., Jovanovic, N., Nhamo, L., 2025. A validation of fruitlook data using eddy covariance in a fully mature and high-density japanese plum orchard in the Western Cape, South Africa. Water (Basel) 17, 324. https://doi.org/10.3390/w17030324.
Menzel, C., Oosthuizen, J., Roe, D., Doogan, V., 1995. Water deficits at anthesis reduce CO2 assimilation and yield of lychee (Litchi chinensis Sonn.) trees. Tree Physiol. 15, 611–617. https://doi.org/10.1093/treephys/15.9.611.
Meyers, T.P., Hollinger, S.E., 2004. An assessment of storage terms in the surface energy balance of maize and soybean. Agric. Meteor. 125, 105–115. https://doi.org/10.1016/j.agrformet.2004.03.001.
Moeletsi, M.E., Myeni, L., Kaempffer, L.C., Vermaak, D., de Nysschen, G., Henningse, C., Nel, I., Rowswell, D., 2022. Climate dataset for South Africa by the agricultural research council. Data 7, 1–7.
Mohammadi, B., Chen, M., Nikoo, M.R., Al-Maktoumi, A., Yu, Y., Yu, R., 2025. Enhancing daily runoff prediction: a hybrid model combining GR6J-CemaNeige with wavelet-based gradient boosting technique. J. Hydrol. 657, 133114. https://doi.org/10.1016/j.jhydrol.2025.133114.
Mpakairi, K.S., Dube, T., Sibanda, M., Mutanga, O., 2025. Leveraging remote sensing for optimised national scale agricultural water management in South Africa. Sci. Total Environ. 974, 179199. https://doi.org/10.1016/j.scitotenv.2025.179199.
Nagro, S., 2025. A stacked ensemble approach for symptom-based monkeypox diagnosis. Comput. Biol. Med 191, 110140. https://doi.org/10.1016/j.compbiomed.2025.110140.
Nel, G.P., Dangare, P., Kleinert, A., Dzikiti, S., 2024. Estimating crop coefficients and water use of a full-bearing mango orchard in north-eastern South Africa using the fraction of vegetation cover and a dual source evapotranspiration model. Sci. Hortic. 336, 113388. https://doi.org/10.1016/j.scienta.2024.113388.
Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P., Righetti, M., 2024. Applications of XGBoost in water resources engineering: a systematic literature review (Dec 2018–May 2023). Environ. Model. Softw. 174, 105971. https://doi.org/10.1016/j.envsoft.2024.105971.
Nohara, Y., Matsumoto, K., Soejima, H., Nakashima, N., 2022. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput. Methods Prog. Biomed. 214, 106584. https://doi.org/10.1016/j.cmpb.2021.106584.
Norman, J.M., Anderson, M.C., Kustas, W.P., French, A.N., Mecikalski, J., Torn, R., Diak, G.R., Schmugge, T.J., Tanner, B.C.W., 2003. Remote sensing of surface energy fluxes at 10^{1} -m pixel resolutions. Water Resour. Res 39. https://doi.org/10.1029/2002WR001775.
Norman, J.M., Kustas, W.P., Humes, K.S., 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. Agric. Meteor. 77, 263–293. https://doi.org/10.1016/0168-1923(95)02265-Y.
Ntshidi, Z., Dzikiti, S., Mazvimavi, D., Mobe, N.T., 2021. Contribution of understorey vegetation to evapotranspiration partitioning in apple orchards under mediterranean climatic conditions in South Africa. Agric. Water Manag 245, 106627. https://doi.org/10.1016/j.agwat.2020.106627.
Pagano, A., Amato, F., Ippolito, M., De Caro, D., Croce, D., Motisi, A., Provenzano, G., Tinnirello, I., 2023. Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation. Ecol. Inf. 76, 102133. https://doi.org/10.1016/j.ecoinf.2023.102133.
Pande, C.B., Radwan, N., Heddam, S., Ahmed, K.O., Alshehri, F., Pal, S.C., Pramanik, M., 2025. Forecasting of monthly air quality index and understanding the air pollution in the urban city, India based on machine learning models and cross-validation. J. Atmos. Chem. 82, 1. https://doi.org/10.1007/s10874-024-09466-x.
Paredes, P., D'Agostino, D., Assif, M., Todorovic, M., Pereira, L.S., 2018. Assessing potato transpiration, yield and water productivity under various water regimes and planting dates using the FAO dual K c approach. Agric. Water Manag. 195, 11–24. https://doi.org/10.1016/j.agwat.2017.09.011.
Paredes, P., Petry, M.T., Oliveira, C.M., Montoya, F., López-Urrea, R., Pereira, L.S., 2024. Single and basal crop coefficients for estimation of water requirements of subtropical and tropical orchards and plantations with consideration of fraction of ground cover, height, and training system. Irrig. Sci. 42, 1059–1097. https://doi.org/10.1007/s00271-024-00925-7.
Paredes, P., Rodrigues, G.C., Alves, I., Pereira, L.S., 2014. Partitioning evapotranspiration, yield prediction and economic returns of maize under various irrigation management strategies. Agric. Water Manag 135, 27–39. https://doi.org/10.1016/j.agwat.2013.12.010.
Paw U, K.T., Jie, Q., Hong-Bing, S., Tomonori, W., Yves, B., 1995. Surface renewal analysis: a new method to obtain scalar fluxes. Agric. Meteor. 74, 119–137. https://doi.org/10.1016/0168-1923(94)02182-J.
Pervez, A., Jamal, A., 2025. Exploring e-scooter risk factors based on interpretable machine learning framework. J. Saf. Res 94, 128–140. https://doi.org/10.1016/j.jsr.2025.06.011.
Pieruschka, R., Huber, G., Berry, J.A., 2010. Control of transpiration by radiation. Proc. Natl. Acad. Sci. 107, 13372–13377. https://doi.org/10.1073/pnas.0913177107.
Qin, G., Zhang, C., Wang, B., Ni, P., Wang, Y., 2025. An interpretable machine learning model for failure pressure prediction of blended hydrogen natural gas pipelines containing a crack-in-dent defect. Energy 320, 135401. https://doi.org/10.1016/j.energy.2025.135401.
Qiu, R., Du, T., Kang, S., Chen, R., Wu, L., 2015. Assessing the SIMDualKc model for estimating evapotranspiration of hot pepper grown in a solar greenhouse in Northwest China. Agric. Syst. 138, 1–9. https://doi.org/10.1016/j.agsy.2015.05.001.
Rahimi, M., Karbasi, M., Jamei, M., Rezaverdinejad, V., Malik, A., Farooque, A.A., Yaseen, Z.M., 2025. Meticulous estimation of maize actual evapotranspiration: a comprehensive explainable CatBoost algorithm reinforced with Jackknife uncertainty paradigm. Comput. Electron Agric. 237, 110599. https://doi.org/10.1016/j.compag.2025.110599.
Rahman, M., Hasan, M.M., Hossain, M.A., Das, U.K., Islam, M.M., Karim, M.R., Faiz, H., Hammad, Z., Sadiq, S., Alam, M., 2025. Integrating deep learning algorithms for forecasting evapotranspiration and assessing crop water stress in agricultural water management. J. Environ. Manag. 375, 124363. https://doi.org/10.1016/j.jenvman.2025.124363.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.
Reinders, F.B., van der Stoep, I., Backeberg, G.R., 2013. Improved efficiency of irrigation water use: a South African framework. Irrig. Drain. 62, 262–272. https://doi.org/10.1002/ird.1742.
Rosa, R.D., Paredes, P., Rodrigues, G.C., Fernando, R.M., Alves, I., Pereira, L.S., Allen, R. G., 2012. Implementing the dual crop coefficient approach in interactive software: 2. Model testing. Agric. Water Manag 103, 62–77. https://doi.org/10.1016/j.agwat.2011.10.018.
Sadig, H.E., Kamal, M., Rehman, M. ur, Habadi, M.I., Alnagar, D.K., Yusuf, M., Musa Mohammed, M.O., Alqasem, O.A., Meraou, M.A., 2025. Advanced time complexity analysis for real-time COVID-19 prediction in Saudi Arabia using LightGBM and XGBoost. J. Radiat. Res Appl. Sci. 18, 101364. https://doi.org/10.1016/j.jrras.2025.101364.
Sadras, V., Alston, J., Aphalo, P., Connor, D., Denison, R.F., Fischer, T., Gray, R., Hayman, P., Kirkegaard, J., Kirchmann, H., Kropff, M., Lafitte, H.R., Langridge, P., Lenne, J., Mínguez, M.I., Passioura, J., Porter, J.R., Reeves, T., Rodriguez, D., Ryan, M., Villalobos, F.J., Wood, D., 2020. Making science more effective for agriculture. pp. 153–177. https://doi.org/10.1016/bs.agron.2020.05.003.
Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J.B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., Zhang, L., 2023. Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau. Remote Sens. Environ. 289, 113519. https://doi.org/10.1016/j.rse.2023.113519.
Shen, Y., Wu, S., Wang, Y., Wang, J., Yang, Z., 2025. Interpretable model for rockburst intensity prediction based on Shapley values-based Optuna-random forest. Undergr. Space 21, 198–214. https://doi.org/10.1016/j.undsp.2024.09.002.
Silva, J.A. da, Rezende, M.K.A., Flumignan, D.L., 2019. Yield response factor (Ky) for winter corn crop in the region of Dourados, MS, Brazil. Eng. Agrícola 39, 573–578. https://doi.org/10.1590/1809-4430-eng.agric.v39n5p573-578/2019.
Sinclair, T.R., Tanner, C.B., Bennett, J.M., 1984. Water-use efficiency in crop production. Bioscience 34, 36–40. https://doi.org/10.2307/1309424.
Snyder, R.L., Spano, D., Pawu, K.T., 1996. Surface renewal analysis for sensible and latent heat flux density. Bound. Layer. Meteor. 77, 249–266. https://doi.org/10.1007/BF00123527.
Stewart, J.I., Hagan, R.M., Pruitt, W.O., Danielson, R.E., Franklin, W.T., Hanks, R.J., Riley, J.P., Jackson, E.B., 1977. Optimizing crop production through control of water and salinity levels in the soil. Rep. Pap. 67.
Sun, D., Zhang, H., Qi, Y., Ren, Y., Zhang, Z., Li, X., Lv, Y., Cheng, M., 2025. A comparative analysis of different algorithms for estimating evapotranspiration with limited observation variables: a case study in Beijing, China. Remote Sens. 17, 636. https://doi.org/10.3390/rs17040636.
Swanson, R.H., Whitfield, D.W.A., 1981. A numerical analysis of heat pulse velocity theory and practice. J. Exp. Bot. 32, 221–239. https://doi.org/10.1093/jxb/32.1.221.
Taheri, M., Bigdeli, M., Imanian, H., Mohammadian, A., 2025. An overview of evapotranspiration estimation models utilizing artificial intelligence. Water 17, 1384. https://doi.org/10.3390/w17091384.
Taheri, M., Mohammadian, A., Ganji, F., Bigdeli, M., Nasseri, M., 2022. Energy-based approaches in estimating actual evapotranspiration focusing on land surface temperature: a review of methods, concepts, and challenges. Energies 15, 1264. https://doi.org/10.3390/en15041264.

Tang, D., Feng, Y., Gong, D., Hao, W., Cui, N., 2018. Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulched maize croplands. Comput. Electron Agric. 152, 375–384. https://doi.org/10.1016/j.compag.2018.07.029.

Tolun, G.G., Tolun, Ö.C., Zor, K., 2025. Advanced machine learning algorithms for reactive power forecasting in electric distribution systems. ePrime Adv. Electr. Eng. Electron. Energy 13, 101019. https://doi.org/10.1016/j.prime.2025.101019.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. Meteor. 103, 279–300. https://doi.org/10.1016/S0168-1923(00)00123-4.

Ubisi, N.R., Kolanisi, U., Jiri, O., 2020. The role of indigenous knowledge systems in rural smallholder farmers' response to climate change: case study of Nkomazi Local Municipality, Mpumalanga, South Africa. J. Asian Afr. Stud. 55, 273–284. https://doi.org/10.1177/0021909619874824.

Ünlu, M., Kanber, R., Koc, D.L., Özekici, B., Kekec, U., Yesiloglu, T., Ortas, İ., Ünlu, F., Kapur, B., Tekin, S., Kathner, J., Gebbers, R., Zude, M., Peeters, A., Ben-Gal, A., 2014. Irrigation scheduling of grapefruit trees in a Mediterranean environment throughout evaluation of plant water status and evapotranspiration. Turk. J. Agric. For. 38, 908–915. https://doi.org/10.3906/tar-1403-58.

Villalobos, F.J., Testi, L., Orgaz, F., García-Tejera, O., Lopez-Bernal, A., González-Dugo, M.V., Ballester-Lurbe, C., Castel, J.R., Alarcón-Cabañero, J.J., Nicolás-Nicolás, E., Girona, J., Marsal, J., Fereres, E., 2013. Modelling canopy conductance and transpiration of fruit trees in Mediterranean areas: a simplified approach. Agric. Meteor. 171–172, 93–103. https://doi.org/10.1016/j.agrformet.2012.11.010.

Wang, Y., Wang, P., Tansey, K., Liu, J., Delaney, B., Quan, W., 2025. An interpretable approach combining Shapley additive explanations and LightGBM based on data augmentation for improving wheat yield estimates. Comput. Electron Agric. 229, 109758. https://doi.org/10.1016/j.compag.2024.109758.

Wang, H., Yilihamu, Q., Yuan, M., Bai, H., Xu, H., Wu, J., 2020. Prediction models of soil heavy metal(loid)s concentration for agricultural land in Dongli: a comparison of regression and random forest. Ecol. Indic. 119, 106801. https://doi.org/10.1016/j.ecolind.2020.106801.

Wei, Y.Z., Zhang, H.N., Li, W.C., Xie, J.H., Wang, Y.C., Liu, L.Q., Shi, S.Y., 2013. Phenological growth stages of lychee (Litchi chinensis Sonn.) using the extended BBCH-scale. Sci. Hortic. 161, 273–277. https://doi.org/10.1016/j.scienta.2013.07.017.

Woldemariam, G.W., Awoke, B.G., Maretto, R.V., 2025. Satellite-Based energy balance for estimating actual sugarcane evapotranspiration in the Ethiopian Rift Valley. ISPRS J. Photogramm. Remote Sens. 223, 109–130. https://doi.org/10.1016/j.isprsjprs.2025.03.003.

Wu, S., Deng, L., Qiao, Q., 2025. Estimating long-term fractional vegetation cover using an improved dimidiate pixel method with UAV-Assisted Satellite Data: a case study in a mining region. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens 18, 4162–4173. https://doi.org/10.1109/JSTARS.2025.3531439.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15, 5015–5030. https://doi.org/10.5194/bg-15-5015-2018.

Xiang, Y., Wang, G., Gessler, A., Sun, X., Lin, S., Tang, Z., Sun, S., Hu, Z., 2025. Long-term variations in the ratio of transpiration to evapotranspiration and their drivers in a humid subalpine forest. Agric. Meteor. 372, 110692. https://doi.org/10.1016/j.agrformet.2025.110692.

Xing, W., Feng, Z., Wei, J., Xu, S., Shao, Q., Wang, W., Shi, X., 2025. Impacts of climate extremes on variations in evergreen forest ecosystem carbon–water fluxes across Southern China. Glob. Planet Change 252, 104867. https://doi.org/10.1016/j.gloplacha.2025.104867.

Yang, K., Wang, JiaMing, Zhao, G., Wang, X., Cong, W., Yuan, M., Luo, J., Dong, X., Wang, JiaRui, Tao, J., 2025. NIDS-CNNRF integrating CNN and random forest for efficient network intrusion detection model. Internet Things 32, 101607. https://doi.org/10.1016/j.iot.2025.101607.

Yang, D., Yang, S., Huang, J., Zhang, Shuyu, Zhang, Sha, Zhang, J., Bai, Y., 2025. Improving time upscaling of instantaneous evapotranspiration based on machine learning models. Big Earth Data 9, 127–154. https://doi.org/10.1080/20964471.2024.2423431.

Yaseen, Z.M., Alhalimi, F.L., 2025. Heavy metal adsorption efficiency prediction using biochar properties: a comparative analysis for ensemble machine learning models. Sci. Rep. 15, 13434. https://doi.org/10.1038/s41598-025-96271-5.

Yilgan, F., Miháliková, M., Kara, R.S., Ustuner, M., 2025. Analysis of the forest fire in the 'Bohemian Switzerland' National Park using Landsat-8 and Sentinel-5P in Google Earth Engine. Nat. Hazards 121, 6133–6154. https://doi.org/10.1007/s11069-024-07052-8.

Yu, Z., Kanwal, Q., Wang, M., Nurdiawati, A., Al-Ghamdi, S.G., 2025. Spatiotemporal dynamics and key drivers of carbon emissions in regional construction sectors: Insights from a Random Forest Model. Clean. Environ. Syst. 16, 100257. https://doi.org/10.1016/j.cesys.2025.100257.

Zhang, Jinglin, Li, G., Du, Z., Bao, S., Li, C., Cao, X., Yuan, C., 2025. Machine learning modeling using XGBoost and LightGBM for predicting the minimum ignition temperature of rice husk dust based on the synergistic effect of dispersion pressure and crushed brown rice. Powder Technol. 453, 120682. https://doi.org/10.1016/j.powtec.2025.120682.

Zhang, X., Yan, G., Li, Q., Li, Z. -L., Wan, H., Guo, Z., 2006. Evaluating the fraction of vegetation cover based on NDVI spatial scale correction model. Int. J. Remote Sens. 27, 5359–5372. https://doi.org/10.1080/01431160600658107.

Zhang, Y., Ye, A., Li, J., Nguyen, P., Analui, B., Hsu, K., Sorooshian, S., 2025. Improve streamflow simulations by combining machine learning pre-processing and post-processing. J. Hydrol. 655, 132904. https://doi.org/10.1016/j.jhydrol.2025.132904.

Zhao, Y., Dong, H., Huang, W., He, S., Zhang, C., 2024. Seamless terrestrial evapotranspiration estimation by machine learning models across the Contiguous United States. Ecol. Indic. 165, 112203. https://doi.org/10.1016/j.ecolind.2024.112203.

Zhao, P., Li, S., Li, F., Du, T., Tong, L., Kang, S., 2015. Comparison of dual crop coefficient method and Shuttleworth–Wallace model in evapotranspiration partitioning in a vineyard of northwest China. Agric. Water Manag 160, 41–56. https://doi.org/10.1016/j.agwat.2015.06.026.

Zheng, G., Zhang, Y., Yue, X., Li, K., 2023. Interpretable prediction of thermal sensation for elderly people based on data sampling, machine learning and SHapley Additive exPlanations (SHAP). Build. Environ. 242, 110602. https://doi.org/10.1016/j.buildenv.2023.110602.