Environ Monit Assess (2023) 195:877
https://doi.org/10.1007/s10661-023-11536-8

RESEARCH

Accurate estimation of sorghum crop water content under different water stress levels using machine learning and hyperspectral data

Emre Tunca · Eyüp Selim Köksal · Elif Öztürk · Hasan Akay · Sakine Çetin Taner

Received: 18 March 2023 / Accepted: 19 June 2023 / Published online: 23 June 2023

Abstract This study investigates the effects of different water stress levels on spectral information, leaf area index (LAI), and the performance of three machine learning (ML) algorithms in estimating crop water content (CWC) of sorghum. The results show that the spectral reflectance of sorghum varies with growth stage and irrigation treatment, but consistent patterns are observed for each treatment. The LAI of sorghum gradually increased throughout the growth stages, with the most significant variation observed during the flowering stage. In this study, three machine learning-based regression models, namely, extreme gradient boosting (XGBoost), random forest (RF), and support vector machine (SVM), were utilized to estimate sorghum CWC using hyperspectral measurements. Recursive feature elimination (RFE) method was used to select the optimal spectral reflectance wavelengths for the ML models, and principal component analysis (PCA) was used to reduce the dimensionality of the hyperspectral data. The results indicated that the RF model achieved the highest R^2 (0.90) and lowest of RMSE (56.05) value using selected wavelengths, while the XGBoost model demonstrated superior accuracy and reliability in estimating CWC using dimensionality-reduced hyperspectral data (r = 0.96, RMSE = 45.77). Also, the study highlights the importance of vegetation index (VI) in CWC estimate. Some VIs, such as NDVI and MSAVI, performed poorly, while others, such as CL_Rededge and EVI, performed better. The study provides valuable insights into the effects of water stress levels on spectral information, LAI, and the performance of ML algorithms in estimating the CWC of sorghum. The findings have significant implications for precision agriculture, as accurate and reliable estimates of CWC can help farmers optimize irrigation and fertilizer applications, leading to improved crop yields and resource efficiency.

Keywords Crop water content · Hyperspectral · ML · LAI · Vegetation indices

E. Tunca (✉)
Department of Biosystem Engineering, Faculty of Agriculture, Düzce University, Düzce, Turkey
e-mail: emretunca@duzce.edu.tr

E. S. Köksal · S. Çetin Taner
Department of Agricultural Structures and Irrigation, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Turkey

E. Öztürk · H. Akay
Department of Field Crops, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Turkey

Introduction

Sorghum is a cereal crop with various uses in different sectors, such as food, feed, energy, and industry (Wanga et al., 2022). However, its optimal yield depends on how often it rains during its growth period, highlighting the importance of effective

irrigation management. Crop water content (CWC) has a direct effect on the growth, yield, and quality of sorghum, as shown by Aydinsakir et al. (2021) and Allamine et al. (2023). CWC can also indirectly indicate soil moisture content (Zhang et al., 2021). Hence, CWC determination is vital for assessing crop growth and irrigation management in a timely and accurate manner.

Previously, trained professionals conducted in-situ measurements or used conventional methods to evaluate CWC (Chivasa et al., 2020). However, utilizing these techniques requires significant effort, high cost and longer time periods, making them impractical for regular and efficient crop monitoring (Yue et al., 2018). Hyperspectral data have demonstrated its effectiveness in monitoring CWC over the decades by providing detailed information on crop reflectance across various wavelengths. Several studies have employed hyperspectral data to enhance the accuracy of CWC estimation for different crops and vegetation types, such as rice (Elsherbiny et al., 2021), wheat (Shi et al., 2022), maize (Meiyan et al., 2022), poplar (Colombo et al., 2008), and natural vegetation (Clevers et al., 2010). Moreover, hyperspectral data has also been applied to other aspects of crop management, such as nutrient concentration (Eshkabilov et al., 2022), soil physical properties (Arslan et al., 2014), crop yield (Irik & Kirnak, 2022) and irrigation (Köksal, 2011) estimation.

Estimating CWC using hyperspectral data is challenging due to the complexity of the relationship between reflectance and CWC, influenced by various factors such as leaf biochemistry, canopy structure, and soil moisture level, making it difficult to directly derive CWC values from hyperspectral measurements. These complexities necessitate a deeper exploration of the feasibility and limitations of using hyperspectral data for accurate CWC estimation. In recent years, machine learning (ML) techniques have emerged as promising solutions to overcome these challenges and extract valuable information from the hyperspectral data. For instance, Jin et al. (2017) applied a support vector machine (SVM) model to estimate grass CWC and achieved high model accuracy ($R^2 = 0.98$). Sibanda et al. (2021) used random forest (RF) to estimate the grasslands CWC, obtaining an R^2 of 0.98 and an RMSE of 9.8 gm^{-2}. In addition to these ML models, partial least squares regression (Das et al., 2021), decision trees (Ji et al., 2014), K-nearest neighbor (Zhang et al., 2021) and artificial neural networks (Elsherbiny et al., 2021) were employed to estimate CWC. The efficacy of the model based on hyperspectral data is depends on two primary determinants: (I) feature selection and (II) model selection (Elsherbiny et al., 2021). Model-based feature (MF) and principal component analysis (PCA) are the commonly employed methods for feature selection. Among them, MF selection technique is characterized by the identification of a feature subset that displays a high level of discriminative power and predictive utility (Beltrán et al., 2005). The MF method offers a feasible means of improving model performance by eliminating redundant features, preventing over-fitting, and retaining the original feature representation, which enhances model interpretability (Guyon & Elisseeff, 2003). Some scholars used this method to choose optimal spectral bands to estimate CWC (Elsherbiny et al., 2021). In addition to these methods, dimension reduction (DR) using PCA is another effective method to reduce the dimension of hyperspectral reflectance. DR is a technique that transforms data into a lower dimensional space while eliminating irrelevant variance and detecting important features. DR extracts critical low-dimensional features (Hasanlou & Samadzadegan, 2012). While other feature selection methods have been widely used for choosing suitable spectral bands, there is little research on using the DR of hyperspectral data in combination with ML methods for CWC estimation. Additionally, there is limited research on using ML models to estimate CWC under different water stress levels, such as fully irrigated, partially stressed, and rainfed conditions. Most previous studies have focused on modeling CWC without water stress, and it is unclear whether these models can be applied to other water stress conditions.

This study aimed to develop a model that utilizes proximal hyperspectral data and ML algorithms to accurately estimate the CWC of sorghum plants grown under different water stress levels. Specifically, sensitive bands and spectral VIs were selected to estimate sorghum CWC. Input variables were optimized using MF-based feature selection and DR methods to identify the best-performing ML algorithms.

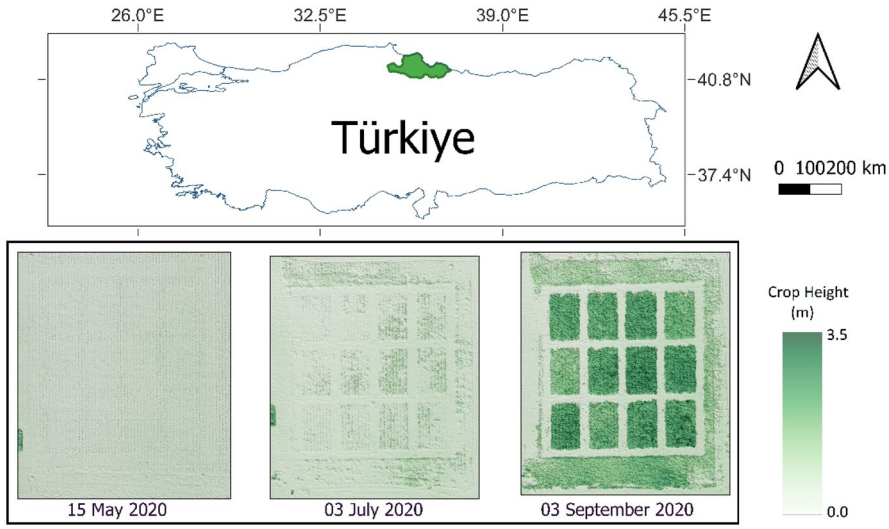


Fig. 1 The location of the study area and spatial and temporal sorghum height changes within the study area during 2020. The study area is in the northern part of Türkiye. The sorghum height was measured using a UAV at a height of 40 m, and the measurements were obtained from a canopy height model (CHM)

Materials and methods

Study area

The study was conducted during the 2020 and 2021 sorghum growing seasons in Samsun, Türkiye. Its geographic coordinates are 41° 36′ N, 33° 55′ E, and its altitude is 16 m above mean sea level (Fig. 1). The study area is located in a semi-humid weather zone with a mean annual temperature of 14.6 °C. Annual mean total precipitation in the region is 717.9 mm per year from 1990 to 2020. The soils in the experimental plots were alluvial and colluvial, with a total available water of 177.2 mm m^{-1}.

Field experiment

In this study, an experimental area of 0.1 hectares was utilized to investigate the effects of four irrigation levels: S1 (full irrigation), S2 (70% of S1), S3 (40% of S1), and S4 (rainfed) to sorghum CWC. The experimental design involved a randomized complete block design with three replications for each irrigation level, resulting in 12 treatment plots. Each plot measured 10 by 6.3 m, with a 2.1-m-wide alley between adjacent plots. Seeds were sown on May 15, 2020, and May 20, 2021, in rows with 0.7 m spacing between rows and 0.05 m spacing between plants. Standard cultivation practices were used to control pests and diseases. Sorghum grains were harvested on September 23, 2020, and September 27, 2021.

Proximal hyperspectral data acquisition

The ASD Field Spec Pro spectrometer was utilized to obtain sorghum canopy spectral reflectance values from a height of 4.5 m above ground level. The measurements were taken at the same height throughout the growing season to detect the changes in vegetation. Spectral measurements were captured at an angle of incidence of sunlight at approximately 45°. Reflectance values were collected in the 325–1075 nm range with a spectral resolution of 1 nm. A white calibration panel of $BaSO_4$ was used to calculate black and baseline reflectance. The spectroradiometer used in this study was calibrated before each measurement to ensure accurate and reliable spectral data. To minimize the impact of external factors such as weather and field conditions, 3 points within each plot were measured and averaged to generate a single value. Measurements were collected at multiple dates corresponding to key growth stages of sorghum (Table 1).

Table 1 Dates of proximal spectral reflectance measurements during the 2020 and 2021 sorghum growing season

	2020	2021
June	8, 11, 18, 23, 29	15
July	1, 3, 6, 10, 27	4, 6, 10, 13, 17, 21, 24, 26, 28
August	3, 6, 12, 17, 19, 24, 28	2, 7, 21, 29
September	3,15	1, 12, 25

Canopy water content measurements

The data were collected from the experimental field on multiple dates in 2020 and 2021. Specifically, the dates were 4, 14, 24 July, 5, 25 August, and 14 September 2020, which corresponded to 50, 60, 70, 82, 102, and 122 days after sowing (DAS). Additionally, data were collected on 26 July, 5, 16, 25 August, and 3, 23 September 2021, which corresponded to 72, 82, 93, 102, 111, and 131 DAS, respectively. A representative selection of sorghum plants was obtained from 12 treatment plots for destructive sampling, and the total leaf area of each sample was measured using the method given by Tunca et al. (2022). The sorghum's fresh weight (F_W) was weighed, and dry weight (D_W) was obtained by drying plants at 80 °C until a constant weight was reached. Plant water status was calculated using Eq. 1.

$$Plant\ Water\ Status\ (\%) = \left(\frac{F_w - D_w}{F_w}\right) \times 100 \quad (1)$$

The leaf area index (LAI), a key parameter for characterizing crop growth (Tunca et al., 2022), was calculated using Eq. 2 as follows:

$$LAI = \frac{Total\ Leaf\ Area}{(Row \times Distance)} \quad (2)$$

The LAI is closely related to photosynthesis and transpiration in crops, as noted by Meiyan et al. (2022) and Xu et al. (2019). Therefore, compared to plant water status, revised version using LAI (CWC) has the advantage of showing the difference between crops with various water stress levels. In this study, CWC was used to assess the sorghum water content, and calculated using Eq. 3.

$$CWC = Plant\ Water\ Status \times LAI \quad (3)$$

Feature selection

Recursive feature elimination (RFE) method

Recursive feature elimination (RFE) is a feature selection technique commonly employed in ML to choose the most relevant variables for a given model. RFE performs by iteratively eliminating variables from a dataset and constructing a model from the remaining features. The significance of each feature is then graded based on how much the model's performance degrades when the feature is eliminated. The process continues until the desired number of features is reached (Granitto et al., 2006). RFE is especially beneficial when the number of variables is large, and there is a risk of model overfitting. By picking the most valuable variables, RFE can increase the model's accuracy and interpretability while lowering the data's dimensionality (Kilincer et al., 2023). Details of the RFE method were given in Ilniyaz et al. (2022)

Dimension reduction using PCA analysis

PCA is a frequently applied method for dimension reduction in data analysis. First introduced by Karl Pearson in 1901 (Pearson, 1901), this technique utilizes mathematical algorithms to transform a collection of interrelated variables into a reduced set of uncorrelated variables known as principal components. DR is used with the PCA method to identify patterns and relationships in high-dimensional data by reducing the number of dimensions while retaining most of the original variability. Principal components are ranked according to the variance they explain in the data (Hotelling, 1933). This allows us to determine which variables contribute most to the variation in the data. By transforming the original data into a lower-dimensional space, PCA can simplify the analysis and visualization of complex data sets (Meier et al., 2022).

Machine learning models

In this study, three commonly used ML methods, namely RF, SVM, and XGBoost, were identified and tested for estimating the CWC of sorghum. These models were selected based on their proven performance and widespread use in machine learning applications. RF can process complex relationships and provide accurate predictions through an ensemble of decision trees (Guan et al., 2022). XGBoost is known for its efficient implementation, parallelization, and ability to process both linear and nonlinear relationships (Gao et al., 2022), while SVM is more effective at processing high-dimensional data and nonlinear relationships (Yildirim et al., 2023). The ML models and processing steps were implemented using the Python programming language, version 3.9, and the sci-kit learn library.

Random forest

RF is a widely used ML algorithm for regression and classification tasks. It is an ensemble method that combines multiple decision trees to make a prediction. In RF, each tree is constructed by randomly selecting a subset of the training data and a subset of the input features. This randomization reduces the risk of overfitting and improves the model's accuracy. The final prediction of the RF model is the average prediction of all individual trees (Küçüktopcu, 2023). In this study, 1000 decision trees were applied to the RF model.

Support vector machine regression

SVM regression is an application of machine learning for regression tasks. SVM is a non-parametric model that identifies the optimal linear or nonlinear boundary between classes of data points. SVM regression aims to identify a hyperplane that maximizes the difference between the predicted and actual values. The margin is the distance between the hyperplane and the data points that are closest to it. The hyperplane is chosen to minimize the difference between the predicted and actual values while increasing the margin. One of the advantages of SVM is its ability to handle high-dimensional data with a relatively small number of samples. In addition, SVM can handle data with nonlinear relationships by using kernel functions to transform the data into a higher-dimensional space where the relationship is linear (Cemek et al., 2022). This study employed the SVM regression method using the radial basis kernel function. The model parameter epsilon and C of the SVM were estimated on all datasets using a grid search method with cross-validation.

eXtreme gradient boosting (XGBoost)

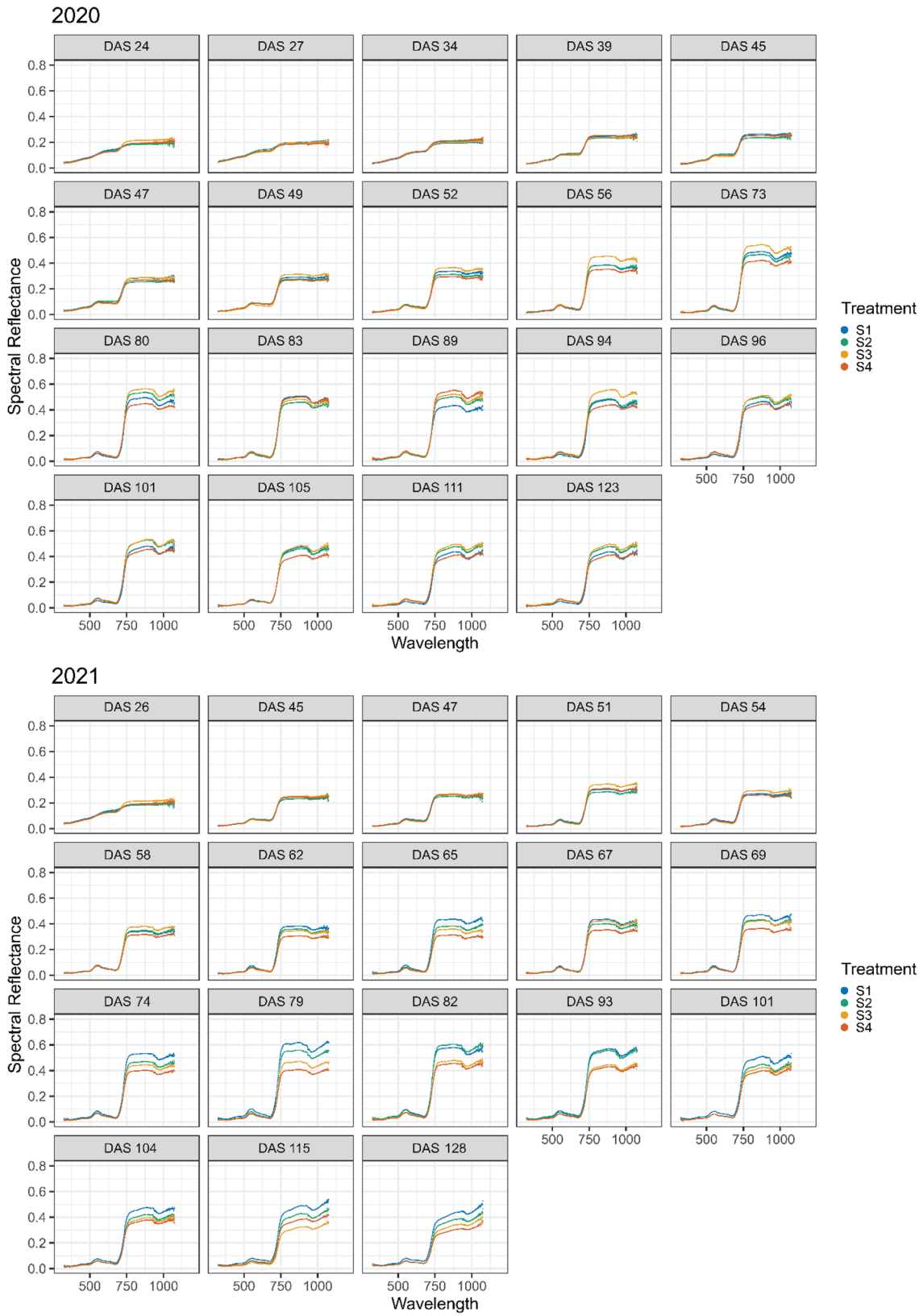
XGBoost is a decision-tree-based ensemble machine learning algorithm. The XGBoost model has a gradient-boosting framework that efficiently handles linear and nonlinear relationships between input variables and output predictions. One significant advantage of XGBoost over other commonly used machine learning approaches, such as RF and SVM, is its ability to produce results faster due to its parallelized implementation and efficient use of hardware resources. XGBoost model can provide the importance of each input variable, allowing users to determine the most important features contributing to the estimation accuracy (Gao et al., 2021).

Vegetation indices

Sorghum CWC was estimated using RF, SVM, and XGBoost with 12 different published spectral VIs that were previously reported to be highly correlated

Table 2 List of spectral vegetation indices used in this study to estimate sorghum CWC

Vegetation index	Formula	Reference
Difference vegetation index (DVI)	$\lambda_{800} - \lambda_{670}$	(Perry Jr and Lautenschlager, 1984)
Enhanced vegetation index (EVI)	$2.5 \times \frac{(\lambda_{800}-\lambda_{680})}{(\lambda_{800}+6*\lambda_{680}-7.5*\lambda_{450}+1)}$	(Buschmann & Nagel, 1993)
Greenness index (G)	$\frac{\lambda_{554}}{\lambda_{677}}$	(Smith et al., 1995)
Optimized soil adjusted vegetation index (OSAVI)	$(1+0.16) \times \frac{(\lambda_{800}+\lambda_{670})}{(\lambda_{800}+\lambda_{670}+0.61)}$	(Smith et al., 1995)
Simple ratio (SR.)	$\frac{\lambda_{800}}{\lambda_{670}}$	(Jordan, 1969)
Modified triangular vegetation index (MTVI)	$1.2 \times (1.2 \times (\lambda_{800} - \lambda_{550}) - 2.5 * (\lambda_{670} - \lambda_{550}))$	(Haboudane et al., 2004)
Normalized difference vegetation index (NDVI)	$\frac{\lambda_{800}-\lambda_{680}}{\lambda_{800}+\lambda_{680}}$	(Lichtenthaler et al., 1996)
Photochemical reflectance index (PRI)	$\frac{\lambda_{531}-\lambda_{570}}{\lambda_{531}+\lambda_{570}}$	(Gamon et al., 1992)
Triangular veg index (TVI)	$0.5 \times (120 \times ((\lambda_{750} - \lambda_{550}) - 200 \times (\lambda_{670} - \lambda_{550}))$	(Gitelson et al., 2002)
Vegetation stress ratio (VS)	$\frac{\lambda_{725}}{\lambda_{702}}$	(White et al., 2008)
Water band index (WBI)	$\frac{\lambda_{950}}{\lambda_{970}}$	(Sims & Gamon, 2003)
Chlorophyll index red edge (CL_Rededge)	$\frac{\lambda_{800}}{\lambda_{720}}$	(Gitelson et al., 2003)


2020
DAS 24
DAS 27
DAS 34
DAS 39
DAS 45
DAS 47
DAS 49
DAS 52
DAS 56
DAS 73
DAS 80
DAS 83
DAS 89
DAS 94
DAS 96
DAS 101
DAS 105
DAS 111
DAS 123
Spectral Reflectance
Wavelength
Treatment
S1
S2
S3
S4
2021
DAS 26
DAS 45
DAS 47
DAS 51
DAS 54
DAS 58
DAS 62
DAS 65
DAS 67
DAS 69
DAS 74
DAS 79
DAS 82
DAS 93
DAS 101
DAS 104
DAS 115
DAS 128
Spectral Reflectance
Wavelength
Treatment
S1
S2
S3
S4

◂**Fig. 2** Temporal changes in sorghum spectral reflectance of each treatment during the 2020 and 2021 growing seasons. S1, S2, S3, and S4 show the average spectral reflectance of sorghum crops grown under different irrigation treatments: S1 is the full irrigation, S2 is the %70 of S1, S3 is the %40 of S1 and S4 is rainfed

with vegetation structure. The VIs were selected based on their effectiveness in estimating CWC. The 12VIs included two band indices (e.g., DVI, G) and three band indices (e.g., EVI, MTVI). The formulas of the VIs used in this study are presented in Table 2.

Model evaluations

During 2 years of sorghum growth, 108 samples were collected. These samples were randomly split into a training set of 75 and a testing set of 33, with a 7:3 ratio. ML regression models were built to estimate CWC using full spectra, feature-selected data, dimensionally reduced data, and spectral VIs. The accuracy of these models was evaluated using three metrics: the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE).

Results and discussion

Effect of different water stress levels on spectral information

Figure 2 shows the average spectral reflectance of sorghum under different water stress levels for the 2020 and 2021 growing seasons, respectively. The spectral reflectance of sorghum varied with growth stage and irrigation treatment, but similar patterns were observed for each treatment. The peak and valley regions of the spectral curves were consistent for each treatment, suggesting that canopy reflection was governed by the same principles regardless of canopy structure changes. This result is in line with Meiyan et al. (2022), who found a strong correlation between the peak and trough positions during maize growth stages despite changes in canopy structure. Similarly, El-Hendawy et al. (2019) reported similar troughs around 680, 1200, 1450, and 1950 nm for different cultivars and salinity stress levels. As the growth stages progressed, its spectral reflectance gradually increased in the near-infrared (NIR) region (700–1100 nm). In addition, noticeable differences in spectral curves due to the irrigation treatments were observed among the treatments after DAS 51 in 2021. Generally, fully irrigated sorghum had higher NIR reflectance values, indicating healthier and more vigorous crops. Previous studies have also reported a positive relationship between NIR reflectance and crop health. For instance, Ren et al. (2022) found a significant correlation between the NIR reflectance and winter wheat growth. Likewise, other studies have shown that higher NIR reflectance values are associated with healthier crops and higher productivity (Köksal, 2011; Zhu et al., 2014). However, in 2020, the S3 treatment exhibited healthier sorghum crops compared to the S1 and S2. This outcome could be possibly attributed to the S3 treatment, despite receiving less water than the other treatments (S1 and S2), having favorable soil conditions that allowed for better water retention and distribution within the root zone. Additionally, the weather conditions during the 2020 season might have been more conducive to the growth and development of the sorghum crops, compensating for the reduced irrigation input in the S3 treatment.

Sorghum LAI measurement results

LAI of sorghum was measured at various developmental stages. Figure 3 presents box plots of the LAI values for each growth phase, while the fundamental statistical parameters of the sorghum LAI values throughout the growing period are displayed in Table 3. The mean LAI value showed a gradual increase throughout the growth stages, with the crop leaves developing rapidly from the establishment to yield formation, achieving a maximum mean LAI value of 8.22 m^2m^{-2}. The mean LAI value decreased slightly to 8.00 m^2m^{-2} during the ripening period. This trend is consistent with previous studies that reported an increase in LAI as the canopy expands and becomes denser during the growing season (Olson et al., 2012; Qiao et al., 2022).

STD standard deviation, *CV* coefficient of variance.

Table 3 shows the range, mean, standard deviation (STD), and coefficient of variance (CV) of sorghum LAI values during the establishment, vegetative, flowering, yield formation and ripening stages. The

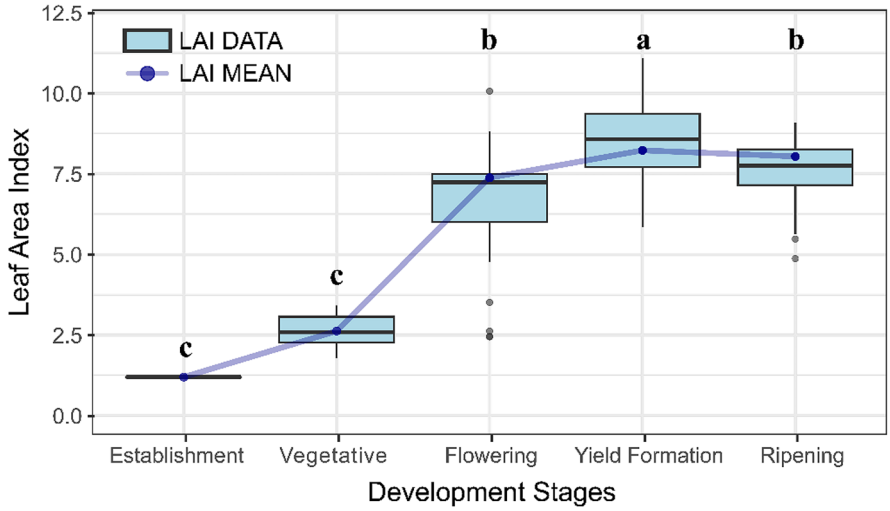


Fig. 3 The changes of LAI and mean LAI values during the sorghum growing periods for all irrigation treatment plots. Statistical significance groups (a, b, c) indicate the results of the Tukey test, highlighting the stages that differ significantly from each other

measurements revealed that the range of LAI values gradually increased from 1.19 to 1.20 (CV = 0.14%) in the establishment period to 4.90–11.19 (CV = 15.14%) in the ripening period. The same results were reported by Bendorf et al. (2022), who found that maximum sorghum LAI values were 9.51 (± 1.45) m^2m^{-2}. A statistical analysis of the sorghum LAI values throughout the different growth periods indicated that the most significant variation in LAI values occurred during the flowering stage of growth (Fig. 3). This finding is consistent with previous studies in wheat by Zhang et al. (2022), in cotton by Ma et al. (2022) and in rice by Yamaguchi et al. (2020), which showed that the highest variability in LAI values occurred during the flowering period.

Performance of the three ML algorithms using full spectra

Table 4 presents the performance of three ML models (RF, SVM, and XGBoost) in estimating the water content of sorghum using full spectra (consisting of 751 spectral wavelengths). Figure 4 compares measured and estimated sorghum CWC values using each ML model. The results of this study demonstrated that full spectra data could effectively estimate the CWC of sorghum. The highest R^2 and lowest RMSE values were obtained using the XGBoost model, with an R^2 of 0.95 and an RMSE of 33.29. RF was the second-best model, with R^2 and RMSE of 0.92 and 42.16, respectively. The SVM model had the lowest performance, with an R^2 of 0.40 and an RMSE of 132.44. Based on these results, the XGBoost model was the most effective ML algorithm for determining sorghum CWC using full spectra hyperspectral measurements, followed by RF and SVM. The superior performance of the XGBoost compared to RF and SVM models can be attributed to its adaptability algorithm, which belongs to the advanced gradient boosting system. XGBoost can correct residual inaccuracies, unlike RF models, by building upon the previous tree and generating a new set (Chen & Guestrin, 2016; Friedman, 2002). These findings align with a previous study that reported XGBoost as a robust algorithm for estimating CWC using remote sensing data (Zhou et al., 2022). On the other hand, SVM was the most ineffective ML model among the models used in this study. This could be explained by the number of samples used in this study. Hoi et al. (2009) stated that SVM performs poorly on small-size datasets. Additionally, SVM is highly sensitive to dataset outliers. This could be the potential reason for the poor SVM performance (Guo et al., 2023).

Table 3 Variation of LAI values during the sorghum growth periods

Growth periods	LAI Dataset			
	Range	**Mean**	**STD**	**CV (%)**
Establishment	1.19–1.20	1.19	0.001	0.14
Vegetative	1.79–3.42	2.62	0.53	20.22
Flowering	2.45–11.04	7.38	1.80	24.48
Yield Formation	4.87–11.09	8.22	1.29	15.77
Ripening	4.90–11.10	8.00	1.21	15.14

Table 4 The performance of ML model used in this study to estimate sorghum CWC using full spectra for training, testing, and whole dataset

Models	Training set			Test set			All dataset		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
RF	0.95	34.09	26.74	0.82	64.68	47.51	0.92	42.16	30.97
SVM	0.40	129.22	94.97	0.37	144.33	110.71	0.40	132.44	98.18
XGBoost	0.98	16.50	10.76	0.83	66.15	47.55	0.95	33.29	18.26

Feature selection using RFE

Three machine learning-based regression models (RF, SVM, and XGBoost) were optimized by applying the RFE feature selection method. The maximum number of selected variables was set to 30, and selected variables were exhibited in Table 5. Figure 5 illustrates that the RMSE gradually decreased with an increasing number of added wavelength bands until minimum RMSE values were obtained. The optimized models were established using reflectance values of these selected wavelengths. Among the ML models, the RF model, with 19 features selected, achieved the highest R^2 value of 0.90 and the lowest RMSE value of 56.05, followed by the XGBoost model, using 22 features, achieved an R^2 value of 0.86 and an RMSE value of 66.55. Using only three selected features, the SVM model had the lowest performance, with an R^2 value of 0.50 and an RMSE value of 127.64 (Fig. 5). According to the results, the XGBoost model demonstrated less accuracy than the RF model. This difference in performance could be attributed to the inherent differences in the XGBoost and RF algorithms, dataset characteristics, and the specific interactions and relationships captured by each algorithm during the feature selection process (Li et al., 2023). Overall, the results indicate that ML algorithms based on selected spectral bands performed comparably to those that used the whole spectrum, which is consistent with previous studies (Elsherbiny et al., 2021; Meiyan et al., 2022; Ndlovu et al., 2021; Zhao et al., 2022).

Estimation of canopy water content using dimension reduction via *PCA*

The PCA method was performed on the proximal hyperspectral data to obtain a lower-dimensional subspace. The number of components was selected based

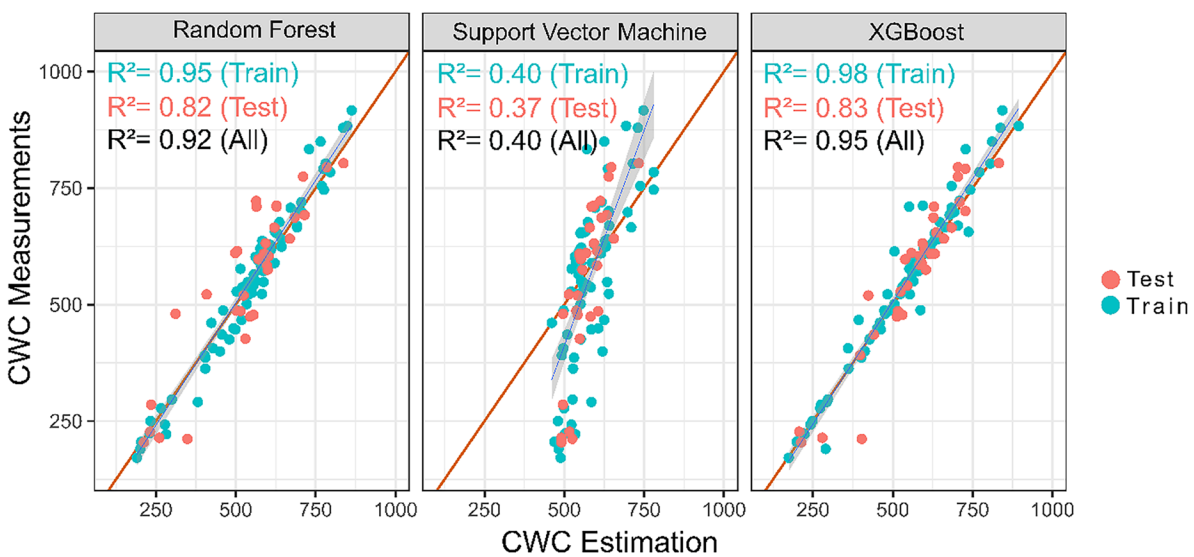


Fig. 4 Comparison of measured and estimated crop water content using RF, SVM, and XGBoost models. The x-axis represents the estimation of sorghum CWC values, while y-axis represents the measured CWC values obtained through direct field measurements. Green and Red color indicate the train and the test points, respectively. The R^2 value of train, test and all points used in ML model are shown in the figure

Table 5 Specific wavelengths selected by each ML model through the RFE method for estimating crop water content

Models	Number of wavelengths	Selected spectral reflectance wavelengths
RF	19	344, 347, 353, 367, 381, 394, 428, 560, 902, 916, 924, 930, 959, 963, 972, 981, 990, 1041, 1057
SVM	3	963, 972, 981
XGBoost	22	344, 347, 353, 367, 381, 394, 428, 560, 763, 902, 916, 924, 930, 947, 959, 963, 972, 981, 990, 1041, 1057, 1067

on the cumulative variance ratio, where only those components with a cumulative variance greater than 95% were selected. A higher threshold would have required more components and increased the complexity and redundancy of the data. A lower threshold would have resulted in more information loss and reduced accuracy and reliability of the data. The choice of 95% as the optimal threshold is consistent with previous studies that used PCA to process hyperspectral data for various applications (Salata et al., 2022; Winder et al., 2009). Figure 6 shows the cumulative contribution rate of each component calculated with PCA analysis. Based on the DR via PCA analysis, the number of components in the hyperspectral data was set to 4. The first component had the highest explaining variance ratio of 91.12%, while the first four components, up to 95.76%, explained the original hyperspectral data, indicating that only 4.24% of the hyperspectral information was not explained.

RF, XGBoost, and SVM models were developed using dimensional reduced proximal hyperspectral data obtained through PCA analysis. The performance of these models was evaluated using a Taylor diagram (Fig. 7), where the purple square represents the reference standard. The Taylor diagram shows that the XGBoost model has the lowest RMSE (45.77) values among the three ML models used in this study, indicating its superior reliability in estimating sorghum CWC using dimensional reduced proximal hyperspectral data. Therefore, XGBoost could be considered a reliable and accurate tool for advancing precision agriculture by estimating CWC. This result is consistent with a previous study that reported the advantage of XGBoost over the other ML methods for hyperspectral classification tasks (Samat et al., 2020).

Performance of published vegetation indices

In this study, three ML models (RF, SVM, and XGBoost) were employed to estimate the CWC of sorghum using 12 different spectral VIs, including

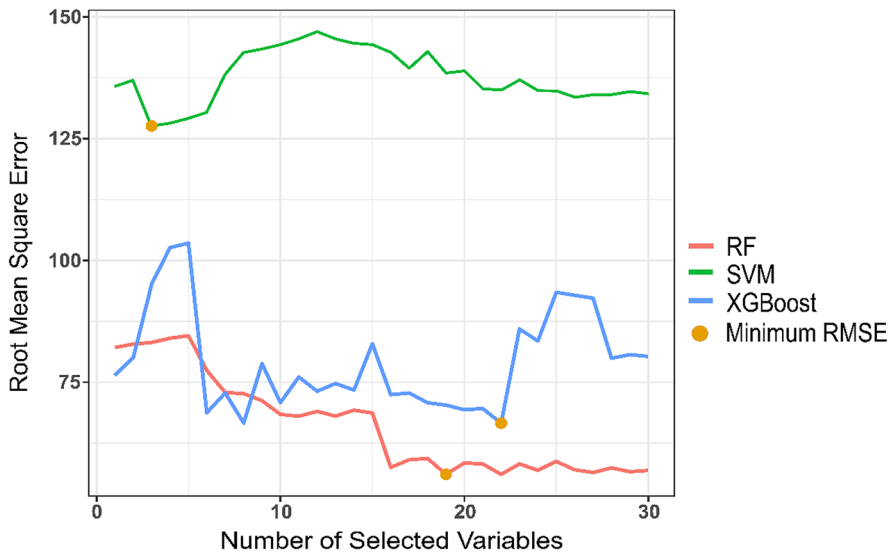


Fig. 5 Relationship between the number of selected variables and root mean square error values. Each data point on the line plot represents a specific number of selected variables and its corresponding RMSE value. The colors used in the lines represent the ML models, while the orange points indicate the achieved minimum RMSE value

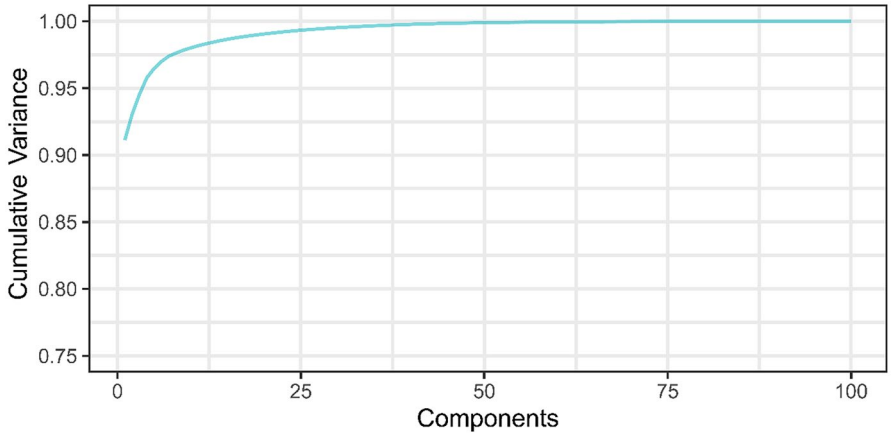


Fig. 6 Cumulative variance rate of each component. The *x*-axis represents the components, numbered sequentially, while the y-axis represents the cumulative variance rate

DVI, EVI, G, OSAVI, SR, MTVI, NDVI, PRI, TVI, VS, WBI, and CL_Rededge. Figure 8 and Table 6 show the performance of ML models using different spectral VIs. The results of the SVM analysis indicated that most VIs exhibited similar performance when applied to both training and test sets. Among the VIs used in this study, EVI had the highest R^2 values of 0.53 and 0.51 for training and test sets, respectively, while OSAVI followed closely with R^2 values of 0.44 and 0.45, respectively. However, G, PRI and WBI performed poorly with near-zero R^2 values.

In contrast to the SVM algorithm, RF and XGBoost ML algorithms performed significantly better on all VIs with high R^2 values except for G and VS. Among the VIs used in this study, CL_Rededge and EVI showed the best performance on both algorithms for the test sets. Overall, XGBoost performs slightly better than RF based on R^2 values, with CL_Rededge, MTVI, and EVI being the best-performing VIs, while CL_Rededge, EVI, and OSAVI are the best-performing VIs in RF. However, the RMSE and MAE values for some indices are still quite high, suggesting that there is still room for improvement. The results suggest that the choice of VI can significantly impact the reliability of sorghum CWC estimation. In particular, VIs that include the blue and red-edge region of the electromagnetic spectrum (e.g., CL_Rededge) tend to perform better than those that rely only on the visible and near-infrared regions (e.g., NDVI and TVI). These findings align with previous research by Zhang et al. (2021), which identified the water-associated spectral bands. Gómez-Candón et al. (2021) stated that the red-edge region is mainly influenced by canopy structure and chlorophyll content, which is highly correlated with CWC. Also, Wang et al. (2015) indicated that red-edge-based VIs could indicate water stress effectively.

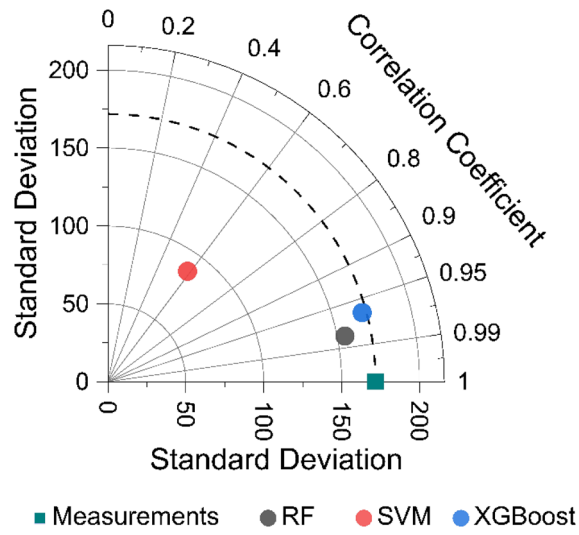


Fig. 7 Taylor Diagram results for sorghum CWC estimation with RF, SVM and XGBoost ML models used in this study

Fig. 8 Comparison of machine learning performance for RF, SVM, and XGBoost using different vegetation index and radar chart visualization of R^2, RMSE, and MAE metrics

Table 6 Comparison of ML model (RF, SVMand XGBoost) performance on vegetation indices for train and test data

ML Model	Vegetation Index	Train			Test		
		R^2	RMSE	MAE	R^2	RMSE	MAE
RF	DVI	0.91	52.54	43.49	0.35	135.3	98.02
	EVI	0.9	54	42.95	0.58	109.2	79.17
	G	0.76	85.26	68.15	0.19	150.89	123.74
	OSAVI	0.9	54.15	42.85	0.52	116.8	84.19
	SR	0.91	50.85	39.54	0.08	161.43	115.65
	MTVI	0.88	59.81	48.73	0.48	121.24	89.44
	NDVI	0.9	53.72	41.2	0.09	175.6	116.63
	PRI	0.66	99.88	76.7	0.32	138.95	111.6
	TVI	0.87	61.12	50.58	0.46	123.42	97.11
	VS	0.8	76.84	58.86	0.08	174.43	131.47
	WBI	0.67	98.85	73.31	0.01	168.23	129.19
	CL_Rededge	0.69	96.36	68.57	**0.66**	**97.98**	**77.7**
SVM	DVI	0.48	123.84	95.59	0.46	123.84	98.41
	EVI	0.53	118.59	91.12	**0.51**	**117.99**	**93.81**
	G	0.21	153.1	114.97	0.01	169.1	137.92
	OSAVI	0.44	129.04	98.22	0.45	124.36	97.96
	SR	0.23	151.57	114.36	0.17	153.03	124.37
	MTVI	0.44	128.41	95.63	0.39	130.84	98.88
	NDVI	0.43	129.8	98.31	0.42	128.06	104.72
	PRI	0.02	173.89	130.44	0.02	170.07	130.19
	TVI	0.4	134.04	102.18	0.33	137.08	104.48
	VS	0.26	148.22	111.81	0.05	163.53	131.94
	WBI	0.02	174.31	130.18	0.04	171.47	133.18
	CL_Rededge	0.02	170.24	130.71	0.03	165.16	130.68
XGBoost	DVI	0.76	84.61	62.03	0.47	121.75	99.04
	EVI	0.81	74.33	56.33	0.54	113.54	90.45
	G	0.73	89.17	69.66	0.23	147.47	121.42
	OSAVI	0.87	61.37	46.00	0.47	122.04	87.26
	SR	0.78	81.45	60.06	0.18	152.25	113.44
	MTVI	0.75	86.28	67.37	0.59	106.96	88.78
	NDVI	0.67	99.66	75.24	0.19	151.40	120.24
	PRI	0.61	108.28	81.86	0.29	141.56	118.00
	TVI	0.77	82.06	64.21	0.49	119.52	95.86
	VS	0.66	99.81	75.05	0.02	168.00	132.71
	WBI	0.55	115.09	85.21	0.19	150.95	120.76
	CL_Rededge	0.68	97.90	67.98	**0.72**	**88.93**	**67.11**

Conclusion

This study investigated the effects of different water stress levels on spectral information, LAI, and the performance of three ML algorithms in estimating the CWC of sorghum. The results show that sorghum's spectral reflectance varies with growth stage and irrigation treatment, but consistent patterns are observed for each treatment. Moreover, fully irrigated sorghum crops had higher near-infrared (NIR) reflectance values, indicating better health and more vigorous growth. The LAI of sorghum gradually increased throughout the growth stages, with the most significant variation observed during the flowering stage.

XGBoost was the most effective ML algorithm for determining sorghum CWC using full spectra

hyperspectral measurements, followed by RF and SVM. The RFE method was used to select the optimal spectral reflectance wavelengths for the models, and PCA was used to reduce the dimensionality of the hyperspectral data. The results revealed that the RF model achieved the highest R^2 and the lowest RMSE values, while the XGBoost model demonstrated superior accuracy and reliability in estimating CWC using dimensionality-reduced hyperspectral data. These findings have significant implications for precision agriculture, as accurate and reliable estimates of CWC can help farmers optimize irrigation and fertilizer applications, leading to improved crop yields and resource efficiency.

The study also highlights that the performance of ML models for estimating sorghum CWC varies significantly depending on the VI used. While some VIs performed poorly, such as the NDVI and the MSAVI, others, such as the CL_Rededge and EVI, performed better. Furthermore, XGBoost performed slightly better than RF and SVM among the three ML models tested. These findings suggest that the choice of VI is crucial for accurate CWC estimation, and VIs that include the red-edge region tend to perform better. However, further improvements are necessary to increase the accuracy of sorghum CWC estimation.

In summary, this study provides valuable insights into the effects of water stress levels on spectral information, LAI, and the performance of ML algorithms in estimating the CWC of sorghum. The findings have important implications for precision agriculture, and future research can further explore the potential of ML algorithms and hyperspectral data for estimating other crop parameters to support sustainable and efficient agricultural practices.

Author contribution Emre Tunca, designed the study, data collection, analyzed the data, and wrote the manuscript. Eyüp Selim Köksal, conceived the research project, participated in the data collection and analysis, and provided critical feedback on the manuscript. Elif Öztürk, contributed to the data collection. Hasan Akay and Sakine Çetin Taner assisted in the interpretation of the results. All authors reviewed and approved the final version of the manuscript.

Funding This study was supported by The Scientific and Technological Research Council of Turkey (118O831).

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Allamine, H. M., Buyuktas, D., Karaca, C., Aydinsakir, K., & Erdurmus, C. (2023). Effect of regulated deficit irrigation on productivity, evapotranspiration and quality of grain sorghum. *Irrigation Science*, 1–17. https://doi.org/10.1007/s00271-022-00844-5

Arslan, H., Tasan, M., Yildirim, D., Koksal, E. S., & Cemek, B. (2014). Predicting field capacity, wilting point, and the other physical properties of soils using hyperspectral reflectance spectroscopy: Two different statistical approaches. *Environmental Monitoring and Assessment, 186*, 5077–5088. https://doi.org/10.1007/s10661-014-3761-2

Aydinsakir, K., Buyuktas, D., Dinç, N., Erdurmus, C., Bayram, E., & Yegin, A. B. (2021). Yield and bioethanol productivity of sorghum under surface and subsurface drip irrigation. *Agricultural Water Management, 243*, 106452. https://doi.org/10.1016/j.agwat.2020.106452

Beltrán, N. H., Duarte-Mermoud, M., Salah, S., Bustos, M., Peña-Neira, A. I., Loyola, E., & Jalocha, J. (2005). Feature selection algorithms using Chilean wine chromatograms as examples. *Journal of Food Engineering, 67*, 483–490. https://doi.org/10.1016/j.jfoodeng.2004.05.015

Bendorf, J., Heaton, E., Hartman, T., Aslan-Sungur, G., & VanLoocke, A. (2022). Agroecosystem model simulations reveal spatial variability in relative productivity in biomass sorghum and maize in Iowa, USA. *GCB Bioenergy, 14*, 1336–1360. https://doi.org/10.1111/gcbb.13004

Buschmann, C., & Nagel, E. (1993). In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing, 14*, 711–722. https://doi.org/10.1080/01431169308904370

Cemek, B., Arslan, H., Küçüktopcu, E., & Simsek, H. (2022). Comparative analysis of machine learning techniques for estimating groundwater deuterium and oxygen-18 isotopes. *Stochastic Environmental Research and Risk Assessment, 36*, 4271–4285. https://doi.org/10.1007/s00477-022-02262-7

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* (pp. 785–794)

Chivasa, W., Mutanga, O., & Biradar, C. (2020). UAV-based multispectral phenotyping for disease resistance to accelerate crop improvement under changing climate conditions. *Remote Sensing, 12*, 2445. https://doi.org/10.3390/rs12152445

Clevers, J. G., Kooistra, L., & Schaepman, M. E. (2010). Estimating canopy water content using hyperspectral remote sensing data. *International Journal of Applied Earth Observation and Geoinformation, 12*, 119–125. https://doi.org/10.1016/j.jag.2010.01.007

Colombo, R., Meroni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C., & Panigada, C. (2008). Estimation of

leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. *Remote Sensing of Environment, 112*, 1820–1834. https://doi.org/10.1016/j.rse.2007.09.005

Das, B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Viswanathan, C., Sehgal, V. K., & Gupta, V. K. (2021). Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 247*, 119104. https://doi.org/10.1016/j.saa.2020.119104

El-Hendawy, S. E., Al-Suhaibani, N. A., Hassan, W. M., Dewir, Y. H., Elsayed, S., Al-Ashkar, I., Abdella, K. A., & Schmidhalter, U. (2019). Evaluation of wavelengths and spectral reflectance indices for high-throughput assessment of growth, water relations and ion contents of wheat irrigated with saline water. *Agricultural Water Management, 212*, 358–377. https://doi.org/10.1016/j.agwat.2018.09.009

Elsherbiny, O., Fan, Y., Zhou, L., & Qiu, Z. (2021). Fusion of feature selection methods and regression algorithms for predicting the canopy water content of rice based on hyperspectral data. *Agriculture, 11*, 51.

Eshkabilov, S., Stenger, J., Knutson, E. N., Küçüktopcu, E., Simsek, H., & Lee, C. W. (2022). Hyperspectral image data and waveband indexing methods to estimate nutrient concentration on lettuce (Lactuca sativa L.) cultivars. *Sensors, 22*, 8158. https://doi.org/10.3390/s22218158

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment, 41*, 35–44. https://doi.org/10.1016/0034-4257(92)90059-S

Gao, R., Torres-Rua, A., Nassar, A., Alfieri, J., Aboutalebi, M., Hipps, L., Ortiz, N. B., McElrone, A. J., Coopmans, C., & Kustas, W. (2021). Evapotranspiration partitioning assessment using a machine-learning-based leaf area index and the two-source energy balance model with sUAV information. In, *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping VI* (pp. 106–129): SPIE

Gao, R., Torres-Rua, A. F., Aboutalebi, M., White, W. A., Anderson, M., Kustas, W. P., Agam, N., Alsina, M. M., Alfieri, J., & Hipps, L. (2022). LAI estimation across California vineyards using sUAS multi-seasonal multi-spectral, thermal, and elevation information and machine learning. *Irrigation Science, 40*, 731–759. https://doi.org/10.1007/s00271-022-00776-0

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology, 160*, 271–282.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment, 80*, 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gómez-Candón, D., Bellvert, J., & Royo, C. (2021). Performance of the two-source energy balance (TSEB) model as a tool for monitoring the response of durum wheat to drought by high-throughput field phenotyping. *Frontiers in plant science, 12*, 658357. https://doi.org/10.3389/fpls.2021.658357

Granitto, P. M., Furlanello, C., Biasioli, F., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems, 83*, 83–90. https://doi.org/10.1016/j.chemolab.2006.01.007

Guan, Y., Grote, K., Schott, J., & Leverett, K. (2022). Prediction of soil water content and electrical conductivity using random forest methods with UAV multispectral and ground-coupled geophysical data. *Remote Sensing, 14*, 1023. https://doi.org/10.3390/rs14041023

Guo, Q., Li, J., Zhou, F., Li, G., & Lin, J. (2023). An open-set fault diagnosis framework for MMCs based on optimized temporal convolutional network. *Applied Soft Computing, 133*, 109959. https://doi.org/10.1016/j.asoc.2022.109959

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment, 90*, 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Hasanlou, M., & Samadzadegan, F. (2012). Comparative study of intrinsic dimensionality estimation and dimension reduction techniques on hyperspectral images using K-NN classifier. *IEEE Geoscience and Remote Sensing Letters, 9*, 1046–1050. https://doi.org/10.1109/LGRS.2012.2189547

Hoi, S. C., Jin, R., Zhu, J., & Lyu, M. R. (2009). Semisupervised svm batch mode active learning with applications to image retrieval. *ACM Transactions on Information Systems (TOIS), 27*, 1–29. https://doi.org/10.1145/1508850.1508854

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417. https://doi.org/10.1037/h0071325

Ilniyaz, O., Kurban, A., & Du, Q. (2022). Leaf area index estimation of pergola-trained vineyards in arid regions based on UAV RGB and multispectral data using machine learning methods. *Remote Sensing, 14*, 415. https://doi.org/10.3390/rs14020415

Irik, H. A., & Kirnak, H. (2022). Evaluation of spectral vegetation indices for drip irrigated pumpkin seed under semi-arid conditions. *Arabian Journal of Geosciences, 15*, 861.

Ji, H.-Y., Wang, X.-Z., He, Y.-L., & Li, W.-L. (2014). A study on relationships between heuristics and optimal cuts in decision tree induction. *Computers & Electrical Engineering, 40*, 1429–1438. https://doi.org/10.1016/j.compeleceng.2013.11.030

Jin, X., Shi, C., Yu, C. Y., Yamada, T., & Sacks, E. J. (2017). Determination of leaf water content by visible and near-infrared spectrometry and multivariate calibration in Miscanthus. *Frontiers in Plant Science, 8*, 721. https://doi.org/10.3389/fpls.2017.00721

Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology, 50*, 663–666. https://doi.org/10.2307/1936256

Kilincer, I. F., Ertam, F., Sengur, A., Tan, R.-S., & Acharya, U. R. (2023). Automated detection of cybersecurity attacks

in healthcare systems with recursive feature elimination and multilayer perceptron optimization. *Biocybernetics and Biomedical Engineering, 43*, 30–41. https://doi.org/10.1016/j.bbe.2022.11.005

Köksal, E. S. (2011). Hyperspectral reflectance data processing through cluster and principal component analysis for estimating irrigation and yield related indicators. *Agricultural Water Management, 98*, 1317–1328. https://doi.org/10.1016/j.agwat.2011.03.014

Küçüktopcu, E. (2023). Comparative analysis of data-driven techniques to predict heating and cooling energy requirements of poultry buildings. *Buildings, 13*, 142. https://doi.org/10.3390/buildings13010142

Li, Y.-f., Xu, Z.-h., Hao, Z.-b., Yao, X., Zhang, Q., Huang, X.-y., Li, B., He, A.-q., Li, Z.-l., & Guo, X.-y. (2023). A comparative study of the performances of joint RFE with machine learning algorithms for extracting Moso bamboo (Phyllostachys pubescens) forest based on UAV hyperspectral images. *Geocarto International*, 1–21. https://doi.org/10.1080/10106049.2023.2207550

Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F., & Miehe, J. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology, 148*, 599–612. https://doi.org/10.1016/S0176-1617(96)80081-2

Ma, Y., Zhang, Q., Yi, X., Ma, L., Zhang, L., Huang, C., Zhang, Z., & Lv, X. (2022). Estimation of cotton leaf area index (LAI) based on spectral transformation and vegetation index. *Remote Sensing, 14*, 136. https://doi.org/10.3390/rs14010136

Meier, M., Kittle, J. D., & Yee, X. C. (2022). Supervised dimension reduction for optical vapor sensing. *RSC Advances, 12*, 9579–9586. https://doi.org/10.1039/D1RA08774F

Meiyan, S., Qizhou, D., ShuaiPeng, F., Xiaohong, Y., Jinyu, Z., Lei, M., Baoguo, L., & Yuntao, M. (2022). Improved estimation of canopy water status in maize using UAV-based digital and hyperspectral images. *Computers and Electronics in Agriculture, 197*, 106982. https://doi.org/10.1016/j.compag.2022.106982

Ndlovu, H. S., Odindi, J., Sibanda, M., Mutanga, O., Clulow, A., Chimonyo, V. G., & Mabhaudhi, T. (2021). A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (UAV)-based proximal and remotely sensed data. *Remote Sensing, 13*, 4091. https://doi.org/10.3390/rs13204091

Olson, S. N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B. A., Zhang, Y., Hall, S., Packer, D., & Mullet, J. (2012). High biomass yield energy sorghum: Developing a genetic model for C4 grass bioenergy crops. *Biofuels, Bioproducts and Biorefining, 6*, 640–655. https://doi.org/10.1002/bbb.1357

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*, 559–572. https://doi.org/10.1080/14786440109462720

Perry, C. R., Jr., & Lautenschlager, L. F. (1984). Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment, 14*, 169–182. https://doi.org/10.1016/0034-4257(84)90013-0

Qiao, L., Zhao, R., Tang, W., An, L., Sun, H., Li, M., Wang, N., Liu, Y., & Liu, G. (2022). Estimating maize LAI by exploring deep features of vegetation index map from UAV multispectral images. *Field Crops Research, 289*, 108739. https://doi.org/10.1016/j.fcr.2022.108739

Ren, S., Guo, B., Wang, Z., Wang, J., Fang, Q., & Wang, J. (2022). Optimized spectral index models for accurately retrieving soil moisture (SM) of winter wheat under water stress. *Agricultural Water Management, 261*, 107333, https://doi.org/10.1016/j.agwat.2021.107333

Salata, S., Ozkavaf-Senalp, S., Velibeyoğlu, K., & Elburz, Z. (2022). Land suitability analysis for vineyard cultivation in the izmir metropolitan area. *Land, 11*, 416.

Samat, A., Li, E., Wang, W., Liu, S., Lin, C., & Abuduwaili, J. (2020). Meta-XGBoost for hyperspectral image classification using extended MSER-guided morphological profiles. *Remote Sensing, 12*, 1973. https://doi.org/10.3390/rs12121973

Shi, B., Yuan, Y., Zhuang, T., Xu, X., Schmidhalter, U., Ata-UI-Karim, S.T., Zhao, B., Liu, X., Tian, Y., & Zhu, Y. (2022). Improving water status prediction of winter wheat using multi-source data with machine learning. *European Journal of Agronomy, 139*, 126548. https://doi.org/10.1016/j.eja.2022.126548

Sibanda, M., Onisimo, M., Dube, T., & Mabhaudhi, T. (2021). Quantitative assessment of grassland foliar moisture parameters as an inference on rangeland condition in the mesic rangelands of southern Africa. *International Journal of Remote Sensing, 42*, 1474–1491. https://doi.org/10.1080/01431161.2020.1832282

Sims, D. A., & Gamon, J. A. (2003). Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: A comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sensing of Environment, 84*, 526–537. https://doi.org/10.1016/S0034-4257(02)00151-7

Smith, R., Adams, J., Stephens, D., & Hick, P. (1995). Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Australian Journal of Agricultural Research, 46*, 113–125.

Tunca, E., Köksal, E. S., Torres-Rua, A., Kustas, W. P., & Nieto, H. (2022). Estimation of bell pepper evapotranspiration using two-source energy balance model based on high-resolution thermal and visible imagery from unmanned aerial vehicles. *Journal of Applied Remote Sensing, 16*, 022204–022204. https://doi.org/10.1117/1.JRS.16.022204

Wang, X., Zhao, C., Guo, N., Li, Y., Jian, S., & Yu, K. (2015). Determining the canopy water stress for spring wheat using canopy hyperspectral reflectance data in loess plateau semiarid regions. *Spectroscopy Letters, 48*, 492–498. https://doi.org/10.1080/00387010.2014.909495

Wanga, M. A., Shimelis, H., & Mengistu, G. (2022). Sorghum production in northern namibia: Farmers' perceived constraints and trait preferences. *Sustainability, 14*, 10266. https://doi.org/10.3390/su141610266

White, D., Williams, M., & Barr, S. (2008). Detecting subsurface soil disturbance using hyperspectral first derivative band rations of associated vegetation stress. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 27*, 243–248.

Winder, S., Hua, G., & Brown, M. (2009). Picking the best daisy. In, *2009 IEEE conference on computer vision and pattern recognition* (pp. 178–185): IEEE

Xu, X., Lu, J., Zhang, N., Yang, T., He, J., Yao, X., Cheng, T., Zhu, Y., Cao, W., & Tian, Y. (2019). Inversion of rice

canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. *ISPRS Journal of Photogrammetry and Remote Sensing, 150*, 185–196. https://doi.org/10.1016/j.isprsjprs.2019.02.013

Yamaguchi, T., Tanaka, Y., Imachi, Y., Yamashita, M., & Katsura, K. (2020). Feasibility of combining deep learning and RGB images obtained by unmanned aerial vehicle for leaf area index estimation in rice. *Remote Sensing, 13*, 84. https://doi.org/10.3390/rs13010084

Yildirim, D., Küçüktopcu, E., Cemek, B., & Simsek, H. (2023). Comparison of machine learning techniques and spatial distribution of daily reference evapotranspiration in Türkiye. *Applied Water Science, 13*, 107. https://doi.org/10.1007/s13201-023-01912-7

Yue, J., Feng, H., Jin, X., Yuan, H., Li, Z., Zhou, C., Yang, G., & Tian, Q. (2018). A comparison of crop parameters estimation using images from UAV-mounted snapshot hyperspectral sensor and high-definition digital camera. *Remote Sensing, 10*, 1138. https://doi.org/10.3390/rs10071138

Zhang, J., Zhang, W., Xiong, S., Song, Z., Tian, W., Shi, L., & Ma, X. (2021). Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content. *Plant Methods, 17*, 1–14. https://doi.org/10.1186/s13007-021-00737-2

Zhang, X., Zhang, K., Wu, S., Shi, H., Sun, Y., Zhao, Y., Fu, E., Chen, S., Bian, C., & Ban, W. (2022). An investigation of winter wheat leaf area index fitting model using spectral and canopy height model data from unmanned aerial vehicle imagery. *Remote Sensing, 14*, 5087. https://doi.org/10.3390/rs14205087

Zhao, J., Li, H., Chen, C., Pang, Y., & Zhu, X. (2022). Detection of water content in lettuce canopies based on hyperspectral imaging technology under outdoor conditions. *Agriculture, 12*, 1796. https://doi.org/10.3390/agriculture12111796

Zhou, X., Yang, W., Luo, K., & Tang, X. (2022). Estimation of aboveground vegetation water storage in natural forests in jiuzhaigou national nature reserve of China using machine learning and the combination of landsat 8 and sentinel-2 data. *Forests, 13*, 507. https://doi.org/10.3390/f13040507

Zhu, L., Chen, Z., Wang, J., Ding, J., Yu, Y., Li, J., Xiao, N., Jiang, L., Zheng, Y., & Rimmington, G. M. (2014). Monitoring plant response to phenanthrene using the red edge of canopy hyperspectral reflectance. *Marine pollution bulletin, 86*, 332–341. https://doi.org/10.1016/j.marpolbul.2014.06.046