ISSN 1807-1929
Revista Brasileira de Engenharia Agrícola e Ambiental
Brazilian Journal of Agricultural and Environmental Engineering
v.29, n.4, e285794, 2025
Campina Grande, PB – http://www.agriambi.com.br – http://www.scielo.br/rbeaa

DOI: http://dx.doi.org/10.1590/1807-1929/agriambi.v29n4e285794

Original Article

Relationship between solar radiation and meteorological variables in predictive models for crop yields[1]

Relação entre a radiação solar e variáveis meteorológicas em modelo preditivo de produtividade agrícola

Abdelkarem M. Adam[2*] & Yuan Zheng[3]

[1] Research developed at College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China
[2] College of Water Conservancy and Hydropower Engineering/Hohai University, Nanjing, China
[3] Renewable Energy Power Generation Engineering Research/School of Water Resources and Hydropower/Hohai University, Nanjing, China

HIGHLIGHTS:

From 2011 to 2021, the correlation between solar radiation and yield was inverse for sorghum and positive for sesame.
A significant upward trend in solar radiation was observed from 2011 to 2021, especially in August and September.
The Mann-Kendall test revealed a significant rise in solar radiation from 2011-2021, with positive Z values and $p < 0.05$.

ABSTRACT: Knowledge of the complicated correlation between meteorological variables and crop yield is crucial for food security and agricultural sustainability. This study aimed to investigate how incident solar radiation has affected crop production in the Gadarif region of Sudan over the last 41 years. Using a predictive framework, trends in annual incident solar radiation and temporal variations during sorghum and sesame growing seasons were examined and machine learning (ML) with Extreme Gradient Boosting (XGBoost), Boosted Regression Forest (BRF), and K-Nearest Neighbors (K-NN) was used to predict crop yield. Significant relationships between incident solar radiation indicators and crop yields were identified via detrending approaches and correlation analyses. Results indicate a significant inverse correlation between solar radiation and sorghum yield, and a positive correlation between sesame yield and solar radiation. For both sorghum and sesame yield, K-NN was the most accurate model, demonstrating the significance of incident solar radiation and temperature in predicting crop yield. These findings highlight the potential of ML to improve agricultural forecasting models and inform adaptive agricultural practices in the region. In general, this study provides valuable insights into the dynamic relationship between incident solar radiation and crop yield, emphasizing the importance of considering meteorological factors in agricultural planning and management.

Key words: solar radiation, crop yield, Gadarif region, Sudan, machine learning

RESUMO: O conhecimento da complicada correlação entre as variáveis meteorológicas e o rendimento das culturas é crucial para a segurança alimentar e a sustentabilidade agrícola. Este estudo está centrado na investigação de como a radiação solar incidente afetou a produção agrícola na região de Gadarif, no Sudão, nos últimos quarenta anos. Usando uma estrutura preditiva, a pesquisa avalia tendências recentes na radiação solar incidente anual, examina variações temporais durante as estações de cultivo de sorgo e gergelim e utiliza técnicas de aprendizado de máquina para prever o rendimento das culturas. Além disso, ML, incluindo Extreme Gradient Boosting (XGBoost), Boosted Regression Forest (BRF) e K-Nearest Neighbours (K-NN), foram empregados para previsão de rendimento. Através de abordagens de redução de tendências e análises de correlação, foram identificadas relações significativas entre os indicadores de radiação solar incidente e o rendimento das culturas. Os resultados indicam uma correlação inversa substancial entre a radiação solar e a produção de sorgo, enquanto a produção de gergelim demonstra uma correlação positiva com a radiação solar. Tanto para o rendimento do sorgo como do gergelim, o K-NN surge como o modelo mais preciso, mostrando a importância da radiação solar incidente e da temperatura na previsão do rendimento das culturas. Estas descobertas destacam o potencial da aprendizagem de máquina para melhorar os modelos de previsão agrícola e informar as práticas agrícolas adaptativas na região. Em geral, este estudo fornece informações valiosas sobre a relação dinâmica entre a radiação solar incidente e o rendimento das culturas, enfatizando a importância de considerar fatores meteorológicos no planeamento e gestão agrícola.

Palavras-chave: radiação solar, rendimento agrícola, região de Gadarif, Sudão, aprendizado de máquina

• Ref. 285794 – Received 22 Apr, 2024
* Corresponding author - E-mail: abdoadam7878@gmail.com
• Accepted 14 Oct, 2024 • Published 29 Oct, 2024
Editors: Toshik Iarley da Silva & Carlos Alberto Vieira de Azevedo

Introduction

Agriculture is the cornerstone of Sudan's economy, with crops such as sesame (*Sesamum indicum*) and sorghum (*Sorghum bicolor*) playing a critical role in both food security and traditional agricultural practices (Elramlawi et al., 2019). In the Gadarif region, these crops occupy a substantial portion of the agricultural landscape, making it vital to understand the factors that influence yield. Crop productivity is particularly sensitive to climate variables, incident solar radiation being a key factor in growth and yield (Mannava, 2023).

Solar radiation is a critical component of the energy balance that drives evapotranspiration processes, affecting water and nutrient transport in plants and, consequently, crop yield (Baur et al., 2024). Different forms of solar radiation, including top-of-atmosphere, incident, reflected, and absorbed solar radiation, play different roles in plant energy balance and growth (Lu et al., 2024). However, the present study focuses on incident solar radiation, which is the direct measure of solar energy reaching the Earth's surface.

Despite extensive research on the effect of climate variables on crop growth to determine crop evapotranspiration and crop coefficients (Kc), there is still a need for region-specific studies that investigate the influence of incident solar radiation and other meteorological factors on crop yields during different growth stages in Sudan. Previous research has demonstrated correlations between weather variables and crop yield (Musa et al., 2021).

This study aimed to analyze the relationship between incident solar radiation and crop yields in the Gadarif region of Sudan, using Pearson's and Spearman's correlation analyses, and predict crop yields based on climate variables such as incident solar radiation, temperature, and rainfall, using advanced machine learning (ML) models. The XGBoost, Boosted Regression Forest (BRF), and K-Nearest Neighbors (K-NN) algorithms were chosen for their ability to capture complex interactions between climate variables and crop yield.

Material and Methods

The study was conducted in the Gadarif region of Sudan, the largest area for mechanized rain-fed sorghum and sesame cultivation in the country. Gadarif covers an area of approximately 78,000 km² and is situated within a semi-arid climate zone, 33 to 37° E and 12 to 16° N. With average annual rainfall of 450 mm, primarily from June to September, and temperatures ranging from 21 °C in January to 37 °C in April and May, the region plays a crucial role in Sudan's agricultural output, making it the focal point of this study (Sulieman & Ahmed, 2013). Figure 1 shows the study area within the Gadarif region.

Meteorological data, including daily incident solar radiation, temperature, relative air humidity, wind speed, and rainfall were collected from the Gadarif weather station. These data were aggregated into monthly averages and analyzed during the July to October growing season, across four decades: 1981–1990, 1991–2000, 2001–2010, and 2011–2021. Concurrently, data on sorghum and sesame yields from 1981 to 2021 were obtained for analysis from the Ministry of Agriculture and Irrigation in Gadarif state, providing essential yield measurements (in kg) and harvested areas (in ha).

To provide a comprehensive overview of climate conditions throughout the study period, meteorological data were collected daily from 1981 to 2021, and then aggregated into monthly averages for analysis of long-term trends. Table 1 presents a summary by decade of the monthly values for the variables incident solar radiation (measured in MJ m^{-2} per day), temperature, relative air humidity, wind speed, and rainfall, all essential for understanding the environmental conditions in Gadarif during the sorghum and sesame growing seasons.

These relationships were analyzed by focusing on anomalies, using Pearson's correlation and nonparametric Spearman's rank correlation. Prior to analysis, the Shapiro–Wilk normality test was conducted on the yield anomalies to ensure normal distribution, confirming that the detrending process was valid. All datasets met the normality assumption ($p \leq 0.05$).

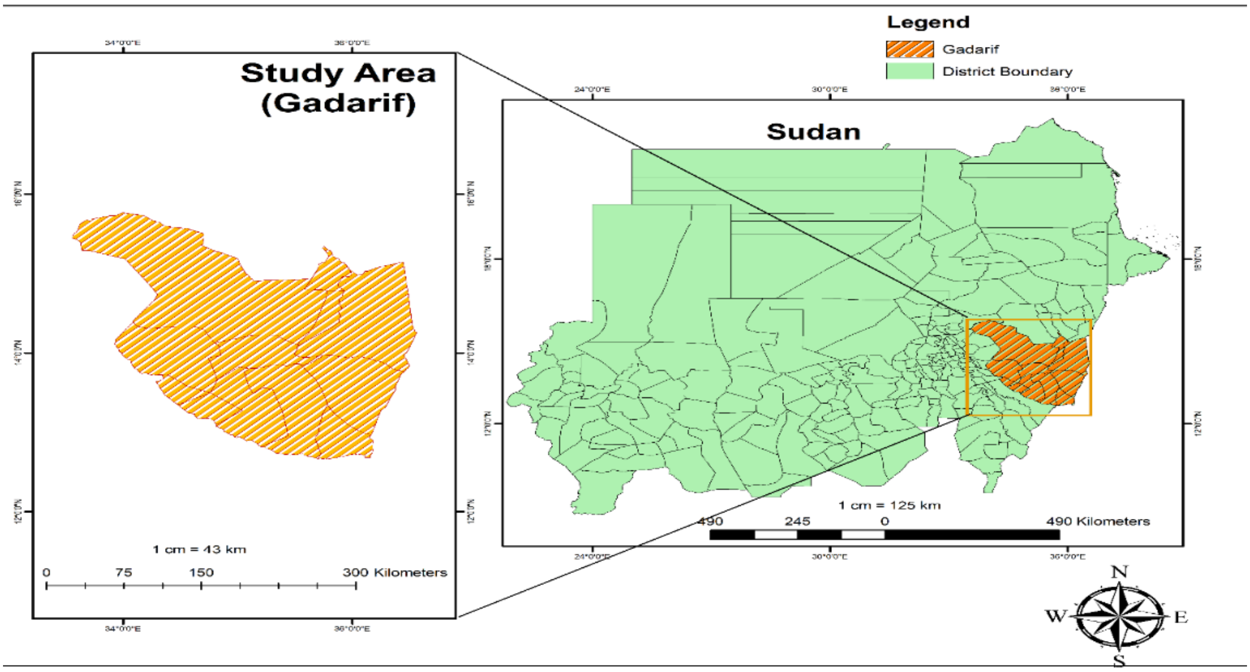


Figure 1. Study area in the Gadarif region of Sudan

Table 1. Average values by decade for key climate variables during sorghum and sesame growing seasons (1981-2021)

Decade	Incident solar radiation (MJ m^{-2} per day)	MaxT (°C)	MinT (°C)	RH (%)	Wind speed (m s^{-1})	Rainfall (mm)
1981-1990	166.2	33.16	21.7	58.59	4	92
1991-2000	171.2	32.96	21.8	58.96	5	102
2001- 2010	154.6	37.27	22.7	42.94	3	66
2011-2021	177.2	34.26	21.9	53.97	3	100

MaxT (°C) - Maximum temperature; MinT (°C) - Minimum temperature; RH (%) - Relative air humidity

In addition to correlation analyses, the ML models Extreme Gradient Boosting (XGBoost), Boosted Regression Forest (BRF), and K-Nearest Neighbors (K-NN) were used to predict crop yields based on the meteorological data. The models were trained and tested using an 80/20 data split. To further clarify the mathematical foundation of the ML models, the following equations illustrate core algorithm use.

XGBoost is a scalable and efficient implementation of gradient boosting machines, developed by Chen & Guestrin (2016), that constructs models by sequentially adding weak learners to minimize the loss function. The general form of the objective function is expressed as follows: the XGBoost model makes predictions f(x) by additive training, sequentially combining the outputs of individual learners' $f_t(x)$ (Eq. 1):

$$f_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \quad (1)$$

where:

x_i - is the training data; and,

$f_t(x)$ - represents the incremental learner fit at stage t.

Typically, simple regression trees are used as the base learners. Additive training minimizes the following regularized objective functions (Eq. 2):

$$\text{Obj}^{(t)} = \sum_{k=1}^{n} l(\overline{y}_i, y_i) + \sum_{k=1}^{t} \Omega(f_i) \quad (2)$$

This equation serves two purposes, namely minimizing the empirical training error measured by the loss function l (y_i, y_i) between the predicted, y_i and target, y_i values, and controlling model complexity through the regularization term Ω (f). The complexity of regularization is defined as (Eq. 3):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2 \quad (3)$$

where:

T - is the number of leaves;

ω - are the leaf weights; and,

λ and γ - control the degree of regularization.

This limits the complexity of the individual tree models to prevent overfitting.

On the other hand, BRF combines regression trees with boosting techniques, as described by Elith et al. (2008), making it particularly effective in modeling non-linear relationships between environmental factors and crop yields. The BRF algorithm builds sequential regression tree models, with each successive model learning from the prediction errors of its predecessor to incrementally improve accuracy. BRF training begins with a basic regression tree and additional trees are subsequently incorporated to fit the errors from the initial model and minimize the loss function. This process continues, with each tree focusing on minimizing the residuals, until convergence or the predefined number of trees is reached. The final BRF model is an additive combination of the sequentially trained regression trees (Eq. 4).

$$f(x) = \sum_{m=1}^{M} w_m \cdot f_m(x) \quad (4)$$

where:

f(x) - denotes the comprehensive prediction;

m - is the number of trees;

w_m - is the weight assigned to the m^{-th} tree; and,

$f_m(x)$ - is the prediction made by the m^{-th} tree.

Finally, the KNN method, first introduced by Evelyn Fix and Joseph Hodges (Fix & Hodges, 1989) and later expanded on by Kramer (2013), is a nonparametric classification technique used for combined data classification and regression tasks. The approach uses a dataset in either scenario and considers the 'k' closest training samples as the input. The KNN method involves querying a database to identify data points that closely resemble the observed data, which are typically the nearest neighbors of the current data. In this study, KNN was applied to predict the most closely related testing stations based on the training station. Eq. 5 summarizes the KNN regression function, as follows:

$$f_{KNN}(x') = \frac{1}{K} \sum_{i \in N_K(x')} y_i \quad (5)$$

In KNN regression, when confronted with an unknown pattern x-′, the algorithm computes the mean of the function values obtained from its K-closest neighbors. The set NK(x) includes the indices of the nearest K neighbors of x-′. The idea of localized functions in both the data and label spaces is the core principle of the averaging process in KNN. Essentially, within the close vicinity of xi, patterns -′ are expected to exhibit similar continuous labels, with f (xi) approximating y_i (Kramer, 2013).

The four most common statistical indicators used to assess model performance are: (1) coefficient of determination (R^2), which measures the proportion of variance in the dependent variable explained by the independent variable(s), with higher values indicating a better fit; (2) mean absolute error (MAE), the average of absolute differences between predicted and actual values, with low values denoting better performance; (3) root mean square error (RMSE), the square root of the average of squared differences between predicted and actual values, whereby low values indicate better accuracy; and (4) mean absolute percentage error (MAPE), the average of absolute differences between predicted and actual values, expressed as a percentage of actual values, where low values demonstrate better performance. These indicators are measured by equations that incorporate actual and predicted values, and the number of observations (Eqs. 6, 7, 8, and 9).

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - X_i\right| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i\right)^2} \quad (8)$$

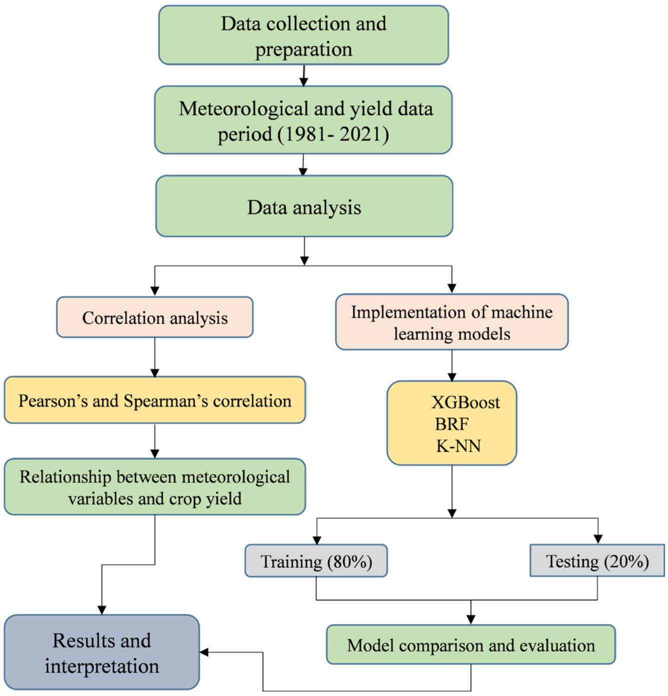
$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - X_i}{Yi}\right| \times 100\% \quad (9)$$

For every time step, X_i and Y_i denote the actual and forecasted crop yield values, respectively, and X_i and Y_i their respective means.

The dataset was partitioned into training and testing sets, allocated at 80 and 20%, respectively, to balance model learning and evaluation. This ratio was deemed optimal via iterative experimentation, starting with equal proportions and adjusting for increased training and decreased testing. Fine-tuning of hyper parameters, including learning rate and regularization strength, was conducted using randomized search cross-validation (CV) to efficiently explore parameter space and identify optimal settings for model training. This technique systematically explores a wide range of combinations, thus enhancing model performance. By rigorously optimizing hyper parameters, the present study highlights the importance of methodological rigor in improving the predictive performance of ML models, providing valuable insights for future research (Kumbure et al., 2022).

K-fold cross-validation was applied to ensure reliable training results, testing k values of 3, 5, and 10 to determine the optimal value for accurate predictions, with minimal differences in outcome. A value of k = 5 was chosen for its reduced bias compared to k = 3 and lower computational requirements in relation to k = 10 (Rodriguez et al., 2009).

Figure 2 presents the methodological framework used to analyze the impact of climate variables on sorghum and sesame yields, from data collection to analysis and interpretation, applying both traditional statistical methods and machine learning techniques.



This flowchart outlines the process from data collection and analysis to interpretation, combining both statistical methods and machine learning techniques

Figure 2. Methodological framework for assessing the impact of climate variables on sorghum and sesame yields in Gadarif, Sudan (1981-2021)

Results and Discussion

Coefficients of variation (CV) for yield, rainfall, relative air humidity (RH), wind speed (WS), minimum (Tmin) and maximum temperature (Tmax), and incident solar radiation (H) were analyzed (Table 2).

The Mann-Kendall test was used to identify patterns in incident solar radiation over time, detecting monotonic upward or downward trends. The results are shown in Table 3, including

Table 2. Statistical characteristics for meteorological parameters at the Gadarif weather station for the entire study period (1981-2021)

Variable	Xmin	Xmax	Xmean	CV	SD
Tmax (°C)	32.54	35.39	33.717	0.0202	0.6804
Tmin (°C)	20.42	22.38	21.42	0.0227	0.4856
Rain (mm)	322	910.7	600.16	0.2197	131.84
H (MJ m^{-2} per day)	140.21	201.23	167.56	0.0956	16.024
WS (m s^{-1})	2.46	7.52	4.9688	0.2449	1.217
RH (%)	31	81.88	53.473	0.2353	12.584
Sesame yield (kg ha^{-1})	330.85	1025.2	601.1	0.2597	156.1
Sorghum yield (kg ha^{-1})	254.96	1015.4	514.63	0.3788	194.96

Xmin - Minimum actual value in the dataset; Xmax - Maximum actual value in the dataset; Xmean - Mean actual value in the dataset; SD - Standard deviation; CV - Coefficient of variation; H – Incident solar radiation; WS - Wind speed; RH – Relative air humidity

Table 3. Analysis of monthly incident solar radiation patterns at Gadarif weather station, Sudan, using the Z statistic from the Mann-Kendall test for sorghum and sesame growing seasons (1981–2021)

Time period	Jul		Aug		Sept		Oct		Growing season	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value	Z	p-value
1981-1990	1.25	0.21	0.44	0.65	0.44	0.65	1.52	0.12	0.80	0.42
1991-2000	-0.71	0.47	0.08	0.92	-0.53	0.59	0.62	0.53	0.44	0.65
2001-2010	1.52	0.12	0.53	0.59	1.52	0.12	1.43	0.15	1.34	0.17
2011-2021	1.96	0.04	2.68	≤ 0.007	1.78	≤ 0.01	1.69	0.08	2.23	≤ 0.01

the test statistic (Z) and associated p-value. Positive Z values indicate an upward trend, negative values a downward trend, and $p \leq 0.05$ a statistically significant trend.

Notably, a significant increasing trend was identified for August in the 2011–2021 period. For both 2001-2010 and 2011-2021, there was an obvious increasing trend in September, with no significant trends for October. Considering the entire growing season, a significant increasing trend was observed for 2011–2021. In summary, the most recent 2011–2021 timeframe showed noteworthy upward trends in incident solar radiation for August, September, and the season as a whole. Conversely, earlier periods displayed fewer significant trends, suggesting a trend of increasing solar radiation during the late summer and fall months over the past decade.

Figure 3 illustrates the trends in sorghum and sesame yields over four decades, from 1981 to 2021. For 1981–1990 and 1991–2000, sorghum yield decreased by 1.14%, or 68.05 kg ha^{-1}, continuing to decline by 1.23% between 1991–2000 and 2001–2010, corresponding to a reduction of 90.07 kg ha^{-1}. However, in the last decade (2011–2021), sorghum yield increased by 1.53%, equivalent to 212 kg ha^{-1}, which was attributed to the rise in incident solar radiation. Between 1981–1990 and 1991–2000, sesame yields fell by 1.17% or 90.82 kg ha^{-1}, followed by a 0.99% reduction (3.97 kg ha^{-1}) in 1991–2000 and 2001–2010, and an increase of 0.80% or 135.96 kg ha^{-1} in the most recent decade (2011–2021). While this last period coincided with an increase in incident solar radiation, it is important to underscore that multiple factors

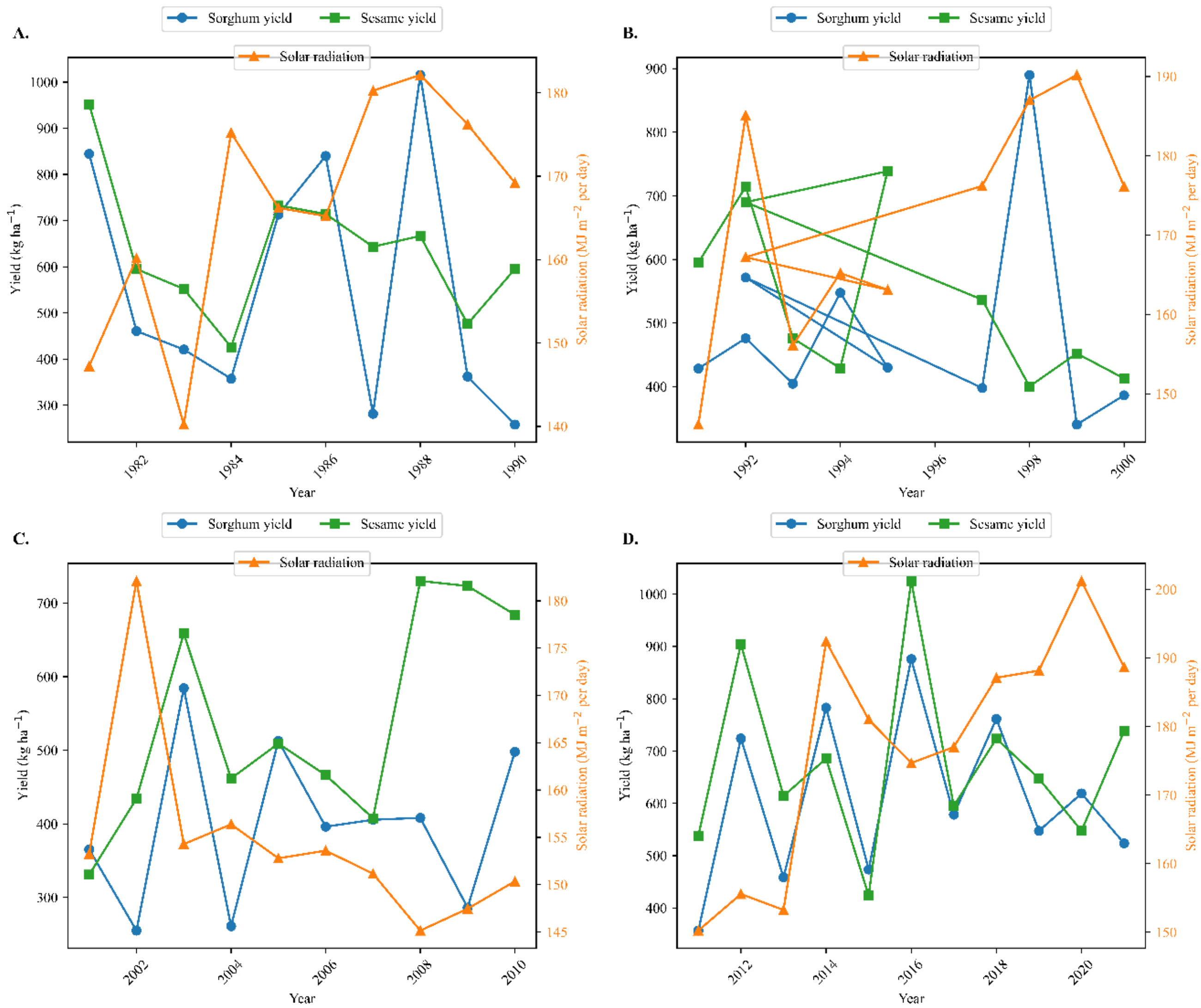


Figure 3. Patterns of incident solar radiation and crop yield across four decades (1981-2021), based on data from the Gadarif weather station

likely contributed to these higher yields, including genetic improvements in cultivars, advancements in agricultural practices, better nutrition, irrigation techniques, and other agronomic interventions. Without comprehensive evidence directly linking the yield increase to incident solar radiation, it is important to acknowledge the potential influence of these additional variables.

Further analysis demonstrated that while RMSE values only increased slightly between K-NN training and testing (from 15.4 to 16.2 kg ha^{-1} for sorghum), the scatter plots (Figures 4 and 5) suggest more pronounced deviations from the 1 × 1 line during testing. This discrepancy can be attributed to the non-parametric nature of the K-NN model, which is particularly sensitive to local variations in data distribution. Moreover, residual analysis and further inspection revealed larger prediction errors for specific outliers or regions of the input space during testing, which are visibly more pronounced in the scatter plots than the RMSE metric alone suggests. Additional error metrics such as MAE and residual distribution plots were used to provide a better understanding, highlighting the complex nature of model performance across different datasets.

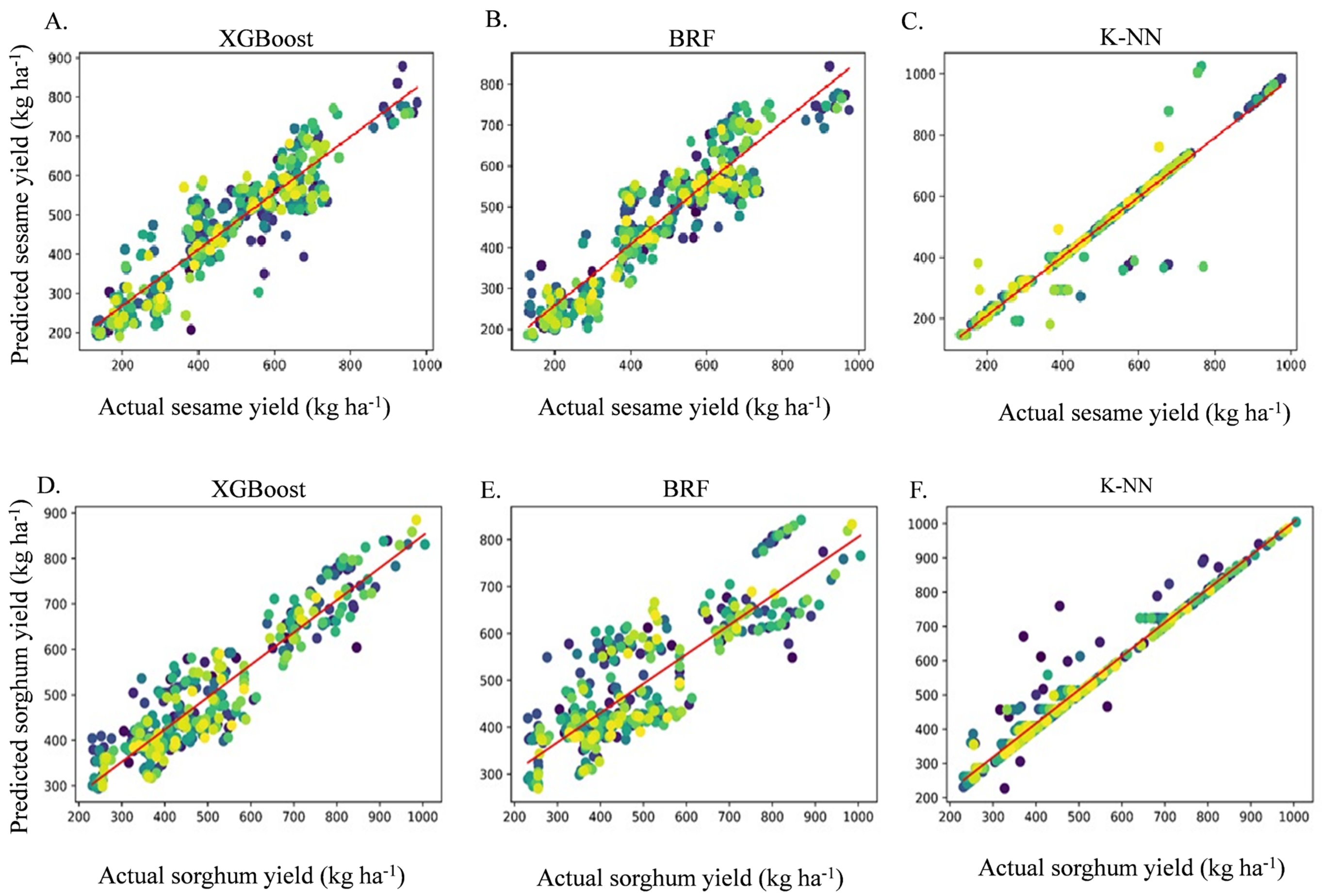
The performance of the three ML algorithms (XGBoost, BRF, and K-NN) in predicting sorghum yield was assessed (Table 4). K-NN was the most accurate, with an average R^2 of 0.89 across different test datasets. BRF and XGBoost also exhibited satisfactory performance, with R^2 values of 0.85 and 0.82, respectively.

Analysis of K-NN indicated that incident solar radiation and average temperature during the growing season were the most influential factors in predicting sorghum yield. This is consistent with a previous study that emphasizes the influence of weather-related variables on sorghum yield (Affoh et al., 2022).

In parallel, the same three ML techniques (XGBoost, BRF, and K-NN) were used to predict sesame yield. Notably, the K-NN model demonstrated superior accuracy, obtaining a R^2 of 0.90 across the validation datasets, while BRF and XGBoost also performed well, yielding respective R^2 values of 0.88 and 0.83.

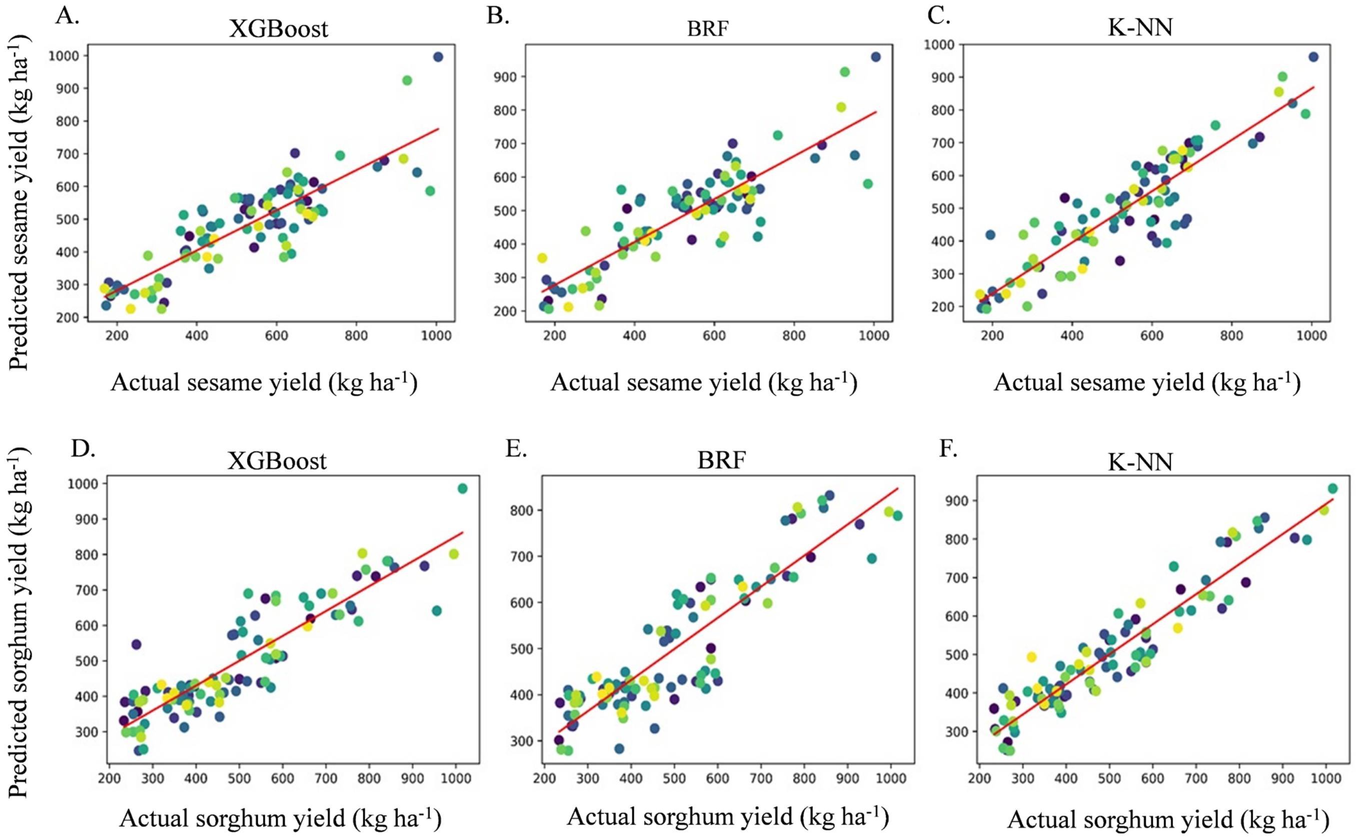
Temperature and incident solar radiation were important predictors of sesame yield, corroborating an earlier study (Zhou et al., 2023). The findings obtained here are consistent with previous investigations, demonstrating the effectiveness of ML methods in predicting crop yields (Gonzalez-Sanchez et al., 2014; Pandith et al., 2020). However, the present study provides new insights by identifying particular variables that significantly influence sorghum and sesame yields within the Gadarif region of Sudan.

For both crops, K-NN consistently outperformed XGBoost and BRF in terms of R^2 in the training and testing datasets, and generally obtained the lowest MAE, MAPE, and RMSE across these datasets, indicating better accuracy and smaller prediction errors compared to XGBoost and BRF. XGBoost tended to exhibit the highest prediction errors (MAE, MAPE, and RMSE) among the models, particularly in the testing



XGBoost - Extreme Gradient Boosting model; BRF - Boosted Regression Forest model; K-NN - K-Nearest Neighbors model

Figure 4. Crop yields predicted by different ML models (XGBoost, BRF, K-NN) compared to actual values for sesame (A, B, and C) and sorghum (D, E, and F) during the training phase, from 1981 to 2013



XGBoost - Extreme Gradient Boosting model; BRF - Boosted Regression Forest model; K-NN - K-Nearest Neighbors model

Figure 5. Crop yields predicted by different machine learning models compared to actual values for sesame (A, B, and C) and sorghum (D, E, and F) during the testing phase, from 2014 to 2021

Table 4. Performance of machine learning models against actual sorghum and sesame yield data during training and testing periods for the Gadarif region

Crop	Models	Training				Testing			
		R^2	MAE	MAPE	RMSE (kg ha^{-1})	R^2	MAE	MAPE	RMSE (kg ha^{-1})
Sorghum	XGBoost	0.871	21.1	9.3	17.5	0.828	20.2	10.6	19.5
	K-NN	0.937	16.2	4.4	15.4	0.892	19.1	8.5	16.2
	BRF	0.898	19.1	7.1	16.2	0.854	21.2	9.3	18.3
Sesame	XGBoost	0.882	22.5	8.1	17.3	0.831	21.3	9.1	19.1
	K-NN	0.951	10.2	4.04	13.1	0.906	12.2	5.2	14.2
	BRF	0.901	18.2	6.3	16.4	0.881	20.7	6.5	17.3

R^2 - Coefficient of determination; MAE - Mean absolute error; MAPE - Mean absolute percentage error; RMSE - Root mean square error

datasets for both crops. These results suggest that, for the given datasets and features, K-NN is better suited to predicting crop yields than XGBoost and BRF (Table 4).

Based on global climate models, Ciavarella et al. (2021), provides evidence that in the 140-year record, 8 out of the 10 warmest years globally occurred after 2010. Similarly, the four warmest years in Africa have all been recorded since 2015. The authors also highlight that the annual temperature increase between 1981 and 1921 is more than twice that observed from 1910 to 1921, rising at rates of 0.31 and 0.12 °C per decade, respectively. These findings demonstrate a rising trend in annual incident solar radiation for the past four decades in a specific region of Sudan (Figure 6). This corroborates the findings of Mohammad & Othman (2022), who reported the potential benefits of predictive models for solar radiation by providing valuable insights to optimize crop yield. A study conducted in the southern portion of the Upper Blue Nile Basin in northwestern Ethiopia supports these findings, highlighting increasing trends in annual Tmin and Tmax from 1981 to 2010, with per decade rises of 0.1 to 0.15 °C. These temperature changes directly influence incident solar radiation, as observed by Mengistu et al. (2014). The magnitude and duration of solar radiation are key factors in crop development. These findings

Figure 6. Time series data on incident solar radiation from 1981 to 2021 in Gadarif, Sudan

Table 5. Pearson's and Spearman's rank correlation between sorghum and sesame yield anomalies and the corresponding incident solar radiation indicators (monthly for July, August, September, October, and seasonal) recorded at the Gadarif weather station

Crop	Time period	Jul		Aug		Sept		Oct		Growing season	
		r	ρ	r	ρ	r	ρ	r	ρ	r	ρ
Sorghum	1981-1990	0.10^{ns}	0.06^{ns}	0.1^{ns}	0.09^{ns}	0.02^{ns}	-0.04^{ns}	0.52^{ns}	0.64^{ns}	0.26^{ns}	0.46^{ns}
	1991-2000	0.03^{ns}	-0.2^{ns}	0.39^{ns}	0.23^{ns}	0.05^{ns}	-0.17^{ns}	0.06^{ns}	0.13^{ns}	0.24^{ns}	0.15^{ns}
	2001-2010	-0.36^{*}	-0.31^{*}	-0.38^{**}	-0.34^{*}	-0.43^{**}	-0.53^{**}	0.19^{ns}	0.03^{ns}	-0.5^{**}	-0.47^{**}
	2011-2021	-0.55^{**}	0.58^{ns}	-0.24^{ns}	-0.31^{*}	-0.35^{*}	-0.22^{ns}	-0.36^{*}	0.41^{**}	0.04^{ns}	-0.06^{ns}
Sesame	1981-1990	-0.16^{ns}	0.28^{ns}	-0.04^{ns}	-0.05^{ns}	-0.15^{ns}	-0.19^{ns}	0.63^{ns}	0.53^{ns}	0.15^{ns}	0.17^{ns}
	1991-2000	0.68^{**}	0.57^{**}	-0.04^{ns}	-0.07^{ns}	-0.11^{ns}	-0.14^{ns}	-0.14^{ns}	0.11^{ns}	-0.33^{*}	-0.23^{ns}
	2001-2010	-0.37^{*}	0.45^{**}	-0.23^{ns}	-0.41^{**}	0.32^{ns}	0.06^{ns}	0.4^{ns}	0.2^{ns}	0.19^{ns}	0.23^{ns}
	2011-2021	0.03^{ns}	0.14^{ns}	0.61^{ns}	0.67^{ns}	-0.03^{ns}	-0.17^{ns}	0.13^{ns}	0.12^{ns}	0.08^{ns}	-0.12^{ns}

r - Pearson's correlation coefficient; ρ - Spearman's rank correlation coefficient; * - Statistically significant at $p \leq 0.05$; ** - Statistically significant at $p \leq 0.01$; ns - Not statistically significant

are reinforced by Villa et al. (2022), who underscored the dependence of plant growth and development on the intensity and duration of incident solar radiation.

As shown in Table 5, for 2001–2010, a statistically significant ($p \leq 0.05$) correlation was observed between incident solar radiation and sorghum yield throughout the growing season, with r values of -0.36, -0.38, and -0.43 for July, August, and September, respectively (p = -0.31, -0.34, and -0.53 for the same months). This correlation persisted from 2011 to 2021, with r = -0.55, -0.35, and -0.36 for July, September, and October, and p-values of -0.31 and -0.41 for August and October, respectively. Additionally, for sesame, there was a significant inverse relationship between incident solar radiation and crop yield from 1991 to 2000. This is consistent with the findings of Holzman et al. (2018) and was particularly evident in July (r = -0.68) and across the growing season (r = -0.33), with both correlations statistically significant ($p \leq 0.05$). The correlation continued from 2001 to 2010, with a coefficient of -0.37 in July and respective p-values of -0.45 and -0.41 for July and August.

The analyses conducted here revealed a number of interesting insights into the interaction between incident solar radiation, other meteorological variables, and crop yield and the implications for agricultural practices.

The results confirm the role of solar radiation in determining crop yield. There is a positive relationship between incident solar radiation and crop yield, indicating that greater exposure to sunlight improves photosynthesis and plant growth. This is well-established in the literature, highlighting the pivotal role of incident solar radiation in increasing crop yield (Holzman et al., 2018; Yang et al., 2019). Farmers in areas with abundant solar radiation can benefit from this knowledge to improve crop planting times and further increase yield potential.

Conclusions

1. Analysis of incident solar radiation trends in Gadarif state, Sudan, over the past four decades reveals significant patterns correlated with crop yields. Notably, there was a marked increase in incident solar radiation during the late summer and fall months from 2011 to 2021, specifically August and September.

2. The machine learning models Extreme Gradient Boosting (XGBoost), Boosted Regression Forest (BRF), and K-Nearest Neighbors (K-NN) were used to predict crop yield. The results demonstrated that the models effectively captured the complex interactions between incident solar radiation and crop yields. K-NN was the most accurate, underscoring the significant impact of incident solar radiation and temperature on yield predictions.

3. Overall, this study highlights the importance of advanced machine learning techniques in improving agricultural forecasting models. These insights are crucial for informing adaptive agricultural practices, improving food security, and ensuring agricultural sustainability in regions with variable meteorological conditions.

Contribution of authors: Abdelkarem Mohmoud Adam was responsible for the conception and design of the study, data acquisition, data analysis and interpretation, and drafting the manuscript. Both Abdelkarem Mohmoud Adam and Yuan Zheng critically revised the manuscript for important intellectual content. Additionally, Abdelkarem Mohmoud Adam and Yuan Zheng approved the final version of the manuscript for publication.

Supplementary documents: There are no supplementary sources.

Conflict of interest: The authors declare no conflict of interest.

Financing statement: There is no funding to declare.

Acknowledgments: We extend our sincere thanks to the Ministry of Agriculture and Gadarif Weather Station, Sudan, for supplying the crop yield and weather data.

Literature Cited

Affoh, R.; Zheng, H.; Zhang, X.; Yu, W.; Qu, C. Influences of meteorological factors on maize and sorghum yield in Togo, West Africa. Land, v.12, e123, 2022. https://doi: 10.3390/land12010123.

Baur, S.; Sanderson, B. M.; Seferian, R.; Terray, L. Solar radiation modification challenges decarbonization with renewable solar energy. Earth System Dynamics, v.15, p.307-322, 2024. http://dx.doi.org/10.5194/esd-15-307-2024.

Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, v.3, p.785-794, 2016. http://dx.doi.org/10.1145/2939672.2939785

Ciavarella, A.; Cotterill, D.; Stott, P.; Kew, S.; Philip, S.; Oldenborgh, G. J. V.; Skalevag, A.; Lorenz, P.; Robin, Y.; Otto, F.; Hauser, M.; Seneviratne, S. I.; Lehner, F.; Zolina, O. Prolonged siberian heat of 2020 almost impossible without human influence. Climatic Change, v.166, p.1-18, 2021. http://dx.doi.org/10.1007/s10584-021-03052-w.

Elith, J.; Leathwick, J. R.; Hastie, T. A working guide to boosted regression trees. Journal of Animal Ecology, v.77, p.802–813, 2008. http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x.

Elramlawi, H. R.; Mohammed, H. I.; Elamin, A. W.; Abdallah, O. A.; Taha, A. A. A. M. Adaptation of sorghum (*Sorghum bicolor L. Moench)* crop yield to climate change in eastern dryland of Sudan. In: Handbook of climate change resilience, p.2549-2573, 2019. http://dx.doi.org/10.1007/978-3-319-71025-9_157-1.

Fix, E.; Hodges, J. L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. International Statistical Review, v.57, e238, 1989. http://dx.doi.org/10.2307/1403797.

Gonzalez-Sanchez, A.; Frausto-Solis, J.; Ojeda-Bustamante, W. Predictive ability of machine learning methods for massive crop yield prediction. Spanish Journal of Agricultural Research, v.12, e313, 2014. http://dx.doi.org/10.5424/sjar/2014122-4439.

Holzman, M. E.; Carmona, F.; Rivas, R.; Niclòs, R. Early assessment of crop yield from remotely sensed water stress and solar radiation data. ISPRS Journal of Photogrammetry and Remote Sensing, v.145, p.297-308, 2018. http://dx.doi.org/10.1016/j.isprsjprs.2018.03.014.

Kramer, O. Dimensionality Reduction with Unsupervised Nearest Neighbors. Berlin: Springer Berlin Heidelberg. Intelligent Systems Reference Library, v.51, 132p, 2013. https://doi.org/10.1007/978-3-642-38652-7

Kumbure, M. M.; Lohrmann, C.; Luukka, P.; Porras, J. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, v.197, e116659 2022. http://dx.doi.org/10.1016/j.eswa.2022.116659.

Lu, Z.; Gao, J.; Wang, Q.; Ning, Z.; Tan, X.; Lei, Y.; Zhang, J.; Zou, J.; Lingxuan, W.; Yang, C.; Yang, W.; Yang, F. Light energy utilization and measurement methods in crop production. Crop and Environment, v.3, p.91-100, 2024. http://dx.doi.org/10.1016/j.crope.2024.02.003.

Mannava, S. Importance of solar radiation and the need for improved respect to Sun by Agrometeorologists. Journal of Agrometeorology, v.25, p.51-60, 2023. http://dx.doi.org/10.54386/jam.v25i1.1971.

Mengistu, D.; Bewket, W.; Lal, R. Recent spatiotemporal temperature and rainfall variability and trends over the Upper Blue Nile River Basin, Ethiopia. International journal of climatology, v.34, p.2278-2292, 2014. http://dx.doi.org/10.1002/joc.3837.

Mohammad, G.; Othman, A. Design of an artificial neural network-based model for prediction solar radiation utilizing measured weather datasets. WSEAS Transactions on Power Systems, v.17, p.132-140, 2022. http://dx.doi.org/10.37394/232016.2022.17.14.

Musa, A. I. I.; Tsubo, M.; Ali-Babiker, I. E. A.; Lizumi, T.; Kurosaki, Y.; Ibaraki, Y.; Tsujimoto, H. Relationship of irrigated wheat yield with temperature in hot environments of Sudan. Theoretical and Applied Climatology, v.145, p.1113-1125, 2021. http://dx.doi.org/10.1007/s00704-021-03690-1.

Pandith, V.; Kour, H.; Singh, S.; Manhas, J.; Sharma, V. Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis. Journal of Scientific Research, v.64, p.394-398, 2020. http://dx.doi.org/10.37398/jsr.2020.640254.

Rodriguez, J. D.; Perez, A.; Lozano, J. A. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, v.32, p.569-575, 2009. http://dx.doi.org/10.1109/TPAMI.2009.187.

Sulieman, H. M.; Ahmed, A. G. M. Monitoring changes in pastoral resources in eastern Sudan: A synthesis of remote sensing and local knowledge. Pastoralism: Research, Policy and Practice, v.3, p.1-16, 2013. http://dx.doi.org/10.1186/2041-7136-3-22.

Villa, B. D.; Petry, M. T.; Santos, M. S. N. D.; Martins, J. D.; Lago, I.; Moura, M. B. D.; Castro, R. P. Effects of Minimum and Maximum Limits of Solar Radiation and Its Temporal and Geographic Interactions. Journal of Agricultural Science, v.14, e173, 2022. http://dx.doi.org/10.5539/jas.v14n8p173.

Yang, Y.; Xu, W.; Hou, P.; Liu, G.; Liu, W.; Wang, Y.; Li, S. Improving maize grain yield by matching maize growth and solar radiation. Scientific Reports, v.9, e3635, 2019. http://dx.doi.org/10.1038/s41598-019-40081-z.

Zhou, M.; Liu, H.; Zhang, J.; Li, G.; Zang, H.; Qiu, Y.; Zheng, G. Attribution analysis on the changing trend of sesame yield data in southern Henan under climate change. In: International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022), Guangzhou, China. SPIE. v.89, e2674617, 2023. https://doi.org/10.1117/12.2674617.