University of Birjand

Water Harvesting Research (2025) Vol. 8, No. 2: 228-237

10.22077/jwhr.2025.10040.1187

ORIGINAL ARTICLE

Predicting the Gray Water Footprint and Water Use Efficiency in Farms Using ML Models (Case Study: Lorestan Province)

Athare Khakshour[a], Masoud Shakarami[b*], Mohammad Nazeri Tahroudi[b], Seyed Yaghoub Karimi[b]

[a]MSc Student, Department of Water Engineering, Lorestan University, Khorramabad, Iran.
[b]Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran.

[*]Corresponding Author E-mail address: shakarami.mas@lu.ac.ir
Received: 05 August 2025, **Revised**: 19 September 2025, **Accepted**: 06 October 2025

Abstract

This study aims to (1) evaluate the Crop Water Productivity (*CWP*) and gray Water Footprint (WF_{Gray}) for key agricultural systems in Lorestan province, Iran, to identify hotspots of inefficiency and pollution, and (2) develop and compare Machine Learning (ML) models for predicting these metrics to aid in management and forecasting. We calculated *CWP* and WF_{Gray} for major crops (including forage corn, wheat, beans, potatoes and vegetables) across multiple meteorological stations in Lorestan province. Furthermore, we employed two ML algorithms including Random Forest (RF) and Support Vector Machine (SVM) to model and predict these indices. Model performance was evaluated using the Mean Absolute Error (MAE). The assessment revealed significant regional and crop-specific disparities. Forage corn was the most efficient and sustainable system (*CWP*: 2.173 kg/m^3, WF_{Gray}: 0.05 m^3/kg), whereas bean production was the least efficient (*CWP*: 0.064 kg/m^3). Spatially, stations like Azna (potato) demonstrated best practices, while Kuhdasht was identified as a critical area of concern due to low efficiency and high fertilizer pollution. In modeling, the optimal algorithm was target-dependent: RF was superior for predicting *CWP* (MAE: 0.236), while SVM performed relatively better for the more complex WF_{Gray}. This study concludes that addressing water security and agricultural pollution in the region requires tailored, crop-specific interventions and improved farm management practices. Furthermore, while ML model (particularly RF) proves to be a powerful tool for forecasting water productivity, accurately modeling the environmental impact (WF_{Gray}) remains a challenge, highlighting the need for more robust data and further research in this domain.

Keywords: Agricultural Water Management, Crop Water Productivity, Gray Water Footprint, Random Forest, Nitrogen Pollution.

1. Introduction

Iran is located in an arid and semi-arid belt, making it a predominantly dry country. Therefore, the problem of water shortage has become a major challenge in the country. The The agricultural sector consumes a large share of water resources to provide food for the growing population (Dehghanpir et al., 2023).

The decline in groundwater levels, the drying up of rivers, and the increase in water pollution are indicators of water resource scarcity (Hoekstra, 2008). Furthermore, there is an unfavorable temporal mismatch between precipitation patterns and irrigation seasons. Additionally, rainfall amounts are highly inconsistent from year to year and season to season. This issue has caused various problems in recent years for different sectors, particularly agriculture, and has inflicted significant losses on this sector (Behmanesh, 2016). One practical approach to water resource management is assessing crop water requirements and determining the volume of water consumed during different production stages (Piri and Sarani, 2020).

The growing competition for water, the need to feed an increasing population, and escalating water scarcity worldwide are essential reasons that necessitate a focus on water consumption management.

Geographically, countries in the Middle East and Central Asia are facing the greatest reduction in physical water resources relative to their societal consumption. On the other hand, projections regarding changes in resources and the intensification of the global freshwater crisis, based on the risk level of unsustainable economic development by 2050, indicate that most countries in Central Asia, North of Africa, and North America face a risk of over 40% reduction in water resources. Countries in the Middle East, particularly Iran, with a risk of over 50% reduction in freshwater resources, are among the most vulnerable regions in the world confronting the water crisis (Vörösmarty et al., 2010; Madani, 2014; Mekonnen and Hoekstra, 2016).

Agriculture is a major consumer of water resources as a primary input for production. Meanwhile, the development of this sector is highly significant for the economic development of developing countries like Iran, in terms of employment, ensuring food security, supplying raw materials for industries, and generating income (Goodarzi et al., 2023).

The water footprint index serves as a global indicator representing the actual volume of water consumed based on the climate and conditions of each region. Understanding and evaluating the actual water used for various agricultural products is of great importance, and this assessment can be highly beneficial for identifying and proposing appropriate strategies to reduce water consumption in the agricultural sector (Aligholinia et al., 2017).

Examining the volume of water directly or indirectly consumed to produce a good or deliver a service (referred to as the water footprint) has significant potential for water management in agriculture (Dehghanpir et al., 2023). The concept of the water footprint, introduced by Hoekstra and Chapagain (2011), enables the analysis of relationships between water use and the allocation of freshwater resources. Furthermore, Hoekstra (2008) defined the gray water footprint as the volume of freshwater required to dilute pollutants generated during the production process of a product, based on established water quality standards.

The total water used throughout all stages of growth and production of agricultural products is referred to as virtual water (Piri and Sarani, 2020). The gray water footprint (GWF) is a critical concept in water resource management, representing the volume of freshwater required to dilute pollutants to meet water quality standards. Understanding and predicting the GWF is essential for sustainable agricultural practices, efficient water management, and environmental protection.

Recent studies have emphasized the importance of accurately modeling the GWF in relation to agricultural practices. Serrano et al. (2016) provided a comparative analysis of organic farming versus conventional rice farming, illustrating the substantial differences in gray water usage. The research found that conventional practices, which rely heavily on chemical fertilizers, herbicides, and pesticides, result in a greater GWF due to the increased volume of freshwater necessary to dilute the resultant pollutants. In contrast, organic farming significantly reduces the GWF, suggesting that sustainable agricultural practices are essential for minimizing environmental impacts while enhancing economic returns.

The water footprint represents the actual water consumption of products through three components: green, blue, and gray. It has recently gained attention as part of modern water resource management with an integrated approach. To properly investigate water consumption in agriculture, it is essential to evaluate the water footprint index across different climates (Piri and Sarani, 2020).

Given the necessity of focusing on water resource management, one of the novel concepts in water management is the water footprint. The water footprint index for product production is used for the quantitative and qualitative management of water resources. Virtual water is the total water used to produce one unit of a product, good, or service (Oveisi et al., 2019).

The water footprint encompasses a concept similar to, but broader than, virtual water, as it incorporates temporal and spatial dimensions into the virtual water concept, thereby serving as a link for policy formulation. The water footprint is an indicator developed based on the concept of virtual water; in other words, at the product scale, the water footprint index reflects the concept of virtual water, but at larger scales, the virtual water concept is used as a calculation tool (Ma et al., 2020).

The impacts of climate change on water resources and the GWF are increasingly critical. Ostad-Ali-Askari et al. (2019) examined the effects of management strategies on mitigating climate change impacts on aquifer water resources. Their findings indicate that proactive management can significantly reduce negative impacts, suggesting that effective strategies are essential not only for preserving water quality but also for ensuring sustainable water supplies in the face of climate variability. This highlights a crucial area for further research on how adaptive management practices can influence the GWF.

The water productivity index is one of the most important indicators for analyzing water use efficiency, representing the amount of product produced per cubic meter of water consumed (Piri and Sarani, 2020). Given that the water footprint depends on various parameters, predicting it is challenging. Therefore, one effective approach for predicting and estimating agricultural and hydrological variables is the use of Machine Learning (ML) models (Li et al., 2023).

ML models utilize interconnected information processing units to identify relationships and patterns in data, transforming input data into output. Limited research has been conducted in this area. For example, Lotfy et al. (2024) investigated the application of ML models for estimating the blue and green water footprint (BWFP and GWFP) of wheat in the Nile Delta under varying climatic conditions. The study's aim was to develop and compare a comprehensive suite of models, including single models (XGBoost, Random Forest (RF), LASSO, and CatBoost), eight hybrid models, and stacking ensembles.

These models were evaluated across five different input scenarios incorporating climate, crop, and remote sensing data. A key finding was that hybrid ML models significantly outperformed both single models and stacking ensembles. Remarkably, the hybrid models XGB-LASSO and RF-LASSO achieved a perfect R^2 value of 1.0 (100%) for predicting BWFP and GWFP, respectively, under specific input scenarios. In contrast, the single LASSO model performed poorly, especially with remote sensing data alone (R^2 = 0.16). The study concludes that hybrid ML approaches demonstrate high efficacy and superior predictive accuracy for water footprint estimation, offering a powerful tool for agricultural water management.

Given the limited studies on estimating the water footprint using ML models, the objective of this research is to predict water productivity and the gray water footprint using ML models.

2. Materials and Methods

2.1. Case study

This study, conducted in 2022, aimed to model and predict the gray water footprint and water productivity of key crops in Lorestan province, Iran: potatoes and beans in Azna County; vegetables in Poldokhtar County; and grain corn, forage corn, and wheat in Kouhdasht County.

The geographical locations of the studied stations are presented in Fig. 1. The data used included key agro-climatic variables: average effective rainfall, average crop evapotranspiration, and average net irrigation requirement. These variables were calculated using NETWAT software.

Additionally, data on crop yield per unit area, cultivated area, total production (in tons), and the amount of nitrogen fertilizer applied for each crop during 2022 were also collected. A summary of this data is presented in Table 1.

2.2. Random Forest (RF)

The Random Forest algorithm is one of the most widely used algorithms for addressing classification and multivariate prediction problems. It is less sensitive to multicollinearity and yields relatively stable results in the presence of missing data and class imbalance (Bageri et al., 2023; Nazeri Tahroudi et al., 2023; Saloman et al., 2024).

The following four steps define the RF process:

a) Defining and bootstrapping the training data,

b) Selecting a random subset of features for each bootstrap sample,
c) Assigning a decision tree to each of these samples using its random feature subset and growing the tree,
d) Generating a single, unified decision model by aggregating the predictions from all the decision trees (e.g., through majority voting for classification or averaging for regression).

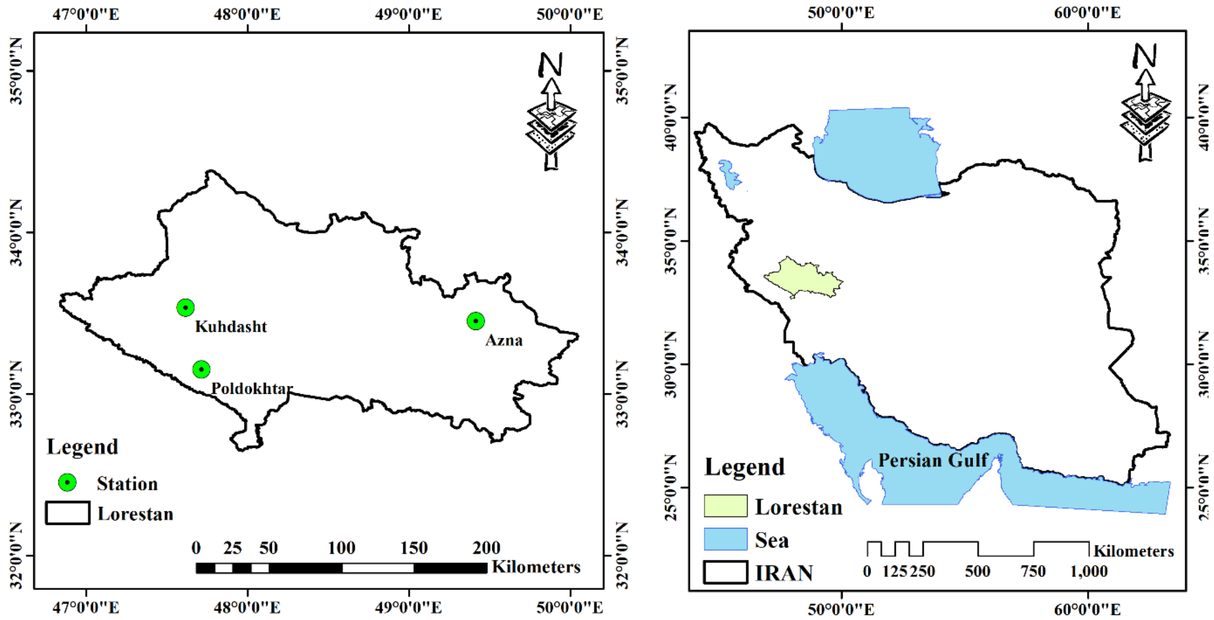

Fig. 1. Geographical location of the studied stations

Table.1. Statistical characteristics of the studied parameters

Province	Product	Average Net Irrigation Requirement	Average Evapotranspiration	Average Effective Rainfall	Cultivated Area (ha)	Nitrogen fertilizer (kg/ha)	Yield (kg/ha)	Total Production (tons)
Azna	Potato	931	941	10	688	400	46773.26	32180
	Bean	751	751	0	6880.75	50	2104.93	14484.1
Poldokhtar	Vegetable	394	443	49	2202.27	200	15904.63	35022
Kohdasht	Grain corn	630	656	26	97.04	400	8000	775.92
	Forage corn	585	611	33	1652.14	600	59803.87	98796
	Wheat	144	177	26	649.8	200	3300.51	21450.2

2.3. Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a supervised ML algorithm (Ding et al., 2017). The SVM method can reduce empirical error, model complexity, and mitigate overfitting (Pisner and Schnyer, 2020). The objective of an SVM is to find the optimal separating hyperplane that maximizes the margin between different classes. In most situations, the hyperplane is defined by a non-linear surface. In such cases, Eq. 1 is used for classifying the datasets.

$$f(x) = \sum_{i=1}^{n} \left(a_i - a_i^*\right) K\left(x_i, x\right) + b, \qquad (1)$$

where a_i and a_i^* are the Lagrange coefficients, K is the kernel function, and b is the bias of the hyperplane from the origin.

2.4. Estimation of Gray Water Footprint (WF_{Gray})

The calculation of the gray water footprint pertains solely to nitrogen fertilizers (Hoekstra and Chapagain, 2011). According to sources, the maximum permissible concentration of nitrogen in surface and groundwater resources is recommended to be 10 mg/L (Su et al., 2018), and it is obtained from Eq. 2:

$$WF_{Gray} = \left(\alpha \times NAR\right) / \left(C_{\max} - C_{nat}\right) \times 1/Y. \qquad (2)$$

In Eq. 2: Y is the crop yield and α is the percentage of nitrogen fertilizer leaching loss. Based on irrigation conditions, this value is 5% for irrigated crops and 10% for rainfed crops (the selected crops in this study are irrigated). The term NAR is the application rate of the fertilizer used per hectare (kg/ha), $C_{\max}$ is the maximum acceptable concentration of the pollutant, and C_{nat} is the natural concentration of the pollutant in the water body.

2.5. Model evaluation

This study employed three criteria to evaluate model performance:

1. Root Mean Square Error (RMSE): Used to measure the distribution of the model's residuals.
2. Mean Absolute Error (MAE): Represents the sum of the average absolute differences between the actual and predicted values. Simply put, MAE indicates how wrong the predictions are, on average.
3. Mean Squared Error (MSE): Calculates the average of the squared differences between the actual and predicted output values (using squares instead of absolute values) before summing them all. These metrics are calculated using Eqs. 3, 4, and 5, respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}, \qquad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \qquad (4)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \qquad (5)$$

where n is the number of observations, y_i is the ith observation, and $\hat{y}_i$ is the predicted value for the ith observation. Also in this study, the Pearson correlation coefficient statistical measure was used to evaluate the correlation between variables. The Pearson coefficient is a symmetric index, meaning the correlation of the dependent variable based on the independent variable is the same as that of the independent variable based on the dependent variable. The value of this correlation index ranges between +1 and -1.

A value exactly equal to +1 indicates a perfect positive correlation, while a value exactly equal to -1 indicates a perfect negative correlation between the two variables.

3. Results and Discussion

The values for the gray water footprint (WF_{Gray}) and Crop Water Productivity (CWP) for the year 2022 are presented in Table 2.

According to Table 2, as the gray water footprint decreases, water productivity increases. This finding is consistent with the results of the study by Piri and Sarani (2020). Potatoes typically have high water productivity and, if managed well, a relatively low nitrogen fertilizer requirement per ton of yield.

Table 2. Values of the gray water footprint (WF_{Gray}) and water productivity (CWP) for the year 2022

Station	Crop	Water Productivity (kg/m^3)	Gray water (m^3/kg)
Azna	Potato	1.159	0.0427
	Beans	0.064	0.118
Poldokhtar	Vegetables	0.883	0.063
Kuhdasht	Grain corn	0.269	0.25
	Wheat	0.487	0.302
	Forage corn	2.173	0.05

Based on Azna-Potato (WF_{Gray}= 0.0427, CWP= 1.159) the results showed that the low WF_{Gray} suggests highly efficient fertilizer use with minimal leaching and runoff in Azna, likely due to good agricultural practices (e.g., drip irrigation, precise fertilizer application). The high CWP indicates that the crop yielded well for the amount of water it consumed.

Bazred on Azna-Beans (WF_{Gray}= 0.118, CWP= 0.064) the results showed thst this is the poorest-performing crop in the dataset. The extremely low CWP (0.064 kg/m³) means a very large amount of water was needed to produce a small yield. This could be due to drought stress, high evaporation, low-yielding varieties, or inefficient irrigation (e.g., flood irrigation). The WF_{Gray} is moderate but is magnified by the very low yield, meaning pollution per kg of product is significant.

Vegetables often have high water requirements but also high yields (Zwart and Bastiaanssen, 2004). the results showed that based on Poldokhtar - Vegetables (WF_{Gray}= 0.063, CWP= 0.883), the high CWP shows good water management. The low WF_{Gray} is impressive for vegetables, which are often heavily fertilized; it points to excellent nutrient management that minimizes pollution.

Corn is a water- and nutrient-intensive crop (Mekonnen and Hoekstra, 2011). Based on Table 2, the results of Kuhdasht - Grain Corn (WF_{Gray}= 0.25, CWP= 0.269) indicates moderately inefficient production. The CWP of 0.269 kg/m³ is on the lower side, suggesting suboptimal yield for the water used, possibly due to water stress or other growth limitations.

The WF_{Gray} is high, indicating significant fertilizer pollution per kilogram of grain produced.

Also based on the results of Kuhdasht - Wheat (WF_{Gray}= 0.302, CWP= 0.487), Wheat has the highest (worst) WF_{Gray}, meaning it generates the most pollution per kg of grain. This is common in systems with high nitrogen fertilizer application. However, its CWP is reasonable for wheat, indicating that its water use efficiency is better than that of the grain corn grown in the same region.

And finally, Forage Corn (WF_{Gray}= 0.05, CWP= 2.173) indicated the best-performing crop in the dataset. Forage corn (harvested for silage) has an outstanding CWP because the entire above-ground biomass is harvested, resulting in a very high yield per unit of water. Its WF_{Gray} is the lowest, as the nitrogen taken up by the plant is removed from the field in the harvest, leaving less to leach away as nitrate.

The data reveals critical insights into the water use efficiency and environmental impact of agriculture in these stations for 2022. There is a massive disparity between the most and least efficient crops. Forage corn is a highly productive and low-pollution system, while bean production appears to be highly unsustainable in its current form, using vast amounts of water for a meager output. The station (location) is a major factor.

Kuhdasht emerges as an area of concern. Both wheat and grain corn show low water productivity and high gray water footprints compared to global averages. This indicates a potential for serious water scarcity and pollution problems in this region. Improving irrigation efficiency (e.g., switching to drip or sprinkler from flood irrigation) and implementing precision nutrient management are critical here. Conversely, Azna (for potatoes) and Poldokhtar (for vegetables) demonstrate that efficient production is achievable. Their combination of relatively high CWP and low WF_{Gray} suggests the adoption of better management practices, such as drip irrigation and careful fertilizer application, which maximize yield while minimizing waste and pollution.

The results highlight the importance of crop selection for regional water security. Growing high-water-productivity crops like potatoes or forage corn is more sustainable in water-scarce environments than low-productivity crops like beans or inefficiently grown cereals.

3.1. Examination of the relationship between the studied variables

Figure 2 illustrates the correlation between the variables. Based on the figure 2, there is a negative correlation between CWP and WF_{Gray}.

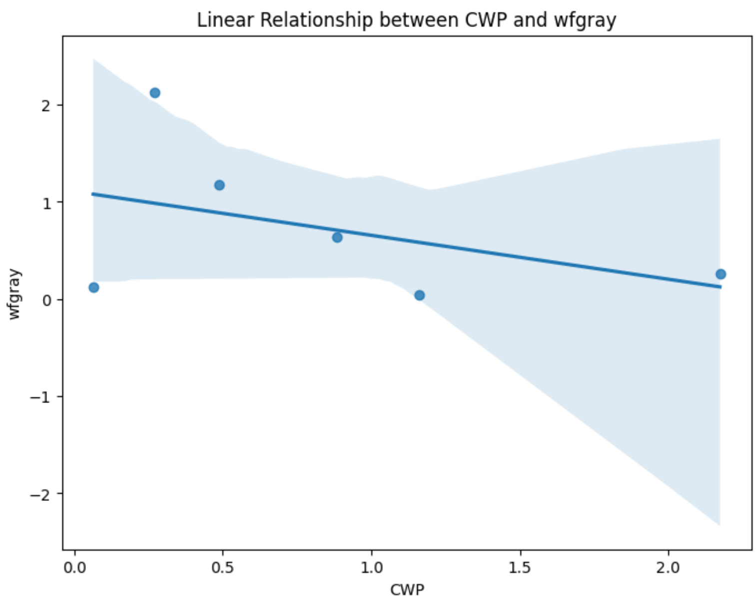

Fig. 2. the correlation between CWP and WF_{Gray}

This indicates that as water use efficiency increases, the gray water footprint decreases.

3.2. Results of gray water footprint and water productivity modeling

Table 3 presents the results obtained from the RF and SVM algorithms. In this study, a comparison was made using 7 variables: crop yield (*Y*), fertilizer application rate per hectare (kg/ha) (*NAR*), nitrogen fertilizer leaching loss (α), maximum acceptable pollutant concentration (C_{max}), natural pollutant concentration (C_{nat}), net irrigation requirement (*IR*), and gross irrigation requirement (*GI*). 80% of the data was used for the training phase and 20% for the testing phase. The error evaluation metrics for the model validation phase are presented in Table 3 for this method. Figures 3 and 4 also illustrate the model's results.

Table 3. Results of SVM-based and RF-based modeling

Metrics	WF_{Gray}_SVM	*CWP*_SVM	WF_{Gray}_RF	*CWP*_*RF*
RMSE	0.722238	0.635046	1.133599	0.301117
MAE	0.637442	0.626510	1.124770	0.236220
MSE	0.521627	0.403283	1.285048	0.090672

Based on the SVM Model in *CWP* prediction (CWP_SVM), the errors are moderate (RMSE: 0.635, MAE: 0.627). This means that, on average, the SVM model's predictions for *CWP* (in kg/m³) are off by about 0.63 units. Also for WF_{Gray} Prediction (WF_{Gray}_SVM), the errors are higher than for *CWP* (RMSE: 0.722, MAE: 0.637). This indicates that predicting WF_{Gray} is a more difficult task for the SVM model, leading to less accurate results.

Based on the RF Model in *CWP* Prediction (CWP_RF), the RF model performs exceptionally well in predicting *CWP*. The errors are low (RMSE: 0.301, MAE: 0.236). An MAE of 0.236 means the model's prediction is, on average, less than a quarter of a unit away from the actual value, which is likely a very good performance given the scale of *CWP* values (which, from the table 3, often range from 0.06 to 2.17). Also in WF_{Gray} Prediction (WF_{Gray}_RF), the performance for WF_{Gray} is significantly worse (RMSE: 1.133, MAE: 1.125). An MAE of over 1.12 means the model's predictions are, on average, off by more than 1 unit. Given that WF_{Gray} values can be as low as 0.05 (Table 3), an error of 1.13 is very large and would render the predictions highly unreliable.

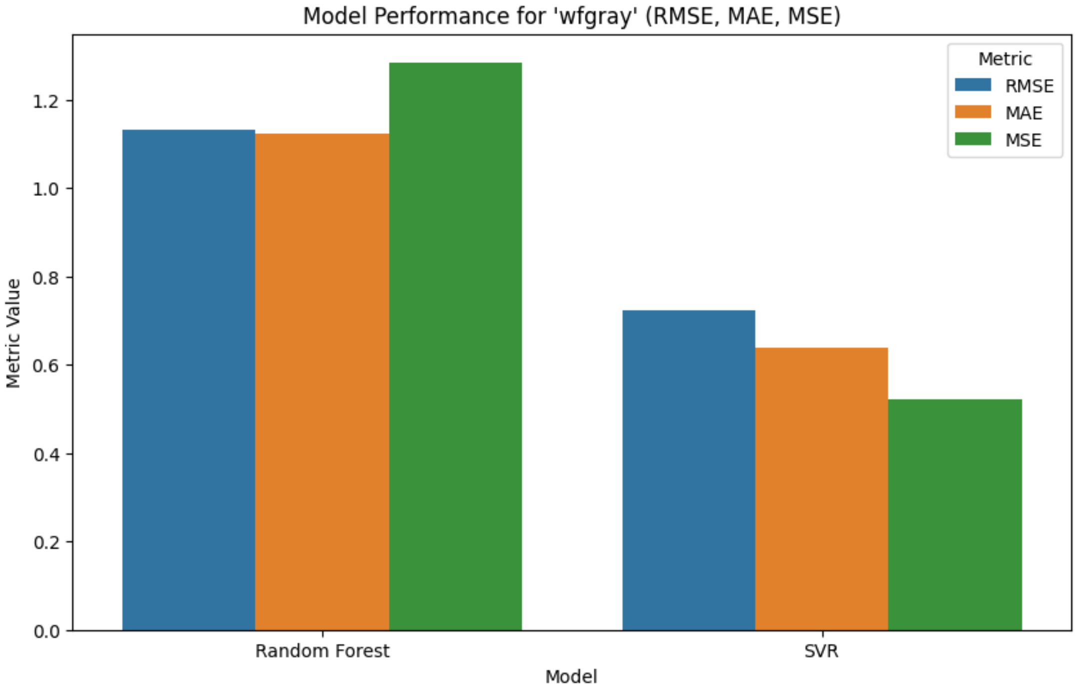

Fig. 3. Results of model performance evaluation based on WF_{Gray} data

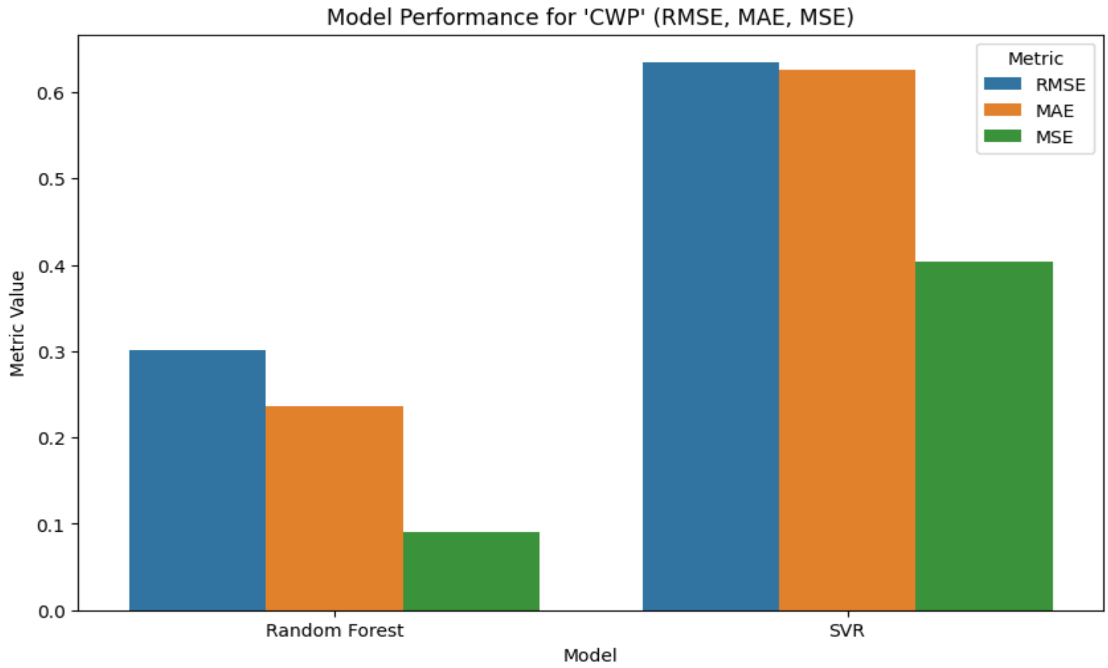

Fig. 4. Results of model performance evaluation based on *CWP* data

Overall For *CWP* prediction, the RF model is decisively superior to the SVM model. All of its error metrics (0.301, 0.236, 0.091) are less than half of those produced by the SVM (0.635, 0.626, 0.403) and for WF_{Gray} prediction, Both models perform poorly, but the SVM model performs relatively better than the RF model. The SVM's errors (~0.72, 0.64, 0.52) are substantially lower than the RF's errors (~1.13, 1.12, 1.29).

based on figure 3, the SVM model achieved lower error values for WF_{Gray}. This suggests that the relationship between the input variables and the complex output (WF_{Gray}) may be one that SVM is particularly well-suited to handle. Since, WF_{Gray}*'s* calculation is highly sensitive to the leaching factor (α), which is non-linear and can be influenced by many complex, interacting factors. SVM, especially with a non-linear kernel (like RBF), can effectively map these inputs to a high-dimensional feature space where a linear separation (or regression) is possible, potentially capturing the complex underlying patterns better than a simpler model.

Based on figure 4, RF's significantly lower errors for *CWP* indicate an excellent fit for this task. *CWP* is fundamentally a ratio of yield to water used. The relationship between water input (*IR*, *GI*) and yield (*Y*) is often strong and can be learned effectively by an ensemble of decision trees. RF excels at capturing such non-linear relationships and interactions between features (e.g., how water and fertilizer interact to affect yield) without overfitting, especially on tabular data, which is very common in agricultural studies. SVM's high errors for *CWP* suggest it was unable to effectively model the relationship from the given data. Furthermore, SVM can be less efficient and effective than tree-based models like RF on datasets with a mix of features and clear, hierarchical decision boundaries that trees can easily exploit.

As is evident from the examination of the table 3 and figures 3 and 4, among the models used, the SVM model, with the lowest mean squared error (MSE) of 0.6979, showed the best performance in predicting WF_{Gray}. For predicting *CWP*, the RF model demonstrated the best performance.

The core finding is that model performance is highly dependent on the prediction task. There is no one-size-fits-all solution. RF is the unequivocally best choice for predicting *CWP*, while SVM is the better choice for modeling the Gray Water Footprint (WF_{Gray}) with this particular dataset.

The model performance reflects the nature of the predicted phenomena. *CWP* is governed by more direct, quantifiable biophysical relationships (water in -> growth out), which RF captures superbly. WF_{Gray} is driven by a more complex, indirect, and volatile process (pollutant leaching and dilution), which appears to be better captured by the high-dimensional approach of SVM in this instance.

Researchers and agronomists can have high confidence in using a RF model to forecast water productivity and optimize irrigation strategies. Predicting environmental impact (WF_{Gray}) remains a challenging task. While SVM performed better here, the relatively higher error rates suggest that predictions should be treated with caution. To improve WF_{Gray} models, future work should focus on incorporating more direct data on soil properties and climate events that drive leaching.

4. Conclusion

This research was conducted with the aim of predicting the water footprint of agricultural products in the counties of Lorestan Province. Accordingly, two ML models were employed, and the gray water footprint was estimated using the Hoekstra and Chapagain framework. The model inputs consisted of seven variables: crop yield (*Y*), fertilizer application rate per hectare (kg/ha) (*NAR*), nitrogen fertilizer leaching loss (α), maximum acceptable pollutant concentration (C_{max}), natural pollutant concentration (C_{nat}), net irrigation requirement (*IR*), and gross irrigation requirement (*GI*). The analysis was performed for three counties.

The results indicate that the share of the gray water footprint in Kouhdasht County for wheat was the highest at 0.302 m^3/kg. The reason for this is the high consumption of chemical fertilizers and the high rate of fertilizer leaching in this region.

Water productivity in Kouhdasht County for forage corn was the highest at 2.173 kg/m^3. The higher the crop production per unit of water consumed, the higher the productivity. Gerkani Nezhad Moshizi et al. (2022), in their research on the water footprint of saffron, achieved similar results, indicating an inverse relationship between water productivity and water footprint, where an increase in the water footprint leads to a decrease in water productivity.

The findings of this research can serve as a decision-making tool for planners and operators to achieve sustainable water management in the agricultural sector. One of the most significant challenges of this study was the lack of sufficient data for use in the ML models. Furthermore, extending the statistical period could aid in better understanding the mechanism of future changes.

5. Disclosure statement

No potential conflict of interest was reported by the authors

6. References

Aligholinia, T., Rezaei, H., Behmanesh, J., & Montaseri, M. (2017). Water Footprint Index Study for dominant crops in Urmia Lake basin and its relationship with irrigation management. *Water and Soil Science*, *27*(4), 37-48.

Bageri, F., Khalili, K., & Nazeri Tahrudi, M. (2023). Evaluation of Entropy Theory Based on Random Forest in Quality Monitoring of Ground Water Network. *Water and Irrigation Management*, *13*(1), 123-139.

Behmanesh, J. (2016). Determination and evaluation of blue and green water footprint of dominant tillage crops in Urmia lake watershed. *Journal of Water and Soil Conservation*, *23*(3), 337-344. doi: 10.22069/jwfst.2016.3203

Dehghanpir, S., Bazrafshan, O., Ramezani Etedali, H., Holisaz, A., & Ababaei, B. (2023). Application of the water footprint concept in the assessment of water scarcity and water stress in the agricultural sector in Hormozgan Province. *Water and Soil Management and Modelling*, *3*(1), 233-248. doi: 10.22098/mmws.2022.11731.1163.

Ding, S., Zhu, Z., & Zhang, X. (2017). An overview on semi-supervised support vector machine. *Neural Computing and Applications*, *28*(5), 969-978.

Gerkani Nezhad Moshizi, Z., Bazrafshan, O., Ramezani Etedali, H., Esmaeilpour, Y., & Collins, B. (2022). The Effect of Past Climate Change on the Water Footprint Trend in Saffron at Homogeneous Agroclimatic Regions of Khorasan. *Journal of Saffron Research*, *10*(2), 295-311. doi: 10.22077/jsr.2022.5742.1199

Goodarzi, M., Abbasi, F., & Hedayatipour, A. (2023). Evaluation of Irrigation Water Application and Water Footprint of Major Agricultural and Horticultural Crops in the Markazi Province. *Water and Soil*, *37*(4), 503-517. doi: 10.22067/jsw.2023.81144.1253

Hoekstra, A. Y. (2008). Water neutral: reducing and offsetting the impacts of water footprints, Value of Water Research Report Series No. 28. *Delft, Netherlands: UNESCO-IHE. Recuperado em*, *10*.

Hoekstra, A. Y., & Chapagain, A. K. (2011). *Globalization of water: Sharing the planet's freshwater resources*. John Wiley & Sons.

Li, Z., Wang, W., Ji, X., Wu, P., & Zhuo, L. (2023). Machine learning modeling of water footprint in crop production distinguishing water supply and irrigation method scenarios. *Journal of Hydrology*, *625*, 130171.

Lotfy, A. A., Abuarab, M. E., Farag, E., Derardja, B., Khadra, R., Abdelmoneim, A. A., & Mokhtar, A. (2024). Forecasting Blue and Green Water Footprint of Wheat Based on Single, Hybrid, and Stacking Ensemble Machine Learning Algorithms Under Diverse Agro-Climatic Conditions in Nile Delta, Egypt. *Remote Sensing*, *16*(22), 4224.

Ma, W., Opp, C., & Yang, D. (2020). Past, present, and future of virtual water and water footprint. *Water*, *12*(11), 3068.

Madani, K. (2014). Water management in Iran: what is causing the looming crisis?. *Journal of environmental studies and sciences*, *4*(4), 315-328.

Mekonnen, M. M., & Hoekstra, A. Y. (2011). The green, blue and grey water footprint of crops and derived crop products. *Hydrology and earth system sciences*, *15*(5), 1577-1600.

Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science advances*, *2*(2), e1500323.

Nazeri Tahroudi, M., Ahmadi, F., & Mirabbasi, R. (2023). Performance comparison of IHACRES, random forest and copula-based models in rainfall-runoff simulation. *Applied Water Science*, *13*(6), 134.

Ostad-Ali-Askari, Kaveh., Kharazi, H. G.., Shayannejad, M.., & Zareian, M. J. (2019). Effect of management strategies on reducing negative impacts of climate change on water resources of the Isfahan–Borkhar aquifer using MODFLOW. *River Research and Applications*, 35, 611-631. http://doi.org/10.1002/rra.3463

Oveisi, F., Fattahi Ardakani, A., & Fehresti Sani, M. (2019). Investigation of Virtual Water and Ecological Footprints of Water in Wheat Fields of Isfahan Province. *Journal of Water and Soil Science, 23*(1), 87-99, http://jstnar.iut.ac.ir/article-1-3636-fa.html.

Piri, H., & Sarani, R. (2020). Investigation of Economic Productivity of Crop Products in Sistan and Baluchestan Province by Water Footprint Approach. *Iranian Journal of Soil and Water Research*, *51*(5), 1093-1104. doi: 10.22059/ijswr.2020.289567.668325

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101-121). Academic Press.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69-79.

Serrano, A., Guan, D., Duarte, R., & Paavola, J. (2016). Virtual water flows in the EU27: a consumption-based approach. *Journal of Industrial Ecology*, *20*(3), 547-558.

Su, H., Kang, W., Xu, Y., & Wang, J. (2018). Assessing groundwater quality and health risks of nitrogen pollution in the Shenfu mining area of Shaanxi Province, northwest China. *Exposure and health*, *10*(2), 77-97.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... & Davies, P. (2010). Global threats to human water security and river biodiversity. *nature*, *467*(7315), 555-561.

Zwart, S. J., & Bastiaanssen, W. G. (2004). Review of measured crop water productivity values for irrigated wheat, rice, cotton and maize. *Agricultural water management*, *69*(2), 115-133.