Shahid Chamran University of Ahvaz

Irrigation Sciences and Engineering (JISE)
Vol. 49, No. 2, Summer 2026, p. 119-142

Research Paper

Journal homepage: https://jise.scu.ac.ir

EXTENDED ABSTRACT

Evaluation of Optimized Hybrid SVR and RF Models for Predicting Actual Evapotranspiration of Sunflower Crop

S. Mohammadzadeh[1], V. Rezaverdinejad[2], J. Bahmanesh[3*] and M. Sharafi[4]

1-M.Sc. Student, Department of Water Science and Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran.
2-Professor, Department of Water Science and Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran.
3- Correponding Author, Professor, Department of Water Science and Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran (j.behmanesh@urmia.ac.ir).*
4- Ph.D. Graduate, Department of Water Science and Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran.

ARTICLE INFO

Artlcle history:
Received: 1 June 2025
Revised: 23 August 2025
Accepted: 15 September 2025
Available online: 05 August 2026

Keywords:
Water requirement, Machine learning, Random forest, Support vector regression.

TO CITE THIS ARTICLE:

Mohammadzadeh, ., Behmanesh, J., Rezaverdinejad, V., Sharafi, M. 2026. 'Evaluation of Optimized Hybrid SVR and RF Models for Predicting Actual Evapotranspiration of Sunflower Crop', *Irrigation Sciences and Engineering*, 49(2), pp. 119-142. doi: 10.22055/jise.2025.49823.2165.

Abstract

Accurate prediction of actual evapotranspiration in response to water resource management problems and increasing the productivity of agricultural products under dry and semi-arid climatic conditions is considered essential. Sunflower, as an economic crop with relative resistance to water deficiency, is a suitable option for cultivation in areas with water limitations. In this study, Support Vector Regression (SVR) and Random Forest (RF) models optimized with the Genetic Algorithm (GA) were used to predict actual evapotranspiration. For this purpose, data from weighing lysimeters in Bushland, Texas, USA, during the 2011–2012 growing season were extracted. A total of 230 data points were selected as model inputs in the form of eight scenarios with different combinations of climatic parameters. The results of this study showed that, in the SVR-GA model, the best performance was obtained for the combination of inputs including daily soil temperature inside and outside the lysimeter at a depth of 1.6 m, relative water storage at midnight, daily average relative water storage, soil temperature at different depths, air temperature, soil heat flux in the rows and the distance between rows, relative humidity, and wind speed, such that the Root Mean Square Error (RMSE) was 1.017 mm/day, the correlation coefficient (R) was 0.95, and the Nash Sutcliffe Efficiently (NS) was 0.908. Also, in the RF-GA model, the best results were obtained with input variables including all the above-mentioned variables except wind speed, with an RMSE of 1.516 mm/day, an R value of 0.90, and an NS value of 0.796. Comparison of the results showed that the use of the Genetic Algorithm in optimizing the models increased prediction accuracy, such that based on the RMSE criterion, the hybrid SVR-GA and RF-GA models experienced reductions of 59.91% and 35.8%, respectively, compared with their corresponding standalone models. Finally, the hybrid models can be used as efficient tools for estimating actual evapotranspiration under conditions where direct measurement using traditional methods is costly or limited.

Introduction

Evapotranspiration is a fundamental component of the hydrological cycle and plays a vital role in water resource management, irrigation design and assessing plant water requirements (Chiew et al., 1995). Since lysimeters provide relatively accurate measurements of evapotranspiration, their application is often limited due to high costs and the need for specialized expertise, particularly in developing regions (Feng & Tian, 2021). Therefore, developing alternative predictive models has

become essential. This study focuses on employing machine learning models, including random forest and support vector regression, optimized by genetic algorithms to estimate evapotranspiration. Despite prior research on reference evapotranspiration prediction, limited attention has been focused to hybrid models, especially for actual evapotranspiration forecasting due to data limitations. This research introduces a hybrid model aimed at improving prediction accuracy, applied in the Bushland region of Texas, with sunflower as a strategic crop. The model's performance is assessed through various evaluation metrics to demonstrate its efficacy compared to existing single-model approaches.

Methodology

This dataset comprised one year of lysimeter weight measurements for sunflowers collected at the USDA-ARS Conservation and Production Laboratory (CPRL), Soil and Water Management Research Unit (SWMRU) in Bushland, Texas, during the 2011–2012 crop year (Evett et al., 2000). The parameters measured include wind speed, air temperature, relative humidity, soil temperature, soil heat flux, daily evapotranspiration, daily averages and relative water storage at specific lysimeter depths. Sensitivity analysis using Pearson's correlation Figure (1) revealed that daily mean soil temperature at 1.6 m depth showed the strongest correlation (r = 0.70), while wind speed exhibited the weakest (r = −0.32). All variables were measured daily over the full crop year (Evett et al., 2018; 2022). In this study, data from two lysimeters located in the northeast and southeast of agricultural fields in Potter City were used as inputs for SVR, RF, RF-GA, and SVR-GA models to predict actual daily evapotranspiration.

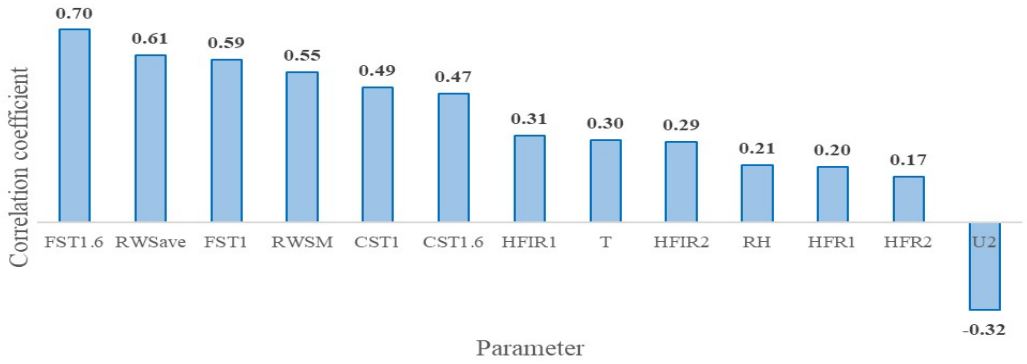


Fig. 1 – Bar chart of correlation coefficient of input parameters

Random Forest (RF)

The Random Forest (RF) model was first introduced by Breiman (2001) as a decision tree-based machine learning method. This algorithm has attracted the attention of researchers in recent years due to its various capabilities, including classification and regression. In Random Forest, a set of decision trees is built based on different parts of the training data.

Support Vector Regression (SVR)

Support Vector Regression is a support vector machine (SVM)-based method used in classification and regression analysis problems. It operates on the basis of binary classification in the desired feature space and is widely used in forecasting problems (Yoon et al. 2011; Jha & Hayashi 2014).

Genetic Algorithm (GA)

The basic concept of a genetic algorithm was first introduced by Holland (1992). This algorithm is an optimization method inspired by the principles of natural selection and genetic theory. In this method, each member of a population, called a chromosome, is considered as a possible answer to the problem.

Results and Discussion

Table (1) presents the evaluation results of the models during the test period based on R, RMSE, NRMSE, NS and WI criteria. Among the SVR scenarios, the seventh scenario showed the best performance with the lowest error and highest correlation, due to the inclusion of HFR_1 and HFR_2 variables. The SVR-GA model further improved accuracy, reducing RMSE by 1.44% compared to the best SVR scenario. These results are consistent with Sharafi et al. (2024), who applied a combined SVR-FOA model for evapotranspiration prediction in Ahvaz with similar improvements. In the RF-GA model, the seventh scenario achieved the highest accuracy, with R = 0.90, RMSE = 1.516 mm/day, NS = 0.796, NRMSE = 0.122, and WI = 0.98. Comparison of RF and RF-GA models indicates that genetic algorithm optimization significantly enhances model performance. This finding aligns with Monavvar Sabegh et al. (2023), who demonstrated improved evapotranspiration estimates at East Azerbaijan stations using a GA-optimized RF model, reporting Wilmot indices of 0.982 and 0.992 in different scenarios for Tabriz station.

Table 1 – Performance of the models during the testing period

Model	R	RMSE (mm/day)	NRMSE	NS	WI
SVR-1	0.69	2.537	0.204	0.43	0.74
SVR-2	0.80	2.044	0.164	0.63	0.89
SVR-3	0.80	2.077	0.167	0.62	0.89
SVR-4	0.84	1.932	0.155	0.67	0.92
SVR-5	0.84	1.837	0.148	0.701	0.93
SVR-6	0.84	1.841	0.148	0.70	0.93
SVR-7	0.86	1.820	0.147	0.706	0.93
SVR-8	0.84	1.835	0.148	0.702	0.93
SVR-GA-1	0.76	2.199	0.176	0.572	0.92
SVR-GA-2	0.87	1.670	0.135	0.753	0.97
SVR-GA-3	0.88	1.610	0.130	0.77	0.98
SVR-GA-4	0.89	1.533	0.124	0.792	0.98
SVR-GA-5	0.92	1.301	0.105	0.85	0.99
SVR-GA-6	0.94	1.151	0.093	0.883	0.99
SVR-GA-7	0.94	1.168	0.094	0.879	0.99
SVR-GA-8	0.95	1.017	0.082	0.908	0.99
RF-1	0.72	2.360	0.190	0.504	0.89
RF-2	0.84	1.810	0.146	0.71	0.96
RF-3	0.84	1.821	0.147	0.706	0.96
RF-4	0.86	1.712	0.138	0.74	0.97
RF-5	0.88	1.602	0.129	0.773	0.98
RF-6	0.88	1.627	0.131	0.765	0.97
RF-7	0.88	1.620	0.130	0.768	0.98
RF-8	0.89	1.618	0.130	0.768	0.98
RF-GA-1	0.76	2.178	0.176	0.580	0.92
RF-GA-2	0.84	1.805	0.146	0.711	0.96
RF-GA-3	0.86	1.755	0.142	0.727	0.96
RF-GA-4	0.87	1.672	0.135	0.752	0.97
RF-GA-5	0.88	1.603	0.129	0.773	0.98
RF-GA-6	0.89	1.551	0.125	0.787	0.98
RF-GA-7	0.90	1.516	0.122	0.796	0.98

Model	R	RMSE (mm/day)	NRMSE	NS	WI
RF-GA-8	0.90	1.539	0.124	0.790	0.98

Conclusion

This study aims to develop an accurate model for predicting actual evapotranspiration. Hybrid models can serve as efficient tools for estimating actual evapotranspiration, especially in situations where direct measurement using traditional methods is costly or limited. In this regard, the SVR, SVR-GA, RF, and RF-GA models were evaluated using available data from two lysimeters in Bushland, Texas, for the years 2011-2012. The results of the evaluations showed that the models optimized with the genetic algorithm had the highest accuracy in predicting actual evapotranspiration and presented the lowest root mean square error (1.516 and 1.017). Also, Future studies should utilize advanced validation methods like cross-validation to ensure robust model evaluation, especially with small datasets such as lysimeter data.

Conflict of Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and publication of this article. Confirmation.

Funding

The authors received no financial support for the research, authorship, and publication of this article.

Data Availability Statements

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethical considerations

The authors avoided data fabrication, falsification, and plagiarism, and any form of misconduct. Confirmation.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors state that they have not used generative AI or AI-assisted technologies in their manuscript.

Authors' contribution

First Author: Saeideh Mohammadzadeh, Writing the manuscript and literature search.
Second Author: Vahid Rezaverdinejad, Scientific supervision and guidance, reviewing the manuscript content, and revisions.
Third Author: Javad Behmanesh, Scientific supervision and guidance, reviewing the manuscript content, and revisions.
Fourth Author: Milad Sharafi, Reviewing the results and data analysis, reviewing the manuscript content, and scientific and editorial revisions.

Acknowledgement

The authors sincerely thank the USDA-ARS Conservation and Production Research Laboratory, Soil and Water Management Research Unit, Bushland, Texas, for providing the one-year sunflower weighing lysimeter data which used in this study.

References

1- Breiman, L. 2001. *Random forests*. Machine learning. 45: 5-32. doi: 10.1023/A:1010933404324
2- Chiew, FHS., Kamaladasa, NN., Malano, H.M. and McMahon, T.A. 1995. Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agricultural Water Management. 28*:9–21. Doi: 10.1016/0378-3774(95)01172-F

3- Evett, S. R., Copeland, K.S., Ruthardt, B.B., Marek, G.W., Colaizzi, P.D., Brauer, D.K. and Howell, T.A. 2022. The Bushland Maize for Grain Datasets: A Machine-readable Resource. 2022 ASABE Annual International Meeting. *American Society of Agricultural and Biological Engineers. 1*(1): 28-41. doi.org/10.13031/aim.202201078

4- Evett, S. R., Howell, T. A., Todd, R. W., Schneider, A. D. and Tolk, J. A. 2000. Alfalfa reference ET measurement and prediction. In Proceedings of the 4th Decennial *National Irrigation Symposium. 17(1)*: 14-16.

5- Evett, S.R., Marek, G.W., Copeland, K.S. and Colaizzi, P.D. 2018. Quality management for research weather data: USDA-ARS, Bushland, TX. *Agrosystems, Geosciences & Environment. 1*(1): 1-18. doi: 10.2134/age2018.09.0036

6- Feng, K. and Tian, J. 2021. Forecasting reference evapotranspiration using data mining and limited climatic data. European *Journal of Remote Sensing. 54*(2): 363-371. doi: 10.1080/22797254.2020.1801355

7- Holland, JH. 1992. *Adaptation in natural and artificial systems*: an introductory analysis with applications to biology, control, and artificial intelligence. MIT Press. doi: 10.7551/mitpress/1090.001.0001

8- Jha, S.K. and Hayashi, K. 2014. A novel odor filtering and sensing system combined with regression analysis for chemical vapor quantification. *Sensors and Actuators B: Chemical. 200*(1): 269-287. doi: 10.1016/j.snb.2014.04.022

9- Monavar Sabegh, S. , Zarehaghi, D. , Samadianfard, S. , Neishabouri, M. R. and Mikaeili, F. 2023. Estimation of Daily Reference Evapotranspiration Using Random Forest Optimized by Genetic Algorithm. *Water and Soil Science*, *33*(4), 33-53 (In Persian). doi: 10.22034/ws.2021.48756.2449. (In Persian). doi**:** 10.22034/ws.2021.48756.2449.

10- Sharafi, M. , Besharat, S. and Zeinalzadeh, K. 2024. A comparative study using a data-driven method versus a hybrid approach to estimate daily reference evapotranspiration in Ahvaz. *Irrigation Sciences and Engineering*, *47*(2), 53-68 (In Persian). doi: 10.22055/jise.2023.43524.2065.

11- Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O. and Lee, K.K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of hydrology. 396*(1-2): 128-138. doi: 10.1016/j.jhydrol.2010.11.002.

علوم و مهندسی آبیاری جلد ۴۹، شماره‌ی۲، تابستان ۱۴۰۵، مقاله پژوهشی، ص. ۱۴۲-۱۱۹

ارزیابی مدل‌های بهینه هیبریدی SVR و RF در پیش‌بینی تبخیر-تعرق واقعی محصول آفتابگردان

سعیده محمدزاده[۱]، وحید رضاوردی نژاد[۲]، جواد بهمنش [۳*] و میلاد شرفی[۴]

۱- دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.

۲- استاد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.

۳*- نویسنده مسئول، استاد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران (j.behmanesh@urmia.ac.ir)

۴- دانش آموخته دکتری، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.

دریافت: ۱۴۰۴/۰۳/۱۱ بازنگری: ۱۴۰۴/۰۶/۰۱ پذیرش: ۱۴۰۴/۰۶/۲۴ انتشار: ۱۴۰۵/۰۵/۱۴

چکیده

پیش‌بینی دقیق تبخیر-تعرق واقعی در پاسخ به مشکلات مدیریت منابع آبی و افزایش بهره‌وری محصولات کشاورزی در شرایط اقلیمی خشک و نیمه‌خشک امری ضروری تلقی می‌شود. محصول آفتابگردان به‌عنوان محصول اقتصادی با دارا بودن مقاومت نسبی به کم‌آبی، گزینه‌ای مناسب برای کشت در مناطق با محدودیت آبی است. در این پژوهش، از مدل‌های رگرسیون بردار پشتیبان (SVR) و جنگل تصادفی (RF) بهینه شده با الگوریتم ژنتیک (GA) برای پیش‌بینی تبخیر-تعرق واقعی استفاده شده است. برای این منظور داده‌های لایسیمترهای وزنی در منطقه بوشلند، تگزاس، ایالات متحده طی سال زراعی ۲۰۱۲-۲۰۱۱ استخراج شد. تعداد ۲۳۰ داده در قالب هشت سناریو با ترکیب‌هایی از پارامترهای اقلیمی مختلف، به‌عنوان ورودی مدل انتخاب شدند. نتایج این تحقیق نشان داد در مدلSVR-GA، بهترین عملکرد مربوط به ترکیب ورودی‌هایی بود که شامل دمای روزانه خاک داخل و خارج لایسیمتر در عمق ۱/٦ متری، ذخیره نسبی آب در نیمه‌شب، میانگین روزانه ذخیره نسبی، دمای خاک در عمق‌های مختلف، دمای هوا، شار حرارتی خاک در ردیف‌ها و فاصله بین‌ردیف‌ها، رطوبت نسبی و سرعت باد بودند، به‌طوری‌که مقدار جذر میانگین مربعات خطا (RMSE)برابر با ۱/۰۱۷ میلی‌متر بر روز، ضریب همبستگی (R) معادل ۰/۹۵ و ضریب نش-ساتکلیف (NS) برابر با ۰/۹۰۸ حاصل شد. همچنین در مدلRF-GA ، بهترین نتایج با متغیرهای ورودی شامل همه موارد فوق به‌جز سرعت باد به‌دست آمد و RMSE برابر با ۱/۵۱٦ میلی‌متر در روز، R معادل ۰/۹۰ و NS برابر با ۰/۷۹٦ بود. مقایسه نتایج نشان داد که استفاده از الگوریتم ژنتیک در بهینه‌سازی مدل‌ها باعث افزایش دقت پیش‌بینی شده است، به‌طوری‌که بر اساس معیار RMSE، مدل‌های هیبریدی SVR-GA و RF-GA نسبت به مدل‌های مستقل خود به‌ترتیب کاهش ۵۹/۹۱ و ۳۵/۸ درصدی را تجربه کردند. در نهایت، مدل‌های هیبریدی می‌توانند به‌عنوان ابزاری کارآمد برای تخمین تبخیر-تعرق واقعی در شرایطی که اندازه‌گیری مستقیم با روش‌های سنتی پرهزینه یا محدود است، مورد استفاده قرار گیرند.

کلید واژه‌ها: نیاز آبی، یادگیری ماشین، جنگل تصادفی، رگرسیون بردار پشتیبان.

مقدمه

تبخیر-تعرق یکی از فرآیندهای اساسی چرخه هیدرولوژیکی به‌شمار می‌آید و برآورد دقیق این پارامتر، برای ارزیابی بیلان هیدرولوژیکی، طراحی سیستم‌های آبیاری، تعیین نیاز آبی گیاهان و مدیریت منابع آب، اهمیت ویژه‌ای دارد (Chiew et al., 1995). گیاه آفتابگردان به‌دلیل ویژگی‌های کلیدی نظیر دوره رشد کوتاه، نیاز آبی نسبتاً پایین، توانایی سازگاری در مقابل خشکسالی و تحمل تنش‌های آبی کوتاه‌مدت، به‌عنوان یکی از محصولات زراعی استراتژیک برای کشت در مناطق خشک و نیمه‌خشک شناخته می‌شود (Lamm et al., 2011; Tolk and Howell, 2012). روش‌های متعددی برای اندازه‌گیری تبخیر-تعرق وجود دارد که از جمله این روش‌ها می‌توان به استفاده از لایسیمترها Dong و Hansen (2023) روش نسبت بوون مبتنی‌بر تعادل انرژی و جرم بخار Xiong et al (2022) و همچنین تشتک‌های تبخیر Choudhury (1999) اشاره کرد. لایسیمترها ابزار مؤثری برای اندازه‌گیری تبخیر-تعرق هستند که می‌توانند به‌طور نسبی داده‌های دقیقی از این فرآیند ارائه دهند. نتایج حاصل از لایسیمترها در بهبود و ارزیابی سایر روش‌ها نیز مورد استفاده قرار می‌گیرد. با این حال، استفاده از لایسیمتر به‌دلیل نیاز به نیروی متخصص و محدودیت دسترسی به این ابزار در برخی کشورهای در حال توسعه، عمدتاً به امور تحقیقاتی محدود می‌شود (Feng & Tian, 2021). در چنین شرایطی، توسعه مدل‌های ریاضی که قادر به پیش‌بینی دقیق روابط پیچیده میان پارامترهای ورودی و خروجی هستند، اهمیت زیادی پیدا می‌کند. این مدل‌ها می‌توانند جایگزینی مؤثر برای روش‌های پرهزینه و زمان‌بر قدیمی باشند و در مطالعه‌های مربوط به تبخیر-تعرق کاربرد گسترده‌ای پیدا

کرده‌اند (Wu et al., 2019). روش‌های مبتنی‌بر هوش مصنوعی (Artificial Intelligence) با داشتن قابلیت‌هایی مانند استفاده از تعداد زیادی متغیر در مدل و توانایی شناسایی پارامترها حتی در شرایط وجود داده‌های ناقص، در مقایسه با روش‌های سنتی برتری قابل‌توجهی دارند (Hashemian et al., 2019). این روش‌ها با داشتن دقت و سرعت بالاتر در پیش‌بینی تبخیر-تعرق از اهمیت ویژه برخوردار می‌باشند (Bayat et al., 2009). جنگل تصادفی (Random forest) یک الگوریتم یادگیری ماشینی است که به‌دلیل عملکرد قوی در طیف گسترده‌ای از مجموعه داده‌ها، در تحقیقات زیادی مورد استفاده قرار گرفته است (Jing et al., 2015 and Da Silva Junior et al., 2019).

در مطالعه‌ای Wang et al (2019) به پیش‌بینی تبخیر-تعرق مرجع با استفاده از مدل‌های جنگل تصادفی و مقایسه آن با روش برنامه‌ریزی ژن (gene-expression programming) در جنوب غربی چین پرداختند. نتایج حاصل از این مطالعه نشان داد که مدل جنگل تصادفی با داشتن ضریب کارایی نش‌ساتکلیف (Nash Sutcliffe coefficient of efficiency) بین ۰/۹۸۶ - ۰/۶۲۶، ریشه میانگین مربع خطا (Root mean square error) بین ۰/۵۶۳- ۰/۱۰۷ میلی‌متر درصد عملکرد خوبی با ترکیب ورودی‌های مختلف در مقایسه با مدل برنامه‌ریزی ژن داشته‌است. Feng et al. (2017) در تحقیقی عملکرد دو مدل یادگیری ماشین از جمله جنگل‌تصادفی و رگرسیون شبکه‌های عصبی (Generalized regression neural networks) در تخمین تبخیر-تعرق مرجع روزانه را در جنوب غربی چین ارزیابی کردند نتایج این مطالعه نشان داد که عملکرد هر دو مدل برای تخمین تبخیر-تعرق روزانه مناسب بوده، اما عملکرد مدل جنگل تصادفی به‌علت داشتن RMSE پایین‌تر (۰/۰۷۵-۰/۰۶۱) بهتر از روش رگرسیون شبکه‌های عصبی بوده است. در پژوهشی Shiri et al (2021) به ارزیابی دقت تخمین تبخیر-تعرق مرجع با استفاده از داده‌های اقلیمی در سه ایستگاه مرطوب کشور با استفاده از مدل جنگل تصادفی و سایر مدل‌های تجربی پرداختند. نتایج این تحقیق نشان داد که مدل جنگل تصادفی نسبت به مدل‌های تجربی واسنجی‌شده و واسنجی‌نشده که دارای متغیرهایی مانند دمای هوا و تابش خورشیدی می‌باشد با داشتن میانگین قدر مطلق خطا برابر ۰/۴ میلی‌متر بر روز و داشتن شاخص پراکندگی ۰/۰۰۸ دقت بیش‌تری را در تخمین ارائه داده است. رگرسیون بردار پشتیبان (Support vector regression) در سال‌های اخیر به‌عنوان یک رویکرد قوی و دقیق در پیش‌بینی تبخیر-تعرق شناخته شده است. Mokari et al. (2022) در تحقیقی به ارزیابی قدرت چند مدل از جمله یادگیری ماشین افراطی (Extreme learning machine)، برنامه‌ریزی ژنتیک (Genetic programming) و (SVR) برای پیش‌بینی تبخیر-تعرق پرداختند. برای انجام این تحقیق از داده‌های هواشناسی بازه ده ساله (۱۳۹۷-۱۳۸۷) در منطقه نیومکزیکو استفاده شد. نتایج نشان داد که مدل رگرسیون بردار پشتیبان با داشتن بالاترین پایداری در مرحله آزمون که در آن کاهشی به‌میزان کمتر از ۱۰درصد داشت عملکرد بهتری نسبت به سایر مدل‌ها نشان داد. Zouzou و Citakoglu (2023) در مطالعه‌ای به پیش‌بینی تبخیر-تعرق با استفاده از مدل‌های یادگیری ماشین ازجمله رگرسیون بردار پشتیبان، جنگل‌تصادفی و رگرسیون فرایند گاوسی (Gaussian process regression) در ترکیه پرداختند نتایج حاصل از این مطالعه نشان داد که مدل SVR موجب کاهش ریشه میانگین مربعات خطا نسبت به سایر مدل‌های یادگیری ماشین گردید. Kumar و Singh (2022) در پژوهشی روش‌های داده‌محور در مدل‌سازی سری زمانی فرآیندهای هیدرولوژیکی را در منطقه Ranichauri هند مقایسه کردند. نتایج این تحقیق نشان داد که مدل SVR با داشتن RMSE بین مقادیر ۰/۶۸۹-۰/۵۴۲ و ضریب NSE بین مقادیر ۰/۹۷۴-۰/۹۵۳، دقت قابل قبولی نسبت به سایر مدل‌ها داشته است. استفاده از مدل‌های هیبریدی موجب افزایش دقت و سرعت در مدل‌های ریاضی در مقایسه با مدل‌های منفرد شده است و مورد توجه محققان زیادی قرار گرفته است (Talukdar et al., 2022; Saravi et al., 2022). در تحقیقی Khosravi et al. (2019) به پیش‌بینی تبخیر-تعرق مرجع در موصل عراق با استفاده از مدل‌های (Adaptive neuro-fuzzy inference system) و مدل هیبریدی ANFIS-GA پرداختند. نتایج این مطالعه نشان داد که مدل هیبریدی ANFIS-GA با داشتن مقادیر (R^2=۰/۹۷ و RMSE= ۳۲/۸۱ mm) بهترین عملکرد را در ایستگاه بغداد در پیش‌بینی تبخیر-تعرق نسبت به سایر مدل‌ها داشته است. در تحقیقی Sharafi و Samadianfard (2024) به پیش‌بینی تبخیر-تعرق روزانه در چهار ایستگاه ساحلی با استفاده از مدل‌های هیبریدی (SVR-FFA) و مدل یادگیری SVR پرداختند. نتایج این تحقیق نشان داد که مدل هیبریدی SVR-FFA با داشتن خطای ۱/۱۳ میلی‌متر بر روز در مقایسه با مدل SVR با خطای ۱/۲۵ میلی‌متر بر روز، از عملکرد و دقت بالایی برخوردار بوده است. در تحقیقی Monavar Sabegh et al. (2023) به ارزیابی عملکرد مدل جنگل تصادفی و مدل بهینه شده جنگل تصادفی با الگوریتم ژنتیک (Genetic algorithm) درکنار معادله‌های تجربی برای پیش‌بینی تبخیر-تعرق در استان آذربایجان شرقی طی دوره ۲۰۰۰-۲۰۲۰ با استفاده از پارامترهای حداقل و حداکثر دما، حداقل و حداکثر رطوبت نسبی، ساعات آفتابی و سرعت باد پرداختند. نتایج این بررسی نشان داد که الگوریتم ژنتیک بهینه شده با کاهش شاخص پراکندگی عملکرد مناسبی داشت.

این تحقیق با ارزیابی و مقایسه عملکرد مدل‌های یادگیری ماشین رگرسیون بردار پشتیبان SVR و جنگل تصادفی RF بهینه‌شده با الگوریتم ژنتیک (GA) به پیش‌بینی تبخیر-تعرق واقعی آفتابگردان با استفاده از اطلاعات لایسیمتری در بوشلند تگزاس می‌پردازد. بهینه‌سازی این مدل‌ها از طریق الگوریتم ژنتیک

DOI: 10.22055/jise.2025.49823.2165

انجام می‌شود تا دقت پیش‌بینی افزایش یافته و حداکثر خطاهای محاسباتی به حداقل برسد. هدف از این پژوهش ارائه مدلی است که از نظر قابلیت اطمینان و کارایی در پیش‌بینی تبخیر–تعرق واقعی عملکرد قابل توجهی داشته باشد. این مدل با بهره‌گیری از یک روش هیبریدی تلاش می‌کند تا دقت پیش‌بینی را نسبت به مدل‌های منفرد افزایش دهد و تخمین‌های دقیق و قابل اعتمادی را از تبخیر–تعرق واقعی ارائه دهد. بررسی جامع مطالعه‌های پیشین نشان می‌دهد که تحقیق‌های محدودی در زمینه مدل‌های هیبریدی برای پیش‌بینی تبخیر–تعرق مرجع انجام شده است. بااین‌حال، اغلب تحقیقات پیشین بر پیش‌بینی تبخیر–تعرق مرجع متمرکز بوده‌اند و تبخیر–تعرق واقعی (ET_a) کمتر مورد توجه قرار گرفته است. همچنین، به‌دلیل محدودیت دسترسی به داده‌های دقیق لایسیمتری، استفاده از این داده‌ها در مدل‌سازی Et_a تاکنون محدود بوده است. تحقیق حاضر با ارائه یک مدل هیبریدی، دقت و عملکرد پیش‌بینی را نسبت به سایر مدل‌های منفرد بهبود می‌بخشد. منطقه بوشلند، تگزاس، ایالات متحده با اقلیم خشک و دسترسی به داده‌های دقیق لایسیمتری، با هدف ارتقای دقت مدل‌سازی و کاهش عدم قطعیت‌ها به‌عنوان منطقه مورد مطالعه انتخاب شد. آفتابگردان به‌عنوان یک محصول با مقاومت نسبی به کم‌آبی، نقش قابل توجهی را در کشاورزی مناطق خشک ایفا می‌کند و پیش‌بینی دقیق تبخیر–تعرق واقعی در این گیاه به مدیریت بهینه آبیاری و افزایش بهره‌وری کمک می‌کند. همچنین، داده‌های لایسیمتری به‌دلیل دقت بالا در اندازه‌گیری تبخیر–تعرق، مبنای معتبری برای توسعه مدل‌های پیش‌بینی هستند. از سوی دیگر، مدل‌های SVR و RF به‌واسطه توانایی در شناسایی روابط غیرخطی، عملکرد مؤثر در تحلیل داده‌های پیچیده و قابلیت تلفیق با الگوریتم ژنتیک، برای این پژوهش انتخاب شدند. نتایج این تحقیق علاوه‌بر بهبود عملکرد و دقت در پیش‌بینی تبخیر–تعرق واقعی موجب کمک به صرفه‌جویی در مصرف آب و توسعه پایدار کشاورزی، کاهش فشار بر منابع آبی و افزایش بهره‌وری این محصول در مناطق خشک و نیمه‌خشک می‌شود.

مواد و روش‌ها

منطقه مورد مطالعه

داده‌های این مجموعه شامل اطلاعات یک‌ساله لایسیمتر وزنی برای گیاه آفتابگردان در آزمایشگاه حفاظت و تولید USDA-ARS (Conservation and Production Laboratory)، واحد تحقیقات مدیریت آب و خاک (Soil and Water Management Research Unit) در بوشلند، تگزاس (عرض جغرافیایی ۳۵/۱۸۶، طول جغرافیایی ۱۰۲/۰۹۴–، ارتفاع ۱۱۷۰ متر بالاتر از میانگین سطح دریا) در سال زراعی ۲۰۱۲–۲۰۱۱ می‌باشد (Evett et al., 2000). موقعیت مکانی منطقه مورد مطالعه در شکل (۱) آورده شده است. گیاه آفتابگردان بر روی دو لایسیمتر بزرگ و با دقت توزین زیاد، هریک در اطراف یک زمین مربعی‌شکل به مساحت ۴/۴۴ هکتار کشت شده است. پارامترهای اندازه‌گیری‌شده شامل سرعت باد در ارتفاع دو متری (U_2)، میانگین روزانه دمای هوا در همان ارتفاع (T)، رطوبت نسبی (RH)، میانگین دمای روزانه خاک در عمق ۱/۶ متری داخل لایسیمتر ($CST_{1.6}$)، میانگین دمای روزانه خاک در عمق ۱/۶ متری خارج از لایسیمتر ($FST_{1.6}$)، میانگین دمای روزانه خاک در عمق یک متری داخل لایسیمتر (CST_1)، میانگین دمای روزانه خاک در عمق یک متری خارج از لایسیمتر (FST_1)، مجموع شار حرارتی خاک در اولین فاصله بین ردیف‌ها ($HFIR_1$)، مجموع شار حرارتی خاک در دومین فاصله بین ردیف‌ها ($HFIR_2$)، مجموع شار حرارتی خاک در اولین ردیف از لبه جنوبی لایسیمتر (HFR_1)، شار حرارتی خاک در ردیف دوم از لبه جنوبی لایسیمتر (HFR_2)، تبخیر–تعرق واقعی روزانه (ET_a)، میانگین روزانه ذخیره نسبی آب در لایسیمتر (RWS_{ave}) و ذخیره نسبی آب در لایسیمتر برحسب میلی‌متر در نیمه‌شب (RWS_M) است. مشخصات کامل این پارامترها در جدول (۱) ارائه شده است. همچنین شکل (۲) آنالیز حساسیت پارامترهای ورودی بر پارامتر هدف را با استفاده ضریب همبستگی پیرسون نشان می‌دهد. مقایسه نتایج آنالیز حساسیت نشان می‌دهد که میانگین دمای روزانه خاک در عمق ۱/۶ متری با داشتن ضریب همبستگی ۰/۷ بیش‌ترین تأثیر و پارامتر سرعت باد با مقدار ۰/۳۲– کم‌ترین تأثیر را بر روی پارامتر هدف داشته‌اند.

تمامی پارامترها در مقیاس روزانه در طول یک سال زراعی اندازه‌گیری شدند (Evett et al., and Evett et al., 2022 2018). همچنین، روش‌های پردازش، نظیر حذف داده‌های پرت و ناقص، پس از جمع‌آوری داده‌ها برای کاهش خطاهای محاسباتی مرتبط با تبخیر–تعرق لایسیمتر به‌کار گرفته شد (Marek et al 2014). در این تحقیق، پارامترهای به‌دست‌آمده از دو لایسیمتر واقع در شمال شرق (Northeast) و جنوب شرق (Southeast) در اراضی کشاورزی شهر پاتر (Potter County) به‌عنوان ورودی مدل‌های SVR، RF، RF-GA و SVR-GA مورد استفاده قرار گرفتند تا مقادیر تبخیر–تعرق واقعی روزانه پیش‌بینی شود.

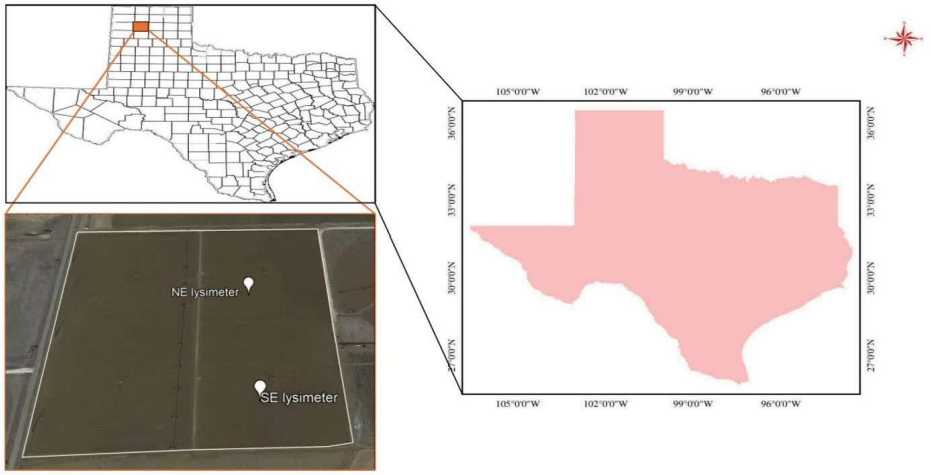


Fig. 1 – Location of the Study Area in Bushland, Texas
شکل ۱- موقعیت منطقه مورد مطالعه در بوشلند، تگزاس

جدول ۱- مشخصات پارامترهای مورد استفاده
Table 1 - Specifications of Parameters Used

Parameter Name	Description
$CST_{1.6}$	Daily average soil temperature at 1.6 m depth inside the lysimeter, measured about 1 meter from the lysimeter wall using a thermocouple.
$FST_{1.6}$	Daily average soil temperature at 1.6 m depth outside the lysimeter, measured about 1 meter from the lysimeter chamber wall using a thermocouple.
CST_1	Daily average soil temperature at 1 m depth inside the lysimeter, measured about 1 meter from the lysimeter wall using a thermocouple.
FST_1	Daily average soil temperature at 1 m depth outside the lysimeter, measured about 1 meter from the lysimeter chamber wall using a thermocouple.
T	Daily average air temperature measured at 2 meters above the soil surface.
ET_a	Daily actual evapotranspiration calculated based on lysimeter mass change, precipitation, irrigation, and other factors.
RH	Daily average relative humidity (%) measured by a Vaisala sensor at 2 meters above the soil surface inside the lysimeter.
$HFIR_2$	Daily total soil heat flux (MJ/m²) measured at 50 mm depth in the second inter-row spacing inside the lysimeter using a heat flux plate, corrected thermometrically to the surface.
HFR_1	Daily total soil heat flux (MJ/m²) measured at 50 mm depth in the first row from the southern edge of the lysimeter using a heat flux plate, corrected thermometrically to the surface.
HFR_2	Daily total soil heat flux (MJ/m²) measured at 50 mm depth in the second row from the southern edge of the lysimeter using a heat flux plate, corrected thermometrically to the surface.
$HFIR_1$	Daily total soil heat flux (MJ/m²) measured at 50 mm depth in the first inter-row spacing inside the lysimeter using a heat flux plate, corrected thermometrically to the surface.
RWS_{ave}	Daily average relative water storage (mm) in the lysimeter based on calibration of lysimeter scale at water depth, related to an initial arbitrary storage value at water depth.
RWS_M	Relative water storage (mm) in the lysimeter at midnight based on calibration of lysimeter scale at water depth, related to an initial arbitrary storage value at water depth.
U_2	Daily average wind speed measured by an anemometer at 2 meters above the soil surface inside the lysimeter.

 DOI: 10.22055/jise.2025.49823.2165

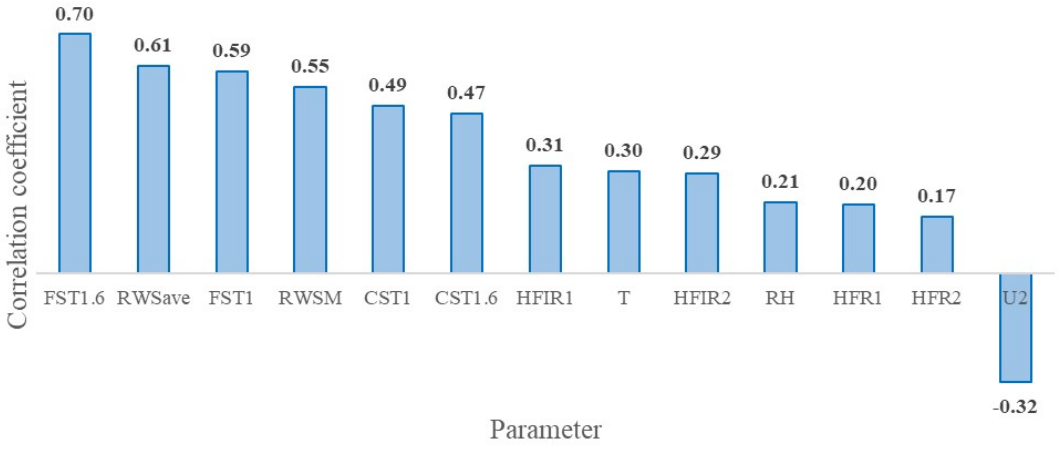


Fig. 2 – Bar chart of correlation coefficient of input parameters
شکل ۲- نمودار میله‌ای ضریب همبستگی پارامترهای ورودی

مدل جنگل تصادفی (RF)

مدل جنگل تصادفی (RF) نخستین بار توسط Breiman (2001) به‌عنوان یک روش یادگیری ماشین مبتنی‌بر درخت تصمیم شناخته شد. این الگوریتم با داشتن قابلیت‌های مختلف از جمله طبقه‌بندی و رگرسیون، در سال‌های اخیر مورد توجه محققان واقع شده است. در این مدل ابتدا ویژگی‌های ورودی منتخب استخراج و پارامترهای مدل تنظیم می‌شوند، سپس داده‌ها به مجموعه‌های آموزش و آزمون تقسیم شده و توسط مدل جنگل تصادفی آموزش داده می‌شود. درآخر خروجی مدل با داده‌های واقعی مقایسه شده و با استفاده از شاخص‌های عملکرد ارزیابی می‌گردد (شکل ۳). در جنگل تصادفی، مجموعه‌ای از درخت‌های تصمیم بر اساس بخش‌های مختلف داده‌های آموزشی ساخته می‌شوند. یکی از ویژگی‌های منحصر به فرد این الگوریتم، توانایی آن در ارزیابی اهمیت پیش‌بینی هر یک از متغیرهای ورودی است. همچنین در این مدل از مجموعه متغیرهایی استفاده می‌شود که می‌توانند به بهبود دقت پیش‌بینی‌کمک کنند (Tyralis et al., 2019). در این روش، هر درخت به‌طور مستقل بر اساس زیرمجموعه‌ای از داده‌ها، آموزش می‌بیند و بدین‌ترتیب مدل‌های مختلفی برای پیش‌بینی ایجاد می‌شود. تنوع درخت‌ها از طریق تولید زیرمجموعه‌های مختلف از داده‌های آموزشی کاهش می‌یابد که این فرایند تحت عنوان طبقه‌بندی شناخته می‌شود. این روش به کاهش پراکندگی مدل و افزایش دقت مدل کمک می‌کند. به‌علاوه، در این فرایند ممکن است برخی از داده‌ها بیش از یک‌بار در مدل‌ها به‌کار گرفته شوند که در نهایت موجب افزایش پایداری مدل می‌شود. این ویژگی باعث می‌شود که مدل در برابر تغییرات کوچک در داده‌ها مقاوم‌تر باشد و دقت پیش‌بینی به‌طور قابل توجهی افزایش یابد. در نهایت، این الگوریتم با استفاده از یک روش تصمیم‌گیری مبتنی‌بر رأی‌گیری، پیش‌بینی نهایی را از طریق انتخاب کلاسه‌ای که بیش‌ترین تعداد رأی را کسب کرده است، انجام می‌دهد.

رگرسیون بردار پشتیبان (SVR)

رگرسیون بردار پشتیبان یک روش مبتنی‌بر ماشین بردار پشتیبان (Support vector machine) است که در دسته‌بندی و تحلیل مسائل رگرسیونی کاربرد دارد. این روش براساس طبقه‌بندی دوگانه در فضای ویژگی‌های مطلوب عمل کرده و به‌طور گسترده در مسائل پیش‌بینی مورد استفاده قرار می‌گیرد (Yoon et al. 2011; Jha & Hayashi 2014). SVR از یک تابع تلفات (insensitive loss function) برای تحمل خطا در داده‌های آموزشی استفاده می‌کند. این تابع به SVR اجازه می‌دهد تا خطاها را در یک محدوده مشخص نادیده بگیرد، و خطاهایی که از این محدوده فراتر می‌روند، در فرآیند یادگیری در نظر گرفته می‌شوند. معادله (۱) تابع SVR را به‌صورت زیر نشان می‌دهد.

$$F(x) = \mathrm{W}^T X + b \qquad (۱)$$

که، W بردار وزن عمود بر ابرصفحه، T ترانهاده، b مقدار بایاس، X بردار ورودی و $F(x)$ بردار خروجی است.

$C > 0$ یک هایپرپارامتر تنظیم است که باید به‌صورت دقیق انتخاب شود، به‌علاوه این پارامتر به‌عنوان شاخصی برای ارزیابی خطای آزمون درنظر گرفته می‌شود. C فاصله میان مقادیر واقعی و مقادیر متناظر در محدوده $\varepsilon - Tube$ را تعریف می‌کند که با ξ_j و ξ_i^* مشخص می‌شود. برای حل مسائل بهینه‌سازی در SVR از متغیرهای لاگرانژی استفاده می‌شود که امکان انعطاف‌پذیری مرزها را در حین حل مسئله بهینه‌سازی فراهم می‌کند. در معادله

(۲) قیود این تابع تعریف شده است که در آن x_i بردار ورودی و y_i بردار خروجی را نشان می‌دهد.

$$\begin{cases} F(x) - y_i \leq \varepsilon + \xi_i^* \\ \mathrm{y_i} - \mathrm{F(x)} \leq \varepsilon + \xi_\mathrm{i} \\ \mathrm{i} = 1, 2, 3, \ldots, \mathrm{N} \\ \xi_\mathrm{i}^*, \xi_j \geq 0 \end{cases} \quad (۳)$$

برای کمینه‌سازی معادله (۱) تحت قیود معادله (۲) تابع زیر بر اساس مطالعه‌های Gun (1998) و Cimen (2008) استخراج شده است.

$$F(x) = \sum_{i=1}^{N}(a_i^* - a_i)K(x, x_i) + b \quad (۳)$$

در معادله (۳)، $K(x, x_i)$ به‌عنوان تابع کرنل تعریف شده است، $a_i, a_i^* \geq 0$ ضرایب لاگرانژی را نشان می‌دهند و b بیانگر جمله بایاس است. به‌طور کلی تابع کرنل در SVR برای حل مسائل بهینه‌سازی مورد استفاده قرار می‌گیرد (Smola and Schölkopf, 2004).

$$K(x, x_i) = exp(\gamma \|x - x_i\|^2) \quad (۴)$$

γ در این معادله به‌معنای عرض تابع کرنل است که نقش کلیدی در تنظیم رفتار تابع کرنل دارد. شکل (۴) فلوچارت مدل رگرسیون بردار پشتیبان را ارائه می‌دهد. در این فرآیند اجرای مدل با پیش‌پردازش داده‌ها آغاز می‌شود تا ویژگی‌ها به‌شکل مناسبی برای تحلیل آماده شوند. سپس، تابع کرنل متناسب با ویژگی‌های داده انتخاب می‌شود تا فضا ویژگی‌ها بهینه شود. پس از آن، مدل با استفاده از داده‌های آموزشی اعتبارسنجی شده و عملکرد آن ارزیابی می‌گردد. در نهایت، هایپرپارامترها با هدف بهینه‌سازی دقت مدل تنظیم شده و فرایند آموزش تکمیل می‌شود.

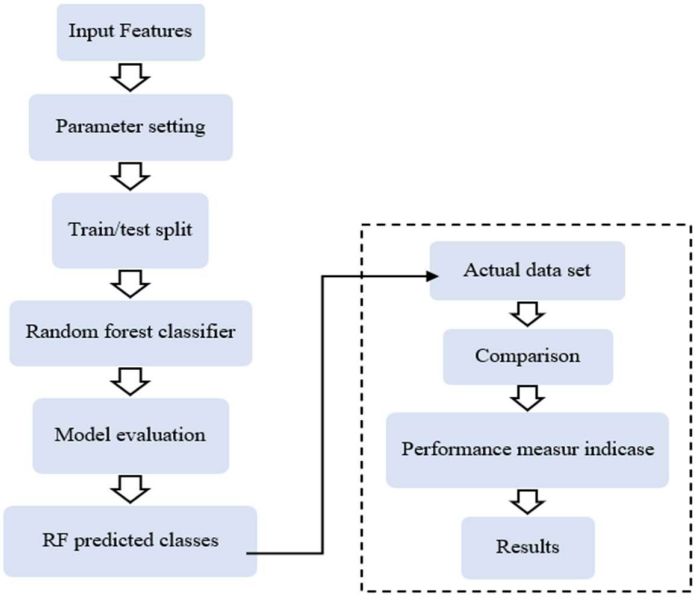


Fig. 3 – Flowchart of the Random Forest (Achite, et al., 2023)
شکل ۳- فلوچارت جنگل تصادفی(Achite, et al., 2023)

DOI: 10.22055/jise.2025.49823.2165

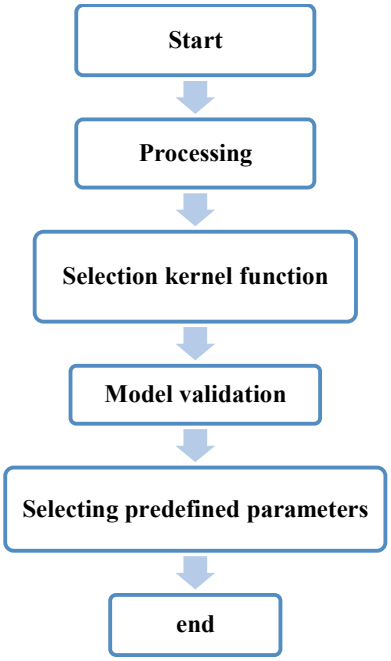


Fig. 4 – Flowchart of Support Vector Regression (Achite, et al., 2024)
شکل ٤- فلوچارت رگرسیون بردار پشتیبان (Achite, et al., 2024)

الگوریتم ژنتیک (GA)

مفهوم اولیه الگوریتم ژنتیک برای نخستین بار توسط Holland (1992) معرفی شد. این الگوریتم یک روش بهینه‌سازی است که از اصول انتخاب طبیعی و نظریه ژنتیک الهام گرفته است. در این روش، هر عضو یک جمعیت به‌نام کروموزوم، به‌عنوان یک پاسخ احتمالی به مسئله در نظر گرفته می شود. جستجو در این الگوریتم با تولید تصادفی جمعیت اولیه شروع می‌شود و نسل‌های بعدی این جمعیت توسط عملگرهای انتخاب، تقاطع و جهش گسترش می‌یابند. در ادامه، به تشریح مراحل اصلی الگوریتم ژنتیک پرداخته شده است.

۱- تولید جمعیت اولیه: الگوریتم با تولید تصادفی یک جمعیت اولیه از کروموزوم‌ها شروع می‌شود. هر کروموزوم یک راه‌حل بالقوه برای مسئله بهینه‌سازی است.

۲- انتخاب (Selection): براساس اصل بقا، کروموزوم‌هایی که دارای شایستگی بیش‌تری هستند (راه حل‌های بهینه‌ای برای مسئله ارائه می دهند) برای تولید نسل بعدی انتخاب می‌شوند.

۳- تقاطع (Crossover): در این مرحله، کروموزوم‌های انتخاب شده باهم ترکیب می‌شوند تا کروموزوم‌های جدیدی ایجاد شوند.

۴- جهش (Mutation): در این مرحله، برخی‌از ژن‌های کروموزوم‌ها به‌طور تصادفی تغییر می‌کنند، این عمل باعث می‌شود که الگوریتم از گیر افتادن در بهینه‌های محلی جلوگیری کند و به‌دنبال راه‌حل‌های جدیدی باشد.

۵- ارزیابی: نسل جدید جمعیت، دوباره ارزیابی می شود تا مشخص شود که کدام کروموزوم‌ها برازش بهتری دارند. این روند تا زمانی‌که یک راه حل بهینه و معیار همگرا برای مسئله پیدا شود، تکرار می‌شود.

بر اساس توضیحاتی که پیش‌تر ارائه شد، شکل (۵)، فلوچارت الگوریتم ژنتیک را نمایش می‌دهد. در این فلوچارت، فرآیند الگوریتم ژنتیک با ایجاد جمعیت اولیه آغاز شده و سپس، هر عضو جمعیت بر اساس معیارهای مسئله مورد ارزیابی قرار می‌گیرد تا کارایی آن مشخص شود. در مرحله بعد، فرآیند انتخاب (Selection) انجام می‌شود تا بهترین اعضا برای تولید نسل بعدی برگزیده شوند. پس از آن، عملیات تقاطع(Crossover) به‌منظور ترکیب ویژگی‌های افراد منتخب و تولید فرزندان جدید اجرا می‌گردد. در ادامه، عمل جهش (Mutation) برای افزایش تنوع ژنتیکی و جلوگیری از همگرایی زودهنگام اعمال می‌شود. این چرخه تکرار می‌شود تا نسل نهایی تشکیل شده و الگوریتم به مرحله خاتمه برسد.

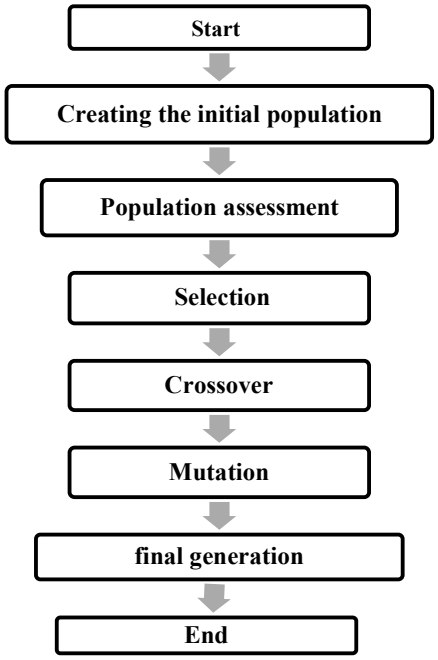


Fig. 5 – Flowchart of the Genetic Algorithm (Achite, et al., 2023)
شکل۵- فلوچارت الگوریتم ژنتیک (Achite, et al., 2023)

توسعه مدل‌ها

در این پژوهش، از دو مدل یادگیری ماشین شامل جنگل تصادفی و رگرسیون بردار پشتیبان برای پیش‌بینی تبخیر-تعرق واقعی استفاده شده است. به منظور دستیابی به بالاترین دقت ممکن، هر دو مدل با استفاده از الگوریتم ژنتیک بهینه‌سازی شده‌اند. الگوریتم ژنتیک با ایجاد یک جمعیت اولیه از ترکیب‌های مختلف هایپرپارامترها آغاز می‌شود و سپس با اعمال عملگرهای انتخاب، تقاطع و جهش، به تدریج ترکیب‌های بهینه را شناسایی می‌کند. در مورد مدل جنگل تصادفی، هایپرپارامترهای مهمی مانند تعداد درختان، حداکثر عمق درختان و حداقل نمونه‌های مورد نیاز برای تقسیم گره‌ها مورد بهینه‌سازی قرار گرفته‌اند. برای مدل SVR نیز پارامترهای کلیدی شامل ضریب تنظیم، نوع تابع هسته و پارامترهای مربوط به آن با دقت تنظیم شده‌اند. فرآیند بهینه‌سازی به گونه‌ای طراحی شده است که همزمان با جلوگیری از بیش‌برازش، بالاترین دقت پیش‌بینی را تضمین کند. الگوریتم ژنتیک در این پژوهش با جمعیت اولیه ۵۰ کروموزوم و طی ۱۰۰ نسل اجرا شده است. معیار ارزیابی هر کروموزوم، خطای مربعات میانگین روی داده‌های اعتبارسنجی متقابل بوده است. عملگر تقاطع با احتمال ۰/۸ و عملگر جهش با احتمال ۰/۱ اعمال شده‌اند تا تنوع لازم در جمعیت حفظ شود. پس از تکمیل فرآیند بهینه‌سازی، بهترین ترکیب‌های هایپرپارامتر برای هر مدل انتخاب شده‌اند. در این تحقیق، هایپرپارامترهای مدل‌ها با دقت انتخاب و تنظیم شدند. برای مدل جنگل تصادفی، تعداد درختان (n_estimators) به‌منظور تعادل بین دقت و پیچیدگی محاسباتی، حداکثر عمق درخت(max_depth) برای جلوگیری از بیش‌برازش، حداقل نمونه‌های مورد نیاز برای تقسیم گره(min_samples_split) به‌منظور کنترل رشد درخت‌ها، و تعداد ویژگی‌های در نظر گرفته شده در هر تقسیم (max_features) برای ایجاد تنوع در بین درخت‌ها تنظیم گردیدند. در مدل ماشین بردار پشتیبان، پارامتر تنظیم (C) برای ایجاد تعادل بین دقت مدل و سادگی آن، نوع تابع هسته (kernel) برای مدل‌سازی روابط غیرخطی، ضریب هسته (gamma) برای تعیین میزان تأثیر هر نمونه آموزشی، و مقدار اپسیلون (epsilon) برای تعیین حد تحمل خطا در پیش‌بینی‌ها بهینه‌سازی شدند. در الگوریتم ژنتیک نیز اندازه جمعیت اولیه (population_size)، تعداد نسل‌ها (generations)، احتمال وقوع عملگر تقاطع (crossover_probability) و احتمال وقوع جهش(mutation_probability) به‌ترتیب برای ایجاد تنوع در جواب‌ها، تعیین مدت زمان اجرای الگوریتم، ترکیب مناسب جواب‌ها و جلوگیری از همگرایی زودرس تنظیم گردیدند. مقادیر نهایی این پارامترها بر اساس آزمایش‌های متعدد و روش‌های بهینه‌سازی استاندارد تعیین شدند تا بهترین عملکرد ممکن حاصل شود. نتایج نشان می‌دهد که این روش هیبریدی منجر به بهبود قابل توجهی در عملکرد مدل‌ها شده است، به‌طوری‌که مدل‌های بهینه‌شده توانسته‌اند با دقت بالاتری تبخیر-تعرق واقعی را پیش‌بینی کنند. این بهبود عملکرد نه تنها در معیارهای ارزیابی عددی، بلکه در توانایی مدل برای تعمیم به داده‌های جدید نیز مشهود بوده است. مقادیر نهایی انتخاب‌شده برای هر یک از پارامترهای GA و SVR در جدول (۲) ارائه شده است.

DOI: 10.22055/jise.2025.49823.2165

در این تحقیق، برای برآورد تبخیر–تعرق واقعی، سیزده پارامتر مختلف که شامل FST_1، $FST_{1.6}$، RWS_M، RWS_{ave}، CST_1، $CST_{1.6}$، T، $HFIR_2$، $HFIR_1$، RH، HFR_1، HFR_2، U_2 در قالب هشت سناریو، به‌عنوان ورودی مدل درنظر گرفته شد (جدول ۳). ترتیب سناریوها براساس ضریب همبستگی پیرسون تنظیم شده است، به‌طوری‌که سناریوی اول بیش‌ترین همبستگی و سناریوی آخر کم‌ترین همبستگی را با پارامتر هدف (تبخیر–تعرق واقعی) دارد.

عملکرد الگوریتم‌های SVR، RF، RF-GA و SVR-GA برروی این سناریوها ارزیابی شد. با نرم‌افزار Rapid Miner Studio، ۶۷ درصد داده‌های آماری برای مرحله آموزش و ۳۳ درصد برای مرحله آزمون مدل‌ها درنظر گرفته شد. دقت مدل‌های مورد استفاده در این دو مرحله، براساس شاخص‌های آماری که شامل ضریب همبستگی(R)، که میزان همبستگی بین مقادیر پیش‌بینی‌شده و مشاهداتی را نشان می‌دهد، مقدار R در بازه‌ای بین ۱– تا ۱ تغییر می‌کند، به‌طوری‌که مقدار نزدیک به یک نشان‌دهنده همبستگی خطی قوی، مقدار نزدیک به ۱– نشان‌دهنده رابطه معکوس و مقدار صفر نشان‌دهنده نبود رابطه خطی میان داده‌هاست. استفاده از این شاخص برای سنجش توانایی مدل در پیروی از روند کلی داده‌های مشاهداتی (واقعی) ضروری است. ریشه میانگین مربعات خطا (RMSE) که بیانگر میزان خطای مدل در پیش‌بینی است، یکی از شاخص‌های متداول برای اندازه‌گیری میزان اختلاف بین مقادیر پیش‌بینی شده و مقادیر واقعی است. این شاخص به خطاهای بزرگ حساس بوده و مقدار کمتر آن نشان‌دهنده دقت بالاتر مدل است. RMSE در واحد متغیر مورد نظر در اینجا که همان تبخیر–تعرق واقعی است، برحسب میلی‌متر بر روز بیان می‌شود. ضریب نش–ساتکلیف (NS) توانایی مدل در شبیه‌سازی تغییرات مقادیر مشاهده‌شده را ارزیابی می‌کند، مقدار آن بین منفی بی‌نهایت تا یک متغییر است. مقدار۱=NS نشان‌دهنده انطباق کامل مدل با داده‌های مشاهداتی است، درحالی‌که مقدار کمتر از صفر بیانگر آن است که مدل از میانگین داده‌ها عملکرد ضعیف‌تری داشته‌است. این شاخص بیانگر میزان دقت مدل در پیش‌بینی نوسانات زمانی داده‌های واقعی است. (NRMSE)، مقدار RMSE را نسبت به میانگین داده‌های واقعی تنظیم می‌کند تا مقایسه بهتری بین سناریوها فراهم شود. مقدار آن بین صفر و یک می‌باشد. هرچه این مقدار به صفر نزدیک‌تر باشد، دقت مدل بالاتر است. شاخص توافق ویلموت (Willmott Index) به‌عنوان یک معیار ترکیبی، میزان نزدیکی پیش‌بینی‌ها به مقادیر مشاهداتی را با درنظر گرفتن پراکندگی آن‌ها اندازه‌گیری می‌نماید، مقدار آن بین صفر تا یک می‌باشد. هرچه این مقدار به یک نزدیک‌تر باشد، عملکرد مدل بهتر بوده است.

جدول ۲- مقادیر هایپر پارامترهای مورد استفاده برای مدل‌های مورد مطالعه

Table 2 – Input variables used in each scenario

Model	Hyperparameter	Optimal value	Description
RF	n_estimators	150	Number of decision trees in the forest
	max_depth	12	Maximum depth of each tree
	min_samples_split	5	Minimum samples required to split a node
	max_features	'sqrt'	Number of features to consider for best split
SVR	C (Regularization)	10	Trade-off between model simplicity and training accuracy
	kernel	'rbf'	Kernel type (Radial Basis Function)
	gamma	0.01	Kernel coefficient for 'rbf'
	epsilon	0.1	Margin of tolerance where no penalty is given
GA (Optimization)	population_size	50	Number of individuals in each generation
	generations	100	Number of iterations for evolution
	crossover_probability	0.8	Probability of combining parents' features
	mutation_probability	0.1	Probability of random hyperparameter changes

Table 3 – Input variables used in each scenario

جدول ۳- متغیرهای مورد استفاده در هر سناریو

Scenario	Input Variables	output
1	FST_1, $FST_{1.6}$, RWS_M	ET_a
2	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}	ET_a
3	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}, CST_1, $CST_{1.6}$	ET_a
4	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}, CST_1, $CST_{1.6}$, T	ET_a
5	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}, CST_1, $CST_{1.6}$, T, $HFIR_2$, $HFIR_1$	ET_a
6	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}, CST_1, $CST_{1.6}$, T, $HFIR_2$, $HFIR_1$, RH	ET_a
7	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}, CST_1, $CST_{1.6}$, T, $HFIR_2$, $HFIR_1$, RH, HFR_1, HFR_2	ET_a
8	FST_1, $FST_{1.6}$, RWS_M, RWS_{ave}, CST_1, $CST_{1.6}$, T, $HFIR_2$, $HFIR_1$, RH, HFR_1, HFR_2, U_2	ET_a

معیارهای ارزیابی

برای ارزیابی دقت روش‌های مورد مطالعه در پیش‌بینی تبخیر-تعرق واقعی، از شاخص‌های ریشه میانگین مربعات خطا (RMSE)، ضریب همبستگی (R)، خطای جذر میانگین مربعات نرمال‌شده (Normalized Root Mean Square Error)، ضریب ویلموت (Willmott's Index of Agreement) و شاخص نش-ساتکلیف (NS) مطابق روابط (۵) تا (۹) استفاده شد.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \quad (۵)$$

$$R = \frac{\left(\sum_{i=1}^{N}(O_i-\bar{O})-\frac{1}{N}(P_i-\bar{P})\right)}{\sqrt{\sum_{i=1}^{N}(O_i-\bar{O})^2\sum_{i=1}^{N}(P_i-\bar{P})^2}} \quad (۶)$$

$$NRMSE = \frac{RMSE}{\bar{O}} \quad (۷)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{N}(O_i-P_i)^2}{\sum_{i=1}^{N}(P_i-\overline{O_i})^2+\sum_{i=1}^{N}(O_i-\overline{O_i})^2}\right] \quad (۸)$$

$$NS = 1 - \left[\frac{\sum_{i=1}^{N}(O_i-P_i)^2}{\sum_{i=1}^{N}(O_i-\bar{O})^2}\right] \quad (۹)$$

در این روابط O_i مقادیر واقعی تبخیر-تعرق واقعی گیاه در منطقه مورد مطالعه، N تعداد داده‌ها، P_i تبخیر-تعرق پیش‌بینی شده با روش‌های مورد مطالعه و $\bar{O}$ میانگین مقادیر واقعی تبخیر-تعرق است که به‌عنوان یک شاخص کلی از داده‌های مشاهداتی استفاده می‌شود. هرچه ریشه میانگین مربعات خطا و خطای جذر میانگین مربعات نرمال‌شده کمتر باشد نمایانگر دقیق‌تر بودن روش می‌باشد. همچنین هرچه میزان ضریب همبستگی، ضریب ویلموت و شاخص نش‌ساتکلیف به عدد یک نزدیکتر باشد، دقت روش بالاتر خواهد بود.

نتایج و بحث

براساس بررسی و مقایسه نتایج حاصل از جدول (۴)، سناریو اول مدل SVR با داشتن ریشه مربعات میانگین خطا به میزان ۲/۵۳۷ (mm/day)، بیش‌ترین خطا را نسبت به سایر سناریوهای این مدل داشته است. در سناریو دوم و سوم به‌ترتیب با کاهش و افزایش ۱۹/۴۳ و ۱/۶۱ درصدی، مقدار RMSE کاهش محسوسی داشته و با افزایش R، عملکرد مدل بهبود نسبی داشته است. در سناریو چهارم مقدار RMSE به ۱/۹۳۲ (mm/day) کاهش یافته است که نشان می‌دهد افزودن پارامتر دما در این سناریو با تأثیر مثبت همراه بوده و منجربه بهبود خطا به‌میزان ۶/۹۸ درصد شده است. درسناریوهای پنجم و ششم تغییرات ناچیز بوده و شاخص‌های ارزیابی مانند ضریب همبستگی و شاخص ویلموت ثابت باقی مانده‌اند. در سناریو هفتم با کاهش مقدار RMSE به میزان ۱/۱۴ درصد در مقایسه با سناریو قبل، دقت مدل افزایش پیدا کرده است. در سناریو هشتم علی‌رغم افزودن پارامتر سرعت باد، مقدار RMSE افزایش یافته است که نشان می‌دهد این متغیر تأثیر قابل‌توجهی بر کاهش میزان خطا نداشته است. با در نظر گرفتن معیارهای ارزیابی، سناریو هفتم (SVR-7) به‌دلیل داشتن کم‌ترین خطا و بیش‌ترین همبستگی، به‌عنوان بهترین سناریو در میان تمام سناریوهای مدل SVR شناخته شد. افزودن متغیرهای HFR_1 و HFR_2 در این سناریو نقش مؤثری در افزایش دقت مدل داشته است. ارزیابی نتایج SVR-GA نشان می‌دهد که سناریو اول با داشتن RMSE به مقدار ۲/۱۹۹ (mm/day) ضعیف‌ترین عملکرد را نسبت به سایر سناریوهای این مدل داشته است. سناریوهای دوم و سوم با بهبود شاخص RMSE به‌ترتیب به میزان ۲۴/۰۵ و ۳/۵۹ درصدی، عملکرد بهتری نسبت به سناریو قبل خود داشته‌اند. در سناریو چهارم مقدار RMSE با کاهش ۴/۷۸ درصدی نسبت به سناریو سوم، دقت مدل را افزایش داد. در سناریو پنجم نیز با افزوده شدن پارامتر $HFIR_1$ و $HFIR_2$ میزان خطا نسبت به سناریو قبلی حدود ۱۵/۱۳ درصد بهبود یافته است. در سناریوهای ششم و هفتم تغییر محسوسی در مقدار شاخص‌های ارزیابی صورت نگرفته است. در سناریو هشتم، با افزودن متغیر U_2، مقدار RMSE (mm/day) به ۱/۰۱۷ کاهش یافت و نسب به سناریو قبل بهبود ۱۲/۹۲ درصدی را تجربه کرد. افزودن این متغیر تأثیر چشمگیری در کاهش خطا و بهبود دقت پیش‌بینی داشته است. به‌طور کلی، افزودن متغیرهای مختلف باعث بهبود دقت مدل شد، با این حال، افزودن برخی از متغیرها مانند CST_1 و $CST_{1.6}$ تأثیر اندکی بر معیارهای ارزیابی داشتند، درحالی‌که متغیرهایی مانند U_2، $HFIR_1$ و $HFIR_2$ باعث بهبود چشمگیر در دقت پیش‌بینی‌ها شدند. تحلیل نتایج نشان می‌دهد که مدل SVR-GA در طی سناریوهای مختلف پیشرفت چشمگیری را تجربه کرده است. بیش‌ترین میزان بهبود در مقادیر R و کاهش RMSE در سناریوهای ابتدایی مشاهده شده و پس از سناریو پنجم، مدل به پایداری نسبی رسیده است. از سناریو ششم به بعد، تغییرات در عملکرد مدل اندک بوده و نشان‌دهنده بهبود نسبی مدل است. درنهایت در سناریو هشتم، مقدار R در حداکثر مقدار خود و مقدار RMSE در کم‌ترین مقدار ممکن قرار دارد و این سناریو بهترین عملکرد را ارائه داده است. مقایسه نتایج دو مدل نشان می‌دهد که استفاده از الگوریتم ژنتیک باعث بهبود عملکرد مدل SVR شده است و به‌میزان قابل‌توجهی خطای مدل را کاهش داده و دقت پیش‌بینی را افزایش داده است. مقدار RMSE در بهترین سناریوی مدل SVR-GA نسبت به بهترین سناریوی مدل SVR کاهشی به میزان ۴۴/۱۲ درصد داشته است. این کاهش نشان‌دهنده قدرت بالای GA در بهینه‌سازی پارامترهای مدل و افزایش دقت پیش‌بینی می‌باشد. نتایج این تحقیق با نتایج Sharafi et al. (2024) که از مدل ترکیبی SVR-FOA برای پیش‌بینی تبخیر-تعرق در اهواز استفاده کردند، هم‌خوانی دارد. آن‌ها نشان دادند که

DOI: 10.22055/jise.2025.49823.2165

ترکیب مدل SVR با یک الگوریتم بهینه‌سازی موجب بهبود دقت پیش‌بینی در مقایسه با مدل مستقل می‌شود. مقایسه سناریوهای اول و دوم مدل RF نشان می‌دهد که جذر ریشه مربعات میانگین خطا حدود ۲۳/۳۰ درصد کاهش یافته و سایر معیارهای ارزیابی نیز به‌صورت قابل‌توجهی بهبود یافته‌اند. دو سناریو سوم و چهارم با کاهش مقدار RMSE معادل ۵/۹۸ درصد و افزایش R همراه بوده‌اند و عملکرد بهتری نسبت به سناریوهای اول و دوم داشته‌اند. افزودن پارامترهای $HFIR_2$، $HFIR_1$ در سناریو پنجم میزان خطای جذر ریشه مربعات میانگین را به میزان ۶/۴۲ درصد کاهش داده است. مقایسه سناریوهای ششم و هفتم نشان می‌دهد که مدل تغییر قابل‌توجهی نداشته و معیارهای ارزیابی به‌صورت تقریباً ثابتی باقی‌مانده است. سناریو اخر(سناریو هشتم) این مدل با کاهش جزئی مقدار RMSE (۰/۱۲ درصد) و افزایش ضریب همبستگی بین مقادیر واقعی و پیش‌بینی شده نسبت به سایر سناریوهای مدل RF عملکرد مناسب‌تری داشته است. مقایسه نتایج دو سناریو اول مدل RF-GA نشان می‌دهد که سناریو دوم سبب بهبود شاخص R و RMSE به‌ترتیب به میزان ۱۰/۵ و ۱۷/۱۲ درصد نسبت به سناریو اول شده است. در سناریوهای سوم و چهارم معیار جذر میانگین مربعات خطا حدود ۴/۷۳ درصد کاهش یافته و سایر معیارهای ارزیابی به‌صورت جزئی بهبود یافته‌اند. برای سناریوهای پنجم و ششم مقدار RMSE حدود ۳/۲۴ درصد کاهش یافته و سایر معیارهای ارزیابی تغییر محسوسی نداشته‌اند. بررسی سناریوهای هفتم و هشتم نشان می‌دهد که افزودن پارامترهای HFR_1, HFR_2 در سناریو هفتم موجب کاهش مقدار RMSE و بهبود ۲/۲۵ درصدی شده ولی در سناریو هشتم افزودن پارامتر سرعت باد موجب افزایش آن و کاهش دقت مدل شده است. مقایسه سناریوهای مدل RF-GA نشان داد که سناریو هفتم با داشتن ضریب همبستگی ۰/۹۰، جذر میانگین مربعات خطا (mm/day) ۱/۵۱۶، ضریب نش–ساتکلیف ۰/۷۹۶، جذر میانگین مربعات خطا نرمال‌شده ۰/۱۲۲ و شاخص ویلموت ۰/۹۸ دقت بالاتری را نسبت به سایر سناریوها داشته است. ارزیابی دو مدل RF و RF-GA نشان می‌دهد که استفاده از الگوریتم ژنتیک برای بهینه‌سازی مدل جنگل تصادفی موجب بهبود دقت مدل شده است. این مدل در اکثر سناریوها RMSE کمتری داشته و شاخص‌های ارزیابی مانند ضریب همبستگی و ضریب ویلموت بهبود یافته‌اند. با این حال، پس از سناریو پنج، تغییرات جزئی‌تر شده و مدل به نقطه بهینه نزدیک شده است. نتایج تحقیق حاضر با نتایج Monavar Sabegh et al. (2023)، که به برآورد تبخیر–تعرق مرجع روزانه در سه ایستگاه آذربایجان شرقی با استفاده از روش جنگل تصادفی بهینه شده با الگوریتم ژنتیک پرداختند و نشان دادند که مدل RF-GA با شاخص ویلموت ۰/۹۸۲ و ۰/۹۹۲ به‌ترتیب در سناریو دوم و چهارم برای ایستگاه تبریز عملکرد بهتری از خود نشان داده است، مطابقت دارد.

جدول ٤- نتایج حاصل از ارزیابی مدل‌ها در دوره آزمون

Table 4– Performance of the models during the testing period

Model	R	RMSE(mm/day)	NRMSE	NS	WI
SVR-1	0.69	2.537	0.204	0.43	0.74
SVR-2	0.80	2.044	0.164	0.63	0.89
SVR-3	0.80	2.077	0.167	0.62	0.89
SVR-4	0.84	1.932	0.155	0.67	0.92
SVR-5	0.84	1.837	0.148	0.701	0.93
SVR-6	0.84	1.841	0.148	0.70	0.93
SVR-7	0.86	1.820	0.147	0.706	0.93
SVR-8	0.84	1.835	0.148	0.702	0.93
SVR-GA-1	0.76	2.199	0.176	0.572	0.92
SVR-GA-2	0.87	1.670	0.135	0.753	0.97
SVR-GA-3	0.88	1.610	0.130	0.77	0.98
SVR-GA-4	0.89	1.533	0.124	0.792	0.98
SVR-GA-5	0.92	1.301	0.105	0.85	0.99
SVR-G A-6	0.94	1.151	0.093	0.883	0.99
SVR-GA-7	0.94	1.168	0.094	0.879	0.99
SVR-GA-8	0.95	1.017	0.082	0.908	0.99
RF-1	0.72	2.360	0.190	0.504	0.89
RF-2	0.84	1.810	0.146	0.71	0.96
RF-3	0.84	1.821	0.147	0.706	0.96
RF-4	0.86	1.712	0.138	0.74	0.97

Model	R	RMSE(mm/day)	NRMSE	NS	WI
RF-5	0.88	1.602	0.129	0.773	0.98
RF-6	0.88	1.627	0.131	0.765	0.97
RF-7	0.88	1.620	0.130	0.768	0.98
RF-8	0.89	1.618	0.130	0.768	0.98
RF-GA-1	0.76	2.178	0.176	0.580	0.92
RF-GA-2	0.84	1.805	0.146	0.711	0.96
RF-GA-3	0.86	1.755	0.142	0.727	0.96
RF-GA-4	0.87	1.672	0.135	0.752	0.97
RF-GA-5	0.88	1.603	0.129	0.773	0.98
RF-GA-6	0.89	1.551	0.125	0.787	0.98
RF-GA-7	0.90	1.516	0.122	0.796	0.98
RF-GA-8	0.90	1.539	0.124	0.790	0.98

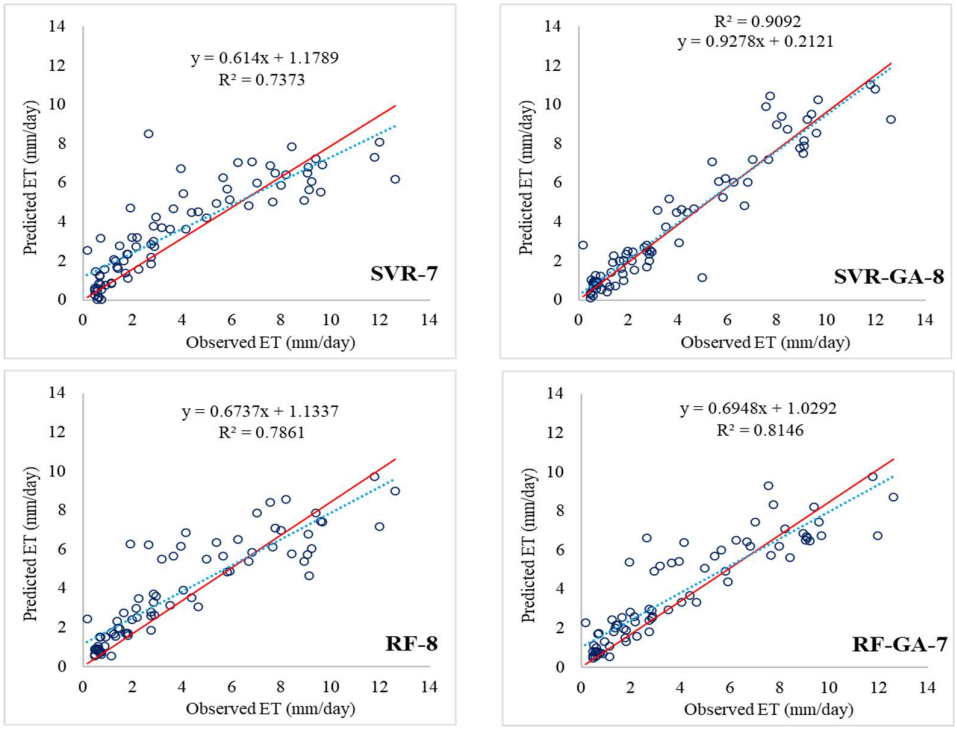


Fig. 6- Scatter plots of observed and predicted evapotranspiration for the best scenario

شکل ٦ – نمودارهای پراکنش مقادیر مشاهداتی و پیش‌بینی شده تبخیر-تعرق برای بهترین سناریو

نمودار پراکنش مقادیر تبخیر-تعرق واقعی روزانه در شکل (۶) که برای بهترین سناریو هرمدل (SVR-7, SVR-GA-8, RF-8, RF-GA-7) می‌باشد، رسم شده است. مقایسه بین دو مدل SVR-7 وSVR-GA-8 نشان می‌دهد که در مدل SVR-GA-8 داده‌های مشاهداتی و پیش‌بینی شده، همبستگی بیشتری نسبت به یکدیگر داشته و ضریب تعیین مدل بهینه‌شده با الگوریتم ژنتیک (SVR-GA-8) بیش‌تر از مدل مستقل (SVR-7) بوده است. هم‌چنین مقایسه دو مدل RF-8 و RF-GA-7 نشان می‌دهد، همبستگی بین داده‌های مشاهداتی و پیش‌بینی شده برای مدل RF-GA-7 بیش‌تر از مدل RF-8 بوده و مدل عملکرد بهتری را از خود نشان داده است.

نمودار تغییرات زمانی مقادیر تبخیر-تعرق واقعی روزانه برای بهترین سناریو هر مدل در شکل (۷) قابل ملاحظه می‌باشد. تحلیل نمودار تغییرات زمانی نشان می‌دهد که مدل ترکیبی SVR-GA-8 تقارب بیشتری میان داده‌های مشاهداتی و پیش‌بینی نسبت به مدل منفرد SVR-7 دارد و این تقارب نمایانگر دقت بالای مدل ترکیبی می‌باشد.

DOI: 10.22055/jise.2025.49823.2165

همچنین مقایسه نمودار سری زمانی دو مدل RF-8 و RF-GA-7 نشان می‌دهد که مدل جنگل تصادفی بهینه شده با الگوریتم ژنتیک توانسته نتایج بهتری در تطابق بین داده‌های مشاهداتی و داده‌های واقعی نسبت به مدل مستقل خود ارائه دهد و در پیش‌بینی عملکرد بهتری داشته باشد.

شکل (۸) عملکرد مدل‌ها را با استفاده از نمودارهای جعبه‌ای نشان می‌دهد. مطابق شکل، مدل RF-8 و RF-GA-7 دقت بالاتری را در پیش‌بینی تبخیر–تعرق واقعی روزانه داشته‌اند. مقایسه خط میانه این دو سناریو با خط میانه تبخیر–تعرق واقعی مشاهداتی نشان می‌دهد که این خطوط در یک راستا قرار دارند. علاوه براین، محدوده بین چارکی دو مدل دارای طول یکسان بوده که نشان دهنده پراکنش تقریبا مشابه بین داده‌ها می‌باشد و این موضوع بیانگر آن است که حداقل و حداکثر مقادیری که محدوده توزیع داده‌ها را تعیین می‌کنند، برای مدل مشاهده شده و پیش‌بینی شده تقریبا مشابه می‌باشند. بنابراین می‌توان تفسیر کرد که این دو سناریو عملکرد بهتری را نسبت به سایر سناریوها داشته‌اند. در نمودار مربوط به مدل رگرسیون بردار پشتیبان نیز دو سناریو SVR-7 و SVR-GA-8 تخمین خوبی از تبخیر–تعرق واقعی ارائه داده‌اند. طول بازوی چارکی این دو سناریو به‌طور تقریبی مشابه با مقادیر تبخیر–تعرق مشاهداتی است. به‌علاوه خط میانه این دو سناریو هم‌سو باخط میانه مقادیر مشاهداتی می‌باشد. لذا می‌توان نتیجه گرفت که این دو سناریو دقت بالاتری را نسبت به سایر سناریوها داشته‌اند.

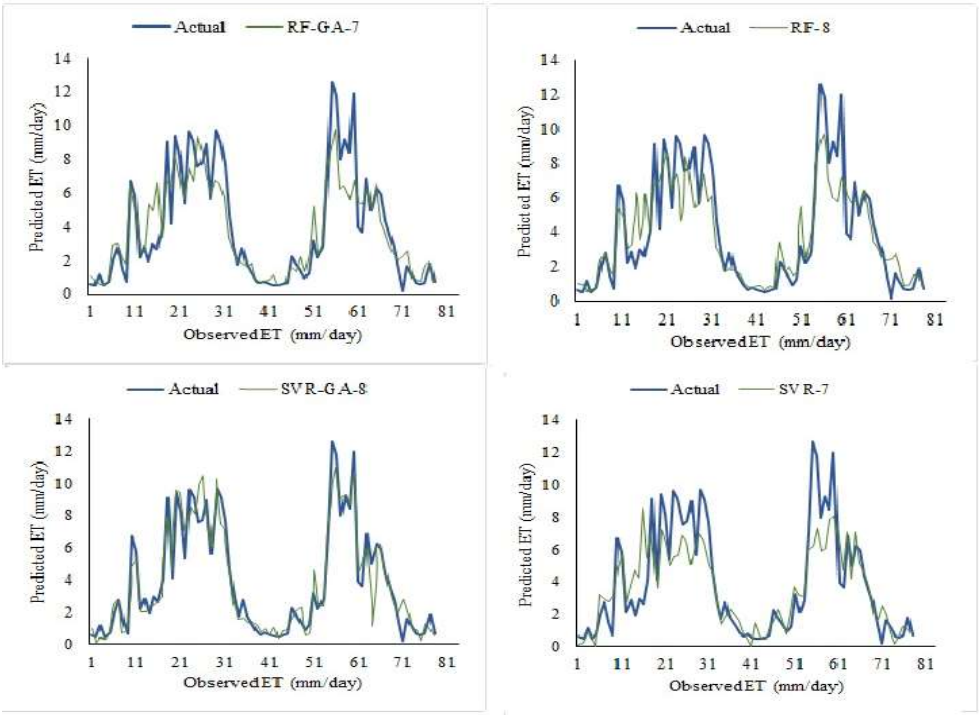


Fig. 7- Actual evapotranpiration time series diagram for the best scenario

شکل۷- نمودار سری زمانی تبخیر-تعرق واقعی برای بهترین سناریو

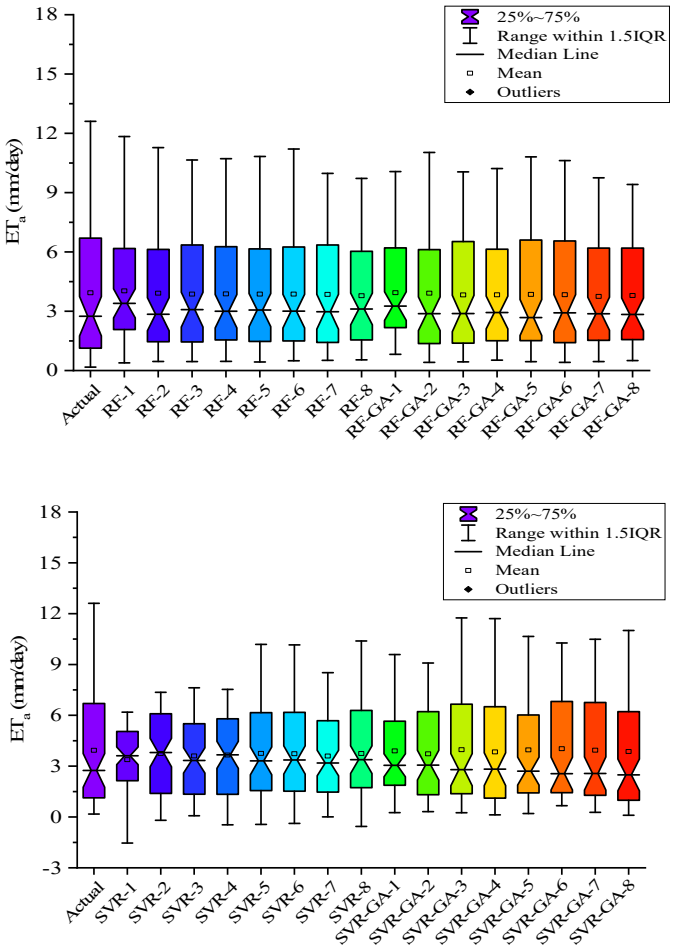


Fig. 8- Box plot predicted values and observed values

شکل ۸- نمودار جعبه‌ای مقادیر پیش‌بینی شده و مقادیر مشاهداتی

شکل (۹) دیاگرام‌های تیلور مدل‌های منفرد و هیبریدی را نشان می‌دهد. مطابق شکل (۹)، مقایسه دیاگرام‌های تیلور مربوط به مدل‌های جنگل تصادفی و مدل بهینه‌شده آن‌ها با الگوریتم ژنتیک (RF-GA) نشان می‌دهد که بهینه‌سازی پارامترهای مدل با استفاده از الگوریتم ژنتیک منجربه بهبود قابل‌توجه عملکرد مدل شده است. به‌طوری‌که، مدل‌های RF-GA به نقطه مرجع (Reference) نزدیک‌تر بوده و دارای ضریب همبستگی بالاتر، انحراف معیار کمتر نسبت‌به مدل‌های اولیه RF بوده‌اند. این موارد نشان‌دهنده دقت بالا و عملکرد مطلوب مدل‌های بهینه‌شده است.

از سوی دیگر، بررسی و مقایسه دیاگرام‌های تیلور مدل‌های رگرسیون بردار پشتیبان و مدل بهینه‌شده آن (SVR-GA) نیز نشان‌دهنده بهبود چشمگیر عملکرد مدل‌ها پس از بهینه‌سازی است، به‌گونه‌ای که مدل‌های SVR-GA به نقطه مرجع نزدیک‌تر بوده، ضریب همبستگی بالاتر، انحراف معیار کمتری در مقایسه با مدل منفرد خود دارد. همگرایی بیش‌تر به خط مرجع و ضریب همبستگی بالاتر نشان می‌دهد که الگوریتم ژنتیک توانسته پارامترهای مدل را به‌خوبی بهینه کند و مدل‌های قابل اعتمادتری بسازد.

DOI: 10.22055/jise.2025.49823.2165

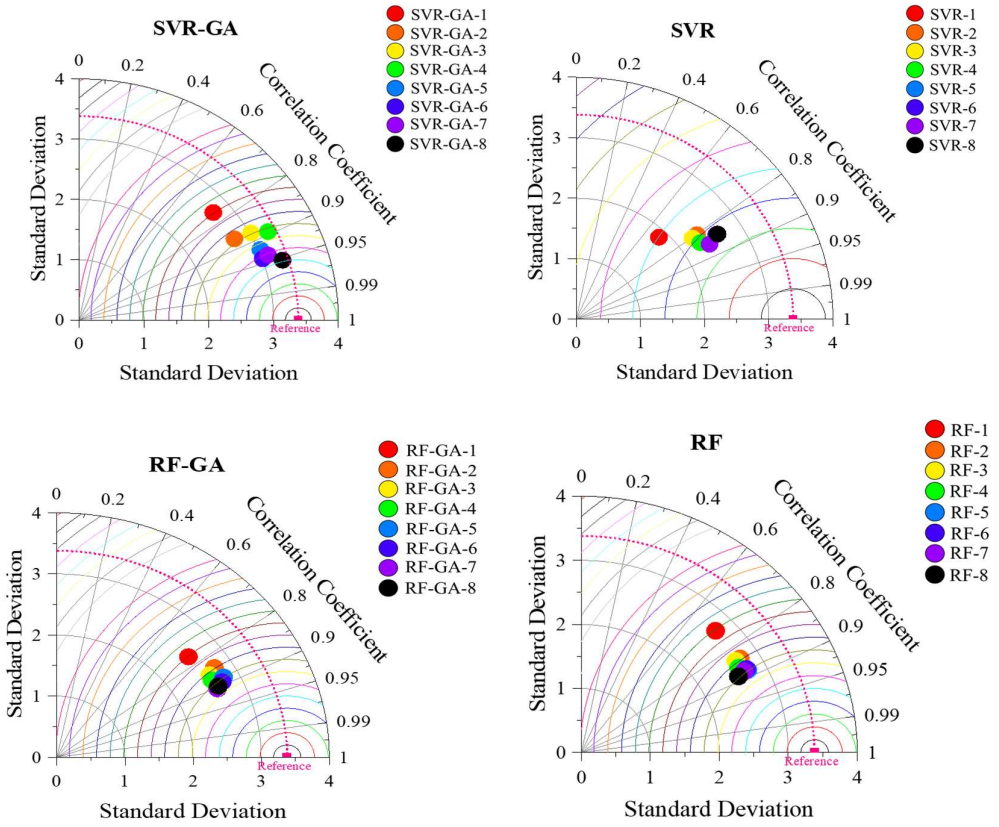


Fig. 9– Taylor diagram of scenarios for observed and predicted values
شکل ۹- دیاگرام تیلور سناریوها برای مقادیر پیش‌بینی شده و مقادیر مشاهداتی

نتیجه‌گیری

کاهش منابع آب و ضرورت مدیریت بهینه در بخش کشاورزی، اهمیت اندازه‌گیری تبخیر–تعرق واقعی را افزایش داده است. یکی از ابزارهای اندازه‌گیری تبخیر–تعرق واقعی لایسیمتر است، اما محدودیت دسترسی به آن موجب ایجاد چالش و کمبود داده‌های ET_a شده است. این پژوهش به توسعه یک مدل دقیق برای پیش‌بینی تبخیر–تعرق واقعی می‌پردازد که می‌تواند جایگزین ابزاری مانند لایسیمتر شود. در این راستا مدل های SVR، SVR-GA، RF و RF-GA با استفاده از داده‌های در دسترس دو لایسیمتر در بوشلند تگزاس نظیر دمای عمق‌های مختلف خاک، شار حرارتی خاک، ذخیره نسبی آب، رطوبت نسبی و سرعت باد می‌باشد که در هشت سناریو و برای سال ۲۰۱۲-۲۰۱۱ ارزیابی شدند. نتایج ارزیابی‌ها نشان داد که مدل‌های بهینه‌شده با الگوریتم ژنتیک بالاترین دقت را در پیش‌بینی تبخیر–تعرق واقعی داشته و کم‌ترین مقدار میانگین جذر مربعات خطا (۱/۵۱۶و ۱/۰۱۷) را ارائه داده است. . هم‌چنین در این پژوهش، تأثیر پارامترهای ورودی مختلف بر دقت پیش‌بینی تبخیر–تعرق واقعی با بهره‌گیری از مدل‌های رگرسیون بردار پشتیبان (SVR)، جنگل تصادفی (RF) و مدل‌های بهینه‌شده آن‌ها با الگوریتم ژنتیک (SVR-GA، RF-GA) به طور جامع مورد ارزیابی قرار گرفت. نتایج نشان داد که پارامترهای پایه شامل میانگین دمای روزانه خاک در عمق‌های مختلف داخل و خارج لایسیمتر، ذخیره نسبی آب در لایسیمتر در نیمه‌شب، میانگین ذخیره نسبی و رطوبت نسبی، اگرچه نقش مهمی در مدل‌سازی دارند، اما به تنهایی برای دستیابی به دقت قابل قبول کافی نیستند. افزودن پارامتر دمای هوا تأثیر چشم‌گیری در کاهش خطا و بهبود ضریب همبستگی داشت. هم‌چنین پارامترهای شار حرارتی خاک، چه در ردیف‌ها و چه در فاصله بین ردیف‌ها ($HFIR_{1,2}$ و $HFR_{1,2}$)، نقش مؤثری در بهبود دقت پیش‌بینی ایفا کردند. سرعت باد به عنوان پارامتری با تأثیر متغیر، در مدل‌های بهینه‌شده توسط الگوریتم ژنتیک موجب افزایش دقت شد، اما در مدل‌های مستقل اثر قابل توجهی نداشت و در برخی موارد حتی موجب افزایش خطا گردید. مقایسه بین نتایج مدل‌ها نشان می‌دهد که مدل‌های ترکیبی عملکرد بهتری نسبت به مدل‌های مستقل دارند.هم‌چنین با تنظیم بهینه پارامترها می‌توان از این مدل‌ها برای پیش‌بینی دقیق تبخیر–تعرق واقعی با استفاده از داده‌های در دسترس بهره برد و وابستگی به روش‌های پرهزینه را کاهش داد. در پژوهش‌های آتی، می‌توان از روش‌های پیشرفته‌تر اعتبارسنجی مانند اعتبارسنجی متقاطع برای ارزیابی پایدارتر مدل‌ها استفاده کرد. این روش‌ها به‌ویژه در مواردی

که حجم داده‌ها محدود است (مانند داده‌های لایسیمتری)، می‌توانند اطمینان بیشتری از عدم وابستگی نتایج به تقسیم‌بندی خاص داده‌ها ایجاد کنند. همچنین، استفاده از تقسیم‌بندی سه‌گانه (Train-Validation-Test) می‌تواند به بهینه‌سازی دقیق‌تر هایپرپارامترها و ارزیابی بی‌طرفانه‌تر مدل‌ها بینجامد. افزودن این روش‌ها در کنار گزارش معیارهای ارزیابی روی داده‌های مستقل آزمون، می‌تواند اعتبار علمی پژوهش را افزایش دهد.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ گونه تضاد منافعی در خصوص نگارش و انتشار مطالب و نتایج این پژوهش ندارند.

منابع مالی

نویسندگان هیچ گونه حمایت مالی برای تحقیق، تالیف و انتشار این مقاله دریافت نکردند.

دسترسی به داده‌ها

مجموعه داده‌های تولید شده و یا تحلیل شده در طول مطالعه حاضر، در صورت درخواست معقول، از نویسنده مسئول در دسترس هستند.

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده اند و این موضوع مورد تأیید همه آنهاست.

اعلامیه هوش مصنوعی مولد و فناوری‌های مبتنی بر هوش مصنوعی در فرایند نگارش

نویسندگان در مقاله خود از هوش مصنوعی مولد و فناوریهای کمکی هوش مصنوعی استفاده نکرده اند.

مشارکت نویسندگان

نویسنده اول: سعیده محمدزاده، نگارش مقاله.

نویسنده دوم: وحید رضاوردی نژاد، بررسی محتوای مقاله و اصلاحات.

نویسنده سوم: جواد بهمنش، بررسی محتوای مقاله و اصلاحات.

نویسنده چهارم: میلاد شرفی، بررسی نتایج و تجزیه‌وتحلیل داده‌ها، بررسی محتوای مقاله و اصلاحات علمی و نگارشی.

تشکر و قدردانی

بدین‌وسیله از آزمایشگاه حفاظت و تولید USDA-ARS، واحد تحقیقات مدیریت آب و خاک در بوشلند، تگزاس به‌دلیل فراهم سازی داده‌های یک‌ساله لایسیمتر وزنی گیاه آفتابگردان برای این پژوهش، قدردانی می‌شود.

References

1- Achite, M., Samadianfard, S., Elshaboury, N. and Sharafi, M., 2023. Modeling and optimization of coagulant dosage in water treatment plants using hybridized random forest model with genetic algorithm optimization. *Environment, Development and Sustainability*, *25*(10), pp.11189-11207.doi: 10.1007/s10668-022-02523-z.

2- Achite, M., Samadianfard, S., Elshaboury, N., Toubal, K.A., Abdelkader, E.M. and Sharafi, M., 2024. A combined support vector regression with a firefly algorithm for prediction of energy consumption in wastewater treatment plants. *Water Science & Technology*, *90*(10), pp.2747-2763. doi: 10.2166/wst.2024.375.

3- Bayat, V.M., Zare, A.H., Maroufi, S., Sabziparvar, A.A. and Soltani, F., 2009. Simulation of reference evapotranspiration using artificial neural method and empirical methods and comparison with experimental lysimeter data in cold semi-arid climate of hamedan.

4- Breiman, L. 2001. Random forests. *Machine learning*. *45*: 5-32. doi: 10.1023/A:1010933404324.

5- Chiew, FHS., Kamaladasa, NN., Malano, H.M. and McMahon, T.A. 1995. Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agricultural Water Management. 28*:9–21. doi: 10.1016/0378-3774(95)01172-F.

6- Choudhury, B. 1999. Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *Journal of Hydrology. 216*(1-2): 99-110. doi: 10.1016/S0022-1694(98)00293-5

7- Cimen, M. 2008. Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal. 53*(3): 656-666. doi: 10.1623/hysj.53.3.656

8- Da Silva Junior, J.C., Medeiros, V., Garrozi, C., Montenegro, A. and Gonçalves, G. E. (2019). Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. *Computers and electronics in agriculture*. 166(1): 1-10. Doi: 10.1016/j.compag.2019.105017

9- Dong, Y. and Hansen, H. 2023. Development and design of an affordable field scale weighing lysimeter using a microcontroller system. *Smart Agricultural Technology. 4*(1): 1-8. doi: 10.1016/j.atech.2022.100147.

10- Evett, S,R,, Copeland, K.S., Ruthardt, B.B., Marek, G.W., Colaizzi, P.D., Brauer, D.K. and Howell, T.A. 2022. The Bushland Maize for Grain Datasets: A Machine-readable Resource. 2022 ASABE Annual International Meeting. *American Society of Agricultural and Biological Engineers. 1*(1): 28-41. doi: 10.13031/aim.202201078.

11- Evett, S. R., Howell, T. A., Todd, R. W., Schneider, A. D. and Tolk, J. A. 2000. Alfalfa reference ET measurement and prediction. *In Proceedings of the 4th Decennial National Irrigation Symposium. 17*(1): 14-16.

12- Evett, S.R., Marek, G.W., Copeland, K.S. and Colaizzi, P.D. 2018. Quality management for research weather data: USDA-ARS, Bushland, TX. Agrosystems, *Geosciences & Environment. 1*(1): 1-18. doi: 10.2134/age2018.09.0036

13- Feng, K. and Tian, J. 2021. Forecasting reference evapotranspiration using data mining and limited climatic data. *European Journal of Remote Sensing. 54*(2): 363-371. doi: 10.1080/22797254.2020.1801355.

14- Feng, Y., Cui, N., Gong, D., Zhang, Q. and Zhao, L. 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management. 193(*1): 163-173. doi: 10.1016/j.agwat.2017.08.003.

15- Gunn, S.R. 1998. Support vector machines for classification and regression, *ISIS Technical Report. 14* (1): 5–16.

16- Hashemian AH, Manochehri S, Afshari D, Manochehri Z, Salari N and Shahsavari S, 2019. Prognosis of multiple sclerosis disease using data mining approaches random forest and support vector machine based on genetic algorithm. *Tehran University Medical Journal 77(1): 33-40.* doi: magiran.com/p1974089

17- Holland, JH. 1992. Adaptation in natural and artificial systems: *an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press. doi: 10.7551/mitpress/1090.001.0001.

18- Jha, S.K. and Hayashi, K. 2014. A novel odor filtering and sensing system combined with regression analysis for chemical vapor quantification. *Sensors and Actuators B: Chemical. 200*(1): 269-287. doi: 10.1016/j.snb.2014.04.022

19- Jing, W., Yang, Y., Yue, X. and Zhao, X. 2015. Mapping Urban Areas with Integration of DMSP/OLS Nighttime Light and MODIS Data Using Machine Learning Techniques. *Remote Sensing. 7*(9): 12419-12439. doi: 10.3390/rs70912419.

20- Khosravi, K., Daggupati, P., Alami, M. T., Awadh, S. M., Ghareb, M. I., Panahi, M. and Yaseen, Z.M. 2019. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Computers and Electronics in Agriculture. 167*(1): 105-121. doi: 10.1016/j.compag.2019.105041.

21- Kumar, P. Singh, A.K. 2022. A comparison between MLR, MARS, SVR, and RF techniques: Hydrological time-series modeling. *Journal of Human, Earth, and Future. 3*(1): 90-98. doi: 10.28991/HEF-2022-03-01-07.

22- Lamm, F.R., Aiken, R.M. and Aboukheira, A.A. 2011. Irrigation research with sunflowers in Kansas. In: Proc. *23rd Annual Central Plains Irrig. Conf., 23rd, Burlington, 17*: 169–180.

22- Marek, G.W., Evett, S.R., Gowda, P.H., Howell, T.A., Copeland, K.S. and Baumhardt, R.L., 2014. Post-processing techniques for reducing errors in weighing lysimeter evapotranspiration (ET) datasets. *Transactions of the ASABE*, *57*(2), pp.499-515.doi: 10.13031/trans.57.10433.

23- Mokari, E., DuBois, D., Samani, Z., Mohebzadeh, H. and Djaman, K. 2022. Estimation of daily reference evapotranspiration with limited climatic data using machine learning approaches across different climate zones in New Mexico. *Theoretical and Applied Climatology. 147*(1): 575-587. doi: 10.1007/s00704-021-03855-y.

24- Monavar Sabegh, S. , Zarehaghi, D. , Samadianfard, S. , Neishabouri, M. R. and Mikaeili, F. 2023. Estimation of Daily Reference Evapotranspiration Using Random Forest Optimized by Genetic Algorithm. *Water and Soil Science*, *33*(4), 33-53 (In Persian). doi: 10.22034/ws.2021.48756.2449.

25- Saravi, B., Hassel, F., Ülkümen, S., Zink, A., Shavlokhova, V., Couillard-Despres, S. and Lang, G.M. 2022. Artificial intelligence-driven prediction modeling and decision making in spine surgery using hybrid machine learning models. *Journal of Personalized Medicine. 12*(4): 509-524. doi: 10.3390/jpm12040509.

26- Sharafi, M. , Besharat, S. and Zeinalzadeh, K. 2024. A comparative study using a data-driven method versus a hybrid approach to estimate daily reference evapotranspiration in Ahvaz. *Irrigation Sciences and Engineering*, *47*(2), 53-68 (In Persian). doi: 10.22055/jise.2023.43524.2065.

27- Sharafi, M. and Samadianfard, S. 2024. Prediction of daily evaporation values using support vector regression coupled with firefly algorithm in the coastal areas of Iran. *Water and Soil Science*, *34*(1), 39-53 (In Persian). doi: 10.22034/ws.2023.54178.2500.

28- Shiri, N. , Kazemi, M. H. and Shiri, J. 2021. Application of empirical and random forest models in estimating reference evapotranspiration of humid regions. *Water and Soil Science*, *31*(2), 1-11 (In Persian). doi: 10.22034/ws.2021.12356.

29- Smola, A. J. and Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and computing. 14*(1): 199-222. doi: 10.1023/B:STCO.0000035301.49549.88.

30- Talukdar, S., Mallick, J., Sarkar, S. K., Roy, S. K., Islam, A. R. M. T., Praveen, B. and Sobnam, M. 2022. Novel hybrid models to enhance the efficiency of groundwater potentiality model. *Applied Water Science. 12*(4): 62-75. doi: 10.1007/s13201-022-01571-0.

31- Tolk, J.A. and Howell, T.A. 2012. Sunflower water productivity in four Great Plains soils. *Field Crops Research. 127*(2): 120-128. doi: 10.1016/j.fcr.2011.11.012.

32- Tyralis, H., Papacharalampous, G. and Langousis, A. 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water. 910*(1):15-36. doi: 10.3390/w11050910.

33- Wang, S., Lian, J., Peng, Y. and Chen. H. 2019. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. *Agricultural Water Management. 221*(1): 220-230. doi: 10.1016/j.agwat.2019.03.027.

34- Wu, L., Huang, G., Fan, J., Ma, X., Zhou, H. and Zeng, W. 2019. Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction. *Computers and Electronics in Agriculture. 168*(1): 105-115. doi: 10.1016/j.compag.2019.105115.

35- Xiong, Y.J., Chen, X., Tang, L. and Wang, H. 2022. Comparison of surface renewal and Bowen ratio derived evapotranspiration measurements in an arid vineyard. *Journal of Hydrology. 613*(1): 22-37. doi: 10.1016/j.jhydrol.2022.128474.

36- Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O. and Lee, K.K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of hydrology. 396*(1-2): 128-138. doi: 10.1016/j.jhydrol.2010.11.002

37- Zouzou, Y. and Citakoglu, H. 2023. General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration. *Acta Geophysica. 71*(2): 927-947. doi: 10.1007/s11600-022-00939-9.