Water and Soil Science, Vol.33, NO.4, Winter 2023, pp. 201-216.

DOI: 10.22034/ws.2021.49033.2492 ISSN:2717-3755 https://water-soil.tabrizu.ac.ir/

**Research Article**

# The Effect of Meteorological Parameters and Satellite Images in Estimating Daily Reference Evapotranspiration Using a Random Forest Model Optimized with Genetic Algorithm

**H Talebi[1], S Samadianfard[2]*, Kh Valizadeh Kamran[3]**

Received: September 10, 2022 Accepted: December 10, 2022
Revised: December 3, 2022 Published online: December 22, 2023

1- PhD student, Dept. of Water Science and Engineering, Faculty of Agriculture, University of Tabriz, Iran
2- Assoc. Prof., Dept. of Water Science and Engineering, Faculty of Agriculture, University of Tabriz, Iran
3- Prof., Dept. of Remote Sensing and GIS, Faculty of Planning and Environmental Sciences, University of Tabriz, Iran.
*Corresponding Author, Email: s.samadian@tabrizu.ac.ir

## Abstract

### Background and Objectives

Water resources management, especially irrigation practices, is heavily reliant on reference evapotranspiration ($ET_0$). $ET_0$ is the rate of evaporation and transpiration from a standard reference surface with a presumed surface resistance of 70 s $m^{-1}$, the height of 0.12 m and an albedo of 0.23. Penman-Monteith FAO-56 (P-M FAO-56) approach is the most commonly used method for calculating $ET_0$. In spite of the fact that FAO-PM is achievable, its implementation remains inconvenient because it requires a large amount of meteorological data, which is derived from standard meteorological observation stations. In the absence of complete climate data, it is highly desirable to have a model with fewer input climatic dates. Therefore, remote sensing methods have been used and improved over time to estimate $ET_0$ at various spatial scales. Alternatively, it has been observed that the research community has become increasingly interested in obtaining data from metaheuristic algorithms that are based on artificial intelligence (AI).

### Methodology

In this research, it has been attempted to estimate the amount of daily reference evapotranspiration ($ET_0$) using two data-driven models, using a combination of inputs from meteorological station data and satellite imagery data from MODIS sensor, by considering different inputs from these sources. The models include the random forest (RF) and hybridized RF with genetic algorithm optimization (GA-RF). Moreover, the correlation of input variables with $ET_0$ is evaluated and the possibility of training a simple and accurate machine learning model in the conditions of lack or absence of meteorological data using satellite image data is investigated. So, this study aimed to determine $ET_0$ in the time period of 2003-2021 using land surface temperature (LST) data and leaf area index (LAI) acquired from MODIS sensor and Tabriz meteorological station data including maximum and minimum air temperatures ($T_{max}$, $T_{min}$), average temperatures (T), wind speeds ($U_2$), average relative humidity (RH), maximum and minimum relative humidity ($RH_{max}$, $RH_{min}$), and sunny hours (n). For the study area, daily LST were extracted from the Terra (MOD11A1) and Aqua (MYD11A1) satellites. Moreover, the LST of Terra and Aqua satellites were combined, since the LST values had missing data due to the presence of clouds. Furthermore, MODIS MCD15A3H version 6.1 using four-day data from Terra and Aqua satellites was used to determine the leaf area index (LAI). The standard P-M FAO-56 method for calculating daily reference evapotranspiration was considered as the base method. The set of input parameters was considered based on the cross-correlation of the parameters with reference evapotranspiration obtained from the FAO-Penman-Monteith equation.

©: The Author(S)
Publisher: University of Tabriz

## Findings

The results of two data-driven models including standalone random forest (RF) and hybridized RF model with genetic algorithm (GA) to estimate $ET_0$ values were compared with calculated $ET_0$ by P-M FAO-56 equation. The results indicated that all of the studied input variables are highly correlated with the target variable. Based on the P-M FAO-56 method, the average air temperature with the highest value ($R^2$=0.903) and the wind speed with the lowest value ($R^2$=0.282) has a high and low correlation with reference evapotranspiration. Also, by comparing LAI and LST MODIS parameters, LST has the highest correlation coefficient with $ET_0$ with $R^2$=0.865. A total of twelve scenarios for estimating $ET_0$ are evaluated, each with a different set of input parameters. Based on the correlation between the parameters and $ET_0$, the first ten scenarios are categorized. Additionally, the eleventh scenario is based only on satellite images, and the twelfth scenario is based solely on weather station data. Based on the results, the GA-RF-10 ($R^2$=0.976, RMSE=0.200, MAPE=11.373, and MBE=0.028), which includes all input parameters, outperforms the other models. There was a greater degree of accuracy with the RF-10 ($R^2$=0.949, RMSE=0.293, MAPE=16.442, and MBE=0.017) when compared with the other random forest models. Based on the comparison of scenario 11 (satellite image data) and scenario 12 (meteorological station data), it appears that scenario 12 is more accurate for both RF ($R^2$=0.922, RMSE=0.357, MAPE=20.712, and MBE=0.009) and GA-RF ($R^2$=0.944, RMSE=0.306, MAPE=17.037, and MBE=0.013) models. Despite the fact that only satellite image parameters did not provide accurate estimation of $ET_0$ compared to independent meteorological parameters, the inclusion of these parameters in the $ET_0$ estimation resulted in more acceptable results, demonstrating the importance of satellite image parameters. Thus, satellite data may be useful and recommended for estimating $ET_0$, particularly in areas without meteorological stations.

## Conclusion

According to the results obtained from the present research, GA-RF-10 model ($R^2$=0.976, RMSE=0.200, MAPE=11.373 and MBE=0.027), which includes all input parameters; It has the best performance among other models. According to the results, although the satellite parameters alone had weaker estimates than the independent meteorological parameters, but by including these parameters in the estimation of $ET_0$, more accurate results are obtained, which shows the importance of satellite image parameters in estimation.

**Keywords:** FAO-Penman-Monteith, Genetic algorithm, Land surface temperature, MODIS sensor, Reference evapotranspiration.

نشریه دانش آب و خاک / جلد ۳۳ شماره ۴ صفحه‌های ۲۰۱ تا ۲۱۶/ سال ۱۴۰۲

**مقاله پژوهشی**

# تاثیر پارامترهای هواشناسی و تصاویر ماهواره‌ای در تخمین تبخیر تعرق مرجع روزانه با استفاده از مدل جنگل تصادفی بهینه شده با الگوریتم ژنتیک

**حامد طالبی[۱]، سعید صمدیان‌فرد*[۲]، خلیل ولیزاده کامران[۳]**

تاریخ دریافت: ۱۴۰۱/۰۶/۱۹ تاریخ پذیرش: ۱۴۰۱/۰۹/۱۹

تاریخ ویرایش:۱۴۰۱/۰۹/۱۲ تاریخ انتشار انلاین :۱۴۰۲/۱۰/۰۱

۱-دانشجوی دکتری، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

۲-دانشیار، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

۳-استاد، گروه سنجش از دور، دانشکده برنامه‌ریزی و علوم محیطی، دانشگاه تبریز

*مسئول مکاتبات، پست الکترونیکی:s.samadian@tabrizu.ac.ir

**چکیده**

تخمین دقیق تبخیر و تعرق مرجع ($ET_0$) برای مدیریت کارآمد آب کشاورزی، مدل‌سازی محصول و برنامه‌ریزی آبیاری بسیار مهم است. در سال‌های اخیر، مدل‌های داده‌محور مبتنی بر هوش مصنوعی توجه محققان در سراسر جهان را به خود جلب کرده‌اند. این مطالعه با هدف تعیین $ET_0$ در زمین های زراعی تبریز برای سال‌های ۱۳۸۱-۱۴۰۰، با استفاده از داده‌های دمای سطح زمین (LST) و شاخص سطح برگ (LAI) از سنجده MODIS و داده‌های ایستگاه هواشناسی تبریز شامل دمای هوای حداکثر و حداقل ($T_{max}$,$T_{min}$)، دمای میانگین (T)، سرعت باد در ارتفاع دو متری ($U_2$)، رطوبت نسبی میانگین (RH)، رطوبت نسبی حداکثر و حداقل ($RH_{max}$, $RH_{min}$) و ساعات آفتابی (n) انجام گرفته است. روش استاندارد فائو-پنمن-مونتیث برای محاسبه تبخیر و تعرق مرجع روزانه به‌عنوان روش مبنا مورد نظر قرار گرفته شد. مجموعه پارامترهای ورودی مدل، براساس همبستگی متقابل پارامترها با تبخیر و تعرق مرجع بدست آمده از معادله فائو-پنمن-مونتیث تقسیم‌بندی شدند. دو مدل داده محور شامل مدل جنگل تصادفی (RF) و مدل جنگل تصادفی بهینه شده با الگوریتم ژنتیک (GA-RF) برای تخمین مقادیر $ET_0$ در نظر گرفته شد و نتایج آنها با $ET_0$ محاسبه شده توسط معادله فائو-پنمن-مونتیث مقایسه گردید. نتایج نشان داد که مدل GA-RF-10 ($R^2$=۰/۹۷۶ ، RMSE=۰/۲۰۰ ، MAPE=۱۱/۳۷۳ و MBE=۰/۰۲۷) که شامل همه پارامترهای ورودی است، بهترین عملکرد را در بین سایر مدل‌ها داشته است. براساس نتایج، دمای هوای میانگین بیشترین ($R^2$=۰/۹۰۳) و سرعت باد ($R^2$=۰/۲۸۲) کمترین همبستگی را با $ET_0$ دارند. همچنین، در همه حالت‌های مورد بررسی، مدل GA-RF نسبت به مدل RF عملکرد بهتری داشت. بنابراین، مدل GA-RF برای تعیین دقیق و مناسب $ET_0$ در شرایط اقلیمی مشابه و کمبود پارامترهای هواشناسی توصیه می‌گردد.

**واژه‌های کلیدی**: الگوریتم ژنتیک، تبخیر و تعرق مرجع، سنجده مادیس، شاخص سطح برگ، فائو-پنمن-مانتیث

**مقدمه**

دسترسی به آب شیرین به یکی از اساسی‌ترین مشکلات در طول قرن بیست و یکم در سراسر جهان تبدیل شده است. بیشتر آب شیرین موجود در سطح زمین توسط بخش کشاورزی مصرف می‌شود (مانیکوماری و همکاران ۲۰۱۷). امروزه تاثیر تغییرات آب و هوایی بر در دسترس بودن و کیفیت منابع آب سطحی و زیرزمینی شناخته شده است. علت اصلی کمبود و بحران آب در کشورهای در حال توسعه مانند ایران، عدم دسترسی به منابع آبی نیست، بلکه بهره برداری نادرست و سوء مدیریت آب است (معروف‌پور و همکاران ۲۰۲۰). تبخیر و تعرق[1] (ET) یکی از فرآیندهای حیاتی چرخه هیدرولوژیکی است و در مدیریت و بهر-برداری صحیح از منابع آبی تاثیر بسزایی دارد. تبخیر و تعرق، به ترکیب تبخیر آب از سطح زمین و گیاهان و تعرق از پوشش گیاهی از طریق روزنه‌های برگ گفته می‌شود (هیلل و همکاران ۲۰۰۵). عوامل متعددی بر میزان تبخیر و تعرق حاکم است که از جمله آنها می-توان به دما، تابش خورشید، رطوبت هوا و سرعت باد اشاره نمود (هلمز ۱۹۸٤). تبخیر و تعرق مرجع[2] ($ET_0$) میزان تبخیر و تعرق از یک سطح مرجع استاندارد با فرض مقاومت سطحی ۷۰ ($m\ s^{-1}$)، ارتفاع ۰/۱۲ (m) و آلبدو ۰/۲۳ (بی بعد) است (آلن و همکاران ۱۹۹۸). برآورد دقیق $ET_0$ اطلاعات مفیدی را برای مدیریت منابع آب، برنامه‌ریزی آبیاری و پایداری محصول فراهم می-کند. لایسمترها[3] معمولاً برای تخمین مستقیم $ET_0$ استفاده می‌شوند. با این حال، استفاده از لایسیمتر برای تخمین $ET_0$ بسیار محدود است، زیرا هزینه‌های نگهداری و عملیاتی بالایی دارند. برای غلبه بر این محدودیت، چندین روش تجربی برای تخمین تبخیر و تعرق توسعه داده شده است. معادله فائو-پنمن-مانتیث[4] (FAO-PM)، به عنوان یک روش معتبر، در تعداد زیادی از کارهای تحقیقاتی به‌عنوان استانداردی برای مقایسه روش‌های مختلف استفاده می‌شود (چیا و همکاران ۲۰۲۰). با این حال، عیب قابل توجه استفاده از این روش، تعداد زیاد متغیرهای اقلیمی و عدم قطعیت در به‌دست آوردن دقیق آن‌ها است. از دیدگاه تغییرات زمانی و مکانی، تراکم ایستگاه‌های هواشناسی (تعداد ایستگاه‌ها در واحد سطح) برای ثبت تغییرات مکانی و زمانی کافی نیست. امروزه، تکنیک‌های سنجش از دور، که اطلاعات سطح زمین را از گستره جغرافیایی وسیع‌تری جمع‌آوری می‌کنند، به عنوان ابزار و روش شناسی مؤثرتری برای استخراج پارامترهای زمین شناخته می‌شوند که می‌توانند برای تخمین $ET_0$ در مقیاس منطقه‌ای مورد استفاده قرار گیرند (یانگ و همکاران ۲۰۱۲). برای این منظور، در مناطقی که با کمبود داده‌های اندازه‌گیری شده مواجه باشند، داده‌های سنجش از دور می‌توانند به طور قابل‌توجهی اطلاعات مورد نیاز مانند شاخص سطح برگ[5] (LAI) و دمای سطح زمین[6] (LST) را تامین کنند. در دهه‌های گذشته، تکنیک‌های یادگیری ماشین به‌طور فزاینده‌ای در کاربردهای هیدرولوژیکی و اکولوژیکی مورد استفاده قرار گرفته‌اند (ستاری و همکاران۲۰۲۱، صمدیان‌فرد و همکاران ۲۰۲۱). به طور خاص، در چند سال اخیر تکنیک‌های یادگیری ماشین برای پیش‌بینی تبخیر و

[1] Evapotranspiration
[2] Reference evapotranspiration
[3] Lysimeter
[4] Penman–Monteith FAO-56
[5] Leaf Area Index
[6] Land Surface Temperature

تعرق مرجع ($ET_0$) در اکوسیستم‌ها و محصولات مختلف و مشخص کردن تأثیرات متغیرهای مرتبط با آن استفاده شده است (موزر و همکاران ۲۰۲۱، ریس روجاس و همکاران ۲۰۲۱). برآورد موفقیت‌آمیز اطلاعات پوشش گیاهی و سطح زمین از تصاویر ماهواره‌ای، راه را برای پیش‌بینی تبخیر و تعرق مرجع ($ET_0$) با استفاده از مدل‌های هوش مصنوعی باز کرده است. منابع اصلی تصاویر ماهواره‌ای را می‌توان از لندست و طیف رادیومتر تصویربرداری با وضوح متوسط[1] (MODIS) به‌دست آورد. داده‌های سنجش از دور امکان ثبت اطلاعات بیشتر توسط تصاویر ماهواره‌ای را فراهم می‌کند. در واقع روش سنجش از دور قابلیت استخراج اطلاعات پوشش گیاهی و تابش را دارد، که برای تخمین $ET_0$ مفید هستند. با ادغام اندازه‌گیری‌ها (از مشاهدات ماهواره‌ای و زمینی) با تخمین‌های مدل، می‌توان نقص در مشاهدات و خطاها در پیش‌بینی مدل را به حداقل رساند (رودل و همکاران ۲۰۰٤). هوانگ و همکاران (۲۰۱۹) روش جنگل تصادفی، رگرسیون بردار پشتیبان و الگوریتم کت بوست[2] را در تخمین مقادیر $ET_0$ روزانه در پنج منطقه مرطوب چین با داده‌های محدود و کامل بررسی کردند. نتایج نشان داد، با کمبود داده، روش‌های ماشین بردار پشتیبان و جنگل تصادفی نتایج رضایت بخشی ارائه می‌دهند. وانگ و همکاران (۲۰۱۹)، در تحقیقی جهت تخمین $ET_0$ روزانه به بررسی توانایی مدل جنگل تصادفی[3] (RF) و مقایسه این مدل با روش برنامه ریزی بیان ژن[4] (GEP)، با ترکیب پارامترهای ورودی‌های مختلف پرداختند. نتایج نشان داد مدل RF با متغییرهای ورودی کامل و ناقص عملکرد بهتری نسبت به مدل GEP داشت. دونا و همکاران (۲۰۲۱) طی تحقیقی توانایی الگوریتم جنگل تصادفی را برای پیش‌بینی $ET_0$ روزانه با استفاده از داده‌ها هواشناسی و محصولات تصاویر ماهواره‌ای MODIS (شاخص سطح برگ و شاخص دمای سطح زمین) بررسی کردند. نتایج نشان داد، پیش‌بینی‌های تبخیر و تعرق روزانه الگوریتم جنگل تصادفی مطابقت بهتری با مقادیر اندازه‌گیری شده نسبت به محصول جهانی MODIS ET نشان می‌دهند. لیو و همکاران (۲۰۲۲) به منظور تخمین $ET_0$ روزانه در جنوب غربی چین از مدل یادگیری ماشین افراطی[5] (ELM) و مدل یادگیری ماشین افراطی بهینه شده با الگوریتم ژنتیک[6] (GA-ELM) استفاده کردند. نتایج نشان داد که؛ مدل GA-ELM با در نظر گرفتن همه پارامترهای ورودی (رطوبت نسبی، دمای حداقل، دمای حداکثر، سرعت باد و تابش خالص) بهترین عملکرد را نسبت به سایر مدل‌ها دارد. در پژوهش حاضر سعی شده است که، با در نظر گرفتن ترکیب ورودی‌های مختلف از داده‌های ایستگاه هواشناسی و تصاویر ماهواره‌ای از حسگر مادیس، مقدار تبخیر و تعرق مرجع روزانه ($ET_0$) به وسیله دو مدل داده محور، شامل جنگل تصادفی (RF) و جنگل تصادفی بهینه شده با الگوریتم ژنتیک (GA-RF) مدل-سازی شود. همچنین همبستگی متغیرهای ورودی با $ET_0$ مورد ارزیابی قرار گرفته و امکان آموزش یک مدل یادگیری ماشین ساده و با دقت مناسب در شرایط کمبود و یا عدم وجود داده‌های هواشناسی با استفاده

[1] Moderate resolution Imaging Spectroradiometer
[2] Cat Boost
[3] Random Forest
[4] Gene-expression programming
[5] Extreme learning machine
[6] Genetic algorithm

از داده‌های تصاویر ماهواره‌ای، مورد بررسی قرار می‌گیرد.

## مواد و روش‌ها

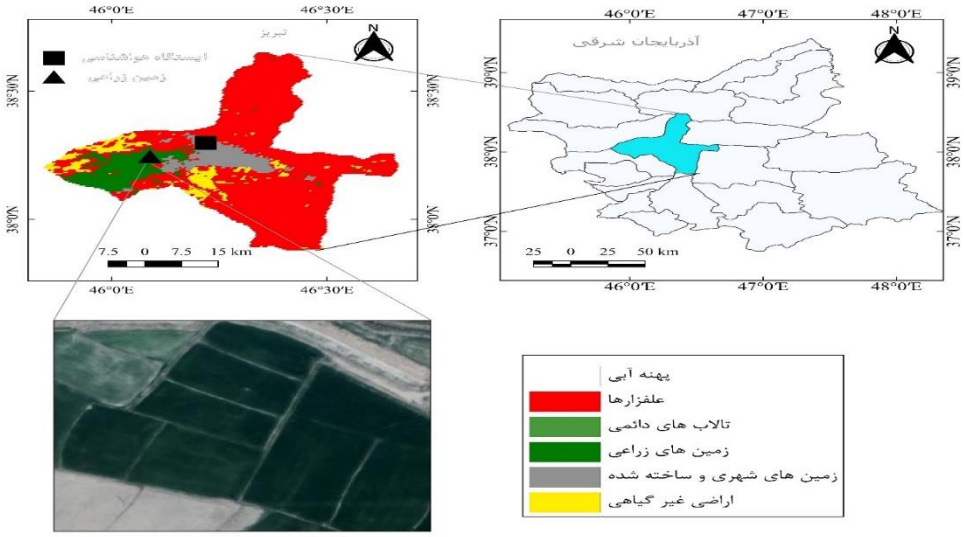
### معرفی منطقه مورد مطالعه

پژوهش حاضر در اراضی کشاورزی شمال غرب ایران در استان آذربایجان شرقی، شهر تبریز انجام شده است. شهر تبریز بین عرض جغرافیایی °۳۷/۸۷۰ درجه تا °۳۸/٦٤۰ درجه شمالی و طول جغرافیایی °٤٥/۸٥۰ درجه تا °٤٦/٦۹۰ درجه شرقی واقع شده است (شکل ۱). برای شناسایی اراضی زراعی شهر تبریز، از محصول نوع پوشش زمین سنجده مادیس (MCD12Q1) استفاده شد. براساس روش اقلیمی دومارتن، اقلیم منطقه در رده نیمه خشک قرار می‌گیرد. میزان بارندگی سالانه در شهر تبریز حدود ۲٦۲ میلی متر است و تابستان خشک و نیمه گرم و بهار خوب و معتدل دارد. مقدار تبخیر و تعرق پتانسیل در حدود ۱۷۳۳ میلی‌متر در سال و دمای هوای متوسط سالانه تقریباً ۱۳ درجه سانتیگراد می‌باشد.

در این پژوهش، روش استاندارد فائو-پنمن-مونتیث برای محاسبه تبخیر و تعرق مرجع روزانه به‌عنوان روش مبنا مورد نظر قرار گرفته است (رابطه ۱).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T_{min} + 273}U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad [۱]$$

در این رابطه، $ET_0$ تبخیر و تعرق گیاه مرجع ($mm\ day^{-1}$)، $\Delta$ شیب منحنی فشار بخار اشباع ($kPa\ °C^{-1}$)، $R_n$ تابش خالص خورشیدی ($MJ\ m^{-2}\ d^{-1}$)، G چگالی شار حرارتی خاک ($MJ\ m^{-2}\ d^{-1}$)، $\gamma$ ثابت سایکرومتری ($kPa\ °C^{-1}$)، $T_{min}$ متوسط دمای هوا (°C)، $U_2$ متوسط سرعت باد در ارتفاع دو متری از سطح زمین ($m\ s^{-1}$)، $e_s$ فشار بخار اشباع (kPa) و $e_a$ نیز فشار بخار واقعی (kPa) می‌باشد.




**شکل ۱**- موقعیت جغرافیایی منطقه مورد مطالعه.

براساس داده‌های مستخرج از ماهواره مادیس و ایستگاه هواشناسی تبریز برای سال‌های ۱٤۰۰-۱۳۸۱ (٪۷۰ آموزش و ٪۳۰ آزمایش)، الگوهای ورودی برای مدل تعریف گردید و مقادیر تبخیر و تعرق گیاه مرجع

حاصل از روش جنگل تصادفی (RF) در دو حالت منفرد و بهینه شده با الگوریتم ژنتیک (GA-RF) مدل‌سازی و در نهایت با بررسی و ارزیابی دقت پارامترهای ورودی و نتایج مدل در برآورد $ET_0$، دقیق‌ترین روش پیشنهاد گردید. در شکل ۲ مراحل روش کار در مطالعه حاضر نشان داده شده است. برای تخمین $ET_0$ روزانه در اقلیم‌های مورد مطالعه، متغیرهای ورودی زیر مورد استفاده قرار گرفت: شاخص سطح برگ (LAI)، دمای روزانه سطح زمین (LST) از سنجده MODIS (جدول ۱) و داده‌های هواشناسی شامل: دمای هوای حداکثر و حداقل($T_{max}$,$T_{min}$) ، دمای میانگین ($T_{mean}$)، سرعت باد در ارتفاع دو متری ($U_2$)، رطوبت نسبی میانگین ($RH_{mean}$)، رطوبت نسبی حداکثر و حداقل ($RH_{max}$, $RH_{min}$) و ساعات آفتابی (n).

## تصاویر ماهواره‌ای

سنجده مادیس که بر روی ماهواره‌های ترا[1] و آکوا[2] قرار دارد، قادر به تولید تصاویر در ۳۶ باند طیفی است. این سنجده قابلیت وضوح فضایی ۲۵۰ متر، ۵۰۰ متر و ۱ کیلومتر را در چند باند دارد. سنجنده‌های مادیس دارای وضوح مکانی ۱ کیلومتر در طیف مادون قرمز حرارتی هستند. در طول یک روز، ماهواره‌ها همزمان در ساعت محلی از روی یک مکان خاص عبور می‌کنند.

ماهواره ترا تقریباً در ساعت ۱۰:۳۰ صبح به وقت محلی از شمال به جنوب از خط استوا می‌گذرد و ماهواره آکوا تقریباً در ساعت ۱:۳۰ بعد از ظهر به وقت محلی از جنوب به شمال از خط استوا عبور می‌کند. داده‌های سنجده مادیس توسط سازمان ملی هوانوردی و فضایی ایالات متحده پردازش و ارائه می‌شود. در این پژوهش به منظور تخمین تبخیر و تعرق مرجع روزانه، از دمای سطح زمین روزانه و شاخص سطح برگ مستخرج از تصاویر ماهواره‌ای استفاده شده است. دمای سطح زمین روزانه (LST) از ماهواره ترا (MOD11A1) و آکوا (MYD11A1) برای منطقه مورد مطالعه استخراج گردید. از آن‌جایی که دمای سطح زمین به‌دلیل وجود ابر، دارای داده‌های گمشده می‌باشد، برای رفع این مورد، دمای سطح زمین ماهواره‌های ترا و آکوا باهم ترکیب شدند. شاخص سطح برگ (LAI) با استفاده از محصول MCD15A3H نسخه ۶.۱ مادیس، بصورت داده‌های چهار روزه، ترکیب ماهواره ترا و آکوا با وضوح مکانی ۵۰۰ متر تعیین شده است.

## جنگل تصادفی

الگوریتم جنگل تصادفی (RF)، اولین بار توسط بریمن پیشنهاد شده است (بریمن ۲۰۰۱). RF یک الگوریتم یادگیری ماشین مبتنی بر درخت تصمیم می‌باشد. این الگوریتم به دلیل سادگی و قابلیت استفاده، هم برای دسته‌بندی و هم رگرسیون مورد استفاده قرار می‌گیرد. در RF، درخت‌های تصمیم در زیر مجموعه‌های مختلف داده‌های آموزشی برازش داده می‌شوند. مزیت الگوریتم RF در مقایسه با سایر الگوریتم‌ها این است که اهمیت هر یک از متغیرهای ورودی را تخمین می‌زند. در الگوریتمRF، بسیاری از متغیرهای ورودی که ممکن است به پیش بینی کمک کنند، می‌توانند استفاده شوند (تیرالیس و همکاران ۲۰۱۹).

## الگوریتم ژنتیک

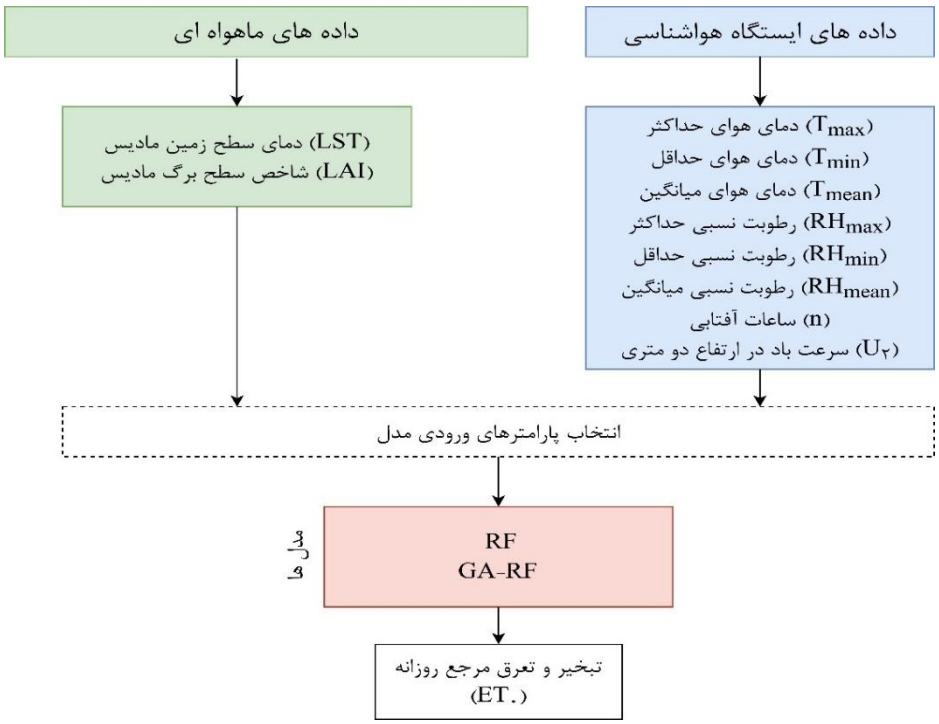
الگوریتم ژنتیک (GA) اولین بار در سال ۱۹۹۲ توسط هلند ارایه شده است (هلند ۱۹۹۲). اگرچه این الگوریتم به‌عنوان یکی از تکنیک‌های فراابتکار قدیمی بر اساس نظریه تکامل داروین شناخته می‌شود، اما یکی از کارآمدترین الگوریتم‌ها برای حل مسائل بهینه‌سازی است. طرح کلی یک الگوریتم GA شامل جمعیتی است که در آن هر یک از اجزای تشکیل دهنده، به نام کروموزوم، پاسخ به هر معما در نظر گرفته می‌شود. جستجو در این الگوریتم با تولید تصادفی جمعیت آغاز

[1] Terra
[2] Aqua

می شود. نسل‌های بعدی این جمعیت توسط عملگرها (انتخاب، متقاطع و جهش) گسترش می‌یابد. بر اساس اصل بقای بهترین‌ها، جمعیت نسل به نسل توسعه می یابد تا راه‌حل های مناسب تری ارائه دهد. مانند تکامل طبیعی، این روش باعث می شود که جمعیت نسل بعدی نسبت به نسل قبلی خود بهتر با محیط سازگار شوند و فرد بهینه در میان جمعیت نسل نهایی را می‌توان به عنوان راه حل بهینه تقریبی برای مشکل حل کرد.

**جنگل تصادفی بهینه شده توسط الگوریتم ژنتیک**

این روش عمدتاً از دو بخش تشکیل شده است: تنظیم پارامترها و بهینه سازی روش جنگل تصادفی. در بخش اول، پارامتر پارامترهای جنگل تصادفی، مانند مقیاس جنگل، تعداد ویژگی‌های تقسیم و حداکثر عمق درخت تصمیم را شناسایی می کند. سپس، در بخش بهینه‌سازی جنگل تصادفی با استفاده از الگوریتم ژنتیک، بهینه‌سازی ترکیب درخت‌های تصمیم با هدف حداکثر کردن امتیاز سود با در نظر گرفتن بازده و زیان واقعی و بالقوه انجام می‌شود (یی و همکاران ۲۰۱۸).




**شکل ۲- فلوچارت روش برآورد $ET_0$ در منطقه مورد مطالعه.**

**جدول ۱-اطلاعات داده‌های سنجش از دور.**

| منبع داده ها | پارامتر | ماهواره | وضوح فضایی | وضوح زمانی |
|---|---|---|---|---|
| سنجده مادیس (MODIS) | دمای سطح زمین LST (MOD11A1) | ترا | ۱ کیلومتر | ۱ روزه |
| | دمای سطح زمین LST (MYD11A1) | آکوا | | |
| | شاخص سطح برگ LAI (MCD15A3H) | ترا/آکوا | ۵۰۰ متر | ٤ روزه |

**توابع عملکرد**

مقادیر خطای بین روش‌های محاسباتی توسط پارامترهای آماری میانگین انحراف خطا[1] (MBE)، جذر

[1] Mean bias error

میانگین مربعات[1] (RMSE)، ضریب تعیین[2] ($R^2$) و درصد میانگین مقادیر مطلق خطا[3](MAPE) و با استفاده از روابط ۲ تا ۵ مورد بررسی قرار گرفت.

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) \qquad [۲]$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \qquad [۳]$$

$$R^2= 1 - \frac{\sum(y_i-x_i)^2}{\sum(x_i-\overline{x_i})^2} \qquad [٤]$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| * 100 \qquad [٥]$$

در ر.ابط فوق، n تعداد داده‌های مشاهده شده، $y_i$ مقادیر تخمینی، $x_i$ مقادیر مشاهده شده و $\overline{x_i}$ میانگین مقادیر مشاهده شده می‌باشد.

## نتایج و بحث

در مطالعه حاضر، مدل‌های RF و RF-GA با ترکیب ورودی‌های مختلف برای تخمین تبخیر و تعرق مرجع مورد استفاده قرار گرفته و دقت آن‌ها ارزیابی شده است. مجموعه داده مورد استفاده برای ساخت مدل شامل ٦٩٤٠ داده (١٣٨١–١٤٠٠) است که ٧٠٪ داده‌ها (١٣٨١–١٣٩٦) برای مرحله آموزش و ٣٠٪ داده‌ها (١٣٩٧–١٤٠٠) برای مرحله آزمایش تقسیم‌بندی شده‌اند. جدول ۲ معیارهای آماری بیشترین، کمترین، میانگین، خطای استاندار میانگین و انحراف معیار را برای مجموعه داده‌ها نشان می‌دهد.

همچنین، همبستگی بین پارامترهای ورودی و هدف در جدول ۳ نشان داده شده است. با توجه به نتایج، همه ورودی‌های مورد مطالعه رابطه قوی با متغیر هدف دارند. دمای هوا میانگین دارای بیشترین مقدار ($R^2$=۰/۹۰۳) و سرعت باد در ارتفاع دو متری کمترین مقدار ($R^2$=۰/۲۸۲) همبستگی را با تبخیر و تعرق مرجع بدست آمده از روش فائو- پنمن- مانیث دارند.

جدول ٤ پارامترهای ورودی برای هر مدل را نشان می‌دهد، که عبارتند از: دمای هوا میانگین، حداقل و حداکثر دمای هوای، سرعت باد در ارتفاع دو متری، ساعات آفتابی، حداقل و حداکثر رطوبت نسبی، میانگین رطوبت نسبی، دمای سطح زمین و شاخص سطح برگ. جدول ٤ نشان می‌دهد که دوازده سناریو ممکن برای تخمین تبخیر و تعرق مرجع هر کدام با مجموعه‌ای از پارامترهای ورودی متفاوت ارزیابی می شوند. ده سناریوی اول براساس همبستگی متقابل پارامترها با تبخیر و تعرق مرجع تقسیم‌بندی شده‌اند. به‌عنوان مثال، سناریوی اول براساس دمای هوای میانگین برای تعیین تبخیر و تعرق مرجع و سناریوی دوم براساس میانیگین دمای هوا و حداکثر دمای هوا تقسیم بندی شده‌اند. بقیه سناریوها را نیز می‌توان به همین صورت تفسیر کرد. علاوه بر این، سناری یازده فقط از داده های ماهواره ای و سناریوی دوازده نیز فقط براساس داده‌های ایستگاه هواشناسی تقسیم‌بندی شده‌اند. هدف از تقسیم‌بندی سناریوی یازده و دوازده، بررسی جداگانه نقش پارامترهای هواشناسی و تصاویر ماهواره‌ای در تخمین تبخیر و تعرق مرجع می‌باشد.

[1] Root mean squared error
[2] Coefficient of determination
[3] Mean absolute percentage error

**جدول ۲- ویژگی‌های آماری داده‌های مورد استفاده برای مرحله آزمایش.**

| پارامتر | واحد | اختصار | بیشترین | کمترین | میانگین | خطای استاندارد | انحراف معیار |
|---|---|---|---|---|---|---|---|
| سرعت باد در ارتفاع دو متری | $m\ s^{-1}$ | U2 | ۹/۳۷ | ۰/۲۵ | ۳/۴۹ | ۰/۴۱ | ۱/۵۹ |
| دمای هوای حداکثر | ◦C | Tmax | ۴۱/۰۰ | −۶/۴۰ | ۱۹/۸۳ | ۰/۲۹ | ۱۱/۲۶ |
| دمای هوای حداقل | ◦C | Tmin | ۳۰/۰۰ | −۱۵/۲۰ | ۸/۵۸ | ۰/۲۴ | ۹/۲۳ |
| دمای هوا میانگین | ◦C | Tmean | ۳۵/۱۰ | −۹/۴۰ | ۱۳/۸۷ | ۰/۲۶ | ۱۰/۲۸ |
| رطوبت نسبی حداکثر | % | RHmax | ۹۹/۰۰ | ۱۵/۰۰ | ۶۷/۲۸ | ۰/۵۰ | ۱۹/۲۶ |
| رطوبت نسبی حداقل | % | RHmin | ۹۱/۰۰ | ۴/۰۰ | ۳۴/۲۶ | ۰/۴۹ | ۱۸/۹۶ |
| رطوبت نسبی میانگین | % | RHmean | ۹۵/۳۷ | ۱۰/۲۵ | ۵۰/۷۶ | ۰/۵۰ | ۱۹/۳۰ |
| ساعات آفتابی | Hr | n | ۱۳/۸۰ | ۰/۰۰ | ۷/۹۱ | ۰/۱۰ | ۳/۹۸ |
| دمای سطح زمین | ◦C | LST | ۴۸/۴۹ | −۶/۴۹ | ۲۴/۰۸ | ۰/۳۳ | ۱۲/۶۱ |
| شاخص سطح برگ | بی‌بعد | LAI | ۷/۰۰ | ۰/۰۰ | ۰/۶۶ | ۰/۰۱ | ۰/۵۳ |
| تبخیر و تعرق مرجع | $mm\ day^{-1}$ | $ET_0$ | ۴/۷۷ | ۰/۱۷ | ۱/۹۴ | ۰/۰۳ | ۱/۲۷ |

**جدول ۲- ضریب همبستگی پارامترهای ورودی با تبخیر و تعرق مرجع فائو-پنمن-مانتیث.**

| پارامتر | Tmean | Tmax | LST | Tmin | RHmean | RHmin | n | RHmax | LAI | U2 |
|---|---|---|---|---|---|---|---|---|---|---|
| همبستگی | ۰/۹۰۳ | ۰/۹۰۱ | ۰/۸۶۵ | ۰/۸۵۳ | −۰/۷۵۵ | −۰/۷۳۱ | ۰/۷۲۰ | −۰/۶۷۵ | ۰/۵۲۷ | ۰/۲۸۲ |

**جدول ۳-ترکیب ورودی‌های مدل مورد مطالعه.**

| شماره | پارامتر | پارامتر های ورودی | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ۱ | ET0 | Tmean | | | | | | | | | |
| ۲ | ET0 | Tmean | Tmax | | | | | | | | |
| ۳ | ET0 | Tmean | Tmax | LST | | | | | | | |
| ٤ | ET0 | Tmean | Tmax | LST | Tmin | | | | | | |
| ٥ | ET0 | Tmean | Tmax | LST | Tmin | RHmean | | | | | |
| ٦ | ET0 | Tmean | Tmax | LST | Tmin | RHmean | RHmin | | | | |
| ۷ | ET0 | Tmean | Tmax | LST | Tmin | RHmean | RHmin | n | | | |
| ۸ | ET0 | Tmean | Tmax | LST | Tmin | RHmean | RHmin | n | RHmax | | |
| ۹ | ET0 | Tmean | Tmax | LST | Tmin | RHmean | RHmin | n | RHmax | LAI | |
| ۱۰ | ET0 | Tmean | Tmax | LST | Tmin | RHmean | RHmin | n | RHmax | LAI | U2 |
| ۱۱ | ET0 | − | − | LST | − | − | − | − | − | LAI | − |
| ۱۲ | ET0 | Tmean | Tmax | − | Tmin | RHmean | RHmin | n | RHmax | − | U2 |

همچنین، جدول ۵ پارامترهای مورد استفاده در توسعه مدل RF و پارامترهای متناظر بهینه شده در مدل‌های GA-RF را با توجه به تعداد درختان، عمق حداکثر، اطمینان، حداقل اندازه برگ، حداقل اندازه برای تقسیم، تعداد جایگزین‌های پیش هرس، و نسبت زیر مجموعه نشان می‌دهد.

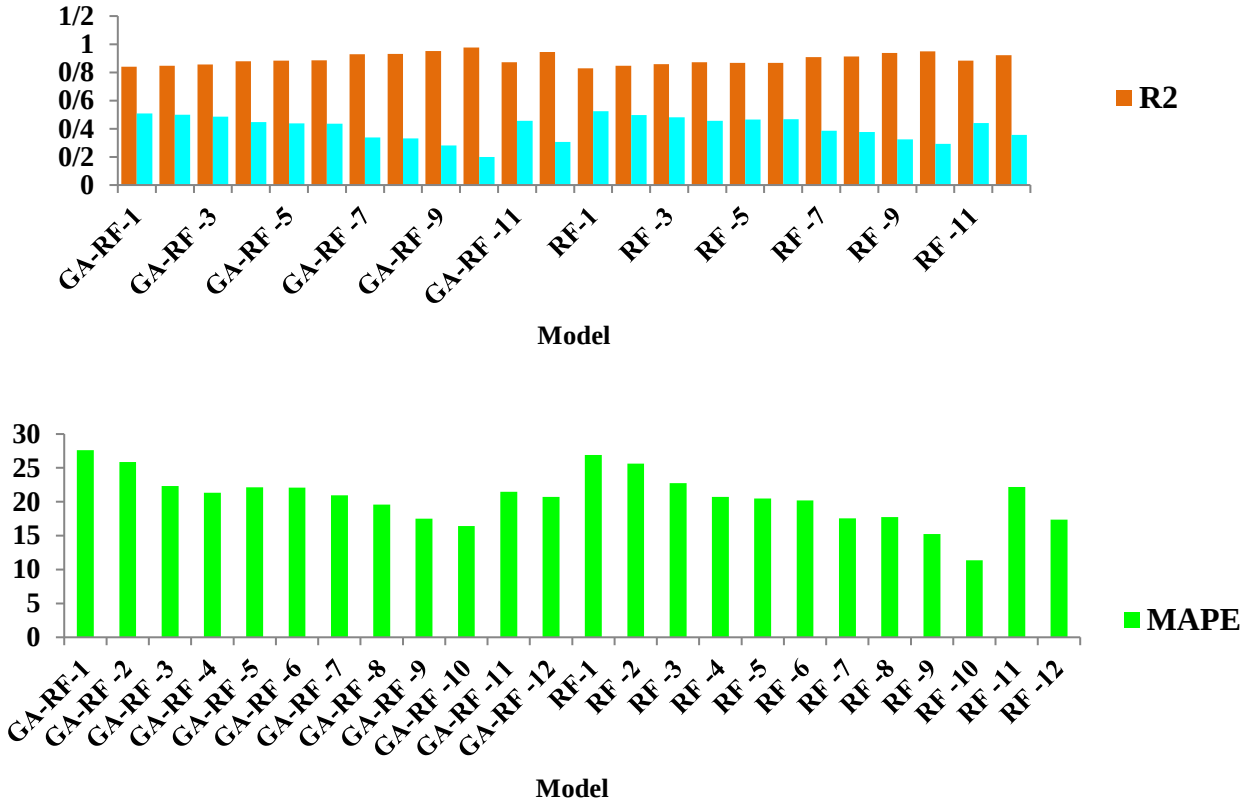
**جدول ۴-پارامترهای مدل های RF و GA-RF.**

| مدل | پارامتر | | | | | | |
|---|---|---|---|---|---|---|---|
| | تعداد درختان | عمق حداکثر | اطمینان | حداقل اندازه برگ | حداقل اندازه برای تقسیم | تعداد جایگزین-های پیش هرس | نسبت زیر مجموعه |
| RF | ۱۰۰ | ۱۰ | ۰/۱۰۰ | ۲ | ٤ | ۳ | ۰/۲۰۰ |
| GA-RF-1 | ۸۱ | ٥ | ۰/۳٥٥ | ٤۱ | ۲۰ | ٥٥ | ۰/۱۳۸ |
| GA-RF -2 | ۹٤ | ٥ | ۰/۳٤٦ | ٤۱ | ۲۰ | ۳ | ۰/۱٤٦ |
| GA-RF -3 | ۹٤ | ٥ | ۰/۳٤۸ | ٤۱ | ۲۰ | ۳ | ۰/۱٥۱ |
| GA-RF -4 | ۹٤ | ۳٦ | ۰/۲۰۷ | ۱ | ۸۰ | ۳ | ۰/۲۸٤ |
| GA-RF -5 | ۸۱ | ۳٦ | ۰/۲۲۸ | ۱ | ۸۰ | ٥٥ | ۰/۳۱٤ |
| GA-RF -6 | ۹٤ | ۳٦ | ۰/۲۱۳ | ۱ | ۸۰ | ۳ | ۰/۲۹٤ |
| GA-RF -7 | ٦۳ | ۸٦ | ۰/۲۸۲ | ۲۳ | ٤۹ | ۱۸ | ۰/۹٦۸ |
| GA-RF -8 | ۹٤ | ۳٦ | ۰/۲۰۸ | ۱ | ۸۰ | ۳ | ۰/۲۷۹ |
| GA-RF -9 | ۷۸ | ٦٥ | ۰/۱٤۸ | ٤۱ | ۲۰ | ۳ | ۰/۲٤٥ |
| GA-RF -10 | ۷۸ | ٦٥ | ۰/۱٥۷ | ۹۸ | ۱۳ | ٦۸ | ۰/۲٥۰ |
| GA-RF -11 | ۹٤ | ٥ | ۰/۳٦۲ | ٤۱ | ۲۰ | ۳ | ۰/۱۲۹ |
| GA-RF -12 | ۹٤ | ۳٦ | ۰/۲۱۳ | ۱ | ۸۰ | ۳ | ۰/۲۹٤ |

علاوه‌براین، نتایج معیارهای ارزیابی برای مدل-های RF و GA-RF در جدول ٦ و شکل ۳ نشان داده شده است. مدل GA-RF-10 در بین مدل‌های بهینه شده با الگوریتم ژنتیک، با $R^2$=۰/۹۷٦ ، RMSE=۰/۲۰۰ ، MAPE=۱۱/۳۷۳ و MBE=۰/۰۲۷ و مدل RF-10 در بین مدل‌های جنگل تصادفی، با $R^2$=۰/۹٤۹ ، RMSE=۰/۲۹۳ ، MAPE=۱٦/٤٤۲ و MBE=۰/۰۱۷ تخمین‌های دقیق با عملکرد خوب ارائه داده‌اند. مقایسه دو مدل مذکور نشان می‌دهد که GA-RF عملکرد بهتری را نسب به مدل RF ثبت کرده است. همچنین، نتایج معیارهای ارزیابی برای سناریوهای یازده و دوازده نشان می‌دهد که در هر دو حالت (GA-RF وRF)، تخمین تبخیر و تعرق مرجع با داده‌های ایستگاه هواشناسی، نتایج بهتری داشته‌اند.

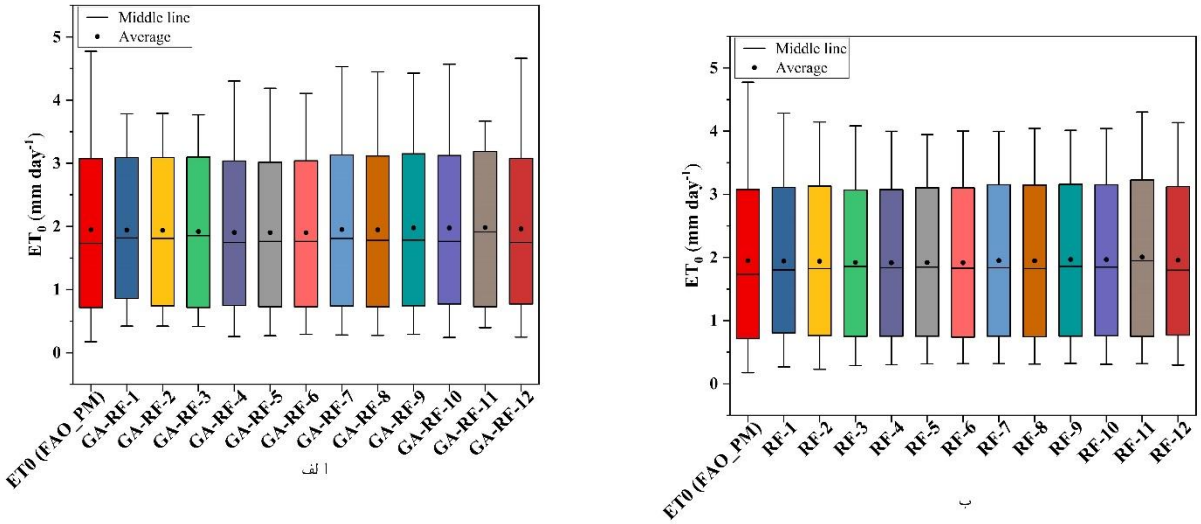
**جدول ۶- نتایج محاسبه‌ها براساس مدل جنگل تصادفی بهینه شده با الگوریتم ژنتیک (GA-RF) و جنگل تصادفی (RF).**

| مدل | پارامترهای آماری | | | |
|---|---|---|---|---|
| | $R^2$ | RMSE | MAPE | MBE |
| RF-1 | ۰/۸۳۰ | ۰/۵۲۵ | ۲۷/۶۲۰ | -۰/۰۰۵ |
| RF -2 | ۰/۸۴۸ | ۰/۴۹۷ | ۲۵/۰۸۷ | -۰/۰۰۹ |
| RF -3 | ۰/۸۵۹ | ۰/۴۸۱ | ۲۲/۳۱۸ | -۰/۰۲۷ |
| RF -4 | ۰/۸۷۲ | ۰/۴۵۷ | ۲۱/۳۲۷ | -۰/۰۳۲ |
| RF -5 | ۰/۸۶۷ | ۰/۴۶۶ | ۲۲/۱۵۳ | -۰/۰۲۸ |
| RF -6 | ۰/۸۶۷ | ۰/۴۶۷ | ۲۲/۰۹۷ | -۰/۰۳۲ |
| RF -7 | ۰/۹۰۹ | ۰/۳۸۶ | ۲۰/۰۸۹ | ۰/۰۰۲ |
| RF -8 | ۰/۹۱۳ | ۰/۳۷۷ | ۱۹/۵۶۳ | -۰/۰۰۱ |
| RF -9 | ۰/۹۳۷ | ۰/۳۲۴ | ۱۷/۵۲۶ | ۰/۰۱۹ |
| RF -10 | ۰/۹۴۹ | ۰/۲۹۳ | ۱۶/۴۴۲ | ۰/۰۱۷ |
| RF -11 | ۰/۸۸۳ | ۰/۴۴۰ | ۲۱/۴۶۲ | ۰/۰۵۷ |
| RF -12 | ۰/۹۲۲ | ۰/۳۵۷ | ۲۰/۷۱۲ | ۰/۰۰۹ |
| GA-RF-1 | ۰/۸۴۱ | ۰/۵۰۹ | ۲۶/۹۱۵ | -۰/۰۰۷ |
| GA-RF -2 | ۰/۸۴۶ | ۰/۵۰۰ | ۲۵/۶۴۷ | -۰/۰۱۰ |
| GA-RF -3 | ۰/۸۵۶ | ۰/۴۸۵ | ۲۲/۷۶۵ | -۰/۰۲۹ |
| GA-RF -4 | ۰/۸۷۸ | ۰/۴۴۸ | ۲۰/۷۳۷ | -۰/۰۴۵ |
| GA-RF -5 | ۰/۸۸۴ | ۰/۴۳۸ | ۲۰/۴۸۳ | -۰/۰۴۷ |
| GA-RF -6 | ۰/۸۸۵ | ۰/۴۳۵ | ۲۰/۱۹۹ | -۰/۰۴۸ |
| GA-RF -7 | ۰/۹۳۰ | ۰/۳۳۸ | ۱۷/۵۴۰ | ۰/۰۰۳ |
| GA-RF -8 | ۰/۹۳۲ | ۰/۳۳۲ | ۱۷/۰۷۴ | -۰/۰۰۱ |
| GA-RF -9 | ۰/۹۵۲ | ۰/۲۸۱ | ۱۵/۲۲۳ | ۰/۰۲۸ |
| GA-RF -10 | ۰/۹۷۶ | ۰/۲۰۰ | ۱۱/۳۷۳ | ۰/۰۲۷ |
| GA-RF -11 | ۰/۸۷۳ | ۰/۴۵۶ | ۲۲/۱۷۴ | ۰/۰۳۵ |
| GA-RF -12 | ۰/۹۴۴ | ۰/۳۰۶ | ۱۷/۰۳۷ | ۰/۰۱۳ |



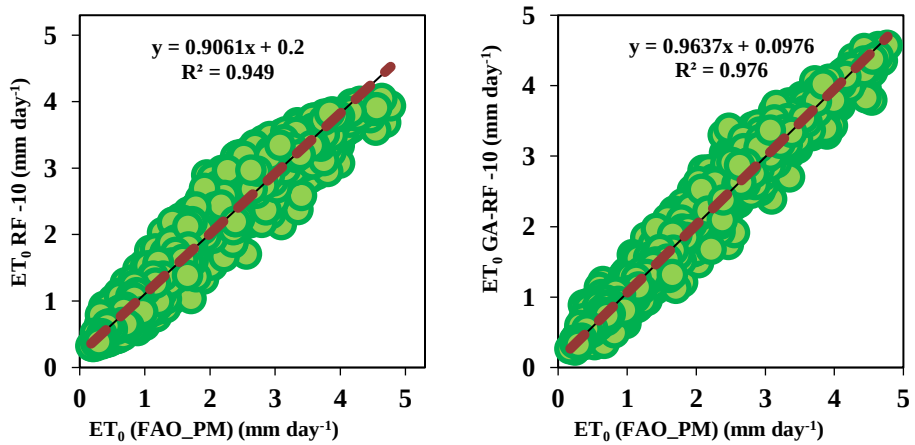

**شکل ۱- نمودارهای ستونی پارامترهای آماری.**

شکل ٤، عملکرد مدل‌های مورد مطالعه را با استفاده از نمودارهای جعبه‌ای نشان می‌دهد. همانطور که در شکل ٤ مشاهده می‌شود، مدل GA-RF-10 و RF-10 نتایج دقیق‌تری را در تخمین تبخیر و تعرق مرجع ارائه می‌دهند. در مقایسه میانه این دو مدل با تبخیر و تعرق مرجع محاسبه شده نتیجه‌گیری می‌شود که مطابق باهم هستند و در یک خط می‌باشند. پس می توان تفسیر کرد که بین این دو مدل تفاوت چندانی وجود ندارد. برای محدوده بین چارکی، هر دو مدل (RF-10 و GA-RF-10) به نظر دارای طول یکسانی هستند که نشان دهنده پراکندگی مشابهی از نظر داده‌ها است. همچنین، طول بازوها نشان دهنده گسترش کلی داده‌ها است. حداقل و حداکثر مقادیری که دامنه توزیع داده‌ها را فراهم می‌کند برای مدل مشاهده شده و پیشنهادی تقریباً یکسان است. باتوجه به نتایج، هرچند پارامترهای ماهواره‌ای به تنهایی برآورد چندان دقیق‌تری نسبت به پارامترهای مستقل هواشناسی نداشتند، ولی با وارد کرد این پارامترها در برآورد تبخیر و تعرق مرجع نتایج قابل قبول‌تری حاصل می‌گردد که نشان از اهمیت پارامترهای تصاویر ماهواره‌ای در تخمین تبخیر و تعرق مرجع به‌خصوص در مکان‌های فاقد ایستگاه‌های هواشناسی است.




**شکل ۲-نمایش بصری مدل ها با استفاده از نمودارهای جعبه ای.**

با مقایسه کلی هر دو مدل نتیجه‌گیری می‌شود که مدل‌های جنگل تصادفی بهینه شده با الگوریتم ژنتیک (GA-RF) با دقت بیشتری تبخیر و تعرق مرجع روزانه را برآورد می‌کنند. دقیق‌ترین مدل‌ها (RF-10 و GA-RF-10) بر اساس نمودارهای پراکندگی در شکل ۵ نشان داده شده‌اند. مدل GA-RF-10 و RF-10 با دمای هوا میانگین، حداقل و حداکثر دمای هوای، سرعت باد در ارتفاع دو متری، ساعات آفتابی، حداقل و حداکثر رطوبت نسبی، میانگین رطوبت نسبی، دمای سطح زمین و شاخص سطح برگ بهترین عملکرد را دارند. علاوه براین، مقایسه نتایج بدست آمده با پژوهش لیو و همکاران (۲۰۲۲)، نشان داد که دقت مدل یادگیری افراطی ماشین بهینه شده با الگوریتم ژنتیک (GA-ELM)، در تخمین $ET_0$ نسبت به مدل منفرد یادگیری افراطی ماشین (ELM) دقت بهتری دارد. همچنین، با مقایسه نتایج بدست آمده از پژوهش آچیت و همکاران (۲۰۲۲) که از روش GA-RF برای مدل‌سازی دوز منعقد کننده در تصفیه خانه‌های آب استفاده شده بود با نتایج این تحقیق همسو می‌باشد.




**شکل ۳- نمودارهای مقادیر مشاهده شده و تخمین شده با مدل GA-RF-10 و RF-10**

## نتیجه‌گیری کلی

تخمین تبخیر وتعرق مرجع ($ET_0$) امری مهم در فعالیت‌های کشاورزی است. از طرفی دیگر، داده‌های استخراج شده از تصاویر ماهواره‌ای فرصت‌های جدیدی را در کشاورزی، بوم شناسی، اپیدمیولوژی و سایر زمینه ها ارائه می دهند. در این خصوص، دمای سطح زمین (LST) و شاخص سطح برگ (LAI) پارامترهای مهم مرتبط در فرآیند تبخیر و تعرق هستند. در این پژوهش، تخمین $ET_0$ روزانه برای زمین‌های زراعی تبریز بر اساس مدل‌های RF و GA-RF، با استفاده از تصاویر ماهواره‌ای و داده‌های ایستگاه هواشناسی ارائه شدند. به منظور توسعه مدل‌های مذکور، از داده‌های سال‌های ۱۳۹٦-۱۳۸۱ (۷۰٪) برای آموزش و داده‌های سال‌های ۱٤۰۰-۱۳۹۷(۳۰٪) برای آزمایش مورد استفاده قرار گرفتند. در این مطالعه از پارامترهای سنجنده مادیس برای مدل‌سازی استفاده گردید. ماهواره‌های آکوا (MYD11A1) و ترا (MOD11A1) برای دانلود دمای سطح زمین استفاده شدند. شاخص سطح برگ از داده‌های مادیس(MCD15A3H) نسخه ٦.۱ محاسبه شد. پارامترهای هواشناسی مورد استفاده شامل دمای هوای حداکثر و حداقل($T_{max}$,$T_{min}$) ، دمای میانگین (T)، سرعت باد در ارتفاع دو متری ($U_2$)، رطوبت نسبی میانگین (RH)، رطوبت نسبی حداکثر و حداقل ($RH_{max}$, $RH_{min}$) و ساعات آفتابی (n) می‌باشند. روش استاندارد فائو-پنمن-مونتیث برای محاسبه تبخیر و تعرق مرجع روزانه به‌عنوان روش مبنا مد نظر قرار گرفت. مجموعه پارامترهای ورودی مدل، براساس همبستگی متقابل پارامترها با تبخیر و تعرق مرجع بدست آمده از معادله فائو-پنمن-مونتیث تقسیم‌بندی شدند. نتایج نشان داد که در بین پارامترهای ورودی به‌ترتیب، دمای هوای میانگین بیشترین ($R^2$=۰/۹۰۳) و سرعت باد، کمترین ($R^2$=۰/۲۸۲) همبستگی را با $ET_0$ دارند. همچنین در بین پارامترهای حاصل از ماهواره مادیس، دمای سطح زمین ($R^2$=۰/۸٦۲)، همبستگی بیشتری نسبت به شاخص سطح برگ ($R^2$=۰/۵۲۷) داشت. با توجه به نتایج به دست آمده از تحقیق حاضر، مدل GA-RF-10 ($R^2$=۰/۹۷٦ ، RMSE=۰/۲۰۰ ، MAPE=۱۱/۳۷۳ و MBE=۰/۰۲۷)، که شامل همه پارامترهای ورودی است؛ بهترین عملکرد را در بین سایر مدل‌ها داشته است. با توجه به نتایج، هرچند پارامترهای ماهواره‌ای به تنهایی برآوردهای ضعیف‌تری نسبت به پارامترهای مستقل هواشناسی داشتند، ولی با وارد کرد این پارامترها در برآورد تبخیر و تعرق مرجع، نتایج دقیق‌تری حاصل می‌گردد که نشان از اهمیت پارامترهای تصاویر ماهواره ای در تخمین تبخیر و تعرق مرجع به‌خصوص در مکان‌های فاقد ایستگاه‌های هواشناسی است.

## منابع مورد استفاده


Achite M, Samadianfard S, Elshaboury N and Sharafi M, 2022. Modeling and optimization of coagulant dosage in water treatment plants using hybridized random forest model with genetic algorithm optimization. Environment, Development and Sustainability 40: 1-19.

Allen G, Pereira S, Raes D and Smith M, 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome.

Breiman L, 2001. Random forests. Machine learning 45: 5-32.

Chia MY, Huang YF, Koo CH and Fung KF, 2020. Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques-a review. Agronomy 101:1-10

Douna V, Barraza V, Grings F, Huete A, Restrepo N and Beringer J, 2021. Towards a remote sensing data based evapotranspiration estimation in Northern Australia using a simple random forest approach. Journal of Arid Environments 191:10-13.

Hillel D and Hatfield JL, 2005. Encyclopedia of Soils in the Environment, Elsevier Amsterdam, The

Netherlands.

Holland JH, 1992. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, MIT Press.

Holmes J, 1984. Measuring evapotranspiration by hydrological methods. Agricultural Water Management 8: 29-40.

Huang G, Wu L, Ma X, Zhang W, Fan J, Yu X, Zeng W and Zhou H, 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. Journal of Hydrology 574: 1029-1041.

Liu Q, Wu Z, Cui N, Zhang W, Wang Y, Hu X, Gong D and Zheng S, 2022. Genetic Algorithm-Optimized Extreme Learning Machine Model for Estimating Daily Reference Evapotranspiration in Southwest China. Atmosphere 13: 971.

Manikumari N, Murugappan A and Vinodhini G, 2017. Time series forecasting of daily reference evapotranspiration by neural network ensemble learning for irrigation system. IOP Conference Series: Earth and Environmental Science. IOP Publishing. 17-18 March, Thanjavur, India.

Maroufpoor S, Bozorg-Haddad O and Maroufpoor E, 2020. Reference evapotranspiration estimating based on optimal input combination and hybrid artificial intelligent model: Hybridization of artificial neural network with grey wolf optimizer algorithm. Journal of Hydrology 588: 50-60.

Moser J and Suarez F, 2021. Actual evapotranspiration estimates in arid cold regions using Machine Learning Algorithms with *in situ* and remote sensing data. Water 136: 80-70.

Reyes Rojas LA, Moletto-Lobos I, Corradini F, Mattar C, Fuster R and Escobar C, 2021. Determining actual evapotranspiration based on machine learning and sinusoidal approaches applied to thermal high-resolution remote sensing imagery in a semi-arid ecosystem. Remote Sensing 13(20): 4105.

Rodell M, Houser P, Jambor U, Gottschalck J, Mitchell K, Meng CJ, Arsenault K, Cosgrove B, Radakovich J and Bosilovich M, 2004. The global land data assimilation system. Bulletin of the American Meteorological Society 85: 381-394.

Samadianfard S, Kargar K, Shadkani S, Hashemi S, Abbaspour A and Safari MJ, 2021. Hybrid models for suspended sediment prediction: Optimized random forest and multi-layer perceptron through genetic algorithm and stochastic gradient descent methods. Neural Computing and Applications 80: 1-19

Sattari MT, Fezie H, Samadianfard S, Falsafian K and Sal Wana E, 2021. Estimation of monthly and seasonal precipitation: A comparative study using data-driven methods versus hybrid approach. Measurement 173: 85-96.

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 910:15-36.

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: 220-230.

Yang Y, Shang S and Jiang L, 2012. Remote sensing temporal and spatial patterns of evapotranspiration and the responses to water management in a large irrigation district of North China. Agricultural and Forest Meteorology 164: 112-122.

Ye X, Dong LA and Ma D, 2018. Loan evaluation in P2P lending based on random forest optimized by genetic algorithm with profit score. Electronic Commerce Research and Applications 32: 23-36.